Addressing Bias in Intervention Research

BETHESDA, MD
FEBRUARY 2001

Workshop Summary

WORKSHOP ORGANZING SPONSOR:

National Institute of Child Health and Human Development, National Institutes of Health
U.S. Department of Health and Human Services

The statements, conclusions, and recommendations contained in this document reflect both individual and collective opinions of the symposium participants and are not intended to represent the official position of the U.S. Department of Health and Human Services, or the National Institutes of Health.

Acknowledgements

Support for this workshop was provided by the NIH Office on AIDS Research and the Institute for Youth Development. This summary was edited by Christine Bachrach and Susan Newcomer, NICHD. The editors thank the panelists, who provided so much material for discussion, and the participants, who were indeed participants, not simply recipients of information. Sarah Brown, Director of the National Campaign to Prevent Teen Pregnancy; Kristin Moore, President of Child Trends; Shepherd Smith, President of the Institute for Youth Development, and Judy Auerbach of the NIH Office on AIDS Research provided invaluable intellectual and practical assistance as this workshop was planned. Joanne Jones of the Institute for Youth Development invested countless hours in the preparation of a full transcript of the workshop, which provided the basis for this summary. The editors are grateful to Shepherd Smith, Gary Thompson, Elaine Draper, Jacques Normand, and Christina Stile, who provided helpful comments on earlier drafts.

Christine Bachrach, Ph.D. and Susan F Newcomer, Ph.D., Demographic and Behavioral Sciences Branch, National Institute of Child Health and Human Development, Building 61e, room 8B07, Bethesda MD 20892, 301-496-1174
FAX 301-496-0962

Abstract:

A central goal of the field of youth intervention research is to conduct rigorous research that identifies which programs and approaches work, which do not, and why. Scientists working in this field face a variety of complex and sometimes controversial issues in the design and implementation of their studies. As a result, the potential for real or perceived bias in many areas of youth intervention research is high. Whether real or perceived, bias can undermine both the conduct of good research, and the process by which good scientific research is brought to bear in contributing to society and informing policy and programs. In February of 2001, NICHD held a workshop, *Addressing Bias in Intervention Research.* This event, was designed to address how these issue relating to bias affect youth intervention research, and, in particular, to address three major goals:

- Increase awareness of the many sources of real and perceived bias in intervention research and of the need to address them in the design and conduct of science;
- Articulate the ways in which investigators can guard against bias, measure it and report on it where it exists, and find ways of validating results when bias cannot be eliminated; and
- Through the above, provide a framework that can help to anchor discussions of potential bias in scientific results in domains where passions and positions are strong.

The concern with bias in intervention research is less one of dishonesty in researchers than the subtle and unwitting sources of bias that permeate life. No evidence exists that researcher dishonesty is more than an extremely isolated, occasional problem. However, “bias” carries a pejorative connotation. Many scientists do not want to identify and acknowledge sources of bias because doing so gives the appearance of being motivated by nonscientific concerns. It may be

more productive to speak in other terms. If we speak of perspective, point of view, or assumptions, it becomes easier to acknowledge, address, and debate the biases that all scientists have. This is a healthy, productive, and accepted part of the scientific process

Participants in the meeting included experts in the fields of prevention research, youth advocacy, pediatrics, adolescent health, reproductive health, substance abuse, violence, sociology, psychology, and research ethics.

KEYWORDS: Bias, Intervention Research, Ethics

Background

Chris Bachrach from the National Institute of Child Health and Human Development (NICHD) introduced the topic with several definitions of bias:

- Deviation of the expected value of a statistical estimate from the quantity it estimates; systematic error introduced into sampling or testing by selecting or encouraging one outcome or answer over others.
- Bent, tendency, an inclination of temperament or outlook; especially: a personal and sometimes unreasoned judgment; prejudice; an instance of such prejudice.[1]

As evident from these definitions, bias has several meanings, both scientific and nonscientific. In scientific terms, an estimate is biased if it systematically distorts the true value of a measured attribute or quantity. This type of bias can result from errors made in the process of observation, collecting data, reporting data, and the process of data processing and transforming data. It can also be built into the design of scientific studies: what is to be observed, what questions are asked, what hypotheses will be tested, how subjects are selected for a study, and how the control groups are conceptualized. This type of bias is always an issue in scientific research. However, bias takes on a pejorative meaning when we think that it results from dishonesty or prejudice. We blame the researcher who selects samples, designs measurements, or analyzes data in ways that systematically bias results toward ideological beliefs or toward results that will benefit something in which the researcher has a financial interest. While recognizing that no science is

value free, most scientists believe that an important value of science is to strive toward an ideal of objectivity. It is important to understand and acknowledge values underlying research and how values might influence the outcome.

Bachrach stressed that the workshop was intended to provide a forum for informal discussion of these issues. The workshop was not intended to suggest or shape NIH policy on any of these issues related to biases, or to express the views of NIH on topics that are not already addressed in formal NIH policy.

Session 1: the social and political context of research as sources of bias

Judith Auerbach, of the NIH Office of AIDS Research (OAR), introduced the first discussion topic: "How the political and cultural context in which interventional research is conducted may lead to bias in approach, implementation, analysis and interpretation of results, and how the structural characteristics of scientific communities may also lead to bias."

She charged the two speakers, Elaine Draper of the NIH Clinical Center's Department of Bioethics, and Jacques Normand, of the National Institute of Drug Abuse, to comment on how the social, political, ethical, and cultural phenomena that surround us affect the way in which we go about doing scientific research. She also stressed the need to address how the structure of scientific communities affects how researchers approach research: what are appropriate investigations to undertake, how decisions are made, and even fundamental understandings about what science is.

[1] Merriam-Webster On-line Dictionary, http://www.m-w.com/dictionary.htm

Elaine Draper focused on three questions. First, what good models do we have for acknowledging bias, for making our own points of view clear? Research on social problems is socially constructed. The social, political, and cultural environments shape how questions are asked and how we structure research to answer the questions. Two examples demonstrate this. First, consider the problem of health risk in the workplace. Do we define the problem of workplace risk in terms of individual risk factors (e.g., genetic susceptibility to risk)? If so, the community would invest its resources and energies in researching individual genetic susceptibility; we would be likely to find genetic susceptibility because we had framed the issue in that way. Or, we could define the problem of workplace risk in terms of workplace exposure hazards. In this case, we might find different answers and it might turn out that various types of structural, economic, or policy factors are critical.

A second example deals with highway deaths. Tens of thousands of people die each year on the highway. Asking "what is it about drivers that leads to all these highways deaths?" leads us to conduct research on drunk driving, human error, and the effects of talking on cell phones while driving 70 miles per hour. Research could just as reasonably concentrate on road design, the lack of public transportation, or auto engineering. Other ways of framing the question lead to very different kinds of research, different kinds of interventions, and different kinds of answers. Draper suggested that we should pay attention to and acknowledge biases up front and encourage open discussion of the effects that values, political perspectives, social organizations, and institutions have on intervention research.

Second, Draper asked whether our current disclosure policies are adequate. She argued that establishing standards will help the scientific community to acknowledge the diverse points of view investigators bring to research. This includes not only financial interests but also the organizational and institutional ties that affect intervention research. Scientific claims are not intrinsically credible, but have to gain credibility through some kind of social process. That social process draws on the credibility and values of the institutions that employ researchers and frame research. We need to examine official representations of research and consider the counterfactual -- that is, what else might have happened under other circumstances? What are the possible alternative constructions of the social problem, alternative theories and formulations of the research, and alternative types of data? What alternatives were considered and why were they abandoned. What alternatives never arose and why?

Draper's third question addressed how we might move beyond disclosure to limit and/or compel some kinds of scientific conduct. Sometimes requiring disclosure is not enough. At times, we must be able to impose sanctions to guard against the inappropriate influence of bias. Draper questioned whether scientists and other researchers have policed themselves adequately in this regard. Two examples illustrate circumstances in which scientists have moved beyond disclosure policies. First, scientific journals increasingly are refusing to publish work where informed consent was not obtained for various reasons rather than accepting disclosure of this information as adequate. Second, NIH now requires investigators to include women and minorities in clinical trials, or to come up with acceptable reasons for omitting those groups. If we accept that disclosure isn't always enough, at least for dealing with bias and conflicting

interests, what conduct do we want to limit, and what conduct do we want to compel, related to bias and research?

Jacques Normand drew on his experience as study director of an Institute of Medicine (IOM) panel that investigated the science around needle exchange programs for preventing blood borne infections, particularly HIV infections. He argued that concerns about bias in research are especially likely when three conditions are present: (1) a sensitive topic (e.g., something illegal or controversial); (2) ideological differences concerning the outcomes being examined; and (3) the research contains uncertainty. One type of uncertainty exists where, for ethical or logistical reasons, the science is unable to implement a gold standard double-blind clinical trial, and scientists must rely on observational or quasi-experimental designs. Another type of uncertainty is the failure of investigators or funders to be clear about the goals of their research. When these conditions exist, there is more room for people to look at the evidence in different ways and to come to different judgments about what it means.

Under these circumstances, it is necessary to take special care to ensure that conclusions drawn from the research are sound. In the IOM needle exchange project, choosing panel members who represented all perspectives allowed the panel to reach conclusions that had practical usefulness attached to them. At times, researchers disregard potentially adverse consequences of interventions they believe are beneficial, focusing instead on the positive outcomes. Opponents of an intervention approach make an important contribution by drawing attention to such overlooked consequences. Also, the IOM study benefited from forcing researchers to defend their findings and to investigate plausible alternative hypotheses. Rigorous debate among

scientists can help to sift through the merits of alternative hypotheses that could plausibly explain observed outcomes.

Normand emphasized that the cultural and political context of research, and the auspices under which it is conducted can have a critical impact on both the ability to implement sound research designs and the interpretation of research findings. For example, local drug paraphernalia laws that prohibit possession of needles prevent the inclusion of many communities in needle exchange research, leaving a biased sample of communities available for studies. Activist groups in some communities exert pressure on researchers to adopt practices that impair the scientific integrity of their studies. In cases where research is conducted for the specific purpose of validating or defending programs rather than for advancing general knowledge (as in the case of a postal service study that Normand described), study designs and procedures may be intentionally circumscribed.

Events external to the research can also jeopardize its credibility. For example, public statements from powerful political leaders or organizations can, intentionally or unintentionally, undermine findings from sound scientific research. At the same time, Normand criticized researchers for having naïve expectations with regard to how science influences policy. Scientists think that policy decisions should be based on science, but in the "real world" it does not always work that way. Scientists might be less frustrated if they were more realistic about the obstacles to basing public policy on scientific grounds.

Discussion

Discussion for Session 1 centered around three general questions. What types of social, structural, and cultural factors introduce bias into research? Can science "rise above" the world in which it is embedded? How can scientists respond to the challenges of addressing bias?

Factors that introduce bias

Influence on scientific research by social, structural, and cultural factors is ubiquitous, but such influences receive little attention. Biases inherent in the culture of science, the culture of politics, and the culture of program implementation may interact in complex ways. Values are created within scientific communities that influence both the questions that are posed and the outcomes that are considered. Social, economic, and cultural factors influence how people sort themselves into occupations such as "scientist" and "evaluator." As a result, there is homogeneity within scientific communities that this makes it more difficult to recognize that any bias exists at all.

The examples of both needle exchange and abstinence research illustrate that one will not necessarily, and probably not likely, find an even distribution of scientists engaged on all sides of an issue. In the case of needle exchange, the people who were widely publishing on the issue were typically proponents. A few of the opponents had addressed the issues in peer-reviewed journals but their numbers were much smaller. People who have been doing research on teen pregnancy and HIV programs tend not to be proponents of abstinence-only programs, but rather of comprehensive approaches to sex education.

Other structural factors matter too. The larger cultural and political context also affects funding streams. If this were a Muslim country, or if the United States had had 20 years of strong funding on abstinence only research, would our conversation about the types of bias in research be the same? Funders are increasingly demanding to see measurable effects of funded projects. Researchers worry about their results. Can one really do a job with integrity if all researchers are being turned into evaluators of their own work, and graded on the results?

The peer review system may also introduce bias. Hierarchies tend to be established among different scientific disciplines as well as among particular conceptual or theoretical frameworks within disciplines. All sorts of social processes go into establishing these hierarchies; the hierarchies have an important influence on the kinds of questions that are asked and the flow of funding to research. The composition of peer review panels also affects the types of research funded. For a number of reasons, very junior investigators are rarely on reviews, yet these investigators may bring a different perspective to the discussion than their more experienced peers.

Scientists need to be aware of the subtle environmental influences that introduce bias. One participant shared an example in which close ties existed between a university and an industry. Although most individual researchers at the university had no direct contractual or financial ties to the industry, the environment in which they worked still had a very heavy influence on their points of view. Such ties are some of the most important sources of bias, yet they don't get put

out on the table. It is often the critics of a researcher who make disclosures about these relationships.

The root of much bias in research is the disciplinary perspective that a researcher brings to a research problem. A study of a work program for low-income African-American adolescents conducted by a team of social psychologists illustrates this problem. Having identified being late for work as a major problem affecting the adolescents' success, the researchers framed a set of questions to study this further. Because of the researchers' psychology framework, they looked at questions such as whether the concept of time differed across groups. However, when other people looked at the problem; they looked at other factors. These other researchers found that transportation problems and the lack of reliable alarm clocks at home were responsible for the kids' tardiness. We often face this problem in intervention research. We start with a fairly constricted conceptual or theoretical framework and fail to even ask the most interesting or fundamental questions.

Finally, the goals of researchers differ which naturally introduces bias, in that what they study depends on their goals. The outcome measures used by a researcher studying abstinence-only programs would differ depending on whether he or she viewed the goal of these programs as pregnancy prevention, prevention of sexual intercourse, or prevention of all sexual activity. Sometimes researchers with different goals think they are in the same discussion, but it is actually not the same discussion because their desired outcomes are different.

Science as a thing apart?

Participants discussed the tension that exists between the goals of objectivity in science and the reality that the scientific enterprise is embedded in society. Can we separate science from its social, political, and cultural context?[2]

Some participants argued that by setting science apart as an independent, objective enterprise there is a danger of creating "spheres" that don't really exist. Scientists are religious; scientists are politicians; scientists vote. Scientists are never independent of their environments. On the other hand, others argued that the particular values, attitudes, and ideologies that scientists hold as individuals should be kept separate from what they do as scientists. Biases or personal prejudices may enter into their interpretations of the science and the implications they draw from it, but should not enter into the basic process of conducting the science. As one participant stated, "I don't think there is any place in scientific research for the right and the left. Scientific research is looking for answers to evidence in an objective way." We have to be sophisticated enough to realize that science and values are all jumbled together, while at the same time we ought to have a goal of keeping the science a distinct and separate component.

Participants recognized the challenges of living up to this standard. In order to frame scientific questions appropriately, we have to understand how different sides are looking at the issues. We have to involve different constituencies and communities; we have to listen to the people who have strong views on all sides of an issue. These challenges are more difficult if the basic

[2] Subsequent to the meeting, participant Kristin Moore provided the following definition of "science": "knowledge acquired by careful observation, by deduction of the laws which govern changes and conditions, and by testing these deductions by experiment; a branch of study, esp. one concerned with facts, principles, and methods; a technique based on training (the science of fencing); natural knowledge or a branch of this."

questions are not clear. It is one thing if the political world presents a clear question and the scientific world addresses it. It is another if the question is rather vague, where part of the scientific process is deciding which question or which outcome is really of interest and to whom.

The tension concerning the place of science in society also affects our view of the role of science in policy-making. Many scientists assume that science will, at some level, ultimately answer the questions relating to policy matters, and that the evidence derived from scientific investigation will prevail over other considerations. However, participants argued that science can never provide all the answers for informing policy. Beyond the scientific answers, there are values, moral issues, and other things that have to be taken into consideration in arriving at a political answer.

However, objective science does contribute in important ways. A group like the National Research Council can address the scientific question: is the evidence clear that the particular intervention is effective or not? The answer to this question can directly influence policy decisions, even if it is not the only factor considered. Also, science and scientific findings influence policy and social life indirectly. Science becomes part of the social process and affects people's attitudes and behavior, and the meaning of social constructs. These influences can have profound, if subtle, effects on policy-making.

Challenges of addressing bias

What are we doing or could we be doing to address some of these issues? Participants discussed two main approaches. We need to train scientists to become aware of their own values so they

consider the possibility that they are probably not thinking about all the possible outcomes of a process, or about the different ways a question could be framed. We need to provide better guidance about both what constitutes conflict and the potential sources of bias. We need good models for acknowledging and getting out on the table the values, bias, political beliefs, and other influences that inform our position. The issues of bias are relatively simple when you are looking at medications or vaccines; that is, you need to disclose whether you have a financial interest in a company that produces them. The issues are more complex when dealing with interventions, especially those dealing with sexual behavior or other sensitive topics. What should be disclosed in these cases, besides whether there is a financial interest?

Participants noted that many scientists do not want to identify and acknowledge sources of bias because it makes them appear motivated by nonscientific concerns. It may be too much to expect that scientists, in addition to saying their name and position, also lay out what brought them to their point of view. But we do need systemic ways of getting that information out as it helps us evaluate what the person is bringing to the table. Open discussion among scientists and other constituencies can help to bring out this information, as others raise these questions in an open forum. However, this strategy alone is unlikely to be sufficient.

In summary, participants agreed that it serves science to clarify our assumptions. Society, politics, and economic realities affect not only our selection of research topics and our framing of the questions, but also how we operationalize variables and how we interpret the data. Rather than see disclosure of biases as an embarrassment and potential source of conflict, we should see it as a process that will strengthen science.

Participants also addressed the challenges of ensuring that all points-of-view are able to inform a scientific project. In some cases, this means bringing scientists with a broad set of perspectives into the discussion. However, putting together a balanced panel of scientists is difficult when there are very few research scientists who have views at one end of the spectrum or another.

In addition, it can be difficult to convince people who hold views that are unpopular in the scientific community to dedicate time and effort to a scientific study. In participating, such individuals can contribute their own points-of-view, but may also risk being associated with a report that disagrees with or undermines their viewpoints. This situation could result in backlash from others who share their views.

Another approach to ensuring that diverse views will inform the scientific process is community consultation. Increasingly, scientists are involving communities in setting the research agenda and identifying the relevant questions to address. Consultations with providers and with people from the community can make the research better and may also help reduce the bias that we bring as scientists to community research and intervention trials.

While recognizing the value of this approach, participants noted several cautions. First, we must recognize that there are hierarchies and social structures within communities, too. When talking about doing community consultation we often go to the people who are in positions of power which also introduces bias. Community consultation must be done in a way that informs research, while recognizing the biases that can come from choosing the gatekeepers as

consultants. Also, community consultation is appropriate to some steps in the research process, but not to others. It can help to inform what questions should be asked and what approaches will be feasible, but should not override scientific methods and principles. Little research has been done on the process of community consultation, but this gap can be filled by studies that examine the implications of different approaches and effective strategies for integrating this concept into the scientific process.

Session 2: sources of individual investigator bias

Christine Bachrach (NICHD) introduced the second session, which focused on the ways in which the financial, ideological, or other interests of the individual investigator could introduce bias into research. The session also touched on existing policies and practices that help to identify and reduce such bias, and on further steps that could be taken. The two panelists were Gary Thompson, director of the Division of Grants Compliance and Oversight in the Office of Policy for Extramural Research Administration at the NIH and Suzanne Randolph, associate professor in the Families Studies Department at the University of Maryland, College Park.

Gary Thompson began with an overview of the regulations governing conflict of interest in research funded by the Public Health Service (PHS) through grants and contracts.[3] These regulations are intended to ensure that the design, conduct, and reporting of PHS-funded research

[3] The Office of Policy for Extramural Research Administration provides information on conflict of interest policy and access to the regulations governing grants (42 CR Part 50) at URL http://grants.nih.gov/grants/policy/emprograms/overview/ep-coi.htm . Regulations governing conflict of interest in contracts (45 CFR Part 94) are similar.

is not biased by a conflicting financial interest of an investigator. The regulations provide for identifying financial conflicts of interest, and ensuring that such conflicts are managed, reduced, or eliminated. Institutions that receive grant or contract funds are responsible for implementing their own institutional standards and procedures, but they must also satisfy the regulations. Experience indicates there is variation in the community as to how the regulations are implemented.

In the last few years, concerns raised by widely publicized incidents involving researcher conflict of interest led the NIH to enhance its oversight of compliance with the regulations. Although the NIH makes 40,000 to 45,000 grants per year, grantee institutions report surprisingly few cases of investigator conflict of interest. To gather information about how institutions were implementing the regulations and how implementation could be strengthened, the NIH conducted site visits to 10 of its major grantee institutions.[4] The site visit teams identified some strengths, particularly at institutions that had standing committees to provide guidance on issues around conflict of interest, and to help identify whether conflicts of interest existed in reported cases. However, they also identified several general problems.

First, there is a real need for improving educational outreach to ensure that faculty members have a clear understanding of their responsibilities. Investigators are required to inform their institutions of potential conflicts by completing a financial disclosure statement. Many faculty members were unclear about how their institutions defined a "significant financial interest."

---

[4] A report on the site visits is available at URL http://grants.nih.gov/grants/compliance/compendium_2000.htm

Some erroneously believed that their detailed financial information would be shared with the NIH.

Second, there is a need to address issues of trust. Faculty members were concerned that their institution or the NIH would misuse their financial information, using it as a rationale for reducing their salary or funding levels. Some faculty members also perceived a certain stigma around the word "conflict," viewing conflicts of interest as blemishes on their records. However, Thompson stressed, it is far worse if a conflict of interest comes to the attention of the public, but had not been properly disclosed. The NIH doesn't want to discourage researchers from engaging in entrepreneurship; there is nothing wrong with having a conflict of interest. However, conflicts must be disclosed in a way that protects the public's welfare, ensures the public trust, and maintains the integrity of the system.

Suzanne Randolph discussed the personal and scientific biases that may influence every aspect of the research process. She reminded the group that the research process is circular, in that it feeds back upon itself. We begin a research project with conceptual frameworks and theories that guide the work, but as we conduct research we reconceptualize on the basis of what we find or don't find. The individual biases we bring to research reflect our histories, including both our personal and scientific backgrounds. Our personal experiences shape our motivations and interests, while directing us down specific research paths and. shaping the theories we select. Whether we take a deficit approach or a deviance approach, whether we strive to understand difference or equivalence between groups, or whether we adopt what is being referred to as

culturally competent approaches, is all fueled by personal history. Our training as scientists also influences the choices we make in the process of conducting research.

Bias can enter into research at all stages. In conceptualizing research questions, Randolph recommended that scientists adopt integrative, cross-disciplinary theoretical models. To minimize bias in planning the research, investigators should involve not only scientists who hold similar views, but also opponents of the theory or approach, and community stakeholders and gatekeepers, all of whom have their own biases. Investigators should also integrate traditional theory with theories of how programs work to produce change. Someone said, "In theory, theory and practice go hand-in-hand, but in practice they do not." The theories and methods we bring to research as scientists may conflict with what is feasible and necessary in real life settings.

Randolph cited an example from her violence prevention intervention work in Washington, D.C. neighborhoods. In these communities, Head Start Centers must be equipped with bullet-proof glass because neighborhood gangs are shooting out the windows. The children at these centers can't go outside and play. Yet our standard approaches to research require observing the child on the playground. Bringing in the voices and ideas of community members and practitioners helps to address these conflicts between theory and practice. This helps not only with framing ideas, but also with making research feasible and realistic within actual contexts. This may extend to involving the community in the choice of intervention model.

The source of funding for a study may depend heavily on a researcher's personal circumstances and prior track record, but this too can have an impact on how the research is designed.

Research methods (e.g., how the study population is defined, how the sample is selected, how key terms are operationalized, what measures are used) are influenced by personal and disciplinary biases. For example, how does one define "parent" in a study of parents? An experiment Randolph conducted with her class yielded 14 different definitions. A major NICHD study of child care excluded families that did not speak English. What disciplinary or personal "frames" guided this decision?

The bias sometimes associated with single measures can be avoided by using multiple measurement approaches (i.e., multiple scales supplemented by observational or videotaped data).

Analysis and interpretation of data is another challenge. Even decisions such as how to group ages or ethnicities can be influenced by personal frames of reference. Involving people with a variety of viewpoints can help us interpret data appropriately. Without the help of multiple perspectives, our biases can lead us to overstate what we've learned.

Finally, there is bias inherent in how we report and disseminate our research: what audiences and publication outlets we target and how we present the findings. Researchers should reach out to non-traditional outlets for publication and presentation to obtain a broader range of reactions and insights about their research. They should also invest time in thinking about the lessons learned from their research and how these lessons can be disseminated, applied, expanded, and extended. In concluding, Randolph quoted an African proverb, "A single bracelet does not jingle." Investigators can do much to prevent the introduction of bias into research by collaborating with others whose training and perspectives differ from their own.

Discussion

Sources of Individual Bias

Discussion returned to two specific types of bias that affect research at the individual investigator level. In response to a question, Randolph identified the theoretical models and frameworks that are embedded in our disciplinary training as one of the most important sources of bias. Investigators have to be careful when developing the theoretical and conceptual aspects of a study because it may lead them down a certain trail. For instance, in the past theories about child development ignored racism and the impact that the experience of racism could have on parenting and children's development. Therefore, racism was not studied until evidence had accumulated to force the expansion of this theory.

Another type of bias is political or ideological. One participant observed that when researchers obtain findings that are consistent with their worldview, they accept them without much scrutiny. When they come up with findings that clash with their values or beliefs, they will often invest a lot of effort into re-analyzing them and attempting to make them "go away." Because most researchers share similar ideological views, it is difficult to develop research teams that provide diverse perspectives. An anecdote shared by one participant told of someone who "dismissed everything that came out of the National Academy of Sciences (NAS)." Despite the tremendous

efforts of the NAS to create balanced *interdisciplinary* panels, this person perceived the panels as unbalanced and biased by a common liberal worldview.

The group returned to the need for training to increase awareness of these issues. We can heighten investigators' awareness about these sources of bias by discussing them in textbooks and in classes. We can give researchers experiences and exposures that help them understand that bias exists, and help them to think critically about it. Objectivity does not have to be compromised by personal bias.

The American Evaluation Association publishes comprehensive standards for evaluation practice.[5] Those particularly relevant to issues of bias include: accurate and detailed descriptions of methods and approaches; discussion of "values, assumptions, theories, methods, results, and analyses that *significantly* affect the interpretation of the evaluative findings"; identification of financial, political, and career interests of the evaluator, client, and other stakeholders; and disclosure of roles or relationships that might pose a significant conflict of interest with the role of evaluator. Workshop participants noted that these standards are not consistently applied in everyday practice.

Institutional sources of bias

The group also stressed that financial conflict of interest operates at the institutional, as well as the individual level. Institutional aspects of bias are a good deal more challenging to deal with

[5] See URL http://www.eval.org

and a good deal more sensitive. We all know that connections make a difference. What are the critical connections? Participants provided many examples of institutional conflicts:

- Investigators at institutions that are dependent on outside funding may experience varying degrees of pressure from funders. Some funders require that investigators sign agreements not to release findings without prior approval. Some funders may try to influence what findings are released and how they are released. Investigators, recognizing that certain research findings may be politically sensitive or unpopular, may self-censor to preserve ongoing relationships with funders. These pressures may occur regardless of whether institutions are dependent on a single funder or on a diverse group of funders.
- The potential for conflict clearly exists when investigators work for organizations dependent on income streams that could be affected by their research results.
- Organizations that produce and market curricula for youth interventions, while also employing researchers who evaluate the curricula, create conflict for their researchers.
- Conflicts also exist where investigators conduct research in areas in which the institutions they work for hold strong philosophical positions.
- New NIH procedures that defer Institutional Review Board (IRB) approval until after the application has been scored also raise the prospect of an institutional conflict. If an application has received a fundable score, the institution will receive the money if their IRB approves the project, but will not if the IRB finds human subjects problems that cannot be corrected.

- Bias may also be injected by non-research institutions that become involved in research. Some schools and churches, for example, will not allow researchers to include measures of sexual behavior in studies in which they collaborate.

Currently, there are no regulations that speak to identifying and managing institutional conflicts of interest. However, the NIH will be exploring these issues as it conducts 10 more site visits to institutions in the coming year. It is incumbent upon the community to start addressing these issues themselves and dealing with them in ways that will satisfy concerns raised by the public and/or by Congress.

Participants had conflicting views on how we deal with these issues. Some participants had serious concerns about the potential conflict between their ethical obligation as scientists to publish findings as they are, and the implications of contracts that require review of findings by funders before publication. Others noted that most major academic institutions and research firms have very strict standards and do not sign contracts that could require the withholding of information from research. Some argued that researchers need to be aware of what "hat" they are wearing when they undertake a research project: are they acting as independent scientists with a mandate to explore a broad range of outcomes and implications, or are they conducting research for a narrower purpose (e.g., advocacy or marketing)? Others questioned the distinctions among different types of research, arguing that the pressures that bias findings exist across all research types. Still others suggested that researchers need to be aware of these pressures and take the extra steps necessary to communicate findings that are unpopular, startling, or contradictory in appropriate ways. If we communicate these results in a

constructive, as opposed to a sensational or negative, way we will be more likely to generate balanced discussion and further research.

All agreed that conflict of interest is a body with many heads. There are individual financial conflicts, institutional conflicts that may be financial or philosophical, individual ideological or political conflicts, commitment or stakeholder conflicts. Pressures from funders or scientific colleagues or reviewers may influence research and introduce bias. No one has a really good handle on many of these sources of bias and how to deal with them. The only regulations that exist are around financial conflict of interest. There is a need to improve our understanding of these other areas of conflict and the issues and problems that they pose in the community.

Evaluating interventions: the meaning and value of independence

Another critical issue for intervention research is the level of independence required for evaluation of a program or intervention, where a wide range of practice exists. In some cases, the same person who develops an intervention also evaluates it. In other cases, evaluation is done by a separate person on the same research team, or by a person working at the same institution. In still others, evaluators are employed by a different institution; such evaluation is funded separately from program implementation. Participants considered many issues that affect how they view the appropriateness of different arrangements.

Lack of independence between those who design and those who evaluate interventions may pose a threat to objective evaluation for several reasons. In the youth intervention area, the dangers

may be less likely to derive from financial stakes in the research than from investigators' emotional or theoretical commitments to a particular intervention design. Interventions that are shown to succeed seldom yield patents that fund their designers' retirements, but are very likely to enhance the reputation and prestige of their designers, as well as the credibility of the theories and methods they advocate. These outcomes lead to both financial and non-financial benefits in the form of promotions and job offers, new research funding, and lucrative speaking engagements. Even in the absence of such benefits, it may be impossible for researchers who have invested their time, effort, and creativity in the design of an intervention to maintain objectivity in its evaluation.

One participant argued that too much independence in program evaluation may have costs as well. Evaluators must be able to understand an intervention thoroughly to be able to evaluate it well.

Many participants felt that there are circumstances under which it is appropriate for the same person or team to develop and evaluate an intervention. When the purpose of the evaluation is to improve a program, rather than to compare its effects to that of an alternative program, it is appropriate for the designer to be involved in evaluation. Even when evaluations are intended to demonstrate program impact, the technical independence of the evaluation may be less important than how it is done. For instance, is it done with objectivity, solid methods, and accurate and complete reporting? It is certainly possible for a researcher to both design a program and evaluate what it does, and to do both very well. The problem is one of monitoring that process; the hazard is that it may not always be done well.

Providing access to accurate and adequate information on the design, execution, and findings of an evaluation, and making these data available to other researchers so that they can reanalyze the data and confirm the findings can provide a check on evaluations. Data archives are one positive way to do this. However, there are limitations: archives can help address the source of bias where an investigator has inappropriately analyzed the data or has drawn conclusions that are not supported by the data; they cannot address prior issues such as the framing of questions, outcomes examined, and use of the right measures. A second drawback is that there is very little reward for going back and analyzing someone else's findings. Funders need to provide support for such secondary analysis, as well as for methodological research that helps to evaluate and correct for the potential biases in evaluation research.

Although a well-designed evaluation by the intervention designer can be a good solid scientific piece of work, and although secondary analysis may help to confirm research findings, there are still inevitable potential sources of bias that can best be ferreted out by independent evaluation. Independent evaluation is particularly important before claims are made that the intervention is ready to go to scale. Participants made a distinction between independent evaluation of an intervention as implemented in its original site and independent replication of the intervention in different sites. Both were considered essential. However, funding for replication is limited. Programs very often stop with the evaluator who develops the program, but are not replicated or adapted for broader use.

One way to get around this problem depends on research that has identified the key principles of effective programs. As these principles are built into new programs, it is possible to assess the success of these key principles and concepts under different circumstances. This process can be enhanced if researchers developing or evaluating programs identify the core elements and principles on which the program relies.

Participants concluded by emphasizing that the concern with bias in intervention research is less one of dishonesty in researchers than the subtle and unwitting sources of bias that permeate life. The fact that we don't have very many findings of huge program impacts is one indicator that dishonesty is not a major problem. The term "bias" is highly pejorative and talking about bias seems like a personal attack on researchers. But if we think in terms of perspective, point of view, or assumptions, it becomes easier to acknowledge, address, and debate the biases that all scientists have. This is a healthy, productive, and accepted part of the scientific process.

Session 3: modes and methods of research as sources of bias

Kristin Moore, President of Child Trends, introduced the third set of discussion topics, which addressed biases related to modes of doing research, including payment of incentives, social desirability, and mode effects. The first panelist was Brian Wilcox, a professor of psychology and law at the University of Nebraska. The second panelist was Matt Stagner, formerly in the Office of the Assistant Secretary for Planning and Evaluation, and currently at the Urban Institute.

Wilcox observed that many researchers have, at one level, recognized issues of bias as being particularly pertinent to intervention research. But, he added, bias is not an issue that has been tackled head-on. He addressed several topics.

The first is the issue of how research designs themselves can result in certain forms of bias. Quasi-experimental designs are a case in point. A common problem with such designs is that of an inappropriate choice of comparison group. The selection itself may result in bias, if the comparison group differs on some unmeasured but essential variable. In general, researchers have not spent enough time analyzing how the selection of participants into intervention and comparison groups might result in bias, nor have they adequately modeled or accounted for potential bias.

Another design issue is differential subject attrition. Youth interventions that are focused on controversial topics present a good likelihood of differential attrition rates in the experimental and comparison or control groups. It is not always easy to predict which way the attrition will occur. Wilcox used as an example a study of abstinence education programs with attrition problems in fully 70 percent of the programs. Some had attrition in the experimental abstinence-only programs, as youth were leaving when their values were not consonant with the messages. There were also cases in which youth whose values were more consistent with the programs were leaving. It is incumbent on researchers to probe and to understand where their differential attrition is coming from and to understand that such attrition is going to be a very significant problem in many studies.

A second class of issues relates to what is and is not measured. Such gaps may be an attempt to avoid controversy, or a constraint of the time frame in which the research must be completed. Even more often, it is a failure to measure unintended effects. For instance, there is a very large literature in the juvenile delinquency area in which interventions have produced effects that are not just unintended, but also iatrogenic. When one assembles a group of youth who have run-ins with the law, group members may learn new bad behaviors from one another. A researcher who is an advocate for the program, theory, and model might not consider the possibility of negative outcomes.

A third set of issues deals with social desirability bias -- frequently ignored but often found in intervention research. A common source is post-test outcome measurements that occur immediately after the intervention. When post-test data are closely linked to the ending of a program, participants may be primed to give responses consistent with the program's goals. Results may be particularly biased when the program staff administers a post-test, when test administration is not perceived as confidential, or in interventions relying on moral persuasion.

This situation can be exacerbated by bias within the research measures themselves. Unbalanced value and attitude scales are one example. When all answers are positively (or negatively) phrased, it creates demand characteristics. A second round of responses, after such a biased one, may reflect a continuing bias. Those who have first taken what might be called the biased scale will show differences in their response to the unbiased scale later. In addition, the placement of

behavioral questions among attitudinal and other questions may result in ordering effects. These kinds of bias are rarely caused intentionally, but they can be important.

Matt Stagner discussed issues related to decision-making on what types of interventions evaluated, and how knowledge is assembled, especially the use of meta-analyses. He encouraged participants to keep the policy context in mind. In many instances the policy context of intervention research is complex and ambiguous. People have different views about what the appropriate interventions and the key outcomes are. Investigators have to be very clear about what research question or questions they are addressing. This concept is complicated when there are multiple agendas: researcher-owned theoretical perspectives that people want to advance, policy perspectives, and practice perspectives, which may or may not line up exactly with the policy perspective, and may or may not be very informed by a theoretical perspective. Balancing these perspectives creates a context that may not make bias more likely, but complicates our thinking about it.

The situation is further complicated because interventions evolve over time, and often the questions we ask about them evolve over time. If resources are committed to an evaluation and the intervention changes over time, is it up to the evaluator to comment on it, to adapt to it, or to combat it? Is it up to the program person to try as much as possible to stay close to that original intervention? Or should people use evaluation and other information and feedback to change the program as they go? Target populations may also evolve and change over time. Do we end up with a bias because we initially had one target population in mind, but that target has shifted over time?

Other issues include the appropriate level of analysis and the generalizability of random-assignment studies. Are the units of analysis (e.g., individuals, classrooms, schools) appropriate to the level at which the intervention is targeted? Are program designers and policy makers clear about the level at which the intervention is targeted? When we design random assignment studies, are we focusing so narrowly on the group that we can comfortably randomly assign that we end up creating a problem of generalizability?

Related to attrition is the payment of incentives, a way to recognize the respondent burden and compensate for it. There is good evidence that payment of incentives works to improve response rates; however, it may also encourage socially desirable responses by changing the relationship between the participant and the study. A related issue is the role of the interviewer in evaluation studies. The relation of the interviewer to the program and to the respondent, the respondent's perception of the interviewer and the confidentiality of the interview, the interviewer's level of involvement in helping the respondent provide answers, and matching of interviewers and respondents by age, sex, and gender are all issues to be considered.

Another issue, especially related to youth risk behaviors, is enhancing questions to make behaviors appear more normative. It is a common questionnaire development practice to ease somebody into a question with something like, "many people do X, have you ever done X?" In trying to elicit a true preponderance of the behavior in a population one wishes to avoid any anxiety about reporting, but if the behavior is made to appear normative does this bias the response? Does it counteract what the program thinks it is doing?

Finally, once there is more than one intervention study in a given arena, we must pay attention to the ways knowledge in a field gets summarized and presented to the policy community. In Stagner's opinion, summaries are more likely to affect policy than are individual studies. However, intervention research itself has improved more rapidly than the science of summarization. A key issue is what to include in a summary. Should summarizers include studies that are marginal in terms of significance or methodology? How do summarizers track down fugitive literature? Investigators may push harder for results to be published if they like the findings, while journals are sometimes reluctant to publish studies that find no differences or show no beneficial results. The answers to these questions and issue lie, in part, in improving the scientific review processes and giving attention to the independence of the reviewers, as well as to both the design issues and sampling issues present in creating these summaries.

Discussion

Discussion focused on challenges to good research, which may bias findings insofar as they lead to poor quality measurements and low response rates. Major topics included biases caused by attrition and faulty measurement, and by unscientific or haphazard approaches to meta-analysis.

Challenges to the integrity of research designs

One important issue related to attrition is the timing of data collection. In a school-based study, a fall pretest will get the largest and most representative population. A post-test conducted in the spring will not generally include those who have dropped out of school since the fall. Those who stay may differ from those who drop out, although one might argue that attrition bias should not differentially affect experimental and comparison groups. Whether or not kids stay in the program is another issue. One may want to ask how much difference a program makes for those who are treated, but that answer has to be interpreted in terms of the take-up rate and the retention rate. Attention paid to tracking at the start of a study can be of some assistance. Use of creative strategies, such as sub-sampling cases lost to follow-up, can help to minimize bias. Also, investigators can anticipate attrition bias before they set out on their research. They can develop a model of what kinds of things might be related to attrition and measure those things. This allows researchers to model what happened to a sample at the time of analysis.

Another type of bias related to study design and execution is contamination. About 10-to-15 percent crossover between experimental and control groups in a study to evaluate an intervention will seriously undermine the evaluation.

Faulty measurement can also bias results. Currently, research that is examining environmental as well as individual influences in behavioral studies is not finding substantial effects of the environmental factors. Is this because the research is not tapping the right environmental constructs and/or not measuring them adequately? Also, some researchers have found that scales using double negative items do not work well for individuals who do not speak standard English

as their primary language. For example, requiring a “yes/no” answer to a “I have never done X” question may confuse the reader.

Some bias can be introduced by the use of “filter questions” that limit the exposure of inexperienced respondents to detailed questions on sensitive and socially disapproved behaviors. Many researchers use these filters as a way of dealing with concerns from IRBs reluctant to approve questions that appear to be sanctioning inappropriate behavior. Filter questions can cause bias if they are based on inappropriate assumptions. For example, 10 years ago it was often assumed that if someone had not had penile-vaginal intercourse it was unlikely that the person had engaged in oral sex. That assumption no longer seems to be the case. Today, the use of vaginal sexual experience as a filter for questions on oral sex would probably produce biased results.

Participants discussed the challenges of identifying and measuring bias caused by methodological approaches. In fact, the discussion suggested that we are often wrong in our assumptions about what kinds of biases we create and why they occur. More often than not we are not thinking about the assumption, but are relying on our own personal cultural knowledge. Success in identifying and testing assumptions will be incremental. However, systematic research is helpful. For example, the National Survey of Family Growth conducted 18 experimental manipulations in its pretest. These types of studies should happen more often and should be published. Also, the NIH should support methodological research on these issues. Such research is difficult to get through peer review and is also complex and expensive.

Meta-Analysis

A second topic discussed was summarizing results of our research studies. One challenge is bias in the populations studied. For example, most studies of condom effectiveness are done in high-risk groups, because these groups provide greater power for detecting changes in incidence associated with condom use. However, there are some legitimate reasons that findings from high-risk groups may not be easily generalizable to the general population. When a number of different studies on varying risk groups exist, it is challenging to compile these into something that makes sense.

However, there are scientific ways to summarize, especially where there are enough studies to enable the capture of the relevant variables. Meta-analysis is research. It is like any other research in that one gathers data, analyzes data, and reports on data. It is subject to the same sort of problems as other research, including sampling problems and over-generalization. In the case of condom effectiveness studies, it would be appropriate to separate high-risk and general population studies for the purposes of the analysis.

Another inherent bias with meta-analysis or less formal summaries is sample selection. Often we are only analyzing those programs that show impacts, that is, those that have been published. As a consequence, we are essentially summarizing the most successful research results and presenting policy makers with summaries that may exaggerate the effects overall. It might be useful to set up some kind of central repository for null findings that can’t get published. Although investigators who design interventions that are shown to have no effect may feel

embarrassed by this, null findings can help in identifying what portion of an intervention approach works. These findings can help to determine whether components of an intervention can be dropped without reducing the intervention's effectiveness.