Journal of Consulting and Clinical Psychology
October 1996 Vol. 64, No. 5, 851-855

How the New NIH Guidelines on Inclusion of Women and Minorities Apply: Efficacy Trials, Effectiveness Trials, and Validity

Ann A. Hohmann
Division of Epidemiology and Services Research National Institute of Mental Health
Delores L. Parron
Division of Epidemiology and Services Research National Institute of Mental Health

ABSTRACT

The *NIH* (National Institutes of Health) *Guidelines on the Inclusion of Women and Minorities as Subjects in Clinical Research* requires investigators applying for NIH research funds to develop and, if funded, implement plans for the inclusion of women and minority populations in their research, when that research involves human participants. It is the purpose of this article to help investigators understand (a) the scientific context and rationale behind the *NIH Guidelines* ; (b) the NIH-defined concepts and the specific content of the *NIH Guidelines* ; and (c) how the intent of the *NIH Guidelines* is related to mental health services research, the most applied area of mental health research. The article also discusses where investigators can find additional information as they work to implement the *NIH Guidelines* .

The opinions expressed in this article are those of the authors and should not be construed as the official position of the National Institutes of Health.

Correspondence may be addressed to Ann A. Hohmann, Division of Epidemiology and Services Research, National Institute of Mental Health, 5600 Fishers Lane, Room 10C-06, Rockville, Maryland, 20857.
Electronic mail may be sent to AHOHMANN@NIH.GOV

Received: June 1, 1995
Revised: November 24, 1995
Accepted: December 1, 1995

Since the publication of the National Institutes of Health's (NIH's) Guidelines on the Inclusion of Women and Minorities as Subjects in Clinical Research (1994) , research investigators from all disciplines engaged in research with human participants have been struggling to understand and implement these

guidelines. It is the purpose of this article to help investigators understand (a) the scientific context and rationale behind the *NIH Guidelines* ; (b) the content of the guidelines and where additional information and assistance can be found; and (c) how the intent of these guidelines is related to mental health services research, the most applied area of mental health research. The other articles in this special section provide suggestions, with rich examples, for conducting outreach efforts to implement the guidelines.

This brief article can neither address all the concerns of investigators who want to apply for NIH funding nor provide all the solutions to difficult sampling and logistical problems that investigators will confront. Investigators are strongly encouraged to read the NIH documents that are cited in this article and that were written to help them understand and implement the guidelines. They are also strongly encouraged to read the guidelines themselves, which were published in the Federal Register (NIH Guidelines, 1994).

The *NIH Guidelines*

In March 1994, the policies of the NIH regarding the inclusion of women and ethnic minorities in study populations were significantly strengthened as a result of requirements stated in Subtitle B of Part 1 of the NIH Revitalization Act of 1993 (Public Law 103-43). The new policy states the following:

> It is the policy of NIH that women and members of minority groups and their subpopulations must be included in all NIH-supported biomedical and behavioral research projects involving human subjects, unless a clear and compelling rationale and justification establishes to the satisfaction of the relevant Institute/Center Director that inclusion is inappropriate with respect to the health of the subjects or the purpose of the research. Exclusion under other circumstances may be made by the Director, NIH, on the recommendation of an Institute/Center Director based on a compelling rationale and justification. Cost is not an acceptable reason for exclusion except when the study would duplicate data from other sources. Women of childbearing potential should not be routinely excluded from participation in clinical research. All NIH-supported biomedical and behavioral research involving human subjects is defined as clinical research. This policy applies to research subjects of all ages. ...
>
> Under this statute, when a Phase III clinical trial ... is proposed, evidence must be reviewed to show whether or not clinically important gender or race/ethnicity differences in the intervention effect are to be expected. ... (NIH Guidelines, 1994, p. 14509)
>
> NIH funding components will not award any grant, cooperative agreement or contract or support any intramural project to be conducted or funded in Fiscal Year 1995 and thereafter which does not comply with this policy. (NIH Guidelines, 1994, p. 14509)

These new guidelines preserve the guidelines of the NIH and the Alcohol, Drug Abuse, and Mental Health Administration published in the NIH Guide for Grants and Contracts in 1990 and add four major

new elements. These elements include the following:

- NIH must ensure that women and members of minorities and their subpopulations are included in all human participant research, not just clinical research.
- For Phase III clinical trials, NIH must ensure that women and minorities and their subpopulations are included so that valid analyses of differences in intervention effect can be achieved.
- NIH will not allow cost as an acceptable reason for excluding these groups.
- NIH must initiate programs and support for outreach efforts to recruit these groups into clinical studies.

There are a number of points about the policy that are fundamentally important for applicants to the NIH to understand.

First, this new policy reflects scientific, not political, concerns. Because the population of the United States is heterogeneous, the health needs and responses to treatment of individuals in the country must be assumed a priori to be just as heterogeneous. Without a scientific base of knowledge regarding human health, disease, and behavior that takes into account the diversity of this country's population, health care delivery, planning, and policy making is compromised because it is based on inadequate information and potentially misleading generalizations.

Second, the policy uses the definition of minority groups used by the Office of Management and Budget (OMB) for federal reporting. The four groups include American Indian/Alaska Native, Asian/Pacific Islander; Black/African American, not of Hispanic origin; and Hispanic. [1] The policy acknowledges the diversity within these broad categories by including the phrase "and their subpopulations" to encourage investigators to think more carefully about the possibly important biological and cultural differences that exist among the people included in those four broad categories.

Third, it is important to note that NIH does not expect all minority groups and subpopulations to be included in every study. The decision regarding what groups to include should be based on the scientific question under study. Broad representation is strongly encouraged, even if multiple centers, clinics, or sites are needed to accomplish it.

Investigators need to be aware that neither the cost of including diverse populations nor the geographic area of the investigator can be used as the sole basis for lack of representation. Investigators must select study participants and justify their inclusion in terms of the purpose of the research and other factors, such as prior research findings; relevant characteristics of and gaps in knowledge about the disease, disorder, or condition; and the feasibility of developing a collaboration or consortium or other arrangements to include minority groups. If there is limited representation of women or minorities, the investigator must provide a rationale satisfactory to the Institute or Center Director that is based on the

health of the participants and the scientific needs of the research being proposed.

This is clearly a difficult issue, particularly for more junior investigators living in geographic areas where the population is primarily majority. This issue and many others related to the implementation of the guidelines are addressed at length in an NIH document, *Questions and Answers Concerning the 1994 NIH Guidelines on the Inclusion of Women and Minorities as Subjects in Clinical Research* (NIH, 1994b). [2] In brief, any investigator living in a geographic area where the population is primarily majority and contemplating a Phase III clinical trial (see Point 5 later) needs to develop collaborative relationships with investigators in other geographic areas, unless similar research is completed or underway elsewhere employing more diverse populations. For other clinical research, investigators are still encouraged to develop collaborative arrangements. Junior investigators, who have yet to establish a track record in conducting research and administering research funds, should consider research questions of more limited scope that would make a sample-limiting rationale scientifically justifiable. [3]

Fourth, every application to NIH that includes human participants must specifically include in the research plan a description of the composition of the proposed study population in terms of gender and minority populations, a rationale for the selection, and a proposed plan for recruiting and retaining women and minority research participants. In addition, the human subjects section of an application must include consent procedures for these populations. Investigators should pay particular attention to the description of their plans for recruitment and retention of women and ethnic minority participants, including such information as prior experience recruiting and retaining the target population, collaborations with investigators who have this experience, and letters from relevant community groups expressing support for the research. Culturally sensitive and feasible outreach plans are critically important in an application and investigators are encouraged to obtain and read the NIH publication on outreach (NIH, 1994a). [4] The outreach plans will be considered by the review group in its assessment of the scientific merit of the application. Investigators should note that the NIH review committees will include all aspects of the subject selection, recruitment, and retention plans and justifications in their assessment of the scientific merit of a proposed study. Their assessment of these aspects will be reflected in the priority score. If the plans or justifications are considered inadequate, the application will be considered inadequate on issues of women and minorities and will not only receive a poorer priority score but will also be barred from funding. The bar cannot be lifted until adequate plans for recruitment and retention or more convincing justifications are developed.

Fifth, for this policy, NIH has developed a special definition of clinical trials that distinguishes this type of research from other types of clinical research supported by NIH and from trials supported by other agencies such as the Food and Drug Administration. NIH-defined clinical trials, for the purposes of this policy, are broad-based, prospective, Phase III clinical investigations, "usually involving several hundred or more human subjects" (NIH Guidelines, 1994, p. 14511), designed to evaluate an experimental intervention in comparison with a standard or control intervention or to compare two or more existing treatments. The aim of these investigations is often "to provide evidence leading to a scientific basis for change in health policy or standard of care" (NIH Guidelines, 1994, p. 14511). NIH-defined clinical trials include pharmacologic, nonpharmacologic, and behavioral interventions given for disease

prevention, prophylaxis, diagnosis, or therapy. [5]

The guidelines, in essence, are making a distinction between clinical efficacy and clinical effectiveness research, the latter most closely matching the guidelines' definition of clinical trial. The goal of both clinical efficacy and clinical effectiveness research is to define a causal relationship, using methods that lead to a valid estimate of the nature of that relationship. For efficacy research, however, the primary concern is internal validity; for effectiveness research, which builds from the results of efficacy research, external validity is primary. [6] This distinction is important, because it has important ramifications for the inclusion of women and minorities.

In designing an NIH-defined, Phase III clinical trial, the investigator must consider whether prior data indicate a potential association between gender or race—ethnicity and treatment outcomes. This association, and the strength of the data supporting or negating it, must be considered in the sampling design of a Phase III clinical trial:

- If the data strongly indicate that a significant clinical or public health difference in treatment effect may exist across gender or racial—ethnic groups, the trial must be designed to address the primary questions or specific aims separately for each group. The study design must allow valid analyses to detect intervention effects that are of clinical or public health importance. A valid analysis is one in which participants are assigned to groups in an unbiased manner, assessment of outcomes is unbiased, and unbiased statistical analyses are used to estimate intervention effects.

- If the data strongly indicate that no significant clinical or public health difference in treatment effect exists across gender or racial—ethnic groups, then gender and race—ethnicity will not be required as a participant selection criterion.

- If the data neither strongly support nor negate a significant clinical or public health difference in treatment effect across gender or racial—ethnic groups, the trial will be required to include sufficient and appropriate entry of gender and racial—ethnic subgroups so that a valid analysis or unbiased assessment of the intervention effect across groups can be made.

Thus, the Phase III clinical trial has the most rigorous standard regarding the inclusion of women and minorities. All studies involving human participants must include women and minority participants. However, a Phase III clinical trial must, in addition, provide valid analyses to measure differences of clinical or public health importance across gender and minority groups. Table 1 summarizes this difference.

The underlying logic for choosing Phase III clinical trials as the focus for the guidelines' most rigorous standard is embedded in the definition of the clinical trial. As previously noted, "Often the aim of such investigation is to provide evidence leading to a scientific basis for consideration of a change in health policy or standard of care" (NIH Guidelines, 1994, p. 14511). This statement makes clear that the ultimate goal of the guidelines is to ensure that research findings that are most likely to be used to change

the way care is delivered are applicable to the broadest range possible of people within the United States.

Health Policy, Clinical Trials, and Services Research

The assumption behind the focus on clinical trials is that only clinical trial data are used to influence health policy or standards of care. Also, although clinical effectiveness trials provide the strongest data for decision-making purposes, they are by no means the only data used or useful to health policy decision making. Clinical effectiveness trials are part of a domain of research called *health services research.*

Health services research, by its very nature, is driven by public health concerns. Much of the research, whether a clinical trials methodology is used or not, is, for the most part, directly relevant to how care is delivered and to health policy, and all of it is relevant to the external validity of future clinical effectiveness trials. The findings from services research help to provide the contextual information that researchers need to conduct meaningful clinical effectiveness trials.

Health services research examines the impact of the organization, financing, management, and delivery of health services on access to, process, cost, and outcomes of care. It also examines how characteristics of the individual, his or her family, and his or her social and cultural environment affect how, when, where, and whether a person will seek care; what types of care are chosen or provided; what happens during the delivery of care; and how satisfied the individual is with that care. Ideally, services research should also examine how the economic, social, political, and cultural environment of the service system and the providers within that system affect the organization, financing, management, and delivery of services and the impact of that interaction on access to, process, cost, and outcomes of care. Like basic behavioral and clinical research, the best services research starts from questions or hypotheses grounded in science and uses replicable methods, psychometrically sound instrumentation, and state-of-the-art analytic techniques. However, services research has one additional requirement: The questions must be as firmly grounded in public health and health policy issues as they are in science.

So an essential question in considering the new guidelines is, How do the new guidelines apply to a field of research that is methodologically much broader than clinical trials but of direct policy relevance and of critical importance to the validity of clinical effectiveness research or Phase III clinical trials? We argue that, to preserve the spirit of the guidelines, across all NIH-funded research, investigators posing services research questions should consider the public health relevance of their questions, the potential importance of their findings to health policy, and perhaps most important, the importance of their findings to future Phase III trials, when developing their sampling strategy. If these considerations are taken into account, then virtually all services research needs to meet the guidelines' Phase III clinical trials standard of sampling.

The need for a rigorous standard for sampling in services research can be easily demonstrated using two areas that do not involve clinical trials but are scientific building blocks for intervention or clinical trial designs: methodological and descriptive research.

Methodological Research

The methodological research vital to services research and, ultimately, to clinical effectiveness trials includes everything from the development of assessment instruments to testing applications of statistical theory. The development or refinement of instruments is an area particularly relevant to public health policy. Psychometrically sound instruments are critical to our understanding of what treatments or services work, for whom, and under what circumstances. If an investigator assessing the validity of an instrument does not specifically include women and minorities in the testing, the long-term consequences to science and public health policy could be profound. If, for example&comma in the development of an instrument, an investigator in Minneapolis did not oversample particular racial—cultural or socioeconomic groups or develop data collection sites outside his or her geographic region to test the validity of the instrument across population groups, the instrument would likely be invalid for important segments of the population. In clinical effectiveness trials, then, the investigators using the instrument would likely misestimate the effect of the intervention for those people for whom the instrument is invalid.

Descriptive Research

Another type of research that forms the foundation for clinical effectiveness trials is descriptive or naturalistic research. As with clinical trials, it is driven by theoretical or conceptual models, empirical data, and clearly articulated hypotheses. Unlike clinical trials, however, the goal of the research is not to test a causal hypothesis but to understand the medical, psychiatric, social, cultural, and economic conditions or context of the persons in need of and those using services. It seeks to illuminate what is actually happening in the service setting, how legislation, regulations, reimbursement methods, and insurance benefit designs affect how people use services and the quality of those services. It examines how clinicians treat patients from diverse backgrounds in various settings, how they respond to signs and symptoms, and how they make decisions about care. This type of contextual information about the patients, clinicians, and service system is imperative if what is learned in clinical efficacy trials is to be transferred to clinical effectiveness trials.

A simple transfer of efficacy results to applied research settings without this information may lead to a misestimation of the effect of the intervention. Although it makes sense, for example, to conduct an effectiveness trial with neuroleptics that have demonstrated clinical efficacy, unless the investigator understands how the service delivery system functions, how long case managers are likely to stay with an agency, the probability of comorbid conditions (such as substance abuse or mental retardation) among the clients, and the effect these conditions have on service use and treatment compliance, he or she may find no effect when in fact there is one. Another possibility is that the investigator may not be able to complete the study or have so much missing data that it would be impossible to draw any conclusions from the data. Thus, an understanding of the people and the context of their lives, which is provided by descriptive research designs, is a fundamental building block of effectiveness research.

If this basic, descriptive work does not include a broad representation of the population, then a distorted

view of the target population for the treatment is likely to result. The work to date in services research strongly suggests that many of these contextual issues do vary across and within cultural and racial groups. For example, different cultural groups interpret and react to symptoms and symptom clusters differently (Guarnaccia, Rubio-Stipec, & Canino, 1989). They have different expectations of service providers, are likely to use services differently, and use alternate sources of care (Hohmann, Richeport, Marriott, Canino, & Rubio-Stipec, 1990 ; Sue & Morishima, 1982). Unless we know what these differences are, the clinical effectiveness trial may fail to show a statistically or clinically significant difference between control and experimental conditions, but for reasons unrelated to the treatment itself. The reason for failure of a clinical trial may be as simple as the assessment method, which, for the target population, is perceived as threatening and intrusive or as complex as the need to provide transportation, baby-sitting services, and assistance in gaining benefits from the local office of Aid to Families With Dependent Children for a depressed, impoverished woman with children.

Conclusion

To conduct policy-relevant, externally valid clinical effectiveness or Phase III trials, the data that come from the broader scope of services research investigations are essential. Nonexperimental services research provides the methodological and contextual information necessary for valid trials through a careful balance of science and public health policy in the aims and design of the investigations. Because this research base forms the foundation of all clinical effectiveness trials, the Phase III clinical trial inclusion standards of the NIH Guidelines (1994) should be adopted by investigators conducting this research.

In essence, the strictest standards of the guidelines ask researchers to acknowledge that one size does not fit all, that there is a tremendous ethnic and cultural diversity in this country, and that to answer the critical clinical effectiveness trial question, "What works for whom, under what circumstances?" researchers must take that diversity into account.

At first examination, the guidelines appear to make the task for researchers more difficult, particularly for those who are interested in treatment effectiveness. It is our hope, however, that on reflection, researchers will see the guidelines and the accompanying outreach notebook (NIH, 1994a) as useful tools to enable them to build a better methodological and contextual base for, and to expand the external validity of, treatment effectiveness research.

References

Cook, T. D. & Campbell, D. T. (1979). *Quasi-experimentation:Design and analysis issues for field settings.* (Chicago: Rand-McNally)

D'Agostino, R. B. & Kwan, H. (1995). Measuring effectiveness:(What to expect without a randomized control group. *Medical Care, 33,* AS95—AS105.)

Guarnaccia, P. J., Rubio-Stipec, M. & Canino, G. (1989). Ataques de nervios in the Puerto Rican Diagnostic Interview Schedule:(The impact of cultural categories on psychiatric epidemiology. *Culture, Medicine, and Psychiatry, 13,* 275—295.)
Hohmann, A. A., Richeport, M., Marriott, B. M., Canino, G. J. & Rubio-Stipec, M. (1990). Spiritism in Puerto Rico: Results of a stratified island-wide community study.(*British Journal of Psychiatry, 156,* 328—335.)
National Institutes of Health. *NIH guide for grants and contracts* (Vol. 19, Issue no. 31, pp. 18—19). Bethesda, MD: Author.
National Institutes of Health Revitalization Act, U.S. Code, Vol. 42, secs. 289a—2.
National Institutes of Health. *Outreach notebook for the NIH guidelines on inclusion of women and minorities as subjects in clinical research.* Bethesda, MD: Author.
National Institutes of Health. *Questions and answers concerning the 1994 NIH guidelines on the inclusion of women and minorities as subjects in clinical research.* Bethesda, MD: Author.
NIH guidelines on the inclusion of women and minorities as subjects in clinical research. (59, Fed. Reg. 14, 508 (Document no. 94-5435)) (1994).
Standards for the classification of federal data on race and ethnicity. (60, Fed. Reg. 44, 673 (Document no. 95-20787)) (1995).
Sue, S. & Morishima, J. K. (1982). *The mental health of Asian Americans.* (San Francisco: Jossey-Bass)

1

Readers should note that the OMB is currently reviewing these racial—ethnic categories, which were adopted in 1977. During the past several years, these categories have come under increasing criticism from those who believe that they do not reflect the increasing diversity of the country's population. Summary of issues and suggestions raised by public comment appear in the *Federal Register* , August 25, 1995, pp. 44674—44693. The issues raised in the OMB's *Federal Register* notice might help investigators understand the diversity of cultural and ethnic identification that exists even within the 1977 race and ethnicity categories.

2

This document is available electronically from the NIH Gopher site under "From the Office of Extramural Research" and from the NIH Grantline: data line, (301) 402-2221semi; John James, moderator, at (301) 594-7270 or via Internet at jqj@cu.nih.gov.

3

Applicants, with specific questions or concerns, can also seek guidance from NIH staff within each institute or center or from the NIH Office of Extramural Research at (301) 594-7270 or via Internet at jqj@cu.nih.gov.

4

The underlying elements of outreach, as defined by the NIH publication, are to (a) understand the study population; (b) establish explicit goals for recruiting and retaining participants; (c) achieve agreement on research plans among researchers, clinicians, and community members; (d) design and conduct an evaluation plan of the recruitment and retention strategies; and (e) establish and maintain communication with all participants in the study (research staff, clinicians, participants, their families, and the community). Each of these is discussed at length in the publication.

5

For most researchers who use the phrase clinical trial to include any controlled research testing the effect of an intervention on human participants, this distinction between "clinical research" and "clinical trials" can be confusing. The intent of the guidelines was to single out the Phase III clinical trial as a special case for the purpose of selecting samples for study; the intent was not to imply that other types of research would not or should not use randomized clinical trial methods or are not, in a more generic sense, clinical trials.

6

The relationship between types of research and types of validity are important in this context. What follows is a brief exposition of these relationships.

To ascertain cause, the investigator needs to know that the variables covary, that the causal variable is temporally prior to the outcome variable, and that other variables can be ruled out as causal. This third element is the essence of internal validity (Cook & Campbell 1979) and should be the primary concern in designing a traditional treatment outcome or efficacy trial. A randomized clinical trial (RCT) is ideal for this purpose, because random selection of participants, who typically possess a limited range of illness and other characteristics, should make the groups equal on all variables except the intervention (D'Agostino & Kwan 1995). External validity becomes the primary concern when the goal moves from evaluating treatment efficacy to evaluating treatment effectiveness. External validity is the degree to which results can be generalized across types of persons, settings, and times. Typically, through the use of quasi-experimental designs and sophisticated design and statistical controls for potential biases, treatment effectiveness research provides an understanding of how population heterogeneity and life context affect the causal relationship between treatment and outcome.

Table 1.

Table 1

Inclusion Requirements in NIH (National Institutes of Health)-Sponsored Studies With Human Participants

Clinical research or trials	Inclusion of women and minorities	Inclusion of minority subpopulations	Required to measure differences
Clinical research	Yes	Yes	No
Clinical trials	Yes	Yes	Yes

Table 1

Inclusion Requirements in NIH (National Institutes of Health)-Sponsored Studies With Human Participants

Clinical research or trials	Inclusion of women and minorities	Inclusion of minority subpopulations	Required to measure differences
Clinical research	Yes	Yes	No
Clinical trials	Yes	Yes	Yes