

## **GUIDANCE ON MAINTAINING PARTICIPANT ANONYMITY AND ENSURING LONG-TERM CONFIDENTIALITY OF INFORMATION**

### **Using “Pseudo-Codes” to Ensure Anonymity**

In study designs in which data collection is accomplished strictly through repeated paper or internet surveying or testing (but not interviewing), it may be possible to anonymously collect and link data from various data collection instruments and/or over a multiple data collection events during an extended time period without the use of a code list that links responses to participants' identities. This can be accomplished by instructing the participant to create his/her own “pseudo-code,” a code that can be accurately recreated each time the participant completes a different questionnaire/test. The pseudo-code is created by instructing the participant to answer questions that he/she is likely to answer the same way each time he/she is asked. For example, questions may include:

- ◆ What is the second letter of your mother's maiden name?
- ◆ What is the fourth digit of your social security number?
- ◆ What is the last letter of the name of the state in which you were born?
- ◆ What is the last digit of your home telephone number?
- ◆ What is the second letter of the month in which you were born?

In this example, Maggie Jean Smith, who was born in Delaware in March 1983, whose mother's maiden name is Markham, whose Social Security Number is 999-22-6512, and whose home telephone is 555-5555, would consistently create a “pseudo-code” of A2E5A.

At least five questions are recommended, and more can be added if the participant population is large and duplication is a concern. If data will be collected over an extended period of time and there is concern that the participant may not be able to recall the pseudo-code at each data collection point, he/she may be given the **option** of writing his/her code on a piece of paper, sealing it in an envelope, writing his/her name on the envelope, and giving the sealed envelope to the researcher for safe-keeping. The paper on which the participant writes his/her name should include a statement explaining that the researcher will keep, but not open, the envelope; that the participant may open his/her envelope at a later date if he/she forgets his/her code, and that anonymity cannot be guaranteed if the envelope has been opened by anyone other than the participant.

Providing participant anonymity by using pseudo-codes is preferred over methods that identify the participant with his/her data if the use of pseudo-code rubric does not affect the quality of the research.

### **Maintaining Long-Term Confidentiality of Information When Code Lists are Used**

Traditionally, researchers have used code lists to link data collected over a period of time and/or through use of different instruments. These code lists normally assign a number to a participant's name and the code number is used during data collection and recording. Often, it is possible to destroy these code lists after all data are collected, transcribed, and/or entered in a database without affecting the quality of the research. Although a study using this type of coding cannot be considered anonymous because, at least at one point in time, the researcher is able to link data to identity, destruction of the code list after data collection and linking is complete does offer future confidentiality protections to

participants. Destruction of the code list de-identifies the data for anyone else who may have access to the data in the future. This practice also allows the researcher to maintain raw data for an extended period without the risk of compromising participant confidentiality.

Destruction of code lists that associate research data with participant identity after data collection, transcription, and/or database entry are complete is the preferred method of maintaining confidentiality in the long-term, providing that de-identification of the data does not negatively affect the quality of the research.