Commentary for GSS Workshop
Methodology &
Technological Innovation and Cyberinfrastructure

Jeremy Freese

Harvard University

and

University of Wisconsin-Madison

Methodology

The General Social Survey is an unquestioned national resource and has become this as a result of the excellent stewardship and continual initiative shown by its principals. As someone who cares about social research methods, my principal complaint about the General Social Survey is perhaps that in some ways it is a victim of its own success, at least within sociology. By which I mean that its very prominence can invite problems of use that result in a number of projects that use the GSS being regrettable from a scientific standpoint in various ways. While I recognize that large-scale data collection efforts are often considered having as their purview being simply getting the highest quality data they can, and pathologies of use are not their problem, I do wonder whether more can be built into the structure of the GSS that would improve the quality of work being done from an analysis side as well as from the data collection side. Three points:

1. Two closely connected problems that I see with the practical use of the GSS is that its familiarity to sociologists leads it to be pressed into service for questions for which other data sources exist that would be better to use and that, despite the mostly descriptive emphasis on documenting trends that was emphasized in the materials we were provided, sociologists sometimes seek to use the GSS to speak to causal questions in ways that cross-sectional surveys are not well-designed to speak to. In this respect, I find special reason to be very excited about three of the developments described in the materials: (1) the continued engagement with ISSP, which will allow theoretical discussions of large-scale causes of social trends to benefit from drawing implications about the timing of changes in different nations (although this is limited by whatever ways the ISSP has unique rather than shared content across years, and so more repetition in the ISSP would be desirable); (2) the development of a panel design to the study, although I would like to know more details about this (which will be desirable to allow inferences about intervening events -- for example, changes in marital status -- especially given the wise decision to attempt to gather three points in time on respondents); (3) the possible expansion to include more administrative record matches (which can have the advantage of extending information about respondents across time). I commend these efforts. In this respect, anything NSF could do to specifically promote use of these components, whether in funding specific projects or in funding infrastructure that would make this data as easy to use as the regular GSS cross-sections are now, could be valuable for improving the overall quality of work done using GSS.

2. Especially given increasing emphasis in the behavioral sciences on interdisciplinary work and on being able to describe "mechanisms" and "pathways," the lack of much good measurement on

psychological constructs in the GSS seems like it might post a threat to its value to interdisciplinary contribution going forward. Such measurement is more commonly conducted by SAQ rather than in-person interviews. The Health and Retirement Study, for instance, has invested now in collecting psychosocial measures using a leave-behind questionnaire. The lack of any kind of leave-behind SAQ with GSS has always been somewhat puzzling to me, and seems especially the case now that respondents will be impaneled. For that matter, even an Internet-administered SAQ for those respondents who have Internet access (even recognizing the limitations of such samples then being limited to only respondents who have Internet access), would provide more strength for getting at psychological moderators and mediators of causal processes than what can presently be done with GSS.

3. In economics, there has been increasing prominence of methods of causal analysis that are based less on structural modeling than on identifying 'natural experiments' within the data, especially those that derive from identifying discontinuities in causes and from identifying sources of exogenous variation (instrumental variables). A less obvious implication of this movement is perhaps the interest of the GSS in making available as much so-called paradata (if I am correct in understanding this term to refer to data on the details of fielding and other characteristics of cases), administrative data matches, and geography as possible, even if its analytic implications are not readily apparent, as such data may have uses for being able to support kinds of inferences along these lines that those running the survey might not themselves anticipate.

Technological Innovation and Cyberinfrastructure

My points are about cyberinfrastucture for users of the General Social Survey. As a broader cyberinfrastructural matter, the General Social Survey's web presence is notoriously confusing; in my own effort this weekend to search for materials, there appear to be two different GSS pages at NORC and two different kinds of GSS pages at ICPSR, which link to different materials. Given what a valuable national resource it is, I am unclear why GSS's web presence seems so much less well-developed than either NES or PSID and worry especially that this may impose a hindrance to investigators outside sociology.

1. *Regarding data availability, when supplemental data collection is done.* The "Generation of Data" material lists some GSS Reinterviews and Auxiliary Data Collection projects. Some of these I was already aware of; some of which I was not. I searched for information online about obtaining these data and came away frustrated and empty-handed. Are these data systematically available to the research community? If not, why not? Can anything be done to make information about the existence of these data more widely known and the data themselves more readily accessible?

2. *Regarding analysis of confidential data.* The advancement of the GSS as a scientific tool I think will require it to be collecting and making more data available to researchers than what disclosure requirements can readily allow to circulate publicly. I would be interested in knowing more about any strategies GSS has regarding the release of data that cannot be made available publicly; indeed, I'm only vaguely and anecdotally aware of what policies are now and was not able to locate anything on the Web describing it. (In this respect, I have *far* greater enthusiasm

for solutions that make data available on a secure basis to be analyzed using any statistical software than solutions that have attempted to develop specialized statistical packages for online use, especially when the latter do not allow analysis to be conducted in any way that provides a record of steps that is subsequently replicable.)

3. *Regarding data agreements.* As the materials we were provided emphasize, one of the great strengths of the GSS is that it is open and public. I have on multiple occasions now checked findings of a paper using GSS either as a reader or reviewer of the GSS by re-analyzing GSS data myself. One of the great advantages of the Internet is that it allows a considerable increase in the transparency it allows for research. Economics has taken advantage of this transparency by having now enforced rules in its most prominent journals that code and data for analyses are to be made available to the maximum extent possible at the time of publication. Sociology, despite once been at the forefront of initiatives for sharing materials, has lagged behind. A costless way for the General Social Survey to show leadership in this area would be to ask its researchers to show the same kind of attitude toward the desirability of openness and public character of work that the GSS itself exemplifies. Namely, as part of its agreement for users downloading the data, the GSS can inform users of the expectation that users will deposit software code sufficient to replicate results from GSS analyses in the ICPSR Publications-Related Archive, or another public archive that accepts code, at the time of the publication of any articles that use the data. In doing so, the GSS can expand its teaching mission as the availability of code for inspection allows new researchers to see first-hand the procedures by which published work was done, and in so doing gain lessons for their own practice.