National Health Ethics Week
Semaine nationale de l'éthique de la santé
March 2 to 8 / 2015

Research Ethics Bulletin #4: IDENTIFIABILITY

Identifiable or non-identifiable?

Where researchers seek to collect, use, share and access different types of information or data about participants, they are expected to determine whether the information or data proposed in research may reasonably be expected to identify an individual.

Information is identifiable if it may reasonably be expected to identify an individual, when used alone or in combination with other available information. Information is non-identifiable if it does not identify an individual, for all practical purposes, when used alone or combined with other available information.

Categories for assessing the extent to which information could be used to identify an individual:

- **Directly identifying information –** the information identifies a specific individual through direct identifiers (i.e. name, social insurance number, personal health number).
- **Indirectly identifying information –** the information can reasonably be expected to identify an individual through a combination of indirect identifiers (i.e. date of birth, place of residence or unique personal characteristic).
- **Coded information –** direct identifiers are removed from the information and replaced with a code. Depending on access to the code, it may be possible to re-identify specific participants (i.e. the principal investigator retains a list that links the participants' code names with their actual names so data can be re-linked if necessary).
- **Anonymized information –** the information is irrevocably stripped of direct identifiers, a code is not kept to allow future re-linkage, and risk of re-identification of individuals from remaining indirect identifiers is low or very low.
- **Anonymous information –** the information never had identifiers associated with it (i.e. anonymous surveys) and risk of identification of individuals is low or very low.

Therefore, anonymous information and human biological materials are distinct from those that have been coded, and also from those that have been anonymized (see Section A of Chapters 5 and 12 of TCPS 2 (2014)).

Identifiable versus Anonymous

Data that once had identifiers attached to it will never be anonymous (it would be considered 'anonymized' once the identifiers were removed). This distinction is critical because only research involving anonymous information or human biological materials are exempt from REB review (as discussed in yesterday's bulletin).

What is de-identified data?

The easiest way to protect participants is through the collection and use of anonymous or anonymized data; although this is not always possible or desirable. For example, after information is anonymized, it is not possible to link new information to individuals within a dataset, or to return results to participants. The 'next best ' alternative is to use de-identified data where the data are provided to the researcher in de-identified form and the key code is accessible only to a custodian or trusted third party who is independent of the researcher.

Health Research Ethics Authority

Canadian Bioethics Society
Société canadienne de bioéthique