Reviewer's report

Title: Identification of Families for Genetic Studies of Birth Defects Using The National Health Interview Survey

Version: 1 Date: 16 October 2003

Reviewer: Sheryl Larson

Reviewer's report:

I think this paper is important but before it can be published, the author's need to provide much more specificity about their methods because the NHIS-D is a very complex set of surveys with many intricate skip patterns.

General
• Is the question posed by the authors new and well defined?

The question is important, and new. However, I thought that the paper was going to show how the people who responded to the NHIS-D were identified for further study through additional contacts rather than simply working within the data provided by the NHIS-D. This is a relatively easy thing to change by modifying the title.

• Are the methods appropriate and well described, and are sufficient details provided to replicate the work?

The methods are not described in sufficient detail. Specifically, I think the authors worked with the condition file. However, that was not specifically stated. Also not described was if and how the authors connected the condition file to the other NHIS-D files for analysis.

• Are the data sound and well controlled?

The NHIS-D is a very good data set compiled by the US Census Bureau. As the authors correctly point out, there are some important limitations in the data set but that is the point of the paper so it is not a limitation per se.

• Does the manuscript adhere to the relevant standards for reporting and data deposition?

The authors conducted secondary analyses of a public data set.

• Are the discussion and conclusions well balanced and adequately supported by the data?

The conclusions are reasonable given the findings.

The authors list several limitations to the data. One additional limitation is that while the NHIS-D provided information about ICD codes in great detail, subsequent years of the NHIS do not provide nearly the amount or detail. It would likely be impossible to even replicate this study in more recent years of the NHIS.

• Do the title and abstract accurately convey what has been found?

The title was a bit misleading because it led me to believe that the authors were going to actually collect new data from people identified by the NHIS-D.

• Is the writing acceptable?
The writing is fine.

Discretionary Revisions (which the author can choose to ignore)

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

The title should be modified so it is clear that the paper uses the NHIS-D to estimate prevalence of various categories of birth defects and households in which one or more family member has a birth defect.

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

This manuscript does not clearly state which of the many NHIS-D files the authors used for the analysis.

The number of people identified as having spina bifida was 55. In another published paper using the NHIS-D, 120 people with spina bifida were identified (Larson, Lakin, Anderson, Kwak, Lee & Anderson, 2001). In that paper, information from the condition file was combined with information from the Phase 1 Disability Supplement and the Core Survey to identify cases. Many people were reported in Phase 1 to have spina bifida who did not have a condition record of that because they did not report having a limitation in daily activity related to their condition. In the NHIS-D, to have a condition listed in the condition file, the person has to report that condition is the primary cause of a limitation on general activities or in one of several specific activities or they have to be in the 1/6th sample that were asked about specific conditions. Some of the conditions the authors studied were mentioned on one of the six the condition lists used in the core survey meaning that 1/6th of all NHIS-D respondents reported directly if they had those conditions. The other 5/6ths of the sample did not answer those questions. Some households or family members could have a condition which is unreported because they were asked a different condition list. In addition some of the condition lists ask if the person has ever had the condition (e.g., congenital heart disease), some ask if the person has had the condition in the last 12 months (e.g., diabetes) and others ask if a person now has a condition (e.g., a cleft palate or harelip). These all affect a researcher's ability to estimate prevalence. The authors should report how they handled these intricacies in the data.

The manuscript does not specify how the weights for the two years were combined so that the population estimates were appropriate. This seems to have been handled ok based on the estimates made but additional text in the methodology section would make it clearer to readers that this has been done.

In the discussion of Table 2, the authors state that there were significant gender differences between the groups with a higher percentage of females in the multiple familial group. What is confusing to me is that they report an odds ratio of 1.8 with a 95% confidence interval of 0.7-5.1. If the confidence interval includes 1.0 how can this difference be statistically significant? The same is true for socio-economic status where the odds ratio is 1.8 and the confidence interval is 0.5-9.0.

Only 8 households (16 individuals) were in the multiple familial group raising the question of whether the sample size is sufficient to make population estimates at all. The authors should provide standard errors or relative standard errors for the population estimates on the tables instead of offering to provide them upon request so that the readers can see if the relative standard error exceeds 30%.
Reference

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

Level of interest: An article of importance in its specialized field or of broad interest

Quality of written English: Acceptable

Statistical review: No