Workshop on Statistical Disclosure Control for Data Confidentiality

10-12 November, Madison, Wisconsin

For statistical agencies and extramural researchers alike, balancing data sharing requirements and concerns about data confidentiality is an increasingly difficult task, in part because implementation of the latest statistical techniques for protecting data confidentiality (called "disclosure limitation techniques" or "statistical disclosure control") in many cases involves complex optimization problems which are difficult to program efficiently. The goal of the CASC project (Computational Aspects of Confidentiality), funded by the European Union, is to address these difficulties by developing user-friendly software that can be used to estimate disclosure risk and implement statistical disclosure control techniques for microdata and aggregate tabular data. A second goal of the project is to undertake the research necessary to support the development of this software.

The project has developed two software programs, μ-Argus and τ-Argus, which can be used to assess disclosure risk and to protect microdata and tabular data, respectively. The software is of use to researchers responsible for creating public-use datasets (microdata or aggregated tabular data) including the major aging-related data resources supported by NIA (e.g., HRS, WLS, NLTCS). It is also useful to those at secure computing facilities who are responsible for reviewing statistical results for confidentiality problems. The latter includes facilities like CDHA's OLDR (Offline Longitudinal Data Repository), MiCDA's Data Enclave, and PARC'S LADR (Limited Access Data Room). In order to ensure that respondent confidentiality is protected, staff from these computing facilities reviews any statistical output that researchers wish to remove from the center. While these rules of thumb may be sufficient to prevent respondent identification, it is likely that they result in unnecessary and sub-optimal suppression of statistical information that, in the presence of a more systematic means of estimating disclosure risk, could be made available to the researcher. In addition, it is important to realize that once the risk of disclosure has been identified as being too high (on the basis either of rules of thumb or more systematic estimation procedures), computing staff need to come up with a way of communicating to researchers what the problem with their analysis is without being so specific as to indirectly disclose the information they seek to protect. The Argus software can be a valuable tool in working with researchers to revise their analyses so as to protect respondent confidentiality while, at the same time, maintaining the usefulness of the analysis in achieving research objectives.

Because the independent development of software tools like Argus is tremendously resource intensive, the need to maximize the returns to publicly funded research by sharing data and analytical results as widely as possible urgent, and the protection of respondents privacy and data confidentiality an ethical and practical necessity, we believe it is critical that researchers and computing staff at the NIA- and NICHD-funded demography centers be exposed to the latest research on disclosure limitation and be given an opportunity to test the Argus software through a hands-on training course. To the extent possible, we would also like to provide researchers and managers from the federal agencies, major data archives, and data-related professional organizations with an opportunity to try out the Argus software. In cooperation with another EU-funded project called AMRADS (Accompanying Measure for Research and Development in Statistics), the CASC project has developed a 3-day training workshop on statistical disclosure control methods and their implementation using the Argus software. To date, however, this training has been available only in Europe. We have invited researchers from AMRADS and CASC to offer a similar training program, tailored to meet the practical needs of the NIA- and NICHD-funded national demographic surveys and demography centers.

The workshop will take place at the Pyle Center on the University of Wisconsin – Madison campus from 10-12 November 2004. Three specialists from Statistics Netherlands, Anco Johannes Hundepool, Eric Schulte Nordholt, and Peter Paul de Wolf, will conduct the training, which will provide participants with an understanding of the theoretical mathematical aspects of statistical disclosure control and the application of these methods using the ARGUS software. The theoretical discussion will be most accessible to researchers with graduate training in statistical methods (e.g., econometrics, statistical methods for sociologists). A portion of the last day of the conference will be devoted to a discussion of participants' work and to the ways in which SDC methods and the ARGUS software might be used to estimate disclosure risk and protect data confidentiality. Participants are encouraged to prepare a brief (1-2 page) description of a study with which they are working, a list of questions and problems related to estimating and controlling disclosure risk in their data, and a sample data set, and to submit these to CDHA by the last week in October so that the trainers will have time to review them and to tailor their presentations and workshop activities to participants' best advantage.

Registration

The workshop will be held at the Pyle Center on the University of Wisconsin – Madison campus. Detailed information is available on the web (http://conferencing.uwex.edu/pyle.cfm).

Because space is limited, invitations have been extended to the NIA- and NICHD-funded survey projects and demography centers, and to selected federal agencies, data archives, and professional organizations. We are asking those institutions who would like to attend to submit a list of candidates ordered by priority so that we can ensure that a space is reserved for at least one participant from each interested project or institution. Interested individuals affiliated with these institutions should contact their director. If your institution or project has not received an invitation and you would like to attend, please contact Janet Eisenhauer Smith, the CDHA Data Archivist (email: jeisenha@ssc.wisc.edu; phone: 608-265-3937) or Mark Schmidt, Assistant to the Director at CDHA (email: mschmidt@ssc.wisc.edu; phone: 608-262-4715).

Accommodations

The conference hotel is the Howard Johnson Plaza Hotel in downtown Madison. The conference hotel is within walking distance of the Pyle Center. The hotel runs a free shuttle service to and from the airport. The shuttle will also transport guests to and from the Pyle Center. Guests at the conference hotel will receive a welcome packet and workshop materials when they check in.

The hotel is located at 525 East Johnson Street, Madison, Wisconsin 53705. A block of rooms has been reserved at the conference rate of $91.00 for a single or double. To make reservations, call the hotel at 608-251-5511. To get the conference rate you will need to use the reservation code "CDHA" or "CDHA Software." More information on the hotel is available on the web (http://www.hojo.com/HowardJohnson/control/home). If you make reservations on-line, please confirm that you are booked at the conference rate.

For participants staying elsewhere: Please be aware that parking on campus is extremely limited. Most spaces are reserved for permit holders, although there is some metered on-campus parking. Those not staying within walking distance of the Pyle Center should contact Mark Schmidt, Assistant to the Director at CDHA (email: mschmidt@ssc.wisc.edu; phone: 608-262-4715), to obtain a temporary parking permit. . There is a municipal public parking ramp near the Pyle Center, but please keep in mind that it fills up

early. Participants who do not stay at the conference hotel can pick up their packet at the Pyle Center at breakfast on Wednesday, 10 November.

Contact Information for Conference Organizers

The conference is being organized by the Center for Demography of Health and Aging. Funding for the Center and for the workshop is provided by the National Institute on Aging.

Interested institutions should submit a prioritized list of candidates to Mark Schmidt, Assistant to the Director, with a copy to Janet Eisenhauer, Data Archivist.

Directions and visitor information will be posted to the CDHA website as they become available. Please direct questions on logistics, hotel registration, and parking to Mark Schmidt.

Questions about the program should be directed to Janet Eisenhauer.

Participants who wish to submit a description of their work and a list of questions or problems they face in applying SDC to their study should do so by October 30. Submissions should be sent to Janet Eisenhauer who will forward them to the conference presenters.

Mark Schmidt, Assistant to the Director
Center for Demography of Health and Aging
University of Wisconsin – Madison
Social Sciences Building, Room 4418
1180 Observatory Drive
Madison, WI 53706

       Phone: 608-262-4715
       Email: mschmidt@ssc.wisc.edu

Janet Eisenhauer Smith, Data Analyst/Archivist
Center for Demography of Health and Aging
University of Wisconsin – Madison
Social Sciences Building, Room 4407
1180 Observatory Drive
Madison, WI 53706

       Phone: 608-265-3937
       Email: jeisenha@ssc.wisc.edu