Anonymous vs. Confidential

Often the difference between the terms anonymous and confidential is not well understood when applied to research studies involving human participants. Yet, these concepts are very important for the protection of the subjects, and they need to be considered by researchers when designing a study and writing the protocol. During protocol review, the level of protection described will be evaluated by the IRB.

Most studies will include some amount of protection of the subjects, as "harm" may come from revealing certain information, e.g., medical records, beliefs, or even school transcripts. In such cases, the studies need to be designed as either "anonymous" or "confidential." No study can be both. However, to be technical, a study may include two different modes of data collection – one an anonymous on-line survey and one a focus group interview - and so it could be described as both in this situation. That distinction is important when initially writing the protocol and then writing the materials and methods section of the thesis or published paper.

A **"strictly anonymous"** study design is one in which it is impossible to trace data or information back to the research subject from whom it was obtained. In other words, the data **cannot** be identified to any particular research participant, not even by the researcher. There is total separation. No study design that involves the creation of a code linking the subject's identity to a pseudonym or a number can be termed an anonymous study, as the identity of the subject can be traced to the data. Additionally, when a written consent form is collected, this consent form has to be separated from the data that the subject provides. The PI (principal investigator) needs to describe in the protocol how this will be accomplished.

Generally, on-line surveys (SurveyMonkey, Zoomerang, and others) are accomplished with anonymity. But, not always. Survey Monkey has a setting that can be set to not collect IP (Internet Protocol) addresses. The PI must assure the IRB in writing the protocol that the process does not collect IP addresses which could identify the computer user. It is also possible that enough identifiers (gender, age, race, work-site, etc.) could identify a person in the dataset when downloaded. Imagine an online survey of professors at CMC – it is possible to associate responses from the female, aged 50-55, Asian-American professor in the "ABC" department who thought she had anonymity.

"Confidential" research participation means that the data from the research subject(s) **can** potentially be identified or linked to a particular individual. Thus, **any** data collected face-to-face (consumer survey, focus groups, standing in front of a classroom, etc.) is automatically considered in the category of being "confidential" as opposed to "anonymous." This is true even when the researcher assigns a coding number to the subject—and this number cannot be traced back to the subject—because the researcher him-/herself knows who provided the data.

It is possible to **de-identify** data that have been collected by a confidential means. This happens when a PI aggregates individual responses into groups and report means and standard deviations. So, using the example above, a researcher could publish a table of teachers and indentify professors only by race and not include age, academic department, etc. This works as long as there is not an N=1 among the demographics. In a sense, then, this makes the data anonymous when it is "processed," but it would be incorrect to say the data collection method is anonymous.

Thus, where a study involves confidential participation by the subject(s), extra measures need to be taken for their protection. These would include at minimum:

- Securing the collected data (e.g., samples and information) in a locked file cabinet or similar environment, to which only the researcher and/or other trained assistants have access.
- If assigning subjects a "key" or "code" that connects them to the data, storing the key in a locked file cabinet separate from the data.
- Informing the research subjects of these measures to ensure confidentiality. The information provided should be included within the written consent form they sign. Further, it should include the plans to destroy the original data at some reasonable point after the research project is completed, typically five years after publication or immediately after transcriptions are done.

Overall, these comments should not be interpreted to mean that a PI should not collect sensitive data. Rather, the CMC IRB wants the PI to understand the fine and subtle details between anonymity and confidentiality, keep them in mind when writing the protocol and interacting with subjects and their data, and collect only the data necessary to fulfill the research objectives and answer the hypothesis of the study.