

A RECOMMENDATION TO CLARIFY THE EXEMPTION FROM RESEARCH ETHICS REVIEW OF SECONDARY ANALYSIS OF STATISTICS CANADA DATA SETS IN THE UNIVERSITY OF ALBERTA RESEARCH DATA CENTRE

*Harvey Krahn, Professor, Sociology
Chair, Arts, Science & Law Research Ethics Board
and
Chuck Humphrey, University of Alberta Data Librarian
Academic Director, University of Alberta Research Data Centre*

April 7, 2003

1. Introduction

Two years ago, a Research Data Centre (RDC) was constructed in the Rutherford Library to allow University of Alberta researchers to conduct secondary analysis of Statistics Canada longitudinal data sets previously unavailable to them. Use of the RDC by University of Alberta researchers has been increasing steadily. However, there has also been some confusion recently among researchers and administrators as to whether secondary analysis of RDC data sets is exempt from research ethics review.

We believe that exemption is appropriate, within GFC Guidelines for research involving human participants, and that this exemption should be clearly stated in any future revision of these Guidelines. In the meantime, agreement among University of Alberta researchers and administrators about such an exemption is required in order to ensure that important research projects are not delayed unnecessarily.

2. GFC Guidelines Regarding Ethics Review for Secondary Analysis

Section 66.10.2 of the University of Alberta's GFC Policy regarding research involving human participants clearly states that secondary analysis of data bases is exempt from research ethics review, unless individuals can be identified in the data base or may be identified inadvertently in published reports because of the way in which the data have been analyzed (see **highlighted** sections of Section 66.10.2 below).

GFC 66.10.2 Secondary Use of Data

Secondary use of data refers to the use in research of data contained in records collected for a purpose other than the proposed research itself. Common examples are patient or school records or biological specimens, originally obtained or produced for therapeutic or educational purposes, but subsequently are proposed for use in research. Secondary use of data also refers to instances in which data were obtained

for one REB approved project, but subsequently are proposed for use in new research. The issue of secondary use primarily becomes of concern when the data can be linked to individuals, and becomes critically important when the possibility exists that individuals could be identified from published reports of the research.

If identifying information is involved, REB approval shall be sought for secondary uses of data. Researchers may gain access to identifying information if they have demonstrated to the satisfaction of the REB that:

- a. Identifying information is essential to the research; and
- b. They shall take appropriate measures to protect the privacy of the individuals, to ensure the confidentiality of the data, and to minimise harms to subjects; and
- c. Individuals to whom the data refer have not objected to secondary use.

Depending on apparent risk of the data collection procedures, an REB may also require that a researcher's access to secondary use of data involving identifying information be dependent on:

- a. The informed consent of those who contributed data or of authorised third parties; and/or
- b. An appropriate strategy for informing the subjects; and/or
- c. Consultation with representatives of those who contributed data.

Researchers who wish to contact individuals to whom secondary use data refer shall seek the authorisation of the REB prior to contact. The REB shall evaluate the adequacy of the means by which original confidentiality guarantees are respected in making such contacts.

The implications of data linkage in which research subjects may be identifiable shall be approved by the REB. Advances in the ability to link databases create both new research opportunities and new threats to privacy. These techniques provide means for addressing previously unanswerable questions and for generating better social and health-related information. The values underlying the ethical obligation to respect privacy require researchers and REBs to exercise caution in the creation and use of data of this kind. Where such issues arise, REBs must inform themselves of relevant statutory frameworks, and the criteria required by government for authorisation of use of data in governmental data banks.

3. The RDC Program: History, Structure and Operating Procedures

The Research Data Centres (RDC) program is part of an initiative by Statistics Canada, SSHRC, and major Canadian universities to help strengthen Canada's social and policy research capacity by providing more opportunities for secondary analysis of complex Statistics Canada data sets. The University of Alberta Research Data Centre, situated on the 3rd floor of Rutherford Library, is one of nine such centres established in major research universities across Canada. Infrastructure costs for most of the centres have been provided by the *Canadian Foundation for Innovation* (CFI) and parallel provincial funding programs (e.g., ASRA). Operating costs are covered by university contributions and base grants from SSHRC. The RDCs have been built to the security specifications of Statistics Canada and serve as local Statistics

Canada offices on these nine university campuses.

Statistics Canada has developed a number of new longitudinal surveys (e.g., the *National Population Health Survey*, the *National Longitudinal Survey of Children and Youth*, the *Survey of Labour and Income Dynamics*, the *Workplace and Employee Survey*) that track the responses of individuals over time. Such data bases are a rich source of information for secondary analysis of social issues, much more so than traditional cross-sectional surveys. However, the problem Statistics Canada has faced with such longitudinal data files has been finding a way to produce public-use microdata files that contain sufficient detail for researchers while, at the same time, safeguard the privacy of respondents. The Research Data Centres are expressly designed to address this problem.

Research Data Centres provide researchers with access, in a physically secure university setting, to microdata from Statistics Canada's population and household surveys (usually, but not always, longitudinal surveys). All data sets available in the RDC (both cross-sectional and longitudinal) have been stripped of personal information (i.e., name, address, telephone number). In addition, each RDC is staffed by a Statistics Canada employee who serves as a 'disclosure analyst,' ensuring that individuals cannot be inadvertently identified in published reports because of the cross-tabulation of specific variables (e.g., female dentists living in Wetaskiwin). The RDCs operate under the provisions of the Statistics Canada Act in accordance with all of its confidentiality rules, and are accessible only to researchers with projects approved by a SSHRC-coordinated peer-review process and who have been sworn in under the Statistics Canada Act as 'deemed employees' of Statistics Canada. A copy of Article 17 of the Statistics Canada Act is appended here.

To conduct secondary analysis using Statistics Canada microdata housed in the RDCs, researchers must submit a project proposal to a peer adjudication committee operating under the auspices of SSHRC and Statistics Canada. Approval of proposals is based on: (1) scientific merit and viability of the proposed research; (2) viability of the methods to be applied—given the data to be analyzed; (3) demonstrated need for access to detailed microdata; and (4) expertise and ability of the researchers to carry out the work. If a proposal is approved and a researcher is given access to a Research Data Centre, she/he can access only the microdata set specified in the approved research project and only for the purpose of completing that project. Re-applications are required for undertaking different projects or accessing different data sets.

Statistics Canada also conducts a security check on any researcher undertaking secondary analysis using an RDC data set. Once project approval is provided and the security check has been completed, the RDC analyst will invite the researcher (or research group) for an orientation session where operating procedures are explained and where researchers sign their contract with Statistics Canada and take the Oath or Affirmation of Office and Secrecy, thus becoming 'deemed employees' of Statistics Canada.

Thus, the university-based RDCs are, essentially, extensions of Statistics Canada offices, with a full-time Statistics Canada employee at each site to screen research output and ensure compliance with confidentiality policies and procedures by deemed employees. The RDCs operate under the same security provisions as any other Statistics Canada offices, including the

use of physical access controls and stand-alone computers with no links outside Statistics Canada. Researchers are subject to all the conditions and penalties of regular Statistics Canada employees, including fines and/or imprisonment for breach of confidentiality. And, as explained above, all research output leaving the RDC is screened by the Statistics Canada disclosure analyst to ensure that confidentiality is not being inadvertently breached (i.e., via multi-variable cross-tabulations).

4. Exempting Secondary Analysis of Anonymized Data Sets in the RDC from Research Ethics Review

We believe that the exemption from research ethics review for data analyses conducted in the Research Data Centre should be clarified, in the short term by a written notification from the University of Alberta Research Policy Coordinator to the Chairs of University of Alberta Research Ethics Boards, and in the longer term by a revision to the GFC Guidelines for research involving human participants (Section 66).

Recognizing that some future secondary analyses conducted in the RDC might involve data bases including identifying information, or might involve data linkage with identifying information, the exemption should apply only to standard secondary analyses conducted in the RDC (i.e., those involving data sets from which personal identifiers have been removed by Statistics Canada personnel before the researcher(s) gain access to the data. The exemption could be written into Section 66 of the GFC Guidelines in the manner **highlighted** below. It might also be useful to include the secondary analysis of survey data sets as an example (see **highlight** below), along with other examples, since this type of secondary analysis is probably much more common than those already listed.

GFC 66.10.2 Secondary Use of Data

Secondary use of data refers to the use in research of data contained in records collected for a purpose other than the proposed research itself. Common examples **are social science or public health survey data sets collected for specific research purposes but then re-used to answer additional research questions.** Other examples are patient or school records or biological specimens, originally obtained or produced for therapeutic or educational purposes, but subsequently are proposed for use in research. Secondary use of data also refers to instances in which data were obtained for one REB approved project, but subsequently are proposed for use in new research. The issue of secondary use primarily becomes of concern when the data can be linked to individuals, and becomes critically important when the possibility exists that individuals could be identified from published reports of the research.

Secondary analysis of data sets housed in the University of Alberta Research Data Centre (RDC) and from which personal identifiers have been removed by Statistics Canada personnel is exempt from research ethics review, since standard RDC operating procedures follow Statistics Canada regulations to ensure that individuals cannot be inadvertently identified in published reports.

If **personal** identifying information is involved, **in RDC-based or any other secondary** access and use of data bases, REB approval shall be sought in advance of access to the data.. Researchers may gain access

to identifying information if they have demonstrated to the satisfaction of the REB

While we believe that GFC Policy should explicitly allow the exemption from research ethics review of secondary analysis of RDC data sets from which all identifying information has been removed prior to researchers gaining access to the data, we nevertheless feel that researchers using the RDC should be aware of the ethical guidelines for research involving human participants and should understand why their secondary analysis is exempt from review. Consequently we propose to include in the "check list" that potential RDC users must sign and submit before being given access to RDC data the following statement that currently appears on the Arts, Science & Law Research Ethics Board application:

I have read the UNIVERSITY OF ALBERTA STANDARDS FOR THE PROTECTION OF HUMAN RESEARCH PARTICIPANTS [GFC Policy Manual, Section 66] and agree to abide by these standards in conducting my secondary analysis of confidential data files maintained in the University of Alberta Research Data Centre.