Reviewer's report

Title: Ethnicity and the ethics of data linkage

Version: 1 Date: 29 August 2007

Reviewer: David R. Nerenz

Reviewer's report:

General

-------------------------------------------------------------------------------

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

There are no major revisions that need to be made. The author has clearly stated several concerns about a specific approach to linking data on ethnicity to data on health care utilization. The concerns are valid, and the author’s recommendations do not go beyond a recommendation for caution in future applications of data linkage technologies.

-------------------------------------------------------------------------------

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

None

-------------------------------------------------------------------------------

Discretionary Revisions (which the author can choose to ignore)

The author is commenting on a data linkage initiative that is set in the context of data sets, ethnic group identification, and privacy and patient consent policies in Scotland. His points are valid in other contexts as well, but some features of other medical settings, and some alternative approaches to data linkage, may deserve further comment. These include:

1. Use of indirect methods for assigning race/ethnicity categories to individuals. In some settings (e.g., areas of high residential segregation) to infer race or ethnicity with reasonable accuracy using information like home address or surname. It is not necessary in these situations to use information on ethnicity provided by people in census returns or in registration for medical care. The authors' concerns about use of personal information in general still apply here, but there are ways to "bypass" the need for use of information on ethnicity per se. It may also be possible to analyze patterns of care to identify racial or ethnic disparities without using any individual-level data at all, but rather using data aggregated at the residential block or region level to develop variables characterizing areas or neighborhoods rather than individual people. Some of
these approaches raise fewer ethical concerns than the approach described here, but all use, in one way or another, information provided in the context of a census, for purposes other than those that people may have originally understood. Is there a clear line we can identify where specific ethical concerns arise, or is any use of data beyond its original stated purpose problematic?

2. Many of the data collection processes and associated privacy and consent issues arise here in the context of a public health program that is effectively a monopoly; that is, individuals for the most part have no choice but to participate. In other countries, though, health insurance and health care provision occur in private settings characterized by considerable freedom of choice. One feature of that choice can be the presence of research activity in the organization, and patient consent to having data used in research. In the U.S., some hospitals and insurers have informed people, at the point of enrollment or registration, that various types of personal information may be used for research purposes without their explicit consent for each project. Joining that health plan, or receiving care at that facility, involves the voluntary agreement to use of personal information for those purposes. Those who object have the option to go elsewhere. Ethical concerns are not eliminated, but are different in both type and extent, in these circumstances.

3. The author may wish to expand the discussion on the extent to which patients who fill out census forms have any ability to refuse to provide information, or to assert any control over how the information is used in the future. I suspect that in most countries, participation is mandatory and that essentially no control is possible over future use of information. If that is generally true, then the author’s points about “fully informed discussion” may not go far enough. The only protections that provide any real privacy or data security may be those formally written into law and public policy, not those just debated in public forums.

What next?: Accept after discretionary revisions

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that I have no competing interests.