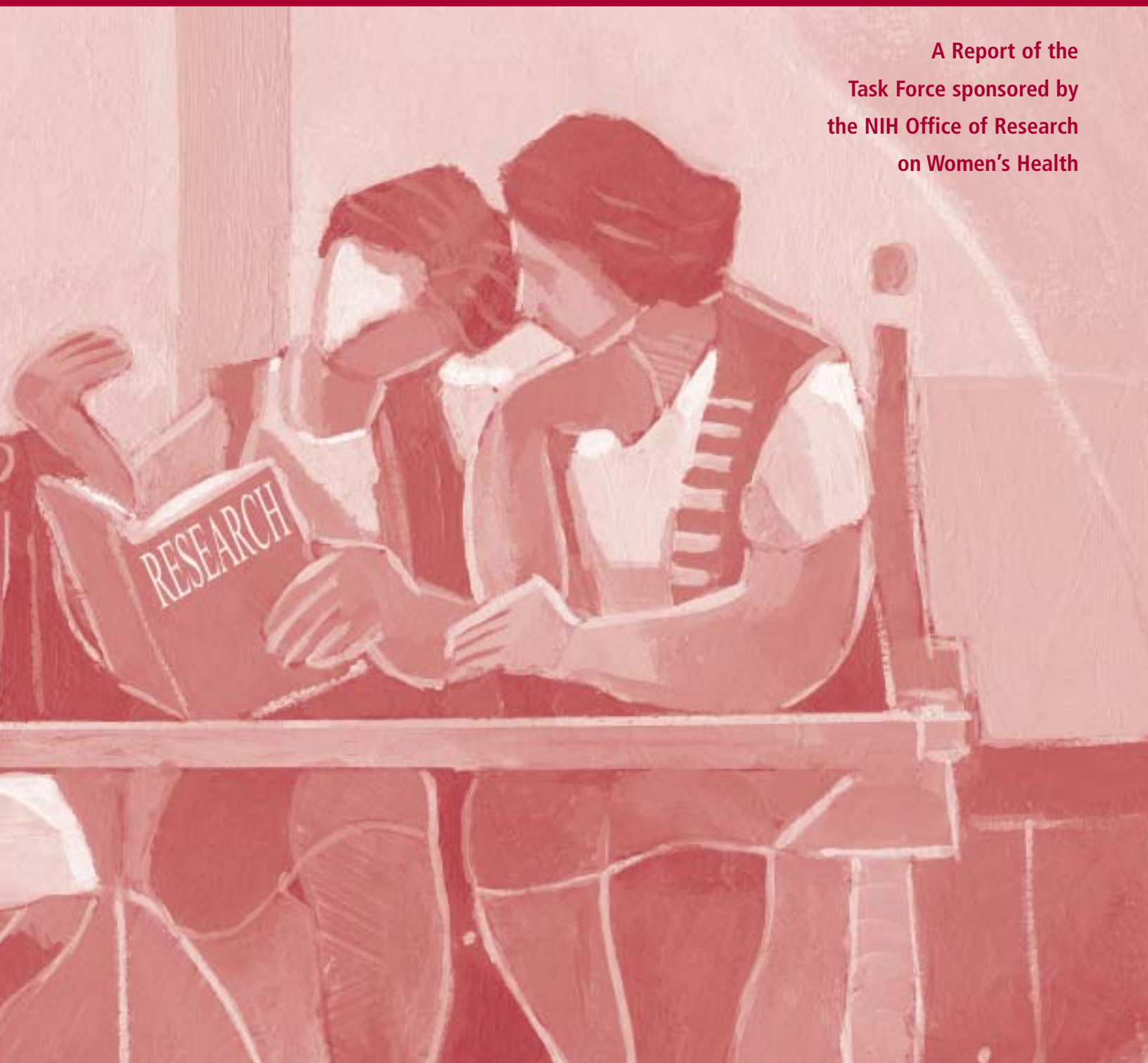


SCIENCE MEETS REALITY:
Recruitment and Retention of Women in Clinical Studies,
and the Critical Role of Relevance

JANUARY 6 – 9, 2003

A Report of the
Task Force sponsored by
the NIH Office of Research
on Women's Health



Contents

1	Preface <i>Vivian W. Pinn, M.D.</i> Associate Director for Research on Women's Health, National Institutes of Health
3	Introduction <i>Viviane W. Pinn, M.D.</i>
7	Introduction from the Cochairs <i>Otis W. Brawley, M.D., and Julie E. Buring, Sc.D.</i>
	Public Statement and Keynote Addresses
9	Public Statement – A Perspective on the Relevance of Women's Health Research <i>Lillian Thiemann</i>
11	Keynote Address 1 – Reality Strikes: Thinking About the Ethics of Inclusion in Clinical Research <i>Jeremy Sugarman, M.D., M.P.H., M.A.</i>
16	Keynote Address 2 – Challenging Sameness: Women in Clinical Trials <i>Bernadine Healy, M.D.</i>
19	Keynote Address 3 – From Research to Clinical Practice: When Have Enough Questions Been Answered? <i>Judy Norsigian</i>
24	Special Invited Address – Inclusion of Women in Research at the NIH: A Look Back and a View to the Future <i>Ruth L. Kirschstein, M.D.</i>
26	Keynote Address 4 – Point and Counterpoint: The Critical Role of Relevance <i>Raynard S. Kington, M.D., Ph.D., and Lewis H. Kuller, M.D., Dr.P.H.</i>

Presentations/Chapters

- 39 **Chapter 1**
**Lessons Learned: The Importance of Recruitment and Imperative
for Retention in Clinical Trials**
Moderators: Marianne J. Legato, M.D., and Joan McGowan, Ph.D.
- 40 Enhancing Recruitment of Women Into AIDS Clinical Trials: Lessons Learned
Susan E. Cohn, M.D., M.P.H.
- 43 Recruitment and Retention: Lessons Learned From the Women's Health Initiative
Marian C. Limacher, M.D.
- 46 Recruitment and Retention of African Americans in Clinical Studies:
Fiction, Myths, and Reality
Kwame Osei, M.D.
- 49 Recruitment and Retention in American Indian/Alaska Native Communities
Barbara V. Howard, Ph.D.
- 51 **Chapter 2**
**Lessons Learned: The Importance of Recruitment and Imperative
for Retention in Prevention Trials and Longitudinal Studies**
Moderator: Amelie G. Ramirez, Dr.P.H.
- 52 Assessing the Building Blocks: Lessons Learned From Recruitment
and Retention in Cancer Prevention Trials
Worta McCaskill-Stevens, M.D.
- 56 Life's Transitions Studies: Establishing and Maintaining
Longitudinal Cohorts With High Respondent Burden
MaryFran R. Sowers, Ph.D.
- 59 **Chapter 3**
**Nuts and Bolts of Sex and Gender Analyses:
Where Sex and Gender Really Matter**
Moderator: Janet Wittes, Ph.D.
- 60 Why Sex Doesn't Matter
Steven Piantadosi, M.D., Ph.D.
- 62 Good for Men, Good for Women, Bad for People
Barry S. Kramer, M.D., M.P.H.
- 64 Gender Analysis in the Regulatory Setting
Karen Weiss, M.D.

67	Chapter 4 Informed Consent: The New Landscape <i>Moderator: Joan Rachlin, J.D., M.P.H.</i>
68	Human Subjects Review Panels and Informed Consent <i>Greg Koski, M.D., Ph.D.</i>
71	Protecting Vulnerable Populations <i>Paula A. Johnson, M.D., M.P.H.</i>
73	Inclusion of Pregnant Women in Clinical Trials: A Case Study <i>D. Heather Watts, M.D.</i>
77	Chapter 5 Recruitment and Retention of Women and Minorities in Clinical Studies <i>Moderators: Edith Mitchell, M.D., and Jackson T. Wright, Jr., M.D., Ph.D.</i>
78	Issues and Strategies in Recruiting Minority Women Into Clinical Studies <i>Diane R. Brown, Ph.D.</i>
81	Recruiting Minority Women Into Clinical Research Trials: The Intersection of Race, Ethnicity, Gender, Community, and Academia <i>JudyAnn Bigby, M.D.</i>
85	Chapter 6 Inclusion: Policy, Ethics, or Justice—Is It Time To Reexamine the Guidelines? <i>Moderator: Darlene Yee-Melichar, Ed.D., CHES</i>
86	Policy on Inclusion of Women in Clinical Studies <i>Ruth J. Katz, J.D., M.P.H.</i>
88	Ethical Dilemmas of Inclusion in Clinical Research <i>Anne Drapkin Lysterly, M.D.</i>
91	Justice in Research <i>Jeffrey P. Kahn, Ph.D., M.P.H.</i>
93	Comments From a Member of a Feminist IRB <i>R. Alta Charo, J.D.</i>
97	Chapter 7 Pharmacogenomics and Implications for Clinical Research: Is It Time for a Paradigm Shift? <i>Moderator: Rosalie Sagraves, Pharm.D., FCCP</i>
98	The Future of Clinical Research: Genomics, Proteomics, and Other "Omics" Meet Medicine <i>Lucio Miele, M.D., Ph.D.</i>

104	From Pharmacokinetics and Dynamics to Pharmacogenomics: Relevance to Women in Clinical Research <i>Leslie Z. Benet, Ph.D.</i>
107	Gender, Race, and Regulation: An FDA Perspective <i>Susan F. Wood, Ph.D.</i>
111	Chapter 8 Recruitment, Retention, and Relevance: Continuing Challenges <i>Moderator: Susan E. Cohn, M.D., M.P.H.</i>
112	Recruitment and Retention in the Real World—Thinking Out of the Box <i>Victoria A. Cargill, M.D., M.S.C.E.</i>
115	Lesbian, Gay, Bisexual, Transgendered, and Queer Populations: Challenges in Recruitment, Retention, and Relevance <i>Sallyann M. Bowman, M.D.</i>
118	Adolescents: Challenges in Recruitment, Relevance, and Research With Youth <i>Jonathan D. Klein, M.D., M.P.H.</i>
121	Rural Populations: Challenges in Recruitment, Retention, and Relevance <i>Kathryn Whetten-Goldstein, Ph.D., M.P.H.</i>
125	Chapter 9 Community-Based Participatory Research <i>Moderators: Cynthia A. Pearson and Glorian Sorenson, Ph.D.</i>
126	Communities as Essential Partners in Research: As Participants and Subjects <i>Karen M. Emmons, Ph.D.</i>
129	Community-Based Participatory Research: Principles, Strategies, and Lessons Learned From the Community Action Against Asthma Project in Detroit, Michigan <i>Barbara A. Israel, Dr.P.H., M.P.H., and Wilma Brakefield-Caldwell, B.S.N.</i>
134	Research and the Community: Shifting the Paradigm <i>Byllye Y. Avery, M.Ed.</i>
137	Chapter 10 Recruitment Is Only Half the Challenge: The Essentials of Retention <i>Moderator: Marian C. Limacher, M.D.</i>
138	The Importance of Retention in Clinical Studies—A Statistical Perspective <i>Daniel Scharfstein, Sc.D.</i>
141	Retention Strategies: Lessons From the Women's Health Initiative <i>Judith Ockene, Ph.D.</i>

- 143 Essentials of Retention: Studies of HIV-Infected Women
 Ruth M. Greenblatt, M.D.
- 148 Retention Strategies:
 Retention of African Americans in Efficacy and Effectiveness Trials
 Shiriki K. Kumanyika, Ph.D., M.P.H.

Concurrent Expert Panels

- 151 **Panel I**
 Mistrust, Skepticism, and Bias in Recruitment and Retention:
 Reality and the Five Core Elements of Outreach From the ORWH Outreach
 Notebook
 Moderator: J. Taylor Harden, Ph.D., R.N., FAAN
- 157 **Panel II**
 The Inclusion of Vulnerable Populations With Questionable
 Capacity To Consent in Clinical Research: Ethical Issues and Approaches
 Moderators: Mary C. Blehar, Ph.D., and David Shore, M.D.
- 164 **Panel III**
 Recruitment and Retention of Children in Clinical Research: Science Meets Reality
 Moderator: George Giacoia, M.D.
- 177 **Panel IV**
 Observational Studies versus Clinical Trials—Why Are the Results Inconsistent?
 Moderator: Barbara Alving, M.D.
- 184 **Panel V**
 Recruitment and Retention of Participants in Drug Abuse Research:
 Incentives, Ethics, and Practical Considerations
 Moderator: Cora Lee Wetherington, Ph.D.

Appendix

- 191 **Task Force List**

Preface

Vivian W. Pinn, M.D.

*Associate Director for Research on Women's Health,
National Institutes of Health*

The Office of Research on Women's Health (ORWH) of the National Institutes of Health (NIH), U.S. Department of Health and Human Services (DHHS), and the Task Force on the Recruitment and Retention of Women in Clinical Research are pleased to present the results of the scientific workshop "Science Meets Reality: Recruitment and Retention of Women in Clinical Studies, and the Critical Role of Relevance," held on January 6-9, 2003. This title was selected after much deliberation, as this was not to be just another conference on 'the inclusion of women in clinical trials'!

The ORWH was established in 1990 to catalyze efforts related to women's health research at the NIH. The ORWH has given the following mission, which has defined its activities over the years of its existence:

The Office of Research on Women's Health is under the direction of a Director who: (a) advises the NIH Director and staff on matters relating to research on women's health; (b) strengthens and enhances research related to diseases, disorders, and conditions that affect women; (c) ensures that research conducted and supported by NIH adequately addresses issues regarding women's health; (d) ensures that women are appropriately represented in biomedical and biobehavioral research studies supported by the NIH; (e) develops opportunities for and supports recruitment, retention, re-entry, and advancement of women in biomedical careers; and (f) supports research on women's health issues.

During the years since the establishment of the ORWH, there has been a significant increase in the attention and resources devoted to women's health and women's health research. One of the major areas of progress has been an increased focus on research studies that include women as

participants and a greater ability to monitor the inclusion of women as study participants in NIH-supported clinical trials. That progress and the challenges that remain constitute the theme for this meeting and report.

The establishment and implementation of policies for the inclusion of women and minorities in clinical research funded by the NIH have their origins in the women's health movement. Following issuance of the report of the Public Health Service Task Force on Women's Health in 1985, the NIH established a policy in 1986 for the inclusion of women in clinical research. This policy, which *urged* the inclusion of women, was first published in the *NIH Guide to Grants and Contracts* in 1987. Minority and other scientists at the NIH recognized the need to address the inclusion of minority populations. Therefore, in a later 1987 version of the *NIH Guide*, a policy *encouraging* the inclusion of minorities in clinical studies was first published. In July 1989 an *NIH Memorandum on Inclusion* stated that research solicitations should encourage inclusion of women and minorities and require a rationale if excluded, and that executive secretaries of scientific review groups should ensure that responsiveness to policy would be addressed and indicated in summary statements.

In 1990 the Congressional Caucus for Women's Issues requested that the U.S. General Accounting Office (GAO) conduct an investigation into the implementation of the Inclusion Guidelines for women by the NIH. This report, delivered in congressional testimony, indicated that implementation of the policy for inclusion of women was slow and not well communicated, that gender analysis was not implemented, and that the impact of this policy could not be determined. The GAO testimony also indicated that there were differences in implementation of the policy recommending the inclusion of minorities and that not all NIH Institutes and Centers (ICs) factored adherence to

these policies into scientific merit review.

In order to ensure that the policies for inclusion were firmly implemented by the NIH, Congress made into public law what had previously been NIH policy, through a section in the NIH Revitalization Act of 1993 (Public Law 103-43), *Women and Minorities as Subjects in Clinical Research*. In 1994 the NIH revised its inclusion policy in response to this law to meet the mandate that women and minorities be included in all of its clinical research studies. The revised Inclusion Guidelines were published in the *Federal Register* in March 1994 and became effective in September 1994. The Revitalization Act essentially reinforced the existing NIH policies, but with four major differences, that

- The NIH ensure that women and minorities and their subpopulations be included in all NIH-supported clinical research
- Women and minorities and their subpopulations be included in Phase III clinical trials in numbers adequate to allow for valid analyses of differences in intervention effect
- Cost is not allowed as an acceptable reason for excluding these groups
- The NIH initiate programs and support for outreach efforts to recruit and retain women and minorities and their subpopulations as participants in clinical studies

The result was that the NIH could not support and would not fund any grant, cooperative agreement, or contract, nor would it support any intramural project, to be conducted or funded in Fiscal Year 1995 and thereafter, that did not comply with this policy, unless a clear and compelling rationale and justification establish that the inclusion is inappropriate with respect to the health or purposes of the research. Research awards covered by this policy require the grantee to report annually on the enrollment of women and men and on the race and ethnicity of research participants so that accrual can be monitored.

Although much progress has been achieved, many issues related to inclusion still need to be addressed, especially those related to the retention of women in clinical studies and the recruitment of populations of women who have been considered difficult to recruit into clinical research. Following the example of the 1995 ORWH Task Force and Workshop on the Recruitment and Retention of Women in Clinical Trials, which had issued a report focusing on innovative and successful methods identified in the mid-1990s to ensure the recruitment and retention of women in clinical studies, the ORWH

established a new Task Force in 2002. With the guidance of this new Task Force, a plan was implemented for a scientific workshop to reexamine the topic of inclusion of women in clinical research. The Task Force identified three overarching goals, to

1. Examine lessons learned concerning the recruitment and retention of women and other participants in clinical prevention and treatment trials, as well as longitudinal cohort studies, conducted over the past decade
2. Address the continuing challenges confronting the recruitment and retention of women, minorities, and other participants to ensure that clinical research is representative, relevant, and targeted to address scientific questions important to the public health.
3. Note the emerging ethical and policy issues that present both challenges and opportunities for women's health research as well as studies that will elucidate sex and gender factors in health and disease.

The Task Force for this effort is cochaired by Otis W. Brawley, M.D., and Julie E. Buring, Sc.D., with the assistance of Margaret A. Chesney, Ph.D., a visiting scientist in the ORWH serving as Conference Coordinator, along with Ms. Joyce Rudick, also of ORWH as Conference Liaison. All of the members of the Task Force have contributed greatly to the success of this endeavor (see Appendix A for a list of Task Force members).

Efforts are moving forward based on what has been learned during the years since the new policies and procedures have been in place. These efforts are needed to ensure that women and minorities are appropriately represented in clinical research funded by the NIH. This report addresses the issues, successful strategies, innovations, and recommendations that were discussed during the workshop.

The ORWH, the NIH, and the members of the Task Force believe that collaboration is critical. As we move forward, we continue to invite the participation and guidance of researchers, study participants, ethicists, clinicians, women's health advocates, and policymakers in this ongoing process and dialog. We welcome and appreciate the ongoing involvement and interest of all those concerned with improving the health and well-being of women through research.

Vivian W. Pinn, M.D.

*Associate Director for Research on Women's Health
Director, Office of Research on Women's Health
National Institutes of Health*

Introduction

Vivian W. Pinn, M.D.

*Associate Director for Research on Women's Health,
National Institutes of Health*

Strategies to ensure uniform implementation of the revised *NIH Policy and Guidelines on the Inclusion of Women and Minorities as Subjects in Clinical Research (NIH Inclusion Guidelines)* across the NIH were developed through the establishment and deliberations of an NIH Tracking/Inclusion Committee, which is made up of representatives from each of the NIH's Institutes and Centers (ICs). This trans-NIH committee is convened by the Office of Research on Women's Health (ORWH) and cochaired by the Director of the ORWH and a senior Committee official, Carl A. Roth, Ph.D., National Heart, Lung, and Blood Institute. It meets on a regular basis, focusing on consistent and widespread adherence to the *NIH Inclusion Guidelines* by all the ICs. Working in collaboration with the Office of Extramural Research, the Office of Intramural Research, and other components of the NIH, the ORWH coordinates the activity of developing and establishing data collection and reporting methodologies to ensure uniform standards and definitions in the reporting of data on women and minority participants in NIH-funded clinical research.

Each year the ORWH and the NIH Tracking/Inclusion Committee issue a report on aggregate data from across the NIH of the numbers of women and men included in clinical trials. Prior to the passage of the NIH Revitalization Act of 1993 (Public Law 103-43) and before implementation of the current system of collecting data (which started with FY 1995), there was no NIH-wide system for collecting data and monitoring the numbers of women and men in clinical studies; therefore, there is no documentation of the numbers of women and men in clinical studies for those years. The annual reports for each year since FY 1995, *Monitoring Adherence to the NIH Policy on the Inclusion of Women and Minorities in Clinical Research*, include complete

histories from the implementation of the NIH Revitalization Act and related issues on inclusion to the present, how NIH has enhanced its efforts to ensure that sex/gender analysis is performed, and the most recent data about women and minorities in clinical studies.

The most recent aggregate data available indicate that many more women participate [in clinical trials] than men.

The most recent aggregate data available, for women and men in all human subject studies funded by the NIH in FY 2000, indicate that many more women participate than men. Approximately 61 percent of clinical research participants were women compared with 38 percent male participants; in Phase III clinical trials, the numbers are even more striking—about 71 percent women versus about 29 percent men. However, when one considers that there are studies addressing those areas of research in which women have not been included before, research on the many gender-specific areas affecting women's health, and studies that include both women and men, the seeming disproportionate numbers of women in clinical studies are better understood. Furthermore, when sex-specific studies are eliminated, the participation of women and men in clinical studies is about equal: 50.2 percent women versus 49.3 percent men.

Following a congressional request for an assessment of the NIH's progress in implementing the 1994 guidelines on including women in clinical research, the U.S. General Accounting Office (GAO) issued another report in May 2000, *Women's Health: NIH Has Increased Its Efforts to Include Women in Research*. The report concludes that, in the past decade, the NIH has made significant progress

in implementing a strengthened policy on including women in clinical research. In addition to updating the *NIH Inclusion Guidelines*, the NIH has conducted extensive training for scientists and reviewers; furthermore, the issue of inclusion of women and minorities is a matter of scientific merit in the review process, thus affecting a proposal's eligibility for funding. Extramural and intramural research programs implement the inclusion policy, and a centralized inclusion data tracking system has been developed to monitor policy implementation.

Although the GAO report applauded the NIH's efforts in furthering women's health issues, it also included two specific recommendations to the Director of NIH to ensure the following, that

- The requirement be implemented that Phase III clinical trials be designed and carried out to allow for the valid analysis of differences between women and men and communicate this requirement to applicants as well as requiring peer review groups to determine whether each proposed Phase III clinical trial is required to have such a study design and that summary statements document the decision of the initial reviewers
- The NIH staff members who transmit data to the inclusion tracking data system receive ongoing training on the requirements and purpose of the system

Immediately following the release of this report, many activities were undertaken to reexamine the NIH's system for tracking data on the inclusion of women and minorities in clinical research, recommend any necessary changes to improve its accuracy and performance, and reiterate the NIH policy related to analyses by sex/gender.

A variety of outreach activities have been initiated to explain the revised policy to the scientific research community.

Several actions resulted that were designed to clarify the requirement that NIH-defined Phase III clinical trials include women and minority groups, if scientifically appropriate, and that analysis of sex/gender and/or racial/ethnic differences be planned and conducted by investigators engaged in NIH-funded research. Specific actions included updating the *NIH Inclusion Guidelines*, revising the *Application for a DHHS Public Health Service Grant* (PHS 398), and creating guidelines for reviewers and review administrators to address and document adherence to the *NIH Inclusion Guidelines*.

- *The NIH Policy and Guidelines on the Inclusion of Women and Minorities as Subjects in Clinical Research and Amended Notice to the Guide for Grants and Contracts* was updated and posted on the ORWH Web site and the *NIH Inclusion Guidelines* Web page. These documents supersede previous iterations of the *NIH Inclusion Guidelines* as published in the 1994 *Federal Register* notice, clarify the definition of clinical research as reported in the 1997 Report of the NIH Director's Panel on Clinical Research, and incorporate the Office of Management and Budget Directive 15, which defines racial and ethnic categories to be used when reporting population data. The updated versions also restate that NIH-defined Phase III clinical trials must be designed and conducted in a manner sufficient to allow for a valid analysis of differences in intervention effect related to sex/gender and/or race/ethnicity.
- The NIH issued the newly revised *Application for a DHHS Public Health Service Grant* (PHS 398, rev. 5/01). The instructions in this document describe the requirements for designing Phase III clinical trials to provide valid analysis by sex/gender and/or race/ethnicity.
- *Guidelines and Instructions for Reviewers and Scientific Review Administrators* was developed to emphasize and clarify the need to review NIH-defined Phase III clinical trials for both inclusion requirements and analyses by sex/gender and/or race/ethnicity. These instructions reiterated the need for proper documentation in the summary statement.

A variety of outreach activities have been initiated to explain the revised policy to the scientific research community and to clarify common misunderstandings about the revised requirements. The NIH prepared a document designed to help researchers comply with the revised *NIH Inclusion Guidelines*, *Outreach Notebook for the NIH Guidelines on Inclusion of Women and Minorities as Subjects in Clinical Research*, which outlines elements of outreach processes, offers practical suggestions, and provides references to additional sources of information. The *Outreach Notebook* is available on the ORWH Web site <<http://www4.od.nih.gov/orwh/outreach.pdf>>. The *Outreach Notebook* Committee worked long and hard, spending more than a year updating this document. Dr. J. Taylor Harden, NIA, and Dr. Mary C. Blehar, NIMH, have served as coauthors of the Outreach Notebook Committee, along with Angela Bates, ORWH; Gladys Melendez, National Institute of Neurological Disorders and Stroke; Dr.

Otis W. Brawley, formerly of the NIH and now of Emory University; Dr. Joan McGowan, National Institute of Arthritis and Musculoskeletal and Skin Diseases; and Dr. Cora Lee Wetherington, National Institute on Drug Abuse. The ORWH has also just released the second edition of the *Women of Color Health Data Book*, which was produced to provide data for those who question the need to examine subpopulations, to look at different minority groups, or to gather data on women.

At the opening session of “Science Meets Reality: Recruitment and Retention of Women in Clinical Studies, and the Critical Role of Relevance,” participants were reminded of the critical role of inclusion in increasing knowledge about the contributions of sex differences and/or similarities to the health and disorders of women and men and minorities. They examined the lessons learned to date concerning the recruitment and retention of women and other participants from clinical prevention and treatment trials, as well as longitudinal cohort studies conducted over the past decade. During the

course of this 4-day meeting, participants identified and discussed the challenges that continue to confront researchers as they endeavor to recruit and retain women, minorities, and other participants in clinical research to ensure that study populations are representative, relevant, and appropriate for addressing scientific questions important to the public health.

Emerging ethical and policy issues present challenges and opportunities for women’s health research.

Workshop participants also considered carefully and at length the emerging ethical and policy issues that present both challenges and opportunities for women’s health research and studies that will elucidate sex and gender factors in health and disease. This report presents a summary of the many hours of thought, collegial debate, and discussion devoted to these issues.

Introduction from the Co Chairs

Otis W. Brawley, M.D., and Julie E. Buring, Sc.D.

It was just 10 years ago that the NIH Office of Research on Women's Health formed a Task Force on the Recruitment and Retention of Women in Clinical Studies. For those of us who were part of that Task Force and who participated in the subsequent public hearing and scientific meeting held in 1993, it was an exciting and exhilarating time, as we shared our own experiences – successful and not – and realized that we were not alone in our efforts. Based on the statements and presentations at that 1993 meeting, the resulting report clearly supported our belief that not only was a much broader inclusion of women in clinical studies scientifically necessary, but that this goal, despite early concerns, was achievable.

In late 2001, the Task Force began planning again, for the meeting in January 2003 entitled “Science Meets Reality: Recruitment and Retention of Women in Clinical Studies, and the Critical Role of Relevance.” This publication is the proceedings from that meeting. Many of the members of this latest Task Force had also served on the first one. What became clear to us was how far we had come in 10 years, and that the innovative and successful methods of the mid-1990s had resulted in great strides in ensuring the recruitment of women into clinical studies. However, it was also clear that it was time to meet together again as a scientific community, so that we could move into the next stage, build on the successes already achieved in the last decade, and address crucial continuing challenges and emerging issues.

Thus, the Task Force agreed that the major goals of this conference should be three-fold: to review our “lessons learned”; to address the continuing challenges confronting the recruitment and retention of women, minorities, and special populations; and to consider the impact and strategies for addressing the emerging ethical

and policy issues regarding research.

A number of important themes emerged from this meeting. With respect to structural or conceptual issues, we learned that:

- Lessons from previous recruitment issues have shown that it is necessary to separate conceptual problems from structural problems and structural solutions.
- Community involvement at every level is necessary, and the culture of the community must be acknowledged.
- Successful studies ask questions of the study cohorts and the community and then *listen to what they say*.
- Incentives such as reimbursement for time, training, continuation of healthcare after the study has ended, and respect for the monetary constraints on vulnerable populations are important, but with the understanding that they must not be coercive.
- Issues of trust between the medical research establishment and vulnerable populations need to be heeded and overcome.

Furthermore, with respect to the epidemiologic issues that inform the design and conduct of these studies, it is clear that:

- Differences between observational studies and clinical trials need to be taken into account in the study design.
- A continually finer sieve of the population is needed. For example, Chinese women born and raised in the United States may be a different population than Chinese women who have emigrated from China: they may live in different communities, go to different churches, and have different early childhood experiences.
- There need to be accurate and generalizable definitions of the population being studied, within the study and

across studies. For example, is a population defined by race, poverty, affluence, or sexual orientation?

- Conflict could exist between the issues of equal access to healthcare and drug trials and the scientific mandate to learn about sex-based differences of these conditions and treatments.
- Sex and race may be surrogates for socioeconomic conditions.

What became clear to us was how far we had come in 10 years, and that it was time to meet together again as a scientific community.

With respect to the crucial issues of ethics, we learned that:

- There has been a steady progression from the ethics of protectionism to the ethics of inclusion to a notion of justice.
- Previously understood differences between clinical care and clinical research have been blurred.

- Terminology may be an important consideration. For example, members of a cohort may perceive “experimentation” differently than “research,” and a “subject” differently than a “patient.”
- Necessary incentives can become problematic lures, such as reimbursement for time, training, and continuation of health care after the study is over.
- Given the current U.S. health care system, does participation in clinical trials substitute for universal healthcare for some populations? What is the role of the researcher when full medical care is not provided?
- Once the study is over, what responsibility, if any, does the research team have to former participants if they have provided the best or only healthcare an underserved population has ever received?
- The thin line between profiling vulnerable populations and women and focusing on providing good healthcare has to be considered explicitly.

What became crystal clear from this meeting is the balance that must exist between good science and reality. The sharing today of our lessons learned will continue to form the foundation of our continued successes during the next decade.

Public Statement and Keynote Addresses

Public Statement:

A Perspective on the Relevance of Women's Health Research

Lillian Thiemann

Ms. Thiemann is Research Workgroup Cochair of the Women's HIV Collaborative of New York and Co-Founder and President of Visionary Health Concepts, a community-driven medical education company specializing in HIV and hepatitis. She was diagnosed with HIV and hepatitis C in 1989 and is a former intravenous drug user. Ms. Thiemann achieved national recognition with her column "Dope Fiend's Corner" for POZ magazine. The popular patient and secondary provider programs she has written and designed include "Double Jeopardy: The HIV/HCV Co-Infection Handbook" and "The Future of HIV Treatment." Her newest educational programs are targeted to physicians and include the continuing medical education-accredited monograph "Explorations in Care: Metabolic Abnormalities in HIV" and "Vital Lines: Management of Clinical Challenges in HIV/HCV Co-Infection." She has been a participant in the New York University cohort of the ACTG 035 study for 8 years.

I am an HIV-positive woman from New York City, and I am here to give you my experience with clinical trials in human immunodeficiency virus (HIV) and to talk to you about their comparative relevance in the pre-HAART and post-HAART eras. (For those of you not working in the HIV field, "HAART" is "highly active antiretroviral therapy.")

Relevance was a much clearer issue in the pre-HAART era than it is today. I was a member of the People With AIDS (PWA) Health Group and was referred there through my case manager at a drug treatment program because I really wanted to attend a group that talked about HIV treatment and women. There was nothing like that happening anywhere except at the PWA Health Group. I joined the Women's Treatment Project, where a continual parade of M.D.s, R.N.s, dietitians, and AIDS activists would talk to a diverse group of about 12 women, some fresh out of prison and some like me, who knew nothing about HIV. Most researchers did not know much about HIV at that time, either.

The PWA Health Group tried its best to answer some of the problems of AIDS by importing drugs from other countries that were not yet approved in the United States. The PWA Health Group accomplished this with the help of U.S. physicians who would write prescriptions for these drugs so that people who represented our group could travel to other countries and obtain the treatments and bring them back into the United States. We did this so that people who were

suffering with opportunistic infections and side effects of the approved AIDS drugs could get some relief from those side effects.

In 1995, 6 years after my HIV diagnosis, Phase III trials for protease inhibitors began enrolling research participants (Roche Labs' saquinavir and Merck's indinavir trials). At that time, if you wanted to participate in a protease inhibitor trial, you put your name in a lottery; thousands of people entered their names (or their doctors did), and names were literally picked out of a hat. For the indinavir trials, a waiting list was compiled.

The longer I was on the trial the more I knew how extremely important it was for women to be enrolled in clinical trials.

So when, at the behest of my health care provider, I called the AIDS Clinical Trial Group (ACTG) at New York University to request enrollment in the indinavir trial, I had five T-cells and a very high HIV viral load, and although I had never suffered a major opportunistic infection (just minor ones like thrush), the researchers and I both knew that I was on the skids and starting to "circle the drain." So I called the study nurse and requested to enroll in the trial. Her response was, "I'm really sorry, but it's fully enrolled." Not to be deterred, I asked if my name could be put on a waiting list. Her response was, "Sure; you are number 2,347." It was a

major psychological setback to hear that number, but I really needed to participate in that trial, so I told her what a fabulous study participant I would be: “I will show up for every appointment, take every test, and stick with it to the very end.” She said, “That’s very nice, but you’re still number 2,347.”

I called her the following week and the week after that; I think I called her twice a week for a month or so. By the fifth or sixth time I called, I only had to say hello, and she would recognize my voice; she would say, “Hello, Lillian; you’re still number 2,347. Thank you for calling.” Finally, I think I had tortured her enough because she said, “Listen, Lillian, I understand you would be a great study participant, that you’ll come to every appointment, take every drug, do every test, stick with it to the end. I have that, but please stop calling.” It was difficult for her every time she had to talk to me. I would like to think that I was not begging, but I am sure that there was an edge in my voice that was slightly like that.

As time went on, I did stop calling, and I accepted the fact that I was going to get sick and die, like many of the people I watched around me. Acceptance is a beautiful thing. Two weeks later my phone rang, and I heard a woman’s voice with a Filipino accent, asking for me; it was Candy, the study nurse. (Her name is Candida, and for those of you who work in HIV, it is an inside joke.) She said, “Lillian, are you still interested in being in the trial?” I said I was. She said, “We’ve had many dropouts. The drugs have side effects. We had a meeting, and we were discussing who we could get who would show up for every appointment, who would take every drug, and who would hang in until the end. And, magically, your name popped into my head!”

Well, I have been a participant in the ACTG 035 study for almost 8 years. Every year, for my work, I go to the Conference on Retroviruses and Opportunistic Infections to listen to the latest data and to get ideas for educational programs. Every time, although it is old hat now, Trip Gulick, the primary investigator for ACTG 035, trots out the latest data. I sit there, and I am proud to hear those data.

When I first entered the study, it was extremely relevant for me because I believed I was going to die. As time went on and more drugs became approved, many people were exiting the trial that I was on, but I wanted to stay because it was still relevant to me—I received very good health care on my clinical trial. They looked at me very closely, and we formed a relationship over time that was very important to me. As time went on, the relevance as far as getting free drugs was no longer there. At the time I was on Medicaid, and I could get anything that was approved in the United States. Yet the longer I

was on the trial and the longer I consorted with HIV treatment activists, the more I knew how extremely important it was for women to be enrolled in clinical trials. I wanted those long-term data to be available; even if not enough women were in my particular cohort to make it statistically significant, I felt that some really important data might come out over time. I stayed, on a dinosaur of a regimen; HIV drug regimens are not like that anymore, and the trial is now being stopped.

What I now understand through my membership at the Women’s HIV Collaborative of New York is that activists did a really wonderful job, with the help of well-placed medical activists at the National Institutes of Health and other agencies in the United States forcing the issue, so that the restrictions on women in clinical trials and the exclusion of women from clinical trials would go away. For the most part, women are now welcomed in trials in which they can enroll. However, women are not showing up. It is not news that women and men are different, and we must all understand that we need to tell the clinical trial story from the “female experience” point of view.

This year when I heard that this conference was going to occur, I took time out of my schedule because I felt it was so important to have a presence here and to see what you all are doing. I also want to find out about other disease states and, for the people who are still working in HIV, where we can make a difference. I want to gather information here to do what I do best, which is produce educational programs to help health care providers and patients learn what they need to in order to move along a continuum that brings everyone’s health up a notch.

I hope I can generate some interest in producing education based on the knowledge that I gain here from all of you, and I thank you for your kind attention and for all your work.

Keynote Address 1

Reality Strikes: Thinking About the Ethics of Inclusion in Clinical Research

Jeremy Sugarman, M.D., M.P.H., M.A.

Dr. Sugarman is Professor of Medicine and Philosophy at Duke University and Founding Director of the Center for the Study of Medical Ethics and Humanities in the Duke University School of Medicine. He has served as a senior policy and research analyst for the Advisory Committee on Human Radiation Experiments and as a consultant to the National Bioethics Advisory Commission. Dr. Sugarman conducts theoretical and empirical research in medical ethics, concentrating on informed consent, research ethics, and ethical issues associated with emerging technologies.

About a month ago, I was leading a plenary session at the Public Responsibility in Medicine and Research meeting, and Marian Gray Secundy, Ph.D., was on the panel. Marian was a leader in medical ethics who had worked at Howard University and then was Founding Director of the National Center for Bioethics in Research and Healthcare at Tuskegee University. I had referred casually to the Tuskegee syphilis study, and Marian, in her resonant voice, turned around to me and said, “I wish you’d get that right. It was the ‘U.S. Public Health Service Tuskegee Study of Untreated Syphilis in the Negro Male’.” She was right! I am dedicating this talk to Marian.

Today I will briefly highlight the early history of subject selection, and then I will discuss the concept of justice, initially describing how justice was conceived as a need for protection in research that shifts to involve claims about access to research. I will then try to make sense of this shift and begin to address some critical questions related to it: Where are we going conceptually, now that we have had experience dealing with problems related to inclusion as well as protections? What are the implications for research?

The early history of subject selection began in a series of scandals. Carol Levine, M.A., summed up research ethics, as being “born in scandal and reared in protectionism.” The international documents that in part drive our understanding of subject selection came out of events like the Nazi experiments, but they also came out of domestic events.

In the Willowbrook State School in New York, a school for (then-termed) “mentally retarded children,” the noted, well-published researcher Saul Krugman, M.D. wanted to study hepatitis. Dr. Krugman knew that all of the children in the Willowbrook State School eventually became infected with hepatitis. To develop treatments for hepatitis, Dr. Krugman reasoned that if the children were going to become infected anyway while in the institution, why not inject them with hepatitis virus as they were admitted and then learn from them? After all, they were in a very controlled environment. So, Dr. Krugman inoculated and injected the children with the stools of infected children to induce hepatitis.

Willowbrook was a doubly troublesome case because the school was full; there were no beds. Services for retarded children in the 1960s were terrible, and they are

The early history of subject selection began in a series of scandals-events like the Nazi experiments but also domestic events.

not great today. If parents wanted their child admitted to the Willowbrook State School, they almost had to sign up for research. Was consent obtained? Sort of. Was it voluntary? Perhaps not. It was a population that was quite vulnerable. Unlike much of the research of its day, this research was prospectively reviewed by its funding agency, the March of Dimes. It was not secret research; it was published research. Dr. Krugman's work was important and is still cited in textbooks on hepatology. Nonetheless, people were alarmed when they learned of such practices.

A second case involved research on whether cancer is infectious, conducted by the Jewish Chronic Disease Hospital in New York, then an affiliate of Memorial Sloan-Kettering Cancer Center. To answer this question, doctors injected live cancer cells into elderly patients without their consent. This was brought to public attention, and people were alarmed about what had happened to researchers—those folks in the white coats, the people they trusted like Marcus Welby. Doctors were undermining the trust in the doctor-patient relationship by doing things in the name of research that were not acceptable.

The third and most egregious case was the so-called Tuskegee syphilis study in Macon County, Alabama. The research was conducted to study the natural history of syphilis among African Americans. Why go to Macon County, Alabama? Why not New York or Washington or Boston or Chicago? Was the decision made to go to a very poor, rural community because people might not ask as many questions? The study began in a place that was poor, although it the research question was not specific to poor people. As Pat King puts it, this is a question of justice. Men were given lumbar punctures to look for the natural history of syphilis, to see whether syphilis had gone into their brains. The spinal taps were performed only for the purposes of research, yet these men were

Researchers and clinicians have a responsibility to look out for the best interests of their clients. The concept goes beyond "do no harm" to "help people if possible."

told that they were receiving "back shots," and they thought these shots were therapeutic and might prevent the spread of syphilis to their brains.

The study began in 1932 and continued well after penicillin had been discovered and was shown to be uniformly effective in treating syphilis. The men in the

study were told that they were not to walk across the street to the health department and receive penicillin. If they did so, they would lose the single benefit they were given to participate in the research—a funeral benefit of \$50 that their family members likely could not have otherwise afforded. A proper funeral, or "homecoming," which would have been completely consistent with what would have been expected or desired, would have bankrupted their families.

This study continued until 1972 when an investigative reporter brought national attention to it and an ad hoc committee was created by the Federal Government to evaluate it. Subsequently, the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (NCPHSBBR) was established by the National Research Act (Public Law 93-348). In its Belmont Report: *Ethical Principles and Guidelines for the Protection of Human Subjects of Research*, the NCPHSBBR outlined the ethical justification for the basic regulatory structure for the protection of human subjects that is still used in large part today. The report outlined the use of three ethical principles that should be applied in understanding the ethics of research: respect for persons, beneficence, and autonomy.

Respect for persons is founded on the right of liberty, the right to be left alone. Almost as a rule, Americans do not like to be touched. Tax us, and we throw your tea overboard. Americans do not like to be touched in person or politically. The notion of not being touched is also relevant in health care: "Don't do that surgery on me unless I give you permission," "Don't use my medical information unless I give you permission," and "Don't do research on me unless I give you permission."

Beneficence refers to favor or benefit, the risks and benefits of research: minimized risks and maximized benefits. Researchers and clinicians have a responsibility to look out for the best interests of their clients. The concept goes beyond "do no harm" to "help people if possible."

Justice is fairness and distribution. The Commission talks about individual justice in the selection of subjects, which would require that researchers exhibit fairness. Thus, they should not offer potentially beneficial research to only some patients favored by the researchers or select only desirable persons for risky research. Social justice received less attention at the time of the NCPHSBBR, and it requires that a distinction be drawn between classes of subjects that ought and ought not to participate in a particular kind of research, based on the ability of members of that class to bear burdens and on the appropriateness of placing additional burdens on already-burdened persons. As suggested in the work of John Rawls, there is an obligation to protect the least well-off members of society.

AIDS activists wanted the experimental drugs and were willing to take responsibility for taking on the risk of doing so. The appeals were so powerful that the FDA created two different tracks to change its system for drug approval.

With such conceptions of justice being prominent, it is not surprising that protective regulations were put in place that include having institutional review boards critique research proposals and special regulations to protect those who are especially vulnerable. For example, although children are vulnerable and it is important to protect them, one of the drawbacks of protecting children in the context of research has meant that many drugs used in children have not been tested in children, and so clinicians and parents have to guess at the effectiveness of these drugs in children. In addition, the only available pill size may be too large a dose for a child.

The Commission determined that prisoners are vulnerable because they are not positioned to make a voluntary choice. Nevertheless, it turns out that the top guys in prisons got to be the subjects in experiments in the prisons. It broke the monotony, and they welcomed it.

Pregnant women are considered a vulnerable population, but that is somewhat puzzling. Why are there special rules about pregnant women? What do they need to be protected from? It might have been a political agenda to protect the fetus, but it was couched in language about women. Pregnant women take on average between four and six drugs during pregnancy. Virtually none of those drugs has been tested during pregnancy, and so the desire to protect the fetus or the pregnant woman has led to the same problem we have with children—not having good information about the safety of drugs used in pregnancy.

The shifting claims about justice came into play in large part after the onset of the AIDS epidemic. A diagnosis of human immunodeficiency virus infection or acquired immunodeficiency syndrome (AIDS) in the 1980s was a death sentence. There was nothing to do, but there were some drugs that were just beginning to be tried. AIDS activists made very powerful appeals. They were dying. They did not care about protective regulations, including those of the Food and Drug Administration (FDA), for establishing safety and efficacy through staged clinical trials. They just wanted the experimental drugs and were willing to take responsibility for taking on the risk of doing so. The appeals were so powerful that the FDA created two different tracks to

change its system for drug approval.

Cancer activists began to make similar appeals. As a result, what has occurred is a change from protection to access; the pendulum has shifted from the notion of fairness as a consideration of vulnerability and protection to that of fairness in access. For the most part, this notion informed a lot of policy changes.

What rationales have been advanced for inclusion rather than protection or access? The rationale for diseases that affect a population subgroup is that, if no research is done with children, there will be no treatments for those children. If the disease affects only pregnant women and no research is done with pregnant women, there will be no treatments for pregnant women. For diseases that affect multiple population subgroups—for example, women and men—there may be treatment differences to consider. Research on single, Caucasian men does not take into account the hormonal cycles of women, which could influence treatment outcomes among other population subgroups. Various rationales have resulted in regulatory changes.

Important claims and regulatory changes have occurred for at least four groups: women, children, minorities, and vulnerable sick people. The initial publication of the Institute of Medicine report on inclusion of women in clinical research led to monumental changes, as did the rules of the National Institutes of Health, the Centers for Disease Control and Prevention, and the FDA. The *Guidelines for the Conduct of Research Involving Human Subjects at the National Institutes of Health* ensures that women and members of minorities and their subpopulations are included in all appropriate human subjects research and also ensures the inclusion of women and minorities and their subpopulations in Phase III clinical trials so that valid analyses of differences in interventions can be accomplished. Cost cannot be an acceptable reason for excluding these groups, and outreach efforts must be made to recruit these groups into clinical studies. So, here we are years later trying to reflect on how well these rules work.

With respect to “vulnerable sick people,” in a recent publication, Amit Shah and I report on a study in which we analyzed the first four studies that have approval to

use a waiver for emergency research. We found that there were expressions of concern about trust, the ability to say no, the ability to opt out, and the extent to which the community's input would be involved in the research design. We can and should learn from such experiences.

Of course, in the swing of the "inclusion pendulum," we have not witnessed such a change with respect to prisoners and students.

What are the challenges to inclusion? To compare population subgroups within the context of the clinical trial, adequate statistical power is required. That typically means more people. More people means a bigger study, which costs more, takes longer, and can be more difficult to conduct. This presents a conflict of interest to academics because they want to publish; they want tenure. Companies want a drug approved because they

Balancing candor and hope in people who are devastatingly ill is extraordinarily difficult. The purpose of research is to meet a social end, not a personal one.

want to bring it to market. The incentive structure is such that the statistical power argument can often be waged as a way to undermine the need for inclusion. This issue demands critical attention, because you do not want to have people participate in research if the resulting data will not be usable.

Recruitment is an enormously tough issue with respect to the challenges of inclusion. In 1992 I walked around Johns Hopkins and asked the medical students if they knew about "Tuskegee." None of the students knew except the medical students of color, the technicians, and the secretaries; they knew something about Tuskegee, but they sometimes thought the issue was that the Government was injecting syphilis into African American men. That did not happen; everybody got it the old-fashioned way, but the legacy and the lack of trust had continued. Giselle Corby-Smith's article last month that included about 600 persons of color (probably the largest such study) revealed the effect of trust and the challenges of recruiting persons of color into research.

So as policy moves to inclusion, there are questions about building up trust that was undermined through bad examples. Retention in clinical trials is also an issue. If the incentive structure is to enroll people for research but not to retain them, it is no surprise that people find participation difficult and drop out. If a good job is not done on informed consent at the front end, if you do not

tell participants what research is really about and remind them along the way, there is no reason for them to stay.

How do we make sense of justice like this? The chapter by Madison Powers, J.D., D.Phil., "Theories of Justice in the Context of Research" in *Beyond Consent: Seeking Justice in Research*, presents some useful concepts about the egalitarian, libertarian, and utilitarian "spheres of justice" that relate to health care, research, and environmental policy, respectively. From the perspective of egalitarian justice, we try to think about how we can equate things and how we can give everybody the decent minimum, just enough. How can we make people equal? We come up with a complicated system in which everybody is treated as an equal. With a libertarian theory of justice in which people are allowed to make their own decisions, protective mechanisms to ensure that everything is designed to ensure that people make their own choices make sense. In contrast, in a utilitarian theory of justice, there are risk-benefit calculations that measure tradeoffs. These three notions of justice become confused in the issues of health care, research, and the environment, which are necessarily intermingled. Dr. Powers' challenge is to move from these different notions of justice to a clear understanding of what is appropriate and when.

Paul Applebaum, M.D., and his colleagues coined the concept of "therapeutic misconception" to describe the mistaken belief that what is done in the context of research is primarily about the research participant. If I enroll in research and believe mistakenly that this research is about me, that is a misconception. As we observed when working on a project for the White House Advisory Committee on Human Radiation Experiments, balancing candor and hope in people who are devastatingly ill is extraordinarily difficult. We have to emphasize to potential participants that the purpose of research is to meet a social end, not a personal one. If you can make people better in the process, that is great, but the research is designed to serve the greater good and the greater numbers, unlike regular medical care.

Now there is evidence from clinical research that people in early-phase trials, especially in oncology, confuse research and treatment. The notion of having access to clinical trials came to bear for endostatin when an article in *The New York Times* reported that endostatin cured cancer in mice. The fact that mice can be cured does not mean that people can be cured. However, at that time, a Phase I trial about the dosing and toxicity of endostatin was announced, and the MD Anderson Cancer Center received more than 2,000 telephone calls for a Phase I trial enrolling fewer than 20

people. To address the problem of whom to enroll, we evaluated why people were calling and reported this in the *Journal of Clinical Oncology* in September.

Language may be contributing to the problem. For example, we now have lots of research that involves gene transfer, popularly known as gene therapy. It is exciting science, but there is little evidence that the early-phase trials in gene transfer experiments are going to be therapeutic, so is it misleading to call it “gene therapy”? How about “novel therapeutics” or “cancer vaccines”? Who would not want a cancer vaccine? We give vaccines to kids (although vaccines are now under new scrutiny). How about a “natural therapy”? Natural things can be harmful, but people think that because something is natural, it must be better.

A related kind of misconception has to do with the meaning of the terms used to describe research. I was at a meeting at a university one day, and a bank president on a university committee turned to me and said that she was so pleased to find, when she reviewed the university’s records, that the university had conducted research but had done no experiments. When asked the difference, she explained that research is cutting edge and is what you want; an experiment is when you do not know what is going on or you are trying to figure something out.

So this became one of several domains we examined in a study in which we asked 1,882 people around the country to compare different terms used for research: medical experiment, medical research, medical study, clinical investigation, and clinical trial. In 103 indepth interviews, we asked people what they meant by the terms “research” and “experiment.” Similar to the bank president’s explanation, respondents said that research is cutting edge, the best treatment you can have; it is what you get when you are really sick. An experiment is when they cut you up and put things inside you, you are treated like a guinea pig, and the doctors do not know what is happening. A study, they said, was when the doctors and nurses get your medical records; they read about you, and they study about your disease. People did not understand clinical investigation: What went wrong? Who’s investigating?

These words are confusing. In the informed consent process, you can tell me all about the risk-benefits, alternatives to participation, whom to contact if the study goes wrong, and whom to contact about my rights as a research subject, but if you say it is a study and not research or an experiment, people may have enrolled for the wrong reason, and they may not show up again. This is a problem of recruitment and retention.

It is important to realize that justice has implications for research at every step. The justice questions begin at the funding decisions: What research is funded, and who decides what will happen? It is not just about the big important trials. Those are critical, but which projects are not being done? Which diseases are not being studied? Is it a profitable drug? Is it for some reason a symbolically important drug or disease? Notice what gets studied and notice what is not. What about research design and justice? Who gets included? When do you include community participation? How do you do that? What constitutes fairness? Industry has an interest, but when does it come to the table? When does government come to the table? Who is being left out? Recruitment: Who are the participants, and how do you pick them? Individuals in the clinic waiting room? Or the people who read something in

A paradigm shift has occurred in how the principle of justice should be applied in research.

the newspaper and fly across the country or call on a phone that autodialers until they get through? Not everybody can do that. Publication: How do the announcements and results get disseminated, and when are they disseminated? Justice issues are everywhere.

In conclusion, a paradigm shift has occurred in how the principle of justice should be applied in research. More conceptual work is needed to figure out those spheres. It is not easy, but we have made some progress. The regulations have mapped this shift, but only in part; some lingering concerns remain. To meet ethical requirements, justice must be considered at every step of the research process.

Keynote Address 2

Challenging Sameness: Women in Clinical Trials

Bernadine Healy, M.D.

Dr. Healy is a physician with a career in medical education, research, and patient care, and has led the Lerner Research Institute at the Cleveland Clinic, the NIH, the College of Medicine and Public Health of Ohio State University, and the American Red Cross. She is now a medical and health columnist and senior writer for U.S. News and World Report and serves on the President's Advisory Committee for Science and Technology.

Women carry most of the burden of chronic disease—they take most of the medications, visit doctors more, and have more surgeries than men. An anecdote: Shortly after I left the NIH, I was the fourth speaker at a large cardiovascular conference. The first speaker was presenting his clinical trial of a particular cardiovascular drug. With slides on the screen and in the darkness of the room he said, “We would have gotten this trial done a lot faster if we didn’t have to include women” and complained that this “new” rule made it harder to do his clinical research. When I got up to speak, I noted how delighted I was to hear that there was a little bit of pain going on out there because it meant that inclusion of women in clinical trials, which NIH had mandated since 1985, was finally taking hold.

Challenging Sameness

Women are different. It is not a particularly outlandish statement, but in fact, that notion challenged the common orthodoxy—the orthodoxy of sameness and the orthodoxy of the mean, which has dominated much of the thinking in medical science. In the world of physiological research, the more uniform your animal model could be, in gender and breed, the better the model was because there were fewer variables. This view often impaired our attitude toward clinical research. Reducing all humans to the proverbial 60-kilogram white male, 40 years of age, and defining that as the normative standard may make the data easier to analyze. Nonetheless, such designs

The attention to gender may have opened our eyes to broader diversity issues in research.

make it harder to extrapolate findings to the diverse population for which the trial was conducted. Even for men, “the average American male” is a fiction.

Genomics give us a window into diversity. Yes, 99.9 percent of the human genome is exactly the same among all humans, but that 0.1 percent creates powerful differences and directly relates to both health and illness. Pharmacogenomics, the new and emerging field that recognizes the importance of tailored treatments, is based on subtle genetic traits that determine personal biochemistry and drug interactions. By definition, it recognizes that no one therapy is necessarily appropriate for all. There may be no more defining genetic difference than that determined by whether one has or lacks a second X chromosome. In retrospect, the exclusion of women from clinical research and the failure to appreciate gender differences in experimental laboratory studies were blind spots that rippled through many years and billions of dollars of medical research.

Indeed, the attention to gender may have opened our eyes to broader diversity issues in research—age, for example. Both ends of the age spectrum have been challenges in biomedical research, particularly in clinical trials. Certainly it is more difficult to conduct research,

and there are bigger ethical issues in a study on a 3-year-old or a frail octogenarian. But if they are going to be the target of medical intervention, we need to gain insight directly from those populations, and from studies that are conducted by more than the National Institute on Aging or the National Institute of Child Health and Human Development. Although the lines are not so crisp, ethnic and racial differences also need to be considered. The Women's Health Initiative (WHI) was a leader in prospectively planning for diversity from the very outset, including calling for clinical sites that would provide a diverse mix of study participants. And, despite much skepticism that this too would be overly difficult, the WHI showed it could be done, and done well.

Clinical Trials as Powerful Research Tools

The importance of humans in medical research has not always been as widely appreciated as it is now or here in this room. This is in part due to the insights and capabilities gleaned from human genetics research. Discovery of human genes carries much more caché than those of *Drosophila* or mouse. And this focus only reinforces the need for direct human investigation at all levels. Years ago when my research was taking me back and forth from the bench to the bedside, the bench research using animal models was typically seen as the higher science. I remember being delighted by a comment in a British medical journal at the time that wryly noted that to understand human disease one should at the least consider data from the human. And a corollary to this: When one designs clinical trials to provide information that will be applied to humans, the human condition must be taken into account. The NIH Office of Research on Women's Health (ORWH) has been a great force in championing clinical trials and focusing on clinical trial design that takes into account half of the human condition. The growth in the number of studies focusing on women's health and the sheer number of women participants in clinical trials is a tribute to these efforts. Part of this symposium is to celebrate and reinforce these efforts.

Women's Health and Suffrage

As we pause to look closely at the great successes of the past decade with regard to women's health research, we

should not forget another human factor at work. In a larger sense these successes have been born out of the sustained efforts of women from all walks of life and over time. Just as our clinical trials must be seen in terms of those we serve—women of today and tomorrow—our ability to fund, design, and execute this growing domain of medical work in women's health should be understood in terms of the women who came before us.

The first of these started more than three generations before 1919, the year women got the right to vote. Women gained a political voice and the right to own themselves and not be owned. It came at a price. The early suffragists recognized the dilemma of women of the day. As Elizabeth Cady Stanton wrote 1868, "To keep a foothold in society, a woman must be as near like a man as possible, reflect his ideas, opinions, virtues, motives, prejudices, and vices." This notion of sameness is often the ticket to advance women's rights. In a second phase of women's suffrage, which many of us lived through in the late 1960s and early 1970s, women fought for economic independence, equal access to education, and better jobs. In so doing, they were so often up against a difficult bar—proving that they could do just as well as men in virtually all the male domains of life. This enabled them to fit in, gain access, and be accepted. Sometimes this meant turning a back on motherhood. Perhaps bra burnings by the "radicals" and male-like neckties for the "conformists" were symbols of the same mindset. So too, unfortunately, was taking up male vices that were largely unknown to women, such as smoking. Indeed, the legacy of "You've come a long way baby," the tobacco ads that linked women's freedoms to tobacco use, is now seen in the female face of the number one cancer killer of both men and women: lung cancer. Women no longer need to make that bargain.

Focusing on women's health, which spotlights how women are different, not the same, might seem contrary to those historical efforts. Actually, it was a necessary followon to them. The women's health movement, and indeed it is a movement, took to task the notion of women being just like men and, without apology, recognized that knowledge about those differences were crucial to women's well being. The knowledge gap of centuries needed correcting. The simple premise underlying women's health is that men are not the normative

The growth in the number of studies focusing on women's health and the sheer number of women participants in clinical trials is a tribute to the ORWH.

standards for behavior, or for mental or physical health. Women learned—and gradually had science to back it up—that they could acknowledge that they are different from men without giving up the rights gained.

Women's health research happened because it was right. But these and the many other women's health efforts were sustained because the time was right.

Women's Health Movement

The inclusion of women in clinical trials, the WHI, numerous fundamental studies on women's health and gender differences, the many activities of the ORWH, and the sprouting up of similar offices throughout other government health agencies were vital parts of this effort. Also important were the many private sector efforts in universities, medical centers, and industry, and among groups like the Society for Women's Health Research, the American Heart Association and the American Cancer Society. Women's health research happened because it was right. But these and the many other women's health efforts were sustained because the time was right. Groups of women who had used the gains of yesterday—political voice, education, and positions of influence in medicine, in Congress, in journalism, and in homes and communities across the land—knew it was time. The commitment to women's health was not to be a passing fancy as some predicted. And the commitment to clinical trials involving women helped solidify and sustain this focus. Surely, as but one example, the surprising early findings of the WHI with regard to hormone replacement therapy (HRT) have turned conventional wisdom on its head, reinforcing the need for more study, not less.

At the time of its inception, the WHI was called by some—and not always approvingly—the “mother of all clinical trials.” It is a delicious title, actually. The WHI, including the controlled trial and the observational arm, was the biggest trial worldwide at that time and met some surprising opposition even among women. Widely touted negative views seemed intent on killing the study. Many saw it as being too big, too expensive, too ambitious, too interdisciplinary and as testing questions that were already answered. HHS Secretary Donna Shalala was pressured to stop the trial, even after the 40 clinical sites were awarded; she had the courage not to do so. For example, one loud objection was that we already knew that HRT was good for the heart and questioned whether a placebo-controlled trial at that point was possible, necessary, or even ethical. Another was that the hypothesis that a high-fat diet was a risk factor for breast cancer was so wrong-headed it should not even be tested—not to mention that the high-fat diet was to be examined as well in terms of its risk for colon cancer and heart disease. This scientific self-assurance looks a little silly now.

We have only begun to learn from the WHI and in the near future we will see more knowledge come forth on matters such as estrogen replacement alone, diet, bone and brain health, cardiovascular health, heart failure, cancer risks, psychosocial variables, aging in the second half of life, and the genetic predisposition for health and disease. The observational part of the study should emerge as a Framingham kind of treasure trove of information and will set the stage for clinical and basic research for generations ahead.

The relatively new medical research focus on women that we are examining here, which includes the creation of clinical trials focusing on women's unique health concerns combined with the inclusion of women in all clinical trials when appropriate, is well past the tipping point. Participation of women will not be an issue again.

Keynote Address 3

From Research to Clinical Practice: When Have Enough Questions Been Answered?

Judy Norsigian

Ms. Norsigian is a founding member of the Boston Women's Health Book Collective, where she served as a board member from 1971 until 1998; she is currently its Executive Director. She is coauthor of Our Bodies, Ourselves, first published in 1970 and now in its seventh edition as Our Bodies, Ourselves for the New Century. She has been active in the areas of sexually transmitted diseases, contraceptive research, misleading advertising of prescription drugs, tobacco use by girls and women, menopausal hormones, and the impact of genetic technologies on women.

It is a pleasure to join you today and to report that, after the October hormone meeting, several of the principal investigators from the WHI have agreed to work with us on the eighth edition of *Our Bodies, Ourselves*, which will be published in 2005. We intend to showcase the WHI, the findings that we will be able to report on by then, and also the history of this whole effort.

This morning during the panels, I was reminded of my late husband's research back in the 1960s. As a medical sociologist, Irving Kenneth Zola studied interactions between doctors and patients at Massachusetts General Hospital and noted significant problems when there was a radical difference between the background of the caregiver and the background of the patient or client (in terms of race, ethnicity, class, and sex). Many of you are now confronting these issues as recruitment goes forward in your own research.

As we take a closer look at how best to recruit, retain, and otherwise involve women in clinical studies, it also presents a good opportunity to step back and reflect on some of our past experiences and lessons learned, for example, when research was conducted poorly or not at all. As long-time women's health educators and activists, those of us at *Our Bodies, Ourselves* have frequently advocated for more and better research relevant to women's health and medical concerns. We have been especially pleased by the WHI and similar efforts that have contributed to a knowledge base that will ultimately improve decisionmaking for both laypersons and health care professionals.

Among the more controversial issues we have confronted during the past 30 years is the undue influence exerted by ideological or corporate interests on the research enterprise and cases where research findings were not appropriately translated into practice. In more recent years, misleading media reports have exacerbated these problems, which is why some of you, if you are not already actively dealing with media reports of research findings, might think about doing so at least in your own community.

Here are some examples from the anecdotal archives of women's health advocates like ourselves, and by keeping these stories in mind, we will be better able to avoid past mistakes.

- Early on, oral contraceptive research containing high dosages of estrogen was conducted in a rather controversial fashion, though it was not considered controversial at that time. Some of you are probably familiar with the infamous quotes of Joseph Goldzieher, M.D., made known to the rest of us by Barbara Seaman's investigative writings in the 1970s. He said something like, "We can't tell Mrs. X that these pills might give her a headache, because then she'll come around the next day and report that she's got a headache." Condescending, paternalistic, and even racist attitudes toward women, especially poor women and women of color, were uncovered.
- Research on the Dalkon Shield intrauterine device showed problems early on with the Nylon-6 used in

the tail string. It was already known from the surgical literature that it would decompose inside the human body, and researchers noted their concern about infection well before marketing of the product. This was a case in which the desire to reap profits from launching a new product on the market took precedence over clear concerns about safety. Quite a bit of literature exists on this episode, since the consequences were substantial—many women

The WHI and similar efforts have contributed to a knowledge base that will improve decisionmaking for laypersons and health care professionals.

experienced serious morbidity, and others died.

- Early published studies of the internal fetal heart monitor were written by physicians with a financial interest in the two major companies marketing these machines. Even by standards at that time, it was premature, on the basis of relatively small studies, to advocate routine use of internal fetal heart monitoring, but an aggressive campaign convinced many obstetricians to employ this equipment regularly. Once Albert Haverkamp and others conducted the more adequate large-scale studies that demonstrated lack of benefits for the vast majority of infants in terms of outcomes, it was extremely difficult to change what had become the standard of care. Those physicians and midwives who knew that ordinary fetal auscultation would be perfectly adequate—and sometimes preferable—often thought that they had to use the internal monitors in case a bad outcome might be blamed on their failure to use this equipment.
- Some of the recent media reports regarding cesarean sections and vaginal births after cesarean (VBACs) are instructive here. Data from the *New England Journal of Medicine* study and other recent research regarding cesareans and VBACs are markedly different from messages in the mainstream media. I am hoping that an American College of Nurse Midwives' op-ed piece on this topic will get published in a major publication in the next week or so. I also hope that the Boston University School of Public Health's Maternal and Child Health Department will publicize its analyses of media reports regarding what is now a clearly unacceptably high cesarean section rate in this country: close to 25 percent and climbing. The suggestion that this is a good

thing boggles the mind when you look closely at all the literature and the evidence.

- Quinacrine, a drug once used to treat malaria, was discovered to be a sclerosing agent and was initially tested as a means of chemical sterilization via insertion as a slurry into the fallopian tubes. Subsequently, studies tested it in the form of pellets, often with questionable results, even in terms of efficacy. Because the basic animal toxicology research on this drug is only now being completed by Family Health International, few researchers have promoted its use in clinical trials in advance of knowing these toxicology findings. Family Health International is also conducting followup research in Vietnam, where thousands of women underwent quinacrine sterilizations that were reported on in the *Lancet* in the early 1990s. In this instance, the population-control zealotry of two individuals—now we are talking about ideology—well funded by private interests, has made it possible for quinacrine to be promoted in a number of less industrialized countries in starkly unethical ways, well in advance of the completion of well-designed research that might demonstrate its safety as well as its efficacy.

To its credit, the U.S. Food and Drug Administration (FDA) has stopped the illegal manufacture and distribution of quinacrine kits in the United States. However, already tens of thousands of women in other countries have been sterilized with quinacrine, as chronicled in an award-winning story in *The Wall Street Journal* in June 1998. When Shree Mulay, Ph.D., a researcher from Montreal, presented on quinacrine before the Planned Parenthood Federation of America medical advisory committee in 1999, she left them with a very important comment: "Science and ethics cannot be placed in two separate compartments."

- The case of breast implants demonstrates how difficult it can be to encourage the conduct of quality research once a product is marketed widely. For more than 20 years, the FDA did not regulate either silicone or saline breast implants, despite repeated calls to do so by women's health advocates, who claimed that informed choices were not possible for women considering implants. Because of the insistence of such groups as the Command Trust Network, the National Women's Health Network, my own organization, and a congressional investigative committee, as well as the principled stance taken by then-FDA Commissioner Dr. David Kessler, silicone implants were limited to clinical trial use with mechanisms established to collect more meaningful

safety data. Unfortunately, the types of symptoms and problems experienced by women with silicone implants have not lent themselves to clear clinical diagnosis. This difficulty with nomenclature has made it that much easier to discredit both the women who have suffered from implants as well as the physicians who recognize that implants have caused serious problems in many women. As better studies are now being done, we are learning, for example, that women who suffered silicone leakage from silicone implants have been at higher risk of developing fibromyalgia, and a recent National Cancer Institute (NCI) study has shown elevated risks of certain cancers.

Now to bring us up to the present. A relatively new science dedicated to the biology of female sexual dysfunction (FSD) is gaining more media attention, especially since a 1999 *Journal of the American Medical Association* study that found 43 percent of women in this country report sexual problems. Although there certainly are women with underlying physiological problems that might respond to drug interventions like Viagra, it is also clear that undue emphasis on a more medicalized and reductionist approach to women's sexual difficulties could do us all a great disservice. So often women's sexual difficulties are the result of lack of knowledge about female sexuality, inadequate communication with partners, unresponsive partners, and other social conditions, rather than, for example, impaired blood flow to the genitals.

This new and growing emphasis on more narrow biomedical solutions may well obscure the findings of earlier research conducted by sociologists, sex therapists, and others. Some women's health advocates and researchers are raising questions about the way in which pharmaceutical companies may be contributing to a distortion of the research process. As Sandra Leiblum, Ph.D., Director of the Center for Sexual and Marital Health at Robert Wood Johnson Medical School, has said, "It is misguided to think a significant number of female sexual dysfunctions are organic." A strongly worded critique of this trend by Ray Moynihan in the January 4, 2002, issue of the *British Medical Journal* has already generated a range of interesting replies; it is available

online for those of you interested in following this debate. The excellent book *A New View of Women's Sexual Problems*, edited by Ellyn Kaschak and Leonore Tiefer, also offers some fascinating discussions.

Menopausal hormone therapy is a good example of how clinical practice can take off in the wrong direction before important questions are thoroughly studied and answered. We are doing that now, and this matter has been the subject of much debate. I highly recommend viewing Susan Dentzer's slide presentation, given at the NIH-sponsored "Scientific Workshop on Menopausal Hormone Therapy," October 23-24, 2002, and also at the Women's Health Summit, November 6, 2002, held by the American Association of Health Plans. If you have not heard her speak, it is well worth getting this presentation.

The use of tamoxifen for risk reduction of breast cancer was yet another case of premature promotion of a drug therapy that subsequently was found to be far less useful than the glowing claims in front-page news stories during spring 1998 would have had us believe. In August 1999, the NCI published a reanalysis of these data and suggested that the benefits were far fewer than previously publicized. This is instructive, because the media did not pay much attention to this very important analysis done by the NCI, and most of the American public was still left with the impressions from April 1998.

Finally, the field of genetics may become the most challenging, as we envision new kinds of clinical trials involving women. As we debate the myriad issues surrounding adult and embryonic stem cell research, embryo cloning for research versus embryo cloning for purposes of human reproduction, human germ line genetic modifications, and other ethically and scientifically complex matters, it will be critical to take the time to involve our friends, families, and communities in these discussions. Already there has been so much confusion in the public's mind from media reports that obfuscate important distinctions or reports that seem to promise embryo stem cell therapies in the very near future, we need to take the time necessary to sort out what we do and do not know.

Throughout the debates about cloning, it will be critical for scientists to articulate why it is totally

Menopausal hormone therapy is a good example of how clinical practice can take off in the wrong direction before important questions are thoroughly studied and answered.

unethical to conduct reproductive cloning experiments on humans. First, there is no ethical way to “get there” from here, given the human experimentation on women and children that would be required. Here I want to credit Rudolph Jaenisch, M.D., and other colleagues in the field who have taken the time to talk to the media about how unethical it would be to conduct this research on human beings.

All the talk about how human reproductive cloning might be acceptable—if we could demonstrate that it can be done safely—is simply absurd. I am particularly troubled that spokespersons for the American Society for Reproductive Medicine could not recently state clearly that they oppose human reproductive cloning but instead suggested that it would be acceptable if it could be proven safe. Equivocation from reputable physicians will only make it more difficult to establish the important distinctions that must be better understood by the public. Stem cell researchers are rightfully concerned that they will be wrongfully lumped in with those trying to create genetic duplicate humans, if we are not forthright about the unacceptability of human reproductive cloning. Furthermore, much of the public continues to confuse embryo cloning—somatic cell nuclear transfer—with embryo stem cell research involving embryos that would otherwise be discarded from in vitro fertilization (IVF) clinics. Researchers need to address this kind of confusion whenever possible.

Recent surveys have shown that women are already less likely to support research involving germ line genetic modifications. Why might it be that women would draw the line at clear medical therapies and say no to genetic modifications that have more to do with enhancements? One of the problems with embryo cloning for research purposes, not reproductive purposes, is that it is the gateway technology to designer babies. Might women be more concerned that there will be no way to limit the development of this technology to purely therapeutic purposes?

One particular concern for women that has received minimal attention in the past year and a half of public debate about embryo cloning is the lack of good-quality, long-term safety data on the drugs that both suppress and hyperstimulate the ovaries. Although women now use these drugs in IVF clinics and some of them sign consent forms that indicate numerous serious risks, even the small risk of death, these women in IVF clinics may realize a benefit that they assess as offsetting any of these risks—that is, the possibility of a successful pregnancy and birth.

For many of us advocates, exposing healthy women to the risks of egg harvesting solely for the purposes of

research cloning is unethical, especially because we are so far from demonstrating that viable therapies are likely from embryo stem cell research. Just look at the many current hurdles that face embryo stem cell researchers. Problems with both the inability to control differentiation of the cells and tumorigenicity need to be worked out. In fact, thus far there has been more progress with fetal tissue research than with embryo stem cell research. At the very least, we need to more closely analyze the several thousand adverse reports to the FDA that have been associated with leuprolide acetate (Lupron, TAP Pharmaceutical Products Inc.), some of which are quite serious and need much closer scrutiny.

During the early years of ORWH’s existence, we joined with many others in calling for a greater role for laypersons in the design and conduct of research. Now we can point to many cases in which the active role of, for example, women who have had breast cancer or a person living with AIDS has made a positive difference in the deliberations about what kinds of questions to ask, how to ask them, and what specific measures would improve the outcome of a clinical trial. Some institutions—like the Seattle-based Program for Appropriate Technology in Health (PATH), which has involved women in developing an improved diaphragm—have reported their experiences to share with others. Similarly, the Population Council, Family Health International, and others have conducted a range of studies in the reproductive health field that have involved laywomen in some very productive ways.

Some current microbicide research is also a good example of incorporating the perspectives of laypersons. Pennsylvania-based ECRI involved lay health advocates in its excellent patient reference guide on high-dosage chemotherapy with autologous bone marrow transplantation for metastatic breast cancer. Updated as recently as 1999 and available on the Web, this resource helped women understand why it was so problematic that some clinicians were discouraging women from being randomized into an important NCI study looking at the effectiveness of ABMT plus high-dosage chemotherapy versus high-dosage chemotherapy alone. It took longer to get the answer from this research because it took longer to get women to enroll.

At the National Institutes of Health a number of meetings and conferences, especially those organized by the ORWH, have included laypersons with a special interest in the topic being addressed, and this has almost always enriched the discussion and debate. Because we see the positive effects from including laypersons, it was disappointing last week to see only doctors, scientists, and professional ethicists named to the Secretary’s

Advisory Committee on Human Research Protection. There are very good people on this Committee, but this offered an opportunity to further involve a knowledgeable layperson with a patient/advocate perspective.

Ultimately, we want to make sure that research is relevant and that findings can be translated into accessible therapies for all who stand to benefit. The World Health Organization's 2002 *World Health Report: Reducing Risks, Promoting Healthy Life* underscores the major worldwide risk factors to health: malnutrition, access to safe water and sanitation, high blood pressure, high cholesterol concentrations, unsafe sex, tobacco use, and alcohol consumption. It is wise to keep these issues front and center as we decide on research priorities. We want to make sure that this research is primarily done in the service of alleviating human suffering and disease—not what some people are now referring to as “boutique medicine”—and both chronic and acute conditions. We also want to ensure that resources are devoted to solutions that will be accessible to more than an elite few. We need more research on the direct consequences of racism and on the health of women of color. This research is equally, if not more, important than research on new drugs, devices, and biologics. However, as we all know, we may have to work harder to secure funding for this kind of research.

What can you do, those of you who are researchers? Certainly be on top of the media, and care about accountability and issues related to public disclosure and potential conflicts of interest. An excellent discussion of how conflict of interest can muddy the research waters and especially the public's understanding of research findings can be found in the January 2003 issue of *The Washington Monthly*—an article called “Hot Flash, Cold Cash.” It is instructive and well worth reading.

All this is not to say that we should not have drug companies funding research. Pharmaceutical company-sponsored research can be as valuable as research funded by other means. However, we have extensive documentation of some of the problems that have resulted from industry-sponsored research, and we have to be watchful that findings from all research are reported responsibly in the media. When there are misleading public relations campaigns, those of us who are advocates must ask that there be corrections and that another point of view be offered.

As an example, I want to offer a wonderful letter that Diana Zuckerman, Ph.D., sent to ABC's “Good Morning America” about a month ago in response to a fairly misleading piece they did on breast implants. She is with the National Center for Policy Research for Women &

Families here in Washington, D.C., and formerly served as staffer to the congressional committee that looked more closely at the breast implant issue. Her letter, sent to Nancy Snyderman, M.D., at “Good Morning America,” is a good model for anyone who wants to think about how best to “correct the record.”

Ultimately, we want to make sure that research is relevant and that findings can be translated into accessible therapies for all who stand to benefit.

I encourage all of you to keep in mind whatever activism and advocacy in which you can participate. Pick up the literature for some of these groups and do not think that there must be a constant wall between the advocacy arena and the research arena. We can and we have to work together, so that good-quality, ethical research is conducted, the reporting of it is accurate, and we implement the findings in a way that make sense.

(On the literature table I left materials from the Council for Responsible Genetics, the Center for Genetics and Society, and a new coalition called Prevention First Incorporated, because I think their materials embody excellent examples of activists, advocates, and scientists working together to raise critical issues for the public.)

Special Invited Address

Inclusion of Women in Research at the NIH: A Look Back and a View to the Future

Ruth L. Kirschstein, M.D.

Dr. Kirschstein is Deputy Director of the National Institutes of Health (NIH) and has served twice as the NIH's Acting Director. From 1957 to 1972 she performed research in experimental pathology at the U.S. Food and Drug Administration (FDA) and helped develop and refine tests to ensure the safety of viral vaccines for such diseases as polio, measles, and rubella; her work on polio led to selection of the Sabin vaccine for public use. Dr. Kirschstein served as Deputy Associate Commissioner for Science at the FDA, and in 1974 she was named Director of the National Institute of General Medical Sciences at the NIH, a position she held for 14 years. Dr. Kirschstein has received numerous awards and honorary degrees for her research and her achievements.

As Deputy Director of the NIH—and as the first Acting Director of the NIH's Office of Research on Women's Health (ORWH), the first office within the Federal Government dedicated to improving the health of women—I am delighted to participate in this scientific conference “Science Meets Reality: The Recruitment and Retention of Women in Clinical Studies, and the Critical Role of Relevance.”

All who are active in the movement to improve women's health must acknowledge the foresight of Dr. Edward N. Brandt, Jr., who established the U.S. Public Health Service (PHS) Task Force on Women's Health in 1983, sparking a new level of Federal commitment to addressing women's health issues. As cochair of the U.S. Department of Health and Human Services (DHHS) Coordinating Committee on Women's Health Issues from 1983 to 1995, I had the privilege of participating in the formulation of policies and programs that embody the Federal Government's commitment to improving the health of women. Nowhere has that commitment been more enthusiastically embraced and sustained than at the NIH.

With the creation of the ORWH in September 1990, the NIH committed itself to fulfilling a promise made to American women and their families—a promise to provide the scientific knowledge needed to improve the health, prolong the lives, and enhance the quality of life of all Americans, regardless of race, creed, age, geographic locality, or sex. Ten years ago, NIH Acting Director Dr. William F. Raub and other NIH leaders recognized that, to increase our understanding

of women's health, the NIH needed to address women's health issues in a more comprehensive and coherent way than had been possible in the past. The ORWH is the embodiment of that recognition and commitment.

Today, we examine the tangible achievements in women's health and the inclusion of women in clinical research made possible by the foresight and vision of Dr. Brandt and Dr. Raub and sustained over the past decade by individuals within the executive branch, including former DHHS Secretary Dr. Donna Shalala; individuals within the PHS such as Dr. James O. Mason, Assistant Secretary for Health, and Dr. Audrey Manley, former Deputy Assistant Secretary for Health; members of the bipartisan Congressional Caucus for Women's Issues; and the ongoing commitment and advocacy of individuals and organizations within the private sector.

For the past 12 years, the ORWH has been an effective and successful catalyst for women's health research. During the ORWH's first decade, we have seen studies related to women's health integrated into the research portfolios of NIH Institutes and Centers to a remarkable degree. This speaks strongly to the effectiveness of the Office under the able leadership of Dr. Vivian Pinn, who became the first full-time Director of the ORWH in November 1991. Under Dr. Pinn's leadership, the ORWH has worked to foster, develop, and increase basic and clinical research on diseases and conditions that affect women; determine gaps in the medical community's knowledge of such conditions and diseases; and identify areas of great scientific promise and pressing public health need.

Building on the work undertaken by the Task Force on Opportunities for Research on Women's Health at Hunt Valley, Maryland, in September 1991, in the past 3 years, the ORWH has identified and assessed the enormous advances in basic and clinical science knowledge and linked them to a research agenda targeted to improve women's health. As a result of all these efforts of the past 12 years, there is now widespread, and largely unquestioned, recognition that researchers and clinicians must understand how differences in sex and cultural, ethnic, and socioeconomic background may influence the causes, diagnoses, rates of progression, and treatments of diseases. This sea change in perceptions and outlook is truly remarkable, and it is occurring at a time of unparalleled opportunity for women and for biomedical science.

At this moment in history, we perceive the light of a new dawn in medicine: Never before have the life sciences held such tremendous promise for alleviating human suffering. From advances in genetics to improvements in understanding the influence of behavior on health, we stand now at the threshold of boundless scientific opportunity and medical promise. Researchers on the NIH campus, along with NIH-supported investigators in every State, are working to ensure that the promise and potential of science are realized through new treatments and cures for a wide array of diseases and conditions that afflict humankind. At the same time, women are playing increasingly significant and visible roles in our national life—as elected officials and as members of professions traditionally dominated by men. Never before has it been more important to safeguard our Nation's health and prosperity by preserving and protecting the health of women and their families.

Working in partnership with other components of the NIH, the ORWH sponsors research aimed at developing strategies to encourage women to protect their health through improved nutrition, exercise, and other lifestyle changes. These studies reach out to women in the communities where they live and work. They help women help themselves by adapting the fundamentals of health promotion and disease prevention for women of diverse cultures, ethnic backgrounds, geographic settings, and socioeconomic circumstances.

Ensuring the inclusion of women of diverse ages, ethnic backgrounds, and socioeconomic circumstances in clinical studies is a crucial part of keeping our promise to America's women. During the past 12 years, the ORWH, in concert with other components of the NIH, has worked to fulfill that promise by implementing the NIH's revised and strengthened guidelines on the inclusion of women and minorities in clinical studies and through a new computerized tracking system to monitor the inclusion of women in NIH-supported

research. The ultimate goal of inclusion is to raise the general level of health enjoyed by all women by addressing the health issues of many diverse groups of women.

It is to all American women that the NIH is pledged and committed. Keeping our promise of ensuring that all women enjoy robust health requires the talents and resources of many individuals and groups. It is a promise that can be fulfilled only by engaging the hearts and minds of men and women alike. By giving men and women scientists an equal opportunity to pursue their hopes and dreams and to fulfill the promise of their natural gifts and talents, the NIH is also fulfilling its pledge to women and its promise to all Americans.

Through a wide variety of programs designed to encourage young women and young men to pursue careers in science and to undertake medical research related to women's health—as well as by initiating and sponsoring programs aimed at helping women advance in their scientific careers—the ORWH plays a key role in fulfilling the NIH's promise of improved health for all Americans through science. From films and computer-based science curricula for middle and high school students to programs that enable investigators at the beginning of their careers to pursue research on women's health issues, the ORWH is helping reshape the demographics and the culture of biomedical science.

When we look at the health of American women today, we see a picture as varied and diverse as the lives and roles of women themselves. Not all American women and their families enjoy the same level of health and health care. With so much that remains to be done to address the many pressing issues in women's health, it is easy to focus only on the challenges ahead, forgetting where we started and how far we have traveled. As Marie Curie once observed, "One never notices what has been done; one can only see what remains to be done."

We must recognize, however, that we have traveled far indeed, thanks to the efforts of thousands of women and men from every sector of American society. In celebrating the 12th anniversary of the ORWH, we acknowledge its contributions and celebrate the power of our individual voices and collective actions to effect real change in government policies and programs.

In the coming decades, the NIH and the ORWH will continue to foster an approach to the study of women's health that encompasses the myriad social, behavioral, and biological factors that influence health over the course of the lifespan. Only by addressing the multiplicity of factors that influence health can we improve the health and well-being of women in the 21st century. That is our challenge—and our promise for the future.

Keynote Address 4

Point and Counterpoint: The Critical Role of Relevance

Raynard S. Kington, M.D., Ph.D.

Dr. Kington is the National Institutes of Health (NIH) Associate Director for Behavioral and Social Science Research; he directs the NIH Office of Behavioral and Social Science Research (OBSSR) in the Office of the Director. Last year Dr. Kington also served as Acting Director of the National Institute on Alcohol Abuse and Alcoholism. His research is focused primarily on the role of social factors, especially socioeconomic status, as determinants of health, including wealth and health status, the health status of U.S. Hispanic populations, determinants of health care utilization in racial and ethnic differences, and the use of long-term care.

I was asked to speak briefly on my perspectives on a few issues in recruitment and retention, and I will focus primarily on issues related to recruitment for minority populations. My experiences in this regard are numerous, complementary, and sometimes conflicting, but perhaps my most influential experiences were in the field, running a community-based research center in south-central Los Angeles that focused on the health of elderly African Americans. After that field experience, I directed the National Health and Nutrition Examination Study (NHANES), which is the only national representative sample study of the health of the American people based on actual physical examinations and extensive clinical and laboratory testing of about 5,000 Americans each year; NHANES is sponsored by the National Center for Health Statistics (NCHS), part of the Centers for Disease Control and Prevention.

My perspective is also informed by my experience as the NIH Associate Director for Behavioral and Social Science Research, where I have been called on to think and act more broadly in our efforts to expand the application of the behavioral and social sciences to promote the health of American people. One important area in which the behavioral and social sciences have been especially valuable has been in the field of health disparities research. Over the past few years, the OBSSR has been deeply involved in helping to sort out, for example, some of the ethical issues related to health research involving behavioral and social science research; concerns about

outreach, community relations, and recruitment have been an important part of those discussions.

Two issues are especially important in trying to make research more relevant for addressing both minority health and women's health. The first issue is the lack of data collection and analysis on participation rates in many studies. The second issue is more subtle and sensitive: how scientists interact with communities as we try to improve the conduct of research in community settings.

At present, scientists and communities have very few data to help guide our understanding of many of recruitment and retention issues, and this lack of data often results in the creation of myths regarding the research process. The primary data problem is a lack of solid empirical demographic information about people who are asked to participate in research studies and clinical trials as well as those who refuse to participate in such studies. In the population sciences—particularly epidemiology, sociology, and demography—the state of the science for large studies demands an understanding of the representativeness of the study population and an understanding of the overall population from which representative samples are selected. In almost all the large epidemiologic and demographic studies, these data are used to calculate sample weights that can be used during analysis to count for nonrandom participation among those who are offered a chance to participate. As you know, participation in research is rarely truly random, and

who enters and completes studies can have a significant impact on research results and, perhaps more importantly, on the interpretation of those results.

Dr. Brawley and myself and several colleagues at the NIH are completing the analysis of data from a number of large studies that are revealing an interesting result. The conventional wisdom in the research community is that minorities are less likely to participate in research in general. We looked at several large, epidemiologic observational studies of varying degrees of invasiveness, from simple interview studies to studies with extensive clinical examination, and we are in the process of comparing participation rates across subgroups. We have found that, when we look at all persons who were offered a chance to participate in these studies, minorities do not consistently have lower participation rates; in fact, in some of the studies, some minority groups have higher participation rates than those of Caucasians.

This finding is not a surprise to me, in light of my field experience. When I directed the NHANES study, our field staff looked forward to going to communities with large minority populations because—and they would often say this—they expected the response rates would be higher and that helped balance response rates for the year. Although we have not looked at these data in particular, these same staff members often believed that the most difficult group to enroll in observational studies was well-off Caucasians. (This was not true for studies in which enrolling gained access to treatment that would otherwise not be available, but only for observational studies.)

In the same analysis, working with Dr. Brawley and others, we were able to examine data from the NCI's minority-based community clinical oncology program, which is a network of 10 sites conducting clinical trials for a range of different cancers, in which they collected careful data on who was offered and who refused enrollment. Again, we found that large differences in response rates were not present, and although the response rates of African Americans were slightly lower than those of Caucasians, the response rates of Hispanics were higher than those of Caucasians.

In another paper that will soon be published in the *Journal of Genetics in Medicine*, colleagues from the NCHS and I examined data from the 1990 and 2000 NHANES studies in which participants in the overall study were also offered the chance to have their deoxyribonucleic acid (DNA) put into a national repository for unspecified future genetic research. At first, we found that African Americans were slightly less likely to agree to have their samples included in the

repository and that Mexican Americans agreed at about the same rates as Caucasians. The differences were relatively small: Approximately 79 percent of African Americans in this study agreed to have their genetic samples put into a DNA repository, as did approximately 85 percent of Mexican Americans and about 87 percent of Caucasians. We also noted that women, on average, were slightly less likely to agree to have their DNA put into a repository.

So what do these data mean? We are currently trying to identify an even larger number of studies to determine whether the response rate data are consistent across a wider array of research settings and a wider range of study types. One of the tentative conclusions we draw is that, at the very least, the conventional wisdom about how difficult it is to gain the participation of minority groups in research endeavors is simplistic at best. Based on the rhetoric about minorities and genetic research, most scientists would have guessed that far fewer African Americans would agree to have their DNA entered into a genetic repository.

Participation in research is rarely truly random, and who enters and completes studies can have a significant impact on research results and interpretation of those results.

Some researchers are able to achieve comparable participation rates across a wide array of groups, and the mythology that has grown up about minority recruitment in particular has perhaps focused too much attention on the problem of minorities who will not participate and not enough attention on what researchers are doing or not doing to achieve adequate response rates. We need to have better data on the details of recruitment, especially in clinical trials, if we are ever to seriously address these issues. In many biomedical studies, no data are maintained on who was offered a chance to participate in the study and who ultimately did not choose to participate. I know it would be costly and difficult to collect these data for every clinical trial, but there are things we can do short of that. One strategy would be to develop a system of regularly sampling clinical trials and collecting more comprehensive data from the sample studies to determine who is refusing to participate and, more importantly, why they are refusing to participate.

I serve on an institutional review board for the NCHS. This committee conducts the ethical reviews for all of the

national surveys that monitor the health of the Nation, and we often have to make judgments based on few or no empirical data. For example, we often discuss what are considered invasive or sensitive questions, and we often have very few data to guide this discussion. It is a constant moving target, and we know very little about how that sensitivity might vary between men and women and across different subgroups. For example, questions about sexual behavior for our current cohort of persons older than 65 might be received differently compared with how they might be received in 10 years when the first wave of substantial numbers of baby boomers appear in that age group. The different populations bring very different life experiences, in this case, in terms of openly discussing sexuality. Yet we still tend to make broad generalizations about what is sensitive and what is not and assume that “one size fits all” across all subgroups. That is simply not the case; as we begin to look at the evidence, some things are the same across groups, but other things are very different—and we need to understand that to make appropriate decisions that operationalize our research.

In our analysis, we found that women were slightly less likely to agree to have their DNA included in a repository. After finding this result, we were basically left to speculate about why that might be the case. Perhaps inclusion of DNA samples for women raises issues about reproduction and the future for their children that may not be as salient for men, but we do not really know. We need more information on how perceptions and concerns might be driving differential and sometimes similar response rates.

In another situation in the OBSSR, we are helping sort out guidelines for researchers. Some investigators made the claim that, because mental illnesses were more stigmatized than other diseases, the standards for designing studies for those illnesses should be more stringent. The question we asked was, “stigmatized for whom?” We also asked what those differences might be across groups and over time. For example, the world has changed dramatically in recent years with respect to the willingness of individuals to openly discuss mental illness. A presidential candidate publicly acknowledged the use of mental health services, which would have been unthinkable not too many years ago, and yet we have very few data to help guide those sensitive areas and how

they might vary across different subgroups. For another example, perceptions about human immunodeficiency virus (HIV) infection have changed dramatically in some communities, and yet important decisions about how and whether HIV research is conducted too often rests on the shaky foundations of speculation and bias, and too often our decisions fail to capture the diversity, preferences, and experiences across population groups. All of this is a plea to gather more data.

The second topic I want to talk about is a little more sensitive but is nonetheless important to raise: the general issue of interaction between scientists and communities, particularly in the context of what is now often referred to as “community-based participatory research.” There is no question that, in the past, scientists had a poor track record for actively and respectfully interacting with communities in the conduct of biomedical, behavioral, and population research; some of those poor track records continue to exist today. The community-based participatory research movement has been an appropriate response to facilitate a change in the terms of interaction between scientists and communities and to make research more relevant to communities.

However, I am concerned that this approach might be operationalized in ways that ultimately can hurt communities and can be disrespectful of the ability of community groups to understand the value of science. For example, in discussions about community-based participatory research, I have often heard the plea from community groups and some well-meaning scientists that the community should decide what its problems are and what topics should be studied within their communities—that research priorities should somehow “float up” from within communities. In my experience, this argument is made almost exclusively when talking about minority and other disenfranchised communities. Communities setting the stage for research agendas and determining research priorities sounds—and in many ways is—good, and I wholeheartedly sanction the idea that community members bring essential insights about their communities’ problems. A simplistic approach to community-based research, I believe, ultimately holds the possibility of denying some communities the benefits of science.

One of the major contributions of epidemiologic studies—appropriately collecting the right information

Communities are fully capable of understanding the value of science, and we do them a great disservice when we underestimate that ability.

and analyzing it appropriately through rigorous application of scientific principles—is that we can get a picture that is not possible to achieve by simply asking people about perceptions. Perceptions on the ground level can be and often are incorrect, and we have had to learn that lesson all too well in the AIDS epidemic. Many communities initially denied, and some still vehemently deny that their communities were disproportionately affected. That denial was understandable, but it may very well have hurt these same communities in addressing the epidemic. Science tells us that individuals, even well-intentioned and well-connected individuals, each see only a slice of reality and often a distorted view of the health of populations. It is disturbing when the suggestion is made that we should use science for some communities to determine the magnitude of their health problems but in other communities we should use folk wisdom.

Folk wisdom does have cultural and in some cases actual health value, but in my opinion, the best science possible (not folk wisdom) will help us figure out the

problems. Often, a double standard is in place—as if some communities cannot possibly understand that the scientific perspective is valuable, different, and important, and, when properly applied, can ultimately save lives. When I was conducting qualitative research in south-central Los Angeles on the perspectives of elderly African Americans on biomedical and public health research, I was impressed by the statements of those who demanded that African Americans be included in research because they wanted to make sure that their children and grandchildren benefited from research.

Communities are fully capable of understanding the value of science, and we do them a great disservice when we underestimate that ability. Science can and should be used to benefit all communities. We will all be better off when we begin to have a more open and honest dialog that recognizes the great advances that can be made when communities and scientists work together acknowledging the assets that each brings to the table to address important public health problems.

Lewis H. Kuller, M.D., Dr.P.H.

Dr. Kuller is Professor of Public Health and former Chair of the Department of Epidemiology at the Graduate School of Public Health at the University of Pittsburgh. He is recognized widely for his contributions to elucidating cardiovascular risk factors and determinants of the development of coronary heart disease in women, including the development of atherosclerosis from premenopause to post-menopause, prevention of risk factors in menopause, and hormone metabolism and the risk of breast cancer and osteoporotic fractures.

The NIH has been very successful in fostering major studies that have provided the basis for better clinical and preventive medicine for women and minorities. These studies have provided an important guidepost for reducing disparities and improving health care outcomes. These important studies have in common the following: (1) a well-defined hypothesis; (2) reasonable sample sizes of women and minorities to test hypotheses of potential differences in specific outcomes; (3) long-term funding to be able to clearly measure outcomes; and (4) a strong emphasis on recruitment, retention, and adherence to various therapies. The results of these studies have generally demonstrated that there are relatively small differences in risk factors for clinical disease or outcomes of intervention by either race or sex for most major diseases. Selected differences

Everything we do must be demonstrated to have a real outcome and demonstrate both efficacy and effectiveness.

both by race and sex can be identified that require additional research.

The primary determinants of health outcomes remain: (1) the determinants of risk factors, the interaction of lifestyles, and host susceptibility (genetics) and (2) adherence to therapies, both pharmacological and nonpharmacological. The strongest determinants of these two variables are the levels of education and some measures of cognitive function.

The study of the etiology and pathogenesis of disease and the development of specific preventive approaches

based on successful studies at the NIH and their application in the community, both in public health and preventive medicine, has by far the greatest impact on the health of women and minorities. This approach requires a strong continued commitment to basic, clinical, and experimental research studies at the NIH to determine the best approaches to reducing morbidity and disability. We must focus on outcome studies and not just on the process. Everything we do must be demonstrated to have a real outcome and demonstrate both efficacy and effectiveness. Doing good does not mean that it is good.

The NIH policy of requiring the inclusion of women and minorities in all NIH studies—including epidemiology, clinical trials, and so forth—has in general been a huge waste of money and probably has detracted rather than improved health and may increase rather than decrease health disparities. The reason for this is that such studies shift resources from productive studies that include reasonable sample sizes and test specific hypotheses to

We need a portfolio of studies that answer specific and important questions regarding women, minorities, and various ethnic groups.

studies that have little or no likelihood of increasing our understanding of either gender or racial differences in health. Studies that include women and minorities place women and minorities potentially in harm's way by exposing them to experiments without any likelihood of determining specific benefits because of small sample sizes and the inability to determine whether there is any real gender or racial difference in the outcome. These studies lack specific power because of the small sample sizes. Thus, they provide no real scientific information or, more likely, provide what are called substantial type 2 errors, potentially the wrong answer. For example, prior to the completion of more large-scale studies, it was reported that women do not benefit from antihypertensive therapy. A recently published article suggested that obesity is not an important health problem for African American compared with Caucasian women. Small studies have reported that American Indians, in spite of high rates of diabetes and cigarette smoking, are immune to coronary artery disease. All of these observations are untrue.

What should we be doing? We need to change. What we need to do is have a portfolio of studies that answer specific and important questions regarding women, minorities, and various ethnic groups. We need to redefine carefully what the term “minority group” means and whether we are talking about genetic differences,

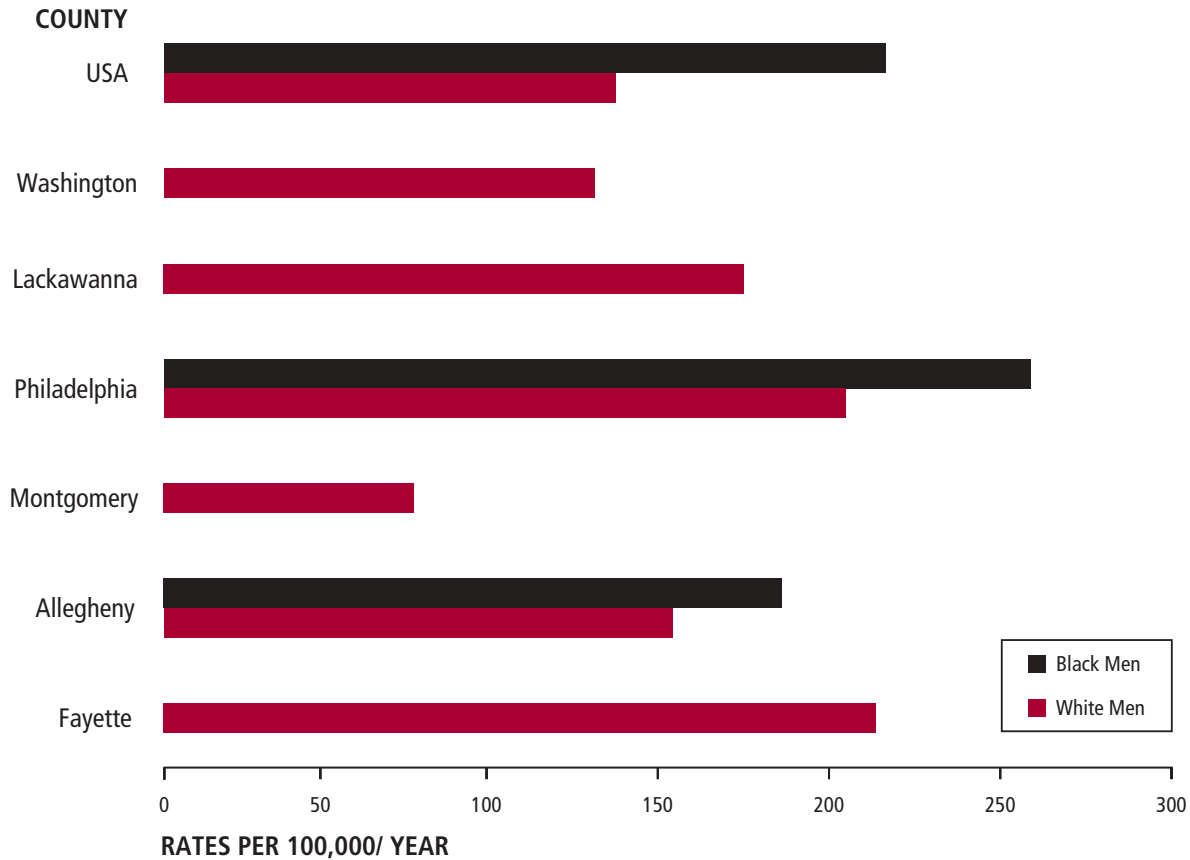
socioeconomic factors, or cultural differences. We must recognize that blaming racial and gender discrimination for inequalities in health status is both counterproductive and a poor excuse for our failures. First, there have always been disparities in health outcomes, and there always will be disparities even in the most homogeneous societies.

Second, in spite of disparities, Dr. Joseph Goldberger wiped out pellagra in the early 1900s, without knowing that nicotinic acid was the cause but by recognizing that this was a dietary-related disease and that modification of the diet of the population would have an effect on the disease independent of other changes in the environment. In addition, syphilis was practically eradicated in the United States. Jonas E. Salk, M.D., and Albert B. Sabin, M.D., successfully developed a vaccine to eliminate polio, and other childhood illnesses have been eliminated. Lead poisoning in children has been greatly reduced, and stroke and coronary heart disease (CHD) mortality has declined dramatically across all gender, racial, and socioeconomic groups. These successes have occurred in spite of continued substantial socioeconomic and educational differences in communities. We clearly must strive to reduce these racial and socioeconomic disparities. This cannot be an excuse for our failure to reduce the incidence and mortality of preventable diseases.

A major goal, therefore, of a successful women and minority program must be effectiveness research and the application of successful scientific research to reduce disparities in health by gender, race, and education. Results to date have been poor. We need a better approach to “catch the ball” and run with the successful positive studies. We have had a very poor performance in modifying ethnic, racial, and gender disparities. Recent Institute of Medicine reports have primarily emphasized health services differences. I believe that this is a mistake. There is no question that there are disparities in the delivery of health services and that they should be reduced. However, they contribute only a small component to the overall health disparities. The biggest problem relates to lifestyles and how to deliver preventive health services effectively.

Here is an extreme example. Fayette County, Pennsylvania, is a rural county about 1½ hours from Pittsburgh. It has plenty of health services. The death rate for Caucasian males is high for CHD, age 45 to 54. From 1989 to 1998 the CHD death rate for Caucasian males in Fayette County, Pennsylvania, was higher than that for African American men and Caucasian men in Allegheny County, Pennsylvania. Montgomery County, Pennsylvania, a suburb of Philadelphia, is a rich community. It has a death rate about one-fourth of that in Fayette County and about one-sixth of the death rate

Age-Specific CHD Death Rates: Age 45-54, 1989-98 per 100,000/year



for African Americans in Philadelphia. The rates for women are the same as those for men, with women in Fayette County having very high CHD death rates and women in Montgomery County having very low CHD death rates. The critical issues are lower SES, high levels of risk factors, and inadequate preventive services.

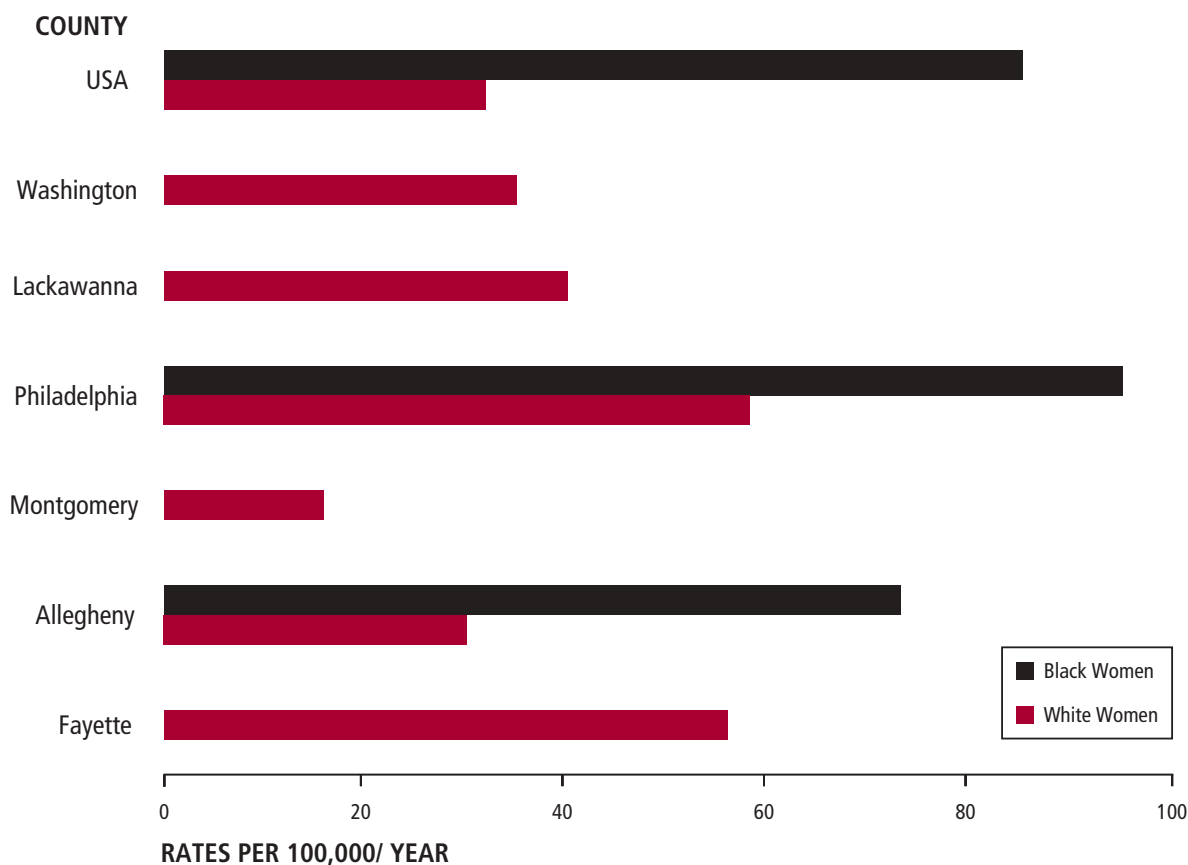
We are not going to solve the fivefold difference in CHD mortality between African American men in Philadelphia and Caucasian men in Montgomery County, Pennsylvania, or Caucasian men in Fayette County, Pennsylvania, or, similar rates for women, by requiring a percentage of African Americans and women in every single clinical epidemiological study and clinical trial.

Low SES, poor education, and poverty are associated with excess morbidity and mortality. We have known that for at least 100 years. We need neither more studies that show that low SES is associated with high rates of disease

nor more studies to find better ways to find the relationship of SES to health and the interrelationship of education and outcome. We know the problems, and we must seek the solutions.

The real focus must be on identification of at-risk populations and much better systematic approaches to maximizing adherence to prescribed therapies such as reduction of elevated blood pressure. There is no excuse for only 30 percent of Americans having their hypertension controlled. We still have a high prevalence of cigarette smoking, especially among minorities and women. We have an epidemic of obesity. Hyperlipidemia is poorly controlled, and the prevalence of atherosclerosis is extraordinarily high among middle-age and older individuals, especially men in the United States, accounting for extremely high incidence of CHD and major cost to the health care system. These problems

Age-Specific CHD Death Rates: Age 45-54, 1989-98 per 100,000/year



cannot be resolved by just adding women and minorities in all NIH studies. I believe that this has become, unfortunately, a rationalization for our failures to implement the successes of our basic clinical and efficacy trials to programs in the community.

Our continued failure to understand the cognitive processes that lead to both adverse risk factor profiles and to good adherence to health behaviors or therapy is a major limitation for the reduction of disparities in health across race, sex, and ethnicity. These adverse cognitive processes almost certainly begin early in life and are only in part measured by education. We must focus on a better understanding of the physical and social environments, nutrition, and infectious agents that affect cognition as well as important host genetic factors. The emphasis on differential behavioral interventions without understanding

the cognitive processes that may be at the root of some of these adverse risk behaviors and adherence to therapies will not solve the problem. We are not measuring intelligence but rather how individuals process and use information and implement changes in behavior.

Individual behavior, as well as the social and physical environments (i.e., the host, the agent, and the environment), is the cornerstone of epidemiology. If we presume that the host is not important and that the brain of that host is not important—how that person thinks and functions and how these factors have evolved since childhood and how they can be modified—then we will not have a winner. We must recognize that the processing of information by the individual is probably the most important variable irrespective of education, race, sex, or ethnicity. We have to determine to some degree why there are variations in how people respond in the community

with respect to health behaviors (i.e., lifestyles and preventive and clinical therapies). We must face the reality that some people do well and some do not do so well. Are there more “pellagras” out there? Is there more lead poisoning out there? Unfortunately, nobody seems to want to find out. We presume that studying cognition and how people process information is a taboo and, therefore, cannot be studied. This is simply wrong. We should ban statistical adjustment of data by SES, education, and race until we understand what we are measuring.

We must base the successes or failures of women and minority programs on real health-related outcomes.

We need to expand and support both women and minority investigators who can compete for basic clinical, epidemiological, and preventive research. These individuals must work in environments with investigators who have solid track records of research. This must be a high priority. You learn by working with accomplished people who are enthusiastic and who will continue to provide the environment for successful training and career development. It is obviously important that women investigators study not only women but also men. If you are a minority investigator, you should study not only minority populations but also all other populations. We may lose extraordinarily competent minority investigators because we insult them by telling them that they should apply for a minority research award or should focus only on minority research. We should be training minority investigators who will become first-rate researchers at the NIH and at our universities.

We must base the successes or failures of women and minority programs on real health-related outcomes and not on “bean counting,” that is, the number and percent of minority or women in each study. This is counterproductive to good science and will not reduce disparities.

QUESTIONS AND ANSWERS FROM “POINT AND COUNTERPOINT”

Dr. Kington’s Response

I agree with much of what you are saying, particularly the question of decomposing rate factors that we know we can intervene upon and that give us a better understanding of the actual cause of pathways that account for the gross differences we see. We part company in several ways. One, the notion that there are racial, educational, and

other clearly delineated factors is a false dichotomy. There is little research about the real interactions across these variables—the ways in which race, ethnicity, and gender interact with a range of other variables.

For example, we are about to release a request for applications (RFA) to fund research to examine the pathways by which education ultimately leads to differences in health status. One big problem, particularly in older populations, is the relatively crude measure of education; a year of education for the average 65-year-old Caucasian is not the same as a year of education for the average 65-year-old African American. We know that most elderly African Americans received their education in the South, and we have documented evidence of huge gaps in the quality and duration of education in the South during the time that many of the older population were receiving their primary education. The idea that these variables can be sorted out neatly is unrealistic, and we need to do a lot more exploring to look at the ways in which these various factors interact.

Some people argue that it is ultimately impossible to disentangle some of these pathways because, for example, racial discrimination can lead to differences in income attainment. In determining health status, it may be possible neither to isolate income and race from other factors nor to separate income and race from each other. We must have a much higher quality of research to address this issue.

I agree that it is time for us to start thinking about not allowing investigators to say, “We don’t have to worry about minorities and women because our population mirrors the percentage of minorities and women in the overall population from which we are sampling.” As we know, for small sample sizes (small populations), that will not give you enough power to make any reasonable conclusive statement. We have to do a better job of moving toward thinking about whether the numbers of minorities and women are adequate to make judgments about whether or not those populations are different in any way in terms of outcomes or processes.

I also have a problem with the last anecdote about the minority candidate whom you encouraged to apply for the minority award. Every minority scientist in this room has had this happen, that is, someone says you will be able to get a particular grant because you are a minority. I think we absolutely need those programs because the playing field is not level. It is wrong to stigmatize those programs as if they are second rate, because they are not, and the goal is to level the playing field, not to fund second-rate science in any way. I know you did not mean to say that, but that is a pathway down which the processes often lead.

Questions and Answers

Q: Dr. Kuller, I think you completely misstated the NIH policy on the inclusion of women and minorities, and I want to clarify that. The NIH does not require the inclusion of women and minorities in any quota-type percentage way in all studies. The NIH does ask that investigators include women and minorities in Phase III studies unless they can provide a scientific justification for not having such an inclusion plan.

DR. KULLER: Let me correct you, because I just got a note from the National Cancer Institute (NCI) last week, and I was told that Emory University Cancer Center is not enrolling enough African Americans. They did not give me a number, but they said we should increase African American enrollment in all our cancer trials.

We ask that scientists design their research and include in their study populations something representative of affected populations.

Q: My understanding of the NIH policy is that if you have a scientific justification for not including women and minorities, that is acceptable. So the question is, can you provide sufficient justification to pass peer review on that question? If what you are saying is true and inclusion of women and minorities is irrelevant, then submit your application that way. It is not an absolute requirement, and it never has been.

DR. KINGTON: I think Dr. Pinn might want to comment on this. The ORWH is responsible for monitoring the inclusion of women. I think you are correct—the rule says you have to justify your decision. Many NIH Institutes and Centers (ICs) (understandably for programmatic reasons) strongly encourage their investigators to enroll substantial numbers of minorities and women in their studies so that they can answer the fundamental questions about what works and what does not work in these populations. There are two issues: the rule and the programmatic decisions in which some ICs are aggressively encouraging their investigators to include appropriate numbers of women and minorities to answer some fundamental questions.

Q: The question and the argument may be with how the rule is being implemented, but I just wanted to clarify what the guideline actually states.

DR. KULLER: Having sat on a large NIH study section that dealt with studies having had problems in dealing with this, having sat on an NIH council, and having been involved in policy implications at several NIH ICs, I can say that, in general, the policy is that researchers cannot

exclude women or minorities in studies. The concept you are raising is correct; unfortunately, the reality of the situation is not. There is no question that, for Phase III studies for which there is value and importance in learning more about women or minorities, you should have a big enough sample size to answer the question. The problem is that the NIH and investigators out in the community are spending an inordinate amount of money to get a study that has 100 men in it also to have 20 women, especially where there is no chance that those 20 women will identify any effect different from the men.

DR. PINN: I want to make a comment, especially since what I spend half my life doing was called an embarrassment (for me, for the NIH, or for you, Dr. Kuller). I have taken the effort and the initiative to do this for the NIH because it is a congressional requirement if we are going to get appropriations, which we need to fund the research. We all know the language of the law, and we know that there was much criticism of the law before we wrote the implementation plan, which was published in March 1994. I still hear many comments about what it was thought the law would include, as opposed to the really sincere efforts that NIH scientists made so that the science would not be affected but could accomplish the intent of the law.

It is amazing to hear your side because, on the other hand, I go to other meetings at which we hear that the NIH is not paying any attention to this and has no requirements! Perhaps it does differ by study section, by who is involved, and by IC, as well as by the attention being paid by the individual IC directors. The NIH Office of Extramural Research ensures there is consistency as much as possible.

The requirement of a percentage is something new. Our policy in this report states very clearly that there are no quotas and that no percentages are required. We ask that scientists design their research and include in their study populations at least something representative of affected populations. If someone is giving you a specific number, as opposed to looking at it from the standpoint of science and if you are trying to determine representative populations, you would likely use that for your justification. You should be able to appreciate why we need some regulations, because you pointed out how you had a study in which you were going to include women, but you were told that it was not necessary. The policy went from recommending and urging to requiring that women and minorities be considered when appropriate, and it was the Congress that said it did not trust researchers to do so without a law. We did not want that law, but that is what happened. Ruth Katz, J.D., M.P.H., and Susan Wood, Ph.D., are two of the people who wrote that law.

At the NCI, Dr. Brawley was one of the people who helped monitor what was going on, so I am amazed that he now reports that somebody told him he had to include more African Americans. The NCI was one of the ICs that had a sharp increase in Caucasians in clinical trials. Can you explain what was said in your letter? Also, I would like Dr. Kuller to explain this requirement of percentages, because I am hearing that the review committees are not paying attention—and nowhere should there be a percentage.

DR. BRAWLEY: The NCI funds institutions to enroll people in a number of clinical trials. For example, the Eastern Cooperative Oncology Group, of which Emory University is a member, has 50 clinical trials that are currently open; Emory is accruing patients to 25 or 30 of those. Program directors at the NCI look at various institutions and notify them that their minority accrual to clinical trials is lower than what the NCI thinks it should be. This is one of the issues I discussed 10 years ago, and I remain concerned about it. I supervise 90 doctors in 4 hospitals whose job it is to accrue people to those 30 clinical trials, as well as to see people in their usual clinical practices. I am very concerned what one of

What is the point of spending this time and money unless we can clearly demonstrate a significant outcome?

those doctors may be saying to my African American patients at one of those four hospitals, because they all know that we are at risk of losing what amounts to about \$2 million a year from the NCI for accrual to clinical trials if we do not increase our minority accrual. I am concerned that they are giving African American patients the “hard sell” and possibly discouraging Caucasian patients from enrolling in the trial. This system has been in operation this way by the NCI since about 1994. It is not new, and no one has ever said, “This is your quota” or “This is your number or percentage.” The program directors look at accrual from the previous year and determine that African American or Hispanic accrual should be higher. They then state that, if those numbers do not improve, they will either decrease grant money by a percentage or remove it altogether; the Mayo Clinic and the Memorial Sloan-Kettering Cancer Center have been hardest hit because they lost money.

DR. KULLER: I think the major problem is that we are talking about process and not outcome. What is the point of having all this process and spending this time and money unless we can clearly demonstrate a significant

outcome using this approach? There is an eightfold difference in one State in CHD morbidity and mortality, and we are doing hardly anything about it.

DR. BRAWLEY: We have lots of research that we cannot hook up directly to substantial improvements in mortality and morbidity. We know this is a long process.

DR. KULLER: You are absolutely right. The studies we should be designing to include women and minorities should state that conducting these studies will provide an answer that will lead to the next step. We should not keep going around in circles simply because we have a mandated number of women and minorities for each study.

Q: I want to give a real-life example. The early studies on cholesterol-lowering medications were conducted in men, and when the therapy guidelines came out, it was recommended that these medications not be used in women because there was no evidence of their efficacy in women—because there were no women in the trials. Those are the kinds of data we want to come out of inclusion of women and minorities in studies. Research backfires if there are no women, and it ultimately affects clinical practice.

DR. KULLER: Until recently, the NIH has not funded any studies for outcome of cholesterol lowering in humans; all these studies have been funded by the drug companies or in Europe. People have done meta-analyses to show that it does work in women, but unfortunately, coronary disease risk factors, at least in terms of low-density lipoprotein cholesterol, are somewhat different in men than in women.

Q: I want to comment on your recommendations, which I see as two separate issues. There is a tremendous need for leadership in disparities research apart from the regulations, and that concerns me. I work in the field of obesity, and at last count, there were something like 35 obesity/weight reduction trials conducted in minority populations, and only one or two of them were randomized—out of thousands of obesity-related trials. Many of these investigators do not feel that it is their mission to look at this issue in the populations, in part because they do not yet have the expertise and do not feel competent to do it. As long as we lack leadership coming from the NIH ICs to take these disparities and come up with a systematic program to try to take them apart, we will not have focused research on the disparities.

Q: If you say something like, “What proportion of obesity research funded by the NIH could address the question about the highest risk populations?” people should be held accountable that some proportion of it—more than the current approximately 1 percent—should go in that direction.

DR. KINGTON: I agree 100 percent.

DR. JOAN DAVIS: I am a physician trained in preventive medicine, and I would argue that there is a compelling abundance of evidence that shows that education and socioeconomic status impact health. My question to you, Dr. Kuller, is if you are arguing that cognition plays a major role in health outcomes, how do you alter someone's cognition without educating them?

DR. KULLER: There are two answers to that. The number-one process is to understand from early life what factors adversely affect people's cognitive processing. The second answer is that the advertising industry knows how to do that, because just take a look at what they have done adversely. The best example, adversely, is tobacco advertising—the “Joe Camel studies”—that focused advertising successfully into low-income minority communities; these ad campaigns were so phenomenally successfully that the only way we could stop them was by legislation. They were able to manipulate lifestyle without changing educational level.

You can have adequate representation and still not be able to answer the question of differences.

The best example of altering cognition without education is the hypertension detection and followup program, in which community samples were obtained in 1981. In the special intervention, they had wiped out the SES or educational gradient in both African Americans and Caucasians for morbidity and mortality and had obtained 70 or 80 percent adherence, including in the African American population. They developed a model for the intervention, which related to the characteristics of the study participants. They looked at some of the issues and trained health educators to go out into the community to work for these people on a day-to-day basis to modify the behavior. It is not modifying cognitive behavior but, rather, understanding it so that the intervention is consistent with the cognitive functioning and cognitive processing of the participants.

DR. PAULA STRICKLAND: I am one of the people who writes those summary statements that have to do with whether your grants get funded or not. I want to make a clarification in terms of the policy for inclusion. In all clinical research, you must show a reasonable plan for the inclusion of women and minorities. If there is a compelling reason for not including them, you can include that in your justification as to why they are not there. The study sections look at the demographics of the

area. There is a policy, and it refers to all phases of clinical research.

DR. KULLER: I am saying that we have matured enough right now that maybe it is time to go back to Congress and maybe it is time for the public and the scientific community to say, “We have some really big problems now that we must solve, and this is the way to do it; our current mess of the process is not solving the problems.”

DR. KINGTON: I agree that this is an important point to make: that you can have adequate representation and still not be able to answer the question of differences or whether there are ways to be informed about how different groups' outcomes are compared with one another. I do not think we should let investigators off the hook by saying that their representative sample of women and minorities is adequate when that sample only exactly mirrors the percentage in the community. We have to be more aggressive; some of that is driving the programmatic staff's concern.

DR. KULLER: Instead of having 10 percent or 15 percent of African American women in 14 different obesity studies, which do not provide any useful answers, why not have the NIH put out an RFA oriented toward reducing obesity in African American women and supporting researchers who delve into that problem?

DR. KINGTON: You can also help answer the question by having a larger sample of minorities in that study, so you can have the power to detect differences. You are assuming that the fundamental approach has to be different, and that may be true in some populations. In other populations, we may see that the same intervention works in different ways in different populations. The point is, you can do both: You can have interventions that oversample minorities and are able to see differences in outcomes for the same intervention, and you can have targeted interventions that target a specific population with a specifically developed intervention for that population.

DR. KULLER: These studies lack power; they cannot tell whether any differences were found. At the end, if you get a big difference, you do not know whether it is real, or if you do not get a difference, the argument is there is no statistical power.

DR. ROY LOU: Let me follow up on the topic of going back to Congress. I want to ask the audience here just a question: Do you feel trapped by the policy statement that cost is not to be considered in terms of excluding women and minorities? As a program director, I look at quality of grants, and I want to fund studies that will give me an answer. Which grant shall I fund? I am a steward of public money. How should we make this judgment? It is a great principle to say that cost should not be considered a

factor, but in reality, cost is everything. Are we going to go back to Congress and address that issue? Considering what was published in *The Wall Street Journal* yesterday—on page 2 it reports that the NIH budget is going to “take a big hit.” What should we do about it?

Q: As a citizen, I certainly hope that the NIH is providing the stewardship of our money. Those guidelines should say that the studies that get funded are the ones that will give us the best information—about all the populations in our society. I do not think it would sit well with citizens to hear that the only studies getting funded are those that only include Caucasian men and, therefore, that we are not getting information about 55 percent of the population (women). As citizens, we expect a higher standard, and we want you to figure out how to make the better judgment.

DR. KULLER: There is no question that our women’s health research has come a long way, and no one in this room would disagree—we have done a phenomenally successful job in women’s research over the past 10 years, for which the NIH deserves a lot of credit. That is not the issue.

DR. KINGTON: Investigators cannot refuse to include women and minorities due to cost or because it is more difficult to enroll them; those reasons have been used as an excuse for years for not addressing difficult problems. The problem is, that runs up against a real cost factor; in some situations, it may be more costly to recruit difficult-to-reach populations of all sorts. My concern would be that, if you do not acknowledge that, differences in costs might appear, or there might be a different type of negative incentive to deal with the problem. This regulation started from a good place, however—the desire to make sure that investigators do not use that excuse not to do anything.

DR. KULLER: All I was arguing was that we have matured beyond that. We have done phenomenally well, but what is the next step?

DR. FLORENCE HAZLETINE: Ten years ago, things looked very different, and we should be excited and pleased that this type of argument is being raised now. It was true that women were not included in studies, and the reasons were often cited as cost, but the reasons also were built on a scientific basis that experiments and research had been done using men, and so it was easier for the investigators to add on. If they had come in with

study sections with women, they would have been thrown out because they would have had to redo the baseline. The Women’s Health Initiative has helped us with that, and there have been a great deal of strides; without public outcry, I do not think any of this would have happened.

I think we still have to deal with an underlying problem; let me point it out by giving you a story. A year or so ago, I was sitting next to Joshua Lederberg, Ph.D., who said to me, “Why all this women’s health stuff? Why do we need to do the research?” I said, “We don’t *need* to do the research, but I want you to look at the following issues. In the past number of years, a large number of drugs have been pulled off the market; 7 or 8 out of the last 10 have been pulled because of adverse effects in women. That’s really rather frightening, because if that research had been done up front, it would not have been a billion-dollar problem later on. We are talking about millions of dollars required to do the study versus billions of dollars that affect the economy at many levels. It also affects the health of our people, and it affects how many people go into hospitals.” If you want to argue on a cost basis, you can argue it that way.

If you want to get at a basic science issue, look around the room and see the many researchers who would never have been involved in this area. The science issues are becoming very interesting, and we are providing a whole new area for our young investigators to fight for funds using this as a criterion. We have encouraged a whole new set of investigations because of what people wanted.

I am a gynecologist, so it was surprising to me that women were not included in clinical trials because we do not include a lot of men in our clinical trials (although they sometimes accidentally appear there). Obstetricians and gynecologists may have to broaden their perspective as well. Where we are right now is predictable and, to those of you who are interested, do not let bureaucratic entanglements discourage you.

Q: As more people with HIV happen to be women, why is it that women represent only 5 percent or 10 percent of people enrolled in important clinical trials about HIV—treatment interruptions, immunological assessments, and vaccine studies? If we find out that these are helpful, we really will not know if they are helpful to women. The Government is trying to ensure that researchers like myself assume that you have to study the population with

It is a great principle to say that cost should not be considered a factor, but in reality, cost is everything.

a process; if 30 percent of the population has that process, they should be involved in the studies to as representative a sample as possible. It will be more costly, and it may involve different assessments, because women have things like menstrual cycles that may affect assessment of drug dosages.

DR. KULLER: What about looking at the diabetes trial as a model? To determine whether there is a difference in outcomes between men and women in relationship to therapy, instead of supporting a lot of little studies, why not conduct one good study that compares therapy for men and women who have HIV or AIDS? It seems more sensible than to have a bunch of small studies or little projects with a small percentage of women.

Q: I completely agree, but the reality, in many situations, is that a new medication or a combination of medications is being introduced, there is a lot more effort to involve women than men, and it is too costly. Researchers want to get the trial completed and figure out whether the new medication works; the resulting reality is that women are being left out.

DR. BRAWLEY: I brought up the issue of black/white,

and many people are talking about male/female in clinical trials; perhaps these are two separate issues. I have a lot more ease with the idea that we need to include women on clinical trials, we need to monitor that extremely carefully, and we need to focus on questions of women versus men. With racial or ethnic questions, it is a little less clear for me. For example, Dr. Pinn and I are very different ethnically, even though we are of the same race; that is my concern.

Q: Politically, we are all African Americans, regardless of what we are genetically. There is denial of services and restriction of services, and only through these studies have we uncovered these inequities. It is important to continue the outreach and the institutional leadership of inclusion. We are all different—we are heterogenic as African Americans—but if I walk into the emergency room with crushing chest pain in South Carolina, I might not get the services that will save my life that night.

Q: I want to reiterate an earlier comment: If we do not look for differences, we will not find them. That is a lesson we learned from all the women's health research in the past few years: If you do not look for it, you will not find it.

It is important to continue the outreach and the institutional leadership of inclusion.

Chapter One

Lessons Learned: The Importance of Recruitment and Imperative for Retention in Clinical Trials

Moderators: Marianne J. Legato, M.D., and Joan McGowan, Ph.D.

The speakers in this session were experienced veterans of clinical research studies who have successfully recruited and retained some extremely challenging populations in clinical studies. They spoke from a variety of experiences with different trials and studies and offered suggestions for achieving the specificity that needs to drive recruitment and retention efforts.

It is important to characterize the nature of the study—whether it is a prevention trial in healthy individuals who need to be motivated to participate when they are not sick and do not anticipate getting sick or a treatment trial in which individuals may question whether they are getting the “right” treatment. Meeting the informational needs of potential and actual research participants requires innovative approaches to initial presentation and to ongoing contact with those participants.

Participants in clinical studies are not a monolith and cannot all be approached in the same manner. The speakers presented examples of enrolling young women with acquired immunodeficiency syndrome (AIDS) in a long-term study as well as recruiting and retaining older women in a prevention study projected to be carried out for more than a decade. They offered approaches to gaining trust in African American and Native American communities and shared their insights into meeting the needs of participants to satisfy the needs of the study.

Enhancing Recruitment of Women Into AIDS Clinical Trials: Lessons Learned

Susan E. Cohn, M.D., M.P.H.

Dr. Cohn is Associate Professor of Medicine at the University of Rochester and former Chair of the AIDS Clinical Trials Group's (ACTG) Women's Health Committee. Her research experience focuses on women with human immunodeficiency virus (HIV), AIDS management, and clinical trials.

I will review the history of the HIV/AIDS epidemic and highlight the role of activism in the research process, focusing on factors leading up to the creation of the Women's Health Committee within the AIDS Clinical Trial Group (ACTG). I will then review the current situation in terms of HIV and women in the United States and then focus back on the ACTG: How well have we done in enrolling women into clinical trials, and what strategies have we put into place to try to improve what was less-than-optimal accrual? What are future directions and programs that we have and hope to have in place?

In 1981 the Centers for Disease Control and Prevention (CDC) warned about a rare disease that eventually became known as AIDS. By 1984 the virus that causes AIDS was identified and later named "HIV, the human immunodeficiency virus." It was 1985 when everyone got a "wake-up call"—Rock Hudson announced that he had AIDS, putting a personal face on this disease.

The first International AIDS Conference increased the awareness of HIV internationally as well as in the United States. At the same time, there was a rising concern in the AIDS community about the lack of leadership and delay in action in the face of a national health crisis. In response, in 1987 the Presidential Commission on the Human Immunodeficiency Virus Epidemic (the Watkins Commission) was established; at the same time, the first AIDS memorial quilt panel was created. In 1987 the U.S. Food and Drug Administration (FDA) approved zidovudine (AZT), the first antiviral agent for the treatment of AIDS. The AIDS Coalition To Unleash Power (ACT UP) was founded, and its motto—"We are not silent"—became a mantra for the time. ACT UP's "Silence = Death" buttons are reappearing as we look at the international epidemic of HIV.

In May 1988 Surgeon General C. Everett Koop distributed 170 million copies of "Understanding AIDS"

to every household in America. The pamphlet provided a clear description of what AIDS is and how it is and is not transmitted. By June 1988 the Presidential Commission on HIV issued its first report, which, to the surprise perhaps of no one, declared that the FDA was not meeting the needs of people with AIDS.

A few months later, on October 11, 1988, more than 1,000 ACT UP demonstrators virtually shut down operations at the FDA, demanding more effective treatment for AIDS. The ACT UP activists were invited by and met with Dr. Anthony S. Fauci, Director of the National Institute of Allergy and Infectious Diseases, and within 8 days, the FDA announced the fast-track drug approval process, revolutionizing the way drugs come to market for this disease.

In 1989 it was believed that more than 10 million people worldwide were already infected with HIV, and a study revealed that AZT not only prolongs the life of people with AIDS but also may slow the progression to full-blown AIDS.

In 1990 Ryan White, a young hemophiliac, died of AIDS, and President George Herbert Walker Bush signed the Ryan White Care Act and established the AIDS Drug Assistance Program, which has been crucially important for the care of people with HIV in our country. Still, there was community unrest. In May 1990 ACT UP New York organized a national storm on the NIH program. A thousand demonstrators demanded more treatment for AIDS and opportunistic infections, more representation of women and people of color in clinical trials, and the formation of a specific women's health committee within the ACTG to focus on these issues.

In 1990 one of the most talked-about documents at the International AIDS Conference was a treatment agenda put forth by activists, which pushed for a women's

health committee. NIAID officials met with women who pushed for a conference on women with HIV, a women's health committee within the ACTG, and a natural history study for women similar to the one that had been conducted for men; Dr. Fauci was very influential, and subsequently, all these requests have been met.

Clearly, there was good partial success. The ACTG actively restructured its organization. It began to incorporate patients and people at risk of AIDS as members of the ACTG Scientific Research Agenda Committee, protocol teams, and at the highest level in the Executive Committee. The ACTG involved affected populations at the local level through community advisory boards and nationally through a Community Constituency Group. However, it became quite clear that some HIV-affected groups, other than gay white men, were not well represented in these activities.

Another turning point in the epidemic came in 1991 when Magic Johnson announced that he had contracted HIV; this announcement significantly increased the awareness of heterosexual transmission of HIV. The red ribbon became an international symbol of AIDS that year, and the Women's Health Committee was formed.

In 1992 many advances took place. Several new drugs were released, including DDC and D4T, which went through the new fast-track mechanism. The definition of AIDS was revised to encompass more of the people who actually had the disease and include some gender-specific diagnoses, specifically cervical cancer. When the AIDS quilt was finally displayed here in Washington, it covered 15 acres, there were 22,000 panels, and a third of a million people came to view it.

In 1994 the landmark study within the ACTG 076, which the Women's Health Committee helped orchestrate and push to completion, showed that AZT was able to reduce mother-to-child transmission from 25 percent to 8 percent, a 66 percent reduction. In 1995 other new drugs were released, including 3TC, which is used heavily today, and the first protease inhibitor (saquinavir), opening up a whole new class of antiretroviral agents. The approval of the viral load test and the ability to detect HIV in the bloodstream advanced the understanding of whether these antiretroviral regimens would be effective. Increasingly, however, the population that was becoming HIV-infected was more disenfranchised and much less

organized and was clearly not benefiting from the initial swell in community activism and services seen in the early years of the epidemic. Women did not have access to medications and had little access to clinical trials, and there was clearly a need for more organization within the various groups of clinical trials, such as the Women's Health Committee, to advance that agenda.

By the end of 2001 (and probably these numbers will be revised upward), there were more than 40 million adults living with HIV worldwide. The prevalence is clearly higher in certain areas of Africa, but areas such as India and Southeast Asia may uncover even more cases in the decade to come. Globally, 1 percent of the population between the ages of 15 and 49 is HIV-infected, and of the 5 million people newly infected in 2001, more than half were women. Not surprisingly, more than 95 percent of the newly infected are in resource-poor countries.

The incidence of AIDS in women in the United States has been increasing dramatically over time. In the early part of the epidemic, before 1985, only 7 percent of all the AIDS cases were women; in 2001 about 25 percent were women. As of 2002 there were more than 816,000 cases of AIDS reported to the CDC. About one-third of people newly identified with HIV are women. This disease has primarily affected women of color, of diverse backgrounds, and of disenfranchised populations; 82 percent of women with HIV are nonwhite compared with 66 percent of men.

More importantly, this is a disease that is heterosexually acquired, both here and around the world; 64 percent of women are thought to have acquired the disease heterosexually. Many of the women who had reported intravenous drug use presumably also acquired the disease heterosexually. The incidence of women with HIV is highest in the South and the northeast areas of the United States, but women with HIV have been reported in every State.

How successful have we been in accruing women into clinical trials within the ACTG? Many more men than women have enrolled in clinical trials, but in 2002 more women enrolled in clinical trials than ever before. Twenty percent of the women were in sex-specific (women-only) studies involving pregnancy and gynecologic evaluations. Even not counting those women, many more women were enrolled last year than in previous years, but clearly we have a long way to go to be reflective of the prevalence

We hope to enhance enrollment of women in AIDS clinical trials and to improve our understanding of how best to manage HIV in women.

of HIV in our country. Perhaps a better indicator is how we are doing in terms of enrolling women into larger ACTG studies that investigate primary therapy for HIV. In all the larger studies, less than 20 percent of participants have been women, and the current antiretroviral therapy trial (A5095), which just completed enrollment and is investigating various initial drug regimens, enrolled only 18 percent women.

What has the ACTG done to try to enhance the accrual of women into clinical trials? First, we wanted to understand the barriers: Why were the sites not able to enroll women into trials, and what were some of the difficulties they were encountering? We also made an effort to revise the Women's Health Committee agenda and to devise some creative strategies to enhance the recruitment and retention of women into clinical trials.

First, we organized a symposium in July 2000 that brought together the pediatric and adult ACTG populations and examined the barriers to recruitment and retention and what we could do to develop creative solutions to this problem. This symposium reviewed site surveys that examined barriers at the site level. We listened to HIV-affected women: They addressed what they saw as the problems, why women were not enrolling, what studies they would find of interest, and in which studies they would be lining up to enroll. We also heard from people in other research clinical trial groups who addressed their understanding of why they have been successful over the years at enrolling and retaining women in their studies. Finally, we distributed our recommendations to the Executive Committee and also distributed them broadly. We also expanded the Women's Health Committee's website and posted these recommendations on the Web.

What were the potential barriers to enrolling women? Many of these barriers are true for almost every clinical study involving women. Some barriers were relatively easy to address, such as making clinic hours more convenient, arranging transportation and childcare, and providing onsite gynecological services so women did not have to go to three different clinics to get the care and study supervision they needed. Other barriers have been much more difficult to address, such as mistrust of the health care system.

In terms of revising the Women's Health Committee agenda, we wanted to make this a group process. We solicited input from all the Scientific Research Agenda Committees within the ACTG and the scientific committees, specifically the Pharmacology Committee, which has been very influential in furthering our understanding of the pharmacokinetics and pharmacodynamics that appear to differ at times between women and men. We also

involved the at-risk populations through the Community Constituency Group and the Patient Care Committees. We had open conference calls, and we arranged an open, interactive session with the chairs of these large committees. Through this process, we revised the Women's Health Committee agenda, which was then widely distributed.

By involving the community, we turned the corner; instead of hearing "You're not doing enough," we heard "We need to do more." It became a group process that was much less antagonistic. For those who are involved in this field, that was a significant turning point.

One of the solutions was to develop a request for proposals from sites to test creative strategies for enhancing recruitment into clinical trials. These proposals had to be hypothesis-driven, have an evaluation component, and ideally be self-sustaining, with the potential for applicability to other sites. The proposals would have an average budget of about \$25,000 for the 1-year intervention. It required a little rallying, but we worked with the Executive Committee, and by December 2001 the Committee agreed to allow us to put forth this recruitment proposal. It was released in spring 2002; a dozen proposals were submitted last summer, four sites were selected for funding, and representatives of the four sites came to our meeting in December and presented their proposals. We look forward to hearing about their progress.

In terms of future directions, the Women's Health Committee has a large agenda with lots of work that needs to be done. The ongoing areas of interest clearly have been in the management of women with HIV, especially during the period when they may be pregnant and beyond. The safety of the various antiretroviral agents for women is another important area, and understanding more about the metabolic abnormalities and body changes that occur is clearly a huge issue for women and may deter them from even starting antiretroviral therapies. An area that may be more problematic for women is breast enlargement, which has become a research agenda item. We have also looked at the issues of adherence and treatment outcomes for women.

This epidemic is international; it is not just an epidemic within the United States. Providing antiretroviral therapy internationally is an important issue, as are preventing heterosexual transmission and preventing mother-to-child transmission.

In summary, with the cooperation of researchers and the affected community, we hope to enhance the enrollment of women in AIDS clinical trials and to improve our understanding of how best to manage HIV in women as well as men.

Recruitment and Retention: Lessons Learned From the Women's Health Initiative

Marian C. Limacher, M.D.

Dr. Limacher is Professor of Medicine in the Division of Cardiovascular Medicine at the University of Florida and past president of the American Society for Preventive Cardiology and current president of the Society of Geriatric Cardiology. She is the principal investigator of the Women's Health Initiative site at the University of Florida

I will review the experience of the Women's Health Initiative (WHI) in recruiting an incredibly large number of women into the four clinical trials and the observational study (OS) of the WHI. Initially, we recruited for two clinical trials: a large dietary modification study enrolling 48,000 women and a hormone study enrolling 27,000. The principal outcome for the diet study is the occurrence of breast cancer, and the secondary outcome is colorectal cancer with cardiovascular disease. For the hormone study, the primary end point is cardiac disease, with the secondary outcomes of hip fracture and breast cancer. The hormone study comprises two separate clinical trials: estrogen with and without progesterone. The total enrollment of the dietary and hormone studies is 68,000 women, with an 11 percent overlap of women participating in both trials.

The WHI OS, which is complementary to the clinical trials, enrolled 93,000 women. Women screened for either of the clinical trials at entry could enroll in the OS if they were not eligible for the clinical trial or if they chose not to participate in a randomized intervention and still wanted to be involved in some part of the process. In the OS and clinical trials combined, more than 161,000 were enrolled. In addition, at 1 year into either of the primary clinical trials, participants were invited to participate in a randomized study of calcium plus vitamin D. The full WHI study, therefore, includes four clinical trials plus the OS.

Enrollment goals were highly ambitious and complex. No one had ever tried to enroll this many women, and enrollment success was deemed not possible. We also worked very hard to enroll a diverse sample representing the U.S. population. Recruitment was challenged by the long duration of the planned studies, from 8 to 10 years.

Recruitment began in an environment of controversy and scrutiny, which relaxed later on. There were budgetary issues—the \$700 million was not the initial budget. On the other hand, there was considerable support, particularly from the Congressional Caucus for Women's Issues, the National Institutes of Health (NIH), consumer advocates, and at least some elements of the medical community.

No one had ever tried to enroll this many women, and enrollment success was deemed not possible.

The national recruitment activities were coordinated through both the project office at the NIH and the Clinical Coordinating Center (CCC) at the Fred Hutchinson Cancer Center in Seattle, Washington. The contract activities included the funding of recruitment incentives. Budgets were tied to recruitment goals, which were highlighted in a number of monitoring areas, and progress was tracked regularly. There was a national public relations hotline. National outreach efforts through magazine, television, and other ads were referred to a toll-free telephone number; the responses were then directed by area code to a local recruiting clinical center. Project coordination was conducted through the NIH project office as well through the investigative team and the administrative structure of the study.

The CCC was critical throughout the recruitment phase and continues to play a key role. The CCC ensured that the recruitment materials were identical across recruiting sites, distributing them to the clinical sites so that the sites did not have to develop their own materials. The CCC conducted central training of recruitment

coordinators and then shared the successes and communicated what was working and what was not working in monthly regional phone calls. Monthly progress reports were distributed.

The most comprehensive national publicity campaign was initiated not at onset but several years into the study. Numerous public service announcements (PSAs) were developed and distributed to national media outlets. The brochures were tweaked throughout the recruitment period, and posters and postcards were developed. Consent and recruitment videos were produced and distributed to the sites for use, and celebrity spokespersons were enlisted. One national spokesperson, Angela Lansbury, did a PSA, and for the New York-New Jersey area, Olympia Dukakis made a PSA.

These activities occurred across 40 study sites. The sites were distributed to reach a representative population, with the caveat that we were looking for specific populations to recruit; for example, Pacific Islanders have been heavily recruited through the Hawaii center, and at least one site focused on the recruitment of Arizona Native Americans. Several recruitment sites concentrate on Hispanics/Latinas. A total of 10 minority recruitment sites were established for African Americans and other less represented populations.

Recruitment began in 1993, when there were 16 sites. They were labeled the “Vanguard Clinical Centers,” and they truly were at the forefront. They had to not only sort out the protocol issues but also begin recruitment. In early 1995, 24 additional sites began recruiting. All the recruitment was completed by the end of 1998, with recruitment for the clinical trial completed in August and for the OS in December.

We advertised to recruit individuals who would understand that their participation would be a legacy for their daughters and granddaughters.

How did it happen? Mailings, mailings, mailings. Distinct from many clinical trials, the WHI was not disease-specific, so we were not looking for physicians and providers to recruit patients to this wonderful study. That would have been great but in fact did not happen; it was an individual woman’s decision whether to participate in a 10-year study of her health. Physician referral yielded very few participants.

We wanted to market directly to the participants, to the older women, and that was accomplished primarily using direct mail, supported by newspaper articles and

ads. All the WHI researchers placed recruitment ads, but there are so many research recruitment ads that you have to consider carefully whether they will attract the kind of people you are interested in. We were not paying people to participate, but the public-interest stories that we were able to get from the newspapers were quite supportive.

The TV and radio ads, articles, and PSAs were supplementary; they did not yield a large number of participants, but they kept our name and presence in front of the public. We developed incentives for the participants to help us recruit friends, neighbors, and so forth. We put brochures everywhere, blanketing communities—in clinics, health care systems, retirement communities, supermarkets, and beauty salons—every place we could think of. In addition, all of us made many presentations, speaking to all the women’s groups in our communities. Most were very receptive, and these presentations were quite effective.

What messages worked for us? First, I think the motto of the whole study, “Be part of the answer,” has been quite effective. The message was to “Participate in something that will be meaningful, look at the gap in knowledge—you can be part of this.” Second, we did not emphasize that there would be personal benefit. This of course leads to selection bias—people who enrolled in the WHI were not individuals seeking to gain from participation in the studies. So we advertised aiming to recruit individuals who would understand that their participation would be a legacy for their daughters and granddaughters. The other message that worked is related to the first: “It’s about time that we are studying women in this very important area.”

How did we do the mailings? We developed lists—drivers license lists from the Department of Motor Vehicles, lists from the Health Care Financing Administration (now the Centers for Medicare & Medicaid Services), voter registration lists, and commercial lists from marketing groups, health maintenance organization members, previous clinical research participants, and other organizations. We sometimes sent brochures only, but usually with a cover letter, all of which were approved by our institutional review boards. The mailings were sent out monthly for a steady flow, with most sites mailing between 1,000 and 5,000 brochures per month for the entire recruitment period; in the final months, some sites mailed up to 50,000 per month. This was expensive, but it was the way to recruit these women.

Response rates from the mailings ranged from 2 percent to 20 percent, with 20 percent being rare. The initial response rates were in the 5 percent to 10 percent range,

and then there was a steady 1 percent to 2 percent response rate during the final years, even from repeat recipients.

It is key in any clinical study to have a high retention rate throughout the study years, so the initial contact was with a recruitment brochure. Candidates responded by sending in a filled-out, prepaid postcard or by telephoning to schedule a brief telephone screening. We call that a first screening visit, an SV-0. An SV-1 was scheduled as an in-person visit in the clinic. We ascertained interest and global eligibility with these two screenings and then scheduled additional, specific screening visits that included some laboratory testing for those who met at least basic entry criteria.

The goal was to make several screening visits to figure out who would be able to stick with the program, rather than enrolling everybody at SV-1. In fact, our biggest dropout was from SV-0 to SV-1. We did not want to retain individuals who were not truly committed. The goal of the screening visits was to have very little dropout after SV-3, and in fact that is how it worked out. The last screening visit to the randomization resulted in very little dropout.

The trajectory of enrollment for all of the studies started out slow, then had rapid pickup through completion. The actual and projected enrollments came quite close. The OS targeted 100,000 participants and enrolled 93,000, a number we were quite satisfied with. We met the target for the dietary and hormone studies, but not for the calcium study, although enrollment was within 10,000 women. It was easier to recruit younger women than older women, but we met our goal, even for the oldest age group. The oldest women responded better to the mailings than the younger women, and the youngest age group responded more to newspaper articles. By race and ethnicity, we came very close to the goal of approximately 20 percent of nonwhite individuals for both the clinical trials and the OS. Recruitment strategies were effective across ethnicities, with mailings delivering most recruits.

Reasons for exclusion are worth examining. Of the total percentage eligible, those who were not interested or did not provide consent were weighted to the hormone study and to the OS. The hormone study required a certain kind of individual, one who understood the research process and was not already wedded to an opinion, for example, that hormones are bad and are going to cause cancer or “hormones are great; I’m never coming off them.” Either of those two types would not work out in a randomized study. Some of the subsequent dropout is attributable to those strongly held opinions.

The OS was quite interesting in that no intervention was involved, and yet a large number of women were not

interested in participating. Dropout from the diet study is also interesting. Any study called “dietary modification” is interpreted as “This is going to help me lose weight,” even though the Dietary Modification Clinical Trial of the WHI is not a weight loss study. Fifty percent were not willing to be randomized. The primary reason for exclusion from the diet study was that individuals did not meet the food frequency eligibility of percentage of fat from calories. We specified that women could not have too little body fat or could not have too much fat in their diets.

It is key in any clinical study to have a high retention rate throughout the study years.

We implemented retention strategies by holding events at each clinic—study “birthday” parties and celebrations. We wanted to acknowledge the contributions of these women and did so on an annual basis at local clinics.

We have prepared a manuscript on recruitment that will be part of a monograph on the initiation of the WHI. Publication of the monograph will be announced on the WHI website.

Recruitment and Retention of African Americans in Clinical Studies: Fiction, Myths, and Reality

Kwame Osei, M.D.

Dr. Osei is Professor of Medicine and Exercise Physiology at the Ohio State University and Director of the Division of Endocrinology and Diabetes and Metabolism.

I am going to reveal my naivete in trying to do a prevention trial. In 1993 we submitted a proposal to the NIH to study the use of oral antidiabetic medications and prevention of diabetes in African Americans. The study was funded, so we called our patients and said, “Your mom has diabetes, and your dad has diabetes; we want you to come to our study.” We thought that would be a slam-dunk, but it never happened. In the first encounter, people said, “Why do I have to be in the study?” So that is when we learned, very quickly, that we had to prepare a whole brochure that would cover the details. We went back to the drawing board.

We designed a brochure that outlines why African Americans and minorities should be part of a study. We explained morbidity and mortality and the prevalence of the disease among African Americans: The mortality and morbidity data show that the chances that an African American will end up on kidney dialysis or die from cardiovascular disease is much higher if they have diabetes. We explained that diabetes is genetic, and therefore, because their parents have diabetes, they are prone to become diabetic.

Environmental factors can be changed to lessen the risk for diabetes—factors of nutrition, exercise, lifestyle, and stress—things they could relate to. We explained about how the disease varies from low to high socioeconomic groups and about social behaviors that can affect the disease. In the brochure, we outlined these in bold, highlighted language they could understand. We addressed questions about constitutional and biological

differences and the many variables of genetic factors. They needed information to understand why they should participate in the study.

The first question they had asked us was why we were going to use different drugs, so we explained that we did not know which drugs would work for which individuals. Three drugs would be administered, and data about one of these—thiazolidinedione (TZD)—were available, so TZD would be given as a comparison base for the effectiveness of the new drugs.

We learned very quickly that there is a lack of understanding of the ethnic differences relevant to the disease of diabetes. Everybody said, “We have diabetes; all populations have diabetes; therefore it is the same disease.” We wanted to impress on them that there may be differences and therefore that we had to do studies that could distinguish the differences so that we could tailor our therapy specifically to African Americans. We wanted to debunk the concept that “one size fits all”—that if we do a study with one population, we can apply the results to another race or another ethnic population. We believed that the one-size-fits-all hypothesis was unethical at that point, and no prevention trial had examined the racial differences.

Another issue that we needed to address was lack of access to health care. We wanted to give study participants the opportunity to bring their families to our facility, so we used that as a tool to help us retain them because we believed that if our study participants did not buy into the family concept, the prevention study would not work.

There is a lack of understanding of the ethnic differences relevant to diabetes.

Is it true that African Americans and minorities do not participate in clinical trials? I was invited to give a lecture on this topic, so I asked investigators of clinical trials how many minority people participate in their studies. I was surprised that only 10 percent of participants were minorities. So, if only 10 percent of participants in a diabetes trial were minorities, it might not be correct to extrapolate the resulting data to minority populations.

We wanted to know why African Americans did not want to be in the study. An important issue was mistrust of medical research, especially in prevention trials. The

An important issue was mistrust of medical research, especially in prevention trials. The community must endorse the project...and...family endorsement is critical.

suspicion was, “If I don’t have the disease and you want to prevent it, then are you going to kill me or use me as a guinea pig?” That was a very common statement from many individuals. Lack of feasibility in the research protocols was also a major issue. You tell the subjects to come in at 9:00 a.m., and if they are not there on time, they have to go home because they are late. If you do that, they are not going to show up. The traditional concept is: We have a study, we invite you, and because we are an “ivory tower,” you come to us. This concept does not work; you must go to the community to be able to include everyone.

Lack of flexibility in work schedules was another major issue for us, especially for young African American men whose work schedules do not allow them to get to the facility at 8:00 a.m. We should be able to make arrangements for them to come at 7:00 a.m., so arranging staff schedules to accommodate participants was very important to us.

Young African American men (ages 20 to 30) did not want to participate in this study for several reasons. They had other issues they were dealing with, and that group needs to be considered before you design a study. We found that 60-year-olds were much easier to recruit.

Also important but rarely discussed is the race or ethnicity of the principal investigator (PI). We did not even think it was an issue, but it became a critical issue. Who is the person who is trying to do this study? What is his name? Does he have a foreign name, like Kwame or Osei? It was clear to us that potential participants took

into account who the PI was; this became evident as we went out to the community. They wanted to know who we were and what we would do in their community. If you have never been in a particular community, explain who you are and what you do in the community. In our case, because we are in the community, they knew us, they knew our names, and they knew our program.

What is the race or ethnicity of the staff? We did not think in the beginning that that was going to be an issue. When we started sending people out to the community, they were told, “Oh yes, I’ll call you, but don’t call me.” So we needed to ensure that our staff would be able to accommodate that difference and have racial and cultural sensitivity to the community. The NIH’s ethnospecific literature about the disease has been helpful.

A number of elements were important in our study design. One thing we learned very quickly is that, for a long-term study such as a prevention trial, ownership is important. The group has to feel that this is their study. The pride of participating in that study was even more important to participants than was the outcome: “We want to be a part, we want to contribute, we want to make history.” Ownership was important to the community.

The community must endorse the project. If you go into the community and people say, “Don’t do it,” “our older brother said don’t,” “sister says don’t do it,” or “the church leader says don’t do it,” then forget it. Do not go to that community; your study will not happen. The community has to endorse your study, and community facilities such as barbershops will be an influence. You would be surprised what they say about your project while getting their hair done!

Advocate for the project. You have to have people out there who are speaking for you, because you are not present in the street all the time. You need a group of people to do that, and their advocacy has to be culturally sensitive.

Family endorsement is critical. If the family says no, you can do whatever you want, but your study will not be able to recruit. The family has to say “yes” to the project. It is almost a family thing to participate in a long-term trial because somebody has to take care of the baby when the participant goes to have blood drawn, somebody may have to drive the participant there; the family has to do that.

When we started, we thought everything would be fine if we went to the churches, but we now know that not all African Americans go to church. Some go to synagogues and some go to other places to worship. You have to find where they are, and you have to work with them at those places.

I want to highlight a few things that improve participation in clinical trials. Language is critical—what you say and how you say it mean so much to different

people. Financial incentives are important—they have to pay their bills, they have to come in, they have to ride the bus—so that is very important in terms of anticipating what is going to happen.

Community activities are important. Every year we have a brunch for the whole group and their family members. We go out to the different organizations, we solicit money, and we have a whole group. We just finished the last one 2 weeks ago at which a State representative gave a talk to a whole group of our participants, telling them about the economics and health care policies in Columbus, Ohio. Every year we bring a national speaker or a State speaker to talk to our study participants about what we are doing in the community. That is part of keeping them in the loop.

We created logos, project T-shirts, and other things that say the participants are part of history. We designed some of these things to help them understand what we were doing and why they should be part of our study. We also send out newsletters, a very important activity. Every 3 to 6 months, our study participants receive some information about exercise, diet, and diabetes among African Americans.

We had to face the big question of where to put the study. A prevention trial can be done very easily in a military institution, or it could be conducted in a community. Where the study is done generates a certain kind of data, with different interpretations and different implications for the community. There are advantages and disadvantages associated with institution-based

projects such as military bases, veterans hospitals, or churches. Advantages include the cultural diversity of the population, subjects who are onsite, and subjects who are a known entity; the subjects already come to the same place, and they know the building and where to go in that building to attend the research study. The setting allows flexibility for scheduling, because participants might not have to travel anywhere, depending on the institutional setting. Participants from an institution might have medical insurance. There are also disadvantages. The data cannot be generalized because they are skewed and are only applicable to certain populations. One thing we learned was that the faculty members at our university who participated in studies would go on the Web to learn about the drugs, including the side effects, being used in the studies.

Studies using a community base, on the other hand, yield data that are generalizable because the population is in the real world, across a spectrum of socioeconomic attributes. The disadvantages are that they are very expensive and staff members must be trained to perform the same responsibilities and duties at all study sites.

African Americans want to be part of research studies; there is no question about that, and I have seen this over and over. But researchers have to understand the population they want to recruit to their community studies and how to do the recruiting. If African Americans understand the research and know who the researchers are, they will participate in the study.

Researchers have to understand the population they want to recruit to their community studies and how to do the recruiting.

Recruitment and Retention in American Indian/Alaska Native Communities

The opinions expressed in this paper are those of the author and do not necessarily reflect the views of the Indian Health Service.

Barbara V. Howard, Ph.D.

Dr. Howard is President of MedStar Research Institute and a principal investigator in the Women's Health Initiative. She is former Associate Chief of the Phoenix Clinical Diabetes and Nutrition Section of the National Institute of Diabetes and Digestive and Kidney Diseases and is Chair of the Steering Committee of the Strong Heart Study.

I have been working on studies in American Indian and Alaska Native communities since 1976. There are about 2 million American Indians/Alaska Natives in the United States, and this is truly an underserved, underprivileged population group with multiple health problems. It is also a group that is largely rural, and therefore, our experience will be applicable to research initiatives in a number of other smaller, ethnically diverse populations in the United States and perhaps in other countries.

I will describe my experience with two studies. The Strong Heart Study (STS), funded by the National Heart, Lung, and Blood Institute (NHLBI), has been in progress since 1988. It is a population-based longitudinal study involving 4,549 men and women, 45 to 74 years old, in 13 communities, and 2,800 members of their families. The Genetics of Coronary Artery Disease in Alaska Natives (GOCADAN) study began 3 years ago in the Norton Sound area of Alaska and involves 15 remote Alaskan villages.

Geography is a significant issue. Three of the STS communities are located in the Sonoran Desert. There are no street signs and no house numbers, and the temperature reaches 120 degrees Fahrenheit in the summer. Three of our Sioux Indian communities in the STS live in the Badlands of South Dakota; again, there

are no street signs or house numbers. Nome, Alaska, is another location, but some of the villages are 1,000 miles from the town. Another issue is poverty, with very difficult living conditions throughout the communities.

Despite these issues, we have had enormous success. In the STS, we averaged 60 percent participation, ranging from 50 percent to 70 percent for population-based recruitment and about 90 percent retention at exams two and three (see table below). Almost no one was lost to followup. For the current funding period for the STS, we are in the process of recruiting 120 families and, again, are having great success.

Recruitment starts slowly in new communities. For the GOCADAN recruitment, we started only a year and half ago. Our goal is to recruit 1,200 adults (10 per week) in Inupiak families in 15 villages; at this point, we have recruited 650 adults. In the previously recruited village, we achieved 100 percent of our goal.

How did we succeed in reaching our recruiting goals? Simply put, we involved the community. It is essential in studies involving communities with either socio-economic or geographic barriers to have continuous community input from the beginning, have as many community members as possible on the staff,

Strong Heart Study Recruitment and Retention (45-74 Yrs Old at Baseline)

	AZ	OK	ND/SD	Total
Exam 1				
Men, N	547	642	660	1849
Women, N	953	885	862	2700
Total, N	1500	1527	1522	4549
Participation, %	71	62	53	62
Exams 2 and 3				
Participation, %	90	89	87	88
Lost to Followup (N)	28	8	51	15

and provide continuous feedback to the community and the participants.

We started the planning process early in both the STS and the GOCADAN study. It takes time for communities to get to know the investigators, to be able to work together, and to understand the objectives and the relevance of the objectives to the community's goals and health problems. We also were very careful to explain the procedures. There was a lot of fear and lack of understanding about what we were doing, and reviewing the questionnaires for understanding and relevance was very important.

Both the GOCADAN study and the STS focus on genetics. Because the word "genetics" engenders fear, we have developed a community-based educational program that we present at community meetings. At those meetings, we explain genetics, deoxyribonucleic acid (DNA), and what it all means for the individual. We emphasize that we will not clone anyone and that we will not look into personal histories to uncover things the individual wants to remain private. Community involvement must be present from the beginning, and this involvement takes time and patience to develop.

We relied heavily on the community for recruitment and retention. All of our recruiters are members of the community; tribal leaders played an enormous role in helping us. Our brochures are community-specific, and we conducted many home visits. We also used local media and community activities for recruitment.

It is also important to involve community members as investigators in the study. When I started in 1976, there were only 25 American Indian physicians and probably fewer Ph.D.s. Now there are many more, and they have become increasingly active as investigators; they are on the steering committees of both studies. I hope that next time there is a talk, one of them will be giving it.

Virtually all of the staff members for these studies are community members. Higher education is not necessary. Even without a high school diploma, people in the

community can learn how to do most of the things done in a typical epidemiologic study or a simple trial.

We also worked very hard to deliver health education to participants. That helps with both recruitment and retention because they feel that they learn something and that we are concerned about them. We have also tried to provide opportunities for community students to go into health care careers via sponsored mentorships and pre-sectorships and to get youths involved at an early age. Most of them do not know that they can have careers in this area, and spending the summer or winter vacation with us can make a difference.

It is very important in an underserved community to report study results to the participants and help them get followup if health care problems are discovered. Residents in some of our communities had not seen a health care provider for 20 years before we arrived to do the study. Obviously, followup for problems was extremely important. We report the results to the communities so that they can use them for health care planning. To reach these communities, it is important to publish the study results not just in prestigious journals but also in publications read by the primary health care providers in the community.

Incentives are important. If you are struggling to survive, a T-shirt can be really valuable; it also is a great public relations tool. We have tried to choose incentives that are useful to the individuals who participate in our studies.

In summary, the bottom line always is: Ask the advice of the community and then take it, even if it is not what you want to hear. Community members and researchers have to be involved in the decisionmaking, even if at first you are not happy with the decision. Translate the findings into the health care environment as quickly as possible. Maximize the opportunities for training community members for careers in the research and medical fields. We are not going to get where we need to go in research unless we gain the interest of children in grammar school and high school and a diversity of individuals in becoming researchers.

**Ask the advice of the community and then
take it, even if it is not what you want to hear.**

Chapter Two

Lessons Learned: The Importance of Recruitment and Imperative for Retention in Prevention Trials and Longitudinal Studies

Moderator: Amelie G. Ramirez, Dr.P.H.

This panel session examined various issues affecting the success of recruitment and retention in two diverse types of clinical research studies. From the perspective of an investigator assessing breast cancer chemoprevention in multiple sites across the country, one presenter addressed the sources of subject accrual and both effective and ineffective minority recruitment strategies. From the perspective of an investigator in a longitudinal cohort study, a second presenter addressed the challenges of minority recruitment and retention in such research and the implications of these challenges with regard to the generalizability and bias of study findings. In both cases, recommendations were offered based on the lessons learned in investigators' efforts to meet accrual demands and maintain the integrity of their research. Both speakers were eminently qualified to address these issues, not only because of their academic and research credentials and backgrounds but also, and more importantly, their personal involvement with large-scale studies reflecting the two divergent research perspectives. Given the expertise and experience of the panel, expectations were met regarding insight into approaches that work (as well as those that are less successful) in recruiting and retaining study subjects, and particularly minority accrual.

The wide-ranging and relevant concurrent issues that emerged included the role in accrual played by individual oncologists and other physicians as well as community and clinical oncology programs (and entities such as corporations and churches); the need for change and improvement in physician education, clinical trial public image, and promotional efforts regarding cancer prevention; the strong impact of the broad array of socioeconomic factors on study participation; the influence and implications of culture on clinical trial participation as well as various facets of research, such as deoxyribonucleic acid (DNA) collection; and the inherent pitfalls in lumping minority populations together in study conclusions as well as in developing recruiting strategies. The net effect of these presentations was an insightful and richly detailed illustration of approaches and perspectives regarding clinical trials and longitudinal cohort studies and acknowledgments that more research and resources are needed. To effectively recruit and maintain special populations in clinical trials, cultural sensitivity to individual population groups must be emphasized, limitations and barriers recognized and addressed, and participation incentives increased.

Assessing the Building Blocks: Lessons Learned From Recruitment and Retention in Cancer Prevention Trials

Worta McCaskill-Stevens, M.D.

Dr. McCaskill-Stevens is a medial oncologist and Program Director of the Community Oncology and Prevention Trials Research Group in the Division of Cancer Prevention at the National Cancer Institute (NCI). She is also Program Director of the NCI-funded National Surgical Adjuvant Breast and Bowel Project's Study of Tamoxifen and Raloxifene (STAR) Trial breast cancer prevention study and the Minority-Based Community Clinical Oncology Program.

Our understanding of the biology of carcinogenesis is rapidly expanding, through programs that identify individuals at high risk of disease and through scientifically sound strategies that result from experimental and clinical research. As a result of recent increases in knowledge, approaches used for clinical trials that are designed for disease prevention and to reduce the cancer burden have been validated. From the significant building blocks cited during the Recruitment of Minorities in Clinical Trials held 6 years ago, we can now look back to assess the role of these building blocks and lessons learned from large prevention trials. Three trials will be assessed to determine the lessons we have learned—one screening and two chemoprevention studies.

The first trial is the Prostate, Lung, Colorectal, and Ovarian (PLCO) Screening Trial. The four cancers targeted for intervention are major sources of cancer incidence and mortality in the United States. The primary goal of the study is to assess the impact of screening tests on cause-specific mortality; secondary goals include looking at sensitivity, specificity, and positive predictive values as well as early markers of cancers, etiology, and some ancillary studies.

This study was a large Phase III randomized trial conducted in 10 clinical centers, most of which were large academic centers throughout the country. The protocol for the study included 74,000 women, 37,000 of whom were randomized to the intervention arm. This study compared some screening techniques with the control arm; the 37,000 women in the control arm received routine medical care. For ovarian cancer detection, women were going to have the CA125 test and a transvaginal ultrasound; chest x-ray was the screening modality for lung cancer; and flexible sigmoidoscopy at baseline and at 5 years was also used for screening

purposes. Age eligibility for this trial was 55 to 74 years, and 13 years of followup was planned.

The last individual was accrued to this trial in July 2001. There was an interim analysis last year, and there will be a report and an analysis in 2015, which seems like a very long period of time. The trial was approved in 1991 in the Division of Cancer Prevention, and of significance, it was approved for the concomitant development of a biorepository.

There are some interesting elements to this trial. Approximately 11 percent of enrollees were minorities and about 85 percent were Caucasian. This trial had a pilot recruitment phase in which the participating clinical centers were given an accrual target, which was met successfully. This phase gave the centers time to assess their communities for recruitment and to see how smoothly the interventions could be done and how receptive the population would be to them. Although the clinical centers were successful in obtaining their accrual goals in the pilot stage, this did not necessarily translate at the end, such that the trial was expanded or extended so that the accrual goals could be met.

An eligibility change at the beginning of the trial was of interest: Women who did not have intact ovaries were not allowed to participate. Because the other screening organs involved were the lung and colon, it was quickly noted that this was not going to allow the type of power that was needed to meet the accrual for those organ sites. Therefore, this change was made in 1993, which meant that this trial looked at the whole population. These were women ages 55 to 74 for whom hysterectomy rates ranged between 40 percent and 50 percent. This was an interesting observation, and it is reported that the inclusion of women who did not have intact ovaries had a dramatic impact on the successful accrual and accounted for researchers meeting their goals for recruitment of women in this trial.

If we look at recruitment strategies for this trial, direct mail was quite successful. It has been reported before that direct mail is a successful strategy in screening trials, and although the sites had a pilot period in which they were creative in looking in their communities and discovering what works specifically for the target population, all sites reverted to mass mailings in the end. Most of the sites contracted with mailing houses to reach the very large numbers necessary for this trial.

This trial was important in that 82 percent of participants consented to give a specimen to be added to the biorepository; this is a key component of prevention trials. There are no gender differences in consenting for the biorepository; although some sites had more men than women, overall the consenting was split evenly between the sexes.

I would like to talk about the chemoprevention breast trials. In the mid 1980s, several large randomized trials began both in this country and elsewhere, looking at reduction in the incidence of breast cancer. The rationale for all of this was that the “godfather” of the selected estrogen receptive modulators, tamoxifen, had been found to inhibit the carcinogenic process in breast cancer in animals and in humans. In addition to having a wealth of information from treatment trials about tamoxifen, we found that it could reduce the risks of contralateral breast cancer in a cohort of women who were at increased risk of developing a second primary, having had a primary diagnosis. This result was shown in women who were treated either with placebo or tamoxifen, and there was a 47 percent reduction in the risk of contralateral breast cancer. This was a very exciting rationale to take into the chemoprevention arena.

Chemoprevention trials use either a synthetic or natural compound to try to reverse or suppress the carcinogenic process. Using the rationale of tamoxifen’s success in contralateral reduction, the first U.S. breast cancer chemoprevention trial was designed. We had an agent, so now we needed a cohort, and the cohort of women for this trial were women at increased risk of developing breast cancer. The model used was developed by NCI’s Mitchell H. Gail, M.D., Ph.D., in 1986; this model takes into consideration previous biopsies, a woman’s estrogenic exposure throughout her life, and other variables.

The threshold criterion for being at increased risk was a 1.66 predicted risk over 5 years for developing breast cancer, the average rate for an average 60-year-old

woman. This trial was conducted by the National Surgical Adjuvant and Bowel Project (NSABP), a research group funded by the NCI. It was open to both premenopausal and postmenopausal women, and it randomized thousands of women to receive either tamoxifen or a placebo for 5 years.

This trial was conducted in about 90 primary centers and totaled almost 200, including the subcenters, in the United States, Canada, and Puerto Rico. The objectives were to see whether tamoxifen could effectively reduce the incidence of invasive breast cancer, breast cancer mortality, cardiovascular mortality, and bone fractures. The rationale for this was that the “sera” are rather schizophrenic in that they contain both estrogenic and antiestrogenic components, depending on the organ target.

Results of this trial, published in 1998, show a 49 percent reduction in the risk of developing invasive breast cancer across all of the women with varied 5-year predicted risk as well as age ranges. There was a 70 percent reduction in estrogen-receptive positive breast cancer and a 50 percent reduction in noninvasive breast cancer, and this was the first randomized trial to demonstrate such reductions. Subsequent ancillary studies have also shown that 7 percent of those who had permitted a sample were breast cancer 1 gene (BRCA1) rotation carriers and that, not surprisingly, tamoxifen only affected the BRCA2 carriers.

More than 60 percent of participants were older than age 60, about 7 percent were older than 70, and about 40 percent were between the ages of 35 and 49. Approximately 78 percent of participants had at least some college education, which was consistent with what we were seeing across many of the large randomized prevention trials, and only 3 percent of participants were African American. During this study, we did not have NCI Surveillance, Epidemiology, and End Results incidence rates for Latinas, so Latinas were included in the category of “other races” along with other populations.

What were the sources of accrual for the Breast Cancer Prevention Trial (BCPT)? When a woman filled out her risk assessment form to determine eligibility, she was asked how she learned about the trial. Unquestionably, the one-on-one interaction with a health care provider—a physician or a nurse—had the most significant impact on accrual to BCPT. Also of note were TV and newspaper ads and program announcements for BCPT.

We can now look back to assess the role of lessons learned from large prevention trials.

What about the participating physicians? We were taking patients who were not showing up in the oncologist's office; not many oncologists were participating in the recruitment of these studies. A study conducted at the MD Anderson Cancer Center looked at the effect of advice or recommendation from a primary care physician; a psychosocial survey was given to women who agreed to come to an informational focus meeting at MD Anderson. During the first 2 years of recruitment to the BCPT, 175 women completed the questionnaire. This study demonstrated that women whose physicians advised them to participate or enroll in this study were 13 times more likely to enroll. This result was not modified by their age, by their history of breast cancer, or by their interest in or desire to continue with hormone replacement therapy (HRT).

In the mid 1980s, several large randomized trials began looking at reduction in the incidence of breast cancer.

Community and clinical oncology programs, which were designed to take oncology out into the community, were responsible for one-third of the recruitment to BCPT. So when we asked about the most difficult factors to overcome in establishing recruitment programs for BCPT, answers included educating physicians, preference for HRT by women, the "pro-estrogen" physicians who believed that women had to have estrogen to live, health politics within the cities and internal politics within institutions, and sufficient time and staffing to enroll women into the study.

What was the demand on time? Each woman who completed a risk assessment form was given an individual risk assessment profile. There was a lot of discussion about risk and benefit: her individual risk, the minimal eligibility risk to participate in the trial, and the average woman's risk. In this way we were asking physicians to think about prevention and to take time out to discuss this with their patients. There were 98,000 risk assessment forms completed to enroll 13,000 women in BCPT, indicating that this process is quite labor intensive, and we had to engage the physician in a new dialog about risk and benefits. If women opted not to participate in the trial or if they were not eligible from their risk assessment form, we needed to decide what to do with them. Such women have expressed an interest in their health, so we have learned that we have to give something back to these women. We have to tell them that they are now "in the loop" in terms of prevention and that they should make sure they are informed about what they should be doing

to ensure their breast health.

The most common ineffective minority recruitment strategy reported to us was the health fairs. Neither brochures nor the churches were found to be effective, although the previous speaker found that perhaps churches were effective in recruitment. Other trials, with men, indicate that some approaches utilizing churches have been effective, but for this study, that was not the case.

The most effective minority recruitment strategies for the BCPT were related to outreach coordinators. These coordinators were at the various sites, and their duties were to go into the communities to educate women about breast prevention, to bring them back to the center with completed risk assessment forms, and to walk them through the process. During the time of the BCPT, there were at least five centers, including Howard University, Indiana University, Fox Chase Cancer Center, and a few others, that used outreach coordinators or were funded to assist specific institutions in outreach coordinating. Four sites in the BCPT program recruited 40 percent of the Latina population.

At the completion of this trial and publication of the data, various conferences and seminars were convened to help physicians develop ways to counsel women about risk and benefits resulting from the trial's findings. We knew that we now had potential risk, specifically endometrial cancer and some of the thromboembolic events. In addition, there was another agent that might be more attractive because it did not have the significant impact and increase of endometrial cancer—raloxifene. This was the second-generation serum that had been studied in postmenopausal women for osteoporosis and had shown about 70 percent reduction in the risk of breast cancer in a non-high-risk population, so this was a logical agent to compare with tamoxifen in a study. In this trial, we enrolled 22,000 postmenopausal women using the Gail Model criteria, with the exception that "age of 60" was not, independently, an ineligibility criterion.

Comparing the accrual of BCPT and the STAR study, during the first and second quarters in the first year, approximately 1,600 women were accrued to BCPT but only about 800 in STAR, and this trend continues in the second quarter of the first year. The single most important factor for that bolus of enrollment to BCPT had to do with the press release on April 29, 1992, which said that this trial would recruit 16,000 women (the trial was terminated earlier because of the heightened risk to participants).

In contrast, the STAR accrual by month was slow in the beginning of the trial. Several factors should be considered, one of which is that the BCPT was a novel trial based on novel science. We made attempts to make

the applications, risk assessment forms, and the whole process simpler, but there were a lot of things that needed to be changed with the IRB so it took time for the Centers to get started. I also think this was a different trial: We now had two agents, and information was available that forced women to be much more pensive in terms of identifying their risks and benefits. If we look at recruitment of minorities into STAR, there were 10,000 risk assessment forms at 36 months and 291 enrolled as opposed to 98 in the first BCPT.

There are many staff-related factors. The coordinators, who play a key role in recruitment to the breast cancer prevention trials, are committed and compassionate, and they use a nonthreatening approach and are honest with the women who are seeking trial enrollment. They treat the women with respect. We asked the women if they would enroll their family members, and of those recruited by program coordinators, 92 percent said they would.

Population samples have a habit of dwindling when one is seeking them. One of the approaches we used was to seek high-risk individuals from corporations. We looked at General Motors Corporation because it was in Detroit and because it was associated with the United Autoworkers (UAW). General Motors employs about 25 percent minorities, so NSABP collaborated with General Motors and mailed out 133,000 risk assessment forms. The company had a history of assessing risk in various health variables among their salaried employees. We were unable to get the endorsement of the UAW.

We received 9,000 risk assessment forms, which is about 7 percent return, and my colleagues in marketing say that is a fabulous outcome. Approximately 39 percent were risk eligible, 92 percent were risk eligible by menopause status, 64 percent were interested in STAR, and 90 individuals enrolled. Is that a good yield for such a massive mailing? Looking at our most successful sites,

this is certainly a competitive enrollment number. Other corporations have just begun to collaborate with the NSABP, so we do not yet have strong results for them.

We need to continue to change the public image about clinical trials and about prevention in particular. We need to educate physicians about risk-benefit counseling. We need to fund recruitment. We need to improve national promotion of cancer prevention.

We do not know the impact of HRT, which was a peril for us in terms of the BCPT accrual. I think it has certainly been thought provoking, but I think it was too early for us to report on the impact of the data released from the Women's Health Initiative on clinical trial accrual to breast cancer prevention trials.

It is a challenge to narrow the gap for outreach coordinators and institutions; they are working hard, but they can only be as successful as the institutions are receptive to them. We need to continue to change the public image about clinical trials and about prevention in particular. We need to educate physicians about risk-benefit counseling. We need to promote the science of clinical trials and provide academic recognition for those individuals seeking to become clinical trial researchers or trying to publish data on clinical trial retention and recruitment. We need to fund recruitment. We need to improve national promotion of cancer prevention to complement the local and regional efforts, and we need to continue to seek better identification of high-risk individuals and less toxic chemoprevention agents.

Life's Transitions Studies: Establishing and Maintaining Longitudinal Cohorts With High Respondent Burden

MaryFran R. Sowers, Ph.D.

Dr. Sowers is Professor of Epidemiology at the School of Public Health at the University of Michigan and Adjunct Professor in the Departments of Internal Medicine and Obstetrics and Gynecology. She directs the University of Michigan Center for the Study of Complex Diseases. Dr. Sowers has extensive experience in recruiting for and conducting longitudinal studies. Her experience includes recruiting and following two cohorts of women through pregnancy and 2 years postpartum, two cohorts of women through menopausal transition, and a cohort of women between the ages of 18 and 80 at initiation, and she has been following up for the past 15 years.

This presentation will move from clinical trials to longitudinal cohort studies, and there is a different mentality and rationale for how to approach these longitudinal studies. I would like to talk about a unique longitudinal cohort study and the lessons learned from the recruitment and retention in the Study of Women's Health Across the Nation (SWAN). SWAN is a prospective longitudinal study examining the experience of the menopausal transition and the decline in ovarian function.

The importance of conducting these kinds of longitudinal studies is well exemplified in SWAN, where we understand that the menopausal transition may play an important key physiological role in setting the stage for subsequent disease such as Alzheimer's disease, cardiovascular disease, and osteoporosis. However, most of the studies that have focused around this area have examined HRT or estrogen replacement therapy (ERT) and have not examined the concurrent physiological, psychological, and sociological processes. Furthermore, much of our information about this area comes from nonrepresentative samples, usually clinically derived and usually of Caucasian women.

SWAN is organized at seven sites, including a coordinating center, a core laboratory, and a DNA repository. SWAN has a rather unique hybrid study design: It was initiated as a cross-sectional study to identify sampling frames, and then the longitudinal component was based on those sampling frames. There are no lists for sampling that identify whether women are premenopausal, perimenopausal, or postmenopausal, so each of the seven sites conducted a cross-sectional survey to identify a representative sample of women. I would like to underscore the issue of representativeness, because all women experience the menopausal transition. This is not a unique, high-risk population but is representative of the

population of women, so ensuring generalizability was critical to us. We conducted this cross-sectional survey in women ages 40 to 55 and ultimately conducted interviews with more than 16,000 women.

Phase 2 was the longitudinal phase, of which we are in the seventh year of annual examinations. The aim of the longitudinal phase was to characterize the menopausal transitions in an ethnically diverse cohort. The women at initiation were 42 to 52 years old, and each site was responsible for organizing the efforts to understand the transition in more than 450 women. The eligibility characteristics of those women were that they (1) had a menses in the previous 3 months, (2) had an intact uterus, and (3) were not using hormones. SWAN was organized to recruit both the majority population Caucasian women as well as minority women; recruiting techniques are different where both groups are represented as opposed to recruiting exclusively one group.

SWAN is a prospective longitudinal study examining the experience of the menopausal transition and the decline in ovarian function.

Regarding lessons learned, a decision must be made to conduct either a population-based or community-based study. Data from population-based studies increase generalizability, but there are extraordinarily high costs associated with the data collection. Data from population-based studies are less likely to be biased because of the absence of selection factors. However, the more a study is population based and the researchers must follow people over time, the more difficult retention is.

Throughout SWAN, better participation occurs among women with more positive health, and poorer participation occurs among those who have poorer health.

This slide represents data from the 16,000 women who were seen in the cross-sectional study; we recruited our cohorts from that pool of women. The success with which we were able to recruit women for the cohort from that population pool varied greatly depending on several factors, the most notable of which were educational background, difficulty in paying for basic life necessities, self-reported health status, and to a degree less than expected, ethnicity. This result is not new, but we have to ask ourselves whether we have addressed these issues effectively in dealing with both recruitment and retention when we are looking at multiethnic populations.

SWAN is a great burden for participants. The total examination is 2½ to 4 hours long, depending on the particular site, the woman, and the year. It includes keeping menstrual calendars on a monthly basis and clinical assessments for bone densitometry. It includes having blood drawn for a number of different parameters. It includes urine collection. It also includes a daily hormone study in which we ask one-third of our 3,300 women participants to collect their urine daily throughout a menstrual cycle or for 50 days, whichever comes first, and to keep a daily record. This imposes terrific burdens on this community-based population. Any time we are dealing with physiological studies tied to hormonal studies, we deal with the cyclic hormonal variation in the menstrual cycles and with collecting data that we can interpret. The protocol asks us to see these women on days three through six after menstrual bleeding commences, and it mandates the use of diaries to characterize menstrual variability or for the daily urine collections.

SWAN is very important, but it creates time demands and financial challenges, particularly for women who are already economically challenged. But it also provides interesting social, emotional, and cultural challenges. It is difficult for many of these women at the Michigan SWAN site to pay for the basics of heat, housing, and food; among Caucasian and African American women enrolled here, approximately 40 percent report “some” to “substantial” difficulty in paying for basics. Interestingly enough, both groups have the same participation, and we believe that this is because both groups are equally economically challenged. Study-wide, difficulty in paying for basics varies greatly by ethnicity. Among Hispanic women, more than 80 percent report difficulty in paying for basics.

What is the impact of this challenge? Economically challenged women cannot be expected to participate when their participation costs them money. They want to participate, but if these women are employed, they may not have benefits that allow them to take the day off or they may not have sick leave. The relative cost of participation in a longitudinal, burdensome study like SWAN is regressive. Our participant fees are not an incentive, even though our IRBs may believe they are, because these fees do not cover the cost of participation for economically challenged women. Money speaks in our population; it pays bills.

The economic burden affects our studies and participants in many unique ways. If a research component requires giving some resources to the woman to support the research, such as freezers to hold the urine specimens collected throughout one menstrual cycle, the freezers may be sold, and then she puts her urine samples next to the ice cream. If we give participants watches so we can collect the daily salivary hormone studies in a time-related manner, watches might be sold for cash to pay the heating bill. Providing transportation is an ongoing struggle; it is not merely a matter of having taxicabs come to the door, since some taxicabs will not venture into some neighborhoods. If our results letters tell women to see their health care provider about an elevated glucose measure, can they afford the copay for this trip to their health care provider? What if they do not have any health insurance? At my site, 15 percent of the women do not have health insurance, and proportionately more Caucasian women than African American enrollees do not have health insurance.

Other factors are also important to remember. Self-reported health status varies by ethnicity, and among our study enrollees, two groups are more likely to self-report poor health status: African American and Hispanic women. It alters our participation rate; throughout SWAN, better participation occurs among women with more positive health, and poorer participation occurs among those who have poorer health. The same is true at the Michigan site. Why do I care about this? We are trying to describe something that is representative of all women, and this kind of participation pattern may bias our findings.

How might bias occur? Our cross-sectional study found that women who self-report that they have heart disease have on average an earlier age at menopause by almost 1_ years, even after adjusting for smoking and other factors. If large numbers of women who are ill—and some with heart disease—drop out of the longitudinal study, then the age at menopause that the study will report is later than if all women, including the less healthy ones, had continued to participate. This kind of result makes you question whether we are providing information that will be a true measure of age at menopause, which is one of the key determinants in SWAN when we do not have information about the full spectrum of the population. As in other longitudinal studies, we distill the population of people who will participate. The individuals distilled out may be those who are more economically disadvantaged and less likely to be healthy; it affects certain ethnic groups, particularly African American and Hispanic women, and may give rise to inappropriate conclusions about the findings from SWAN as a whole—and possible inaccuracies in any ethno-specific attributions that are made.

SWAN has also provided some very interesting challenges relative to culture. We do the study in four languages, and like everybody, we know it is important for participants to see themselves in the staff and the principal investigator. But it creates other very interesting constructs. For example, is eligibility based on the Chinese calendar or the Gregorian calendar? There are implications of culture for blood drawing or for DNA

collection. If a culture believes that providing blood means providing the essence of oneself, and one is reluctant to do this without relevance or rationality, then the woman may not participate. There are diverse and interesting cultural perceptions of the value of research, the role of health, and the cultural roles of women and their empowerment in a health care system.

Simple constructs of race and ethnicity can impose multiple and important subgroups that must be addressed. For example, are émigré Chinese the same as Chinese who have resided in the United States since the 1890s? Are the African American women at the Michigan site, whose families have been residents of Michigan since it became a State in 1837 and who were active participants in the Underground Railroad, the same cultural group as those African Americans who came to the area in the 1940s to work in the automotive plants? These groups live in different places, go to different churches and different schools, and are culturally different.

It appears that, in longitudinal cohort studies, it is inappropriate to lump together minority populations and think in terms of a single research recruitment and retention strategy. Themes about participation commonly attributable to minority groups may reflect the disincentive that we have placed there with respect to participation based on economic status and health status. We need to expand our methodology of recruitment, data collection, and data analysis to consider different aspects of recruitment and retention, particularly if we understand how these differences may bias our findings.

Simple constructs of race and ethnicity can impose multiple and important subgroups that must be addressed.

Chapter Three

Nuts and Bolts of Sex and Gender Analyses: Where Sex and Gender Really Matter

Moderator: Janet Wittes, Ph.D.

Clinical trials and epidemiologic studies funded by the National Institutes of Health (NIH) and by industry routinely recruit both women and men. The NIH maintains guidelines for the representation of men and women in clinical trials, but for industry trials that aspire to U.S. Food and Drug Administration (FDA) approval, the demographics of the study disease dictate the trial's gender mix. This session addressed the use to which recruitment and analysis by gender assists in interpreting data from such studies. The three presenters looked at the issue from three different vantage points.

Dr. Piantodosi argued that scientific questions, rather than general guidelines, should drive recruitment and analysis schemes. In observational studies, Dr. Kramer pointed to the importance of analysis by gender and the possibility of erroneous conclusions when the numbers of men and women differ greatly and the event rates differ markedly by gender. Dr. Weiss described how reviewers at the FDA analyze and interpret data by gender. All three presenters stressed the need for reasoned, thoughtful approaches to data analysis from clinical studies and stated that formulaic analysis would not likely yield important insights.

Why Sex Doesn't Matter

Steven Piantadosi, M.D., Ph.D.

Dr. Piantadosi is Professor of Oncology at the Johns Hopkins University and Director of Biostatistics at the Johns Hopkins Oncology Center. He has designed, implemented, and interpreted many clinical trials.

I question some of the foundations of the rationale for gender representation in clinical trials, and I do so from what I consider to be first principles. First, the foundations of subset mandates, particularly those for sex and race representation in clinical trials, have a political rather than a scientific origin. Second, the belief about underrepresentation of women in clinical trials was at the time of the NIH Revitalization Act of 1993 (Public Law 103-43) and remains today, a myth. There is a minor amount of ambiguity associated with this claim, but probably not as much as one might think. Third, race and ethnicity are sociodemographic constructs, not biological ones. Experience indicates overwhelmingly therapeutic clinical trials that gender and sex are not biological constructs.

The following heuristic argument underlies my strong claim. Imagine that a disease occurs in both sexes, albeit with some difference in frequency. That fact alone is strong evidence that the disease is produced by mechanisms that are independent of sex. Why, then, would one presuppose that therapy for such a disease should differ according to sex? In general, one would not. Would one therefore routinely design clinical studies to detect effects that the most reasonable biological approach would suggest are not present?

I do not think that this skepticism should be very surprising. One of my colleagues has suggested a genomic argument. Males and females share a minimum of 35/37ths of the human genome. That is an underestimate because it assumes that all 2,000 genes on the X and Y chromosomes are different, which is probably not the case.

It is easy to find circumstances in which we thought previously there might be gender-specific effects due to treatment but that have now evaporated. Breast cancer is a wonderful example because it is so unbalanced, with a very small minority of cases occurring in men, yet there

do not appear to be clinically important therapeutic interactions between treatment and the sex of the patient. Cardiovascular disease and prevention are another example in which we thought for many years there were important sex effects. We are discovering new facts about incidence, but the therapeutic effects appear not to interact with sex.

Sex is a surrogate for some factors that may turn out to be important. There are psychosocial differences; differences in body mass, lifestyle, and some environmental exposures, and even neuropsychiatric differences correlated with sex. So there might be sex effects in some important areas. One area that seems possible is treatment for drug abuse, for which some of these factors may partially dictate the kinds of therapies needed. It is often said that sex appears to modulate the therapeutic actions of drugs, but probably mostly through the kinds of surrogacy just mentioned. I have never heard the claim that sex modulated the effect of a skill-dependent therapy or a physical therapy such as surgery, a medical device, or radiotherapy. That is a very odd fact. Why not, if there are prominent interactions between sex and pharmacologic therapeutics? Of course there will be occasional circumstances where treatment decisions are affected by sex, but I am arguing that these will be greatly in the minority.

What are the implications of these ideas for representation in clinical trials? I will present a few concepts that I refer to as “representation fallacies.” The first fallacy is that the external validity of the trial is a consequence of similarities between the study cohort and the target population. This is not true. External validity is a consequence of biological knowledge about the effect of the therapy and the likely biological effect in the target population. Superficial similarities or dissimilarities

between the target population and the study cohort are irrelevant unless they modulate the effects of the therapy.

A second fallacy is that proportionate representation will facilitate detecting differences based on subsets. In fact, subset comparisons, when they are required, are not necessarily efficient or optimal with proportionate representation. Generally, we would require equal representation of the two subsets to have full efficiency to detect these effects. A third fallacy is that overt differences between cohorts invalidate comparisons. This is not true, for the same reason as the first point. Fourth, biological differences between cohorts invalidate comparisons, which is also not true. A biological difference need not modulate the effect of the treatment, and if it does not, it would not invalidate the comparison.

The final fallacy is that heterogeneous trial cohorts are more representative. This is similarly not true. All trial cohorts are self-selected and therefore superficially are nonrepresentative. Reliable biological knowledge needs to be applied to know whether a study cohort is substantively nonrepresentative compared with some reference cohort. Even those that appear marginally representative (one factor at a time) can either be biologically nonrepresentative or can differ in a multivariate statistical sense.

I am not saying that participation is not a worthwhile goal. In fact, I think it is a worthwhile goal for the following reasons. Increased participation speeds our answers to important questions. Generally the care of patients on research studies is as good as or better than that on community standard therapy. Participation does allow us to learn about potentially important biological effects, and it permits uncovering the large interaction between treatment and covariates. It does satisfy societal and ethical concerns about risks inherent in research and the benefits of that research. I waffle (probably on a weekly basis) between thinking that research is a benefit and thinking that it is a risk, and I am not sure which of those concepts is operative.

In reality, it is quite difficult and expensive to control the composition of a study cohort. What it boils down to is interaction. I want to sketch two types of interaction. An interaction occurs when the magnitude of the treatment effect depends on the presence or absence of a second factor. This is exactly the circumstance that we might worry about with sex. It is important to know about large interactions, but clinically consequential interactions tend to be rather uncommon, particularly for sex. Furthermore, even if there is a difference in magnitude of the therapeutic effect as a function of sex, usually the therapeutic decision will not depend on it.

For example, low-dose aspirin is good for men, and it may be slightly better or worse for women, but the general effect is that it is good for the arteries. So the therapeutic decision to use aspirin would not depend *per se* on sex, but it might depend on other risk factors.

A second kind of interaction has more of a therapeutic consequence. If the treatment effect changes depending on another factor, then the interaction is said to be qualitative. In other words, treatment A might favor one subset, but treatment B would favor the other subset, which would be a qualitative interaction. These interactions are exceptionally rare but would be important to know about. They are difficult to detect and usually require a minimum of fourfold increases in sample size. Unless there is a strong suspicion that such an interaction is present or is absolutely biologically vital to detect, on a routine basis we would not increase the size of clinical trials fourfold to try to anticipate these interactions.

Sex is a surrogate for some factors that may turn out to be important. Psychosocial differences; differences in body mass, lifestyle, and environmental exposures; and neuropsychiatric differences correlate with sex.

When the treatment effect in two subsets, men and women for example, is the same, there is no interaction and Treatment A would be favored. In a quantitative interaction in which women have a larger treatment effect than men, Treatment A would still be favored. In a qualitative interaction in which the treatment effect is positive in men and in the opposite direction in women, presumably men and women would be given different treatments.

In conclusion, increased study participation (e.g., diversity) is a worthwhile goal, provided it does not reduce the number or scope of clinical trials. Representation *per se* is not a worthy goal except in the rare circumstances in which it is required to test interactions, and then we would seek representation that is dependent on statistical criteria rather than sociodemographic criteria. Increased biological knowledge rather than empirical similarity is what enhances the validity of the results of clinical investigations. It is biological knowledge that allows inferences to cross species, ethnic, and gender boundaries.

Good for Men, Good for Women, Bad for People

Barry S. Kramer, M.D., M.P.H.

Dr. Kramer is Associate Director for Disease Prevention at the National Institutes of Health and Editor-in-Chief of the Journal of the National Cancer Institute (an independent journal of the Oxford University Press).

By way of disclosures, I would like to point out that although I am a Federal employee, any opinions I express here do not necessarily reflect official positions of the Government. I am going to discuss a statistical phenomenon named “Simpson’s Paradox.” I am unaware of any official position taken by the Federal Government on Simpson’s Paradox.

Second, I am not a statistician. Stuart Baker is the first author of an article we published last year in the *Journal of Women’s Health & Gender-Based Medicine*, titled “Good for Women, Good for Men, Bad for People: Simpson’s Paradox and the Importance of Sex-Specific Analysis in Observational Studies.” Stuart is the intellectual firepower behind some of these concepts. I picked up on Simpson’s Paradox by reading lay books. I read the book *A Mathematician Reads the Newspaper*, which mentions Simpson’s Paradox. Stuart and I often talk about statistical issues in the design, conduct, and analysis of clinical trials, and that led to the article I will discuss here.

This is an explanation of how the paradox can occur. In observational studies, a cohort or a series of people receive treatment A, and another cohort receives treatment B. It is important to point out that this applies to observational studies, not randomized clinical trials. (One of the reasons that randomized clinical trials are so important is that they avoid the possibility of Simpson’s Paradox.) The dilemma or paradox is that treatment A results in a survival rate of 60 percent for men; treatment B results in a survival rate of 50 percent for men. So treatment A is better than treatment B for men. For women, treatment A results in a 95 percent survival rate for women and treatment B results in an 85 percent survival rate. So treatment A is superior to treatment B for women as well—it is good for women, and it is good for men. But when the two cohorts of men and women

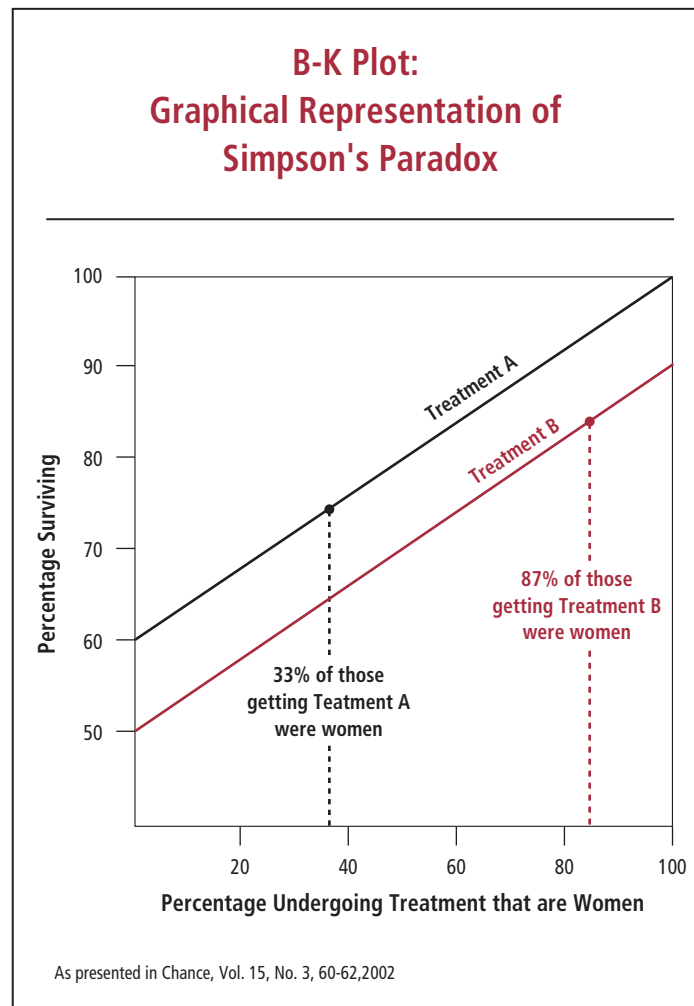
are combined into one category of people, treatment A has a survival rate of 72 percent and treatment B results in a survival rate of 80 percent. So, paradoxically, treatment B is better than treatment A!

It is key to know that the percentage of women given treatment A versus treatment B is quite different, and this is at the core of Simpson’s Paradox. The percentage of women getting treatment A is 33 percent but the percentage of women getting treatment B is 87 percent, and the natural history of the disease is more indolent in women than in men. Those are the two key issues. The new twist in our publication was that we came across a graphical presentation that explains the concept better for the nonstatistician.

The X-axis is the percentage of people undergoing treatment who are women; the far left side is 100 percent men and 0 percent women, and the far right is 100 percent women. The cohort may have any composition of men versus women. In the example presented, treatment A is better for women than for men; the survival rate is 60 percent versus 50 percent. Likewise, in a cohort of all men, the overall survival rate is higher for treatment A than for treatment B (95 percent versus 85 percent).

However, as illustrated in the figure above, in the first cohort only one-third of the people being treated are women, and in the second cohort the majority being treated are women. In each case, treatment A is about 10 percent better than treatment B, but when the two cohorts are compared, the cohort that received treatment B did better than the cohort that received treatment A. That, graphically, is Simpson’s Paradox.

To reiterate, there are two necessary conditions for Simpson’s Paradox to occur. First, for each treatment, the survival rate is better for women than for men; that is, the parallel lines on the graph slope upward from left to



right. Second, a larger fraction of women receive treatment B than receive treatment A. Put together, the results will produce Simpson's Paradox.

A randomized controlled trial is not subject to this problem because the percentage of women in each arm of the trial will be the same by study design; that is what randomization does. At any point in the curve under those circumstances, as long as the design is a randomized, internally controlled trial, the conclusion will be correct that treatment A is superior to treatment B. The implication is that, in observational studies, to avoid being fooled by Simpson's Paradox, one should adjust for sex in a regression model or present separate analyses for women and for men. Where appropriate, clinical and epidemiologic studies should be analyzed to determine whether there is an effect of sex on any of the major ethnic groups. If there is no effect, it should be so stated in the results section of the report.

In closing, I will point out that I have called these the B-K plot, and there is a little story behind that: B-K stands for Baker-Kramer. I am proud of this because this is the

only phenomenon to my knowledge that has ever been named after me. Howard Wainer, former editor of *CHANCE*, came across this graphical illustration of Simpson's Paradox, liked it, and wanted to write a column about it. He called us and said he wanted to label this the Baker-Kramer plot, the B-K plot. The problem was that Stuart Baker, who follows the statistical literature far better than I, had by now come across a similar presentation in a 1987 issue of the *Journal of the Korean Statistical Society*, published long before we thought of this independently. Stuart informed Howard Wainer that it was actually Dr. Jean and colleagues who came up with the conceptualization. Despite that, Howard Wainer decided to label this the Baker-Kramer plot, and the reason he gave was Stigler's Law of Eponymy: No invention or discovery is ever named after the correct person. That is certainly the case here; other examples are Simpson's Paradox, first described by Yule nearly half a century before Simpson, and Stigler's Law itself, first described by Merton. So all is right in the world.

Gender Analysis in the Regulatory Setting

Karen Weiss, M.D.

Dr. Weiss is Director of the Division of Clinical Trial Design and Analysis at the U.S. Food and Drug Administration's Center for Biologics Evaluation and Research.

Even though the focus of my presentation is on gender, from my experience gender is just one aspect of potentially larger subgroup issues with which we must contend. The authors of *The Fundamentals of Clinical Trials* (Friedman, Furberg, and DeMets) state “a leap of faith is always required when applying any study findings to the population with the condition. In taking this jump, one must always strike a balance between making unjustifiably broad generalizations and being too conservative in one’s claims.” This is the essence of what I deal with on a daily basis at the U.S. Food and Drug Administration (FDA).

When we review proposed clinical trials at the FDA, we always look at the inclusion and exclusion criteria to be sure they are appropriate. When the study is completed, the results are analyzed, and a product is approved for marketing, we wrestle with the labeling issues and with the issue of extrapolation from the study population to the larger population, which may or may not derive the same benefit from use of the product.

A few, but not many, regulations are specific for gender and subgroups. The following regulation is about clinical hold (meaning to stop or to prohibit initiation of a trial) and is in addition to the other FDA regulations about clinical hold:

Clinical Hold – 21 CFR 312.42 (b) (v)

The IND is for the study of an investigational drug intended to treat a life-threatening disease or condition that affects both genders, and men or women with reproductive potential who have the disease or condition being studied are excluded from eligibility because of a risk or potential risk from use of the investigational drug of reproductive toxicity or developmental toxicity.

This regulation probably was a result of earlier guidances and the impressions that women of child-bearing potential should be excluded from research, particularly from early clinical trials. When analyses were conducted, a concern arose that this regulation was overly paternalistic and rigid and that it had the unintended effect of inappropriately excluding women from participating in trials and leading to loss of valuable information in women who have certain diseases, particularly diseases that are severe or life threatening.

I have never personally invoked this particular hold criterion, and I am not aware that it has come up as an issue in the oversight of clinical trials relevant to the products regulated in my office. Many of our products are being proposed for study in serious life-threatening diseases, and to my knowledge, in those settings, there has never been a proposal to exclude women, even in early clinical trials. However, the criterion has been put in place to try to rectify some situations from the past.

Another provision in the FDA regulations—21 CFR 314.50(d)(5)(v)—states that (a) the effectiveness data should be presented by various types of subgroups (such as by gender, age, and race), (b) the data should identify any needed modifications, and (c) the data from other subgroups, such as patients with renal failure or with different levels of disease severity, shall also be presented when appropriate. It does not state specifically that the trial must enroll certain numbers or certain proportions of any particular group; rather, it states that those data should be analyzed when the trials are done and clinical data are submitted to the FDA in the form of a marketing application.

Manufacturers almost always perform those types of subgroup analyses, and reviewers at the FDA will also perform similar subgroup analyses and look for any differences or trends. However, as statisticians warn,

analyses of many different subgroups will often detect some aberrant effect that is somewhat different than the effect for the overall population. Questions then arise: What does it mean, is this real, is this something that needs to be followed up, and is this something that is relevant to put into the labeling? Those issues are never easy to address.

Now I will highlight the FDA's labeling regulations. Among the various sections of the label is a clinical pharmacology section. This section is to include information on pharmacokinetics and on special populations, including women, elderly persons, those in renal failure, and so forth. Myriad different populations exist for whom pharmacokinetics information might be different and important to include in the product label, particularly if those differences have important clinical consequences. The inclusion of detailed information about subtle differences in pharmacokinetic data that had no clinical sequelae and that do not lead to changes in the dosing or monitoring confuses people unnecessarily and clutters the label. Therefore, it is important to identify the key issues and to know what to include on the label, and that is not always easy.

A section in the FDA labeling regulations that address the indication and use section of the label states that, "if evidence exists to support the safety and effectiveness only in selected subgroups of the larger population, the labeling should describe the available evidence and state the limitations of the drug's usefulness." Product approval is often the point at which we have intensive discussions, negotiations, and sometimes battles with manufacturers about what and for whom to indicate a product. As a rule, manufacturers tend to want indication statements that are as broad as possible, for a lot of different reasons. Many consumers and third-party payers will look at the label to determine issues such as reimbursement. The scientific point of view at the FDA needs to ensure that the label is as appropriate as possible in identifying who is likely to benefit from use of the product.

The label also has a precaution section that provides information on certain subgroups, such as use for pediatric patients (if there are data to allow pediatric use) and geriatric patients. The FDA recommends specific boilerplate statements to be used in labeling to describe some of these populations.

Gender is an important subgroup of interest for a number of different reasons. For example, gender is sometimes a surrogate for weight, and for most of the products I deal with in biologicals, dosages are based on a milligram-per-kilogram basis or per-square-meter basis. The differences in pharmacokinetics often result from

differences in weight or lean body mass, and a drug may be administered by injection because of the ratio of muscle to fat tissues. Even so, identified pharmacokinetic differences often do not translate into pharmacodynamic or clinical differences.

Elderly persons are a group of interest. Guidance documents currently are available regarding individuals 65 and older and 75 and older, and it is recommended that labeling describe data for those populations. Usually, elderly research subjects represent relatively small numbers in those subsets compared with the larger group in the study, and as a result, it is difficult to draw firm conclusions.

A frequent issue is different degrees of disease severity: Does this work better or worse in people with more or less severe extremes of a disease?

People with various comorbidities constitute another group. In many clinical trials, research participants are selected carefully who do not have certain types of comorbid conditions, such as renal failure, hepatic failure, or use of concomitant medications, and that is one reason our postmarketing surveillance program is so extensive and important once a product is marketed. It is known that the drug will be used in people who are not as carefully selected as those who participate in clinical trials.

Demographic and baseline characteristics might be important to identify differences in populations. We struggle frequently with continental differences (the United States versus Europe); for example, we attempt to discern whether clinical data generated solely from outside the United States are relevant to the U.S. population. Differences exist among countries in the practice of medicine, use of medications, standard of care, and access to medical systems, and these differences raise questions about whether data can be generalized from one country to another.

I will now describe three examples of subgroup issues that have arisen in my experience at the FDA. The first is a product called Synagis®, a monoclonal antibody that was licensed in 1998. Synagis® is used as prophylaxis for infants at high risk for developing severe respiratory syncytial virus (RSV) infection. The primary outcome of that study was the incidence of RSV hospitalization. The overall effect was quite impressive: The incidence of RSV hospitalization in placebo patients was approximately 11 percent; in the treated group it was 5 percent. It cut the incidence of RSV hospitalization in half, and the significance was $p < .0001$. We looked at various types of subgroups and noted a difference in genders: The incidence of RSV hospitalization for males was 13 percent with placebo and 4 percent with Synagis®; for females, although there was still a benefit, the effect was

much smaller. There was no biological reason for the difference except maybe that premature girls tend to have better outcomes compared with premature boys, so that there was less obvious benefit with the drug, but there was no certain reason to restrict the indication to boys and not allow it to be indicated for girls, because a benefit still exists overall and in both genders.

Of perhaps more interest was that the infants studied in this trial either had either bronchial pulmonary dysplasia (with or without prematurity) or only prematurity (less than 32 weeks gestation). These two conditions are ones in which RSV infection can be particularly serious. Infants with congenital heart disease were excluded from the trial. One reason for this exclusion was that in earlier trials with RespiGam®, a polyclonal serum also indicated for prevention of RSV hospitalization, a higher rate of adverse events was seen among infants with congenital heart disease. That manufacturer subsequently conducted a separate trial in only children with congenital heart disease and did not find any obvious benefits of the polyclonal serum for those children. Part of the reason for this result might be that the overall high incidence of adverse events and adverse outcomes in infants with heart disease overshadowed any potential benefit of preventing RSV infection.

Thus, the manufacturer of Synagis® (the same manufacturer of RespiGam®, the polyclonal serum) chose to specify that infants with congenital heart disease was an exclusion criterion. When the Synagis® trial was completed and efficacy was proven, we had no data on infants with congenital heart disease, an important subgroup of infants who were also at high risk for developing complications from RSV infection. We were unsure whether benefit would accrue to this subgroup, but there was no plausible biological reason why they would not benefit from Synagis®. We struggled with this issue and decided not to restrict the labeling, but rather we indicated the drug for all infants at high risk for developing RSV infection. This decision was somewhat radical in that most label and indication statements reflect only the population that was studied.

It turned out that the American Academy of Pediatrics had enough concerns about use of products for prophylaxis of RSV infection in the setting of congenital heart disease that, despite the lack of restriction in the indication, they did not recommend its use in infants with congenital heart disease. Pediatricians tend to take seriously the advice of the American Academy of Pediatrics. The manufacturer is now undertaking a large controlled trial only in infants with congenital heart disease. Had infants with congenital heart disease been

enrolled in the first trial, they probably would have been examined as a separate subgroup, and although there probably would not have been enough statistical power to detect a difference, there might have been sufficient experience with Synagis® in infants with congenital heart disease that the Academy may have been less restrictive in its recommendations.

Another example is Xigris®, which is an activated protein C for the treatment of severe sepsis. We licensed this product in November 2001. The overall effect of this product was that it showed a significant difference in mortality: The death rate with Xigris® treatment was 25 percent and with placebo, 31 percent. Because the treatment difference between Xigris® and placebo crossed a preset boundary for stopping based on efficacy, the trial was stopped at the interim analysis. However, an analysis of subgroups was performed based on acute physiology and chronic health evaluation (Apache II disease severity) quartiles. The Apache system is a scoring system originally developed as a prognostic indicator among patients with sepsis. Subgroup analysis based on Apache quartile was a prespecified covariate in this trial. That analysis showed that individuals who fell within the first and second Apache quartiles had no overall treatment benefit with Xigris® compared with placebo. All the benefit was confined to those who were in the third and fourth Apache quartiles; in other words, the people with the most severe degree of impairment from their sepsis and who were at the highest risk of death had the most benefit from Xigris®. Of particular concern in this analysis was that, although the individuals who were in the first Apache quartile had the lowest overall risk of death from sepsis, those who were randomized to receive Xigris® had a higher death rate than those on placebo.

Xigris® is a product with a narrow therapeutic index. There is a significant risk of bleeding and, as a result, the possibility that it could lead to severe intracranial hemorrhage and other serious bleeding events. It appeared that people with lower Apache scores experienced the same clinical adverse events of significant bleeding; therefore, the labeling restricted use to only individuals with higher Apache scores. We have asked the manufacturer to evaluate this product in patients with lower Apache quartiles, and it is conducting a large controlled clinical trial with those individuals with lower Apache quartiles.

These examples are not specifically focused on gender, but they also represent other subgroup issues that we have evaluated and considered at the FDA.

Chapter Four

Informed Consent: The New Landscape

Moderator: Joan Rachlin, J.D., M.P.H.

This panel considered a range of bioethical issues that arise when women are included in clinical trials, including how to prevent breaches of trust similar to that created by the U.S. Public Health Service (PHS) Tuskegee Study of Untreated Syphilis in the Negro Male; whether women are more vulnerable than men and thus deserving of higher levels of protection relevant to clinical trials; and whether pregnant women should participate in clinical trials.

Dr. Koski presented from the perspective of one who has conducted research, participated in research, chaired and sat on an institutional review board (IRB), and led the Federal agency responsible for ensuring that research subjects are adequately protected. He spoke of the need for more caring in research and of a corresponding desire that subjects and clinical investigators form a partnership.

Dr. Johnson has been an outspoken and credible advocate for the ethical treatment of women and minorities in research, two communities that historically have experienced disparities in treatment and that have often been less-than-informed research subjects. She spoke of the importance of research subjects better understanding the nuances of coercion. Dr. Johnson suggested that the treatment of subjects would improve if IRBs and investigators formed a more robust partnership, so that subjects

have the benefit of more accurate consent information delivered in a more respectful and responsive manner.

Dr. Watts has worked extensively in the fields of acquired immunodeficiency syndrome (AIDS) clinical trials and has a special interest in issues surrounding the inclusion of pregnant women in trials and issues involving maternal/fetal decisionmaking. She spoke about a range of ethical issues when involving pregnant women in clinical trials and discussed who gives consent when doing research in this population, how risk assessment is conducted, and what level of effort is necessary to locate the father when paternal permission is required.

Each panelist directly addressed the issues that block truly informed consent and suggested strategies for improving the current manner in which consent is obtained. The panel framed the ongoing challenges that exist when enrolling women, particularly pregnant women and women from communities of color, in clinical trials, and urged those present to redouble their efforts to ensure fairness and adherence to the three Principles laid out in *The Belmont Report: Ethical Principles of Human Subjects of Research*: “Respect for Persons, Beneficence, and Justice” and the proposed fourth principle of caring.

Human Subjects Review Panels and Informed Consent

Greg Koski, M.D., Ph.D.

Dr. Koski is Associate Professor of Anesthesia at Harvard Medical School and a senior scientist at the Institute of Health Policy at Massachusetts General Hospital. He was the first Director of the newly constituted Office for Human Research Protections (OHRP). Under his leadership, the concept of shared goals and shared responsibilities has taken hold, moving away from a reactive, compliance-based system of oversight to a more proactive paradigm of prevention, using education as its cornerstone.

The theme of this conference, “science meets reality,” knows no gender. I would suggest that there are two orders of reality that science and society have encountered during the past 50 years. Unfortunately, our recognition and acceptance of these two realities perhaps came out of order; had these confrontations with reality transpired differently, we might have seen a very different progression in the events that have dominated human research over these past 50 years.

When breaches have occurred, too frequently the victims have been the least fortunate, most vulnerable members of our society.

That human research ethics was, as Carol Levine, M.A., so aptly put it, “born in scandal and reared in protectionism,” leads us to the first reality, and it is not one in which we can take a great deal of pride. This reality is a lesson learned from repeated breaches of ethical behavior and responsible conduct in human research. We have learned over the years that some individuals in science are too willing and too able to allow their personal interests and the interests of science to take precedence over the interests of those who are being studied in the course of their work. When these breaches have occurred, too frequently the victims have been the least fortunate, most vulnerable members of our society: minorities, prisoners, elderly persons, infirm and terminally ill individuals, those with mental incapacity, young people, and women.

Justice is at the heart of this issue and a fundamental value in our society. We are, after all, “one nation under

God, with liberty and justice for all.” In reality, we have learned that it is far easier to talk about the principle of “justice for all” than to realize that goal. The same is probably true for liberty, since, in most societies, liberty and justice go hand in hand. Those individuals who are able to make their own decisions, those who are empowered to pursue courses of action in their own interest, are more likely to achieve justice and to be treated justly in any system, whether in science or in another endeavor.

A national commission discussed principles of justice—absolute justice, individual justice, deliberative justice, and distributive justice—and at this conference we have heard about egalitarian and utilitarian justice. An additional aspect of justice that should be considered is relative or value-based justice. Simply stated, relative justice is the extent to which any group or society is willing to expend its limited and sometimes precious resources to take the actions necessary to ensure that justice will be done; in part this is determined by the perceived value of those who are likely to be subject to injustice.

In recognizing this aspect of justice, a realization that came from the disclosures of ethical lapses years ago, the Federal Government moved to take steps to prevent injustice to human beings in research. First, the PHS, and later the Congress, acted to implement a system for protection of human research subjects. The system advanced first through policy and then through regulations and was implemented through what have been called the “twin pillars of protection”—review boards and informed consent. This is a system that was imposed on the research community, and its acceptance has been less than ideal in many instances. For more than three decades, we have been trying to make this system work—and progress has been made—but clearly we still have a

long way to go. Few people at this point would argue that either IRBs or the informed consent process are working as they should. Still, too many scientists and too many sponsors view these twin pillars of protection of human research subjects as twin barriers to their research progress.

In fact, many people wonder whether the current system of IRBs and informed consent can ever, in their present practice and form, achieve the desired goals. Many initiatives are under way to try to improve the performance of these processes. These initiatives are a direct consequence of the recent and rather harsh encounter of science with a second reality: that human research simply cannot progress, in fact it cannot even proceed, unless people are willing to volunteer to be participants in research studies.

In clinical research particularly, the primary limitation to progress is the number of people who are willing to enroll in studies and continue participation through to completion. Clear evidence in support of this comes from industry's analysis of the amount of time required to enroll and complete trials. This harsh reality occurs in all areas of research, not just women's health research. Even in the war against cancer, the plight is the same—too few people are willing to participate. This is particularly ironic given the excitement generated any time there is yet another news account of a “breakthrough” in the treatment, diagnosis, or prevention of disease.

In his book, *Retrospectroscope*, Julius Comroe discusses the enhanced acuity and clarity of hindsight. Consider for a moment how things might have been different today if science had faced these realities early on. Suppose that a national commission had delineated the ethical principles for conducting human research before the PHS began its study on the progression of untreated syphilis. Suppose that the scientists engaged in that study had been formally trained in human research ethics and responsible conduct before Tuskegee. Had the scientific community recognized and accepted in the first instance its absolute dependence on the willingness of volunteers to participate in research; had scientists more clearly appreciated the critical importance of confidence and trust in the scientific process; had scientists applied the principles of respect, beneficence, and justice early on, we would not today be struggling with how to rebuild the confidence of the African American community in the research process. This lack of trust poses a serious impediment to research in the African American community. The same may be said for other groups.

Suppose for a moment that the reports from the Office of the Inspector General regarding the status of our

national system for protection of human subjects had been issued before the NIH Office of Protection from Research Risks (now the U.S. Department of Health and Human Services OHRP) began its site visits to academic research institutions, rather than after. Might this have prevented some of the suspensions of research activities that have been so costly in so many ways to so many people? Could the negative consequences have been avoided? Perhaps, but only if the academic scientific community had acted decisively and proactively to ensure that their responsibilities were fulfilled.

Today, we find ourselves, both scientists and policy-makers, victims of our own shortsightedness and of our failing to ensure justice for all research participants. We have failed to give appropriate priority to the value of those on whom we are absolutely dependent for the very ability to conduct our research activities. From the outset more than 30 years ago, our efforts to protect people participating in research have suffered from inadequate commitments to them, both financially and intellectually. Our actions have not fulfilled the promise of our words. We have allowed a system to evolve that has failed to provide reliably the protections that were intended. And as a consequence, the scientific community has failed to earn the confidence and trust essential for willing participation of human subjects in research.

Many people wonder whether the current system of IRBs and informed consent can ever, in their present practice and form, achieve the desired goals.

Fortunately, this dismal landscape is changing. The scientific community is now coming to accept these realities, and progress is being made toward a more effective, more appropriately resourced system that seeks to prevent harm and to promote responsible conduct. We clearly have a long way to go, but at least we have started down this less traveled road with all hope that it continues. Let us hope that a new focus on prevention—through quality improvement and education—will continue to predominate over the inadequacy of simply achieving minimal regulatory compliance as we work together toward our common goals. The move toward excellence and building trust through responsible conduct is under way, and the scientific community must embrace it to avoid another cycle of abuse disclosure and reactive hyperprotectionism that could further stifle valuable and much needed progress in human research.

We are only beginning to address the challenges of the informed consent process in any meaningful way. Regulations tell us what we have to say in a consent form—what elements must be there, who has to sign it, and when it has to be signed. Investigators and sponsors try their best to write appropriate consent forms to meet these requirements, but those forms rarely meet the criteria of IRBs. Consequently, IRBs, which already have inadequate time and resources to do their substantive reviews of research protocols, spend too much of their time in trivial editing of consent forms (wordsmithing), which contributes in no meaningful way to anyone's goals.

Already we know that these forms are too long, too technical, and too legalistic and that they emphasize far too heavily documentation to the detriment of the effectiveness of the entire process. There is universal agreement that the consent process as routinely practiced today simply fails to achieve its desired goal. Virtually everyone says that we must focus more on process rather than paperwork, but despite this recognition, we seem to be able to do little or nothing to get out of the bind in which we have placed ourselves and prospective participants. Clearly, this requires more attention. Some progress is being made, particularly when IRBs, investigators, and sponsors begin to take a more holistic and integrated approach to informed decisionmaking in contrast to an exclusive focus on consent forms and documentation. Attention to this process avoids a preconceived notion that the desired outcome of the process is consent and instead views the goal of the process as being to help people make an informed decision whether or not to participate.

This progress needs to continue. At the same time, more effective efforts to attract participants are needed. One concept for advancing recruitment and strengthening the consent process is to develop, as part of the research plan, a formal protocol for identifying potential subjects, contacting them, recruiting them, putting it all together, and engaging members of the community in the process of designing the mechanisms for reaching out to the entire community, to build a relationship of trust that can be the foundation necessary for enrollment and continuing participation. The use of new supplementary tools to provide additional information—different kinds

of audiovisual presentations, computers, and even evaluative techniques—is emerging and hopefully will have a positive effect.

The need for new guidance regarding the process of shared, informed decisionmaking is evident. There may also be a need for significant changes in the regulations as we find that regulatory requirements become impediments to achieving the true goal of informed decisionmaking. That is an area the research community would do well to pursue on its own proactively, rather than wait for the government to mandate how it should be done. The research community must recognize that it—and not a Government regulatory agency—bears the primary responsibility for protecting human subjects and that it must take the lead for what has to be done. This is the single most important message I share with you today. Responsibility for protecting and promoting the interests and safety of research participants rightly lies with the research community; it should not, and indeed cannot, be vested solely in the hands of government.

During the past 20 years, we have seen the ethical and regulatory perspectives regarding the participation of women in research change from one that was originally intended to protect vulnerable individuals to one that was viewed as a paternalistic impediment to women's participation. Despite regulatory change that has established inclusion of women as the norm, we now find that sponsors and IRBs are reluctant in some cases to allow women to participate in research out of fear of potential harm and legal liability. Our perspectives can and do change, and the scientific community should be leading the way, not waiting to be told what it must do.

A lot of work needs to be done, and “caring” is emerging as a theme. Caring is a concept that came from feminist moral philosophy, from individuals such as Carole Gilligan, Ph.D., and Nel Noddings, Ph.D.,—the notion that we put others' interests, the interests of those for whom we care, ahead of our own. If we could add caring to the principles explicated in the Belmont Report, and if we begin to practice caring within our research activities, we may not only be able to conduct research responsibly without harming people but also to facilitate the overall conduct of our research to more fully and expeditiously realize the promise and benefits of science.

Progress is being made toward a more effective, more appropriately resourced system that seeks to prevent harm and to promote responsible conduct.

Protecting Vulnerable Populations

Paula A. Johnson, M.D., M.P.H.

Dr. Johnson is Executive Director of the Connors Center for Women's Health and Gender Biology at the Brigham and Women's Hospital in Boston. She is also Chief of the Division of Women's Health at Brigham, where she is developing an interdisciplinary research, clinical, and educational program in women's health.

There are more questions than answers on the topic of including and protecting vulnerable populations in research and on the role of IRBs. To begin with, what is a vulnerable population, and what is meant by “vulnerable”? One aspect of vulnerability refers to physical health status, which includes not only physically impaired health but also pregnancy. Another aspect is psychological health status—mental illness and other psychological states. Other aspects include economic status, literacy, command of the English language, minority status, and caregiver status. The big question is whether women are more “vulnerable” than men. A number of people believe that women in general do not fit the definition of vulnerable.

There are three essential elements to the informed consent process: The subject must have all the relevant information, must have the capacity to give consent, and must be free from coercion.

Let us review some facts obtained from the Jacobs Institute of Women's Health and the Henry J. Kaiser Family Foundation that address whether women should be considered a “vulnerable population.” The Kaiser Foundation's Women's Health Survey provides helpful information: 32 percent of the women surveyed reported a chronic health condition, as did 26 percent of men; 21 percent of women experienced a major depression in their lives, a psychological health state that might put one in the “vulnerable” category. Women are more likely than men to be poor or near poor, a condition that is greater for minority women and for women with poor health status.

Women are more likely than men to graduate from college, but they still earn only 76 percent of what men earn, an economic disadvantage that could be classified as making them “vulnerable.” Women are more likely than men to be caregivers for sick or elderly family members, they are more likely to be single heads of household, and the women who perform these duties are more likely to be poor. Men clearly have much more power and hold more powerful positions than women, and therefore there is a power differential. Looking overall across these categories, one could argue that women are more likely to fall into the “vulnerable” category.

One of the main responsibilities of an IRB is to review the types of studies being performed to determine whether there is less than minimal harm or more than minimal harm. If we think about those definitions of vulnerability and about how an IRB functions, do these definitions change when characterizing vulnerable populations? The answer is probably yes. Investigators as well as IRBs must think about the impact of a particular study on its subjects. Another responsibility of IRBs is to safeguard the consent process and to promote open and free communication between the researcher and the research participants. This is not an easy task. There are three essential elements to the informed consent process: The subject must have all the relevant information, must have the capacity to give consent, and must be free from coercion.

First, the subject must have all the relevant information. In the Kaiser survey, 30 percent of the women in fair or poor health said that when they left the doctor's office they did not understand or remember some of the information given; 18 percent said the doctor did not usually take time to answer all of their questions. These women who are at increased risk, who are probably

candidates for many of our clinical trials and studies, are already in a situation of having a poor understanding of their health problems from their health care providers and of feeling dissatisfied and unable to obtain the information they need. Issues of power, literacy, cultural competence, and gender-specific qualities of communication must be addressed as part of the responsibility of providing relevant information to research participants.

IRBs have the responsibility to ensure that any special vulnerabilities of participants are accommodated and handled appropriately.

Regarding freedom from coercion, one must understand the nuances of coercion. Overt coercion is relatively easy to identify, but subtle coercion is much more difficult to grasp. Consider coercion in the context of quality of health care. This might be controversial, but does participation in a trial improve the quality of health care? What is the quality of health care in general for women? In the Kaiser study, only 35 percent of women ages 50 to 64 had been screened for colon cancer in the past 2 years; 56 percent of women ages 18 to 64 had undergone blood cholesterol screening. Screening is much less likely among uninsured persons and for individuals on Medicaid, and the number of uninsured and underinsured individuals is increasing. Does the fact that inclusion in a trial may improve the quality of care function as a form of coercion when the number of underinsured and uninsured people, especially women, is increasing significantly in the United States?

IRBs have the responsibility to ensure that any special vulnerabilities of participants are accommodated and handled appropriately. How are issues for vulnerable

populations addressed by IRBs? By regulation, the community must be represented on IRBs, but how is this accomplished? Who is the “community,” especially regarding vulnerable populations? What about the time and effort required to involve community members, and how can the lack of pay be addressed? Once again the question arises of coercion versus the need for reimbursement. There are no clear answers, but there are important issues to consider in terms of how IRBs function.

The use of representative consultants has been put forth as a recommendation, but how often is this undertaken and how realistic is it? Those who have tried it have found it is a somewhat difficult thing to do. The available representatives tend to be retired individuals of reasonable means, but do they truly represent the community of vulnerable populations?

How can IRBs or investigators understand some of these issues? Most of the issues must be addressed prior to developing the informed consent document. Informed consent is a process, and it is important to consider whether IRBs have the ability to become more of a resource for investigators regarding working with vulnerable populations, rather than merely being reviewers. Several IRBs around the country do live up to this resource standard. If we move in that direction, we may have a more robust partnership between investigators and IRBs that allows the consent process to act as it should.

What is protecting a population, and should vulnerable populations be excluded from clinical trials? I think we all know the answer is no, but we have gone down this road before. Exclusion led to significant negative impacts on the health of vulnerable populations, especially women. However, in recognizing that history, including vulnerable populations in clinical research brings with it the responsibility to ensure that they have the ability to undertake all three aspects of informed consent. The question is: Can there be true informed choice?

Inclusion of Pregnant Women in Clinical Trials: A Case Study

D. Heather Watts, M.D.

Dr. Watts is Medical Officer in the Pediatric, Adolescent, and Maternal AIDS Branch in the Center for Research for Mothers and Children at the National Institute of Child Health and Human Development. She has worked extensively with women with HIV/AIDS and has chaired a number of task forces and committees on those topics.

Over time, there has been a shift from presumed exclusion to presumed inclusion of pregnant women and women of reproductive potential if certain conditions are met; these conditions are specified in subpart B of the Code of Federal Regulations. Where scientifically appropriate, we must have preclinical and clinical studies to enable an assessment of risks. If the risk to the fetus is greater than minimal, there must be some prospect of benefit, and only the least possible risk should be included that still enables achievement of the research objectives. A relatively recent requirement is that the informed consent include reasonably foreseeable impacts of research on the fetus. Subpart B is in addition to the other requirements for informed consent and other aspects that IRBs look for and investigators should consider.

Where scientifically appropriate, we must have preclinical and clinical studies to enable an assessment of risks.

Who must give consent? If there is benefit to the woman, whether or not there is benefit to the fetus, or if the risk is minimal, only maternal consent is required. However, if the benefit is deemed primarily to the fetus and the risk is greater than minimal, paternal permission is also required unless the father is unavailable or incompetent or the pregnancy was the result of rape or incest. There can be no inducements for termination of pregnancies, and the researchers are not allowed to be involved in decisions regarding termination or viability of the fetus.

To examine the various aspects of inclusion of pregnant women in trials, I have chosen to use a case study approach. The study, known as AIDS Clinical Trials Group (ACTG) 076, was a randomized, double-blind,

placebo-controlled trial of the efficacy, safety, and tolerance of zidovudine (ZDV) for prevention of maternal-to-fetal HIV transmission. I was involved in implementing this trial at the University of Washington, and it was a fairly outlandish idea at the time, according to some people, to think about using ZDV in pregnant women. Although at the time this was controversial, it is now standard practice.

The trial enrolled pregnant women between 14 and 34 weeks gestation. The subjects had to have a CD4+ lymphocyte count greater than 200, which is important because it was felt that if women had CD4 counts less than 200, they should be receiving ZDV for their own health. Thus, it would be unethical to randomize those individuals to a placebo arm. The flip side of that argument meant that we were doing this trial primarily to prevent transmission and not for the benefit of the mother. Treatment was given during pregnancy, intrapartum to the mother, and then for 6 weeks to the infant. Enrollment began in April 1991.

One of the requirements of the regulations is to have data to assess risks for the fetus and for the woman. What data did we have? We had a small amount of data from studies in nonhuman animals. This issue regarding preclinical data occurs with many drugs that might be used in pregnancy: What responsibility does a drug development company have to provide animal data? In the case of ZDV, we had more data than are often available when a drug is prescribed during pregnancy, whether as part of a trial or in a clinical situation. We knew that ZDV was not teratogenic in mice or rats, and there was no evidence of impaired fertility. There was an increased incidence of nonmetastasizing vaginal tumors in rats, which was probably due to the anatomy of the rats and to the concentration of ZDV in the urine, which

was much higher than in humans. So it was believed that this was not an issue for use in women, although we had to consider it.

What human data did we have? We knew that ZDV is a nucleoside analog and interferes with DNA polymerase. It is fairly specific for viral DNA polymerase, but at high enough doses it will affect human DNA polymerase as well. The data that we had on its use in nonpregnant individuals were primarily in men. The studies that had been done at that point had included more than 90 percent men and indicated that nausea and anemia were the most common side effects.

There was a background risk that 25 percent to 30 percent of infants born to infected mothers would be infected, and reducing that risk would be a significant benefit.

We had pharmacokinetic data from about a dozen women by the time the trial started. Because the IRBs and investigators were concerned about the lack of data, a retrospective survey of data was conducted from AIDS clinical trials sites that had treated pregnant women with ZDV. The survey yielded 43 cases, so that was our database on use in human pregnancy at the time the trial was getting under way. There were no data on the potential impact of ZDV on pregnancy outcomes such as preterm birth or on the long-term effects of maternal therapy on the women or on the infants.

Balance these risks against the potential benefit. We were not talking about a drug for the common cold; we were talking about HIV. There was a background risk that 25 percent to 30 percent of infants born to infected mothers would be infected, and reducing that risk would be a significant benefit. No test was available to determine the individual risk of transmission, so we could not target risk. The median survival time for infected infants then was only 38 months.

The study was powered to detect a reduction in transmission from 30 percent to 20 percent—a one-third reduction, a fairly large effect. We did not want to have a sample size of thousands because we wanted to have an answer about efficacy fairly quickly. Reviews were conducted on a regular basis by a data and safety monitoring board (DSMB) to evaluate the side effects, adverse events, and transmission rate in each arm of the study. Indeed, a larger effect on transmission was found, and the DSMB recommended early closure of the study to

allow all of the pregnant women to receive ZDV.

What about the issue of the least possible risk to achieve objectives? Why not just treat the babies? The initial thought was to treat the babies after they were born to prevent infection. However, the data indicated that there were several animal models suggesting that treatment of pregnant animals with other retroviral infections (e.g., feline leukemia virus) prevented infection in a substantial proportion of the infants. Instead, if treatment was delayed until right after birth, manifestations of infection were delayed, but infection was not prevented. So the justification was solid that this drug needed to be administered during pregnancy to try to block transmission. Treatment was begun after the first trimester to minimize any risk of birth defects related to drug exposure. The mother and fetus were monitored intensively during pregnancy, and the neonate was monitored for drug toxicity since limited data were available on toxicity.

The requirement to include in the consent form any foreseeable impact on the fetus was added to the regulations in 2001. I looked back at the sample informed consent document from PACTG 076, which was six single-spaced pages long. It emphasized that we did not know whether this participation would benefit the woman or the fetus/child and that we had no idea of the long-term effects of this drug on the fetus or, subsequently, on the infant/child. Most of the risk discussion focused on the side effects of ZDV and the risk of blood draws.

What about paternal permission? Then and now, paternal permission would be required for this trial based on the current code of Federal regulations. There was clearly greater than minimal risk involved in using this drug. Since the trial was designed to exclude women who had their own indications for ZDV, it would be potentially of benefit to the fetus but not necessarily to the mother. There could even have been some potential risk to the mother's own health later from the short-course ZDV in pregnancy.

One of the issues that came up (and I am speaking as a clinician who was enrolling people to the trial, not as a Government employee) was who decides what “available” means regarding locating the father. Wide variations of practice occurred among the sites. In some sites, if the mother was sitting in the exam room and the father of the baby was sitting in the waiting room, and the staff person asked if the father of the baby is available to sign this consent and the mother said no, the father would not be asked. In other sites, the staff tried to track the fathers down in prison and get the mothers to take the consent forms to the fathers to sign. This is a big issue and could be a source of much debate. It clearly was an impediment to enrolling some women into PACTG 076 and could be in

others. Trials with subjects who have HIV infection have the added issue of disclosure of the HIV status to the partner by virtue of seeking their consent to the trial. I do not advocate that people should not inform their partners of their HIV status, but it is not an easy thing to do. Sometimes the woman is no longer involved with the father of the baby, or she might have serious concerns about abuse or abandonment if she discloses her HIV status.

Additional issues are not necessarily covered in the regulations, but I think they should be discussed. There is the issue of equipoise, which surfaced earlier today in the discussions of the hormone replacement trials. Physicians were saying that women should be on hormone therapy, but now the results do not support that. Regarding PACTG 076, some asked how we could treat all these pregnant women when “only 30 percent” of the babies are infected. These protests rose to the level of impeding meetings at which we were trying to work on implementing the trial; there were so many protests that we were forced to adjourn the meeting and try again another day. The other side questioned how we could do a placebo-controlled trial if we really thought this worked. We clearly were at the point where we needed to do the trial.

An important lesson from experience with clinical trials is the need to engage the communities involved. We learned this with PACTG 076. To engage rather than antagonize the communities, it is necessary to include them from the beginning in the design of the trial and in

its implementation—to get the word out to the communities about why the trial is important, why it is being done, and what are the pros and cons.

A second major lesson also occurred in the setting of the ACTG, which to that point had conducted trials primarily with gay Caucasian men who showed up for appointments, had their own cars, and did not have children. To enroll women into this trial, sites needed to provide childcare and transportation, which would cost more than the ACTG 016 or 019 trials of ZDV had cost.

A third lesson is the importance of rapid availability of therapy once a treatment is proven efficacious. Groups were fairly forward thinking and tried to rapidly develop guidelines for use of ZDV and make sure that everybody knew about it. The guidelines specified the doses and explained that intravenous ZDV needed to be available in the pharmacies at hospitals that conduct deliveries.

In conclusion, it is obvious that pregnant women can and should be included in clinical trials. The study designs must be based on the best available data to both minimize risk and maximize benefit. Representatives of the communities must be involved in the study design and implementation. The paternal permission requirement remains a barrier to participation in clinical trials, and it is not consistent with guidelines regarding consent for participation in trials, which says that if there is potential benefit and greater than minimal risk for the child, only one parent must sign the consent form.

An important lesson from experience with clinical trials is the need to engage the communities involved. To engage rather than antagonize the communities, it is necessary to include them from the beginning.

Chapter Five

Recruitment and Retention of Women and Minorities in Clinical Studies

Moderators: Edith Mitchell, M.D., and Jackson T. Wright, Jr., M.D., Ph.D.

Recruiting and retaining women and minorities in clinical trials is a continuing conundrum for scientific research. Both presenters in this session discussed their studies of recruiting women of color and offered recommendations about what can be done to improve minority participation in clinical trials.

Dr. Brown presented recruitment issues and strategies using a health services research approach that included predisposing characteristics, enabling factors, and illness need factors that affect help seeking and clinical trial participation. In another survey of 1,200 community residents to ascertain understanding and perceptions of human subject protections in research, her study found that African Americans were both significantly less likely to believe that scientists followed those protection guidelines and significantly less likely to believe that African Americans received the same quality of health care as did Caucasians.

Dr. Bigby discussed factors that have been conceptual and structural barriers to minority women participating in clinical trials and other studies and what can be done to overcome them. Based on her studies, she noted that although much is known about the barriers and facilitators to participation in research for women of color as well as the different approaches that work, better methods are needed to institutionalize those successful approaches.

Issues and Strategies in Recruiting Minority Women Into Clinical Studies

Diane R. Brown, Ph.D.

Dr. Brown is Executive Director of the Institute for the Elimination of Health Disparities at the University of Medicine and Dentistry of New Jersey (UMDNJ). She is also Professor of Health Education and Health Behavior at the UMDNJ School of Public Health.

My presentation covers issues and strategies in recruiting minority women into clinical trials using a health services research approach. It is based on research that we have been doing on clinical trial participation. The purposes of the study, funded by the National Cancer Institute (NCI), were to examine the decisionmaking processes for breast cancer treatments and to assess race differences. We also wanted to understand the patient and health care system factors that affect the participation of African American women in breast cancer treatment trials. We used a health services utilization model developed by Ronald M. Andersen, Ph.D., University of California, Los Angeles. It includes predisposing characteristics, enabling factors, and illness need factors, all of which affect help seeking and clinical trial participation.

Our project involved three data collection methodologies. First, we conducted personal interviews with new breast cancer patients coming into a treatment facility. We interviewed them about their decisionmaking process related to treatment. We asked them if they had had a discussion with their physician about taking part in a clinical trial. Then we asked them about the other predisposing, enabling, and illness need characteristics identified as part of the theoretical model. Our second source of data consisted of medical information. It encompassed a clinician survey as well as medical records extraction. Specifically, we asked the 10 physicians on the breast cancer service to complete forms on each patient. We also asked the physicians to indicate whether they had offered a clinical trial to the patient and, if they had not, why not. Our third source of data came from the clinical trials office. Those data were used to validate enrollment in a clinical trial. We linked all of these together for our data analysis.

The study site was an academic medical center with a comprehensive cancer center in a major urban area with a large African American population. Thus, we assumed that we would have a large number of African American breast cancer patients coming through the treatment facilities. Our target population consisted of all newly diagnosed breast cancer patients seen during our 12-month accrual period.

Differences between African American and Caucasian breast cancer patients were uncovered with regard to predisposing factors, enabling, illness need, and help-seeking factors.

Of the 319 new patients identified during the 12-month accrual period, 195 took part in the survey; 134 of these were Caucasian, and 61 were African American. At the time of the study, there were 10 active clinical trials for invasive breast cancer, which included 4 adjuvant trials, 1 dietary intervention, 4 stage IV trials, and 1 radiation trial for local recurrence.

In our initial descriptive analysis, differences between African American and Caucasian breast cancer patients were uncovered with regard to predisposing factors, enabling, illness need, and help-seeking factors. We found that African American women were less likely to be married and more likely than Caucasian breast cancer patients to be heads of household. We also found racial differences in enabling factors: African American women had less education and lower incomes. They were less likely to be familiar with the term “clinical trial”; they did not necessarily understand the concept of research; and

they were less likely to know someone who had participated in a clinical trial. Also, African American women were less likely to have private health insurance coverage and more likely to have Medicaid.

Overall, the breast cancer patients expressed high levels of confidence in their physicians. However, there was a significant racial difference, probably linked to issues of trust. African American women said that they did not have as much confidence in their physicians as did Caucasian women. The percentages were 83 percent for African American women compared with 93 percent for Caucasians.

African American women were less likely to say that they had discussed a trial with a physician. The question we asked was, “Did you have a discussion with your physician about clinical trial participation?” We did not ask who initiated that discussion, but we wanted to know if they perceived that they had had such a discussion. There was a significant racial difference of almost two to one, with Caucasian women being more likely to state that they had had such a discussion with their physician. In the analyses, we also examined whether or not there were differences by race in being offered a clinical trial by a physician. The responses to this question came from the physicians. When we analyzed the responses from the physicians, it was apparent that they were more likely to offer clinical trials to Caucasian breast cancer patients than to African American breast cancer patients.

Racial differences were also evident with regard to illness need factors, which was pretty much as expected. African American women were more likely to have advanced disease (e.g., Stage 3 and Stage 4); they were significantly more likely to have poor performance status.

In terms of help-seeking factors, Caucasian breast cancer patients were more likely to get a second or third opinion and were more likely to seek alternative therapies. On the other hand, African American patients were less likely to have conducted research on available breast cancer treatments. With regard to other help-seeking resources, African American patients were less likely than Caucasian women to telephone the Cancer Information Service for assistance. Moreover, treatment decisions by African American patients were more likely to be influenced by relatives and friends than by husbands and physicians. That is not to say that husbands and physicians did not influence those decisions, but family members and friends were also important.

A multivariate logistic regression was conducted to examine the predictors of clinical trial participation. Two factors emerged as significant while controlling for race, age, stage of disease, education, and other factors: (1) the

patient having had a discussion with a physician about clinical trial participation and (2) the physician offering a trial. We conducted additional analyses to better understand what was occurring. Of the 195 women who completed the survey, 70 (35.9 percent) indicated that they had had this discussion with their physician; more Caucasians than African Americans so indicated. Of those 70 women who had discussed it with a physician, 59 (84.3 percent) said that a trial was offered to them, thus indicating that the likelihood of being offered a trial increases in accordance with having had such a discussion with the physician. Regarding clinical trial enrollment, of those to whom a trial was offered, 18 (out of 59 or 30.5 percent) were actually enrolled in a clinical trial.

Enabling factors appear to be most important in terms of clinical trial enrollment. Other predisposing factors and help-seeking factors were less important.

The physicians were queried about the reasons for not offering a clinical trial to particular patients. Among the most frequently mentioned response was that there were no protocols available for these patients. This applied to 29.1 percent of African American and 30.9 percent of Caucasian patients with breast cancer. Physicians also indicated that many patients were ineligible for the trials that were available (African Americans, 49.5 percent, Caucasians 40.9 percent). However, there were racial differences regarding ineligibility. Almost 50 percent of African American breast cancer patients were deemed ineligible for available trials compared with about 41 percent of Caucasian women. The reasons for ineligibility varied by race. Physicians generally said that the African American breast cancer patients were ineligible because of poor performance status or poor organ function or because they were likely to be noncompliant. For Caucasian breast cancer patients, physicians reported that their current treatment precluded participation or they did not need any additional treatment.

In conclusion, enabling factors appear to be most important in terms of clinical trial enrollment. Other predisposing factors and help-seeking factors were less important. Clinical trial enrollment is more likely to occur when women have discussed the trial with their physician and the physician offers the opportunity to enroll. However, our analyses show that African American women were less likely to have a discussion with their physician even when age, cancer stage, and education are

controlled. However, there were no racial differences in being offered a trial when age, stage, and education are controlled.

The findings from our analyses suggest that when research is part of a system of health care delivery, we must ensure that African American and other minority women are not excluded from research because of their perceived socioeconomic barriers or prejudgments of potential noncompliance. At the same time, when research is part of a system of health care delivery, we also need to ensure that African American and other minority women are included for the right reasons—that it is in the best interests of their health and well-being. It is important to mention again that African American and other minority women tend to have issues that may present challenges to their recruitment and participation in clinical research. We need to face that up front and devote the time and effort required to recruit these women and keep them in our studies. We need to support the instrumental needs of minority women in terms of transportation, employment, and childcare, and we need to recognize the cultural dynamics around religion, spirituality, family support, health beliefs, and other factors that may influence their clinical trial participation.

As researchers, we also need to be concerned about the recruitment presentation and the sincerity and respect displayed in ongoing interpersonal connections between the research staff and patients. Trust and trustworthiness are essential elements of the scientist-patient relationship. Examples of the importance of trust and racial differences in perceptions of trust are evidenced in other work that we are doing. Specifically, as researchers and health care providers, we are aware of the human subjects regulations

for research subjects. All of us have probably taken and completed the training that describes the Federal regulations for the protection of human subjects and how these regulations should be followed. But how many people here believe that most scientists follow those regulations? Most of us feel that way. Yet, the general public, and in particular minorities, do not necessarily have the same perceptions.

We conducted a survey of 1,200 community residents to ascertain their understanding and perceptions of human subjects protections in research. The results were as follows: 38.7 percent of African Americans and 40.5 percent of Caucasians stated that they were aware of the Federal regulations for human subjects protection. The racial differences were minor and not significant. We then asked them if they believed scientists followed those regulations. Surprisingly, less than 50 percent of African Americans believed that scientists follow these regulations in contrast to almost 80 percent of Caucasian respondents. Another indicator of racial differences in levels of trust was evidenced when we asked survey participants to indicate the extent to which they believed that African Americans receive the same quality of health care as Caucasians. Responses indicated that only 43.2 percent of African Americans perceived that African Americans receive the same quality of healthcare as Caucasians. On the other hand, almost two-thirds of Caucasians believe that African Americans receive the same quality of health care as Caucasians. These are clear examples of where science meets reality and underscore the continuing challenges to increasing the participation of minorities in clinical trials.

We must ensure that African American and other minority women are not excluded from research because of their perceived socioeconomic barriers or prejudgments of potential noncompliance.

Recruiting Minority Women Into Clinical Research Trials: The Intersection of Race, Ethnicity, Gender, Community, and Academia

JudyAnn Bigby, M.D.

Dr. Bigby is Medical Director of Community Health Programs at Brigham and Women's Hospital, Director of the Harvard Medical School Center of Excellence in Women's Health, and Associate Professor of Medicine at Harvard Medical School.

I will talk about factors that have been barriers to minority women participating in clinical trials and other types of studies and what can be done to overcome these barriers. To set the stage for understanding why these barriers exist, I offer a historical context of the health care system and women of color. As William Faulkner has written, “The past is not dead; it’s not even past,” and that is why I believe it is important to examine these issues. Regarding the history of exploitation of women, especially minority women, by the health care system, we can go back to Dr. J. Marion Simms, who has been referred to as the father of modern gynecology. Dr. Simms perfected his surgical techniques on African slaves whom he purchased solely for the opportunity to practice surgical techniques; some people say he did not use adequate anesthesia.

In 1929 the Supreme Court ruled that poor women could be sterilized without their consent. The rationale offered by Oliver Wendell Holmes was that it was in the best interest of the state to sterilize women with hereditary defects, because it was believed that poverty was the consequence of a person’s inability to function in society due to hereditary deficit. During the 1960s the U.S. Government endorsed a sterilization campaign on the island of Puerto Rico. I recommend the film *La Operación*, which documents the campaign. Puerto Rico had the highest concentration of sterilized women of any area in the Western world as a result of this campaign.

In the 1970s headlines in the *Boston Globe* and *The New York Times* described the excessive rates of hysterectomies among African American women. Medical students at prominent medical schools in Boston and New York went to the press suggesting that women were undergoing hysterectomies without informed consent and for the purpose of providing medical residents with the

opportunity to practice their surgical techniques. We also know that many hormonal contraceptives have been tested in women of color in developing countries, such as Bangladesh and Mexico. In the 1980s and 1990s pregnant African American women disproportionately underwent involuntary drug testing for cocaine because of the crack epidemic. Epidemiological studies showed that pregnant African American women were no more likely than pregnant Caucasian women to use drugs. This history provides a lot of ground to understand mistrust of the health care system.

The definition of what constitutes research is a significant issue. There are discordant beliefs about that between researchers and potential participants.

Many people have already talked about the barriers to minority women’s participation in clinical trials. I have divided these barriers into conceptual barriers and structural barriers. We have already heard that mistrust is a major problem; Dr. Brown presented some of her data, and other studies indicate that the level of mistrust is significant. Mouton and colleagues found that 32 percent of African American women say that they do not trust researchers compared with only 4 percent of Caucasian women.

Lack of knowledge about clinical trials, which may be due to lack of information made available, is a major problem. Negative beliefs about research—the belief that research is not helpful to individuals and that it cannot advance care for individuals—are also a significant problem. In some of the work we have done, women indicated that they believed that the informed consent

process was there to protect the researchers rather than to provide informed consent to participants. Fear of health consequences is also a problem—the concern that research is harmful as opposed to being beneficial or even neutral.

We found discordant beliefs about the meaning of research. One African American woman participant in one of our focus groups described the fact that her provider prescribed hormone replacement therapy (HRT) to her, as a form of research. She was aware of the fact that, at that time, published case control studies did not include many women of color. She believed that no data existed to suggest that she, personally, would benefit from HRT. Other women have described “completing patient satisfaction surveys” as research to which they did not consent. The definition of what constitutes research is a significant issue. There are discordant beliefs about that between researchers and potential participants.

We found that women of color are not so concerned about increasing the percentage of minorities who participate in clinical trials. They believe that other priorities are more important, such as increasing the number of minority health care providers, increasing the quality of care that minorities receive, and increasing resources for health care in their communities. They also have differing perceptions about health and wellness and what is important. Their major concern is that, if minorities are not represented in the design and implementation of the research, they are not sure that research results will be valid for them.

Structural barriers have been mentioned in other presentations at this meeting: the logistics of transportation and childcare, conflicting roles, the role of caregiver, and being a vulnerable population. We know that women of color are much more likely to be caretakers of an elder parent, a spouse, or a child in their home than are Caucasian women. Other structural barriers include language and literacy, lack of flexibility, burdensome procedures, and the fact that women simply are not asked to participate.

Another issue is inadequate data to identify potential populations. We are struggling with this problem in most institutional databases because we do not have accurate or complete information about patients’ races and ethnicities. Many recruitment procedures are built on lists of people with specific diseases or who engage in particular health care activities, so we often do not know the race or ethnicity of potential participants. This lack of information is an impediment to recruiting underrepresented populations.

We also know that some factors facilitate participation. Some women of color expressed the desire to “do good”

Reasons Black Women Participated in Women’s Health Demonstration Project

- Easy access to study site
- Wanted to find out more about own health
- Research staff recommended participation
- Provider recommended participation
- “Researcher” of own age and ethnicity talked about study

and to serve their family or community as a reason for participating in research. Other factors that facilitate participation included improved access to care or that a trusted physician endorsed the research. This raises an issue for physicians in academic medical centers, in which the primary care or treating physician is asked to cosign (with the principal investigator) letters recruiting patients. When patients receive multiple letters signed by their doctors, what does that do to the trust between the treating physician and the patient? Having community leaders endorse research also facilitates participation, as does previous experience with research.

About 18 months ago, the Harvard Center of Excellence in Women’s Health sponsored a 1-day meeting between Harvard researchers and representatives from communities of color. This meeting focused primarily on women in Boston. Participants spent the day examining factors that would facilitate minority women’s participation in research. From our conversation with the community, the Center of Excellence developed some specific recommendations for recruiting minority women into clinical trials. One possible solution is a centralized process for identifying underrepresented populations. A centralized unit would conduct outreach, develop recruitment materials, modify informed consent documents, explore alternative ways of delivering informed consent such as producing videotapes, and study how the forms are experienced conceptually by underrepresented populations. The level of expertise required to accomplish this centralized recruiting may not be available in most departments or research units, and it may be unrealistic and inefficient to ask every researcher to develop expertise in these methods.

Some of our researchers have developed patient registries, specifically of people of color who are willing to be contacted for research related to different diseases. This took a long time to develop, working with a coalition of multicultural organizations. The advantage to the

community is that its members became more involved in outreach and in educating their community about different health issues, such as Alzheimer's disease, diabetes, and other topics.

Lack of coordination efforts is another barrier. Women told us that it is confusing when researcher A who is doing a breast cancer study and researcher B who is doing a breast cancer study and researcher C who is also doing a breast cancer study all try to recruit them at different times. They do not know how the research studies differ, and having multiple investigators who are not working together approach the same population is detrimental in terms of building trust.

Another significant recommendation is that study protocols should reflect an understanding of the structural barriers related to participants' time and transportation. Scheduling interviews and other activities required for participation should take into account geography and times that are convenient.

The meeting participants also saw a major role for institutional review boards (IRBs), which should pay more attention to protocol design and to informed consent language. They concluded that IRBs could do a better job to ensure they promoted research designs and included language in informed consents that benefited minority women. The participants also recommended that researchers and the research staff should reflect the population being studied and that the number of minority women researchers should be increased. Institutions should develop specific goals for improving the representation of minority women in research trials and should track their results and evaluate their strategies to understand why they are failing or succeeding. They emphasized that what works for recruiting one population of minority women may not work for others; one size does not fit all.

The meeting participants recommended that researchers and the populations they want to study develop partnerships to address the barriers. Advisory committees of minority women are helpful in providing advice about research design, interpretation of data, and dissemination of results. Minority women should be more involved in identifying important research questions so that these can be tagged on to the research efforts already planned. All instruments and procedures should be tested

in the community, on multiple populations, before being finalized. Women of color should be recruited and trained to staff clinical trials at all levels of staffing.

Meeting participants thought that addressing the sustainability of research trials was an important issue and that it should be addressed up front, not just at the conclusion of the trial. For example, if you are asking people to participate in a trial that studies the effects of a drug, who is going to pay for that drug after the study period is completed? Other potential poststudy benefits should be pursued, in terms of resources that could be contributed to the community once the study is over.

We are currently involved in a collaborative project to look at disparities in breast and cervical cancer mortality among African American women in Boston. This research is being conducted in collaboration with the Boston University Center of Excellence in Women's Health and the City Health Department in Boston. The goal is to identify factors that contribute to disparities at the personal, institutional, and community levels. One of the studies looks at personal contributions to disparities, involving six clinical sites. In this study we are trying to recruit women between the ages of 18 and 75 who self-identify as African American. However, the population that we want to study is women who have traditionally not been connected to the health care system. To get at that population, we are looking at women who have frequent emergency department visits, who have not kept health care appointments, who often walk in for care, and who are out of compliance with pap smear and mammogram screening—not an easy population to recruit into clinical trials. Potential participants have to agree to an initial hourlong interview and to do followup interviews over 4 years, so they have to be reachable. We want access to their medical records, and we also give them the opportunity to have access to a case manager. By administering a complicated and long assessment tool, we identify social and medical risk factors that may be impeding their ability to participate in health care.

The research staff in this project includes seven African American women of different ethnic backgrounds, who act as research assistants and help with recruitment at each of the clinical sites. We also have five African American women outreach workers who have community connections and who help to get the word out about this

Researchers and the research staff should reflect the population being studied and the number of minority women researchers should be increased.

study. These women are multilingual and multicultural, and they include women who are of African American, Haitian, Cape Verdean, and Somalian descent, which represents the diversity of the community of African American women in Boston.

The outreach workers go to churches, health fairs, people's homes, nail salons, and hair salons, to talk—not necessarily initially about this study but about health issues facing African American women and breast and cervical cancer disparities. Clinicians involved in the study accompany the outreach workers so they can address clinical issues or health concerns that come up. All of these women underwent special training that included understanding research design and implementation, ethics, IRB issues, cultural competency, antiracism training, communication skills, and understanding breast and cervical cancer.

Our goal is to recruit 1,200 women, and we have recruited nearly 500 women in less than a year, 94 percent of whom are African American. Other characteristics of the study population include the following: 95 percent had never participated in research before, 63 percent were recruited by the research assistants, and 23 percent were referred by their providers. The reasons for participating included easy access to the study site, belief that the study would help them find out more about their own health, trust in the research staff members who recommended participation, recommendation from their providers, or the fact that the research staff represents the population we are recruiting in terms of age and ethnicity.

There are many implications arising from experience and recommendations. Researchers need more time up front for planning and working with different groups to organize the research design and strategies. The project timeline needs to reflect this, and the research may take longer to complete. We heard Dr. Healy mention this morning that a researcher complained that it took longer to complete a cardiovascular study because they had to recruit women. When you are talking about women of color, it can take even longer. In the end, the research is more appropriate, and the results more valid. Researchers require additional financial resources to recruit women of color. The budget should include resources for paying outreach workers and for the training.

When researchers take the time and effort to recruit women of color, it takes longer to complete projects.

From a personal point of view, when researchers take the time and effort to recruit women of color, it takes longer to complete projects. This translates to a longer time until publication and other end points of success, which may affect some people's academic careers.

We know a lot about the barriers to participation in research for women of color. We also know a lot about facilitators. And we also know that different approaches do work. We just need better ways to institutionalize them.

Chapter Six

Inclusion:

Policy, Ethics, or Justice—Is It Time To Reexamine the Guidelines?

Moderator: Darlene Yee-Melichar, Ed.D., CHES

Inclusion considerations are fundamental to the design of any research investigation involving human subjects. Institutional review boards (IRBs) are charged with responsibilities for ethical review and oversight of the use of human subjects in research protocols. The primary principle governing the IRB's review and action is the protection of human subjects from risks while permitting the advancement of research. IRBs are faced today with the emerging ethical and policy issues that present both challenges and opportunities for women's health research.

This session focused on inclusion policy, ethics, and justice in women's health research. The panel provided an authoritative and provocative discussion of challenging issues and opportunities for inclusion in women's health studies. Ms. Katz reflected on her extensive experience with policy and inclusion of women in clinical studies, Dr. Lyerly shared her perspectives on ethical dilemmas of inclusion in clinical research, Dr. Kahn discussed his thoughts on justice in research, and Ms. Charo provided comments from the perspective of a feminist IRB member.

Policy on Inclusion of Women in Clinical Studies

Ruth J. Katz, J.D., M.P.H.

Dr. Katz is Associate Dean for Administration at Yale University. In the early 1990s she served as a congressional staff member on Capitol Hill where she worked on various women's health issues and, in particular, on the legislation that led to the National Institutes of Health (NIH) guidelines on including women in clinical trials.

The NIH policy on including women in clinical trials can be viewed from three perspectives: (1) the history behind the 1993 Revitalization Act and the NIH inclusion policy, (2) Congress' intent in mandating that women and minorities be included in NIH-sponsored clinical trials, and (3) whether we have succeeded in meeting both the letter and the spirit of the inclusion policy in that 1993 law.

In 1990 when the debate over women's health was getting under way on Capitol Hill, members of Congress had three overriding concerns. Among men and women members of Congress a strong interest was growing about women's health research at the NIH: research on breast, cervical, and ovarian cancers; research on osteoporosis; and research on autoimmune diseases such as multiple sclerosis, all of which primarily affect women. Members of Congress were concerned about the famous—some would argue “infamous”—1990 Government Accounting Office (GAO) study that looked at women's health from the perspective of including women in NIH clinical trials. Members of Congress had become increasingly concerned about the fact that the 1986 NIH Guidelines, which were the subject of the 1990 GAO study, were only voluntary and could not be enforced.

The 1990 GAO study had a major effect on policy makers. It came on the heels of an important study on aspirin and heart disease, a study that included only men despite the fact that heart disease affects both sexes significantly. The GAO study presented three major findings: (1) the NIH had not adequately implemented its own 1986 policy that simply “urged” researchers to include women in clinical trials, (2) the NIH did not have the tools in place to measure the impact of its own 1986 policy, and (3) the 1986 policy applied only to NIH extramural, not intramural, research. Response to this

GAO study was far greater than anyone anticipated; it was overwhelming and made national headlines and there were many calls from Capitol Hill.

The 1993 Revitalization Act included five major requirements regarding NIH clinical research and women and minorities in particular: that women and minorities be included in all research sponsored by the NIH, that clinical trials be designed and conducted in a manner that would indicate any differences between men and women in the variables studied, that the NIH institute outreach programs to recruit women into its clinical trials, that the NIH issue guidelines that explain how the new inclusion policy would be carried out, and that the NIH establish a data system to provide information relevant to women's health research.

The 1990 GAO study had a major effect on policy makers. It came on the heels of an important study on aspirin and heart disease that included only men, despite the fact that heart disease affects both sexes significantly.

The law was followed in March 1994 by the guidelines, published in the Federal Register. The introductory comments to the guidelines captured the spirit of Congress' intentions and then laid out in detail how the inclusion policy would be carried out. This raises the second policy question: What did Congress actually intend in enacting the inclusion requirements? The language in the House of Representatives committee report that accompanied the 1993 law expresses what Congress was seeking: It makes clear that Congress intended to create a

presumption that women and minorities are to be included in NIH-sponsored clinical trials, and that not doing so could only be based on science, not cost.

Moreover, Congress made it clear that simply including women in clinical trials was not enough. Congress also intended that studies describe the implications of their findings for both sexes and, therefore, Congress mandated that clinical trials include a “valid analysis” that would measure any differences that might be discovered during the studies. Congress left the definition of “valid analysis” up to the NIH, and it was spelled out in the March 1994 guidelines. As this language from the House committee report makes clear, a valid analysis includes certain basic parameters; it must answer basic questions such as whether differences between men and women exist, what those differences are, and whether those differences are significant. Congress also made clear what it did not intend in establishing the inclusion policy: that this policy was not and is not a policy about quotas (the argument used to try to stop the original 1992 bill from moving forward). Congress intended a general principle of inclusion.

What has been the progress during the last decade? Has the NIH progressed to where Congress intended it to be? Some examples of NIH-sponsored research that demonstrate different responses between men and women are symptoms of heart disease, biological mechanisms involved in drug abuse and dependence, response to pain relief, and manifestations of HIV disease (especially viral load). This is just a small sample of the work that the NIH has sponsored during the last 10 years. The 1993 inclusion policy is directly responsible for these studies occurring in the first place, and for the results that were ultimately discovered.

I would say the same thing about the study results that come from work sponsored by my own institution,

the Yale School of Medicine, through its Women’s Health Research Program. Studies on recovery from heart bypass surgery, efficacy of smoking cessation programs, causes of depression, prevalence of undiagnosed diabetes, and genetic predisposing to alcoholism were funded with non-NIH dollars; however, the NIH inclusion policy drove these studies to be undertaken in the first place. I do not believe such studies would have been conducted at all without the 1993 NIH Revitalization Act. That law not only pushed the NIH to do the right thing, it also pushed other research institutions to take on women’s health research and to conduct it in a way that provides meaningful results for men and women alike.

A more recent GAO study, undertaken in 2000, has confirmed that important progress has been made in this area. However, the study also indicates that the NIH has not yet attained the goals set by Congress. In particular, additional work must be done to better implement the valid analysis requirement. The 2001 Institute of Medicine study entitled *Exploring the Biological Contributions for Human Health, Does Sex Matter?* underscores the need for additional work in this area.

In conclusion, regardless of GAO findings or Congressional reports, it is clear that the work related to women’s health research is far from complete. We should review and refine the NIH Guidelines if that is appropriate. In so doing, we cannot and should not forget how difficult it was to get where we are today, nor should we forget that some people would like to cut short this important work or end it altogether. The overall goal of improving women’s health is certainly within our reach, but only if the tools are in place to make it happen. The NIH Guidelines are one of those very important tools, so I suggest that we be very careful about how we think about the future of that document. The NIH Guidelines on Inclusion are too important to lose.

We cannot and should not forget how difficult it was to get where we are today, nor should we forget that some people would like to cut short this important work or end it altogether.

Ethical Dilemmas of Inclusion in Clinical Research

Anne Drapkin Lyerly, M.D.

Dr. Lyerly is Assistant Professor in the Department of Obstetrics and Gynecology and Faculty Associate in the Center for the Study of Medical Ethics and Humanities at Duke University.

Women can and sometimes do get pregnant. This fact has had a persistent but changing effect on women's health research, which has become particularly interesting in light of current technology and public policy. With advancing technology, growing interest in research on the fetus means the inclusion of more pregnant women in clinical studies. At the same time, there is another tendency: When the health and well-being of the fetus are subjects of concern, the woman has become at best blurry and at worst invisible in terms of research objectives. The outcomes relevant to pregnant women participating in research studies in which the health of the fetus is the primary subject of inquiry are not adequately measured.

The revised subpart B of the Federal Regulations defines the fetus as the product of conception from implantation until delivery. This means that a fetus by definition is always connected to and dependent on a pregnant woman for survival. Professor Mary Briody Mahowald, Ph.D., talks about a concept called the "fallacy of abstraction," which is the consideration of a concept as if its meaning were adequately expressed apart from other concepts in which that meaning is necessarily imbedded. Applied to the research noted above, this means that talking about a fetus without explicitly recognizing the pregnant woman in whom it resides is not only inaccurate but also practically and morally misleading and has implications for ethics and public policy regarding the health and well-being of women.

In the past, policy regarding human research subjects was in part driven by public outcry and tragedy surrounding the experiments at the Willowbrook State School and the U.S. Public Health Service Tuskegee Study of Untreated Syphilis in the Negro Male, which led to protectionist policies. With women in the 1960s, this manifested itself as a fear of the horrors associated with thalidomide, a drug that was approved for marketing in Europe for the prevention of nausea but that resulted in some devastating birth defects. As a result, policies were developed to exclude women from clinical research to protect their fetuses from the potentially teratogenic effects of experimental drugs. For instance, the U.S. Food and Drug Administration, until 1993, excluded all women of childbearing potential from early-phase clinical trials.

The result of such policies has been not protection but rather underrepresentation of women in the name of fetal protection. This is both morally and practically problematic: morally, because it is not acceptable to place a higher value on fetal life and well-being than on women's lives and well-being, and practically, because the policies did not achieve any meaningful goal. Pregnant women get sick, and they need to use pharmaceuticals, but little is known about what those drugs do in terms of harm or benefit to women or to fetuses. With the focus on the fetus, the woman becomes invisible; nobody is protected.

A shift from protectionist to inclusionary policies occurred during the past decade, and requirements for inclusion of women in clinical research emerged. However,

Applied to research, talking about a fetus without explicitly recognizing the pregnant woman in whom it resides is not only inaccurate but also practically and morally misleading.

something else has also happened; as one surgeon has said, a “whole new patient population” has emerged—those yet to be born. With inclusionary policies in place, the result is a new population of potential research subjects.

A comprehensive literature search revealed that, among the hundreds of articles that have been written about maternal fetal surgery, only two articles primarily address outcomes regarding women.

Available technology contributes to the misconception that a fetus can be considered as separate or separable from the pregnant woman. For example, 3-D ultrasound enables viewing the fetus in striking detail. The technology eventually may be helpful in diagnosing structural abnormalities such as cleft lip and spina bifida, but its impact now is more symbolic than clinical, sharpening the focus on the fetus and possibly blurring the focus on pregnant women even more.

Technology alone cannot be blamed for this perspective. Policy is being made with a similar thrust. In November 2002 the State Children’s Health Insurance Program (SCHIP) was revised to expand and clarify the term “child” so that a State may elect to make individuals from conception to birth eligible for coverage under the State plan. This sounds like a good idea—women can get prenatal care under an already existing program—but there is a problem: The unborn somehow emerges as a separate or separable individual from the pregnant woman, within whom it necessarily resides. Clarification of the new rule specifies conditions that would not be covered, such as medical conditions not directly related to fetal well-being. For example, if a pregnant woman breaks her leg, that is not covered under the plan. Postpartum complications and issues, hemorrhage, infection, counseling, screening for postpartum depression, contraception, family planning, and so on, are not covered. Anesthesia is covered because, without anesthesia, some problems could occur with labor and delivery that would affect the fetus (e.g., with a C-section). Again, the focus is on the fetus or baby, and the interests of the pregnant woman are ignored.

One justification for expanding the rule was that some treatments are unique to the fetus separate and apart from the pregnant woman, in particular techniques known as fetal surgery. (The term “maternal fetal surgery” has been advocated by several authors because it is more accurate, as it explicitly recognizes that surgery on the fetus neces-

sitates surgery on the pregnant woman.) This justification for expansion of SCHIP coverage was problematic for two reasons: First, maternal fetal surgery is research, not treatment; and second, it is practically and conceptually impossible to separate the fetus from the pregnant woman. When surgical interventions are directed at the fetus, the woman undergoes surgery, too.

A well-circulated image showing maternal fetal surgery depicts the surgeon’s hand, the fetal arm, and the bloody background of a woman’s uterus. The woman is there, we know she is there, but she may not be thought of as a patient or a research participant. Surgery to correct fetal anatomic abnormalities has been available for nearly three decades, but the practice has received much attention in the past 3 to 4 years. In 1999 two centers published a case series of in utero repair of spinal cord defects known as spina bifida. These spina bifida surgeries received attention because surgeons were repairing conditions that were not lethal to fetuses, unlike prior surgeries to correct fetal anatomic abnormalities. In addition, significant problems had been detected in the research procedures used by the surgeons and reported by the centers.

Maternal fetal surgery involves making an incision in the pregnant women’s abdomen and uterus, exposing the back of the fetus, repairing the spinal cord defect, sewing up the pregnant woman, putting her on prolonged bed rest and an anticontraction medicine, and then delivering the baby by C-section near term. The studies had many problems that were addressed at an NIH workshop in July 2000. One significant problem was that, despite the fact that pregnant women were clearly research subjects in their own right because they had undergone at least two traditionally unnecessary surgeries, maternal outcomes were not considered when the risks and benefits of intervention were balanced to justify the research endeavor, and no data regarding maternal outcomes were systematically collected according to the published reports. Again, the focus on the fetus made the woman invisible. It was at this conference that the term “maternal fetal surgery” was first used.

Ultimately, the randomized controlled trial proposed to study this intervention will correct for the failure to collect outcomes regarding women, but this was actually not a new problem. A comprehensive literature search revealed that, among the hundreds of articles that have been written about maternal fetal surgery, only two articles primarily address outcomes regarding women. The imagery and vocabulary involved in these experimental therapies show no recognition, explicit or otherwise, that a woman is present.

Remedies

- Explicit recognition that research on the fetus is research on a pregnant woman
- Requirement for measurement of outcomes regarding pregnant women who are undergoing research aimed at fetal well being
- Studies regarding the attitudes of women regarding research on their own products of conception

In summary, the tendency to focus on fetuses means that the interests of women are not considered in the conduct of clinical research. Earlier, with protectionist policies, women were not enrolled in studies at all. Now, with inclusionary policies, data are not collected regarding the effect of interventions on women who are enrolled in clinical research for fetal indications.

I propose a few remedies, a “wish list”: (1) Research on the fetus should be explicitly recognized as research on a pregnant woman, (2) a requirement should be established for measuring outcomes regarding pregnant women who are undergoing research aimed at fetal well-being, and (3) studies should be conducted on women’s attitudes regarding research on their own products of conception, whether that is a fetus that resides inside them or embryos for use in stem cell research.

Justice in Research

Jeffrey P. Kahn, Ph.D., M.P.H.

Dr. Kahn is Director and Professor at the Center for Bioethics at the University of Minnesota.

The concept of justice as applied to human subjects protection has evolved since the middle of the 1970s, a period known as the Belmont period because of The Belmont Report (1978), which was issued by the National Commission on Protection of Subjects in Biomedical and Behavioral Research and which first articulated the ethical principles for protection of research subjects in the United States.

What was the perspective on justice in the Belmont era? Policy was driven then by a concern to protect research subjects from harm and exploitation growing out of the unfortunate cases of the misuse of subjects experienced in the U.S. Public Health Service Tuskegee Study of Untreated Syphilis in the Negro Male, which was conducted in Macon County, Alabama, and of the children in the Willowbrook State School hepatitis study, among others. The subjects in these and other studies either were deceived into participating and/or did not give adequate informed consent for their participation. From those experiences came a sense that we needed to do a better job of protecting those who were considered “vulnerable subjects.” The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research talks about justice as being in part about the fair distribution of the risks and benefits of research, but the emphasis was on risks, and that was incorporated into policy as special protection in the Federal regulations for prisoners, for children, and for pregnant women as research subjects. These three populations were deemed particularly vulnerable, in part due to the experiences of the past.

Only 15 to 20 years later, a sea change occurred in the way we have come to think about justice to increasingly emphasize access to the benefits of research, the flip side of concerns about risks entailed by research.

Thus, in the later sense, fairness is about access to the benefits and, therefore, inclusion in the research enterprise. That perspective came through clearly in advocacy efforts around issues such as HIV/AIDS and in quite famous slogans by advocacy groups such as AIDS

We have come to think about justice to increasingly emphasize access to the benefits of research, the flip side of concerns about risks entailed by research.

Coalition to Unleash Power (ACT-UP), which organized protests in large cities around the country, where marchers carried placards that said, “Clinical trials are health care, too.” It was a very strong statement about wanting and demanding access to research. We can talk about whether we think that statement is true, but for people who did not have any other access to therapy for a fatal disease, it was true, and it makes a strong statement about wanting and pushing for policy that guarantees access to research trials and to the benefits that research offers. Similarly, there was strong advocacy for research on breast cancer and the ability to participate in those trials for parallel reasons.

Over the same 20-year period, there was a warranted perception that research, in many cases—for instance, for childhood cancer—offered truly cutting-edge therapy, and there was a push to tell people that the best way to get treated for some diseases was to participate in a clinical trial. As a result, the public began to believe that participation in clinical trials was the way to obtain the best and most up-to-date medical care. This led to an increasing sense—expressed in the scholarly literature—

that policies had led to an effective overprotection of some groups. R. Alta Charo, J.D., wrote that research policies had effectively protected women to death. By keeping women out of research, they were protected from the risks, but they did not receive the benefits yielded by research participation.

This change in public perception, scholarly debate, and advocacy led to a change in the way we thought about the policy picture, and the emphasis changed to include the benefits of research for individual subjects and the groups to which they belong. The individual's perspective was that, by participating in a clinical trial, she would have a chance of realizing some benefit. Obviously and importantly, we also learned much more about women's health when women participated in clinical trials. This history culminated in the policies we all know about. In the early 1990s policy required that women and minorities be included in research, with justification required when they are excluded. Similarly, there is an ongoing effort and much discussion about including children in research because we do not know a lot about how children metabolize drugs. To round out the policy perspective, during the budget allocation discussion before Congress in 2000, Richard Klausner, M.D., then Director of the National Cancer Institute (NCI), said (and this was quoted in *The New York Times*) that one reason the NCI portion of the NIH budget request was necessary was to ensure that all people who wish to participate in a clinical trial are able to do so—research participation for all. That is a remarkable statement in a country where there is no such guarantee for access to health care.

Other policy changes increased research participation. In 2000 President Clinton issued an Executive Order requiring the Health Care Financing Administration (now the Center for Medicare Services) to allow Medicare

recipients to participate in clinical trials at Government expense. That is now Federal policy. Even earlier, policy changes in 1993 and 1994 allow research to be conducted in an emergency context without the informed consent of the individuals participating. This provision was the effective stepping away from that cornerstone requirement of informed consent, partly justified on the grounds that the benefits of research participation were deemed to be so valuable that we needed to let people who cannot give consent be research subjects.

Given these developments, are we now in a new era of protection for human subjects? Certainly it is important to include individuals and groups as subjects of research, to make the benefits of research available to them. But at the same time, we must remember that protection should not be far from our minds and that it is not that long ago when we were worried about exploitation. Research inherently carries risk and does not always—and often does not at all—provide benefit. So the goal must be to ensure an appropriate balance. We have heard a lot today about the confusion of “patients” and “subjects.” Patients are people who receive clinical care; “patient-subjects” is the more appropriate term for subjects who suffer from an illness or disease and participate in clinical trials trying to develop new approaches for treating them. “Research” and “therapy” are not the same thing, and there is much misunderstanding about that. The people who care for patients are often researchers as well, and that is a conflict that leads to confusion on the part of individuals recruited to participate in research.

In conclusion, we have to ensure equitable access to research participation and the benefits it offers, at the same time protecting the rights and interests of the research subjects. If we do not do that, we risk undermining the trust of the public, and that is far too high a price to pay.

We have to ensure equitable access to research participation and the benefits it offers, at the same time protecting the rights and interests of the research subjects.

Comments From a Member of a Feminist IRB

R. Alta Charo, J.D.

Ms. Charo is Professor of Law and Bioethics at the University of Wisconsin, Madison (UW), where she is on the faculty of the UW Law School and the UW Medical School's Department of Medical History and Bioethics. She teaches courses on health law, bioethics and biotechnology law, food and drug law, medical ethics, reproductive rights, torts, and legislative drafting. Ms. Charo has served on the UW Hospital and Clinics clinical ethics committee, the institutional review board (IRB), and UW's Bioethics Advisory Committee.

During the past 12 years at the UW, I have had the privilege of being first a feminist as a member of the IRB and then, more recently and more gratifyingly, a member of a feminist IRB. This transformation is telling because it is part of a slow trend across the United States of how we view the inclusion of women in research trials, particularly fertile women, and, hopefully in the future, pregnant women. I will offer some observations about that history and where I hope it will go.

The major challenge to an IRB member who has a feminist agenda is that every other area of expertise you might have is likely to be obscured by the perception that you are a one-issue member and that your comments will always be filtered through that lens. A touchy problem as you plan to be an advocate for a point of view is learning how to sense when people start tuning you out. At times, you even develop humorous ways of signaling the fact that yes, once again, you are going to dissent from a vote and force everybody to record the ayes, the nays, and the abstentions instead of simply raising hands in anonymity.

What are some effective approaches to dealing with this challenge? Being on good terms with your colleagues turns out to be very important in letting your message get through and not having it blocked by irritation at your politics. Educating new IRB members as they arrive is important. It has been my observation that, as the older members of the physician staff at my institution join the IRB, each one needs to be educated about the normality of including women (especially fertile women) in research, because each new member begins with the notion that there should be zero risk to fetuses, that is, that there is no particular value to including women in research, therefore, there is nothing to offset even a microscopic risk to the fetuses.

In addition, many new IRB members are nervous and

uneducated about the prospect of personal or institutional liability in the event of fetal injury. They are completely ignorant, most often, of the amount of accumulated evidence that indicates real differences between men and women in how they respond in clinical trials. Taking the time to describe these phenomena—such as the wonderful anecdote about the pain killers in the dentist office several years ago and statistical anomalies such as Simpson's Paradox—to the new members, whether outside the meeting or in the course of a discussion, is

The major challenge to an IRB member who has a feminist agenda is that every other area of expertise you might have is likely to be obscured by the perception that your comments will always be filtered through that lens.

very helpful in bringing them forward to the IRB's level of comfort about including women in clinical trials. These arguments are particularly important in the context of nontherapeutic research when an IRB tries to ensure that the generalizable knowledge gained reaches all people, thus maximizing the benefit of the research. Initial thinking is often that, if the benefits outweigh the risks for any population, the standard has been met; researchers are supposed to maximize this ratio to its best advantage, which means producing knowledge that is generalizable to the full population, including women. The argument that there is potential therapeutic advantage for women who participate in a trial is less persuasive in educating new members because, quite

rightly, it is pointed out that the trials exist because there is no certainty about any benefit whatsoever.

After you have educated these new members about how important it is to include women, it is important to avoid including women indiscriminately without paying attention to the special issues that arise because of their presence in the trial. For example, the inclusion of adolescent girls is a situation that deserves some special attention; it is hardly antifeminist to mention the possibility that adolescents might get pregnant but be reluctant to reveal this to study coordinators, so this possibility should be considered in the review of the study protocol. Nor is it antifeminist to know that adolescents can get pregnant when they do not expect to and that therefore one ought to plan for adventitious pregnancy in adolescents when they are enrolled in studies. These events often happen after the study protocol and design have been completed and the IRB discussion has concluded. It becomes a kind of minicrisis as the investigators and their staff members try to figure out how to handle it. Emergency meetings are called with the agency or with the IRB staff or chair. It is far better if these kinds of events can be anticipated and some routine provision can be made for how to handle the event, particularly when interested parents are involved.

These kinds of banal, everyday observations about life as a woman often can make the difference between consent that is informed and consent that is only perceived to be informed.

A third and more banal continuing challenge is explaining to an IRB (whose membership is predominantly men) those aspects of women's lives that will be relevant to the consent process and ensuring that the consent process addresses the issues and information women need to make a decision. For example, in describing risks and benefits, often what are considered to be low-level harms (such as fatigue) are omitted or glossed over. However, fatigue that makes a woman too tired to handle work and childcare is probably a very significant factor in her decision about whether to enter a trial. These kinds of banal, everyday observations about life as a woman often can make the difference between consent that is informed and consent that is only perceived to be informed.

There are other "big-ticket items" with which my own IRB has struggled. My most recent dissenting vote had to do with a protocol that involved oophorectomy for the

prevention of breast cancer. I voted this way because of the difficulty of effectively communicating the degree to which sterilization is a significant life event for women and therefore should not be treated casually. Protocols that involve sterilization must be discussed and reviewed with the greatest possible rigor.

Those are some of the existing challenges. In addition, there are three areas where I would like to see the IRB move in the future. The first one concerns not the IRB but rather the field: When is the inclusion of women in a trial genuinely important, and when could it be foregone, based on data available on men or from sterilized or post-menopausal women? In other words, how can we use data predictively to know in which areas or for which substances or devices one is likely to find some difference based on sex? With that information, one might take a more nuanced approach to the inclusion of women and the issue of possible fetal exposure. Without it, we will have to choose between two draconian responses, the overall inclusion or the overall exclusion of fertile women, neither of which is entirely satisfactory.

The second area seems simple and yet has been surprisingly controversial in many settings: to provide contraceptives onsite when recruiting women who are fertile but who should not get pregnant during the clinical trial; or, if a physical examination is required that is beyond what is available at the site, to refer individuals to a local clinic that provides these exams and contraceptives. I have been shocked at the resistance I have received to the suggestion that, if the protocol says "you should not get pregnant," the trial's recruiter should have a bowl of condoms and, if being conducted in a setting in which this is appropriate, should be qualified to provide Depo-Provera shots that will provide secure contraception for several months, a prescription for a contraceptive pill, or a referral to a local clinic. Yet this kind of inclusion seems to raise hackles and excuses, such as "this is not our job," "this is too controversial," and "this is federally funded so I do not think they would let us do this." It seems to me that, if there is sincere interest in both recruiting women and minimizing the possibility of unintended fetal exposures of unknown risk, this approach is appropriate.

The third, and most controversial, area is to encourage the testing of drugs and devices on pregnant women. I am referring to learning how drugs and devices operate in the bodies of pregnant women so that you and I and the rest of our friends can finally take Sudafed for our head colds while we are pregnant, instead of being terrified of anything that comes in pill form. For most women, pregnancy now means that there will be 1 to 3 years in her life where she is unable to take care of routine medical problems;

only the most severe conditions will be treated because only then will the physician and the woman feel comfortable that, even with fetal exposure, there is ample reason to take medications. This approach reflects our current absence of information and the continuing fear concerning fetal exposures.

Buried in our regulations is an assumption that fetuses that are destined to go to term should be treated exactly the same as fetuses that are destined to be aborted. This means that we cannot take advantage of what is undoubtedly a provocative and controversial area of research, which is to recruit women who have decided to have an abortion. In many States, women must now wait 24 hours before having an abortion, thus providing a

window of time in which to recruit them into clinical trials. In those trials, researchers could examine basic aspects such as the rate of metabolism of various substances to understand how these drugs operate in women's bodies and begin to make some predictions and some better medical decisions about the use of these drugs for pregnant women. This may not be the highest priority area, but it is perhaps the last big challenge to getting past the notion that women are simply too complicated or atypical for inclusion in clinical studies. It is time to bring pregnant women back into the fold and accept that they, too, are part of the normal variations of humankind for whom drugs and devices and biologics are being developed.

It is time to bring pregnant women back into the fold and accept that they, too, are part of the normal variations of humankind for whom drugs and devices and biologics are being developed.

Chapter Seven

Pharmacogenomics and Implications for Clinical Research: Is It Time for a Paradigm Shift?

Moderator: Rosalie Sagraves, Pharm.D., FCCP

Our understanding of the human genome is opening up new avenues of science at a rapid rate. Developing technologies such as pharmacogenomics and other “omics”—including proteomics, interactomics, lipomics, and glycomics—will affect the future conduct of clinical trials and, therefore, the recruitment and retention of research subjects. These types of research also will require the development of large databases through bioinformatics. Dr. Miele presented an overview of the various “omics” and how they will help change the way clinical research is performed. Dr. Benet shared his research in pharmacokinetics and pharmacodynamics in which he has demonstrated how women and men respond differently to certain medications, especially those that are hepatically metabolized via enzymes. More recently, his research has explored the genetic variation of drug transporters that may explain differences in individuals’ responses to drugs. As science moves forward and clinicians are able to conduct subpopulation analysis, this may change how patients are recruited to and monitored in clinical trials. Dr. Wood presented the U.S. Food and Drug Administration’s views about bringing regulation together with science and public issues that affect the health of U.S. women.

These presentations exposed participants to new sciences and technologies that should enable scientists, clinicians, and regulators to determine for whom to use currently available medications to obtain the best clinical outcomes with the lowest possible numbers and rates of adverse effects. New medications may also be developed that are targeted to appropriate populations with specific diseases.

The Future of Clinical Research: Genomics, Proteomics, and Other “Omics” Meet Medicine

Lucio Miele, M.D., Ph.D.

Dr. Miele is Associate Professor of Pharmacology and Molecular Genetics at the College of Pharmacy and at the Cancer Center at the University of Illinois, Chicago (UIC). He is also the leader of the thematic program on cell signaling at the UIC Cancer Center and a member of the National Cancer Institute Experimental Therapeutics Committee One, Study Section. Dr. Miele is Assistant Editor of Women's Oncology Reviews and Journal of Cellular Biochemistry.

A set of technologies is revolutionizing basic and clinical science. They can be referred to collectively as “system biology approaches” because they look at biological variables in a systematic fashion and because they look at the “big picture.” These approaches analyze multiple variables at the same time, generally on a large scale and using high-tech methods, and they seek to identify patterns that are clinically informative and that can be used for the diagnosis, risk assessment, and prognosis of human disease and for the stratification of patient populations in future trials.

We keep hearing about these “omics”—genomics, proteomics, interactomics, lipomics, glycomics—but what are these “omics”? Genomics looks at viability in genes and gene expression. Proteomics looks at biological viability in proteins. Interactomics looks at how different genes and proteins interact with one another and is essentially a bioinformatic discipline. Lipomics is a term that I heard for the first time about a month ago; it refers to the analysis of complex lipid mixtures in clinical samples. Someone has come up with the term “glycomics,” which looks at complex mixtures of sugars in clinical samples.

These techniques provide surrogate efficacy end points that can shorten or simplify patient followup.

What do these technologies have in common? Because they look at many variables at the same time, they generate very large data sets, which means that the most important hurdle to getting clinically informative data becomes bioinformatics—the combination of statistical and information technology tools that allows you to mine clinically useful information from these large

data sets. However, you cannot merely take a clinical sample, put it into some wonderful machine or give it to a team of scientists, and out comes a clinically informative answer. The questions posed will determine whether the resulting answers are clinically informative. Clinical researchers must understand how these techniques work to be able to use them efficiently in their research.

Why does this matter to subject recruitment and retention? At present, these techniques provide surrogate efficacy end points that can shorten or simplify patient followup. For example, consider an ovarian cancer trial for a new drug. If the end point is overall mortality, there is a waiting period of a few years before the end of the trial, which means retention problems and implies that the trial becomes more expensive. To shorten the trial, a genetic or protein biological pattern could be used that predicts outcome, once it is appropriately validated, instead of mortality outcome.

These techniques can also be used, once appropriately validated, as primary or secondary end points for clinical trials to look at diagnostic patterns, prognostic patterns, or risk of disease development progression patterns. An example of this is the recent paper from the St. Jude's Group. They looked at childhood lymphoblastic leukemia, and from gene expression pattern were able to predict which children would develop a secondary leukemia as a result of treatment. This prediction is an outstanding advance that would not have been possible using conventional laboratory techniques.

What will we be able to do in the future, once we know more about these gene expression patterns? We will stratify the recruitment of patients based not on variables such as ethnicity or eye color but on genetic or protein patterns, and we will be able to enter patients into trials who will offer more informative results.

Why does this matter to women's health research? System biology approaches will allow identification of gender-specific gene expression or protein, lipid, or sugar patterns that will predict response to treatment, adverse events, delayed complications, or gender-specific complications—for example, endometrial cancer in women who take selective estrogen receptor modulators.

How do patients differ from one another in ways that can be analyzed by these new techniques? It is far more than just looking at gene sequence or race. Patients can differ in the number of copies of a given gene, or many genes, or in the sequence of specific genes within a specified coding region or within regulatory regions of deoxyribonucleic acid (DNA). Patients differ in gene expression, for levels of messenger ribonucleic acid (mRNA) expression and for the kinetics of mRNA metabolism. From the same gene, we can often make different mRNAs—and, therefore, different proteins—by splicing variants, which is now a growing field of interest in genomics. There are also epigenetic changes in which the sequence of DNA is not changed, but the DNA or the proteins that control accessibility of genomic DNA are modified in ways that change the gene expression pattern.

What about proteomics? The primary variable here is the relative abundance of individual proteins or groups of proteins in tissue samples or in biological fluids, the important message being that these are not always correlated with the levels of expression of RNA. If you only do a gene expression study, unless you also look at the protein, you may or may not get data that predict how much of the protein that gene makes will be expressed. Then there is an entire list of biochemical modifications of proteins that affect their function, and therefore are potentially important from a biological standpoint, that can be studied by these techniques. Phosphorylation is an example—a protein-phosphorylating enzyme is the primary target of the now famous drug Gleevec. If you know more about how proteins are modified, you will be able to identify more molecular targets for drugs like Gleevec.

How do you get this information? There are many different technologies, not all of which produce the same answers, so researchers must beware of what they use

and what questions they ask. You can get information on DNA sequence and copy number by several methods—from direct sequencing to array-based comparative genomic hybridization.

The study of epigenetic changes is a field in its infancy, but these experiments have to be done on relevant tissues, as do mRNA expression studies and for RNA splicing variants—the answer depends on the tissue used. Most proteomics methods are based on a technique called mass spectrometry, which, when applied to proteins, can provide information on the molecular mass of the protein and its sequence, if the protein is not too large. Mass spectrometry-based methods can be used to look at the amount and modification of proteins. The take-home message is that, because you cannot amplify proteins the same way you can amplify DNA, the clinical sample needed for a proteomics study will be larger—requiring more tissue or more blood.

The quality of the resulting data is affected by a number of biological variables. In working with human tissue, you must ensure that all tissue is stored within the same time period after harvesting the tissue and that the tissue is stored properly—even a brief delay in proper storage affects mRNA levels—and ensure that the sample will provide the answer sought. Such issues are also key for samples from primary cells isolated from blood and from bone marrow. The take-home message is that what you do with the sample matters, as in any other laboratory study, and matters even more in tissue or cells for genomics and proteomics tests.

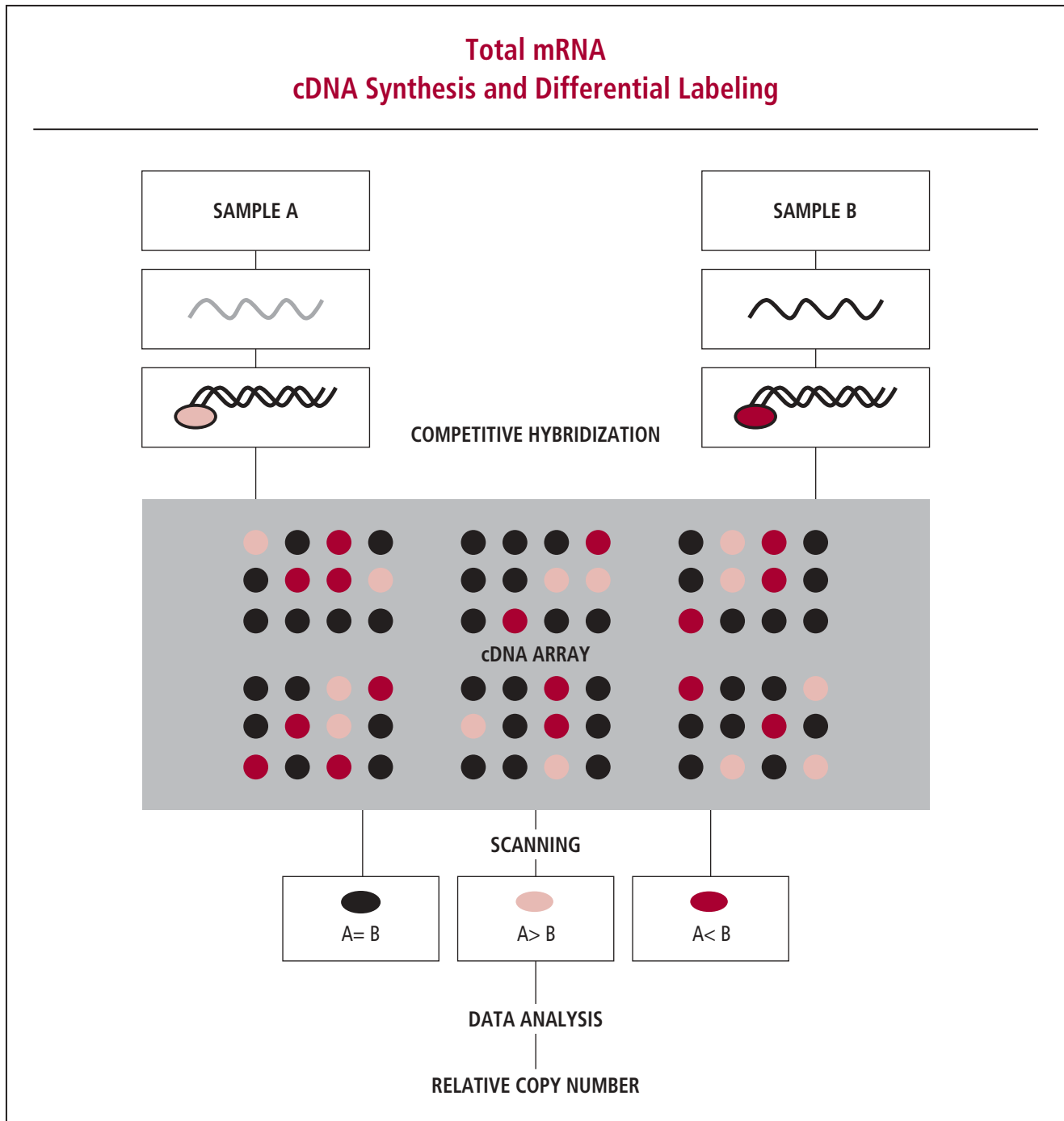
Connective tissue in a tumor should only be analyzed when asking a research question about connective tissue. Researchers used to grind up all of this biological material and look for an answer. What you really have to do, to get a genomics answer that means anything, is to get these sites out of the sample [areas with only cancer cells] and compare them with these other sites [areas with connective stroma]. You can do that by using laser capture microscopy, using a laser light under a microscope to excise only the desired cells from the tissue, separate them, and analyze their gene expression pattern in the informative sites. This technique is important, particularly for mRNA expression studies.

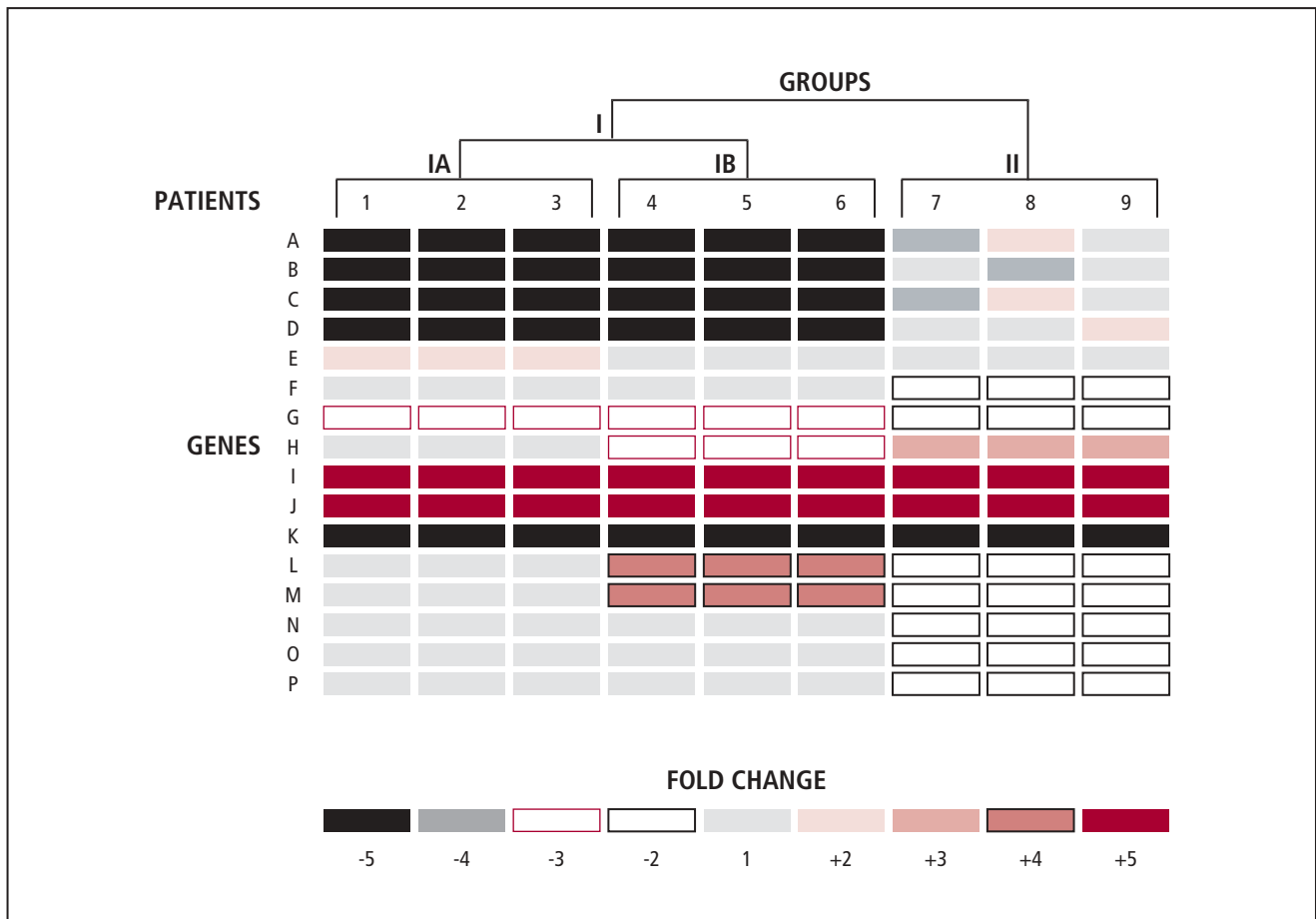
What you do with the sample matters, as in any other laboratory study, and matters even more in tissue or cells for genomics and proteomics tests.

Here are two vignettes about how these studies are conducted. This is a typical mRNA expression genomics study.

To compare two samples (before treatment with after treatment), start by isolating total RNA or mRNA. Then generate complementary DNA copies from this RNA and label them with two different fluorescent markers, for example, a red fluorescent marker for sample A and a green fluorescent marker for sample B. Then mix these together in the same tube and hybridize them at the

same time with a slide that has thousands of dots arrayed generally in square blocks. Each dot contains a target sequence that corresponds to a gene. Genes that are more abundant in sample A will produce redder dots, genes that are more abundant in sample B will produce greener dots, and genes that are equally abundant will produce yellow dots. Then scan and interpret the data, which is a complex step, to derive a relative copy number of mRNA molecules in these samples.





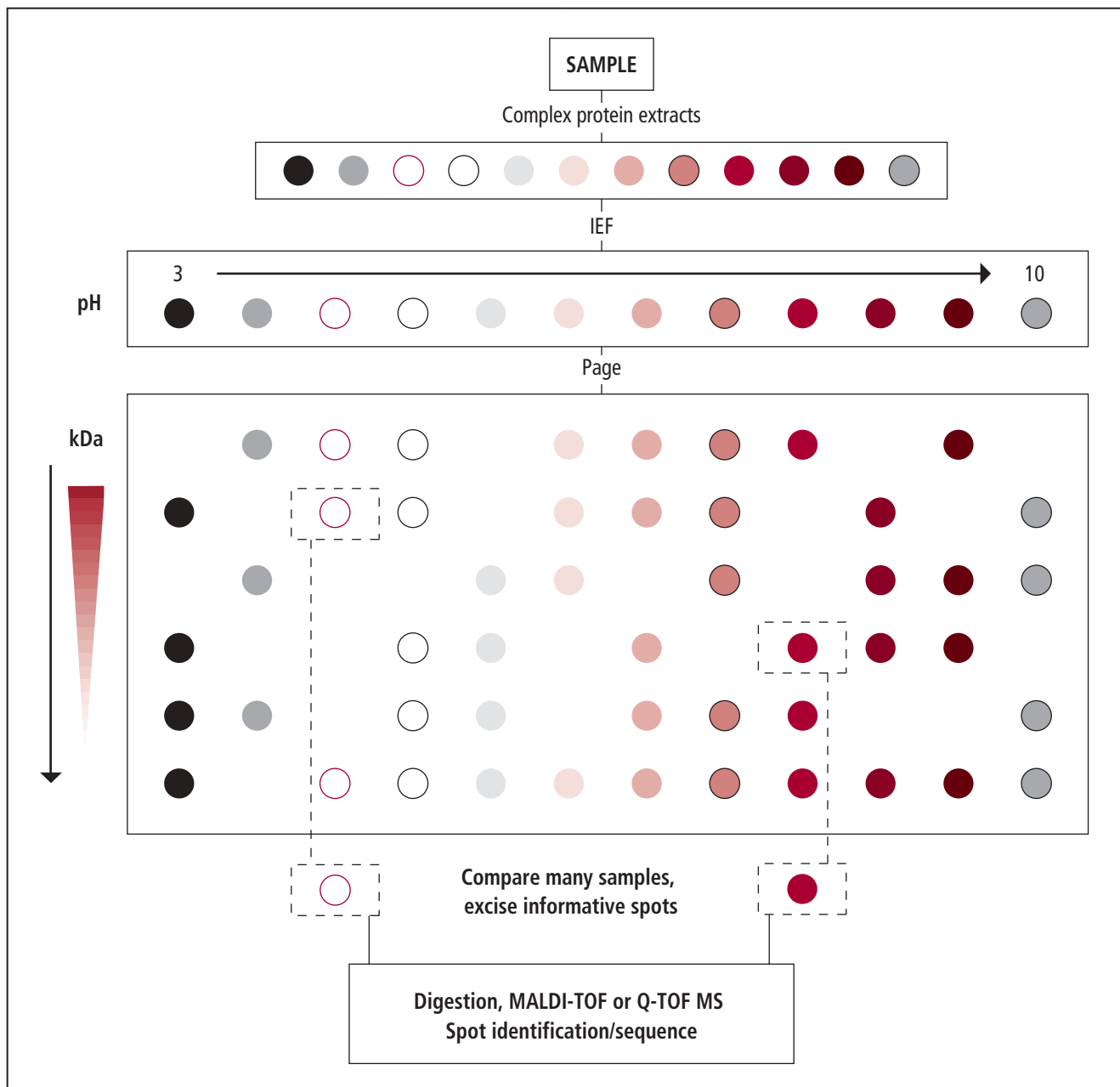
Above is the typical method of representing this data. This depiction is simple, but one can see that patients 1 through 6 are more similar to each other than they are to patients 7 through 9 and, patients 1 through 3 are more similar to each other than to patients 4 through 6.

Hierarchical clustering can be used to analyze the data. This diagram is called a heat map, which indicates degrees of gene underexpression or overexpression (depending on the color). Patients or samples can be classified based on the colors by a large variety of fairly complex cluster analyses.

Below is a typical proteomics experiment. The first step is to extract a complex mixture of proteins from a clinical sample. In the most common application, the sample is separated into two dimensions. The first separation is based on pH—separating the proteins into groups of similar isoelectric points—and then each group is further separated based on molecular mass. Patterns of spots are produced that allow comparison across many samples and indicate which spots differ among the many samples being studied. These spots then could be excised out of the gel and analyzed by mass spectrometry, which

tells which proteins they are, whether or not they are modified, and why they are important.

The right-hand page shows a hypothesis generation machine. Starting with a clinical sample, the first stage is sampling and phenotyping; in some cases, the cells will have to be cultured but in other cases that step will not be necessary. Once the right sample is obtained, the molecules of interest can be extracted—RNA, DNA, genomic DNA, proteins, lipids, etc.—and then the results of the analysis can be analyzed and validated before interpretation. The racial identification is validated using

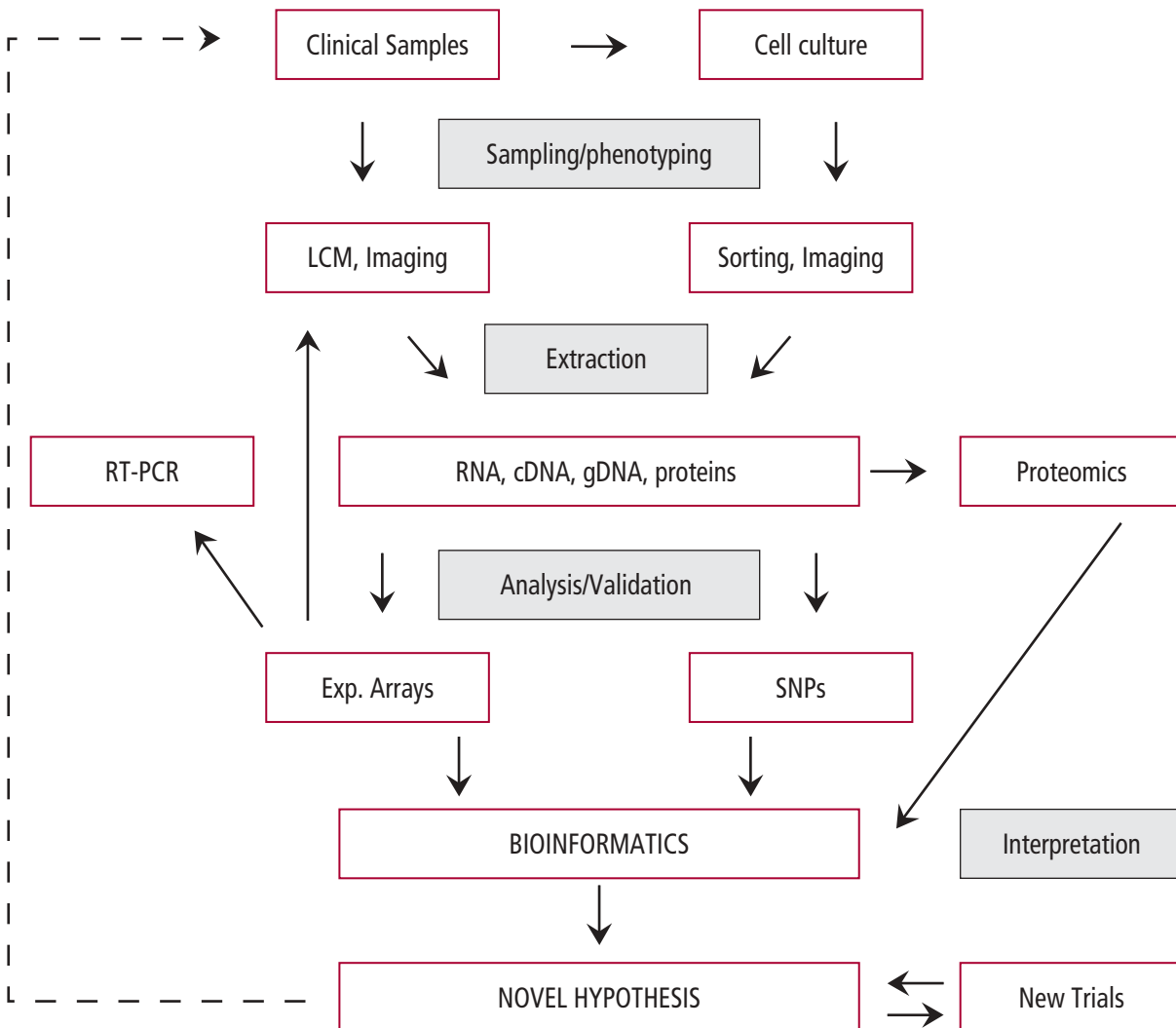


real-time polymerase chain reaction and the needs are validated by resequencing additional PCR reactions; immunostaining is used to obtain the proteomics. After obtaining a validated set of results, interpretation is the next and key step. Make a bioinformatics expert your best friend if you intend to go into this area of research!

This analysis provides a new clinical hypothesis. This is not a hypothesis-free method of conducting scientific research; it is a hypothesis generation machine indicating that pattern X is important in disease Y, which means that new trials must be conducted to analyze prospectively

the hypothesis that the study generated, which provides more clinical samples, and then the cycle starts over again. This point is extremely important. Simply collecting data in a more or less random fashion and analyzing them to reach clinically meaningful conclusions is very likely to lead to wrong conclusions. As in any well-designed clinical study, once a hypothesis is generated, it is imperative that it be tested prospectively in a new trial.

The Flow of Information from Clinic to Molecular Lab and Back



From Pharmacokinetics and Dynamics to Pharmacogenomics: Relevance to Women in Clinical Research

Leslie Z. Benet, Ph.D.

Dr. Benet is Professor and former chair of the Department of Bio-Pharmaceutical Sciences at the University of California, San Francisco (UCSF). He also chairs the boards of two companies, AZMACS, Inc., and OxoN Medica, and is a member of the Institute of Medicine. Dr. Benet is a founding member and the first president of the American Association of Pharmaceutical Sciences.

The reason I am here is that, when the Women's Health Initiative started, the ORWH wanted speakers in all areas. They were looking for somebody in drug metabolism. After looking through the literature and seeing that I had carried out studies and had investigated women's health issues related to drug metabolism, they said, "Let's invite her—Leslie Benet." So they called me and I consented.

Although it has been known for many years that genetic polymorphisms of metabolizing enzymes exist and can lead to marked differences in pharmacokinetics and pharmacodynamics, the importance of drug transporters and potential genetic differences only recently has been recognized. At the UCSF we have a large National Institutes of Health (NIH)-funded grant related to transporters. One transporter, called MDR1, is an efflux transporter—it pumps drugs out of cells. It got its name from cancer chemotherapy: Resistance develops in cancer chemotherapy because this transporter is upregulated and pumps the cancer drugs out of the cells, thus causing lack of efficacy of drugs.

This transporter is found in many places in the body—adrenals, kidney, liver, brain, and gut. It is also found in the vagina, and that is the basis of some of my talk today. This transporter has a lot of genetic variance; in our laboratory we look at 6 of the approximately 30 transporters that exist. These genetic variants can produce differences in how people will respond, because the transporter has different activities. Most of the clinical work has been done with transporter 26, which is a C-to-T variant at 3435.

The meeting materials included an interesting paper from *The Lancet*, January 2002, on response of anti-retroviral treatment of HIV-infected individuals with allelic variants of the MDR1 transporter. The protease

inhibitors nelfinavir and efavirenz are nonnucleoside reverse transcriptase inhibitors. Lower levels of nelfinavir and efavirenz both occur in the TT homozygous variant; however, protease inhibitor efficacy is higher in the TT variant population.

The importance of drug transporters and potential genetic differences only recently has been recognized.

When people talk about pharmacogenomics, they talk about the variant in terms of its activity, and the TT variant is more active than the wild-type transporter, but there is more of the wild-type transporter than the variant. It is a quantitative versus qualitative issue: The TT variant is more active, but there is less of it in that patient population. So the genetic relationship is not simple, and both the variant and the phenotype must be reviewed—how the patients respond. Because a lot of response variation exists, this will be a confounding factor.

From a grant with my colleague Deanna L. Kroetz, Ph.D., at UCSF, we looked at the ethnic breakdown of the frequency of the MDR1 at exon 26, the TT variant, which is the more active variant in African Americans, is very low, about 5 percent, whereas in Caucasians, it is 33 percent. On the other hand, the CC variant (the wild type) in African Americans is very high, but in Caucasians it is relatively low and shows a lot of heterozygotes. So there is an ethnic difference in how patients respond when this transporter is important for drug disposition and how the body handles the drug.

Some of these data are familiar. African Americans have poor clinical outcomes in the general transplant population due to bioavailability, which is significantly

lower for the immunosuppressive cyclosporin; African American patients require higher cyclosporin dosage to achieve drug levels similar to those in Caucasian patients. Two years later, we carried out a similar study with tacrolimus, FK506, and found that African Americans exhibit significantly lower peak concentrations and bioavailability, just like with cyclosporin. When we gave the drug intravenously, we saw no difference, because that transporter is in the gut and this is all a gut effect; all of this is much more important than we thought.

When we look at this relationship, African Americans exhibit poor bioavailability and levels of cyclosporin and tacrolimus and often poorer outcomes than Caucasian transplant patients. African Americans exhibit a significantly lower prevalence of the TT variant of this transporter than Caucasians, so there is a reasonable probability that the metabolism outcome differences found in African Americans are related to a genetic polymorphism of P-glycoprotein. This is the hypothesis of our laboratory, but I cannot yet say that this is definitely the answer.

All of those drugs that I study—the HIV protease inhibitors, the immunosuppressants, the anticancer agents—are all substrates for this transporter, but they are also substrates for the major enzyme for metabolizing drugs in humans, cytochrome P4503A4, and they are all the overlap of P-glycoprotein and 3A4 substrates. We wanted to develop a cell system that would allow us to look at this interaction, and our paper was published a year and a half ago regarding the development of the cell system.

The paper we published earlier this year in the *Journal of Pharmacology and Experimental Therapeutics* described how we looked at the interaction. In a cell that contains P-glycoprotein and 3A, the P-glycoprotein is on the apical side, not on the basolateral side. On the apical side, it pumps the drug molecule out, so when the drug comes into the intestine, that transporter pumps it back out. In the brain, the transporter pumps the drug so that it does not get into the brain. In the liver, the transporter pumps it into the bile. We looked at the extraction ratio—how much metabolism you would have if you came from the apical side or the basolateral side for a drug that was a substrate for both—and there is a big difference. From the apical side, a lot of the drug gets metabolized; from the basolateral side, only a small amount of the drug is metabolized.

We then inhibited the transporter and showed that the metabolism is exactly the same if we came from either side, if only the enzyme was active. We also showed that this transporter was affecting metabolism. When we come from the apical side, we hit the transporter first and then the enzyme; however, when we come from the basolateral side, we hit the enzyme first and then the transporter. The implication is that this is exactly the model for the intestine and the liver: In the intestine you hit the transporter first and then the enzyme, and in the liver you hit the enzyme first and then the transporter.

There are many sex differences in drug metabolism. If we concentrate on cytochrome P4503A4, more clinical evidence emerges that CYP3A4 shows the sex effects (men versus women). In vivo, it appears that the ability to eliminate drugs in premenopausal women is greater than it is in men, but in liver bank samples of enzymes from men and women, no difference is evident. The enzyme looks the same in men and women when the enzyme is examined, but there is a definite difference in vivo for drugs that are substrates for 3A and P-glycoprotein. The answer is that P-glycoprotein is different in the liver in men than it is in women; women have lower levels of P-glycoprotein than do men.

We hypothesize that, if the liver shows lower levels of P-glycoprotein, it is not pumped out—the drug remains, and the enzymes can metabolize it. We hypothesize that the difference between men and women and how they metabolize CYP3A4 substrates is not due to an enzyme difference, even though that is the end point; rather, it is due to a transporter difference affecting the access of the drug to the enzyme. In our paper in *Journal of Clinical Pharmacology and Therapeutics* from November 2002 (included in the meeting materials), we show that this is probably what is happening. For drugs that are substrates for the enzyme and the transporter, women eliminate those drugs more quickly; for drugs that are substrates only for the enzyme, there is no difference between men and women; and for drugs that are substrates only for the transporter, men eliminate those drugs faster than women.

I received funding 2 years ago to test my hypothesis that the vagina is a sanctuary for HIV. During the midluteal phase when progesterone levels are high, P-glycoprotein becomes high in the endometrial tissue. My hypothesis is that the P-glycoprotein is pumping out the protease

P-glycoprotein is different in the liver in men than it is in women; women have lower levels of P-glycoprotein than do men.

inhibitors during that half of the menstrual cycle, and therefore HIV cannot recognize the drug. We are testing this hypothesis in 12 HIV-positive women and 12 HIV-negative women, and we are looking at Caucasians versus African Americans because of that genetic difference in the transporter. It is a complicated study: Women come in during the follicular phase, and we take a vaginal endometrial biopsy and a gut biopsy; then we give them a protease inhibitor, study its pharmacokinetics and look at their lymphocytes. These women then return during their midluteal phase for more biopsies.

Our recruitment has been successful, which we think is amazing. We have data completed on five women who have gone through the study, and we are in the midst of completing the study on others. Early data confirm our

hypothesis about the induction of MDR1 in the luteal phase: We see twofold to twelvefold increases in P-glycoprotein during the midluteal phase versus the follicular phase in these women. We also see some intestinal differences, but not as much difference as we see in the endometrial tissue. We do the snip, and the MDR variants in the African American participants are almost always the wild-type CC, and the Caucasian participants are almost always the TT, as expected. We see a correlation between the pharmacokinetics of nelfinavir and the intestinal P-glycoprotein, and then we measure phenotype to evaluate transport.

Genetics, gender, and drugs are what I came to talk to you about. I want to thank my collaborators—the graduate students and fellows in my lab, where I hope to be for a while so you can keep inviting me back.

Early data confirm our hypothesis about the induction of MDR1 in the luteal phase: We see twofold to twelvefold increases in P-glycoprotein during the midluteal phase versus the follicular phase.

Gender, Race, and Regulation: An FDA Perspective

Susan F. Wood, Ph.D.

Dr. Wood has been Director of the U.S. Food and Drug Administration's (FDA) Office on Women's Health since November 2000. She provides leadership to the FDA about science-driven policy and direction and support of regulatory, scientific, and public health issues that affect women in this country. Previously, Dr. Wood was at the U.S. Department of Health and Human Services and, prior to that, worked for the Congressional Caucus for Women's Issues.

Today I will provide you with an FDA perspective on the complex regulatory issues related to subpopulation data from proteomic, pharmacogenomic, and traditional clinical study data collected on medical products. I will provide a brief introduction to the FDA's mission and role, in particular as they pertain to drugs. I will then discuss the importance of collecting subpopulation data, primarily by discussing what is known about racial and ethnic differences in response to drugs. Finally, I will discuss past and present FDA actions to encourage the collection of subgroup data through policies related to participation and data analysis.

The FDA's mission is to ensure that regulated products are safe and efficacious; honestly, accurately, and informatively represented; and in compliance with laws and regulations. For those who are unfamiliar with medical product development, there is a typical process for product development. In the preclinical phase, researchers analyze a drug's main physical and chemical properties and study its effects in animals. If the results of these studies are promising, the sponsor begins testing the drug in humans in Phase I safety studies. If there are no significant safety issues, efficacy is evaluated in Phase II and III studies. After a product is approved, the FDA monitors safety through several postmarket mechanisms, including Phase IV studies such as registries, and passive surveillance, such as MedWatch. A key component in the regulation of medical products is the premarket approval process. Prior to allowing certain products—such as drugs, biologics, and certain devices—on the market, the sponsors must provide information to the FDA that demonstrates that a product is safe and effective for its intended use, otherwise known as the indication. The information gathered from preclinical and clinical studies during the product development process is utilized to

make such a determination. It is important to note that current FDA policy states that, to demonstrate that a product is safe and effective, it should be tested in a population that is representative of those who will be using the product once it is on the market. Therefore, the product should be tested in a population that has gender, racial, and ethnic diversity.

Current FDA policy states that, to demonstrate that a product is safe and effective, it should be tested in a population that is representative of those who will be using the product once it is on the market.

Why would the FDA require a diverse population? Certainly there are a variety of ethical and scientific reasons. The short answer is that we want to know whether the effects of drugs are different in the various subpopulations that would be using them—such as differences in adverse events and efficacy/effectiveness. A variety of factors contribute to variations in response, including environmental, cultural or psychosocial, and biological factors. To illustrate the biological variability, I will focus on the biological influences, in particular the influence of genetics on drug metabolism.

What is known about racial and ethnic differences in response to drugs? I will provide examples of differences in pharmacokinetic, pharmacodynamic, and clinical responses in ethnic subpopulations in the context of pharmacogenetic variations. Finally, I will discuss what is known about the participation of different racial/ethnic groups in FDA-regulated trials and how that information affects the product label.

As previously stated, it is critical that we have adequate representation of people of different genders, races, and ethnic groups because we know that there can be genetic differences in different groups' responses to drugs. The source of genetic difference can be found in disease pathway, drug targets, and drug metabolism. For today's discussion, I will focus on metabolism. One of the first cases of a phenotype demonstrating a difference in response to drugs was observed in World War II service men. When given antimalarial drugs, men of African American, Mediterranean, and Asian descent experienced a rare but serious hemolytic anemia more frequently than Caucasian men. This was due to a sex-linked recessive phenotype known as glucose-6-phosphate dehydrogenase (G6PD) deficiency. There are over 400 genetic variants, and it is estimated to affect 400 million people worldwide. The prevalence of the variants is different among different ethnic groups. For example, Africans frequently express the African-type variant (G6PDA) and G6PDA(-). The G6PDA allows for normal red blood cell activity, while the G6PDA(-) provides only 10 percent of the activity and is unstable in vivo. Risk of the severity of hemolysis is dependent on a variety of factors (e.g., dose, duration of therapy, and environmental factors) so genetics alone does not predict clinical outcome. Currently, more than two dozen drugs—including primaquine, sulfones, sulfonamides, nitrofurans, and vitamin K analogs—are known to cause hemolytic anemia in G6PD-deficient patients.

A trait discovered in the 1950s was the hereditary variation in a conjugating enzyme N-acetyl transferase found in the liver and certain other tissues. There are more than 20 reported alleles with frequencies that are not uniformly or randomly distributed across different populations. There are two phenotypes: the rapid acetylator and the slow acetylator phenotypes. Determining phenotypic status has acknowledged prognostic value, as the slow phenotype generally demonstrates a toxicity from certain drugs, and the rapid phenotype may not respond to therapy.

A large number of drugs are metabolized by the CYP3A family. CYP3A4 is responsible for metabolizing approximately 50 percent of oxidatively metabolized drugs. The distribution of clearance for these drugs is unimodal, which suggests that genetic variance is not responsible for the observed variation. CYP3A4 is found in the liver and intestine of most Caucasians and roughly 50 percent of African Americans. It is believed that the remaining 50 percent of African Americans predominate CYP3A5. The importance of this difference is illustrated by Midazolam, which clears 30 percent more slowly in African Americans with CYP3A5.

There can be genetic differences in different groups' responses to drugs. The source of genetic difference can be found in disease pathway, drug targets, and drug metabolism.

Frequently used drugs, such as dextromethorphan, beta-blockers, antiarrhythmics, antidepressants, antipsychotics, and morphine derivatives are metabolized by the CYP2D6 enzymes. Unlike CYP3A4, it is a noninducible enzyme so the genotype is predictive of metabolism to a large extent. Poor metabolizers have two nonfunctional alleles and are likely to have adverse events. Ultrarapid metabolizers have as many as 13 copies of alleles and often experience therapeutic failure. An example of this phenomenon is the observation that African Americans and Asian Americans have differing responses to beta-blockers.

Drugs such as S-mephenytoin, omeprazole, diazepam and propranolol are metabolized by CYP2D19. A poor metabolizer is characterized by two nonfunctional alleles. The frequency of poor metabolizers varies among racial and ethnic subgroups, being most frequent in Asian people. Few drugs are metabolized through this enzyme, but there are pronounced pharmacodynamic effects. For example, in the case of diazepam, the poor metabolizer has a higher risk for toxicity. Therefore, it has been concluded that caution should be used when giving diazepam to Asian populations.

Finally, there are also different frequencies of poor metabolizer phenotypes of CYP2C9 as well, which metabolizes roughly 20 percent of hepatically cleared drugs such as phenytoin, S-warfarin, and nonsteroidal anti-inflammatory drugs. The poor metabolizer frequency is more frequent in Caucasian populations and may be important for drugs with therapeutic ranges, such as warfarin.

We have discussed the genetic variations and examples where subgroup differences have been seen, so now we will discuss the state of racial and ethnic subgroup participation in trials and analysis of the data collected in those trials. Although data about subgroup differences are increasing, there is a need for greater participation of underrepresented racial and ethnic subgroups, particularly in the early phases when dose-response is determined, so that more information can be obtained. The FDA has completed two studies to determine subgroup participation in clinical trials that were conducted in support of marketing applications.

The Office of Clinical Pharmacology and Biopharmaceutics in the FDA's Center for Drug Evaluation and

Research conducted a study to determine the representation of African Americans in early-phase clinical studies. A random subset of New Drug Applications (NDAs) submitted from 1996 to 2000 was reviewed for inclusion of data on race, particularly the participation of African Americans. In addition, the Physician's Desk Reference (PDR) was searched to determine how many entries included race as a covariate in population pharmacokinetic (PK) studies and how many reported that PK was influenced by race. The results indicate that 8.4 percent of the subjects in early-phase studies in 28 NDAs were African American. The percent of African Americans in PK studies was greater than in dose-response studies, and the percentage of African Americans participating in tolerance studies was least.

A key word search was done on 2,250 entries in the PDR. Only 37 entries reported population PK study results. Ten entries reported analyzing race as a covariate. No entry reported race as an important factor influencing PK. Finally, eight entries reported race-based PK differences with an observed trend that African Americans were having a higher exposure level. One entry recommended a race-based dosage adjustment.

In 2001 the FDA Office of Special Health Issues (OSHI) published the results of a study that looked at drug applications of new molecular entities that were approved between 1995 and 1999. Enrollment data were obtained and tabulated according to race/ethnicity, and the approved product label was searched for statements related to product testing in racial and ethnic subgroups. The results indicated that racial and ethnic groups participated in trials to varying degrees. Race or ethnicity could be determined for only 53 percent of the participants. Of that group, African Americans participated the most. However, participation declined from 12 percent in 1995 to 6 percent in 1999. In trials conducted only in the United States, the participation of African Americans is comparable to their representation in the general population. One percent of participants were Asian, Pacific Islander, or Native Hawaiian. Three percent were Hispanic/Latino, and less than 1 percent were American Indian or Alaska Native. Some differences in participation of all subgroups are seen when comparisons are made from year to year and by product class.

Labeling from 45 percent of the products contained some statement about race, including statements that no studies were conducted, too few patients were studied, or no differences were seen. Thirty percent of the statements indicated that no differences were found. Eight percent of the labels described differences related to race. Roughly half of those described PK effects, 39 percent were efficacy, and 11 percent were safety. One product label recommended a change in dosage based on racial

differences. The majority of these differences were noted in cardiorenal, analgesic, neuropharmacological, and metabolic/endocrine products.

In reviewing past and present actions, the FDA has published a series of guidance documents and regulations. It is important to note that the guidance and regulation discuss issues of inclusion and analysis, but there are no specific requirements regarding absolute numbers or methods of analysis as they relate to subgroup analysis.

With the current regulations and guidance in place, the FDA believes it has taken the necessary first steps to encourage the participation of all demographic subgroups in clinical trials and promote the analysis of the data. However, this may not be enough. The FDA continues to evaluate current data to determine future directions through initiatives such as its activities with the U.S. Department of Health and Human Services (DHHS) and its sister agencies to standardize the definitions of racial and ethnic subgroups. In addition, the FDA has various initiatives to track enrollment trends, such as the development of a demographic database to capture data on subgroup participation in clinical trials and data resulting from subgroup analysis.

Regarding the concept of standardizing subgroup definitions, the DHHS has formulated a standard method of analyzing and presenting data with regard to race and ethnicity based on the U.S. Census, on Healthy People 2010, and on Centers for Disease Control and Prevention

2001 OSHI Study: Participation

Race and/or Ethnicity determined for 53% of participants

Of those with determined subgroup:

- Caucasian, 88%
- Black, 8%
- Asian, Pacific Islander, Hawaiian, 1%
- Hispanic/Latino, 3%
- American Indian or Alaska Native, <1%

Of those in US only trials:

- Caucasian, 83%
- Black, 13%
- Asian, Pacific Islander, Hawaiian, 1%
- Hispanic/Latino, 3%
- American Indian or Alaska Native, <1%

2001 OSHI Study: Labeling

- 45% of product labels contained some type of statement related to race
- 30% of the statements indicated that no differences were found
- 8% described differences related to race
 - 50% pharmacokinetic
 - 39% efficacy
 - 11% safety
- The majority of race related labeling differences occurred in cardio-renal p, analgesics, neuropharmacology, and metabolic/endocrine products

definitions. However, at this point, the FDA does not have guidance that specifies data reporting by race and ethnicity. Although the FDA calls for those data, not having guidance leads to the data being presented in inconsistent formats, which leads to analysis difficulties. [A draft guidance was

issued after the workshop and is currently being finalized.]

The FDA has been challenged to develop a demographic database, but such a database must be filled with relevant data. When a standardized form is not linked to the demographic variables or when the data that come from pharmacogenetic or proteomic data sets and traditional clinical study data, this represents a real challenge. Inclusion in clinical trials and having the information linked to individuals and available for analysis are critical issues.

We will not succeed in obtaining better reviews or management systems or improving our knowledge base unless we have access to usable data. That is the challenge we face at the FDA.

We will not succeed in obtaining better reviews or management systems or improving our knowledge base unless we have access to usable data. That is the challenge we face at the FDA. To meet that challenge will take the concerted effort of the academic community and the research community as well as the regulatory community.

Chapter Eight

Recruitment, Retention, and Relevance: Continuing Challenges

Moderator: Susan E. Cohn, M.D., M.P.H.

This session highlighted special populations that are underrepresented in current clinical trials and presented policy options and guideline considerations for the National Institutes of Health (NIH) and other agencies to encourage and facilitate diversity among research participants. Each speaker also addressed the importance of balancing the need for enhanced enrollment of special populations with the need to protect these participants from potential harm.

Dr. Cargill addressed the issue of recruitment and retention of ethnic minority women in the “real world.” She urged conference attendees to make clinical trials relevant to women in inner-city settings by assisting research participants with logistical needs such as childcare, reimbursement for transportation, compensation for time spent during study visits, and flexible clinic hours. She also urged researchers to develop collaborations and partnerships with community leaders and local service agencies, provide participants with necessary assistance, and improve investigators’ understanding of these women’s lives and community impressions about research participation.

Dr. Bowman addressed special challenges to participation in clinical trials for lesbian, gay, bisexual, transgendered, and queer populations. She highlighted the importance of accurately assessing sexual orientation and the need for sensitivity and awareness in defining populations and in measuring sexual feelings and behaviors. Patients may fear rejection and retaliation by medical providers, the research community, their families,

and others through disclosure of their sexual orientation; confronting these issues directly will likely result in enhanced recruitment of these vulnerable populations.

Dr. Klein discussed the challenges of recruiting and retaining adolescents in research studies. He reviewed the policy issues surrounding adolescents’ access to health care, participation in research, and concerns about confidentiality. Federal guidelines concerning informed consent generally assume that children and adolescents are not competent to provide informed consent. As with other vulnerable groups, adolescents may fear disclosure of their behaviors and concerns, in this case to their parents. Federal guidelines exist but are applied inconsistently; clearer guidance from the NIH would be of assistance to researchers and institutional review boards (IRBs) to better involve youth in research.

Dr. Whetten-Goldstein addressed the ongoing challenges of recruiting and retaining rural populations in studies. Compared with urban women, rural women are more likely to live in poverty, to be less educated, and to travel longer distances to participate in clinical trials. Rural women are also more likely to be less healthy than urban women, yet they have fewer medical and support services. Confidentiality is difficult to maintain in rural areas, and many rural populations harbor a distrust of health care systems, deterring their participation in trials. Listening to and understanding the concerns of rural communities will help bridge the cultural divide and engender trust, thereby decreasing the barriers to care and participation in clinical studies.

Recruitment and Retention in the Real World—Thinking Out of the Box

Victoria A. Cargill, M.D., M.S.C.E.

Dr. Cargill is Director of Minority Studies and Clinical Research at the NIH Office of AIDS Research and is former Professor of Medicine at Case Western Reserve University School of Medicine.

A clinical trial has a powerful research design, and it can help answer important treatment questions. A major problem is recruiting marginalized or hard-to-reach populations to participate in clinical trials. These trials often fail to meet the needs of marginalized populations, leaving significant gaps in the scientific knowledge base. A clinical trial can be of too little relevance for those of us who take care of those hard-to-reach populations, and from the perspective of the patients, it is sometimes the providers that represent the hard-to-reach population. The trial may not address questions that community providers face, such as how street drugs interact with human immunodeficiency virus (HIV) treatment drugs. Trials are often done to the community—what I call “drive-by research.” Stereotyping may lead us to exclude populations. Several articles in the *Journal of the National Medical Association* have addressed this issue. In some settings, African American providers are as guilty as Caucasian providers of stereotyping their African American patients.

The challenge is to make clinical trials relevant. Consider the following scenario, which is a real case. A clinical center has a clinical trial protocol open for women with HIV infection and their children. Mildred gets her care at the center, and she is homeless. She shows up, occasionally intoxicated, but she does keep most of her appointments; she has heard about the study from discussion in the waiting room. Mildred is fairly

well informed about HIV infection. Although she has no consistent address, she has extensive community networks. She wants to know more about the study, admits that her motivation is in part to get better care for her child who is also seropositive, and agrees to talk to a study screening nurse.

Then she learns that a pharmacokinetic (PK) study will be part of this trial and that it requires multiple blood draws, requiring her to stay in a unit all day. She also finds out that the drug has some side effects that she would be responsible for financially, for example, if she were to need a blood transfusion. The protocol visits at the clinic are currently scheduled for Monday through Friday, 9:00 a.m. to 5:00 p.m. She needs to provide a telephone number where she can be contacted in an emergency.

Should the study enroll Mildred or not? The choices would be to decline to enroll Mildred because it requires a lot from her—and she has no resources—or to enroll Mildred ensuring that she understands in no uncertain terms what she is expected to do, when she is expected to do it, when she is supposed to show up, and how often. Another option is to enroll Mildred and attend to what can be called “the extras”—arrangements that are responsive to the participant’s life circumstances. I am troubled by how much we undervalue the time of the people who participate in these studies. Some may have to take three buses in the rain with sick kids; others must worry about leaving their homes and stepping over junkies in the hallways. We should not be arguing about paying somebody \$50 to cover expenses. A staff person would have to work with Mildred, establish a relationship, and extend the hours for protocol visits, except the PK study for which there is less flexibility. Mildred would also need transportation to and from the trial site, a “study buddy,” and lunch and dinner vouchers for the long days.

Clinical trials often fail to meet the needs of marginalized populations, leaving significant gaps in the scientific knowledge base.

In our study, we provided these extras to Mildred because we had to balance her needs with her reality. We tried to reduce the structural barriers. Until we could arrange everything, the white gay man who told her about the study and why she needed to get her child into care drove into the neighborhood every day and picked her up when she had to come for her studies. She was not stereotyped, we respected the fact that she had some needs, we used case management, we worked

closely with community-based agencies, and we tried to balance between the issue of coercion and respect for the time commitment it takes to do something like this.

How did it work out with Mildred? She kept 90 percent of her study visits, she recruited many other women in her community network to our study, she became a study outreach worker, and she was no longer homeless but she was still poor. Mildred had what I call a “Ph.D. of the Street,” without which no one from the street was coming to the study. A “Ph.D. of the Street” has to be honored as such with appropriate compensation.

Thinking outside of the box is equivalent to acting outside of the box: addressing individual barriers, identifying issues, trying to reduce and modify these wherever possible, and including staff members who reflect the patient population at all levels. Mildred and many of her referrals would say to me when they came in that I looked like them, but that was only one day out of five. So many times they would say, “Why does nobody here look like me? Where are they all? How come I am the only brown face in here?”

We reduced some of the structural barriers by providing childcare. We actually changed the protocol and subsequently, through our clinical practice, offered moms’ and children’s clinics. This is not new now, but back then it was. We routinely provided transportation and were able to get small grants to pay for that.

We started doing what we called “study buddies.” We ran town hall meetings at which we explained the

different changes in HIV care, the nature of HIV infection, and how our protocols were designed to learn more about it. We had panel discussions by people who had participated in some of the studies. They described how they

felt about it, the positive aspects and the challenges, so that if people wanted to sign up and be on the study or even wanted to talk to someone to get information from someone other than the provider, there was someone else to whom they could talk. These people committed to

buddying up with new candidates, and it was a very powerful thing. Some of those relationships went on until people died. Study buddies stayed with them right until the end.

Some unique collaborations and partnerships were formed. Mildred had to have a phone number where she could be contacted, which is a challenge for a homeless person. We partnered with a colleague at Cleveland State University who had designed a unique bridge: A person could call in on a toll-free number and access voice mailboxes for people who were difficult to contact—homeless individuals, injection drug users, or people in various different locales. Each person was given a unique password, so he or she could call in and pick up voice mail and, by using that same password, could call into the bridge and leave a question for an HIV physician. Each of us took turns answering those kinds of questions. Those responses were tape-recorded, but the name of the person who left the question was erased so they could not be identified after the question was answered. The answers were eventually cataloged as a sort of HIV library so that people could call in to get information about HIV care. They could just hit the button and listen to topics such as HIV infection transmission, children, antiretrovirals, protease inhibitors, and opportunistic infections. The system became very heavily used.

Special populations are not aliens. They all have the same needs and desires that everyone has, but barriers, challenges, and issues in their lives make things more difficult.

Thinking Outside of the Box = Acting Outside of the Box

Address individual barriers

- Identify issues
- Reduce/modify where possible
- Include staff that reflect the patient population (at all levels)

Reduction of structural barriers

- Childcare
- Transportation
- Study buddy

Unique collaborations and partnerships

- Tech specialist provided voice mail
- Call in computer became teacher

Special populations are NOT aliens. At core they have the same desires and needs.

It is essential to avoid stereotyping the participants or providers, but it is equally important to be prepared to listen to the people because sometimes what you hear is painful, sometimes what you hear is hard to fix, sometimes what you hear means your whole Friday afternoon just went up in flames trying to fix this problem. The greatest demonstration of respect and care for the individual is to listen. We cannot assume that all communities are

monolithic, whether homeless, injection drug user, Caribbean, Latino, Hispanic, African American, and so forth. Allow the community to educate you. If you can make it a two-way educational process, you will be amazed at what you can learn. Over the years of doing this, I have come to believe that study without reflection is a waste of time, but reflection without study is dangerous.

Special populations have the same needs and desires that everyone has, but barriers, challenges, and issues in their lives make things more difficult. Allow the community to educate you.

Lesbian, Gay, Bisexual, Transgendered, and Queer Populations: Challenges in Recruitment, Retention, and Relevance

Sallyann M. Bowman, M.D.

Dr. Bowman is Staff Internist at the Student Health Services at the University of Pennsylvania and the former Director of Clinical Services at the Medical College of Pennsylvania/Hanneman Center for Women's Health. She has also been actively involved for many years with the Women in Medicine, Lesbian Physician's retreat and currently serves as a board member for the organization.

Why should there be concern for and inclusion of lesbian, gay, bisexual, transgendered, and queer (LGBTQ) individuals in clinical research? The data show that between 1 percent and 10 percent of individuals characterize themselves as LGBTQ and between 1 percent and 4 percent of women report a lifetime incidence of lesbian sexual relationships. Since most of my background and awareness is in the area of lesbian health, I will refer to lesbian health issues but will try to make associations with LGBTQ. (The “Q” represents younger gay and lesbian people, who tend to prefer the term “queer” to “lesbian” or “gay.”)

Studies indicate that lesbians may have higher rates of certain risk behaviors. Early studies indicated that lesbians smoke more, drink more heavily, and have higher body mass indices than other women. More recent studies indicate that lesbians have lower rates of adherence to certain preventive health screening measures such as mammography and cervical cancer screening. Since available data are inconsistent and sketchy, additional research regarding preventive health practices among lesbians is needed.

Issues around access to health care become part of the sociopolitical fabric of what it means to be lesbian. Work-based health insurance benefits of the employed domestic partner are not available to an unemployed partner, and the unemployed domestic partner's health care might be compromised because the employed partner cannot take caregiver leave. Literature from the early years has stated that people are uncomfortable “coming out” to their health care providers for fear of rejection and mistreatment. How does that change the assessment of risks? How does it change the assessment of strategies for care if there is no full understanding of the context and fabric of the patient's life? Health

professionals must take into consideration all of the psychosocial factors affecting health care, including sexual orientation.

Strategies for Inclusion of LGBTQ issues into Women's Health Research

- Ask the LGBTQ question
- Ask the research questions – what are the potential ramifications, outcomes, etc. if Lesbian?
- Extend outreach into LGBTQ communities
- Examine existing databases
- Report Results
- Fund the research

The 1999 Institute of Medicine (IOM) report on lesbian health made eight recommendations. One recommendation refers to the validity and reliability for measuring sexual orientation. Articles on LGBTQ status note the different ways of asking the question about sexual orientation and sexual behaviors. Each investigator reports how their study asked the questions and, at the end, apologizes for not being able to generalize the data because the populations may not be comparable due to the way in which the questions were asked. Therefore, the IOM recommended conducting research to develop the tools that will validate and make reliable the questions asked to correctly identify members of sexual minorities and to separate, when appropriate, sexual behavior from sexual identity.

A second challenge is defining “lesbian.” Difficulties with the nomenclature include that some studies are

focusing on sexual behavior to define lesbian, gay, transgendered, transsexual, and queer, while other studies are defining sexual orientation. How an individual self-identifies is influenced by whether or not the individual is “out” to themselves or to their community. A person’s reluctance to disclose her orientation or behavior can be clouded by her personal comfort with her identity: There are sociopolitical ramifications to disclosure as a lesbian, especially in conservative communities.

The Valanis article in the *Archives of Family Medicine* in 2000 defined a population of women who were 50 to 74 years old (from the Women’s Health Initiative) by asking people to self-describe as having no adult sex, sex only with men, sex with men and women partners, sex only with women, and sex only with women after age 45. They stand counted as lesbian those women who had only had sex with women and those who had only had sex with women after age 45. The group in the middle was counted as bisexual. This nomenclature exemplifies the concepts that sexuality, sexual behavior, and orientation are lifespan issues.

Issues around access to health care become part of the sociopolitical fabric of what it means to be lesbian.

Another IOM recommendation is to review existing databases for issues of relevance to LGBTQ populations. Although there is not enough money to address the needs of the many vulnerable populations, several large databases already exist from which the appropriate question could be asked regarding sexual orientation. For example, there are likely to be 920 to 3,680 lesbians among the 92,000 nurses in the Nurses’ Health Study cohort, assuming the statistics regarding prevalence apply. Valuable information regarding lesbian health is lying dormant in those data, so studies to mine those data should be designed and funded.

What are some of the other challenges researchers face in recruiting lesbians into research? First, all lesbians are not alike. The stereotyping of what is a lesbian, who looks like a lesbian, who acts like a lesbian, or who may be a lesbian may keep us from understanding the needs of lesbians and how each lesbian woman reacts and presents herself to the medical community for care. Not everybody is equally averse to participating in clinical trials, but there is no single location that is considered “the lesbian place” to recruit research participants. Recruitment appeals must be across the board. The recruitment materials should speak to lesbian

or gay individuals, indicating that the research study applies to them and to the issues in their communities. Inclusion language will motivate people’s interest and might appeal to their need to get involved on a personal level, which is an important incentive to research participation. Though there is not one single LGBTQ community, there are avenues for getting the word out—through community-based organizations, gay and lesbian publications and community centers (which can be located in any phone book in any town), the Gay and Lesbian Medical Association, the Lesbian Health Fund and the Internet serve as examples.

Another challenge is fear—homophobia. Patients and researchers friendly to LGBTQ health issues fear rejection and retaliation by medical providers, by the research community, by their families, and by others. They also fear disclosure and whether the data gatherers and researchers can be trusted with the information, which is critical information about the person. For example, would a schoolteacher who is a lesbian really want to disclose her orientation in a clinical trial, even though she might want to participate? If yes, it would obviously require a lot of thought before she checks that box, if she fears that the information would be shared with employers or the community. All researchers have the responsibility to assure and guarantee as much as possible the anonymity and sanctity of the information provided by their research participants.

Research protocols not only must ask the question “are you lesbian, gay, transgendered, bisexual, or queer” but also must tailor the question to the potential outcomes. Is it relevant if the study subjects are LGBTQ, and will the study outcomes be different or the same for LGBTQ individuals? We must at least ask that question at the research table and in the design phase and support that kind of inclusion at the institutional level. Certainly IRBs and review committees can address the question of impact for LGBTQ people.

Results must be reported. The Nurses’ Health Study started asking the LGBTQ question in 1973, but not one article has yet been published that reports the results for lesbians in that study population.

Numerous opportunities for data analysis exist. For example, the U.S. Census 2000 asked the LGBTQ question, and Healthy People 2010 asked behavioral questions to tease out sexual orientation. The research priorities of the NIH Office of Research on Women’s Health (ORWH) can be made relevant to lesbians; those priorities include gender differences in treatment choice, compliance and adverse effects in healthy living, and prevention of chronic disorders. In addition, the

The research priorities of the ORWH can be made relevant to lesbians; those priorities include gender differences in treatment choice, compliance and adverse effects in healthy living, and prevention of chronic disorders.

literature on smoking, alcohol, and obesity provides additional research opportunities.

On a more positive note, the information base is expanding. An article published in 2002 in the *American Journal of Public Health* reviewed abstracts of publications searched on MEDLINE that mentioned lesbian, gay, transgendered, and transsexual. The number of abstracts that referred to LGBTQ issues increased significantly

since 1984, with 442 articles relevant to lesbian issues from 1995 to 1999. In contrast, between 1980 and 1984, only 86 articles were published.

The objectives of the ORWH include several areas of interest to lesbian health that could be integrated into overall study design and reporting. If even one study adopts these suggested approaches, then the objectives of this panel will have been met.

Adolescents: Challenges in Recruitment, Relevance, and Research With Youth

Jonathan D. Klein, M.D., M.P.H.

Dr. Klein is Associate Professor of Pediatrics (Adolescent Medicine) and Associate Professor of Community and Preventive Medicine at the University of Rochester Medical Center. He is an expert in adolescent medicine and in child and adolescent health services research.

I will describe briefly the adolescent population, what we know about their health care, how confidentiality relates to them, and what that means for research endeavors. As a context in which to think about the challenges of recruiting youths into research studies, imagine trying to communicate with 15- and 16-year-olds about anything and then imagine asking questions about their personal behaviors. Also, consider the genetic differences, age differences, and ethnic differences among youths and that the evidence base for children and adolescents is almost nonexistent on how they behave or the way they react to therapies.

How this country thinks about teenagers has changed somewhat since the 1950s, which can be observed partly in how the media reflect teenagers and partly in the kind of health issues that are discussed in relation to them. There are about 40 million adolescents today and that portion of the population is not growing as fast as the overall population. The ethnicity mix is changing: The largest group now is 64 percent Caucasian non-Hispanic, but that will drop to 58 percent by 2002 and continue to shrink.

The major causes of death for adolescents and young adults are injury, homicide, suicide, and HIV, in that order, except for young African American men, for whom homicide is the number-one cause of death. Most of these causes are potentially preventable.

What about adolescent behaviors? The Centers for Disease Control and Prevention monitors these as types of risk behaviors on the Youth Risk Behavior Surveillance System, and risk behaviors are also described in Healthy People 2010: alcohol, tobacco, and drug use; depression and suicide; injury and violence; school failure and learning issues; and sexually transmitted diseases (STDs), unsafe sexual behavior, and pregnancy. About

10 percent of adolescent girls get pregnant in most years, a number that is falling only a little bit. Lifestyle issues such as obesity, eating disorders, and cardiovascular disease risk all start during the teenage years. Asking adolescents about these behaviors or studying therapies or counseling interventions that might intervene may seem overwhelming.

How this country thinks about teenagers has changed somewhat since the 1950s, which can be observed partly in how the media reflect teenagers and partly in the kind of health issues that are discussed in relation to them.

If adolescents are seen as problems, restrictive policy approaches are likely, in social settings as well as in health care settings. However, if adolescents are seen as assets and as positive elements of their communities, policy solutions are more likely to be supportive and nurturing. So the other way to approach this is to flip the statistics around: 65 percent of adolescents are physically active, 80 percent are not overweight, 50 percent do not drink alcohol regularly, and 65 percent were not sexually active in the past 3 months. I always hesitate to put sex on the same list with those other risk behaviors because it is a normal expectation to become a sexual person in relationships. However, the way in which many young people engage in sex does put them at risk; only about half of them used barrier contraception the last time they had heterosexual intercourse. Again, the issue is how to study these behaviors and what that means for the way we think about teens.

A broad range of issues is part of the research agenda for child and adolescent health. The biopsychosocial model helps ground the issues for adolescents in terms of the biology of puberty, the psychology of cognitive development, and the social development of relationships, networks, and intimacy. It also grounds the entire research agenda in thinking about the biology, molecular genetics, physiology, and cellular level of development as well as the psychological, cognitive, and neuroendocrine and immunologic brain and cognition development of adolescents and young adults. It encompasses the social issues: individual health behaviors, the relationship of young people to health services, and the family, school, and community influences that address how some of this works.

The evidence base is extremely limited. I have been on the U.S. Preventive Services Task Force for the last couple of years, and we have yet to review a topic for which the evidence base is sufficient to make a strong recommendation for children or youth. Hardly any trials include adolescents, and although we generally perceive adolescents to be healthy in society, about 10 percent of teens have a definable chronic illness, and about half of them have some impairment of activity of daily living from that problem.

When is extrapolation among populations appropriate? There are some examples when it is clearly not appropriate and others where it might be. Regarding counseling for obesity, if someone were physiologically mature, would it seem reasonable to extrapolate from adult evidence? Maybe or maybe not, depending on what lens is used to look at that. From nonhuman animal models, we know that it sometimes might not be appropriate. Rat data suggest that the brains of adolescent rats develop nicotine receptors at higher rates than do adult rat brains, but it is difficult to extrapolate from those studies to humans. On the other hand, it does suggest that the tobacco companies know what they are doing when they try to get 11- to 14-year-olds to initiate smoking before they are cognitively mature enough to know what the long-term effects might be.

There are some challenges and limitations in methods as well. Who should you ask about access to care or about utilization: parents, teens, or both? Who is a valid reporter? If adolescents are alone with their clinicians in a room—and we hope that they are at least for part of the visits—how can parents report on what happened during those visits? Some data from the Commonwealth Fund's survey of the health of adolescent girls (which actually was about 6,000 boys and girls) surveyed in school show that adolescents think they have a usual source of care. This is the first time this question was asked nationally of teens themselves, although this question has generally

been asked of parents through the National Health Interview Survey. Teenagers believe they have access to care because they reported having a health care visit that was not the result of illness; this is true of parent reports, too, although the quality data from health plans, the Health Plan Employer Data and Information Set (HEDIS),

If adolescents are seen as assets and as positive elements of their communities, policy solutions are more likely to be supportive and nurturing.

suggest that the number is only about 50 percent, which raises the validity issue. Only 62 percent of the boys and 53 percent of the girls ever spoke privately, one on one, with their physician or health care provider, and a substantial portion say they “ever missed” needed care and that they were “ever too embarrassed” to discuss a topic. The main reason was that they did not want their parents to find out. This was much more of an issue for young women than for young men, but it was still the highest reason noted by the boys as well.

All the recommendations from national organizations and elsewhere from the Federal Government state that services for adolescents should be confidential. What does this mean? This is not absolute secrecy and privacy; it is conditional confidentiality. A duty to provide some protected services exists in most States, whether the treatment is related to STD, mental health, or abuse, but some things must be disclosed because a duty to warn and mandated reporting also are present in most child and adolescent services.

About half of adolescents confide in their parents about sensitive care issues, but about 15 percent to 20 percent of adolescents reported in some studies have used care without their parents' knowledge. The Youth Risk Behavior Survey, conducted every other year in schools, reveals that one in three teens has had sex and only a little more than half of them used barrier protection. Some data from Reddy (2002) and colleagues suggest that teens say that, although parent involvement is desirable, they would not seek care if they needed parental consent. On the other hand, 99 percent of the teens in this study said they would still keep having sex.

That is what the teens say. Some people say that fire extinguishers encourage young people to play with matches, that airbags encourage head-on collisions, or that condoms encourage people to have sex. But Pandora's box does not exist when it comes to health behavior. If it were so easy to change behavior—by

talking about it once or by just thinking about the issue—we would not have the health problems that we have talked about so far!

What about parents? What about rights? Whose rights and responsibilities do we care about in our country? U.S. public policy tends to regard the sanctity of the family as a central concept and hesitates to interfere with family rights. That policy permeates the way we think about how government acts on families, but some evolution has occurred over the years regarding protection for children and youths. In English common law, children were property; that was also true for women back then. In the early 1900s there was some recognition that children have some limited autonomy from their family's ownership of them, and in the 1960s we began to see some child abuse and due process legislation that addressed this specifically. The current effort to define fetal rights is progress in that area and is a reflection of the current tone.

The American Academy of Pediatrics has a policy statement on confidentiality in adolescent health care, which says that adolescents need to know that health care professionals will provide them with the best possible care and counseling if they choose to seek treatment. It encourages including parents in decisions, but it also recognizes that some laws and regulations are unduly restrictive and in need of revision. State laws differ, and there are many jurisdictions. No consistency exists regarding STDs, HIV, mental health, substance abuse, reproductive health, and abuse—some of these are protected in some places and not in others. It also matters whether the adolescent is married, is in military service, is a parent, is a mature minor, or is an emancipated minor. Some States define an age; others do not. In other cases it is the clinician's judgment.

Research requires a higher standard. The Belmont Principles of personhood, beneficence, and justice, and the requirements for informed consent and the safety of participants require that we think about the level of potential benefits and harms to any particular research activity. Adolescents' capacity to consent is an empirical question. Adults are assumed to be competent, but adolescents are generally assumed not to be competent when it comes to consent for studies. This ability relates to one's cognitive ability and to one's life experience, and

in general, cognitive development and developmental psychology tell us that those who are about age 14 or older have formal operational thought. Individuals at about age 14 can think about risks, benefits, and consequences, and they have an ability to understand research that is comparable to that of adults.

The Federal guidelines specifically talk about waivers to parental permission when parental permission is not a reasonable requirement and the waiver would not be inconsistent with Federal, State, or local laws. That is the only specific language within the Federal guidelines. The Belmont Commission permits adolescent consent in several other areas, and a 1997 report on children from the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research talked about other areas for which adolescents can receive treatment without parental consent, if the teens would be assuming minimal risk. For instances when parents have designated the children as in need of supervision or when the parents are incompetent to decide, the Belmont Report suggests an alternative mechanism for supervised consent.

Because consent can precipitate or result in disclosure, we need to think about how to make it possible to study certain questions—care that is protected, behaviors that are hidden, activities that are illegal, and especially things that are status offenses because of age. IRBs are inconsistent, and it would help to have some specific guidance. One study that looked at 180 IRBs found that 70 percent of them required consent for all research with minors, 71 percent required consent for anonymous HIV seroprevalence, and 52 percent required consent for a survey on satisfaction with care (Mammel and Kaplan, 1995).

We need some Federal guidance here, not just local custom, because Federal and the National Commission report language states what we can and should be able to study, but IRBs do not follow this consistently. It would be nice if agencies and the NIH promulgated that guidance in the form of specific recommendations. On the other hand, there is a trend toward requiring active consent for any kind of study, rather than waived documentation of consent or adolescent consent. How we deal with this issue is very important when it comes to what kind of evidence we want to deliver as a result.

Research requires a higher standard. The Belmont Principles require that we think about the level of potential benefits and harms to any particular research activity.

Rural Populations: Challenges in Recruitment, Retention, and Relevance

Kathryn Whetten-Goldstein, Ph.D., M.P.H.

Dr. Whetten-Goldstein is Director of the Health Inequalities Program at Duke University, Assistant Professor at Duke's Terry Sanford Institute of Public Policy, Assistant Professor of Community and Family Medicine at the Duke University Medical Center, and Senior Research Fellow at the Center for the Study of Aging and Human Development.

Most clinical trials are conducted in urban areas and policies are made in urban areas; therefore, it is easy to forget about the rural areas and to stereotype the people who live there. One-third of all women in the United States live in rural areas. Rural women are more likely to live in poverty than their urban counterparts, and they are less likely to have high school diplomas—25 percent do not. Rural women are more likely to be mothers, and they are more likely to have children who are home, yet they must travel greater distances if they want to get to a clinical trial.

Rural women are more than twice as likely as urban women to suffer from depression and anxiety, according to the American Psychological Association. Recent data on the high rates of mental health disorders in rural areas have been somewhat surprising: 40 percent of rural women are reported to suffer from depression and anxiety. Suicide rates are three times higher for women in rural areas than for some of their urban counterparts, particularly those women in the Midwest. Rural women are more likely to suffer from chronic diseases, and HIV/AIDS is increasing most rapidly in rural areas.

When a map of the mortality and morbidity distributions is superimposed on a map of availability of care, it becomes apparent that care is less available across rural areas than elsewhere.

Teen pregnancy rates are higher in rural areas. Spousal abuse is as high in rural areas as it is in urban areas, but the necessary support services are not available. Women with disabilities have more difficulty obtaining services in

rural areas, and they have more difficulty becoming involved in clinical trials. Social stigmas are greater for lesbians, gay men, and bisexuals in rural areas, a fact that is being recognized within the AIDS epidemic by those of us providing care for men in the South. Many men in the South who are having sex with men are married and have children, and they are not telling their health care providers that they have sex with men. We need to be able to talk about that, but the stigma is so great that it is very difficult, particularly because there are multiple layers of stigma. When people who are poor, African American, or gay, live closely with their families in small communities, being open is very difficult because of the potential effects on so many others.

The concentration of rural populations with the highest age-adjusted mortality are in the South and the Southeast, primarily in North Carolina, South Carolina, Louisiana, Alabama, Mississippi, Georgia, parts of Florida, and Arkansas. This is also true for stroke, heart disease, diabetes, syphilis, gonorrhea, and now for HIV/AIDS. Are the populations in these areas getting care? When a map of the mortality and morbidity distributions is superimposed on a map of availability of care, it becomes apparent that the availability of care does not match up. Care is less available in the needy regions—across rural areas—than elsewhere. Thus, there are some easy answers to why rural women are not as healthy as urban women: Not only are services lacking, but where they do exist, the population may lack awareness of the availability of those services. Providing care in rural areas is costly, but it is also costly for rural women to go to academic medical centers or urban areas to participate in clinical trials.

There are also not-so-simple answers about why rural people are not as healthy; I propose these as hypotheses.

The social fabric of society is particularly rigid in rural areas, and particularly in the South. People in rural areas are less able to move from one community to another because of the social fabric of the individual, the individual's family, and the people with whom they live. Individuals represent their families and the areas in which they grew up. Families have known one another for generations, and particularly in the South, there is a long history of a strong social network, for African Americans as well as for European Americans.

There have been incidences that have made rural people distrust medicine and the government in general. We asked our sample of people who were poor in the South about where they think HIV came from, and the majority of people believe that the Government created HIV.

The cultural divide in the United States between people who live in the land of opportunity and those with no opportunity is fantastically large. People in rural areas often see their urban counterparts—and see physicians—as coming from a completely different place. They are not people with whom they can engage in conversation, who they believe will respect them. Confidentiality is incredibly difficult to maintain. Often the families of providers know the patients' families; thus, individual women do not want to go to their local providers with sensitive medical problems because they know that it is difficult for that provider to keep confidentiality.

At present, I work primarily with people who are HIV-infected and with people who live in rural areas. At Duke University and at the University of North Carolina, 75 percent of patients come from rural counties. We have been collecting both case studies and large data sets; we have followed 700 people for 3 years and are now following 900 people in six States for another 3 to 4 years. We are asking about confidentiality, and we are hearing stories across the board about providers breaking confidentiality, in part because they worry about their own children and the health of their families. For example, one story is about a nurse at the health department who told her daughter that her patient was HIV-positive because she did not want the daughters to play together. The daughter of the HIV-positive patient had a daughter who did not know that her mother was HIV-positive. She found out in school the next day because every child in the school knew that

the mother was HIV-positive. As a result, they had to move to another town.

Another woman in our study moved to another town because people in her workplace found out that she was HIV-positive. A transcriber who was working in one of our rural hospitals found out that one of her neighbors was HIV-positive, and she believed that she and her family were in danger, so she told the community. We take an oath of keeping confidentiality, but we do not address the issues that people face in rural areas where they know one another and where much misinformation exists about HIV and other diseases. In rural areas, there is a whole different level of knowledge about diseases.

There have been incidences that have made rural people distrust medicine and the government in general. They know about Tuskegee and that our Government has been denying money to African American farmers systematically and forever. When I ask my students at Duke whether they know about Tuskegee, the African American students always do, as do people from rural small towns. It is the European American urban students who have never heard of these things, and they are shocked. The issue of trauma and abuse in rural areas: Many people were abused as children, which leads to distrust of authorities and reaction to a medical system that is not the same as for their urban counterparts. As children, rural residents are not usually regularly engaged in medical care, so they are not familiar with it even from an early age.

We asked our sample of people who were poor in the South about where they think HIV came from, and we found that the majority of people believe that the Government created HIV. This reflects a fundamental distrust of systems and the belief that there is another "something" out there that is trying to "get" them. This type of study has been done among African American populations with the assumption that minority groups would be more likely to mistrust the Government. However, we found no differences by race among poor people living in rural areas. More than half of the people believe that the Government created HIV, and those people believe there is a treatment for HIV that they are not receiving.

We conducted a case study of about 25 people who were HIV-positive. All of them—European American, African American, gay, straight—had stopped using their medications and had taken drug holidays. Not one of them had talked to their physician, and there were no racial differences.

What can we do if we decide that we want rural populations to be involved in our clinical studies? First,

we need to take a step back and articulate why we want to include rural residents in the studies. We then need to go into communities and talk about distrust and what a medical care system is and what research is—and we need to do that with the respect of knowing that people’s beliefs emanate from valid reasons, which include their own life experiences and historical knowledge.

People come from long distances to participate in our clinical trials. In North Carolina, a person coming to one of our trials, who lives like 75 percent of the rural population, is there for the day. The individual gets picked up by a van, comes in, sits there all day, and returns home when everyone is ready to go back. If people have children, it is almost impossible to spend an entire day away from home because your children must be picked up and babysitters are not affordable. People should be respected for the time they give to clinical trial participation. They have taken a day out of their lives to do this with you, and they are doing this while having to negotiate issues with children; rural women are more likely to have children at home. We can set up childcare within our clinics, but most of those children are not going to come in for childcare; they are going to go to school, and they need to go home afterward. So we need to think about those issues and compensate more fairly.

People in rural areas do not have the same level of education and knowledge about the medical care system.

They have not been engaged regularly in health care since childhood, as is more likely to be the case with urban residents. In a clinical trial, participants need to show up regularly for care in what is somewhat a foreign environment, and they must interact with people who have backgrounds very different from their own.

I do not think we need to have a different kind of provider for each different cultural group, but we do need to bridge the cultural divide by developing understanding and listening to people. We conducted focus groups with African American women in the eastern half of North Carolina, asking about the race and ethnicity of their providers and if they would be more likely to get care if the provider were African American. They said that they might be more likely to trust that person initially, but that they would also likely be harder on the provider over time if it turned out that that person was not able to listen to them. With a European American provider, there is more distance in the beginning. What patients are looking for is someone who listens to them and understands. They will trust that person.

We need to understand that people are different, that they have different needs, that we can hear them, that those needs do not make them “aliens,” and that we can work with different populations. We need funding to accomplish this because it is difficult to work with communities to decrease barriers to care and bring rural residents into clinical trials.

Patients are looking for someone who listens to them and understands. They will trust that person.

Chapter Nine

Community-Based Participatory Research

Moderators: Cynthia A. Pearson and Glorian Sorenson, Ph.D.

To increase recruitment into and retention to research trials, researchers must establish close working relationships with the community. This session provided three perspectives on collaborating with communities.

Dr. Emmons' presentation provided an important scientific rationale for a public health approach to risk reduction: A large number of people exposed to a low risk may generate more cases than a smaller number of people exposed to a high risk. Interventions should include a continuum of approaches, from downstream strategies that include individually focused interventions such as patient education to upstream approaches that target policies. Interventions also need to address the social contextual factors accompanying risk-related behaviors, such as everyday responsibilities, material circumstances, and social support networks.

Dr. Israel and Ms. Brakefield-Caldwell presented on the key principles of community-based participatory research, which provides a means of tailoring interventions to the concerns and cultures of participants by equitably involving community members, organizational representatives, and researchers in all aspects of the research process, including intervention design, implementation, and evaluation.

Ms. Avery discussed sources of distrust within communities of African American women that may influence bridge building between researchers and communities. Researchers may benefit from collaborations with community health activists, who may serve as effective spokespersons and can help build linkages between researchers and the community. Collaborations with the community must be built on an understanding of community needs and an appreciation of the community's available resources and sources of strength.

Communities as Essential Partners in Research: As Participants and Subjects

Karen M. Emmons, Ph.D.

Dr. Emmons is Professor of Health and Social Behavior at Harvard School of Public Health and the Dana Farber Cancer Institute. Her research focuses primarily on community-based approaches to cancer prevention and control among low-income and underserved populations. Dr. Emmons' presentation was prepared in collaboration with Dr. Glorian Sorensen, Professor of Health and Social Behavior at Harvard School of Public Health, and the Dana Farber Cancer Institute and Director of the Dana Farber's Center for Community-Based Research.

In the past few years, increasing realization has occurred regarding the role of behavioral risk factors in chronic disease. Between 60 percent and 70 percent of cancer deaths in the United States are estimated to be attributable to behavioral risk factors—primarily tobacco, adult diet and obesity, and sedentary lifestyle. These same behavioral risk factors are responsible for the morbidity and mortality associated with many other diseases.

This evidence has resulted in increasing attention to the prevention of disease, focusing on behavioral risk factor reduction. Two different approaches have been used to address behavioral risk factors. Early studies such as the Lipid Research Clinics Coronary Primary Prevention Trial and the Multiple Risk Factor Intervention Trial were designed primarily as clinical trials to test strategies to reduce risk among medically high-risk individuals. Although this approach was an important starting point, it had some limitations from a public health perspective because it ignores the underlying cause of high risk and because it is improbable that the entire population can be reached by using a strategy of identifying high-risk individuals one at a time.

Community intervention trials represent a radical departure from this model. This approach has included a set of studies on coronary heart disease followed by efforts targeting specific risk factors, such as the Community Intervention Trial for Smoking Cessation and American Stop Smoking Intervention Study trials, as well as strategies targeting specific settings, such as work sites and schools. A major construct underlying community intervention trials is that of population attributable risk, based on the work of Geoffrey Rose. Our current definitions of risk are different from when this work was conducted, but the major point is that the great contribution to morbidity and mortality at the population level is not

associated with those who have the highest level of risk but rather with those who have more moderate levels of risk, in large part because of the distribution of risk in the population.

The great contribution to morbidity and mortality at the population level is not associated with those who have the highest level of risk but with those who have more moderate levels of risk.

In his early work, Rose pointed out the importance of shifting the distribution of risk to the left and the resulting impact on disease morbidity and mortality. In the figure below, the high-risk approach can be viewed as targeting the right hand tail of these distributions; the alternative is to attempt to shift everybody to the left. The result is that everybody moves a smaller amount, but everybody is moving, not just a few people. This approach has the potential to improve population health to a much greater extent than that of a high-risk approach, while at the same time reducing the costs of identifying high-risk people. On the other hand, the cost of providing an intervention to an entire population can be higher than providing it only to people in the right-hand tail of the distribution. Which approach is the most cost-effective in any given setting depends in large part on the prevalence of high risk in the population and on the cost of identifying those high-risk people, compared with the cost of the available risk reduction strategies.

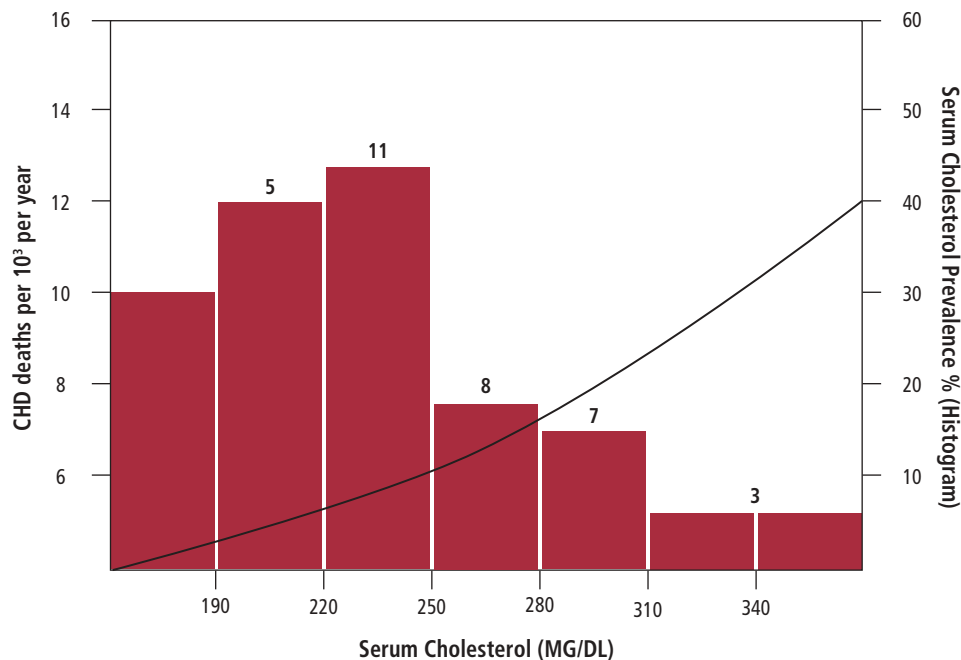
Following this reasoning, Rose concluded that a large number of people exposed to a small risk may generate more cases than a smaller number of people exposed to a

MHHP Population Attributable Risk

Attributable or Excess CHD Death/1000/10 years (Excess Risk x Population Exposed)		
TC	% Population	% Excess Deaths
220	45	15
220-310	50	75
310	5	10

Rose: Framingham, 26: 1970
WHO Expert Committee Report, 1932.

Relative and Attributable Risk with Cholesterol Level



high risk. This principle is illustrated in work conducted by Tosteson and colleagues, who estimated the cost-effectiveness of population-wide strategies to reduce serum cholesterol using the changes achieved in the North Karelia Project and Stanford Five-City Project as standards for comparison. They concluded that community-based approaches to reducing serum cholesterol would be cost-effective if cholesterol levels were reduced by only 2 percent or more.

Risk reduction (see figure below) is a continuum that ranges from individually oriented approaches—such as those targeting patient education or individual behavior change—to midstream approaches that focus more on advertising in the environment, all the way to upstream approaches such as policies or taxes. The impact is quite different for each of these approaches. The downstream approaches target specific individuals, some of whom are high risk and some of whom are not, on a one-to-one level.

Midstream approaches target networks and communities, and upstream approaches address broad populations such as whole communities or States.

Adopting a society and health perspective means raising questions about how social structure may affect personal choice and health. For example, the Alameda County Study data demonstrate a clustering of risk factors associated with low income, including not only behavioral risk factors such as sedentary lifestyle, smoking, and obesity but also unemployment, lack of instrumental support, living in an unsafe neighborhood, and having unmet needs for food and medical care.

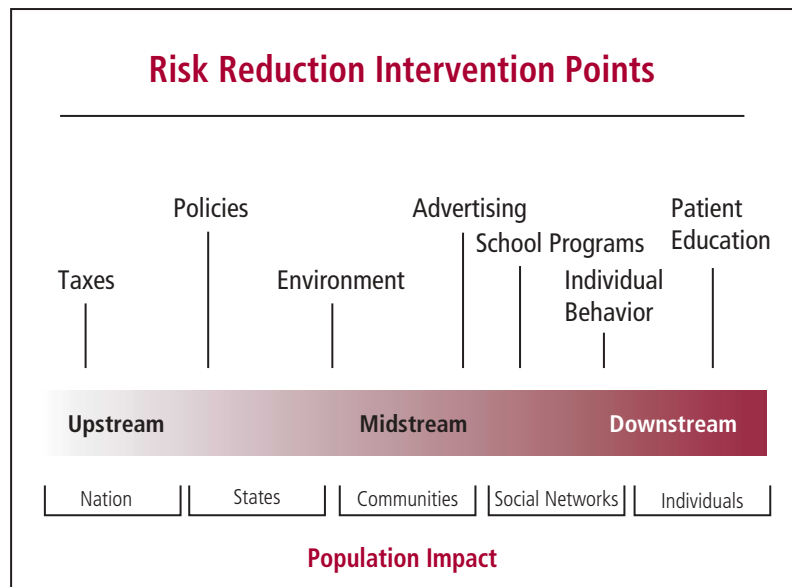
Hillary Graham in England, who studied smoking among low-income women, has conducted some of the most profound work looking at some of these social contextual factors.

She found that social contextual factors associated with low income were particularly relevant for smoking patterns, which is one type of risk-related behavior. Graham concluded that different dynamics drive the smoking habits of low-income women compared with those in upper- and middle-income classes. In particular, low-income women use smoking as a means of coping with their economic pressures and the resulting demands placed on them to care for others. Graham has categorized these different influences, including everyday responsibilities such as childcare and elder care; patterns of paid work; material circumstances such as housing circumstances, debt and budgeting, and access to a car; social support and social networks, including the people around you and their behaviors; and personal and health resources such as patterns of health-related behavior and alternative coping strategies.

Even among low-income women, which we often think of as a homogeneous group, Graham found that smoking rates were highest among those women with the fewest resources and the most overall responsibilities. A key point of Graham's work is that the cumulative exposure to disadvantage increases the risk of smoking among

women, which means that gradients of risk are present even within that low-income group. In one study conducted in England during the early 1990s, Graham evaluated smoking prevalence among women of various levels of educational attainment. About 46 percent of women who did not have a high school education were smokers; within that group of women, those who were in a low-skill job had a smoking rate of about 50 percent. Among women with no high school education in a low-skill job who lived in subsidized housing, 67 percent smoked. Among women who lived in subsidized housing and received income support, 73 percent smoked. All of that compared with a 22 percent smoking prevalence among women who had completed high school. Multiple logistic regression has confirmed that each level of risk within low-income sub-

groups had an independent effect on the relative risk of being a smoker. From her large body of work, Graham has concluded that smoking follows the pathways that lead to low education and low-skill work, to public housing subsidies, and to income support. These data highlight the imperative of adopting a community and population perspective to health.



Social epidemiology has helped identify socially determined factors that may have profound influences on health and health behaviors. However, little work has been conducted to date that considers how these factors influence people's interest in health behavior change or their ability to reduce their risks. If we are to make a meaningful impact on chronic disease, morbidity, and mortality, it is critical that we begin to address the intersection between health and these societal-level contextual factors.

In summary, it is important to remember Rose's seminal words about the importance of a population- and community-based perspective: "A large number of people exposed to small risk may generate more cases than a small number exposed to a high risk." Without such an approach, it is unlikely that we will be able to make true progress on chronic disease, morbidity, and mortality.

Community-Based Participatory Research: Principles, Strategies, and Lessons Learned From the Community Action Against Asthma Project in Detroit, Michigan

Barbara A. Israel, Dr.P.H., M.P.H., and Wilma Brakefield-Caldwell, B.S.N.

Dr. Israel is Professor and former Chair of the Department of Health Behavior and Health Education, School of Public Health, University of Michigan, and has extensive experience in conducting community-based participatory research in collaboration with partners in diverse ethnic communities.

Ms. Brakefield-Caldwell is a member of the Steering Committee of Community Action Against Asthma and a former administrator with the Detroit Health Department.

I**SRAEL:** We will share some of the results of our experiences in conducting intervention and basic research aimed at increasing our understanding and reducing the environmental triggers of childhood asthma. The caregivers of the children involved in the study are predominantly African American and Latina women. Although this is not a clinical trial, our aim is to examine the use of a community-based participatory research (CBPR) approach, with an emphasis on the benefits of this approach and the lessons learned that are applicable to recruitment and retention of women and people of color in scientific investigations.

Asthma is the most common disease of childhood affecting approximately 5 million children in the United States. Childhood asthma is particularly prevalent in low-income communities, urban communities, and communities of color.

We will begin with a brief overview of childhood asthma prevalence and the risk factors associated with childhood asthma. We then will describe the implications for research and interventions, present a definition and key principles of CBPR, and provide a description analysis of the work we are doing in Detroit.

Asthma is the most common disease of childhood in the developed world, affecting approximately 5 million children younger than 18 years of age in the United States. From 1982 to 1994 the prevalence rate of pediatric asthma in the United States increased by 61 percent, and the mortality rate from childhood asthma for persons 19 years

old and younger increased by 78 percent from 1980 to 1993. Childhood asthma is particularly prevalent in low-income communities, urban communities, and communities of color. The national trends showing increases in childhood asthma are similar to those found in the city of Detroit.

The causation and aggravation of pediatric asthma are complex and multifactorial. A number of risk factors are associated with childhood asthma, including genetic predisposition, demographic factors such as socioeconomic status, indoor environmental exposures such as dust mite and cockroach allergens and tobacco smoke, outdoor environmental exposures such as particulate matter and ozone, and psychosocial stressors such as violence, crime, and lack of community resources.

BRAKEFIELD-CALDWELL: Based on our current understanding of the factors associated with childhood asthma, a number of implications for research and interventions arose. We wanted to address the complex set of factors associated with childhood asthma, which is a major challenge for researchers, practitioners, and affected communities. Historically, research has rarely directly benefited, and sometimes has actually harmed, the communities involved and has excluded them from influence over the research process. As Dr. Sugarman alluded, we must always remember the U.S. Public Health Service Research Tuskegee study and the Willowbrook hepatitis study. Often, interventions have not been as effective as they could be because neither were they tailored to the concerns and cultures of the participants nor did researchers include the participants in all aspects of the intervention design, implementation, and evaluation. There have been increasing calls for more participatory and comprehensive approaches to public health research and interventions to address these issues.

ISRAEL: One approach to CBPR is a partnership approach that equitably involves, for example, community members, organizational representatives, and researchers in all aspects of the research process, with all partners contributing their expertise and sharing responsibility and ownership. This level of participation enhances understanding of the given phenomenon and allows integration of the knowledge gained with intervention to improve health. The emphasis here is on knowledge generation and benefit to the community and on the active involvement of the community in all aspects of the research process.

BRAKEFIELD-CALDWELL: The following key principles seek to capture the key elements of this approach: (1) CBPR recognizes community as a unit of identity, which is characterized by a sense of identification and emotional connection to other members, common symbol systems, shared values and norms, mutual influence, common interest, and commitment to meeting shared needs. Communities of identity may be centered on a defined geographic neighborhood or on a geographically dispersed group with a sense of common identity and shared fate. CBPR attempts to identify and work with existing communities of identity and/or to strengthen a sense of community through collective engagement.

(2) CBPR builds on the strengths and resources within a community, such as the skills and assets of individuals, social networks, and mediating structures such as churches and other organizations where community members come together to address their health concerns. (3) CBPR facilitates a collaborative, equitable partnership in all phases of the research via an empowering and power-sharing process that attends to social inequalities. CBPR partnerships focus on issues and concerns, identified by community members, and create processes that enable all parties to participate and share influence in the research. (4) CBPR promotes colearning and capacity building among all involved partners; for example, researchers learn from the knowledge and local theories of community members about how to effectively retain study participants, and community members acquire additional skills in how to conduct research. The emphasis is on enhancing the capacity of all people involved.

ISRAEL: (5) CBPR integrates knowledge generation and intervention for the mutual benefit of all partners. It may not always involve an intervention component, but there is a commitment to translating and integrating research results with community change efforts, with the intention that the partners involved will benefit from the research.

(6) CBPR emphasizes the local relevance of public health problems and the multiple determinants of health and disease, for example, biomedical, social, economic, physical, and environmental factors. (7) CBPR involves a cyclical and iterative process that includes partnership development and maintenance, defining the problem, deciding what kind of information will be collected, how to collect the data, how to analyze and interpret the data, how to use and disseminate the resulting information, and how to develop an intervention if such a component is warranted.

One approach to CBPR is a partnership approach. The emphasis here is on knowledge generation and benefit to the community and on the active involvement of the community in all aspects of the research process.

(8) CBPR seeks to disseminate findings and knowledge gained to all partners and involves all partners in the dissemination process, for example, involving community partners as coauthors and reviewers of publications and as copresenters at meetings and conferences, such as Wilma and I are doing today. (9) CBPR involves a long-term process and commitment. To establish and maintain the trust required for successful CBPR efforts, this long-term commitment must extend beyond a single research project and beyond a single funding period.

BRAKEFIELD-CALDWELL: We will now discuss a partnership in which we are involved that follows CBPR principles. The Michigan Center for the Environment and Children's Health (MCECH) was born from the Detroit Community-Academic Urban Research Center (URC), which was the original partnership. The URC Board identified diseases related to environmental concerns, including asthma, as a priority area; that was not an easy process. It took more than a year for people to agree on what they thought was important. The URC partnership applied for and received funding from the National Institute of Environmental Health Sciences and the U.S. Environmental Protection Agency, through their Centers of Excellence initiative.

The MCECH is governed by community-based participatory research principles. Some of our partners include community-based organizations, health and human services organizations like the Detroit Health

Department, the Henry Ford Health System, and faculty from the University of Michigan School of Public Health and Medical School. In addition to some of the community-based organizations involved in the URC, we have people from an environmental justice organization partnering with us, as well as people who belong to the housing coalition, so we have a representative group of people who are really interested in the subject of asthma and the environment.

ISRAEL: The overall goal of the MCECH is to investigate the environmental, pathophysiological, and clinical mechanisms of childhood asthma and to implement and evaluate comprehensive community- and household-level interventions aimed at reducing asthma-related environmental threats to children, families, and neighborhoods.

The MCECH has three core research projects, the first two of which are integrated into what we call the Community Action Against Asthma (CAAA):

1. Household- and neighborhood-level interventions that focus on reducing environmental triggers of childhood asthma.
2. An exposure assessment to assess the separate and possible intervention effects of outdoor and indoor air quality on the exacerbation of asthma in children. These two studies have been combined into the CAAA.
3. A murine model based at the University of Michigan School of Medicine to determine whether the mechanisms of chronic pulmonary inflammation due to repeated exposure to allergens is mediated by excessive local production of chemokines.

BRAKEFIELD-CALDWELL: I will give you a description of the intervention and exposure assessment project, which is guided by a steering committee comprising representatives from each of the partnership organizations. The CAAA involves participants from two geographic areas within the city of Detroit: The east side is 95 percent African American, and the southwest side is 50 percent African American, 40 percent Latino, and 10 percent non-Latino white. The east and the west sides of Detroit are like two different cities, and these two different groups had never worked together before.

We enrolled 300 families with at least one child between 7 and 11 years old with probable or known asthma. Ninety-five percent of the caregivers participating were women, 81 percent were African American, 12 percent were Latina, and 5 percent were white. Forty-six percent of the annual household income was below \$10,000, and 32 percent of the annual income was between \$10,000 and \$20,000. Recruiting was accomplished using screening questionnaires distributed through the mail and at school. Of the more than 9,000 distributed questionnaires, approximately 3,000 were completed and returned. Of that number, approximately 50 percent of the returned questionnaires described a child with asthma of any severity, 12.9 percent had moderate to severe asthma, and of these children, approximately 30 percent had never been diagnosed by a physician.

The objectives for the household-level intervention activities for the CAAA were to increase the knowledge and perceived self-efficacy of participants about asthma and encourage behaviors to reduce environmental triggers. We wanted to increase behaviors to reduce indoor environmental triggers by explaining to parents about vacuuming and damp mopping, and we wanted to reduce our indoor exposures to environmental triggers by decreasing the dust levels and the cockroach allergens. One of the successful activities was to give these participants a vacuum cleaner at the beginning because it seemed that many of these participants did not know how to clean their homes in ways that reduce asthma triggers. We also wanted to strengthen psychosocial factors associated with asthma-related health status, for example, social support. We wanted to improve their asthma-related health status—quality of life, functional status, and symptom severity—and we wanted to reduce unscheduled asthma-related health service utilization.

The activities included randomizing the number of people who were in the research design, so the 300 families were randomly assigned to either wave 1 or wave 2. The wave 1 families received intensive intervention the first year and the wave 2 families received intensive intervention the second year—the 2-year intervention

The objectives for the household-level intervention activities were to increase the knowledge and perceived self-efficacy of participants about asthma and encourage behaviors to reduce environmental triggers.

offered in intensive and less intensive phases. A minimum of nine visits were conducted by our community environmental specialists, who were outreach workers hired and trained by the partnership, in the intensive phase, and three visits were conducted in year 2, the less intensive phase.

Recognizing that some of the environmental triggers are beyond the ability of any one individual to control, we also have a community-level intervention.

The household intervention was tailored to the needs of each child and family based on what the child was allergic to and what allergens were in the dust. The intervention activities included education—about dusting, cockroaches, and cleaning the house. We distributed materials from the Dollar Store such as buckets and sponges for dusting. We did not spend a lot of money so they would know that they could go to the Dollar Store and purchase the same materials to dust the furniture, wipe it down, and make it clean, especially in the child's bedroom. We also offered integrated pest management in about 100 homes. In addition, we helped people with housing and obtaining city services such as health care; for example, one Christmas somebody donated two truckloads of furniture that was distributed among the families that needed it.

ISRAEL: Regarding the evaluation research methods, we collected several types of data. We skin-tested all the children before they were enrolled, and we conducted annual measurements of both the caregiver and the children, using questionnaires to assess, for example, psychosocial factors and health factors. We conducted household dust sampling and provided an environmental checklist in the household annually. We also did a neighborhood environmental checklist—we hired community members to walk each block and fill out a checklist on the factors in the environment that might have a negative impact on health status.

We collected seasonal physical, environmental, and health outcome data as part of both the evaluation and the exposure assessment, which was 2 weeks long, four times a year. I want to give you a sense of how much we were asking of these participants, which is really relevant for thinking about clinical trials. For 2 weeks, four times a year, for 2½ years, we asked the caregivers to keep a daily diary of symptoms, we asked the children two times a day to blow into a hand-held peak flow monitor, and we

asked the caregivers to make a list of all the medications their children were using. We also conducted daily ambient measures of particulate matter, ozone, and other meteorological variables on the rooftops of two schools, and we conducted indoor daily measures of particulate matter and vapor-phased nicotine in a subset of 15 homes. In the same 15 homes, we also asked the children during those 2 weeks to carry a backpack 24 hours a day; the backpack had a monitoring system that sampled the air quality around them everywhere they went for 2 weeks.

Preliminary results show improvement in the use of asthma controller medicines and improved asthma symptoms, specifically persistent cough, wheeze with cold, and cough with exercise. We have also seen a reduction in cat, dog, and dust mite allergen concentration, a reduction in caregiver depressive symptoms, and an increase in perceived social support on the part of the caregiver.

BRAKEFIELD-CALDWELL: In addition to the household-level intervention and recognizing that some of the environmental triggers are beyond the ability of any one individual to control, we also have a community-level intervention that began within the past year. The objectives of this intervention are to increase knowledge about asthma and environmental triggers within the community, increase the capacity of neighborhoods to reduce environmental triggers, and reduce physical and environmental hazards in the neighborhoods involved (e.g., illegal dumping, air pollution).

Benefits of Using a CBPR Approach: Contributions to Communities and Science

- Enhances relevance and use of data
- Increases quality and validity of research
- Improves research and intervention design and implementation
 - Randomized staggered research design
 - Selection and training of outreach workers
 - Recruitment
 - Retention

ISRAEL: I want to emphasize a number of benefits that we consider from using a CBPR approach, benefits both to science and to the communities involved. CBPR enhances the relevance and use of the resulting data. For example, community members encouraged us to measure diesel exhaust, based on their concerns about truck traffic through the communities. We have been able to

obtain subsequent funding to do that and are finding quite promising results. We also included survey questions on stressors that were identified by community members, stressors that we never would have known to ask about. We are now using those stressors in our community-level intervention to set priorities for which actions to take.

CBPR increases the quality and validity of the resulting data. For example, the survey questions were revised a number of times by the steering committee to be more culturally appropriate. We hired and trained community members as interviewers, and they were more easily able to develop rapport with and obtain quality data from the participants.

CBPR enhances the research and intervention design and implementation. Given the understandable distrust of research in historically marginalized communities, coupled with the requirement that research conducted must benefit the communities involved, the use of a traditional control group is considered unacceptable in the communities in which we work. However, given the trust that we were able to establish with our CAAA partners, we agreed on the randomized staggered design described here, which enabled wave 2 participants to serve as a “control group” since they did not receive the intervention during the first year.

BRAKEFIELD-CALDWELL: The steering committee selected and trained the outreach workers, and we chose individuals from the community who were knowledgeable and sensitive to the needs, values, and cultures of community members and which intervention strategies were likely to work. The steering committee also was instrumental in hiring a community member to handle all the recruitment and contacts with the schools and

community organizations, and a community member visited homes to recruit participants to use the indoor air monitors. The steering committee played a critical role in designing the skin test fairs, which were conducted in the community rather than at a health care facility and which included activities for the entire family—food, games, and prizes.

Regarding retention, we send out birthday cards to the participating children, we send out holiday cards, we give them calendars and magnets and newspapers, we have newsletters with contests so the children can get involved—all of this based on the ideas of the steering committee. We developed a field trip for the children and families involved in the indoor air sampling to go to the University of Michigan, because these children had never been on a university campus. This trip took place one summer, and they visited the labs where the data from their houses are analyzed; it was a lot of fun for the children, the parents, and the project staff. We gave all participants the results of the skin tests and lung functioning tests, and handed out numerous incentives such as coupons to local restaurants.

ISRAEL: Regarding retention, after the first year of the study, of the 300 families initially enrolled, 237 families remained. Of the 63 families who were no longer participating, approximately 50 percent had moved, 40 percent were lost due to not being able to locate them, and only 10 percent of the families left because they no longer wanted to participate. We believe that the active involvement of the steering committee throughout the process and the recommendations it made are a major reason for participants’ continuation with this project.

The steering committee selected and trained the outreach workers, and we chose individuals from the community who were knowledgeable and sensitive to the needs, values, and cultures of community members.

Research and the Community: Shifting the Paradigm

Byllye Y. Avery, M.Ed.

Ms. Avery is the founder of the National Black Women's Health Project (now called the Black Women's Health Imperative) and has been a health care activist for more than 25 years. A dreamer, visionary, and grassroots realist, Ms. Avery has just started a new organization that also addresses the health needs of African American women, the Avery Institute for Social Change.

I am not a researcher; I am a community activist, so that is the perspective I bring to you. Most everything I am going to say has already been said, so you have gotten these messages. It is affirming to know that many of us see this issue in very much the same way.

We have heard a lot of talk about Tuskegee and the mistrust that it engendered, but I say to you there were many other things that have also fostered mistrust. I remember when we started the National Black Women's Health Project in 1981, one of the first things we learned was that Depo-Provera had been given to a whole bunch of women at Grady Memorial Hospital in Atlanta, Georgia, who had not signed a consent form. All the side effects made Depo-Provera the perfect drug not to be given to a population of African American women: It caused weight gain, brought on high blood pressure (which brought on diabetes), caused women to lose their hair and their libido, and basically made everybody “nuts.” The thing about bad news is that everybody gets it; everybody heard about this when it happened.

I also think about the early days when J. Marion Sims called himself the “architect of the vagina”—the way he operated on slave women without anesthesia to effect the operation for vaginal fistula. I think about that every time I pass his statue in Central Park. A lot of things have happened to us, and so therefore it has been imprinted in our brains.

I was looking at Julie Dash's film *Daughters of the Dust* the other day. They talked about how a lot of the bad things that happened to us remain imprinted and come to us through the “old souls.” Sometimes you do not know exactly where it comes from, but most of it is not without basis; we live in a society with a lot of bad stuff going on all the time, so people are going to be mistrustful. Mistrust can be overcome by addressing it

from the start and by openly beginning to build a positive relationship from which everyone involved will grow and change. If the work you are doing is not making you grow and change, then you need to question what you are doing, and more importantly, you need to question why are you there. So we should expect growth and change.

Think about building relationships with community health activists. There are many of us out there, and many of us will not back down from pushing back on you and challenging you.

We ask that you think about looking at building relationships with community health activists. There are many of us out there, and many of us will not back down from pushing back on you and challenging you. But it is through that challenge that change happens. Looking at the women's health movement, the reason that we are able to sit here today was because of the “crazy” women in the late 1970s and the 1980s—we were told we were totally crazy for looking at our cervix, for questioning the medical deity, and for talking about the research. When we found out that Vivian Pinn was going to be the head of the NIH Office of Research on Women's Health, the first thing we did was march right in there and talk with her. Those are the kinds of people you need to align yourself with, because sometimes messages might need to be delivered and you might not be the best person to deliver them. You might be tied in other ways, and most of us activists are not going to get tenure or a lot of other things that sometimes keeps you from being active. We can be effective spokespeople, and we can also help

bridge the way you enter the community because, if we were not the trusted people, we would not be there.

You need to look at who is the expert here. We really denounced this whole idea of the experts coming in and “doing” unto us. We are all experts, and I love the idea of Mildred with her street Ph.D., because you have to be able to look at the strength of a population. People talk about homeless people; if people can survive on the streets of Boston in the wintertime, they are strong people. I think I might last only 15 minutes because I cannot stand to be cold. But I am saying that you need to look at the population, who are the real experts, who are the teachers, and who are the students and to know that in some sense we are all of those at different times—and have respect for that.

It is also very important to adhere to an old concept that has been around a long time: The right message and the right messenger can take you a long way.

Another suggestion as you start to do your work is to educate research participants about their family histories. When people start to understand that some of this stuff is predictable in families, they will understand why they should be involved in research and why they need to become the activist, the educator, the person in their families to work that way. In the African American community, a lot of families have family reunions, and they go down the family tree—I went to one for my husband, and it took them 20 minutes before they got to where we were in that family. You have the family tree historian, but what about the family health historian? Some of the workshops offered could be on diseases that are prevalent in these families, and that whole approach needs to be studied as a viable way to do health education.

Another big thing that gets complained about in communities when researchers come in is that they put in a program that is sexy and wonderful, but when the money runs out, the programs are gone and the community still has unmet needs. That angers people more than anything else, so much so that when they see researchers coming they say, “Now what do you want?” We need programs that, from the beginning, will be successful and have a plan for continuing if it is needed in the community once you have finished your research. This was very common among a variety of prevention programs for teenagers;

people worked hard to come up with stuff that worked, but the money ran out and they were left empty-handed because they did not know how to fundraise although they knew how to run the program.

We have serious health care problems and health care needs. We continue to measure the success of the U.S. health care system on the medical part and on all the wonderful technology, but I maintain that that is the wrong end of the stick. We need to measure it on how well we get health care to the people when we have millions of people without health care, all these millions of people without access; how can we say we have the best health care system in the world, when in fact we do not? We have a very excellent medical system, but when it comes to health, we are losing.

So we need you to be the research activist, we need you to continue to think out of the box when doing all the stuff everybody else has told you to think about. We need you to figure out how you can work on yourself, how you can muster up the courage to stand tall for what it is we need and see how we can make it happen. So that when we talk about being the greatest health care system, we really are—because we are getting services.

The isolation that a lot of people of color face in institutions is ludicrous. I am talking about involving people of color right from the beginning, to show how you will treat the community—on the basis of equality. I am not talking about window dressing—just because you want to work with the Latino population, you need a Latino person, so you get somebody and they are essentially window dressing because you are doing all the work and getting all the credit. We do not need that. One of the things that a lot of people talk about is feeling isolated in their universe, so get yourself together and reach out to and learn from and work with each other.

The last thing I want to leave you with is to be strong, because we are witnessing around reproductive health where politics is trumping science. If that continues to happen, then we are right back where we began with the mistrust. Instead of having it happen among certain affected communities, it will then happen to everyone. We do not need that to happen, so please be strong and maintain your integrity. Our lives depend on it.

How can we say we have the best health care system in the world, when in fact we do not? We have an excellent medical system, but when it comes to health, we are losing.

Chapter Ten

Recruitment Is Only Half the Challenge: The Essentials of Retention

Moderator: Marian C. Limacher, M.D.

Although many clinical trial investigators place the highest priority on recruitment of adequate, appropriate participants into studies, all trials that require followup are subject to loss of participants. Once recruitment is completed, attention to retention is paramount. The panelists for this session covered a spectrum of issues relevant to the importance of retaining participants in clinical trials. Dr. Scharfstein addressed the key concepts of analysis based on “intention to treat” as opposed to “as treated.” He illustrated the potential for bias when censoring and the underlying reasons for losing participants to followup are not accounted for. From a statistical perspective, high rates of retention are mandatory for obtaining reliable, credible, and translatable results. Dr. Greenblatt provided her perspective on conducting clinical trials among women in the most challenging circumstances: HIV-exposed women without adequate social support, income, health coverage, or finances. She reported remarkable followup success rates when clinical centers pay attention to issues that matter most to participants; providing basic needs such as travel and meal reimbursement are critical to retention of participants who do not have adequate resources. Dr. Ockene offered the experience of the large-scale hormone

trial of the Women’s Health Initiative. High rates of retention have been accomplished by an attentive staff. Supplemental contacts, reminders, and support materials are employed routinely to maintain participant connection with the study and the clinic staff. Compared with pill-taking interventions, lifestyle and behavioral change interventions constitute even greater challenges to retention. Dr. Kumanyika described several clinical trials of diet, exercise, and weight loss for African American women. She identified the tension inherent in clinical trial design between enrolling representative participants while risking high losses and restricting enrollment to participants most likely to adhere to and maintain followup while achieving results that might not be applicable in a real-world situation. Each of the speakers offered important considerations and practical suggestions. Clinical trials will always be challenging. The process can benefit from communicating those challenges and sharing successful and less-than-successful strategies.

The Importance of Retention in Clinical Studies— A Statistical Perspective

Daniel Scharfstein, Sc.D.

Dr. Scharfstein is Associate Professor for Biostatistics at the Johns Hopkins Bloomberg School of Public Health.

His research focuses on developing statistical methods for reporting the results of randomized studies.

To explain why it is so important to retain participants in clinical studies I will first compare and contrast the analytic objectives for reporting the results of clinical studies: Intent-to-treat versus as-treated analyses. Next I will cover the potential impact of loss to followup and make a few recommendations. I will give an example of an AIDS clinical trial that was designed to test treatments for patients with HIV and in which survival time was the end point. In this trial, there was a lot of censoring—the survival time for many subjects was not observed directly. In some cases, the censoring was due to premature loss to followup, and in others it was due to end of study.

An intent-to-treat analysis compares subjects in their initially assigned treatment groups and ignores all information on adherence to treatment. It is considered by many clinical trial investigators to be the primary analytic objective. It is called the pragmatic analysis and is also called an effectiveness analysis. Intent-to-treat looks at what would happen if everyone were assigned treatment versus everyone not assigned treatment, and it ignores information on adherence. An as-treated analysis compares subjects based on the treatment actually received. It addresses the fact that poor compliance in the trial can dilute the intent-to-treat effect, so the analysis might not detect an intent-to-treat effect because of that poor compliance.

Results of these analyses can be biased because the prognostic balance, which was guaranteed by randomization, is likely to be disturbed. The disturbance occurs because the comparisons are of groups of people based on something that happened after baseline, which results in a prognostic imbalance. A lot of recent research activity has led to the development of statistical methods to help eliminate this kind of bias, but these analyses tend to rely on strong assumptions that are not testable from the data

at hand. Many investigators consider this the secondary analytic objective because it can help explain the results of an intent-to-treat analysis.

So both the intent-to-treat analysis and the as-treated analysis require complete followup on all subjects. If dropout is related to the outcome under investigation—if the people who are dropping out tend to be sicker or if they tend to be healthier—an inference based on standard techniques may be biased, yielding the wrong answer. The results will rely on untestable assumptions that cannot be validated from the data at hand, such as the terms “missing at random” or “explainable censoring.” A nonparametric, flexible kind of analysis will not be possible; some form of modeling will be required. The final inference will depend on the extent to which substantive experts believe these underlying modeling assumptions, and there will be a loss of power to detect treatment effects because there are fewer data than originally planned.

Recent research activity has led to the development of statistical methods to help eliminate bias.

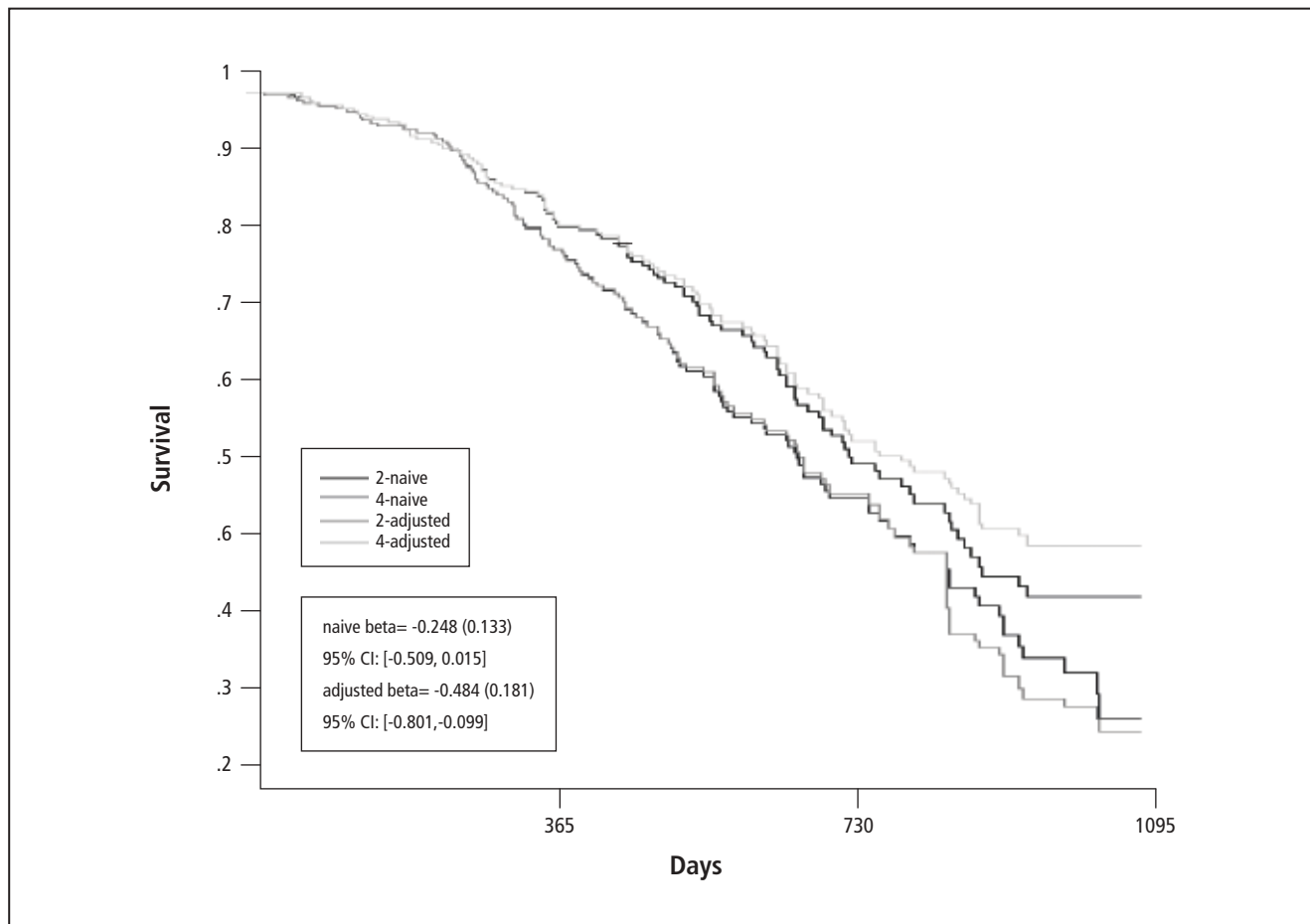
The first recommendation is to minimize loss to followup. If some participants fail to comply, keep following them and do not give up on them. The second recommendation is to collect information even after noncompliance occurs. It is also useful to collect auxiliary information that is jointly prognostic for dropout and outcomes to help make assumptions more tenable. Collect information on factors that you think might be prognostic for dropout, for example, additional health status information that can help make the assumptions

more tenable and more believable by the scientific community. Do not just do one analysis if you think your dropout is really informative; perform sensitivity analysis and try several different methods.

As an example, the AIDS clinical trial group (ACTG) 193 was a randomized, double-blind, multicenter trial that compared the efficacy of four regimes for treating HIV-infected individuals with advanced disease. The primary event was survival, and the primary analytic objective was intent to treat. Treatment two and treatment four are the two that I will compare here. In treatment two, 123 subjects were observed to die, 56 (18 percent) were prematurely lost to followup (we do not know what happened to them), and 137 of them were administratively censored, that is, they made it to the end of the study, but they still had not experienced their event of interest. In treatment four there were 105 deaths, 60 subjects were prematurely lost to followup, and 163 subjects were administratively censored. During

accrual in this study, the protocol was changed to allow some healthier subjects to enroll because of problems recruiting subjects. Therefore, subjects who were coming in later tended to be healthier than those who came on study earlier; this induces an informative censoring problem because the people who were coming later were more likely to be administratively censored.

The data show that, among subjects who were at risk for premature loss to followup at any point in time, those with lower CD4 counts were more likely to be lost—so the people being lost were the sicker subjects, with lower CD4 counts. Among subjects who were at risk for administrative censoring at any point in time, those with higher CD4 counts and no prior use of nucleoside therapy were more likely to be censored—so the healthier subjects were more likely to be administratively censored. If premature loss and administrative censoring were truly noninformative, which are the assumptions that underlie standard analyses such as the log rank test and Kaplan-



Meier curves, then these associations should not be observed. As a result, estimates of survival curves using, for example, Kaplan-Meier curves, will be biased.

We used a method to adjust for some of this bias, and it relies on assumptions. In the graph on the previous page, the “2-naive” curve is the Kaplan-Meier curve for the survival distribution for treatment group two. Roughly 50 percent of subjects survive past 700 days. The “4-naive”

Retention is critical so that we can avoid having to model and make untestable assumptions.

curve is the Kaplan-Meier estimator for treatment group four. These curves may be too pessimistic because sicker subjects are lost prematurely and healthier subjects are more likely to be censored administratively. When this was adjusted by using modeling and untestable assump-

tions, the curve for treatment group 2 moved from the “2-naive” to the “2-adjusted,” the curve for treatment group 4 moved from the “4-naive” to the “4-adjusted,” and there was a greater separation between those curves.

The default analysis, which compares the “4-naive” with the “2-naive,” gives a log hazard ratio of -0.25 , and the confidence interval contains 0, so there is not a statistically significant effect. Adjusting for the informative censoring, we now compare the “4-adjusted” to the “2-adjusted” curves. There is now greater separation, the log hazard ratio moves up from -0.25 to about -0.5 , and the confidence interval no longer contains 0. This illustrates that, by failing to account for informative censoring, treatment effects could be missed or treatment effects could be concluded.

From a statistical perspective, retention is critical so that we can avoid having to move to modeling and having to make untestable assumptions. This makes the analysis more valid and more credible to the scientific community.

Retention Strategies: Lessons From the Women's Health Initiative

Judith Ockene, Ph.D.

Dr. Ockene is Professor of Medicine at the University of Massachusetts Medical School in Worcester and Chief of the Division of Preventive and Behavioral Medicine. She is also a principal investigator for the University of Massachusetts site for the Women's Health Initiative

The hormone study of the Women's Health Initiative (WHI) consists of two hormone treatment trials: one testing estrogen and the other testing estrogen plus progestin. We recruited more than 27,000 women initially between the ages of 50 and 79 and, in the seventh year of followup, the oldest participants are now 86 years old. Our populations are aging and elderly women, so some of our concerns include how to retain women in their 80s.

Followup of the women in the hormone trials included a telephone contact 6 weeks after they entered the trial to see how they were doing with regard to adherence and safety in the use of their medications (hormones or placebo). After that, we contact them semiannually to check on adherence, safety, and outcomes. Annual protocol visits are scheduled for the women in the hormone trials. For women who are/were having problems adhering to use of the medication, we offer individual adherence programs in which we develop individualized approaches for these women. No matter which population, it is important to direct recruitment and followup strategies to the individual.

About half of the women were 60 to 69 years old, and 23 percent were 70 to 79 years old at trial entry, and we purposely recruited by age cohort. With regard to race, 80 percent of our women are Caucasian, 10

percent are African American, 6 percent are Hispanic, and 4 percent are Native American or Alaska Native, Asian American, or of undeclared ethnic background.

With regard to education, this is a relatively well-educated group of women: 71 percent had more than a high school education. These women also are of a relatively high level of income. In our recruiting, it was important (because of the type of study) to make sure that the women would adhere to the treatment; we were testing the use of the hormones, so we had to screen out, for example, women who were substance abusers or women with a strong history of depression or anxiety.

In the WHI, we have retention definitions that include a set of categories that define retention. However, retention definitions are not standardized across different studies. Definitions in the WHI include the women who come in for visits who are considered fully retained. Some women refuse followup for 1 year but may be open to followup

the next year, so we continue to contact them each year. Some women refuse followup completely and do not respond to contact. Some women are lost to followup because such participants may have moved without providing a forwarding address. Despite difficulties, we continue to pursue all of these individuals because it is important to obtain as much followup data as possible.

HT Retention Summary

Degrees of Followup	National Average	Worcester	Minority Centers
Visits	88%	91%	80%
Visit, Phone, Mail	93%	97%	87.5%
No Followup	1.3%	1.0%	2.4%
Absolutely No Followup	1.8%	0.2%	2.6%
Lost to Followup	0.8%	0%	2.1%

*August 31, 2002

For the national average, 93 percent of the participants were retained for contact, either for a visit or telephone or mail followup; only 2 percent said absolutely no followup, and less than 1 percent were lost to followup. Ten centers in the WHI focused on recruiting women from minority groups or women of color. In that group, 87 percent had follow-up contact, either by a visit or telephone or mail followup. Therefore the rates are not that different across those 10 sites compared with the national average. However, there is a wide distribution among those 10 sites: some sites are more successful in followup no matter what the makeup of their site. The distribution of retention rates among sites is between 70 percent and about 95 percent. Some of the difference may be due to the different methods sites used to retain people in trials.

Retention encompasses a host of activities on the part of researchers and their staff members. Above all, maintain that listening ability with participants and be aware that each participant is a unique individual.

For my site at the University of Massachusetts Medical School, we analyzed the retention for women in the estrogen trial compared with the estrogen plus progestin trial. The retention rates were the same for full followup in both trials. No significant difference was seen between retention of women at different income levels or by age, whether or not women had full followup or other than full followup. The significant difference was that in the 60- to 64-year-old group they were much more likely not to have full followup. I am unsure about the explanation of that difference; perhaps many of these women are still working or have some responsibilities for grandchildren and parents.

Whom did we lose? We discuss this question in our monthly case conferences to understand whom we lose from the study. They are usually women who either do not feel connected to the staff, do not feel connected to the study, or have special difficult life events. Pill taking may have been an issue for them, which is an adherence problem. Retention problems in randomized clinical trials often start with adherence problems, with the next step being that they do not come back; so an adherence problem is a red flag. A participant who is having difficulty adhering to the protocol is probably someone with whom the staff should work more closely. Adherence problems can be a slippery slope down to retention problems.

Some participants present special challenges. They may be having some problems and, especially as people (not just women) age, there are special problems of which we need to be aware while working with them. They may have transportation problems. They may have cognitive difficulties and therefore may need more time to complete tasks. Depression might occur, as might grief because of losses. Physical problems also manifest themselves. We need to be aware of each of these conditions, and we need to address them, particularly in an older population.

A number of approaches are key to retention of women in the WHI. We send postcard reminders, we alter doses when necessary, and we provide incentives. It is important to identify retention challenges early through close monitoring. To do this, we generate reports that help us track missed contacts, low adherence, and stops. Large clinical trials use report forms, which are important because they cue us to be aware that, for example, one person may have a medication problem or that we need to develop a special plan for someone. Being aware of problems motivates us to solve them.

Another critical approach is to have well-trained, skilled staff members. It is not simply a matter of being with the participant physically, but it is also knowing how to interact with her and how to listen to what she is saying. It takes training and practice to learn how to listen and problem-solve.

Retention encompasses a host of activities on the part of researchers and their staff members. Listen to the problems the participants voice, be aware of them, problem-solve with them, help them determine the relevance of participation to them, reinforce the contributions that they have made to the study, offer assistance if needed, and maintain contact. Above all, maintain that listening ability with participants and be aware that each participant is a unique individual.

Essentials of Retention: Studies of HIV-Infected Women

Ruth M. Greenblatt, M.D.

Dr. Greenblatt is Professor of Medicine and Epidemiology and Biostatistics at the University of California, San Francisco. She directs a multicenter primary care program for women living with HIV/AIDS.

The Women's Interagency HIV Study (WIHS) is conducted in six sites and has a data center in the United States. The study is sponsored by a number of NIH Institutes and Centers, including the National Institute on Drug Abuse, and we are fortunate to have the resources to do the types of retention work that we have done.

This study represents the largest cohort in the United States of women living with human immunodeficiency virus (HIV), and it is demographically and behaviorally diverse; the study also includes a similar uninfected HIV comparison group. The original cohort was recruited in 1993 and 1994, and additional recruitment occurred in 2001 and 2002. This observational cohort includes detailed and mandatory self-report data, a physical examination, and specimen collection. The table to the right lists the types of data collected at semiannual visits.

Part of the examination is a full-body habitus examination to identify the potential effects of HIV or its treatment on body fat distribution and bone density. Our repository now contains about 2 million specimens, including blood, white cells, urine, cervical vaginal fluid, and saliva.

The cohort is quite diverse. The following table shows the baseline data of the participants for the first recruited group. The more recently recruited portion was younger with a median age of around 30. There is a fairly high rate of intravenous drug users, women who have either multiple sexual partners or a known HIV-positive sexual partner, and, as is true in the general population of women with HIV, a relatively high rate of women who have no particular risk factor.

Relatively few of these women have ever been employed outside the home. The median per capita income in the group is very low, around \$5,000 a year.

A large number of the women who are HIV-negative have no health insurance; women with HIV infection often qualify for State health insurance through Medicaid. A history of physical and sexual abuse is common in both groups of women, as are symptoms of depression. Median CD4 cell count in the HIV-infected women at enrollment

Semi-Annual Visit

Interview Administered Questionnaires:

- Behavior
- Health Services
- Medical abd OB/GYN History
- Demographics/Psychosocial

Physical and Gynecological Examination (colposcopy, dental examination)

Lipodystrophy exam

(body measures, skin folds, BIA, DXA, fasting glucose and lipids)

Medications

(antiretroviral and other)

Participant Samples:

- Blood (Virologic, Immunologic, fasting lipodystrophy markers, etc.)
- Other (CVL, Urine, Saliva, cervical and anal cytology)
- 12-24 hour pharmacokinetic protocols
- Cervical tissue (lesional and normal)
- Oral tissue

was much lower than in the HIV-negative women, who of course had no detectable HIV ribonucleic acid (RNA).

The original cohort included 2,070 HIV-positive women and 550 women who were HIV-negative; in October 2001 the study included more than 9,000 person-years among the positives and almost 3,000 among the negatives. We have continued to follow the cohort, so we now have another year of followup.

We published a paper on retention in this cohort (Nancy Hessel et al.*). We considered a subject “lost” if she made no study visit in the past 12 months, but not if her death was verified. We considered women “found” if they were previously considered lost but returned for follow-up, and we did not include in this analysis those women who were serial converters—women who began as HIV negatives and became positive (of which there have only been 11 in the cohort to date). The following graphs show the retention rates, which are relatively high especially if you consider the median income, the

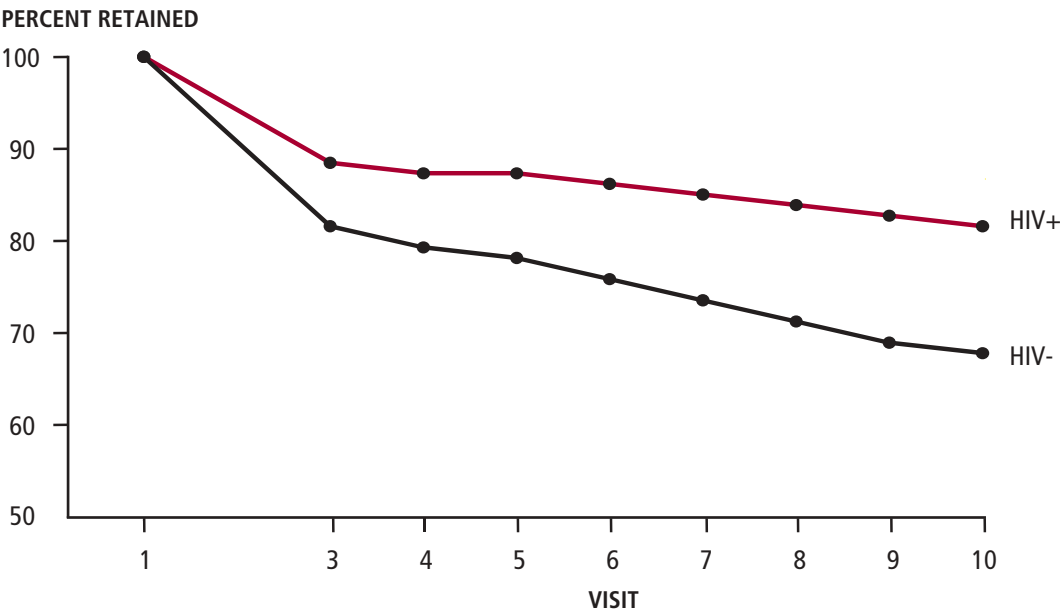
Baseline Characteristics

(Barkan, Melnick, Preston-Martin, et al.,
Epidemiology 1998; 9:117-125)

	HIV+	HIV-
Median age	36	34
Race/ethnicity:		
African-American	56%	54%
Latina	23%	28%
Exposure Category:		
Intravenous drug use	34%	28%
Heterosexual risk	42%	26%
Transfusion risk	4%	3%
No identified risk	20%	43%

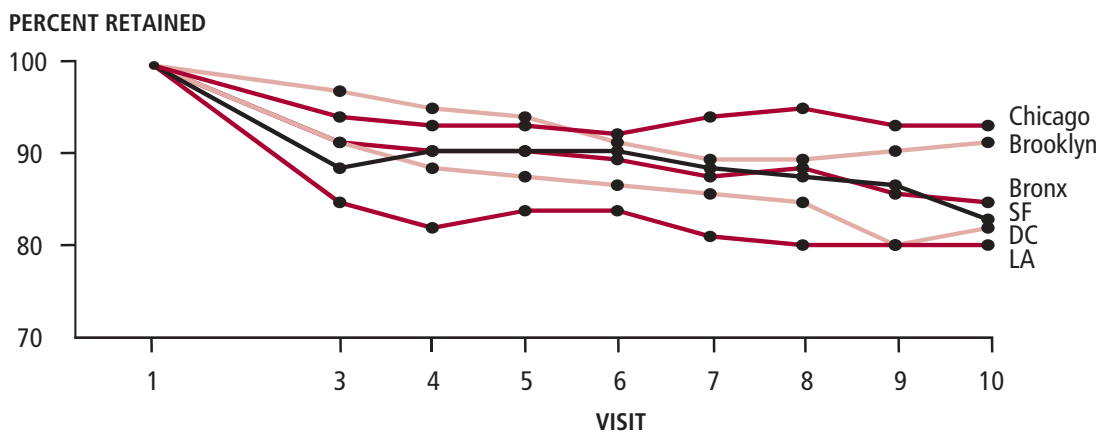
Longitudinal Retention Rates

HIV+	100	90	89	89	88	87	86	85	84
HIV-	100	84	81	81	79	77	75	73	72



Longitudinal Retention Rates for HIV+ (Women by Site)

Bronx	100	91	90	90	89	87	88	85	84
Brooklyn	100	97	95	94	91	89	89	90	91
DC	100	91	88	87	86	85	84	79	81
LA	100	84	81	83	83	80	79	79	79
SF	100	88	90	90	90	88	87	86	82
Chicago	100	94	93	93	92	94	95	93	93



number of injection drug users, and that some of these women have been in and out of jail. We did not recruit at jails, but if the jail will let them come to a study visit, we do follow women who are incarcerated.

Our retention rate is slightly higher for the HIV-positive women—more than 80 percent—than for the HIV-negative women, with some variation in retention at the different sites. Some sites recruited mostly women who were receiving care in clinics at the site, and some recruited women off the street. In our San Francisco cohort, for example, we have a lot of homeless women, but the relative retention rate at 10 years is nonetheless high, close to 80 percent.

Women are the fastest growing demographic group in the United States with HIV infection. They are quite different from many of the men, particularly men who have been enrolled in clinical research. These women tend to be poor and members of minority groups, have relatively little social support, are more likely to use drugs or live in a drug use environment, have dependents, be substantially less well educated than, for example, the

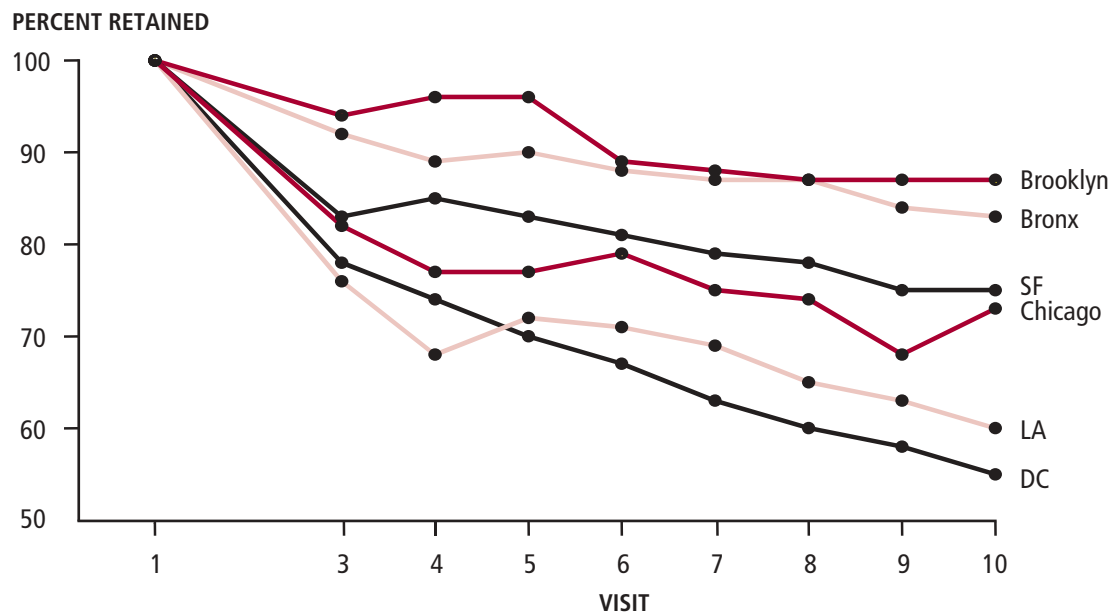
gay men in San Francisco, are less likely to be employed, and in general are much less informed about HIV; many of these women fear ostracism.

What did we do to achieve this retention rate? The first thing we did to help us retain this population was to be responsive to participants' needs. We have local and national advisory boards, and we provide compensation for costs and inconvenience. In general, our approach is not to do something extra burdensome but to support people appropriately for the effort they are making by participating in our research. To some degree, clinical researchers have gotten away with not being supportive enough of participants in their studies.

We support participants' travel by giving them tokens for the MUNI system, for example. If a visit is long and occurs over a meal, we provide food; we provide little snack crackers at all our sites. We reimburse for time, but not excessively. At least at our site, we provide an on-time incentive: When people are late, the timing of our operation is disturbed, so we give them a Wendy's burger coupon if they show up before their scheduled appointment or within

Longitudinal Retention Rates for HIV- (Women by Site)

Bronx	100	92	90	91	88	88	87	85	83
Brooklyn	100	94	96	96	89	88	87	87	87
DC	100	78	73	69	66	62	60	57	55
LA	100	76	68	72	71	69	65	63	60
SF	100	83	85	83	81	79	78	76	76
Chicago	100	83	76	77	79	75	75	68	72



15 minutes of it. We also give out little prizes if they complete five visits in a row. We give “red carpet” service to our participants, including we give out birthday and seasonal cards, with such goodies as movie passes or museum passes inserted. For some of our participants, we are the only ones sending them greeting cards.

We all have site-based newsletters written in simple language describing big issues in HIV or issues of interest to women and sometimes introducing the study staff members. We always have a personal and respectful atmosphere: The study staff has time to talk with participants, and increasingly over the years, participants call when they have conflicts and crises in their lives because they see the study staff as a resource and as people who will listen.

We report all results to our participants and their providers, if that is what the participants want. It is my duty as a researcher to be able to explain to participants what we are doing. Even if the assay does not work or results are not interpretable, I believe I should be able to explain that information. I have a problem with the Clinical Laboratory Improvement Amendments (CLIA) regulations. This will be the first time we are being prohibited from doing what our community advisory board wants us to do.

We actively track our participants, and we obtain their contact information, including backup contacts. We ask for privacy preferences—how we should contact them and whether we can leave messages. We regularly mail our newsletter, which allows us to collect change-of-

address notifications. We have permission to use existing databases and we have permission to talk to primary providers. We go out in the field to find people—checking jails, homeless facilities, and parks—and we do a lot of reminder calls. We participate in the community: We have partnered with community-based organizations, sponsored events, and given out food at our community advisory board meetings. People leave with bags full of food, a gesture that is both important and meaningful with this group of participants. The overall retention rate in our study is 82 percent; we have made efforts to boost our retention rate, and we feel that those efforts have been rewarded.

Recruitment is the challenge. Data presented in a paper at the Barcelona AIDS meeting on 49 randomized controlled trials of antiretroviral therapy showed that

women were in only 12.25 percent of the trials and that no outcome analyses were conducted by sex.

We have also looked at some of the eligibility criteria for U.S. trials of antiretrovirals, which tend to be fairly extensive, and we applied these within our cohort. We looked at 1,280 women who were on antiretroviral therapy at WIHS visit 14. We did not apply any of the trial's antiretroviral use or failure criteria; we just looked at things such as liver function tests and other medications; we did not have glycosuria or proteinuria data available. Nevertheless, 81 percent of the WIHS participants were not even eligible for participation in these studies. Those studies do not reflect these women, and I do not know how to take the findings of those studies and use them in my clinic, when women represented 9 percent of this one trial.

The Women's Interagency HIV Study represents the largest cohort in the United States of women living with HIV, and it is demographically and behaviorally diverse.

Retention Strategies: Retention of African Americans in Efficacy and Effectiveness Trials

Shiriki K. Kumanyika, Ph.D., M.P.H.

Dr. Kumanyika is Professor of Epidemiology at the University of Pennsylvania. Her research focuses on the epidemiology and management of obesity in African American women.

Lifestyle studies are qualitatively different from the kinds of studies in which participants take pills. Both types of studies are critical for risk reduction and management of chronic disease. In a lifestyle study, the intervention has been defined generally, but the participants have to fit the behavioral change into their lives. This is almost the equivalent of sending a participant into the lab to figure out how to formulate the pill. This puts a lot of burden on the individual participant. A second problem in lifestyle studies is the relevance gap. Even if a lifestyle trial has no missing data, participants all adhere to the diet, and the result is beautiful—for example, the participants all lose 15 pounds and the weight stays off—that trial might have no relationship to reality. You cannot necessarily replicate the behaviors and the results of a lifestyle trial in a routine clinical setting or in the community in the same way that you can disseminate drug therapy.

The efficacy trial for lifestyle is one in which you create an artificial scenario for lifestyle change (e.g., weight, sodium, and fat reduction), with fairly expensive measures, both to intervene and to maximize retention. The staff members are usually highly paid and well-trained, especially in techniques such as behavioral counseling and other aspects in which the average dietitian might not be trained. A variety of measures maintain participation—postcards, birthday cards, extra staff members, and telephone calls—almost everything possible to keep the people in the study, so you do not have any missing data and you do not have to rely on modeling to estimate results in the absence of data for the complete set of participants.

The effectiveness trial, on the other hand, is a combination of feasibility and efficacy. In these trials, you are interested in the outcome—whether you helped a

particular patient population or a particular group of people. Efficacy trial investigators will proudly tell you that efficacy trials do not need to have external validity as long as they have internal validity and a balance between the randomization arms; the scientific rigor is the standard for that kind of trial. For effectiveness, although the trial should be rigorous, the key issue is whether you helped the people in the population, not just whether you did a good study.

There is a tradeoff between relevance and rigor. A false assumption applied to lifestyle efficacy trials means that, if you find something that is efficacious, you can then figure out later how to make it work in the real world. That probably is a reasonable approach for studies in which the intervention involves pill-taking, because if you know that the pill works in the human body, later on you can give that pill to people. But this does not necessarily work with lifestyle interventions, because what worked is not necessarily applicable to other individuals in the real world.

For lifestyle studies, I use three definitions for retention. One is attendance at the final visit, no matter what happened during the rest of the study. The second is attendance at the interim data collection visit, which is critical because more than two data points are needed if repeated measures are used. The third is engagement in the intervention itself—participation. Adherence and retention are related. Once adherence slips, dropouts begin to occur, including a type of “front-end” dropout that probably would have been screened out of the randomized trial. Front-end dropouts are people who enroll but never participate or only participate in the initial day or week of intervention. Some participants are partial attendees; their attendance is sporadic, and adherence is poor. Some are not doing well and so do not

want to show up again, especially for weight control. Maintaining adherence in a weight control study means that you have to be able to personally outrun the obesity epidemic and stick to your diet, which nobody else could do, but you are in this trial and you are supposed to be able to do it. These trials are difficult to do.

For example, two weight loss trials in which I have been involved are the TONE Study (Trials of Non-pharmacologic Intervention in the Elderly)¹ and the HELP Study (Healthy Eating & Lifestyle Program)². The main eligibility criteria for enrollment in TONE were ages 60 to 80 years and blood pressure well-controlled on a single medication. We screened out people who were very obese (body mass index [BMI] greater than 33 for men and greater than 37 for women). There was some geographical diversity, since TONE participants were recruited in Maryland, New Jersey, North Carolina, and Tennessee. We created a special study clinic that was independent of any type of usual-care scenario. In overweight participants we tested weight loss with or without sodium reduction versus usual care or sodium reduction alone. The intensive part of the program with weekly contacts lasted 16 weeks, and we continued with some level of intervention through 15 to 30 months of followup. The weight arms of the trial enrolled 585 people, about 50 percent men and 50 percent women and about 25 percent African Americans.

Attendance and retention were good in the TONE study, better than in most community-based studies. Attendance at followup was 90 percent to 95 percent for African Americans and 82 percent to 93 percent for Caucasians. A total of 95 percent of African Americans and 90 percent of Caucasians completed end-of-study followup. Attendance at initial intervention contacts was lower for African Americans than for Caucasians.

In this type of trial, you can get high retention among the ethnic minority participants because of selectivity on the front end. In a paper I wrote with Roberta Ness and J.A. Grisso, we point out that, to attain high retention, you might have selected differently from the African American population than from the Caucasian population. In other words, the screening procedures for trials may select African Americans whose adherence characteristics are similar to those of Caucasian participants who enroll in the trial. However, this subset of African Americans may be relatively less representative on adherence characteristics than African Americans in general.

We recently completed the HELP study at the University of Pennsylvania. The study was based in a family medicine practice setting. The eligibility criteria

included being African American, ages 25-74, BMI 30 to 50 with no contraindications to weight loss, and having a University of Pennsylvania Health System physician. Study objectives were weight loss through nutrition and physical activity. HELP was offered through a special clinic attached to the Department of Family Medicine. Unlike TONE, in which men and women with a BMI over 33 and 37, respectively, were not eligible for the study, HELP enrolled a more obese and more representative study population (e.g., 15 percent of African American women have a BMI over 40). The only exclusions related to unstable health status. If the person had had a heart attack within the past 6 months, we would not enroll them, but no other hurdles were put in the way of getting into the study. When it was time to analyze effectiveness, we wanted HELP to look just like a treatment program in a family medicine department.

You cannot necessarily replicate the behaviors and the results of a lifestyle trial in a routine clinical setting or in the community in the same way that you can disseminate drug therapy.

We offered weight loss in the initial program for 10 weeks to everybody, which of course helped recruitment because everybody got the first treatment. At the first, 3-month followup, we offered an additional 10 to 18 months of enrollment in randomized phase involving various approaches to longer term weight loss or maintenance. The first 10 weeks were like a long “run-in” period in that sense; those not interested dropped out during this period. After the initial period, we wanted to retain about 150 participants to randomize. We retained and randomized 128 of the original 237.

Of the 237 initially enrolled participants, 54 never attended any of the nutrition and weight counseling classes. These 54 people went through the steps to enroll (i.e., came to the measurement visit, filled out all the questionnaires, and signed consent forms), but they never showed up for class. It is unclear why they came for screening and did not return. People can have their weight taken anywhere, so they clearly did not come in just for the measurements. A “shopper group” of another 16 dropouts attended the first class but never returned to classes after that—also something of an enigma.

For those who came to more than one class (i.e., who participated in the program), the average attendance of 6 out of the 10 classes is much lower than occurs in

efficacy trials; about 50 percent of the people who were expected at a given class showed up. There were a lot of concrete reasons, aside from the usual ones, given by people who did not come to classes. These were sometimes explained in a long voice mail message, for example, someone had died, someone had gone to jail, or the participant was spending the night in the hospital with her/his mother. As you can see, the whole concept of running a clinical trial, where you think that people are getting your information sequentially over the 10 weeks, is disrupted. Instead, there are big holes in what they are learning in this type of setting. At the end of the longer term phase, compared with the initial enrollment in the study, we retained 35 percent to 50 percent of people (depending on whether you count those people who came to only one class). However, after the first 10 weeks of nonrandomized intervention, retention was 67 percent, that is, 89 of the 128 who enrolled in phase 2 came to the final visit.

We would really like to know how to identify which people who come for the measurements or lab tests and the questionnaires will never come back, as that is the group we seem unable to serve. Those who did come to the program were older, more likely to have a professional versus some other type of occupation, and had more education. They were also less likely to have a history of respiratory problems and were more likely to get their laboratory tests done. We also noticed a marginally higher prior experience with weight control programs. Of interest, these factors differed between dropouts and nondropouts in the first phase but were not different among the completers and noncompleters of the randomized second phase. Differential dropout only affected the first part of the study.

It was directed at fat and sodium reduction, and we recruited participants with elevated blood pressure or cholesterol from local supermarkets by doing onsite screenings. We recruited 330 participants and kept 77 percent over a 1-year followup. It was a low-burden intervention, so there was not a significant adherence issue. There is some influence of the participation burden

in the trial on adherence and retention. In trials with a high participant burden, some people come to realize all the things they have to do to comply and back out of the trial if it demands too much from them.

Lower retention is associated with more permissive eligibility criteria, and investigators who want to get their grants refunded have figured that out. As a result, we are being more selective. However, to the extent that we keep selecting into our studies people who are not like the populations we want to serve, these participation requirements become important and legitimate concerns. Participant burden and the match of the intervention to the participant's needs and capabilities (e.g., literacy level) are also important.

Regarding perceived participant benefits, lifestyle trials are not usually funded to offer significant benefits outside of the treatment (which may be viewed as a free service), so we have to scrape up money for the birthday cards and some of the other rewards in these studies; in the observational studies, the funding agency is more likely to allow paying for incentives because there is no obvious benefit for the participants.

In conclusion, initial selectivity may be the major factor affecting retention. However, even if we want no missing data, we should not work toward that goal at the expense of only taking in people we know are going to come back. We have to figure out how to be more creative in designing studies that mimic natural scenarios. Weight reduction studies may be a special case even within lifestyle studies because of the unique behavioral demands of weight reduction in the current environment. I pose two questions for consideration:

1. If we are too successful at retention (e.g., keeping people in the study who under ordinary circumstances would drop out and not return), are we making the situation so artificial that we cannot generalize?
2. What is the optimum level of selectivity to obtain an answer to the question of interest while at the same time not being so selective that your answer is meaningless?

**We have to figure out how to be more creative
in designing studies that mimic natural scenarios.**

Concurrent Expert Panels

Panel I: Mistrust, Skepticism, and Bias in Recruitment and Retention: Reality and the Five Core Elements of Outreach From the ORWH Outreach Network

Moderator: J. Taylor Harden, Ph.D., R.N., FAAN

This panel focused on challenges in the recruitment and retention of nontraditional research participants in clinical studies in the United States. Emphasized were solutions to these challenges utilized by investigators working in a variety of communities and the completion of a revised ORWH Outreach Notebook to facilitate community-based investigators' recruitment and retention of study participants and understanding of their responsibilities to participants, including under Federal law.

MISTRUST, SKEPTICISM, AND BIAS: SCIENCE MEETS REALITY

Keith Whitfield, Ph.D., Department of Biobehavioral Health, The Pennsylvania State University

Issues in Recruiting Participants

Dr. Whitfield began by commenting on the perplexity researchers experience when minority individuals do not enter clinical studies and explained the necessity of considering the history of the community.

Relevance

The most significant issue has been a lack of investment in the individuals involved in the study and in the community itself. Researchers need to understand that participants are an integral part of their work. Health research is usually conducted in either populations or in communities as a matter of study design, and it is important to do everything possible to recruit people who reflect the particular set of demographics needed for the study. Relevance of the study to the population is important. Researchers may want to study hypertension among African Americans, but for a given individual or community, hypertension may not be of great importance, even if it is a grave health matter

Perceived Risks

Macrolevel issues are an important consideration, such as the economics of the area and the community's perceived risk of harm. For example, individuals might

believe, erroneously, that the information will be given to the government and that it would compromise their survival—economically, or in terms of their living arrangements and their welfare status.

Previous Experience

Previous treatment by the medical community might also be an issue. In a Baltimore study, some participants said immediately, "You're not coming here to give us drugs, are you? We don't need drugs. We need treatment and to talk about our issues."

The elderly population is changing as a cohort, not just in population size but also in issues that arise from personal histories. This change will influence what elderly individuals are willing and not willing to do as subjects in research. For example, many older African Americans were well aware of what had happened in the U.S. Public Health Service study of syphilis in Tuskegee, Alabama, before the President apologized. The cohort that experienced desegregation and the civil rights movement have their own issues. Now there are also cohorts well aware of the deaths of subjects in clinical trials and they do not understand about predisposing conditions but, rather, see the event only as reported on the evening news.

Access to the Community

Researchers should take into consideration gatekeepers to participants, including wives and daughters in caregiving situations and grandparents who have become

custodians for their grandchildren. Cutting out the gatekeepers usually results in no participation.

Minority participation will almost always be increased if minority interviewers conduct the interviews, particularly when language, immigration, or shared ethnicity is an issue. In circumstances of immigration, interviewers need to be sensitive to which topics should not be discussed, such as family members who were left behind. In telephone interviews, African American interviewees have been known to ask, “Are you black?”, and they have indicated that their willingness to participate further depended on their perceptions of shared ethnicity.

Another issue is that researchers come into a community and then do not share what they find and what they know. Community support is necessary, and networks of information sharing should be tapped into. Sometimes investigators assume that, when they go into an African American community, the church will be their primary network for information sharing, although going through the church exclusively can incur risks. For example, churchgoers talk among themselves and within the community, and if something happens that they do not like, it may change rates of participation.

Issues in Retaining Participants in the Study

The Setting for the Study

Although for clinical trials controlled lab settings might be preferred for a number of reasons, investigators should look into sites other than “the big ivory towers” to make participation in the study more appealing to the priority populations. Churches, senior centers, other local community sites, or individuals’ homes might serve well.

Staying in Touch

Retention and tracking in the context of longitudinal studies can be particular challenges. Dr. Whitfield’s study staff continues interacting with the community and with individuals; for example, participants are sent birthday cards, and staff members go to church. Collecting information about a participant’s family members and friends can help in followup tracking of a participant when, for example, that person has moved.

Knowing what the community is thinking is important. Community gossip should be heeded. In his study, community liaisons can tap into the gossip circuit. Gossip can affect recruitment either negatively or positively. Subscribing to local newspapers, including those in different languages, is a useful tool in understanding the climate of the community and what the community thinks about research issues.

Experiences From the CAATSA Project

In the Carolina African American Twin Study of Aging (CAATSA) funded by the National Institute on Aging, the challenge was to recruit two participants (the original participant plus that person’s twin). To find potential participants, the CAATSA project began by looking at birth records at a time when restrictions on access to personal information was at its height. Drivers’ licenses could not be used in North Carolina, the primary State for the study. Credit reports could be used, but requests for a credit report can sometimes trigger a problem and prevent people from considering participation.

To date, 710 interviews have been conducted in the CAATSA—285 of twin pairs and the majority of singletons and siblings 25 to 92 years of age recruited by phone and then interviewed at home. The rejection rate has been only 16 percent. Sixteen health variables were involved in interviews that usually lasted 2_ hours. Some interviews lasted as long as 7 hours because the interviewers talk about a topic if the interviewee asks about it. Some topics are not about the scientific protocol. They sent the study staff Christmas cards. Staff members gave them mugs; if one breaks, they send another.

Recommendations

In clinical trial research, investigators try to avoid connections with the participants because of concerns about bias or contamination of the work, but that thinking should be modified. Details should be provided to participants and the study should be discussed with family members. In the CAATSA, before consent, participants and gatekeepers/caregivers wanted to see descriptions of the project in writing, so that was provided, in eighth-grade language. Someone with the participants’ characteristics should describe the study protocol to them.

A community advisory board should be formed. Members of the board members should be considered an interactive part of the research and asked whether there are issues that they think should be addressed and who in the community can address the issues. People who are really invested in the community, such as local doctors and local council members known for their interest in health issues, should be considered and taken into account in study protocols.

Researchers should not merely say that they have a personal investment in the community; they should also show it. Investigators should ask to serve on community boards. Over time and before recruitment, staff members should attend church services, attend health fairs, and ask community contacts what they need. After recruitment,

they should continue to show personal investment, for example, by printing a newsletter about the study with information about what the study found as well as health promotion information.

Researchers should develop the study design and recruitment plans with adequate time periods. For example, startup should include time to connect with the community through specific activities, and recruitment plans should include enough interviewers to spend more intensive time on recruitment. Recruitment plans should include trying to find out more about who has not agreed to participate and why and what would make a difference—soft “no’s” might change if the recruiter is willing to be flexible and to get to know potential participants. It is also important to pay subjects for their time.

THE RECRUITMENT AND RETENTION OF NONTRADITIONAL RESEARCH PARTICIPANTS: THE HANDLS PILOT EXPERIENCE

*Michele K. Evans, M.D., Principal Investigator,
Healthy Aging in Neighborhoods of Diversity Across
the Life Span Study (HANDLS), and Deputy Scientific
Director, Intramural Research Program, National
Institutes of Health*

The HANDLS Pilot Study

The HANDLS (Healthy Aging in Neighborhoods of Diversity Across the Life Span Study) is a longitudinal, multidisciplinary study whose focus is minority health, especially the effects of race and socioeconomic status (SES) on risk factors for morbidity and mortality, incidence and progression of preclinical disease, development and persistence of health disparities, and longitudinal health status and health risks.

The scientific goals and objectives of the HANDLS are to investigate:

- Whether or how race and SES influence health disparities independently or as an interaction among several factors (e.g., race, environment, or biologic factors and cultural/lifestyle practices).
 - The influence of SES and race on normal age-related declines in function in an urban population.
 - The influence of SES and race on the incidence and natural history of age-related disease.
 - The existence of early biomarkers of age-related health disparities that may enhance our ability to prevent or ameliorate the severity of these diseases.
- In addition to scientific goals and objectives, the HANDLS’s operational goals and objectives are to:
- Enhance training opportunities (through Morgan State University and the University of Maryland) in epidemiology, aging, and health disparities research for Baltimore area students.
 - Enhance participation by minority investigators and minority institutions in clinical research, with a goal of contributing to research capacities at minority institutions.
 - Develop effective community-based methods of recruiting and retaining minority and socioeconomically diverse participants in clinical research.

A pilot study was conducted to assess the feasibility of field research using a mobile medical research vehicle, assess the logistical requirements for operating a vehicle in the field, and develop community presence and partnership with residents, participants, and neighborhood

institutions. For 14 months, pilot study personnel took the vehicles into the neighborhoods and were able to recruit, draw blood, take histories, and conduct physical exams and cognitive and physical tests. There were 442 participants in this stage of the study, 40 percent of whom were men. The pilot drew a fairly broad age range as a sample of convenience. Participants were only required to be 18 years of age or older and to give informed consent.

The design for the main study includes the following criteria:

- Age 30 or older.
- Fixed cohort with no replacements.
- Classifications at two levels of race (African American and Caucasian), two levels of SES that will be divided based on 120 percent of the poverty level, seven baseline age groups, and men and women.
- Four-way factorial design with 56 cells (30 subjects per cell) with a goal of 1,680 subjects at the end of the study, selected from 12 census tracts in Baltimore to meet race by SES by age distribution.
- Visits at each site for 3 months.
- Eight triennial visits.

The Challenge of Recruitment

When researchers are faced with obstacles to recruitment and retention of nontraditional populations, they sometimes think they should simply settle for doing their study with a group they can handle and then apply those results to nontraditional populations. This approach makes biomedical research a part of the problem of differences and disparities in access to health care because the research is not clinically appropriate and relevant. Researchers must consider the situation of the potential participants in the recruitment tactics.

Barriers to Participation in Clinical Research Studies

Several approaches to overcoming barriers to participation in clinical research can be used:

- Explain clearly and truthfully to each participant what clinical research is and what the study will and will not do.
- If a participant needs medical care but does not have insurance, figure out how to “work the system” to obtain the necessary medical attention for that individual.
- Provide a direct benefit to participants by explaining what the research requires of them but also listen to their needs.
- Address logistical issues such as multiple changes of address and/or telephone number information, lack

of transportation, concerns about safety, and needs for childcare.

- Ensure confidentiality and reassure individuals that the information will remain confidential.
- Address economic concerns about time off from work to participate in the study and the costs of getting to the study site.

Barriers Specific to Longitudinal Studies

Barriers specific to longitudinal studies and recommendations for addressing them include:

- Approach participants through community organizations, such as the “Ten from Ten” recruitment drive that involved local churches, tenant associations, neighborhood civic associations, and the local police.
- Recruit in face-to-face interviews or meetings.
- Hire researchers of the same race and gender as the participants.
- Create a sense of investment in the study.
- Design the study with a nonresearch (including nonmedical) benefit for participants.
- Establish a network for participant referral for nonmedical as well as medical issues.
- Select study staff members based on skill level first, then on diversity (keeping the study staff as diverse as the participant population as possible, for example).
- Make sure that the spokesperson for the study in the community is a well-trained lay community coordinator, not the principal investigator.
- Compensate participants for their time (during the HANDLS, the compensation was \$40/visit, and \$100/visit is planned in the study itself).
- Provide free transportation.
- Be flexible in scheduling.
- Provide a newsletter to the community with study updates, health education information, and features on staff members and participants.
- Devise mechanisms for participant feedback, including bringing study participants to the study center for feedback.
- Be involved in the community as good citizens, including participation in health fairs, street festivals, and other social events.
- Stay in touch with participants through the use of alternative contacts, if necessary, and through frequent contact via phone, mail, birthday cards, and reports.

Community Advisory Boards

Community advisory boards should be created specifically to provide advice and consent, not to rubber-stamp. Typically, such a board will have 6 to 28 members,

meet 2 to 12 times per year, and include members of the clergy, school officials, tenant association members, and neighborhood residents, including homemakers, politicians, and community health professionals. For the HANDLS, the investigators are building a citywide board to ensure that, as the study ranges across Baltimore, it meets the needs of participants. Goals for the board are to help build larger communications networks within local neighborhoods, change perceptions that have inhibited the development of substantive working relationships between community members and scientific researchers in the past, and reach a larger segment of the community for research participation and dissemination of health findings and methods for reducing risk.

Cultural Proficiency

HANDLS is training its staff in cultural proficiency. The curriculum, which is currently mandatory for researchers and investigators but eventually will be mandatory for all clinical investigators, is oriented toward recognition of and appropriate responses to key cultural features that affect clinical research and clinical care. For example, the fact that a researcher is of a minority group does not mean that person can relate to people of all SES levels. At present, the HANDLS curriculum includes three 4-hour interactive sessions involving guest lecturers and contract staff members. It is currently being compiled into a textbook. The curriculum presents important factors that influence the way health care is delivered and clinical research is conducted, with the understanding that cultural factors may ultimately influence our ability to do both well. Specifics include:

- Providing the scientific rationale for inclusion of underserved populations and minority groups in research projects and changing diversity dynamics nationwide.
- Describing and explaining the need for cultural competence and sensitivity among clinical researchers.
- Introducing researchers to cross-cultural communication and discussing the concept of cultural humility, advisory boards, and the legacy of distrust that largely surrounds the intersection of African American culture and clinical research.
- Providing background for investigators to develop a framework for effective community-based research. To date, Columbia University, the University of California, San Francisco, and Rush Medical College have developed cultural proficiency courses for medical students and residents.

REALITY AND THE FIVE CORE ELEMENTS OF OUTREACH FROM THE ORWH OUTREACH NOTEBOOK

Angela Bates, M.B.A., Program Analyst, ORWH, Office of the Director, National Institutes of Health, and J. Taylor Harden

The Outreach Notebook

Ms. Bates presented an overview of the “Revised Outreach Notebook for the Inclusion, Recruitment, and Retention of Women and Minorities in Clinical Research,” “which provides guidance for implementing the requirements of the 1993 Revitalization Act and the revised inclusion policy released in October 2001, which require that:

- The NIH will ensure that women and minorities and their subpopulations are included in all clinical research.
- Phase III clinical trials must include women and minorities and their subpopulations in adequate numbers to allow for valid analysis of differences in intervention effect.
- Cost is not an acceptable justification for exclusion.
- The NIH will initiate programs and support for outreach efforts to recruit and retain women and minorities and their subpopulations as participants in clinical research.

To meet these requirements, investigators are urged to develop appropriate and culturally sensitive outreach programs and activities for the recruitment and retention of the most diverse study populations consistent with the aims of the research.

The original Outreach Notebook addressed a number of issues regarding the recruitment and retention of women and minorities in clinical research and provided references, case studies, and discussions of a few of the ethical issues. A draft revision is now available. The revision retains much of the original information and provides expanded sections on consent, contracts, and the solicitation process. The contents of the revised notebook are organized as follows:

- Section 1: NIH Policies Involving the Inclusion of Women and Minorities: Review of the Inclusion Policy
- Section 2: Research Grants: How Does the Amended Policy Impact the Way NIH Does Business?
- Section 3: Research Contracts: How Does the Amended Policy Impact the Way NIH Does Business?
- Section 4: Recruitment and Retention: Elements of Outreach

- Section 5: Human Subjects Protections and Inclusion Issues
- Section 6 (Expanded Appendices): Glossary and Selected References; NIH Policy on Inclusion; NIH Policy on Reporting Population Data; Target and Enrollment Data Tables; Resources Available on the Internet; 45 CFR 46 Subpart B; Slide Presentation: Sex/Gender and Minority Inclusion in NIH Clinical Research

Highlights of the Notebook

Several important issues and recommendations are addressed in the notebook. The policies on recruitment and retention of women and minorities are being taken seriously; for example, investigators and the NIH must report annually on whether investigators are reaching their targeted numbers. The notebook is designed, therefore, to guide investigators through what they need to know and do to meet these policies. In the revised notebook, there is an emphasis on lessons learned, policies and NIH guidance, and strategies for recruitment and retention in clinical research. The notebook's five key strategies are:

- **Involve the community.** When investigators become involved in the community, they need to be prepared to provide a great deal of information as frequently as possible. This free flow of information helps build a relationship with the community, and such a relationship is a critical part of community involvement. That is when individuals are able to say that they actually know a given investigator or will call on him or her when needed. Mistrust of research and researchers by communities has often been related to whether there was community presence as well as impact.
- **Involve the participants.** The revised notebook provides a number of strategies for involving participants in a clinical research study. It discusses a range of incentives, from bus transportation to payments of \$40 per visit. Grant applicants should budget for sufficient resources for participant compensation; as a participant's level of involvement goes up, reimbursement or at least respect for the participant's time needs to increase. At the same time, too much could be perceived as coercion. Coercion can be avoided if consent is appropriately balanced with incentives, and the whole process is iterative.
- **Staff your team right.** Consider the possible need to have correspondence between staff and subjects and to develop the cultural competency of the team. There is a growing body of literature on cultural competency.

- **Address logistical and financial needs.** In addressing logistical and financial needs, researchers need not only to deal with participant incentives and compensation but also how physically to get subjects from point A to point B. Staff should travel the routes they would have participants travel to ascertain whether there is a need to change the transportation design due to concerns such as personal safety. These are real barriers.
- **Improve communications.** The need to improve communications cannot be emphasized enough. The team must be prepared to make reports back to participating physicians, to the community, to the participants, and to each other. This type of effort, including newsletters, should be a strategic point of planning a study.

Benefits to individuals of participating in clinical research studies include:

- Contribution to research that will eventually reduce the impact of a devastating illness.
- Opportunity to track their cognitive status and receive early warning of any problems that may develop.
- Complete diagnostic evaluation at no cost if needed.
- Participation in an optional "Book Club" that provides free books on topics of interest.
- Possible provision of reimbursement for time and travel expenses.

However, there is debate in the literature about how to utilize the potential of benefits to participants. In some studies, the time participants spend with the study physician may be the most time they have spent with a physician in their entire lives and, while that sells participation, it is important not to oversell or coerce. Nonetheless, at higher income levels, participants understand this benefit, which has the potential to sell the research idea and also to be a real benefit.

Panel II: The Inclusion of Vulnerable Populations with Questionable Capacity To Consent in Clinical Research: Ethical Issues and Approaches

Moderators: Mary C. Blehar, Ph.D., and David Shore, M.D.

This panel discussion focused on issues involving individuals with diagnoses of psychiatric disorders, dementia, or Alzheimer's disease—patients who have presented some of the most perplexing problems to researchers in the area of consent in clinical research.

ETHICAL ISSUES IN PSYCHIATRIC RESEARCH

David Shore, M.D., Georgetown University

General Issues

Dr. Shore began by discussing how when performing a risk-benefit analysis of a study, its predictable risks and discomforts must be weighed against the anticipated benefits to the participant as well as to science and society. The key issue is whether the risks are minimized and reasonable in relation to the anticipated benefits. The foreseeable and predictable (not potential) risks should be balanced against anticipated or reasonably expected (not potential) benefits. The criteria for including risks in a consent document are far broader than the criteria for including benefits. This is intentional, but it is not often properly appreciated.

Potential participants must be aware of what the alternatives are to participating in the research protocol. They need to know whether they can obtain the treatments being offered outside the study and which procedures are for research purposes and not for direct clinical benefit. This knowledge helps minimize what is referred to as the “therapeutic misconception”—the misconception some might have that research necessarily represents treatment. However, because research is voluntary and differs from individualized treatment, one alternative is to not participate in a study, and the principal investigator and the institutional review board (IRB) must consider the possibility that disclosure of the risks and alternatives could result in nonparticipation.

Safeguards and monitoring are becoming increasingly important. Inclusion and exclusion criteria are the first line of protection for ensuring that those who might be particularly endangered by participating in a study are not included. But the landscape has changed, and there is a focus now on inclusion—that is, on not excluding people who might be harmed because the generalizability of the research could be compromised. When including people who might have been previously excluded, it is important to ensure that there are adequate clinical assessments, monitoring, safeguards, and interventions. This has become a special focus in treatment studies in which people in one group do substantially better than those in another group. An example might be the Women's Health Initiative, in which the interim results called into question the appropriateness of continuing the study.

The principles of the Belmont Report have become the cornerstone for the Common Rule, which governs all federally funded research in this country. The first principle, autonomy, refers to voluntariness or choice, as opposed to coercion. People must be able to freely decide whether they wish to participate. The second principle, beneficence, refers to “doing the right thing”—making sure the risk-benefit ratio is the most favorable. It is not just enough to show that the benefits slightly outweigh the risks; the job is to maximize the benefits and minimize the risks so that the ratio is as favorable as possible given the scientific purpose of the study. Justice, the third principle, refers to the fair selection of subjects, as opposed to recruiting vulnerable people who perhaps

cannot say “no,” but who will bear all of the risks and recognize none of the benefits of participation.

Information should be given to potential participants so they can decide whether or not they wish to participate. Informed consent should cover all of the issues already discussed and many more.

Special Issues

Unfortunately, the underlying mechanisms for psychiatric disorders are not understood. Although medical studies can look at clear markers (e.g., blood glucose for diabetes and immune response for vaccines), for psychiatric disorders we can only look at clinical factors, which are considerably harder to measure and are not the type of early indicators of change that are needed.

Psychiatric researchers deal with a potentially vulnerable population. Those who are institutionalized as mentally infirm or described as mentally disabled—in other words, people who are severely impaired by their disorder—are considered to be members of this vulnerable population, and they may be impaired by their condition to a degree that leaves them unable to understand some or all of the elements of informed consent.

Some alleged experts have suggested that people with mental disorders would only agree to participate in a study if it provided the prospect of direct benefit to them. But, there is now some empirical evidence showing that those with mental disorders (even severe ones) tend to participate for the same reasons that other people do. That is, they would like the research to benefit them, but if it does not they would like the research to help others. If they will be paid for their participation, that is fine, too. These are not irrational points on which to make a decision.

Surrogate and proxy consent is an issue because some people, especially those with dementia or psychosis, may be too impaired to give informed consent. The regulations allow a close relative, referred to as the “legally authorized representative,” to consent. The question then arises regarding whether the criteria refer to what is in the best interest of the patient, to what he/she would have wanted if he/she had capacity, or to whom the impaired person depends on to make decisions for him/her.

Problem Designs

In many cases, treatment is delayed because certain assessments and biological tests must be conducted first. There is also the prospect of a “washout,” the stopping of an ineffective treatment to ensure that the study’s outcome is related to the experimental intervention and not to other treatments that the person was already

receiving. Treatment might also be stopped to avoid potential drug interactions. Some argue that delays in treatment need not be lengthy as they have been in the past, and perhaps a more efficient protocol design will minimize the time that people are taken off treatments.

In past clinical trials, participants were randomized to continue an effective medication or switch to a placebo. Symptom worsening and relapse would then be measured. The risks and benefits (and their ratio) in a trial differ considerably from those of a discontinuation study for those who have responded to a particular medication, and we are now seeing a separate consent process prior to the discontinuation phase that informs people of the risks of discontinuation.

Symptom challenges are designed to understand the underlying mechanisms behind clinical symptoms and to find new targets for future treatments. Researchers have used medications or behavioral challenges that temporarily worsen the clinical symptoms of these disorders. Some people believe that this approach—while ethically tricky—may be useful in understanding underlying mechanisms or new treatment approaches. Others argue that, although we have used this design for several decades, we have not learned enough to justify its continued use.

After the 8-week acute trial, participants may have a wonderful new medication that is not available elsewhere, a standard treatment that is very effective, or even a placebo that is effective. Will the participants have to wait until after the study is completed and the blind is broken (usually 2 years) to learn what they were taking and how they responded? Is it really a prospect of direct benefit for people to wait for information that might affect their clinical treatment? The NIH Office for Human Research Protections (OHRP) has commented on this question in recent years. Typically, participants are now being told, at the end of the acute trial and by a clinician that was not part of the study, what they were taking and how they responded. In many cases, those who did well are offered continued treatment (free of charge) for a limited time. The onus is then on patients and their families to follow up on referrals and find treatment independently.

During the past several years, our data and safety monitoring boards and Council of Human Subject Research work group have been formally addressing human subject research ethics issues, including a variety of types of studies that involve discomfort, deception, potential coercion, or questionable risk-benefit ratios.

Placebo Issues

Placebo use has been another area of controversy. Research on mental disorders typically uses a placebo arm to assess whether a treatment is truly effective. Participants on placebo may have their symptoms persist or even worsen, so it is necessary to build into the study ongoing evaluations, rescue medications, safeguards, and comparisons with other groups in the study.

The mechanisms of placebo have been the subject of considerable debate. In research on medical disorders, placebo response rates in the range of 30 percent to 60 percent are not surprising. Expectation is a factor in placebo response; if people think they are getting the active drug and they think the active drug is very effective, they are more likely to respond even if they are on placebo. Some interesting studies describe how the response rate increases as the perceived chance of response increases. Some want to know how to increase the placebo response, and others want to know how to minimize it to show a difference between a placebo and an active treatment.

However, if an active treatment were going to work in 4 to 8 weeks, if it is going to work at all, why would people be kept on placebo for 6 to 12 months? One option currently being considered is making the duration of the comparison that which is needed to show whether there is a difference. People who were on placebo should be informed and should be given the opportunity to be on an active treatment, if that treatment turned out to be generally effective. A number of policy statements have been offered on this issue (the National Depressive and Manic-Depressive Association published a report a year or two ago). The National Institute of Mental Health (NIMH), the U.S. Food and Drug Administration (FDA), and many of the major advocacy organizations agree on the need for scientifically valid, placebo-controlled studies—as long as they are well monitored.

Data and Safety Monitoring

Monitoring is important not just for pharmacological studies but also for psychosocial interventions and combined treatments. The following questions about recruitment and retention should be asked: Are enough people being asked to participate? Are the right people being asked to participate? Are people dropping out of the study, and if so, why? Will there be enough participants at the end of the study to test the hypothesis, or is the study actually futile? If there is no chance that a difference can be shown with results to date, why continue the study?

Problems with human subjects should be reported properly (either to the local institutional review board [IRB], OHRP, FDA, or NIMH, depending on the circumstances). If adverse events suggest more toxicity in one group than another or more toxicity than expected, the study may need to be modified or halted and the participants may need additional information.

When one group seems to be doing much worse than another, it is time to decide whether continuing the study can be justified. The WHI study of last summer is a good example of this kind of decision. Our data and safety monitoring board has halted portions of NIMH studies based on this consideration. The NIMH Guide has some information from June 10, 1998, and June 5, 2000, on data monitoring, and the NIMH policy on data monitoring is posted on the NIMH Web site.

In September 2002, the California Assembly passed a bill that describes who can give consent for individuals who are unable to do so for themselves. A number of groups in this country are available that will accredit an IRB function after a fairly detailed review. The next group to follow the National Bioethics Advisory Commission (NBAC) and the National Human Research Protections Advisory Committee (NHRPAC) will be the Secretary's Advisory Committee of Human Research Protections. Ten people have accepted positions, and Dr. Ernest Prentice of the University of Nebraska, a prominent bioethicist, medical researcher, and IRB cochair, will chair the committee. There is controversy about all of these groups, and the status of NBAC's and NHRPAC's recommendations is unclear, because both groups were allowed to "sunset."

All of the issues discussed are sensitive and challenging, as well as difficult to evaluate. They require thoughtful consideration and expertise both scientifically and bioethically in addition to the perspectives of research participants. To paraphrase the late journalist for the Baltimore Sun, H.L. Mencken: "For every complex, difficult problem, there is a simple, easy solution. Which is always wrong."

DEMENTIA AND CONSENT CAPACITY

Scott Kim, M.D., Ph.D., *University of Rochester School of Medicine*

Facts about Alzheimer's Disease

Although this presentation was designed to focus on Alzheimer's disease (AD), comments were generalizable to other situations in which an individual's ability to give consent is questionable. It is probably not necessary to describe the clinical course of AD, because it is so common, and most people have personal experience with the disease. Currently, about four million Americans have AD, including 10 percent of those older than 65 and 50 percent of those older than 85. If effective treatment or prevention is not available by 2050, 14 million Americans will have it. AD probably afflicts women more than men, a controversial fact, although we do know that there are more women with AD in absolute numbers because women live longer than men. The annual societal costs of AD are over \$100 billion, with AD inevitably leading to complete dependent care, and U.S. businesses lose \$36 billion a year as employees take time off from work to care for relatives with AD. These facts underscore the societal imperative to conduct research on this devastating illness. Federal spending on AD research was approximately \$600 million in 2002.

Informed Consent and AD Research

Because AD will inevitably lead to incapacity if the person survives, two Federal requirements must be met. First, informed consent is required if the subject is competent, and second, if the subject is deemed unable to give consent, a legally authorized representative must provide consent. These seem like relatively simple requirements, but they raise many complex issues, including the following:

- Regarding decisional capacity, who should be assessed, and who should conduct the assessment, and by what standards or methods? In a general hospital, the consulting psychiatrist usually conducts capacity assessments; however, should these psychiatrists make judgments that have legal and ethical implications in a research setting? Interestingly, there are no published papers to indicate that psychiatrists, even among themselves, would agree on a particular case.
- Regarding surrogate permission, who is legally authorized to be a surrogate? There is no simple answer to this question.

A useful question to ask is: how common is decisional incapacity in AD. The answer depends on the definition of "decisional incapacity," an evolving quasi-clinical and quasi-ethical legal term. Some States use the term in their statutes, even though some ethicists believe it should be replaced with a nonclinical term such as "competence." In the context of this discussion, a clinical judgment must be made about whether a person has sufficient decisionmaking abilities to give his or her own consent. The real question is whether this person should be allowed to make his or her own decision. Essentially, the term "decisional capacity" stands for that concept.

Capacity is distinct from the diagnosis of AD, because AD does not necessarily render a person incapable of making certain decisions; and it is specific to context, as the threshold for capacity will vary with the risk-benefit profile of the decision at hand. Although this last statement makes some purist-minded philosophers uncomfortable, the risk-related concept is a fairly accepted doctrine in the "policy circle."

If the commonly cited standards for decisional capacity were applied to a study group of people with mild to early moderate AD, roughly half would be seen as having the capacity to consent to a clinical trial. This underscores the statement that a diagnosis of AD does not render a person incapable of giving consent. To determine sensitivity and specificity, the MMSE score (a crude measure of impairment) can be used; it can be applied to the subjects in a pool who have AD to identify those who are unable to consent or to avoid eliminating those who can give consent.

Another complication is the fact that risk perception in those with increasing impairment tends to be preserved. This clinically corresponds to the observation that AD patients preserve emotional or value-laden issues for a longer time than they preserve new or abstract information. When analyzing consents for procedures of varying risk (e.g., blood draw, drug clinical trial, and PET scan or a hypothetical vignette for brain surgery), no major difference is seen between the consent given by people with AD and by people without—even though approximately 50 percent of those with AD would be considered incapable of giving consent. Categories are imposed because decisions have to be made. However, in clinical reality, the concept is dimensional—not categorical.

Practical Considerations

Because incapacity is so common in AD, a statement should be made about the individual's ability to consent. The formality of the statement depends on the risk-benefit profile of the proposed protocol. For studies of minimal risk, an observational study or a questionnaire containing nonsensitive material might suffice. Studies of higher risk, particularly where there is no anticipated benefit, require greater care. The traditional ways that bioethicists and regulations address risk categories is very crude—for example, minimal risk, greater than minimal risk, categories that have some meaning but that ignore the gradations of risk that occur within categories. It is the obligation of investigators to make their IRBs aware of the options within the risk category and to create a flexible menu of options that can be offered for the proposed research. Examples of options include interviews ranging from informal to structured, evaluations using validated instruments, and postdisclosure “quizzes.”

Legally Authorized Representatives

A legally authorized representative can give consent for a subject who is deemed incapable. Federal regulations defer to States' definitions of legally authorized representatives (45 CFR 46.102c). Unfortunately, most States do not have laws that specifically address this issue. If a research protocol has little direct benefit or significant risk, States' policies become even less clear.

Increasing Concern About Research Ethics

On the one hand, research must be conducted on AD because it is such a devastating disease. On the other hand, our culture is becoming much more sensitive to research ethics issues. The deaths of Jesse Gelsinger and Ellen Roch, the shutting down of prominent institutions, reports on conflicts of interest in clinical research, and the filing of lawsuits have increased public scrutiny. Research ethics is no longer crisis driven, however; it is here to stay as a continuing public policy issue.

The FDA is seeing more complaints against researchers. The following recent events are relevant to the issue of proxy consent:

- The Grimes v. Kennedy Krieger Institute case heard by the Maryland Court of Appeals.
- Actions of specific IRBs (e.g., Mount Sinai).
- Temporary moratoriums on surrogate, consent-based research.
- Letters from the OHRP requiring specific legal justification for enrolling subjects via surrogate consent.

The culture has shifted. It used to be that researchers could rely on the public's faith in them to do the right thing, and now risk managers are involved in the process. California provides a good model in being proactive: Researchers, patient advocacy groups, and others joined together and formed a coalition that was successful in getting legislation passed.

Surveys Related to Consent

The NIH surveyed 246 healthy adults with at least one first-degree relative with AD. These individuals were asked whether they would want to participate in a research study whether they were incapacitated by AD. Eighty-one percent preferred giving advance instructions, and 13 percent preferred having a relative decide. However, almost 90 percent indicated that a relative could consent if no advance directive was given. Eighty percent said a relative could consent to research with potential benefit even if an advance directive against participation was given. These results are consistent with research in clinical settings. Regarding the types of research, 92 percent would take experimental medicine with potential personal benefit, and 80 percent would take experimental medicine that might help others. There are some limitations to this study, however, because the descriptions of what they are consenting to are limited.

What To Do for Now?

It is important to know the risk management situation of your institution, both from State and Federal perspectives. Hospital IRB directors and counsel should be involved to forge a thoughtful policy. The OHRP understands the quandary of investigators and wants to ensure that they do not create intrusive or invalid interpretations. Developing a flexible approach with a menu of options that are proactively thought through and discussed with the relevant IRB is the most promising approach; broad, rigid, risk-category-based requirements dictated by the IRB and armchair regulations not based on clinical realities simply do not serve people well.

In conclusion, it must be recognized that enrolling a person in research without his or her direct consent is an ethically sensitive issue. Researchers must become proactive and involved in the debate, which is still evolving.

RESEARCH WITH SUBJECTS AT RISK FOR IMPAIRED DECISIONMAKING CAPACITY: A VIEW FROM THE NATIONAL INSTITUTE OF MENTAL HEALTH INSTITUTIONAL REVIEW BOARD

Donald Rosenstein, M.D.,
National Institute of Mental Health

Dr. Rosenstein's perspective on the subject comes from his experience as a consultation liaison psychiatrist for a hospital, a member of the NIMH IRB, and the father of an autistic child. The state of the art in treating people with serious neuropsychiatric disorders is minimal. Progress in this area is made difficult because the disorders most in need of empirical data are the ones that are so complicated from practical, procedural, policy, and ethical points of view.

Acute Respiratory Distress Syndrome Network Trial

The Acute Respiratory Distress Syndrome Network (ARDSN) trial provides a cautionary tale that supports the comment that researchers can no longer "put their heads in the sand." This study was reported in the *New England Journal of Medicine* in 2000 and produced an amazing number of ripple effects. This multicenter trial of mechanical ventilation in patients with acute lung injury compared "traditional" tidal volumes to lower tidal volumes. The manuscript states that informed consent was obtained from the subjects or surrogates in all but one hospital where this requirement was waived. However, most of the hospitals were not explicit about who actually gave consent. The primary outcome measure was death, so the stakes were as high as they get.

The OHRP, which issues, suspends, and revokes licenses to conduct human subjects research, found in this case that the IRB process was flawed and that "legally effective informed consent of the subject or the subject's legally authorized representative" was not obtained. In its determination letter, the OHRP stated that the subjects, because of their impaired mental state, and their family members, because of the stress of having a critically ill family member in an intensive care unit, were vulnerable to coercion or undue influence, and that the IRB failed to protect their rights and welfare. However, it is a huge and very dangerous leap to suggest that emotional upset translates into decisional incapacity.

Research With Impaired or Potentially Impaired Subjects

The Federal regulations, 45 CFR 46.11, call for additional safeguards for mentally disabled persons, although there is no definition of "mentally disabled" and

no description of how to assess an individual for mental disability. The following are examples of studies that the NIMH IRB has reviewed that fall into the category of research with people who have potentially impaired decisionmaking capacity:

- Medication trial for AD.
- ECT trial for delusional depression.
- Placebo-controlled study in acute mania.
- Magnetic resonance spectroscopy study of a delirium model.
- Establishing cell lines for genetic studies of mental retardation.
- Tryptophan depletion in autism (adults).
- Medication-free studies of schizophrenia.

Many types of studies and problems need to be addressed, which reinforces the notion that there is no one, simple solution. Obviously, the most difficult situation is when subjects cannot provide informed consent and the study offers no prospect of direct medical benefit and involves more than minimal risk. The NIMH IRBs would ask the following questions about such studies: (1) Can the scientific question be answered with capacitated subjects? (2) What are the relevant risks and benefits? and (3) What is the nature of the anticipated decision-making impairment? It is difficult to develop a single policy approach to answering these questions.

Capacity Assessment

Almost the entire literature on decisionmaking capacity is focused on cognitive aspects: memory, attention, concentration, conceptual organization, psychosis and hallucinations, and executive function. However, decisionmaking is a complex human activity, and there are many other determinants besides how well one understands and processes information. We do not know much about how different people make decisions, but trust, intuition, and insight about interactions between two people strongly influence the consent process. In fact, some people do not even read the consent document; it all comes down to whether or not they trust the person providing the information.

Competence is all-or-nothing, and its determination is usually a response to a specific act. There is a presumption that a person is competent unless a judge says he or she is not. Ultimately, an investigator and the IRB will have to make a "yes" or "no" decision. A model (MCAT) by Paul Appelbaum and colleagues looks at four domains: understanding, appreciation, ability to reason, and the ability to express a choice. Appreciation of the

difference between clinical care (doing what is best for the patient) and clinical research (answering a question) can be elusive, and its lack can result in the “therapeutic misconception.”

The additional safeguards of informed consent and monitoring and independent assessment can bleed together. An IRB should see that a third person monitors the informed consent process. At the most passive level, the monitor would confirm that disclosure took place. At the next level, the monitor would confirm that the participant demonstrated a basic understanding of the research. At the highest level, the monitor would make an assessment of the participant’s appreciation. To do this, the monitor should, at the very least, read the consent documents and know how to conduct a clinical interview.

There are two triggers for capacity assessments: (1) concerns about a class of prospective subjects (because the design calls for “at-risk” subjects or may precipitate loss of decisional capacity) and (2) concerns about an

individual. The capacity assessment should be made by a clinician—someone who knows how to talk to patients—and the decision to enroll an individual in a study should be made after the clinical judgment of capacity is made. Traditionally, the investigators have made these judgments, which gives rise to concerns about conflicts of interest. Some believe that adding too many people to the process can destroy the relationship between the investigator and the subject. The NBAC recommended formal, independent capacity assessment for all greater than minimal risk research. The MCAT is considered by some to be the best assessment tool available. The NIMH uses an intermediate, modified MCAT approach that has worked well.

Regarding advanced directives, the NIMH uses a form that assigns durable power of attorney, provides directives for health care, and asks specific questions about the individual’s willingness to participate in different types of research if capacity is lost.

Panel III: Recruitment and Retention of Children in Clinical Research: Science Meets Reality

Moderator: George Giacoia, M.D.

Attendees expressed multiple reasons for participating in this panel session, including interest in the policy aspect of inclusion of children in clinical trials; the impact of pain and fatigue syndromes in the process of including children; a desire to learn how to design a study for children; curiosity about the vertical transmission of AIDS among children, adolescents, and women; health communications and private-sector patient recruitment and retention for studies involving the pharmaceutical industry and NIH Institutes and Centers; and the session's relevance to a pharmaceutical company study using Fosamax (alendronate) with children.

PERSPECTIVES ON RECRUITMENT AND RETENTION OF CHILDREN IN NATIONAL INSTITUTES OF HEALTH CLINICAL STUDIES

George Giacoia, M.D., Special Expert, NICHD, NIH

Background

Dr. Giacoia described activities of the Pediatric Pharmacology Research Unit (PPRU) Network from 1994 to 1999. The PPRU Network was created and is still supported by the NICHD. The mission of this 5-year segment of the research program was to facilitate and promote pediatric labeling of new drugs or drugs already on the market. In this process, the Network strives to foster cooperative and complementary research efforts among academia, industry, and health professionals. The overall goal of the PPRU Network is the safe and effective use of drugs in children, especially to prove whether drug-labeling studies can be done in pediatrics by a network of pediatric clinical pharmacologists in major academic centers.

The studies were conducted either cooperatively with investigators at other units in the Network, collaboratively with pharmaceutical companies, or independently with other support. Two primary purposes of the studies conducted by the PPRU Network include (1) providing clinical data on new drugs and drugs already on the market that are necessary for U.S. Food and Drug Administration (FDA) approval for use in children and (2) investigating the pharmacology of new molecular entities and biopharmaceuticals for use in children. The Network also serves as a resource for training health professionals in

pediatric pharmacology and clinical trials.

The PPRU has played a major role in the implementation of the Food and Drug Administration Modernization Act (FDAMA). Since the passage of the FDAMA, the FDA has issued 145 requests for pediatric studies. Company sponsors have designed and conducted pediatric studies in response to the requests, resulting in 22 marketed drug products being granted 6 months' additional marketing exclusivity. The PPRU has performed all or part of the clinical trials supporting the exclusivity determination for eight of these drugs. From 1998 to 1999 pediatric labeling was added to six marketed products on the basis of studies conducted under the FDAMA provisions. The PPRU conducted studies to support the labeling changes of four of these six products. Drugs for which the PPRU did studies to support exclusivity extension include Enalapril, Ibuprofen (two products), Metformin, Midazolam, Propofol, Ranitidine, and Tramadol. Drugs for which the PPRU did studies to support pediatric labeling include Ibuprofen (two products), Midazolam, and Ranitidine.

From 1994 to 2000 there have been four active PPRU protocols: (1) a sample of 17 children from 1994 to 1997, (2) a sample of 21 children in 1998, (3) a sample of 54 children in 1999, and (4) a sample of 75 children in 2000.

Obstacles to Recruitment

Recruitment and retention of children in drug trials involve various obstacles. First, an important aspect of these studies is the recognition that children are not miniature adults—they are different. Moreover, the field

of pediatric pharmacology is populated with relatively few investigators compared with similar studies of adults.

Recruitment obstacles also include the fact that disease is less common in pediatrics and the presence of unique methodological and logistical problems. Because of these limitations, there is often danger in extrapolating results from adults to children, and researchers need to protect children from overeager study efforts. Researchers also must consider not only the number of pediatric patients but also how widespread problems are among this population.

Study Population Statistics

From 1994 to 2004 the PPRU Network reported 160,000 inpatient admissions per year, 2,290,700 outpatient visits per year, and 27,600 neonatal intensive care unit (NICU) admissions per year. The current expanded PPRU Network's 13 centers focus on the following activities:

- Performance of drug trials leading to labeling
- Advancing clinical trial methodology in pediatrics, including pharmacokinetics and pharmacodynamics and development and validation of end points, surrogate end points, and biomarkers
- Translational research, including ontogeny of drug-metabolizing enzymes, receptors, transporters, and ion channels

Pediatric Participation in Clinical Trials

The many areas of concern surrounding the participation of children in clinical trials include the need for

- Child-friendly protocols
- Age-appropriate sampling procedures
- Minimization of invasive procedures
- Provision of an appropriate environment
- Appropriate assent and consent
- Strict ethical guidelines
- Emphasis on the nature of the trials as pharmacokinetic studies and not for therapeutic benefit

There is also a need for the appropriate technology and media presentations so that the family understands any inherent risks in participating in the study. The use of healthy children in any such study is also a concern. For example, a child who is healthy between episodes of cystic fibrosis will likely benefit, but if blood pressure is borderline, there may be problems. It would be unethical to use “me-too” drugs—those that belong to a therapeutic group with similar effects; however, pharmaceutical companies sometimes use these to capture the market. There must be ethical guidelines in the use of children in such studies.

Other difficulties include having to place infants in the NICU when parents do not understand what is happening. Additional difficulties may arise with pregnant women, who may be concerned about the conditions of the unborn children vis-à-vis their children who are involved in the study. For example, in England, one study was randomized ahead of time, but participants and their families were not informed of this occurrence.

Integrated Recruitment Plan

The PPRU Network includes a survey of all pediatric diseases, not targeted, disease-specific populations. Elements of the PPRU integrated recruitment plan include the following:

- Targets for recruitment
- Referral sources
- Recruitment strategies
- Recruitment methods
- Barriers to participation
- Study objectives and type and size of the study
- Length of time for recruitment
- Study setting
- Number of study centers
- Location of study site
- Constraints (e.g., time, finances)

Selection of the Target Population

The level of participation and the success of recruitment are often related to the complexity of the study. Also, what works in one study site does not necessarily work in another.

In selecting the target population, investigators carried out an extensive protocol review, including determining the inclusion and exclusion criteria. Researchers also educated themselves about the nature of the population under study, including sociodemographic, cultural, and organizational variables. In addition, they took into account whether individual participants were representative of the target population selected for the study.

Referral sources included private physicians; community-based organizations; hospitals, community clinics, and health centers; religious ministries; educational institutions; disease-specific consumer groups; online medical services; and individuals serving as community liaisons.

Participants were recruited via letters, culturally appropriate flyers and brochures distributed in the community, a well-publicized PPRU Network Web page, advertising in newspapers and other media, solicitations through various organizations and registries, and coalition-building.

Specific administrative considerations in recruitment included hiring an experienced recruitment coordinator, projecting goals and timelines for recruitment, determining the project startup time, monitoring recruitment, and budgeting for the cost of recruitment.

Parents' Rationale

A key element in successful recruitment is parents' motivation to enroll their children in research studies. Parental reasons for enrolling their children include the parents' desire to contribute to medical knowledge, please the physician, and obtain additional attention and referral for medical care of other problems. Parents also are more likely to enroll their children if the children of friends are also in the study. In a survey of the PPRU Network, parents enrolled their children for the following reasons: (1) 72 percent so other children would be helped, (2) 71 percent believed their own child would be helped, (3) 42 percent believed that research is a societal responsibility, and (4) 8 percent felt they had no choice.

With regard to parental satisfaction, Aman and Woldford—in their social validity study to assess the social importance and personal benefit attached to familial involvement in drug research for participating families—found an 88 percent satisfaction percentage; a similar survey by the PPRU Network survey yielded a parental satisfaction rate of 75 percent.

Children's Rationale

In an article published in February 1999 in the *Journal of Clinical Pharmacology and Therapeutics*, Johnson and colleagues found that reasons given by children for participating in clinical trials include helping other children (altruism), 46 percent; response to age-appropriate incentives, 28 percent; "thought it would be fun," 12.3 percent; and wanting a unique experience, 9.6 percent. Of these children, 95 percent said they would participate again.

Children also have firm opinions about what constituted for them the worst aspects of participating in such studies. Johnson and colleagues found that issues of needle placement (26.8 percent), blood procurement (11.3 percent), dietary restrictions (8.5 percent), bad-tasting drug (7 percent), and interference with sleep patterns (7 percent) made study participation unpleasant.

Children often express fear and mistrust when contemplating participation in a clinical research study. Researchers need to lower the fear as much as possible and conduct only those clinical trials that are absolutely necessary. For example, it is detrimental to refer to children as "guinea pigs."

Incentives

Among other inducements, some centers offer compensation for parent travel. One audience member commented that institutional review boards are running into problems with compensation for travel, although a thank-you fee may be acceptable. Another audience member, representing a pharmaceutical company, said that at the end of one project, parents were told they could keep a palm pilot that they used for communicating with the research team during the study.

The location of and accessibility to the site are important to parents and children and to retaining children in drug trials. A comfortable, nonthreatening environment and the consistent presence of familiar staff people, who interact with the same patients during all visits, are also important. The setting should also offer age-appropriate toys and games.

EXPERIENCE OF RECRUITMENT AND RETENTION OF CHILDREN IN CANCER CLINICAL TRIALS

Barry Anderson, M.D., Ph.D., Senior Investigator,
Pediatric Section, Clinical Investigations Branch,
National Cancer Institute, NIH

Background

Dr. Anderson is a pediatric oncologist in the Cancer Therapy Evaluation Program at the National Cancer Institute (NCI). Like other research programs and clinical trials, NCI clinical studies also experience recruitment problems. Successful recruitment is aided when the family and child understand that the child's participation in a clinical trial means being involved in research, not receiving treatment for a newly diagnosed cancer.

Current Cancer Survival

During the past 40 years, there has been substantial improvement in the outcome for children diagnosed with a malignancy, which can be an acutely fatal illness. In the 1960s most children diagnosed with a cancer, such as leukemia, died within weeks or months of being diagnosed. From 1960 to 1996 childhood (0 to 15 years of age) cancer survival rates for all cancers increased, and the survival rate for acute lymphoblastic leukemia (ALL) improved from 3 percent to 85 percent. In the 1990s more than 85 percent of children diagnosed with leukemia were considered cured, which means that no cancer was detected 5 years after treatment; beyond 5 years, children have very limited likelihood of relapsing.

Within subgroups of leukemia, however, the survival rates are even higher. Moreover, for childhood cancer overall, there has been a dramatic decrease in the mortality rate; the number of children who die of cancer has decreased dramatically over the past 40 years. From 1950 to 1998 childhood cancer mortality decreased from approximately 80 per million annually to approximately 25 per million. Now, 70 percent of childhood cancers can be cured, although the cure rate depends on age, type of tumor, stage of the disease, and other criteria.

Childhood cancer is relatively rare. In any year in the United States, the approximate numbers of certain types of cancers diagnosed in persons between the ages of 0 and 19 years are as follows:

- ALL—2,400. ALL and brain tumors are the only two cancers for which there are more than 1,000 cases diagnosed each year in the United States.
- Acute myelocytic leukemia—850.
- Central nervous system tumors—2,200.
- Hodgkin's lymphoma—900. This is a lymphoma of the lymphatic system.

- Non-Hodgkin's lymphoma—800. This is also a lymphoma of the lymphatic system.
- Neuroblastoma—650—is a form of nervous system tumor; it is not a brain tumor but occurs elsewhere in the body.
- Hepatoblastoma—100—is a liver tumor found mostly in babies and so is primarily diagnosed in children younger than 5 years of age.
- Osteosarcoma—400—is a bone cancer.
- Ewing's sarcoma—200—is also a bone cancer.
- Rhabdomyosarcoma—350—is a muscle tumor.
- Wilms' tumor—500—is a kidney tumor.

In all, approximately 12,400 children and adolescents ages 0 through 19 are diagnosed with cancer annually in the United States; of these, approximately 2,500 (20 percent) die each year. Mortality rates of childhood cancer compare favorably with adult cancers. For example, breast cancer is diagnosed in 200,000 (mostly) women annually, with 40,000 dying each year (20 percent mortality). Prostate cancer, diagnosed in 190,000 men annually, claims the lives of 30,000 (16 percent) annually.

NCI's Clinical Trials Cooperative Group Program

The NCI determined that for pediatric oncology to make progress, it would be necessary for researchers to work within the structure of cooperative groups. Although there are breast cancer centers with sufficient numbers of patients to do a Phase III study of randomized patients and obtain statistically significant information, pediatric oncologists work within Oncology Cooperative Groups (OCGs) to enable larger numbers of patients to be studied. The NCI provides funding to pediatric oncologists through its Clinical Trials Cooperative Group Program. These oncologists work in a cooperative fashion within OCGs, specifically the Children's Oncology Group (COG) for pediatric oncology.

The NCI's Clinical Trials Cooperative Group Program is designed to promote and support clinical trials of new cancer treatments, explore methods of cancer prevention and early detection, and study quality-of-life (QOL) issues and rehabilitation during and after treatment. The program involves more than 1,500 institutions that accrue patients to OCG-conducted clinical trials. Thousands of individual investigators participate in these studies. OCGs annually place approximately 20,000 new patients into cancer treatment clinical trials.

The key characteristics of the NCI Clinical Trials Cooperative Group Program are as follows:

- Researchers jointly develop and conduct cancer clinical trials in multi-institutional settings, mostly at academic centers.
- The major emphasis is on definitive studies of treatment using combined modalities (i.e., surgery, radiation, chemotherapy).
- Funding is not linked to a specific trial, nor are the trials designed by the NCI. Rather, the NCI provides funding to the cooperative group, which continually generates new trial designs at academic institutions and cancer treatment centers throughout North America.
- Each OCG is supported by NCI staff involvement to continually generate new trials. In NCI's Clinical Investigations Branch, Dr. Anderson works with the COG (see below) as well as with various disease committees to help them develop new trials and determine the kinds of treatment to be included, but the OCGs must keep that process in play.

Many NCI clinical trials are conducted simultaneously, usually at the Phase III level, with moderately large numbers of patients randomized according to the stage of tumor. Dr. Anderson and other NCI staff members are involved with the trials through a cooperative agreement with the members of the OCG. This national effort has been in place since the late 1950s. NCI staff involvement is provided through a Cooperative Agreement funding mechanism.

Childhood Cancer Clinical Research

National efforts are essential for studying specific childhood cancers because of the limited numbers of children with individual types of cancer. Since the 1950s the NCI has supported a nationwide clinical research program specifically designed to improve the outcome and QOL for children with cancer.

Until January 2001 there were four major pediatric OCGs:

- Children's Cancer Group
- Pediatric Oncology Group
- National Wilms' Tumor Study Group
- Intergroup Rhabdomyosarcoma Study Group

The first two groups were general OCGs; the last two study specific tumors. These four groups are now combined in the COG, comprising more than 240 member institutions—including cancer centers at major

universities and teaching hospitals—throughout the United States and Canada as well as sites in Australia, Switzerland, and the Netherlands. COG members include more than 5,000 cancer researchers dedicated to saving the lives of children with cancer. Whereas general medical oncology is predominantly practiced in the private practice setting, the practice of pediatric oncology occurs mostly in academic centers or in an academic-associated practice.

In all, NCI-supported clinical research for children with cancer includes the following:

- More than 240 COG sites.
- The COG Phase I Consortium—20 sites—conducts Phase I studies only. Because of the higher level of monitoring required for Phase I trials and the pharmacokinetics involved, these sites receive extra money for pharmacology studies and radiographic techniques, for example, to monitor for tumor responses.
- The Pediatric Brain Tumor Consortium—10 sites—includes “superspecialists” in neurosurgery and neuroradiation.
- The Childhood Cancer Survivor Study—24 sites—evaluates long-term child cancer survivors, including the effects of treatment.

There are also PO1 and RO1 research projects with childhood cancer objectives, including New Approaches to Neuroblastoma Therapy (NANT). The NANT is an NCI-funded consortium of universities and children's hospitals to test promising new therapies for neuroblastoma, a common childhood solid cancer. Neuroblastoma is often difficult to treat successfully. The NANT was formed to provide a group of closely collaborating investigators who are linked with laboratory programs developing novel therapies for high-risk neuroblastoma. The NANT conducts clinical trials that test new drugs and new combinations of drugs against high-risk neuroblastoma. Drugs with promising results will then be considered for more extensive national testing.

The NCI also supports an intramural pediatric research program at the Clinical Center on the NIH campus. The Pediatric Oncology Branch (POB) includes basic scientists and clinicians who perform laboratory investigations of childhood cancer biology. Translational research, from the laboratory to the clinic, is also pursued through clinical trials conducted by the POB.

Recruitment

Approximately 50 percent of children who are diagnosed with cancer enter a clinical treatment trial. These children represent only 1 percent of the national cancer incidence, or some 12,000 cases annually compared with more than 1 million adults who are diagnosed with breast, prostate, or other cancers each year.

Children ages 0 to 20 years of age represent 27 percent of patients enrolled in NCI-sponsored clinical trials. In all, about 5,000 children enter treatment trials each year.

Of the 87 active COG trials, 15 are Phase I studies, and 18 are Phase II studies, both of which are conducted among patients who have relapsed following initial therapy; 21 are Phase III studies, comprising randomization studies of a standard therapy versus some modification of the standard therapy; 15 are pilot studies; and 18 are biology studies. The Phase I trials also use drugs that have been tested in animals; most trials using these drugs have been started in adults, but this is the first time they have been given to children. Within months of starting the drug in adults, the drug can be brought into relapse studies of children. Pilot studies of small groups of patients are being done in an effort to combine new therapies for relapsed patients or for patients who have been diagnosed with an advanced tumor and have a poor prognosis.

It is becoming more accepted in pediatric oncology to put a child into a clinical trial; general medical oncology is different, since there is often no cure for many cancers or at best a cure rate of 30 percent to 35 percent—a 2-year survival. Childhood clinical oncology trials always seek to improve outcomes, even on an 85 percent long-term survival rate. These studies sometimes go on for years and require hundreds, sometimes thousands, of patients.

History of Success for Specific Childhood Cancers

Acute Lymphoblastic Leukemia

Pediatric oncology patients who do well in clinical studies have brought the survival curve up as a result of slow, incremental changes. Bleyer and colleagues of the Children's Cancer Group (CCG) (precursor to the COG) tracked the survival of children with ALL from 1968 to 1993. During that period, a total of 12,921 patients were treated in CCG ALL studies; by 1993 the survival rate was approximately 80 percent, with the biggest jump—between 1970 and 1975—due to the realization that the central nervous system required (CNS) antileukemia treatment with craniospinal irradiation, whether or not disease was detected at the time of diagnosis.

Before then, a child could be brought into remission, but almost all patients would have a relapse in the CNS because researchers did not understand that all of the children should be treated, even if the diagnosis did not show leukemia in their spinal fluid. In a related matter, The New York Times (2003) recently reported on the late effects of cancer treatment in children, including often devastating intellectual impairment caused by radiation. As a result, the ALL trial instituted a refinement of CNS irradiation, using an entire subgroup of children who do not require radiation. Earlier, prophylactic chemotherapies were not available that could be introduced into the spinal fluid. Currently, intraspinal injections of chemotherapy and systemic chemotherapy regimens have replaced much of the craniospinal irradiation that was used in the past, with continued good outcome.

Wilms' Tumor

For children with Wilms' tumor of the kidney, survival rates increased from 33 percent in 1960 to 93 percent in 1994. Fewer patients are getting radiation, and the duration of chemotherapy for many children has been decreased from more than a year to just 3 or 4 months, which is important, because most of these children are diagnosed between 2 and 5 years of age.

In pediatric oncology, clinical trials are the primary form of treatment. Ethically, however, NCI researchers must remind families about the nature of clinical trials. For example, the mother of a child who is brought to the doctor with a swollen belly may think the child has not been eating right or is constipated, but the when magnetic resonance imaging (MRI) shows Wilms' tumor, the pediatric oncologist will consider the trials for Wilms' tumor that are currently available as a means to provide the best available treatment strategy for the child's tumor. Seeking treatment through a Phase III clinical trial is good, since patients will receive either new treatment or the best available therapy, including adjusting dosage or adding a new drug. However, parents must understand how enrollment works in a clinical trial. For Wilms' tumor, where the survival rate is high, researchers are now looking at taking away certain treatments to reduce the long-term effects and the toxicity. Third-party payers usually reimburse for treatment costs.

Phase III Trials

Classic Phase III trials compare two treatments for a particular kind of cancer, typically comparing an experimental treatment with a standard treatment to determine whether the experimental treatment produces better survival than the traditional one. In some cases the

objective is to show that a treatment with fewer and less serious side effects is at least as good as the standard treatment. In this case, the treatment with fewer side effects does not have to produce better survival than the standard treatment to be considered superior; it only has to be as good.

Phase III trials are used to compare treatments in common use where there is significant uncertainty or controversy over which is better. Uncertainty or controversy does not preclude the evidence favoring one treatment or the other, although this is not always the case. Almost all modern Phase III trials also compare the QOL associated with various treatments. QOL is assessed using standardized questionnaires that ask patients subjective questions about how they are feeling and functioning. QOL can be helpful in balancing the side effects of treatment with any benefits in objective measures, such as survival.

In Phase III trials, patients are randomized to receive the “best available” therapy or to receive new treatment. New treatments are prioritized by their potential to improve outcome (i.e., increase survival, diminish toxicity). Phase III clinical trials generally require hundreds of participants to identify the superior treatment with acceptable statistical significance. The usual primary end points of Phase III trials are event-free survival and survival rates. Phase III trials are used for most of the common types of childhood cancers. In all pediatric clinical cancer trials, the focus is less on the number of patients recruited than on qualitative elements. The NCI must constantly ensure that pediatric oncologists know that, although clinical trials may offer beneficial therapeutic results, their main purpose is to conduct clinical research.

Parental Experience

NCI sponsorship of COG trials also includes bioethical considerations, recognizing that parents must confront many tasks, emotions, and decisions about treatment, sometimes on the same day as or within days of the diagnosis. A child who is referred by the pediatrician who thinks the child may have the flu may ultimately—after being hospitalized and undergoing blood tests—be diagnosed with Wilms’ tumor. After this happens, parents not only must deal with learning of their child’s diagnosis, prognosis, and treatment options but also must provide emotional support to their child, interact with the treatment team—physician, nurses, radiologists, social workers, and many others—and participate directly and constantly in the child’s treatment. Whereas an adult patient can answer yes or no to a nurse’s question, a 6-

month-old or a 3-year-old with leukemia cannot, so parents must be constantly on the scene to interact with the treatment team. Although parents want and are supposed to protect their child from uncomfortable things, they will have to experience the emotional discomfort of dealing with many details—for example, ensuring that the child does not eat for 6 hours before surgery, getting another MRI, and staying with their child overnight in the hospital.

For a family in this emotionally charged environment, it can be difficult to make a treatment decision when faced with all of the information related to possible participation in a clinical trial. On first hearing about the trial, parents may question why the doctors do not seem to know how to treat their child’s illness. They must be educated very quickly about possible treatments, survival chances, and the investigational nature of a clinical trial. Nevertheless, the pediatric oncologist is obligated to make the family aware of any pediatric oncology trials that are available and all related details so that they can make a truly informed decision.

Although this process is difficult and stressful for the parents, it is quite different from recruiting a child patient through an advertisement, where all the criteria are known ahead of time. There also may be cultural barriers, such as families who do not speak English, which can make the situation even more difficult.

Informed Consent

To consent a family to a clinical trial in this emotionally charged atmosphere requires careful attention to bioethical matters, since decisions in terms of treatment must be made within days, or often within a day, of the child’s diagnosis. For parents and children to consent to trial participation, they must make an informed decision. Giving informed consent includes awareness, understanding, and confidence.

Awareness

Parents must be aware of the child’s treatment options and that the options include participation in a clinical research trial. The pediatric oncologist must carefully explain the differences between treatment options and the option of participating in a clinical trial. Because parents may question whether a certain treatment is optional or what those options mean, researchers should discuss treatment options in detail, including clinical trial participation, with the parents.

Parents must sign many consent forms, including consents for surgery, anesthesia, biopsy, central intravenous line placement in the chest, and lumbar

puncture and bone marrow procedures. Awareness also includes the researcher's recognition of the emotional stress these situations create for parents. Although these procedures are reflected in the various consent forms that the family must sign, they really have no choice with many of these procedures since they are diagnostic tools. When the pediatric oncologist presents these many consent forms along with the clinical trial consent form to the parents to consider, some may sign the trial consent assuming it is one of the procedural consent forms listed above. Pediatric oncologists work to avoid such an unfortunate misunderstanding, and the COG has been studying the informed consent process among families participating in Phase III trials to better inform families about the clinical trial process and the effects of trial participation on their individual child.

Understanding

In addition to parents' awareness of what is optional, they must understand all the details involved in clinical trial participation. What is a clinical trial? What are the treatments? Where do the treatments come from? Parents also need to know that the best available care usually comes from the previous clinical trial, usually a randomized Phase III study that was conducted immediately prior to the current one into which their child may be enrolled. For most NCI Phase III trials, the "standard arm" is available "off study." Educating parents about the child's condition includes explaining that the standard arm is the best proven, currently available therapy and that the Phase III trial will attempt to improve survival and toxicity outcomes. Parents who are leery about the "experiment" need reassurance, and researchers must clearly delineate standard therapy from investigational treatments within the trial. The research team must make it clear that only part of the treatment given may be the experimental treatment and that often the majority of the regimen is the standard treatment.

The oncology team also has a responsibility to explain fully to parents that standard therapy may put their child back in the hospital. Parents also must understand that clinical trials offer no "magic drug" that they could not get for their child elsewhere. They also must understand that the trial is an attempt to improve the outcome—regarding survival or toxicity or both—of the previous trial. They must comprehend that this trial most likely will have an inferior and a superior arm and that their child has a 50-50 chance of getting into either one arm or the other, even though standard therapy is interwoven in both arms; however, in a compressed timeframe, it may be difficult for the pediatric oncologist to get this point

across. For example, if Arm B involves a lower dose of radiation than the previous trial, parents may opt to participate, even though several years later it is discovered that 5 percent of those children relapsed, perhaps due to the decreased radiotherapy dose.

Parents sometimes ask whether the researchers are "experimenting" on their child. In a way, this is true; no matter what the oncologist says, the child is participating in an experiment—all the more reason for the pediatric oncologist to distinguish standard therapy from experimental therapy. Parents must also understand that the child will probably have to undergo more tests—more blood drawn and more x-rays taken—just to monitor the child's disease as part of the study.

In this sort of situation, the oncology team has a real responsibility to watch the kind of "spin" they use to explain all of these details. Treatment side effects are another difficult area; most children treated for pediatric malignancies end up back in the hospital, perhaps deathly ill from just the standard therapy.

Confidence

An informed decision by parents should reflect confidence that their decision is appropriate for their child and their family, that the oncology team supports their decision, and that their child's health is the oncology team's primary concern. The risks and benefits of trial participation should be honestly described to parents by the treatment team. The oncology team must also exhibit cultural and language competence to help instill this feeling of parental confidence in the team.

Relapsed Patients

Although parental motivation for Phase I and II trials is always therapeutic benefit, the Phase I scientific goal is to obtain toxicity and dosage information—for example, the kinds of tumors this treatment has helped—and the Phase II goal is to determine agent activity.

Pediatric Phase I trials may allow parents to maintain hope, but they must also maintain a reality-based outlook. With regard to historical risk, pediatric Phase I trials reveal a 2 percent to 21 percent occurrence of dose-limited toxicity and a 0.6 percent to 2.9 percent occurrence of drug-related toxic death. Historical benefit shows that 6 percent to 10 percent of patients show an objective response to the agents they receive during a Phase I study. However, this does not mean that this percentage of patients will be cured; rather it means that their tumors may stop growing, shrink in size, or even disappear; these results, however, are most likely transient, since the effective drug is being given only for a prescribed

period of time. There are only a few cases in which the benefit is more prolonged, for example, using Gleevec (STI 571) for children who have relapsed to leukemia.

National Cancer Mortality Reduction, 1990-1998

In the NCI Surveillance, Epidemiology, and End Results (SEER) study for the years 1990 to 1998, the greatest reductions in cancer mortality occurred among patients between 0 and 19 years of age, with the highest rates for those ages 0 to 14. Patients between the ages of 15 and 34 had higher mortality. The greatest decrease in cancer mortality—22 percent among those between 0 and 4 years of age—was 3 percent among 20- to 24-year-olds. This gap in improvement over time for teenagers and those in their early twenties correlates with the enrollment of adolescents and young adults in cancer clinical trials. Although parents of young children are highly motivated to have their children participate in a clinical trial as a means of access to state-of-the-art medical treatment, a much smaller percentage of adults diagnosed with cancer—only 2 percent to 5 percent—enter a clinical trial. However, being in a clinical trial is not a benefit in and of itself; it is the clinical science of oncology treatment that has been potentially hindered by the fewer patients enrolling in clinical trials. The older patients are getting treatment, but they are not going into adult oncology clinical trials.

In the NCI's work with COG, it is known that medical oncologists see many adolescents, yet these patients never move into the pediatric oncology realm. Thus, a teenager who has a bone tumor and is treated in a small cancer study in a cancer center may receive less than ideal therapy for what pediatric oncologists consider a very treatable tumor. Pediatric oncologists find this situation tragic; a local doctor may refer a patient to a surgeon to remove a tumor when, in the next big city, there may be an academic center with the knowledge of exactly how best to treat this kind of tumor. Notwithstanding the nearby availability of appropriate treatment, parents still need to understand that their child may end up on an inferior arm of a clinical study.

Moreover, pediatric oncologists need to find out where treatment decisions are made; for example, who decides whether a child goes to a pediatric medical center, to an adult medical center, or to the community hospital? The Internet has been a valuable resource and is a good source for teenagers to learn more about clinical trials for their age group. When they search for information, they find out about COG and other treatment possibilities.

Ratio of Cooperative Group Therapeutic Study Entries to Annual Cancer Incidence, U.S., 1997-2001

Montello and colleagues have provided data showing that, from 1997 to 2001, the ratio of therapeutic study entries decreases dramatically among late teenagers and young adults.

Summary

The primary aspects of recruiting and retaining children in cancer clinical trials are:

- Access. Most patients are diagnosed and treated at COG member institutions.
- Quality. The academic environment provides high-quality research and offers well-trained specialists and researchers.
- Commitment. The culture of clinical research places an emphasis on helping patients.
- Infrastructure. The NCI fully supports the COG.

EXPERIENCE OF RECRUITMENT AND RETENTION OF CHILDREN IN MENTAL HEALTH TRIALS

Benedetto Vitiello, M.D., Chief, Child And Adolescent Treatment and Preventive Intervention Research Branch, Division of Services and Intervention Research, National Institute of Mental Health, NIH

Background

Dr. Vitiello works in the National Institute of Mental Health's (NIMH) Child and Adolescent Treatment and Preventive Intervention Branch, which deals with treatment and prevention interventions and is part of the Extramural Program at the NIMH. The Branch supports and conducts research that is done extramurally at universities and the community, not on the NIH campus.

The primary mission of NIMH pediatric clinical trials is to study the efficacy, effectiveness, and safety of interventions to prevent or treat mental illness in children and adolescents up to 18 years of age. In doing so, researchers collect information about the therapeutic value of interventions that can be used to prevent or treat mental illness in children and adolescents. The ages of trial participants range from 0 to 18 years of age, with most of the trials aimed at the 6- to 17-year-old age group. A few of the trials—particularly autism trials—involve children ages 3 to 6, but very few trials involve infants.

NIMH fiscal year 2002 funding for these trials totaled about \$67 million. By research area, funding was as follows (in millions of dollars/percent of total):

- Depression—\$15 million/23 percent.
- Conduct disorder—\$9 million/13 percent.
- Anxiety—\$8 million/12 percent.
- Attention deficit hyperactivity disorder (ADHD)—\$6 million/9 percent.
- Bipolar—\$6 million/9 percent.
- Autism—\$3 million/5 percent.
- Schizophrenia—\$3 million/5 percent.
- Other—\$17 million/24 percent. This last category accounts for a substantial part of the NIMH research effort, because much research—particularly in prevention—is not targeted at a particular disorder but rather at trying to prevent mental illness in general.

Examples of currently funded NIMH multisite pediatric mental health trials include the following:

- The Families and Schools Together (FAST) Track (FAST TRACK) study of early prevention of conduct disorder attempts to identify and address conduct

problems while children are in preschool and elementary school. It has long been recognized that conduct problems are difficult to prevent and treat by the time children reach adolescence. There is a need to identify children with the key symptoms while they are still in preschool and early elementary school and then try to intervene at the family, school, and community levels to try to prevent worsening of the problem. This prevention trial is currently in a followup phase at four sites across the country. It has about 900 child participants.

- The study of treatments of pediatric obsessive-compulsive disorder (OCD) compares the efficacy of pharmacological, psychotherapeutic, and combined treatments for children with OCD, which is a not uncommon and often is an impairing condition among youths.
- The Research Units on Pediatric Psychopharmacology (RUPP) Network conducts multisite studies of treatment interventions with children who have a variety of mental health disorders. Among recently completed studies, a study of fluvoxamine for children with anxiety disorders showed that a medication was more than twice as effective as a placebo. The trial involved 128 children and adolescents ages 6 to 17 over a period of 8 weeks. Symptoms improved in 76 percent of those randomly assigned to take the medication compared with improvement in only 29 percent of those in the placebo group.
- The Treatment for Adolescents with Depression Study (TADS) comprises 13 sites and involves adolescents ages 12 through 17. The purpose of this trial is to compare different treatment modalities. Teens are randomized to one of four groups to receive (1) an antidepressant medication (Prozac), (2) psychotherapy (cognitive behavioral therapy), (3) a combination of the two, or (4) a control (supportive therapy with a placebo). More than 400 teens eventually will be randomized into this study, which has been going on for several years. All of these trials have taken longer than expected, because recruitment is much slower than originally expected.
- Another current study is the Treatment of Resistant Depression in Adolescents (TORDIA) study, which hopes to identify the most effective next step for adolescents with major depression who have shown no improvement after treatment with a selective serotonin reuptake inhibitor antidepressant. The question to be answered is whether another Prozac-like medication, another type of antidepressant, or psychotherapy to boost the antidepressant effect of

the medication should be tried next. The study is running at six sites.

- The Preschoolers with ADHD Treatment Study (PATs) is trying to test the efficacy and safety of medication treatment for preschoolers with symptoms of ADHD. Typically, ADHD is recognized in elementary school students by teachers who notice problems in students' ability to pay attention in class. Retrospectively, however, parents can relate that they noticed these symptoms in children with the most severe problems when they were as young as 3 or 4 years of age. More than likely, the preschool teacher had already had trouble handling these children. Indeed, some of these children have had difficulties attending preschool because they were too disruptive. In the community, methylphenidate (Ritalin) is increasingly used for these children.
- The Treatment of Early-Onset Schizophrenia Study (TEOSS) is examining the effectiveness and safety of three different antipsychotic drugs in the treatment of early-onset schizophrenia and related disorders among youth.
- Treatment of Early-Age Mania (TEAM). This is a new study that is currently being launched at six sites to test the efficacy and safety of different pharmacological treatments for children and adolescents with bipolar disorder.
- The Children Anxiety Multimodal Study (CAMS) is a multisite investigation of the treatment of anxiety disorders in youth to compare the efficacy of medication, cognitive-behavioral therapy, and combined treatment for youths 7 to 17 years old with anxiety disorders.

Major Challenges

Recruitment

The two major obstacles to participation in NIMH child psychiatry clinical trials are slow and limited patient recruitment into funded projects and the small pool of researchers, which results in few applications for new studies. Increased NIH funding in the past 4 years has allowed for the establishment of a large portfolio of trials.

Achieving an adequate sample size for child clinical trials is a major challenge. Some of the trials listed in the previous section not only are delayed but also must settle for a lower number of participants than was originally anticipated in the protocol. The price paid is less statistical power for the primary hypothesis, for any secondary hypothesis(es), and for any sort of subtype analysis.

Another challenge is recruiting samples that are truly representative of the kinds of patients to whom treatment

must be extrapolated. This includes recruitment of ethnic minorities, which is very difficult in psychopharmacology trials. Researchers have found that some ethnic minority families do not like for their children to be treated with medications for mental health problems, preferring psychotherapy instead.

Recruitment is especially difficult in mood disorder trials; it appears that it is much more difficult to conduct studies of children with mood disorders such as depression or mania than studies of children with ADHD, anxiety, or even OCD. Mood disorder trials may experience recruitment challenges because of the recurrent and episodic nature of this condition.

Protocols With Multiple Diagnostic Requirements

Recruitment into protocols that seek to study children with comorbid mental health conditions is usually more difficult than recruiting for studies focused on just one disorder. This happens because, although comorbidity is quite common among children, finding study participants who meet all of the diagnostic criteria for two co-occurring disorders can be challenging. Recruitment is made conditional on the multiple conditions that must be met.

Sequential designs also are more difficult. If, to randomize, a researcher needs patients who have failed previous treatment, recruitment again becomes conditional on having met a previous condition. This happens, for example, in the ongoing study of adolescents who have previously failed a Prozac-like medication for depression. To recruit into the study, researchers must first make sure that these adolescents have received adequate treatment in the community and have not improved as a result.

Screening versus Enrollment

A large number of potential child participants must be screened to obtain an adequate number of enrollees. The ratio between the number screened and the number enrolled varies widely across studies—roughly from 10 percent to 25 percent—as well as across sites; some sites are much more efficient than others. The rate can be as low as 5 percent. Researchers need to compensate for this low ratio by screening a larger initial sample. Usually, the limiting step is to get patients beyond the first screening. Sometimes researchers are concerned that the inclusion criteria are too strict and that the criteria should be broadened. However, by enlarging the inclusion criteria, researchers end up with a more heterogeneous sample.

Exclusion From Trial Participation

Some families do not want to participate in clinical trials, a position at times associated with the family's socioeconomic status (SES) or with ethnicity. Families of low SES are at times hesitant to enroll their children in research involving medication treatment for mental health problems.

The most common reasons for exclusion include the following:

- The child does not meet the full diagnostic criteria for the disorder. The child may have the symptoms of the disorder under study but does not quite meet all of the criteria. In an ADHD study, for example, child participants should meet all of the criteria; they should not just have some symptoms of inattention or hyperactivity.
- The child is receiving a concomitant treatment that is exclusionary. This means that the child is undergoing treatment with another medication or form of psychotherapy that cannot be incorporated into the protocol.
- A common occurrence is that the child has already received one of the study treatments. The NIMH often conducts trials to compare treatment modalities that are already available in the community. Thus, one incentive to participate, namely that of trying a brand-new treatment not yet available to the general public, does not apply. In addition, children may have already tried the treatment elsewhere and thus become ineligible for or not interested in the study. Although participation in an NIMH trial usually offers better quality of care, better monitoring, and free treatment, the types of medications used are not often different from those available outside a research protocol.

Sources of Recruitment

The NIMH uses many avenues to recruit children and adolescents into clinical trials. Advertising to the community is becoming increasingly important in recruiting children into mental health studies, which may create some problems in itself, since the purpose of the research is to gather data that can be extrapolated to clinically referred patients. Direct advertising to the community may recruit patients who are different from those who are filtered through clinics and doctors' offices.

Although it is often difficult for parents to know the symptoms of ADHD and other disorders and it is important to educate the general public so that children do not go untreated, there is a mismatch between community recruitment (e.g., via newspapers) and clinical referrals. In Europe—for industry-supported antidepressant

clinical trials—there is almost no community recruitment through direct advertisement; rather, recruitment is almost always accomplished through clinics. In the United States most recruitment is done through advertisements to the community. Although it is a matter of debate, a higher placebo response or differences in effect size may be explainable by the recruitment method used. There is a very high placebo response in U.S. antidepressant trials, whereas the placebo response rate is somewhat lower in European trials.

Clinical referral through schools also occurs, although this source is on the decline because of the many regulations in school districts that prohibit teachers, school psychologists, and other school personnel from bringing up the subject of treatment and referral to clinical trials. The most school personnel can do is to suggest clinical evaluations of certain students, but they are hesitant about referring student patients to clinical trials.

Clinical referral from other clinicians is the best way to recruit child study participants but it does not happen often. Psychiatrists and psychologists do not frequently refer; pediatricians refer more often, but all of these may be hesitant to lose patients. Overall, there are few referrals of children to NIMH clinical trials from mental health providers.

Descriptions of Two NIMH Child Mental Health Studies

Multimodal Treatment of ADHD (MTA)

This trial was conducted through a cooperative agreement that began a few years ago; the main report was published at the end of 1999. The purpose of the MTA trial was to compare different treatment modalities of children with ADHD. This comparison of the effectiveness of pharmacological, psychotherapeutic, and combined treatments for ADHD attempted to answer the question whether it was better to use psychotherapy (behavioral therapy focusing on improving behavior and attention) or to use Ritalin or another stimulant to treat ADHD.

Study Design. The study used a parallel-group design. A total of 579 children ages 7.0 to 9.9 years of age were randomized at six sites for 14 months—an adequate period of time for establishing treatment effectiveness.

Conclusions. Treatment modalities that included the medication were more effective than modalities that did not include the medication in decreasing symptoms of ADHD; that is, psychotherapy alone was not as effective in controlling the symptoms as was the medication.

Recruitment—Sex Distribution. Of the 579 children recruited, 465 (80 percent) were boys, and 114 (20 percent) were girls; 28 percent were children from ethnic

minority families. The ratio of the occurrence of ADHD among boys compared with girls is about 5 to 1, so the sample generally reflects the distribution of the condition among youth. Initially, 4,541 potential participants were telephone screened. Of the 4,500 who called in, 49 percent were excluded because they were the wrong age (20 percent), they lived too far from the study site (14 percent), the parents refused (9 percent), the family moved (5 percent), or medical reasons (2 percent). Clinicians, pediatricians, and schools referred about two-thirds of the patients; the remaining one-third were recruited via direct advertisements to the community.

Enrollment. Of those remaining, 2,337 were mailed a package with forms about symptoms of ADHD; of these, 60 percent were excluded (22 percent had insufficient ADHD symptoms, 18 percent did not return the complete package, 14 percent of parents refused to let their child participate, 6 percent for other reasons). Of the 929 potential participants evaluated in person, 38 percent were excluded (12 percent of parents refused at that point once they knew more about the planned treatment and trial design, and 9 percent did not meet all of the criteria for ADHD). Of these 929, the final sample contained 579 children.

Treatment Assignment. Of the 579 participants, 144 children were randomized to medication; of these, 13 (9 percent) refused to start medication once they knew they were assigned to that arm of the trial. Of the 144 children randomized to psychotherapy, 38 (26 percent) found that psychotherapy was not good enough for them and crossed over to medications (i.e., started taking Ritalin on their own by getting a prescription from their pediatricians). After informing the researchers of their move to medication, this became a protocol violation, so they were removed from the study. Of the 145 randomized to combined treatment, 5 (3.4 percent) refused medications, and 1 (.07 percent) refused psychotherapy. In addition, 146 were randomly assigned to community care. Thus, although there was high acceptance in the beginning, there was low compliance in the long term.

Research Units on Pediatric Pharmacology Autism Network Study of Risperidone

When a study is being conducted, researchers do a subgroup analysis to determine whether there is a gender effect. Even though statistically there typically is not much statistical power, a visual inspection of the data can be quite revealing. Of the 270 children with autism who were screened for the RUPP study, 169 were excluded; of these, 112 did not meet the study criteria, 50 refused to participate, and 7 were eliminated for other reasons. The remaining 101 participants (37 percent of those screened) were randomized. Of the 101 participants, 80 percent were male, and 33 percent were female, which is not out of line with the general population male-to-female autism distribution rate of 5:1. About 33 percent were minorities.

Results by Gender. Both boys and girls who received medications experienced dramatically reduced symptoms. Girls on average were somewhat less severely affected at baseline, but improvement over time was similar in both gender groups. Thus, gender did not act as a moderator of treatment in this study. This is an example of how subtype analyses can be informative in testing the possible presence of gender effects in clinical trials.

In mental health, there are no biological markers of disease or treatment response that can be used as treatment outcome measures in clinical trials. There are no blood tests or reliable psychological tests that provide a “score” measuring treatment response or effectiveness or that someone does or does not have a disorder anymore. Instead, researchers must rely on observation or on information from nonpatient informants such as parents or teachers. In most cases, the mother is the primary informant for almost all NIMH child clinical health trials. For example, in the MTA trial, 96 percent of the key information for entry and outcome evaluations came from mothers of participants. Even when the father provides some information, key information came from the mother, which is typical in child mental health trials.

Panel IV: Observational Studies versus Clinical Trials— Why Are the Results Inconsistent?

Moderator: Barbara Alving, M.D.

Information flows logically from clinical studies, in which researchers suspect that an agent has an effect, to clinical trials in which researcher test the suspected effect of that agent in a carefully controlled experiment. Thus, observational studies and clinical trials are complementary. Sometimes the conclusions of the observational study differ from those of the clinical trial because researchers want to isolate a single effect from a single agent in a biologically and physiologically complex organism, and because researchers learn more and more as they collect more and more data and as they analyse and reanalyse the data. Education is essential to prevent confusion and to effectively disseminate the results and their implications to the public and the news media, as well as to colleagues and physicians who are not specialists in the particular field.

A COMPARISON OF THE DIFFERENCES, SIMILARITIES, AND ROLES OF OBSERVATIONAL AND CLINICAL TRIALS

*Lawrence Friedman, M.D., National Heart, Lung, and
Blood Institute, NIH*

Characteristics of the Two Types of Studies

Dr. Friedman discussed the similarities and differences of the roles of clinical trials and observational studies, with a focus on how the two types of studies complement each other. He cited two review articles published in 2000 in *The New England Journal of Medicine* in which the authors conclude that results from clinical trials and observational studies are usually similar. (Benson and Hartz 2000, Concato et al. 2000).

The methodologies in a clinical trial and in an observational study are different. A clinical trial is usually prospective; participants are randomized and eligibility criteria are clear. The intervention is usually highly defined—what, how, when, how much, the route of administration—and is done in a structured setting. It addresses one specific primary question, although there may several secondary questions. A clinical trial is often of short duration, with relatively few participants.

An observational study can be either prospective (cohort) or retrospective (case control). Participants receive treatment through their individual care systems, so it is highly variable and based on perceived need, not on random assignment. There is no randomized control group. There are no eligibility criteria for the participating population, which is often very large. The observational

study involves many measurements and often has a long followup (e.g., the Framingham heart study that began more than 50 years ago).

Complementary Outcomes and Implied Issues

Despite these differences the two types of studies usually have similar results because they are well designed and implemented, analyses are done properly, findings are often robust, interpretations are cautious, and scientific strengths usually outweigh study weaknesses.

An example that the two are complementary is the use of beta-blockers after myocardial infarction. Many clinical trials conducted in the 1970s and 1980s showed benefit from this treatment; however, observations were made on small numbers of elderly persons, women, minorities, and those with other serious illnesses. Despite the results of these trials, beta-blockers were underutilized because of concern that the benefits were unproven for various groups. Furthermore, drug companies were not pushing beta-blockers because these drugs have side effects and were not covered by patent. In 1999 the results of observational studies of beta-blocker therapy for acute myocardial infarction in elderly patients (Krumholz et al. 1999) and for myocardial infarction in elderly diabetic patients (Chen et al. 1999) showed benefits of the therapy. The conclusion was that beta-blockers could extend the benefits of the clinical trials done in rather narrow populations because of the results of the observational studies.

Occasionally, results differ between the two types of studies because of unrecognized and, therefore,

unexpected confounders, different durations of intervention, different interventions, different populations, or interventions used at different times in the course of the condition or disease. Four questions are raised by these differences:

- How do we decide when to trust observational study results?
- Can we know, in advance, which observational study results we should question and, therefore, whether clinical trials are needed?
- If we cannot, do all questions need clinical trials?
- Why do we need to pit one against the other?

There are many kinds of clinical research, all of which have strengths and weaknesses. No single study—clinical trial or observational or any other kind—stands alone; the findings must be put in the context of other research. Usually, the result is not “observational studies versus clinical trials,” but rather “observational studies plus clinical trials.” When the studies yield opposite conclusions, the reason for the difference must be examined, but generally the results of a clinical trial should be accepted if the study was large enough and well conducted.

EXAMPLES OF RESULTS FROM OBSERVATIONAL AND CLINICAL TRIALS WITH RESPECT TO BETA-CAROTENE AND VITAMIN E

Julie E. Buring, Sc.D., Brigham and Women's Hospital/Harvard Medical School

The unexpected results of the Women's Health Initiative (WHI) have raised the question of the relationship between the results of observational and clinical studies and whether there is a way to better predict the outcomes of an intervention. Dr. Buring described two examples of other discrepancies—beta-carotene and vitamin E.

Beta-Carotene

In the 1980s beta-carotene seemed to hold promise as a nutritional supplement that could be effective in the prevention of cancer. Nonhuman animal and laboratory studies had shown that beta-carotene could block the carcinogenic process and inhibit specific tumor growth. There seemed to be a plausible, though unidentified, biological mechanism. A large body of observational epidemiologic evidence had consistently demonstrated that people with a high intake of fruits and vegetables rich in beta-carotene over long periods had a lower risk of cancer, especially lung cancer, and of cardiovascular disease. A higher serum beta-carotene level also indicated a lower risk of cancer and heart disease. These results came from very good epidemiologic studies, such as the Nurses' Health Study, in which the sample size was large, the dietary data were good, and the methodology was solid.

In contrast to the results from observational studies, two large-scale clinical trials in well-nourished populations showed no benefit of beta-carotene on the development of lung cancer and in fact found an increased risk of lung cancer among heavy smokers who were given high-dose beta-carotene supplementation. The Alpha-Tocopherol Beta-Carotene Cancer Prevention Trial studied vitamin E and beta-carotene in a factorial design in Finnish male smokers. The Carotene and Retinol Efficacy Trial studied beta-carotene and vitamin A in combination in those at high risk of lung cancer because of smoking or asbestos exposure. A third study, the Physicians' Health Study, found no benefit or harm after more than 13 years of dietary supplementation, but the study population had a low proportion of heavy smokers, and so the risk for lung cancer was not high.

There are a number of explanations for why a benefit was not seen in the trials although a benefit had been suggested by the observational studies. First, the observational studies could not adequately control for

confounding factors. The relative risk was only about 1.8, and high intake of fruits and vegetables has uncountable millions of correlates. Second, the hypothesis might not have been well grounded. The Nurses' Health Study was not designed to look at the relationship between beta-carotene intake and cancer; it was designed to look at the adverse effects of oral contraceptive use. Among myriad factors that were looked at were diet, and in diet multitude of nutrients were looked at; beta-carotene was the one that looked the most promising. The leap to a supplement of beta-carotene was rather large. Third, it might be that beta-carotene acts synergistically with other nutrients, rather than alone, as a preventive agent. Fourth, the duration of 3 to 5 years might have been too short. Finally, the dose may have been wrong.

The discrepancy between the studies led back to basic research. Basic research showed that in nonhuman animal models beta-carotene acts as an anticarcinogen but that its oxidized products can facilitate carcinogenesis. What happened in the trial was an aberrant metabolism of beta-carotene at high doses in the presence of smoking. Also, it has been found that other antioxidants, such as vitamin C, act as stabilizers for beta-carotene, but vitamin C was never included in the trials.

Vitamin E

The case was strong for conducting a clinical trial with vitamin E. It had a better biological mechanism than beta-carotene, it is a good antioxidant, and in vitro and nonhuman animal studies provided a large and compelling body of evidence that oxidation of low-density lipoprotein and related oxidative mechanisms play a critical role in the initiation and progression of atherosclerosis. Most observational studies had shown that there was a lower risk of heart disease with higher intake of vitamin E, either through diet or supplements; that the lowest risk was seen with vitamin E at a level that could only be realistically achieved with supplements; and that the effects were seen in a short period of time.

Nevertheless, the main clinical trials to date have found no benefit of vitamin E on clinical coronary heart disease. These are good trials—GISSI from Italy, HOPE from Canada, and Heart Prevention Trial from England—large, well-conducted, with an adequate methodology to find an effect if present but of secondary prevention. No study has found harm, but none has found benefit. Perhaps, again, the problem is the focus on one agent, or, more reasonably, the problem is that these are secondary prevention trials.

The difference between the trials and the observational studies is in the biology. The animal models had suggested

that vitamin E affects early lesions but provided no information about advanced lesions. The other problem is that people who have heart disease are extremely well treated. In both groups, cholesterol is lowered, blood pressure is lowered, and every other treatment that is now considered good care is provided. So, the question is, What does vitamin E add to good medical care in the secondary prevention of heart disease? The difference between the two groups is much smaller than would occur in a situation without underlying disease and for vitamin E versus placebo in the presence of no other care. Other reasons might be healthy user effect, healthy complier effect, or uncontrolled and uncontrollable confounding.

Nobody has measured oxidative stress, which would seem important if the agent is going to act by oxidative stress and the outcome measured is heart disease. Nobody has even determined whether baseline oxidative stress had an effect on vitamin E. An analogy would be to give people an antihypertensive drug to see whether it makes a difference on stroke without ever measuring whether the people had hypertension and whether it worked differently in people who had or did not have hypertension and who had different levels of blood pressure. These factors never even got into the model.

Secondary prevention trials that show no effect of vitamin E for 3 to 5 years on advanced coronary heart disease do not answer the question raised by the animal models and observational studies. More relevant would be studies of primary prevention, and better would be the early development of lesions, but that is very difficult to do. Another approach might be to look at vitamin E in conjunction with other dietary elements. The recommendation must remain to increase consumption of fruits and vegetables, but the specific recommendation to take vitamin E or any other antioxidant vitamin supplement would not be warranted at this time.

Conclusion

In considering the WHI, nothing went wrong! Trials are conducted because there is some previous evidence of a benefit. It is important to look at the difference between the rationale and design of the observational study, and the design of the trial. The examples of beta-carotene and vitamin E illustrate that, if differences exist between the two, it is important to consider carefully whether there is reason to believe that those differences could affect the action of the agent or the expected difference.

RESULTS FROM THE HEART AND ESTROGEN/PROGESTIN REPLACEMENT STUDY AND THE WOMEN'S HEALTH INITIATIVE: DO WE KNOW WHY THEY DIFFER FROM THE EXPECTED?

Marian C. Limacher, M.D., University of Florida/Women's Health Initiative

Dr. Limacher focused her remarks on the results of the Nurses' Health Study, the Heart and Estrogen/Progestin Replacement Study (HERS), and the WHI, comparing them with the observational studies that led to those studies.

The underlying hypothesis of all the estrogen and cardiovascular disease studies is that replacing estrogen levels might delay cardiovascular disease or prevent it. This reasoning derives from the fact that heart attacks largely begin to occur among men in their 50s but among women 5 to 10 years later. During those 5 to 10 years, menopause occurs. Evidence for biological effects at the lipid level and at the blood vessel level has been largely positive, although not completely. Nonhuman animal models have been largely positive, although not completely. Observational studies have been entirely supportive. A number of well-done observational studies, cross-sectional studies, and angiographic studies 10 years ago were analyzed and summarized for the effect of supplemental estrogen for protecting women against heart disease. The overall effect from the observational data was a 46 percent reduction and an odds ratio (OR) of 0.54 for all of the analyzed studies.

Nurses' Health Study

A large component of the observational data is from the Nurses' Health Study. Beginning in 1976, 70,000 women, ages 30 to 55 at entry, were followed. A questionnaire and blood sampling were done locally and sent to Boston. The questionnaire asked about duration, dose, and type of hormone replacement therapy (HRT) and whether myocardial infarction or stroke had occurred, as well as other outcomes; hospital records were acquired to confirm these statements. The 20-year followup shows that past use of HRT reduces the risk of cardiovascular disease by 18 percent, and for current users, substantially more—similar to the 0.54 OR seen in the summary. Use for more than 10 years has less of a reduction than use for less than 10 years.

The bulk of the estrogen supplement was conjugated estrogen alone. The estrogen+progestin group also had a reduction, not quite at the same level, but very similar. Some of the questions that have subsequently been raised are available in the Nurses' Health Study, which

did demonstrate an increased risk of stroke, particularly for current users. For past users, there was very little difference. Conjugated estrogen users had an increased risk although confidence intervals overlapped, but the estrogen+progestin group in the Nurses' Health Study had the highest risk for stroke. Thus, the WHI report of increased risk of stroke is not totally surprising. The predominant type of stroke in the Nurses Health Study was ischemic.

Heart and Estrogen/Progestin Replacement Study

The HERS was the first randomized, blinded, placebo-controlled trial of HRT in women, with 2,763 women under the age of 80 (mean 66.7 years). The intervention was the Prempro formulation of 0.625 mg of conjugated estrogen plus 2.5 mg medroxyprogesterone acetate versus placebo. A 4.1-year follow-up showed no difference in cardiovascular outcomes, with an increase, although not statistically significant, in cardiovascular death. All cardiac secondary outcomes showed no difference. The risk at 4 years showed no difference, and there was actually increased risk at 1 year for the combination therapy and placebo. That finding was the first sign of inconsistency in the theory.

Women's Health Initiative

Participants in the WHI were postmenopausal women, ages 50 to 79. Cardiovascular disease did not exclude a participant, but she could not have active disease, that is, no myocardial infarction within the past 6 months. The mean age was 63, somewhat younger than in the HERS, with the age distribution predominantly in the 60s but with a substantial number in the 50s as well as the 70s. The percentage of participants being treated for diabetes at baseline was quite low and not different between the two studies. The proportion of women who were hypertensive or on antihypertension therapy was a substantial proportion, although lower than the known heart disease population, which tends to be in the 50 percent or higher range. Women with known high cholesterol or who were on statin therapy constitute a low proportion, and the proportion was not different between the two studies. The percentages with coronary disease or stroke were quite small. So for some, this was a true secondary prevention trial.

The outcomes for coronary heart disease rates were higher for the treatment group than for the placebo group; the two curves separate within several years and maintain separation, although the separation is smaller in the last few years with smaller numbers reaching those years of followup. The overall hazard ratio was 1.29. The

nominal confidence intervals were in the direction of the adjusted confidence intervals; they overlapped and were therefore not statistically significant. They also did not support benefit. Stroke rates were similar for the first 2 years, and then risk increased. The risk of pulmonary embolism increased throughout.

Challenges to Interpretation

Several considerations exist about elements that might confound the interpretation of results, and unexpected results are an opportunity to look into the biological explanations. When the WHI and the HERS are plotted with the cohort summaries, the direction for benefit shifts to the other side of the confidence interval: for the HERS, an OR of 0.99, but no benefit; for the WHI, a hazard ratio of 1.29, but on the other side of the OR. How the results are interpreted and what is concluded, particularly for the observational studies, necessitate clinical trials.

The timing of applying the treatment is a factor, because in the natural setting postmenopausal women use estrogen for menopausal symptoms. As also happened in the WHI, the intervention was applied postmenopausally but not for menopausal symptoms. That is a big difference.

Risk level is an important factor. Cardiovascular risk factors must be looked at specifically—age, age at onset, age at exposure. The WHI population was heavier but included fewer smokers. The risk levels between the primary and secondary prevention, as the HERS has been traditionally labeled, are the event rates for estrogen+ progestin in the HERS and placebo, but the HERS and the WHI sampled different populations.

Bias could be an issue. The healthy user bias is easy to demonstrate in HRT. Women who use HRT are younger, better educated, leaner, more likely to use alcohol, more physically active, less likely to smoke cigarettes, and less likely to have diabetes and have a lower-risk family history. Also, the placebo effect in clinical trials has been demonstrated in several studies, including the Beta-blocker Heart Attack Trial (60 percent risk reduction for compliant placebo takers), and the Coronary Drug Project (30 percent risk reduction for compliant placebo takers). Is that because the placebo works? Not likely. The more acceptable explanation is that this result is associated with a healthier profile in general. These are people who pay attention to their health and to their symptoms and seek treatment earlier. Diabetes is a true confounder, which cannot be accounted for in an observational study.

The timing in the Nurses' Health Study was different from that in the WHI. Using HRT close to menopause

may occur at a time when atherosclerotic lesions are in earlier development and more amenable to improvement; some nonhuman animal model data suggest that this is true. There is also the issue of early risk, late benefit. The women were tested every 2 years. There would be a survival bias for those who had early events, who are now saying they are still on their hormones and they are better. These women are different from those tested in the clinical trial in which they are exposed to early risk.

There is truth in the results of both studies and the opportunity to analyze the differences to gain an understanding of the mechanisms. It is clearly not as simple as "throwing pills at people" because they seem to be good. The flow of observational study to clinical trial is logical and should be maintained. Clinical trials are still the best level of treatment recommendations, but both types of studies should be used together as a basis for learning what is really going on in the biology of the disease.

BREAST CANCER PREVENTION: RESULTS FROM OBSERVATIONAL VERSUS CLINICAL TRIALS

Karen Johnson, Ph.D., M.D., M.P.H., National Cancer Institute, NIH

The Breast Cancer Prevention Trial (BCPT) was conducted by the National Surgical Adjuvant Breast and Bowel Project. The women who entered this trial had a high level of risk, a 1.66 or greater chance of developing breast cancer within 5 years of starting the trial. They were randomized to receive tamoxifen (20 mg) or placebo for 5 years. Incidence of breast cancer in the tamoxifen intervention group versus placebo showed a 49 percent reduction in incidence, fairly consistent across age groups (44 percent in the 35 to 49 age group and 55 percent in the older age group). The dramatic reduction plus secondary end points from previous clinical intervention trials led to the rationale for a Phase III retrial without Phase I and II trials for the intervention.

The U.S. Food and Drug Administration (FDA) and the National Cancer Institute (NCI) agreed to and published principles and conditions that would justify starting early-phase clinical trials (Kelloff et al. 1995). One emphasis was on nonhuman animal models. In the case of breast cancer, often that is a mouse model, and a typical carcinogen would be methylnitrosourea. Another type of evidence described is compelling epidemiologic evidence or evidence from observational studies, for example, the effectiveness of a specific agent, such as beta-carotene, vitamin E, and calcium in a target tissue, or of a detriment involving a specific exposure. For the latter, there might be surrogate markers, such as on estrogen receptors and their connection with the increase of breast cancer incidence. The Gail statistical model of breast cancer incidence was used to help select women for the BCPT. Within that model, there are surrogate markers of estrogen exposure (e.g., age at menarche and hormonal events connected with first live birth). The Gail model was foreshadowed by well-known models that rely on age intervals that reflect exposure. Some of those intervals depend on early menarche and late menopause, intervals that reflect estrogen exposure and predict increased risk of breast cancer. Also there are results concerning HRT and its association with the increased risk of breast cancer. Those data come not only from cohort studies, case control studies, but also from randomized clinical trials. Those data are very consistent.

Dietary fat is a potential marker of estrogen exposure. This information comes from the Women's Health Trial (the precursor of the WHI). A total of 73 postmenopausal

women changed their diet to receive 20 percent of their calories from fat. The women who were checked had been on the intervention for 10 to 22 weeks. At that time, the baseline estradiol level had decreased about 17 percent. This is an example of how the dietary intervention can be connected with estrogen exposure.

Looking at estrogen exposure is a little less specific than looking at the intervention of tamoxifen. When using observational information, researchers must consider the complexity of the situation. For example, the serum level reflects the dose exposure of a detrimental factor, but not the availability. Binding globulins may keep the effect of the estrogen from being what it otherwise would be. Also, the duration of the exposure is important: The density changes over time. Recent information is that peripheral estrogens are not reliable predictors of risk. Half of breast cancers have intracellular mechanisms for producing estrogens, and local estrogen may override the peripheral effect measured in observational studies. This complexity relates to the interpretation of results: The most obvious conclusions may not always correspond to what is happening on the biological level.

One study that compares a low-fat diet with a high-fat diet is a pooled analysis of 12 case-control studies. The finding was that the relative risk of breast cancer was 1.46 when comparing highest with lowest quintiles for saturated fat (Howe et al. 1990). This is an example of observational data that might be interpreted to suggest that a low-fat diet could be an intervention to prevent breast cancer. However, prospectively collected data from seven cohort studies show no evidence for a positive association between total dietary fat and breast cancer risk (Hunter et al. 1996). The question is still open as to whether a low-fat diet can lead to a lower incidence of breast cancer. The WHI clinical trial will help clarify issues surrounding these questions. That trial involves postmenopausal women using a low-fat dietary eating pattern so that about 20 percent of calories come from fat. Five fruits and vegetables and six servings of grain products a day are encouraged. The end point is reduction in breast cancer incidence. For a 14 percent reduction in breast cancer incidence, the trial would have an 86 percent power to result in that kind of difference over 9 years; that is a reasonable expectation to go into a trial like this.

The question about vitamin D and breast cancer incidence remains open. This too is part of the WHI clinical trial. The epidemiologic data used to justify the WHI trial reflected a latitudinal difference in breast cancer incidence. It might have been due to sun

exposure, a chain of events that makes it harder to connect the incidence difference with vitamin D. When using a serum factor, one question is whether perhaps premalignant lesions may remove that factor from serum. If so, the association observed between low levels and risk of invasive cancer may be a so-called protopathic bias, or a change related to the developing disease.

It is rare to go into a clinical trial without multiple lines of evidence. In addition to observational studies, it is always useful to have information about the mechanism of action. In the case of vitamin D, most breast tumors and invasive cancers have vitamin D receptor expression. There is some information about downstream effects because studies with tumor cell cultures and vitamin D or vitamin D analogs show antiproliferative and proapoptotic effects. The intervention in the WHI clinical trial is 400 IU of vitamin D3 daily along with 1,000 mg of calcium. It is predominantly an osteoporosis intervention, so there is no discussion in the trial design of a breast cancer endpoint. Nevertheless, the results could be of interest to those involved in breast cancer and its prevention.

Summary of the Cytogenetic Model of Breast Cancer

The evolving cytogenetic model of breast cancer can be summarized as follows. Normal and developing breast cells and those that go on to form abnormal cells can be fingerprinted. The fingerprint for the stem cell compartment is identified; there are two lines of development, and stem cells progress through myoepithelial stages and luminal cell stages to set up a structure where the luminal (or ductless) cells are surrounded by myoepithelial cells in the tissue. The fingerprinting has indicated that a major component of ductal carcinoma in situ and atypical ductal hyperplasia arises from cells that are already skewed in differentiation toward the luminal cell pathway of development. A few precursors in cancer do come from the stem cell compartment before those stem cells express factors that align them with other differentiation pathways. There are underlying molecular events that contribute to the etiology of cancer. The challenge is to develop observational studies and look at things in a way that is more specific to the etiologic process.

Panel V: Recruitment and Retention of Participants in Drug Abuse Research: Incentives, Ethics, and Practical Considerations

Moderator: Cora Lee Wetherington, Ph.D.

The panel members described strategies for recruiting women and adolescents, including minority groups, that they have found effective in their own programs and studies.

LESSONS FROM THE FIELD: RECRUITMENT, RETENTION, AND ETHICAL CONSIDERATIONS WHEN WORKING WITH SUBSTANCE-INVOLVED WOMEN AND ADOLESCENT GIRLS

Sally J. Stevens, Ph.D., Southwest Institute for Research on Women/University of Arizona

Dr. Stevens' presentation focused on the lessons she has learned in four areas of intervention research through her experiences in working with research funded by the National Institute on Drug Abuse (NIDA) and the Substance Abuse and Mental Health Services Administration (SAMHSA):

- Reaching and recruitment issues
- Engagement and retention in interventions
- Gender-specific interventions
- Practical and ethical considerations

Reaching and Recruitment Issues

Reaching and recruiting adolescent girls into intervention research involves issues that overlap with but are different from issues for adult women. There is a high correlation between drug abuse and sexual abuse among adolescent girls, and girls know that if they disclose sexual abuse to a counselor, the abuser will be reported. As a result, they do not come forward. Feelings of shame and guilt also keep them from coming forward. Girls in this population tend to date and live with older men, on average 3.2 years older, who are usually using or selling drugs. These girls tend to believe that “everybody

does drugs.” They have a narrow view of life framed by their schools, peer networks, and neighborhood.

In studies conducted by Dr. Stevens, 18 percent of girls deny having a drug problem, despite having been expelled from school or arrested for drug use, having run away from home, and often feeling suicidal. A similar finding is seen in both the Adolescent Treatment Models Program and the Cannabis Use Treatment Program, in which 15 percent to 22 percent of adolescent girls entering treatment deny having a drug problem. Nationally, about 80 percent of referrals to treatment come from the criminal justice system.

Several approaches can be used for reaching and recruiting adolescent girls into intervention research programs for substance abuse:

- Intervene at younger ages by working with schools to identify risk behaviors such as truancy and daytime sleepiness.
- Recognize that work may be a risk behavior for drug use. Adolescent girls who have dropped out of school often find work in settings that expose them to drugs.
- Focus on “network research,” working with peers to change behavioral norms and provide a support system.
- Work with older male partners and their probation officers to encourage them to refer girls for treatment.

Key issues in reaching and recruiting adult women are the stigma associated with drug use, the fear of AIDS, unequal power relationships with men, and childcare

responsibilities. For example, significant others may prevent women from entering treatment. In the Tucson area, issues of citizenship status and lack of fluency in English are also major barriers; many women fear they will be deported if they enter a treatment program. Homelessness is a major issue. Adolescent girls are often living undercover, afraid they will be reported and sent home.

Mistrust of other women can be an additional barrier; women involved in the sex trade tend to view other women as “competition.” Creating an environment that fosters trust is very important. In Tucson, a NIDA-funded storefront in the “murder quarter” has been operating since 1988. The staff waits for a “moment of opportunity” when circumstances make a woman ready to enter treatment (e.g., she has an overdose, her partner is “busted,” her child is taken into protective custody, she is raped or assaulted). When that moment comes, immediate access to treatment must be available; a wait of even 3 days may mean the moment of opportunity has passed.

Engagement and Retention Issues in Interventions

Adolescent girls need appropriate staff and role models, young, ethnically diverse people whom the girls can relate to. For adult women, it is more important to have role models who have come through recovery themselves. Adolescent girls will sometimes say, “Look at Susie. She’s 24 now, and she used to be addicted, but she’s fine now. So I can keep using—I can get through it and be like Susie.” Adult women have a different attitude; they are more likely to say, “It’s amazing, Susie’s been clean for 2 years.”

The engagement and retention approaches used in the Tucson program include:

- Offer activities that will attract the girls’ interest and surround those activities with substance abuse treatment, such as opportunities to learn hip-hop dancing or attend a peer-leadership training retreat. Alternative activities can also play an important role in bringing women in voluntarily who have not been mandated by the criminal justice system to enter a treatment program.
- “Girl time” is also important. The time the girls spend together in the car going to the treatment center is special bonding time, when they can talk and share the ups and downs of their lives. Peer networks are important, too.
- Provide transportation to the program the minute school or work is over, so the girls have no chance to drift off.
- Maintain a significant level of contact, whether a phone call or a visit to the treatment program to remind a patient of her next scheduled visit.

The NIDA-funded outreach program offers a social support group but was having difficulty getting anyone to attend. Then one of the researchers received supplemental funds to start a reading group. The group met for 2 hours a week to read and discuss challenging reading material such as the novel *White Oleander*, Maya Angelou’s *I Know Why the Caged Bird Sings*, and other stories, books, and poetry dealing with themes of prejudice, sexual abuse, drug use, identity, and relationships. Members of the group would find themselves relating to one or more of the characters they were reading about. Attendance at a reading group was more acceptable to their significant others than attendance at a “drug abuse class” would have been.

Women like to volunteer and to feel involved in a larger cause. The Tucson program works with a community food bank to distribute food and has a clothing program in which volunteers sort and distribute the clothing. When a woman has volunteered for 40 hours at a program site, she receives a letter certifying that she has done this volunteer work.

Gender-Specific Interventions

Providing a woman-friendly environment is important for both adolescent girls and adult women. Programs for women and girls must be grounded in an understanding of the differences in male and female development. The typical male developmental model is one of moving from attachment to autonomy, whereas women tend to remain relationship oriented. It is also crucial to do research with women rather than on them. Drug-using women feel oppressed and alone. For these women, learning about women’s struggle to gain the right to vote and to own land, about women writers forced to publish under male pseudonyms, and about Rosa Parks and other women who participated in the civil rights movement helps them place their oppression into a larger context.

The Desert Willow Residential Treatment Program for pregnant women and women with children offered a 1-hour parenting classes twice a week, which participants consistently rated as the most boring, untherapeutic session they had to attend. The staff gradually came to understand that many of these women had failed in school, where the day is typically broken up into classes of 50 minutes to an hour in duration. Native American women may not have experienced boarding school themselves, but their ancestors did. Many of the Hispanic

women come from families of miners or laborers, who were told they could not learn and did not need to go to school. For them, the conventional 1-hour class was not an effective approach to learning.

As an alternative, the staff offered a 2-day workshop that included cognitive learning, therapeutic play, and sharing of emotions concerning parental relationships. The participants were given exercises to take home to do with their children and to discuss in the workshop the following day. This format earned outstanding scores from the patients and staff for both satisfaction and therapeutic effectiveness.

A survey of men and women patients to find out what aspects of therapy were most important to them revealed that that girls and women tended to want more family involvement and liked groups better than the boys and men did.

Practical and Ethical Considerations

Ethical issues include:

- Subjects should be paid for participating in research; however, payment raises ethical considerations. Men may demand the money from women subjects, and subjects may use the money they receive to buy drugs.
- Confidentiality is another sensitive issue. A patient may insist on having her 4-year-old in the interview room while she is asked about her sexual history. Consider providing daycare for children older than 18 months.
- Studies are needed to determine whether better interview data are acquired by trained interviewers or by clinicians.
- A clear policy is needed in studies about which data will be shared with clinicians and other agency staff members.
- It is important that to share study findings with the subjects and give them a voice in interpreting the data. Adolescent girls and women are smart, insightful, articulate, and thankful to be included in the process.

RETAINING WOMEN IN TREATMENT: LESSONS LEARNED FROM OUTPATIENT RESEARCH ON COCAINE DEPENDENCE

Stephen T. Higgins, Ph.D., University of Vermont

Dr. Higgins' presentation focused on outpatient treatment of persons with cocaine dependence, with special attention to the retention of women. More information about the topics he covered is available in *A Community Reinforcement Plus Vouchers Approach: Treating Cocaine Addiction* (NIDA 1998).

Scope of the Problem

The prevalence of cocaine use has decreased since the mid-1980s, but most of that decrease has occurred among lighter users; a core group of about 600,000 heavy users has not decreased significantly. An estimated 400,000 individuals smoke cocaine. About 1.7 million Americans age 12 and older admit in surveys to having used cocaine within the past 30 days; about two-thirds of these individuals are male.

Each year about 200,000 individuals are admitted to publicly funded treatment centers. Men are more likely to enter treatment through the judicial system, whereas women are more likely to enter treatment voluntarily. Overall, cocaine-dependent women are more likely than men to be engaged in treatment; this gender difference is even more pronounced among users of smoked cocaine. The relevance of treatment does not appear to be a problem among cocaine-dependent women, but retention in outpatient treatment remains a daunting challenge.

Retention rates among cocaine-dependent outpatients are equally poor for men and women; both drop out at precipitous rates unless special interventions are used to reduce attrition. Dr. Higgins described empirically supported strategies derived from 13 years of experience in his clinic and elsewhere.

Community Reinforcement Plus Vouchers

Using vouchers is an incentive strategy. In substance abuse treatment, the period from intake assessment through the first weeks of treatment is when the dropout rate is highest. The most intensive intervention occurs during weeks 1 to 12, when patients attend counseling sessions twice a week and undergo urinalysis three times a week. During Weeks 13 to 24, counseling sessions decrease to one per week and urinalysis to twice per week. During months 7 to 12, aftercare is offered, during which check-in with a counselor and random urinalysis is recommended on a once-a-month basis.

The voucher program is built around urinalysis. A patient who submits a cocaine-negative urine specimen earns a voucher that is exchangeable for a retail item. The value of these vouchers increases with each consecutive negative specimen. However, if a patient submits a positive specimen or fails to attend for a scheduled urinalysis, the value of the earned vouchers is reset to zero. The vouchers have no intrinsic value. The program staff makes all purchases. When patients have earned several vouchers, they tell us what they would like to purchase (e.g., a gift certificate at a toy store) and staff members purchase the item within a couple of days. Patients receive no cash, and the staff can veto inappropriate purchases, such as guns. Patients must come to the clinic to submit their urine specimens, which helps retention. They have an incentive for abstinence because they lose their earned vouchers if they resume drug use. While they are doing well, they come to the clinic, show through an objective measure that they are not using cocaine, and receive an incentive. If they “fall off the wagon,” the incentive system shuts down.

When this approach was initiated in Dr. Higgins’ program in the late 1980s, large incentives were awarded—up to \$1,000 over 12 weeks for perfect performance. On average, the cost per patient for such incentive programs is approximately \$500.

Evidence for Effect on Retention

Several elements of the community reinforcement approach (CRA) affect retention.

- **Accelerated intake.** Performing the intake assessment immediately or within 24 hours of the initial call improves engagement by about one-third.
- **Practical needs.** Using community resources to resolve problems such as homelessness, childcare, or other issues likely to affect engagement in treatment improves retention.
- **Outreach and case management.** Intensive followup through phone calls and home visits (or to other sites) reduces the attrition rate.
- **Other drug use.** Preventing patients from drinking alcohol decreases attrition.

Dr. Higgins’ program assessed the “CRA plus vouchers” behavioral approach with advantages over usual care. Of the 19 patients who received behavioral treatment, 58 percent completed 24 weeks of treatment compared with 11 percent of patients who received usual care. Women constituted only about 10 percent of the study population and retention was 100 percent in the behavioral

treatment group and 0 percent in the usual care group (Higgins et al. 1993). A second study assessed the impact of incentives on the outcome of intensive counseling alone or intensive counseling plus vouchers that were contingent on cocaine-negative urinalysis results in weeks 1 to 12. After 12 weeks, retention was 90 percent in the group that received vouchers, compared with 65 percent in the group that received counseling alone. After 24 weeks, retention was 80 percent in the vouchers group versus 40 percent for counseling alone. In this study, about one-third of the study participants were women. Retention was 100 percent among women in the vouchers group compared with 45 percent in the counseling-only group.

In a recent study of the behavioral element, with a larger population of women, retention at 24 weeks was about 65 percent in the CRA-plus-vouchers group compared with about 40 percent for vouchers only. Retention with vouchers only was similar to the retention rate with counseling only in the earlier study. Both elements of the intervention_vouchers and CRA behavioral counseling_are contributing to the outcome.

Studies of Accelerated Intake

In the first study by Festinger and colleagues, in an outpatient cocaine treatment clinic, 59 percent of those assigned to accelerated intake attended the intake assessment compared with 33 percent of those assigned to usual care. Subsequent retention rates were similar in the two groups, which means that accelerated intake is not bringing in unmotivated people who then drop out at a higher rate. On the contrary, it engages people at the crucial moment of opportunity to which Dr. Stevens referred, and these people do just as well as those who enter treatment on a more conventional intake assessment schedule.

The same group replicated these results in a subsequent study. In the followup study, callers were randomly offered intake assessment within 1 day, 3 days, or 7 days. Attendance was 72 percent for the 1-day group, 41 percent for the 3-day group, and 38 percent for the 7-day group. This illustrates again that if patients are told “Come in right away,” they are much more likely to engage in treatment than if they have to wait even a few days for an appointment (Festinger et al. 2002).

Role of Disulfiram

Disulfiram is a medication that interferes with the metabolism of alcohol; subjects become ill if they ingest alcohol while using the medication. There is evidence that it improves both retention and abstinence. Carroll

and colleagues conducted a randomized trial in which 122 cocaine- and alcohol-dependent outpatients (27 percent of whom were women) were assigned to one of five types of psychosocial counseling with or without disulfiram therapy. Patients who received disulfiram were retained for 8.4 weeks on average versus 5.8 weeks for those who received counseling alone (Carroll et al. 1998).

Child Live-In Program

Only one trial in the cocaine literature explicitly addresses the retention of women. In an 18-month residential treatment program, women who were allowed to bring one or two of their children were retained for average of 300 days compared with 102 days for women who were unable to bring their children.

Conclusions

The following strategies improve retention of cocaine- and other drug-dependent women in treatment, even among the most severely drug-abusing populations: accelerated intake, voucher-based incentives, helping women with childcare, and decreasing alcohol use.

CULTURAL ISSUES IN THE RECRUITMENT AND RETENTION OF WOMEN OF COLOR IN DRUG ABUSE RESEARCH

Kathy Sanders-Phillips, Ph.D., Howard University

Dr. Sanders-Phillips discussed some of the challenges in recruiting and retaining women of color, especially African American women, in clinical studies. Her comments were based on the findings of other investigators and her own experience. During the past 15 years, she has conducted several studies of health behaviors and outcomes in African American and Latina women whose children were enrolled in Head Start programs in south-central Los Angeles and two health promotion intervention programs involving African American and Latina Head Start mothers. A significant proportion of these women were using drugs. All of the studies were based on a conceptual model that examines women's health behaviors in the context of the social, political, and economic realities of their lives, realities such as exposure to violence and racism, which are particularly significant determinants of health in low-income women of color. At the beginning of the study, an advisory board of Head Start mothers and staff members was formed. All protocols and procedures were presented to the advisory board for review. This approach brought suggestions for recruiting and retaining participants in the studies that would not have otherwise surfaced.

Recruitment and Retention Strategies

For the study, Head Start mothers were trained to assist in recruiting other participants. This approach was highly successful. Also, Head Start teachers were trained to conduct the interventions, and Head Start mothers were trained to serve as peer counselors who provided "booster sessions" for participants. At 18 months, retention was significantly higher in the group that included booster sessions led by Head Start mothers than in the group that did not include these sessions. Use of these strategies resulted in participation rates that ranged from 75 percent to 90 percent. After data collection, the researchers presented the findings to the Head Start parents and the staff for review, and they left a manual of study procedures and interventions at each Head Start program.

It was more difficult in general to recruit African American women than Latinas. To understand these differences and their implications for the recruitment of women of color, it is necessary to examine several social and historical factors that influence the participation of people of color in research studies in the United States.

Barriers to Participation

Differences among groups of women may partially explain the different levels of participation in research. For example, we concluded that the higher participation rates for Latina women may have been because they were first-generation immigrants to this country and had come here with a sense of hope and empowerment for a better future. In this respect, the sample is biased, since it consists entirely of those who have immigrated. In contrast, African American women seemed to have lost hope and a sense of a better future because of their inability after many generations to fully integrate into American society. A significant proportion of African American women rarely participate in any activities outside their homes.

The United States has an unfortunate history of deception, exploitation, and breach of trust involving participation in research by women of color. Killian and colleagues have argued that these experiences reflect the vulnerability of women of color as well as their unequal status and power compared with those conducting research and developing policy regarding scientific experimentation. Killian and coworkers have identified two categories of barriers to the recruitment of women of color into clinical studies: structural and conceptual. The following are three types of structural barriers.

- **Availability.** Women of color lack opportunities to participate in clinical trials in the places where they live and at hours convenient to their lifestyles.
- **Accessibility.** Safety concerns, needs for transportation and childcare, and low literacy and language abilities deter women of color from participating in clinical research.
- **Acceptability.** Historically, women of color tend to have negative experiences of and attitudes toward research.

Several other factors limited participation in clinical programs:

- **Safety.** Fear of violence in neighborhoods limited participation in the study protocols, and the staff had to build participation around times when women felt safe.
- **Violence in the community.** For example, in 1992, the year of the riots in south-central Los Angeles, we experienced the lowest participation rate. The general level of community chaos and looting of Head Start programs had a significant impact on participation.
- **Ability of the research staff to respond to questions and treat participants with respect and cultural sensitivity.** Respect is paramount. Women stated clearly that they avoided any health care

program where they felt disrespected or where they sensed they would encounter racism.

Recruiting drug-abusing women poses special challenges. Whereas for other populations the perception of risk is a likely motivation for participation in research studies, risk is a normal daily reality in the lives of drug-abusing women, and it does not motivate them to participate in drug abuse research. Drug-using women are likely to be hampered by multiple psychological, social, and economic difficulties. A vicious cycle often occurs. Initial drug use is usually related to trauma associated with violence; as women become more enmeshed in the drug lifestyle, they experience more violence. They are dependent on a male partner in a way that a male partner is not dependent on a woman. All these factors increase their feelings of alienation from the larger society and sap their motivation to protect their health or participate at any level in clinical research.

Conclusions

Developing theoretical models is critical, but those models must reflect a comprehensive understanding of the complexities of life for women of color. Empowerment models of intervention are particularly effective in recruiting and retaining women of color. Similarly, programs that focus on promoting health by promoting a more positive ethnic identity seem to be particularly helpful.

It is important to identify and include community stakeholders in the development and implementation of research; to identify ecological variables, such as the level of community violence, that may influence motivation and/or participation in studies; and to include investigators of color as members of the study team.

Low-income women of color, whose lives are often characterized by poverty, sexism, racism, and violence, are apt to feel hopeless and powerless, and their day-to-day experiences tend to reinforce these feelings. Investigators who hope to recruit these women for clinical studies must acknowledge and overcome these barriers if they are to be successful in encouraging participation in research protocols.

CHALLENGES AND LESSONS LEARNED IN THE RECRUITMENT AND RETENTION OF RURAL, INCARCERATED, AND VICTIMIZED WOMEN

T.K. Logan, Ph.D., University of Kentucky

[Dr. Logan was unable to participate in the panel due to illness. A handout summarizing her presentation was distributed to attendees.]

Appendix

Taxk Force Members

Cochairs

Otis W. Brawley, M.D.

*Associate Director, Cancer Control,
Winship Cancer Institute
Emory University
Glenn Building
Grady Campus
69 Jesse Hill Jr. Drive
Atlanta, GA 30303
404-489-9185
404-616-0705 Fax
otis_brawley@emoryhealthcare.org*

Julie E. Buring, Sc.D.

*Professor of Ambulatory Care and Prevention,
Brigham and Women's Hospital
Harvard Medical School
900 Commonwealth Avenue, East
Boston, MA 02215
617-278-0863
617-731-3843 Fax
jburing@rics.bwh.harvard.edu*

Members

Mary C. Blehar, Ph.D.

*Chief, Women's Mental Health Program
National Institute of Mental Health,
National Institutes of Health
Neuroscience Center
6001 Executive Boulevard, Room 8125
MSC 9659
Bethesda, MD 20892
301-443-2847
301 443-8022 Fax
mblehar@mail.nih.gov*

Victoria A. Cargill, M.D., M.S.C.E.

*Director of Minority Research and
Director of Clinical Studies
Office of AIDS Research
Office of the Director,
National Institutes of Health
2 Center Drive
Bethesda, MD 20892
301-402-2932
301-496-4843 Fax
cargillv@od.nih.gov*

Susan E. Cohn, M.D., M.P.H.

*Associate Professor of Medicine, Medical Center
University of Rochester
601 Elmwood Avenue, Box 689
Rochester, NY 14642
585-275-2590
585-273-1055 Fax
susan_cohn@urmc.rochester.edu*

Lorraine Anne Fitzpatrick, M.D.

*Professor of Medicine and Director
Women's Health Fellowship,
Mayo Clinic and Foundation
Joseph Building, Room 5-194
Rochester, MN 55905
507-255-6663
507-255-4828 Fax
fitz@mayo.edu*

Gil Friedell, M.D.

*Kentucky Cancer Registry
2365 Harrisburg Road, Suite A230
Lexington, KY 40504
859-219-0772
859-219-0548 Fax
gil@kcp.uky.edu*

Susan L. Hendrix, D.O.

*Assistant Professor, Department of
Obstetrics, Hutzel Hospital
School of Medicine
Wayne State University
4704 St. Antoine Boulevard, GYN-5 Center
Detroit, MI 48201
313-966-9101
313-993-8504 Fax
shendrix@med.wayne.edu*

Barbara V. Howard, Ph.D.

*President,
MedStar Research Institute
6495 New Hampshire Avenue, Suite 201
Hyattsville, MD 20783
301-853-7502
301-853-7591 Fax
barbara.v.howard@medstar.net*

Ruth J. Katz, J.D., M.P.H.

*Associate Dean for Administration, School of Medicine
Yale University
333 Cedar Street
P.O. Box 208055
New Haven, CT 06520
203-785-4672
203-785-7437 Fax
ruth.katz@yale.edu*

Raynard S. Kington, M.D., Ph.D.

*Associate Director for Behavioral and
Social Sciences Research
Office of the Director
National Institutes of Health
1 Center Drive, Building 1, Room 256
MSC 0183
Bethesda, MD 20892-0183
301-402-1146
301-402-1150 Fax
kingtonr@od.nih.gov*

Lewis H. Kuller, M.D., Dr.P.H.

*University Professor of Public Health
Department of Epidemiology
University of Pittsburgh
Bellefield Professional Building
130 North Bellefield Avenue, Room 550
Pittsburgh, PA 15213
412-383-1895
412-383-1956 Fax
kullerl@edc.pitt.edu*

Marian C. Limacher, M.D.

*Professor of Medicine, Division of
Cardiovascular Medicine
College of Medicine
University of Florida
1600 Southwest Archer Road
P.O. Box 100277
Gainesville, FL 32610-0277
352-846-1228
352-846-1217 Fax
limacmc@medicine.ufl.edu*

Worta McCaskill-Stevens, M.D.

*Program Director, Division of Cancer Protection
National Cancer Institute
National Institutes of Health
Executive Plaza North
6130 Executive Boulevard, Room 300
Bethesda, MD 20892
301-496-8541
301-496-8667 Fax
wm57h@nih.gov*

Joan McGowan, Ph.D.

*Chief, Musculoskeletal Diseases Branch
National Institute of Arthritis and
Musculoskeletal and Skin Diseases,
National Institutes of Health
1 Democracy Plaza
6701 Democracy Boulevard
MSC 4872
Bethesda, MD 20892-4872
301-594-5055
301-480-4543 Fax
joan_mcgowan@nih.gov*

Ruth B. Merkatz, Ph.D., R.N., FAAN
Director and Team Leader, Women's Health
Pfizer, Inc.,
235 East 42nd Street, Seventh Floor, MS 7
New York, NY 10017
212-573-2721
212-309-0862 Fax
merkar@pfizer.com

Edith Mitchell, M.D.
Clinical Professor of Medicine
School of Medicine
Thomas Jefferson University
1025 Walnut Street
College Building, Room 1014,
Philadelphia, PA 19107
215-955-4652
215-923-9131 Fax
edith.mitchell@mail.tju.edu

John Molina, M.D.
Executive Director,
Las Fuentes Health Clinic of Guadalupe
8625 South Avenida del Yaqui, Suite 2
Guadalupe, AZ 85283
480-777-2263
480-777-2264 Fax
jwmolina@pol.net

Linda C. Niessen, D.M.D., M.P.H.
Vice President, Clinical Education
DENTSPLY International
3428 Amherst Avenue
Dallas, TX 75225
717-849-4465
717-849-4471 Fax
lniessen@dentsply.com

Estella Parrott, M.D., M.P.H.
Program Director, Reproductive Sciences Branch,
Center for Population Research
National Institute of Child Health and Human
Development,
National Institutes of Health
6100 Executive Boulevard
Building 61E, Room 8B-01
MSC 7510
Bethesda, MD 20892-7510
301-496-6515
301-496-0962 Fax
ep61h@nih.gov

Cynthia A. Pearson
Executive Director,
National Women's Health Network
514 10th Street, NW, Suite 400
Washington, DC 20004
202-347-1140
202-347-1168 Fax
cpearson@womenshealthnetwork.org

Andrew M. Pope, Ph.D.
Director, Board on Health Sciences Policy
The National Academies Institute of Medicine
2101 Constitution Avenue, NW
Washington, DC 20418
202-334-1739
202-334-1329 Fax
apope@nas.edu

Amelie G. Ramirez, Dr.P.H.
Associate Professor and Center Deputy Director
Chronic Disease Prevention and Control Research Center,
Baylor College of Medicine
8207 Callaghan Road, Suite 110
San Antonio, TX 78230
210-348-0255
210-348-0554 Fax
aramirez@bcm.tmc.edu

Roxanne Rodney, M.D.
Director, Clinical Research
Cordis Corporation
7 Powder Horn Drive
Warren, NY 07059
908-412-3902
908-412-3083 Fax
rrodney@cordus.jnj.com

Barbara Ross-Lee, D.O., Ph.D.
Vice President, Health Sciences and Medical Affairs
New York Institute of Technology
Northern Boulevard, Room 113
NYCOM I
Old Westbury, NY 11568-8000
516-686-3722
516-686-3830 Fax
brosslee@nyit.edu

Rosalie Sagraves, Pharm.D., FCCP

Dean and Professor, College of Pharmacy
University of Illinois
833 South Wood Street
M/C 874
Chicago, IL 60612
312-996-2497
312-996-3272 Fax
sagraves@uic.edu

Glorian Sorensen, Ph.D.

Professor, Department of Health and Social Behavior
Harvard University
677 Huntington Avenue
SPH3-705
Boston, MA 02115
617-432-4673
617-632-1999 Fax
glorian_sorensen@dfci.harvard.edu

Jeremy Sugarman, M.D., M.P.H., M.A.

*Director, Center for the Study of
Medical Ethics and Humanities*
Professor of Medicine and Philosophy,
Duke University Medical Center
P.O. Box 3040
Durham, NC 27710
919-668-9000
919-668-1789 Fax
sugar001@mc.duke.edu

Sandra Millon Underwood, Ph.D., R.N., FAAN

*American Cancer Society Oncology Nursing Professor,
School of Nursing*
University of Wisconsin, Milwaukee
1921 East Hartford
Milwaukee, WI 53211
414-229-6076
414-229-6474 Fax
underwoo@csd.uwm.edu

Martina Vogel-Taylor, Sc.B., MT(ASCP)

*Senior Advisor for Disease Prevention, Office of
Prevention*
Office of the Director,
National Institutes of Health
Building 31, Room 1B-03
MSC 2082
Bethesda, MD 20892-2082
301-496-6614
301-480-9654 Fax
martinav@nih.gov

Karina Walters, Ph.D., M.S.W.

Associate Professor, School of Social Work
University of Washington, Seattle
4101 15th Avenue, NE
Room 211G, Box 354900
Seattle, WA 98195-4900
206-543-5647
206-543-1228 Fax
kw5@u.washington.edu

Gerson Weiss, M.D.

*Professor and Chair, Department of Obstetrics,
Gynecology and Women's Health*
*Chairman, American Board of Obstetrics and
Gynecology*
*Department of Obstetrics,
Gynecology, and Women's Health*
New Jersey Medical School
University of Medicine and Dentistry of New Jersey
185 South Orange Avenue
Newark, NJ 07130-2714
973-972-5266
973-972-4574 Fax
weissgc@umdnj.edu

Robert A. Wild, M.D., M.P.H.

Professor and Chief,
Section of Research and Education in Women's Health
*Department of Obstetrics and Gynecology, Health
Sciences Center,*
University of Oklahoma
William Pavilion
920 Stanton L. Young Boulevard, Room 2410
Oklahoma City, OK 73104
405-271-5277
405-271-1956 Fax
robert-wild@ouhsc.edu

Janet Wittes, Ph.D.*President,*

Statistics Collaborative, Inc.

1710 Rhode Island Avenue, NW, Suite 200

Washington, DC 20036

202-429-9267

202-429-9268 Fax

janet@statcollab.com

Susan F. Wood, Ph.D.*Director, Office of Women's Health,*

U.S. Food and Drug Administration

Parklawn Building

5600 Fishers Lane, Room 16-65 HF-8

Rockville, MD 20857

301-827-0350

301-827-0926 Fax

Jackson T. Wright, Jr., M.D., Ph.D.*Professor of Medicine, Division of Hypertension,**Department of Medicine*

Case Western Reserve University School of Medicine

RB&C Building

11100 Euclid Avenue, Room 7311

Cleveland, OH 44106-5041

216-844-5174

216-844-1522 Fax

jxw20@po.cwru.edu

Darlene Yee-Melichar, Ed.D., CHES*Professor, Gerontology Programs**College of Health and Human Services,*

San Francisco State University

1600 Holloway Avenue

San Francisco, CA 94132

415-338-3558

415-338-3556 Fax

dyee@sfsu.edu

ORWH Liaisons**Margaret A. Chesney, Ph.D.***Conference Scientific Coordinator**Professor of Medicine,**University of California, San Francisco**Visiting Scientist, Office of Research on Women's Health**Department of Medicine,*

University of California, San Francisco

74 New Montgomery Street, Suite 600

San Francisco, CA 94105

415-597-9163

415-597-9213 Fax

mchesney@psg.ucsf.edu

Vivian W. Pinn, M.D.*Associate Director for Research on Women's Health**Director, Office of Research on Women's Health**Office of the Director,*

National Institutes of Health

Building 1

1 Center Drive, Room 201

MSC 0161

Bethesda, MD 20892-0161

301-402-1770

301-402-1798 Fax

pinnv@od.nih.gov

Joyce Rudick*Director, Programs and Management,**Office of Research on Women's Health**Office of the Director,*

National Institutes of Health

Building 1

1 Center Drive, Room 201

Bethesda, MD 20892

301-402-1770

301-402-1798 Fax

rudickj@od.nih.gov



graves fowler associates

d e s i g n a p p r o v a l — p r i n t

Please proof the attached carefully and sign below. Graves Fowler Associates will not be responsible for typos, changes, or corrections which are not marked.

job description **NIH: Women's Health Book**

job# **03NIH165**

- ☐ okay as is
- ☐ okay with corrections marked
- ☐ another set of revisions requested

signature

date



NIH Publication No. 03-5403

