
THE SOCIOMETRICS SOCIAL SCIENCE DATA AND INSTRUMENT LIBRARY: A NATIONAL RESOURCE FOR TEEN PREGNANCY RESEARCH AND EVALUATION

The conducting and publishing of research based on secondary analysis of existing data sets has significant potential for advancing the field of teen pregnancy research. Existing data can be reanalyzed to validate the work of the original investigator. In addition, good data often lend themselves to extension of the original work by examination of topics other than those pursued by the original investigator or by examination of the same topic from the perspective of additional data sets. Because the data already exist, pressing issues of policy or practice can be addressed without the long delay necessitated by original data collection. In addition, the use of the same data set by multiple investigators as well as the use of multiple data sources by a single investigator, foster the building of a network of researchers, thereby strengthening the research community.

Existing data can also be used for evaluation purposes. For example, several national data sets (such as the 1988 and 1990 National Surveys of Family Growth; the 1988 and 1990-91 National Surveys of Adolescent Males; the 1988 National Health Interview Survey on Child Health; and the 1990 and 1991 National Household Surveys on Drug Abuse) can be analyzed to derive comparison standards against which to compare outcome data from prevention programs aimed at reducing teen pregnancy, HIV/AIDS, tobacco use, alcohol use, and drug abuse. Such comparison statistics can be used in lieu of, or in conjunction with, collecting new data from a local comparison or control group (Card & Reagan, 1989; Schott,

Peterson, & Card, 1993). Because the data yielding the comparison standards already exist, evaluation costs are minimized. This is a crucial consideration when evaluation-related funds are limited, as they almost always are.

Finally, the design, instruments, and data from model research and intervention programs can be studied to help guide the development, implementation, and evaluation of new programs.

This article introduces teen pregnancy researchers and evaluators to the Sociometrics Data Library, a rich source of high-quality health and social science data sets. The Data Library currently consists of five topically-focused data archives:

- the Data Archive on Adolescent Pregnancy and Pregnancy Prevention (129 studies comprising 196 data sets and over 70,000 variables);
- the American Family Data Archive (14 studies comprising 36 data sets and over 20,000 variables);
- the Maternal Drug Abuse Data Archive (7 studies comprising 13 data sets and over 5,000 variables);
- the AIDS/STD Data Archive (12 studies comprising 20 data sets and over 16,000 variables); and
- the Data Archive of Social Research on Aging (3 studies comprising 22 data sets and over 19,000 variables).

The first four of these data archives can speak in powerful fashion to teen pregnancy research and evaluation issues.

The data collections comprising the Sociometrics Data Library have been designed to include only data sets of the highest quality and to facilitate analysis of these exemplary data sets by both experienced and novice analysts. Each of the over 200 data sets in the Data Library has been selected for inclusion by a National Advisory Panel of experts in the archive's topical focus. Strict scientific criteria of technical quality, substantive utility, policy relevance, and potential for secondary data

analysis have been used to select data sets.

Each data set has been processed and documented in a standard way, facilitating use. Once a user has worked with one data set in the Sociometrics Data Library, it is easy for him or her to work with any of the others. Each data set is made publicly available with five standard machine-readable files: (File 1) a raw data file; (Files 2 and 3) machine-readable SPSS and SAS program statements that fully document the variables and values in the data file; (File 4) an SPSS data dictionary; and (File 5) SPSS frequencies.

A printed and bound User's Guide accompanies each data set. The Guide is itself composed of a standard set of sections and subsections that describe the survey from which the data originate; display all variable names and labels, organized by Topic and Type; display a crosstabulation of Topics and types in the data set; provide frequencies for approximately 20 key variables that are included in most data sets (e.g., race, gender, marital status, etc.); provide the results of illustrative data completeness and consistence checks; and provide additional documentation and user notes as needed. Optional purchase of the original instrument and codebook is offered for each data set.

Search, retrieval, and extract software accompanying CD-ROM versions of the data sets allow a user to identify variables of interest within and across data sets, to save located variables as a search set, and to create smaller analysis files with only the selected variables of interest, thus minimizing hard-disk storage requirements and maximizing speed of analysis turn-around.

A list and description of the data sets in the Sociometrics Data Library can be obtained from the author. Free technical assistance on use of the Data Library is also available.

*J. J. Card, Ph.D., President
Sociometrics Corporation
Tel. 415-949-3282*