

## CCSU Institutional Review Board

### Guidelines for Collection of Demographic Data in Research Studies

The concern with collecting demographic data along with sensitive private information lies in the potential for a breach of confidentiality. Sensitive private information is information, that if disclosed outside the research, could reasonably place subjects at risk of criminal or civil liability or be damaging to the subject's financial standing, employability, insurability, reputation, or be stigmatizing.<sup>1</sup> A breach of confidentiality is the inadvertent disclosure of private information that was divulged by subjects for research purposes. Identifiable private information is private information for which the identity of the subject is or may readily be ascertained by the investigator or associated with the information. Subjects can be potentially identified based on a single data point if they represent a group that has small representation in the population (e.g., including Pacific Islander as a response category when only one individual in the population fits that category) or based on a combination of data points that isolate a particular individual in the sample (e.g., 55 year old, Latina, Philosophy major).

In collecting demographic information, researchers must balance the need to maintain research integrity with the need to protect subject privacy. Researchers should also recognize that the risk of a confidentiality breach can occur at any point in the research process from data collection through to the reporting and dissemination of the results. Below are different strategies that can be used prior to data collection as well as after data collection to minimize a potential breach of confidentiality.

#### Strategies for minimizing risk of confidentiality breach prior to data collection:

1. Collect the minimum demographic information needed to fulfill research and reporting requirements. Researchers should determine what information is needed to effectively evaluate the representativeness of the sample and to adequately answer the research questions. Researchers should consider minimizing risks by avoiding collecting information beyond that which is essential to the research and being able to demonstrate that the information to be collected has a legitimate research purpose. For additional guidance, see Applebaum et al. (2018) and Hughes et al. (2016).
2. Collect demographic data separately from other data associated with the study. If the researchers only need demographic information to describe the sample and not to address any specific research questions, the researchers should consider collecting the demographic information separate from any potentially sensitive private information. For example, if the researchers are administering a paper-and-pencil survey on experience with sexual victimization, they can collect demographic information on a form separate from the responses to questions about sexual victimization. The two measures could be placed in separate data collection envelopes. If the researchers are administering the survey electronically, the data can be collected in separate files. However, in order for the electronic survey to be considered anonymous, the data cannot involve the collection of IP addresses.

---

<sup>1</sup> Refer to 45 CFR 46.104 (d)(2) and <http://www.irb.pitt.edu/BenignBehavioralIntervention> for further information about the risk of disclosure of sensitive private information.

3. Use broad categories of response options when collecting demographic data. If the researchers are sampling from a population for which representation among certain groups may be limited, the researchers should try to avoid response categories that may lead certain individuals to be readily identifiable. For example, if less than 5% of the population is over 50, then instead of asking subjects to report their specific age, the researcher can use age ranges (e.g., 21-30, 31-40, 41-50, 50 and over). See Fernandez et al. (2016) for further suggestions.
4. Inform prospective participants that their individual responses are potentially identifiable. In some cases researchers need to collect detailed demographic information and be able to link that information to sensitive private information in order to answer research questions. In such cases, the researchers should inform prospective participants that their identity may be linked to their responses. The researchers should also inform prospective participants about the methods that will be used to reduce the potential for a breach of confidentiality (e.g., data security protocol). The following is sample language that can be included in the informed consent document: “Any information that is obtained in connection with this study and that can identify you will remain confidential. It will be disclosed only with your permission or as required by law. Confidentiality will be maintained by means of ... [describe coding procedures and plans to safeguard data, including where data will be kept, who will have access to it, etc.]”<sup>2</sup>

#### **Strategies for minimizing risk of confidentiality breach during data cleaning and dissemination:**

1. Use data security procedures for maintaining confidentiality during data storage and analysis. Researchers are responsible for ensuring that all data files are stored in a secure location both physically and electronically. Researchers also are responsible for ensuring all staff that will be handling the data are properly trained in research ethics and study-specific data security procedures. See Sieber (2006) for further discussion.
2. Replace identifiers with codes and/or pseudonyms. Researchers should consider replacing direct identifiers with numeric codes or pseudonyms unless such identifiers are needed for follow up with the participants. The removal of direct identifiers, however, may not guarantee that the identities of participants cannot be inferred from other information in the data. Researchers should familiarize themselves with the risk of deductive disclosure<sup>3</sup> associated with data sharing and data dissemination. Refer to Rogers and Nolte (2006) and Zarate and Zayatz (2006) for further information on deductive disclosure.
3. Avoid reporting personally or contextually identifying information. Researchers should consider whether certain information about the individual (e.g., race, ethnicity, occupation) along with the

---

<sup>2</sup> Sample language adopted from UCLA Office of Human Research Protection.

<http://ora.research.ucla.edu/OHRPP/Pages/ConsentTemplates.aspx>

<sup>3</sup> Also known as “internal confidentiality” and “inferential disclosure”, *deductive disclosure* “is the discerning of an individual respondent’s identity and responses through the use of known characteristics of that individual...a person who is known to have participated in ANY survey may be identified by a combination of personal characteristics, allowing identification of that person’s record” (retrieved from <https://www.cpc.unc.edu/projects/addhealth/contracts/dedisclosure> on January 31, 2019).

study setting (e.g., city, state, organization) may result in deductive disclosure. In order to minimize deductive disclosure in quantitative studies,<sup>4</sup> researchers should consider aggregating across small groups and disguising high risk variables by only showing that a variable is greater than x amount, or less than x amount. In order to minimize deductive disclosure in qualitative studies, researchers should consider reporting broad descriptors, (e.g., an administrator in a large urban school district in the Southwest), providing generalized commentary rather than quotes to illustrate a point when identity cannot be masked, and changing inconsequential facts to avoid identification. For further discussion see Bickford and Nisker (2015) and Padget (2008).

4. Consider using a post-interview confidentiality form and member checking. Researchers should consider whether removing potential identifying information from the data, particularly with qualitative studies, might alter the original meaning of the data or affect the authenticity of the results. In such cases, researchers should consider allowing the participant to have a voice in how their data is used. This can involve allowing them to review transcripts for accuracy and using a second consent form pertaining to the use of their data in any reporting. See Kaiser (2009) and Petrova et al. (2016) for more information.
5. Review draft reports to identify potential confidentiality breaches. Researchers should consult independent reviewers, or faculty advisors in the case of student research, to review draft results to assess the potential for re-identification or deductive disclosure.

If you are still unsure if your report could potentially reveal the identities of any participants, consult the CCSU IRB for further guidance.

---

<sup>4</sup> Retrieved from <https://www.icpsr.umich.edu/icpsrweb/content/DSDR/disclosure.html> on January 31, 2019.

## References

- Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist, 73*(1), 3–25.  
<https://doi.org/10.1037/amp0000191>
- Bickford, J., & Nisker, J. (2015). Tensions between anonymity and thick description when “studying up” in genetics research. *Qualitative Health Research, 25*(2), 276–282.  
<https://doi.org/10.1177/1049732314552194>
- Fernandez, T., Godwin, A., Doyle, J., Verdin, D., Boone, H., Kirn, A., Benson, L., & Potvin, G. (2016). More comprehensive and inclusive approaches to demographic data collection. *School of Engineering Education Graduate Student Series*. Paper 60. <http://docs.lib.purdue.edu/enegs/60>
- Kaiser K. (2009). Protecting respondent confidentiality in qualitative research. *Qualitative Health Research, 19*(11), 1632–1641. <https://doi.org/10.1177/1049732309350879>
- Hughes, J. L., Camden, A. A., & Yangchen, T. (2016). Rethinking and updating demographic questions: Guidance to improve descriptions of research samples. *Psi Chi Journal of Psychological Research, 21*(3), 138–151. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2016-45623-001&site=ehost-live&scope=site>
- Padgett, D.K. (2008). *Ethical issues in qualitative research. Qualitative methods in social work research* (2nd ed., pp. 67-68). Thousand Oaks, CA: Sage Publications, Inc.
- Petrova, E., Dewing, J., & Camilleri, M. (2016). Confidentiality in participatory research: Challenges from one study. *Nursing Ethics, 23*(4), 442–454. <https://doi.org/10.1177/0969733014564909>
- Rodgers, W., & Nolte, M. (2006). Solving problems of disclosure risk in an academic setting: Using a combination of restricted data and restricted access methods. *Journal of Empirical Research on Human Research Ethics, 1*(3), 85–97. <https://doi.org/10.1525/jer.2006.1.3.85>
- Sieber, J. E. (2006). Introduction: Data sharing and disclosure limitation techniques. *Journal of Empirical Research on Human Research Ethics, 1*(3), 47–49. <https://doi.org/10.1525/jer.2006.1.3.47>
- Zarate, A. O., & Zayatz, L. (2006). Essentials of the disclosure review process: A federal perspective. *Journal of Empirical Research on Human Research Ethics, 1*(3), 51–62.  
<https://doi.org/10.1525/jer.2006.1.3.51>