

**Data Confidentiality Review**  
**Fragile Families and Child Well-Being Study**  
**Baseline and Wave 1, Mother and Father**

Felicia B. LeClere, Ph.D.  
Director, Data Sharing for Demographic Research  
Inter-University Consortium for Political and Social Research

**ICPSR staff contributors**

Russel Hathaway, Ph.D.  
Brian Madden, M.A.  
Corey Colyer, Ph.D.

## Overview

In this document, we describe the results of our review of the documents, description and data file associated with the first two waves (baseline and first wave) of data collection for the Fragile Families and Child Well-Being Study. This data file is based on data collection from 4700 births in 20 cities from a sampling frame of 75 hospitals. The data were collected from women immediately following the birth of a child and from the fathers of those children. Our review includes only data from the baseline interview (which occurred near the time of the birth of the child) and the first year follow up. These files are already in the public domain so much of the analysis presented here is concerned with data that has already been exposed to the probability of disclosure risk. Because this data is already in the public domain, it has already been stripped of direct identifiers such as name, address, telephone numbers Medicare, Medicaid, and Social Security numbers. We will be describing only the disclosure risk associated with indirect identifiers available on the data file, documentation, or description of the project.

Regular disclosure analyses include two steps (see <http://www.icpsr.umich.edu/HSP/>). The first, which we undertake in this document, is to review the study description, documentation, and data for potential sources of disclosure risk when data files are being prepared for public release. The method we used parallels that recommended for all data released to the public by the Federal government (details are available at <http://www.fcsrm.gov/committees/cdac/resources.html>). The second step is to recommend and carry out methods to reduce the risk of disclosure. This is a process that must occur with the full participation of the substantive researchers as it involves an evaluation of the trade-offs between disclosure risk, analytic value, and the analytic consequences of the methods used to mask or remove data. In this confidentiality overview we do provide some simple recommendations for variables that increase the probability of disclosure but leave the decision making and implementation to the data producers. A detailed bibliography is included with this document that will assist the data archivist and other members of the research team.

There are two primary types of disclosure risk. Both assume that those individuals seeking information from the data file have the technical knowledge to conduct the activities described. The first type of risk is **direct identity disclosure**. Information from the data file can be matched either directly or by some form of propensity scoring to external data sources that include names and/or other direct identifiers. Combinations of variables on the data file may make this possible by providing a unique profile of an individual that can be matched to information from public records. The second form of disclosure risk is **attribute disclosure**, which is the process of finding the unique record of the individual who you know to be in the file. From this identification, you may gain knowledge about that individual that was private and held in confidence. We have reviewed the materials for both potential sources of disclosure risk.

In order to assess the risk of direct identification, we use dates of events and characteristics of individuals to develop a matching profile. We then identify external sources of data that contain similar information. The sources of external match data are listed in Appendix 1. The probability of direct matching is very small without additional information about individual residential geography. Fragile Families has identified in published research and descriptive documents, the geography of the sample and that the sampling unit is city hospitals. The published public profiles (Indianapolis example) in particular which contain descriptive profiles increase the direct matching risk. Attribute disclosure evaluation takes two forms: (1) examination of survey documentation and (2) assessment of outliers. We flag outliers largely when the identification of unique cases may have substantial impact on the individuals if re-identification occurs. Criminal activity and sensitive behavior pose this type of risk in particular.

The disclosure analysis follows in three parts. The first is the result of our examination of the questionnaire and survey documentation. The second part describes the result of our attempt to match

survey respondents to external documents. Finally, we present the results of the analysis of outliers. We provide a short summary of recommendations in the conclusion. Appendix 2 includes a complete review of the variable list with potential disclosure risk. This disclosure review is not exhaustive of all potential disclosure and re-identification risk. Data producers are always advised to continuously reassess risk with new data collections or new analyses of the data set.

## I. Documentation Evaluation

### Mother's Data Dictionaries

#### Baseline Mothers' Data Dictionary

This dataset includes information about baby birth dates and baby birth years (Appendix 2 item #1). This presents specific confidential information not only directly, but also could be linked to external data (public birth records) that could further reveal confidential information about the mother, children, father, and family status. The disclosure risk is increased with the presence of birth weights of the children. Race/ethnicity variables could disclose additional private information in that, for example, the "Hispanic" category is broken down as *Mexican, Puerto Rican, Cuban or other Hispanic*. Though this ethnic breakdown is important, the investigators may have to weigh the added benefit of that breakdown to the confidentiality risk posed given the smaller numbers in each category. In addition, the dataset also has information about the respondents' citizenship, and with only 17% indicating non-citizenship status, the likelihood of disclosure is further increased given the small number. In addition, the combination of the citizenship information and detailed ethnic breakdowns are particularly problematic.

Item #5 in the Appendix 2 specifies the earnings variables that could present a disclosure risk. Given the respondents specify income received from public assistance or welfare, such information when combined with external datasets could indirectly identify individuals, in particular, if connected to geographic location. The earnings variable can increase disclosure risk when one considers the ethnic breakdowns – that is, external information on welfare statistics and recipients contain racial, gender, and income statistics that when combined with the Fragile Families variables could further specify individuals and increase the disclosure risk.

#### One-Year Mothers' Data Dictionary

Many of the disclosure and confidentiality risks in this dataset are similar to those found in the Baseline Mothers' Data. This dataset also includes baby and children birthdates, birth month, and birth year variables (#1 in One Year Mother's Appendix). In the One Year dataset there are also four variables that identify government agency funds received – on their own they do not pose a risk, but potential disclosure risk increases if merged with external data.

The highest risk concerns information about the father, in particular, specific dates identifying when child support agreements were reached as well as information about the father's specific crimes and dates of incarceration. It is possible to deduce from combinations of those variables, for example, connecting incarceration dates, offenses (e.g., theft, assault, etc.), incarceration length, and release dates, the identities of some of the men. #3 in One Year Mother's Appendix.

Disclosure risk is present also in the questions related to respondent's receipt of government assistance, welfare, and TANF. The date identifies the type of assistance, assistance amounts, and dates of receipt and if combined with other data, it could be possible to identify the respondents.

The disclosure risk in this dataset is more substantive than that of the Baseline Mothers' Data in that there are more variables with specific information. The dataset includes variables specifying marriage dates, child support legal agreement dates along with specific month and year information concerning fathers' incarceration. Again, many of these variables on their own do not pose a significant disclosure risk, but when taken together and/or linked with external data (e.g., state incarceration statistics); respondents' and their information could be at a pronounced disclosure risk.

## Fathers' Data Dictionaries

### Baseline Fathers' Data Dictionary

This dataset includes extensive salary information on the respondent, including information on total salary, off the books salary, hustle salary as well as monetary amounts received from unemployment and public assistance. In addition, this information is at disclosure risk from external data given that much of this type of information (e.g., public assistance, welfare) is a matter of public record. In addition, the dataset includes geographic codes that could identify the city and state in which respondents' lived and possibly received such public assistance.

### One-Year Fathers' Data Dictionary

Similar to the One-Year Mothers' Data, this dataset poses more confidentiality and disclosure risk than the Baseline Fathers' Data. The dataset includes specific confidential information about respondents' child support monthly payments, criminal charges pending as well as month and year of recent criminal convictions. Disclosure risk is increased with the potential of linkages to external public data and records. Connecting this dataset to public information concerning child support and prison records and in combination with other variables in the dataset could increase the possibility of identifying an individual male respondent. This dataset includes several variables specifying respondent salary and income, including variables that breakdown salary by job, source, and illegal activities.

Given that the four datasets are designed to be merged together to link mother and father variables, disclosure risk is further increased because more specific information on each can be derived through triangulation.

Other data producers restrict release of variables such as wage and self-employment data, detailed health care provider or transaction data, birth date and death date of family members and/or relatives, benefits data (e.g., Supplemental Security Income), and occupation or industry data at Census 2-digit level or greater as well as any variables derived from such restricted variables. For example, in one dataset, it states "any variables or fields **derived from** the data mentioned in items a.-h. (restricted) above, including data linked to an HRS dataset using the data mentioned in items a.-h. (restricted) as linking or matching variables."

## II. External Linkage

All the information necessary to identify these participants is already available to the public. However, the probability of a significant confidentiality breach is mitigated by that fact that would require fairly significant motivation and resources to identify individual participants as well as more information at lower levels of geography.

The following is a partial list of information in the data set that is also available in public records:

- Marriage date, found on marriage licenses
- Date of giving birth
- Public Housing records
- Salary, if employed by government agency

- Criminal records
- Citizenship

Consider the following hypothetical situation: a researcher obtains this data set from ICPSR and decides that he would like to contact the participants with his own follow-up questions. ICPSR and/or the original researcher deny him permission to do so. The researcher decides to obtain alternative lists of information within a given time frame on two or more of the above-mentioned variables from publicly-available databases. Using standard statistical software, he writes simple syntax that runs a comparison between the dataset and the information he has gathered. He discovers that there is a Michael Jones who was arrested for marijuana possession on a given date, a Michael Jones who was married on a given date, and a Michael Jones who became a father on a given date. These dates correspond to the dates given on these variables by participant #123456. Therefore, it is likely that participant #123456 is Michael Jones. While the exact dates are not given in the data set (only month and year are given), multiple matches across several variables would indicate that the probability of a match was high. Many such matches would undoubtedly occur when these large data sets are compared, and this researcher may need to locate as few as 30 participants to obtain significant results for his follow-up questions.

The main obstacle preventing this researcher from locating the participants is the difficulty in getting a list of all marriage licenses, all birth records, and all records for arrest for marijuana possession, even if it could be restricted to chronologically and spatially. Currently, he would have to go to each of these offices, make a written request, and pay a fee to obtain this information. Furthermore, it is entirely possible that at least some of these variables may be freely available on the web in as soon as the next few years. As computer technology improves, the cost of making this information available drops. Similar web-accessible databases (most notably, the sex-offender registries that most states now maintain) are already in existence. The additional step of restricting the data file by time and geography may serve to prevent such linkage.

Another potential misuse of this data might come from someone the participant knows. Consider another hypothetical situation: A man with a child from a previous relationship participates in the study. The child is told about or discovers the man's participation and passes this information along to his mother. His mother suspects that the man is underreporting his income to minimize his child support payment. The mother decides to search for this study online and discovers the data set. Using information (especially such as birth date of child, weight of child at birth, as well as date he moved, ethnic background, etc) she knows from her previous relationship with the man, she is able to determine his ID number. She can then find the income he reported during the study and check if it matches the income he reported to the family court. If it does not, she can take this information to the court and request an increase in his payments. Whether or not the court would accept this information as proof is debatable, but it still would create undue stress and inconvenience for the participant. However, it is important to keep in mind that the amount of time and effort involved in making these links decreases the probability that a participant's identity will be disclosed or confidentiality breached.

While both of these scenarios are possible, neither is likely. Currently it would be rather difficult for our unethical researcher to obtain the databases necessary to identify participants. And the likelihood that a former partner in the situation described above would find out about the study and be able to match data points to her child's father is quite low. On the other hand, the researcher may be able to find some of the databases online in the future, and a former partner seeking increased child support may turn to a private detective – who in turn would be more likely to identify the participant in question. Still, it appears that the risk of breaching confidentiality by publishing this data is not significant.

Summary

While the information necessary to identify participants is potentially available, it would require a fairly substantial amount of effort, time, and resources to obtain this information. Therefore, the risk of disclosure is low.

### III. Analysis of outliers

We identified variables that generate substantial risk of disclosure through the previous steps. Of the collection of variables that posed the most substantial risk, father's incarceration poses the largest consequential risk.

This is a table of year and month of the most recent incarceration ---the timing variables create very small sample sizes.

y e a r o f m o s t r e c e n t i n c a r c e r a t i o n

G	M	Year												Total		
		0	1	2	3	4	5	6	7	8	9	10	11			
W	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
W	2	2	2	4	3	7	5	6	1	8	4	0	0	0	0	8
W	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
W	3	1	1	1	1	2	1	0	6	5	0	5	0	5	0	4
W	0	2	1	0	0	3	3	6	7	8	5	5	5	5	5	5
W	0	1	3	1	5	2	0	1	7	7	6	6	5	4	4	4
W	0	2	1	1	1	2	2	4	8	9	3	5	5	5	5	5
W	2	0	5	3	1	0	1	3	8	5	9	4	4	4	4	4
W	2	0	0	0	4	5	3	5	8	1	6	6	6	6	6	6
W	2	0	3	3	1	6	3	2	0	4	4	4	4	4	4	4
W	0	2	1	2	0	7	2	2	0	2	2	2	2	2	2	2
W	0	1	0	1	1	1	1	3	5	3	9	1	1	1	1	1
W	0	0	3	3	4	1	4	4	9	1	2	2	2	2	2	2
W	1	1	1	2	2	0	3	8	6	5	1	1	1	1	1	1
W	2	1	1	1	0	6	5	4	4	8	1	0	0	0	0	0
W		3	5	2	0	0	3	2	0	0	2	4	0	0	0	0

M o n t h o f m o s t r e c e n t i n c a r c e r a t i o n

G	M	Month												Total		
		0	1	2	3	4	5	6	7	8	9	10	11			
W	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
W	-	2	2	4	2	7	5	5	0	8	4	0	0	0	0	9
W	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
W	J	3	0	1	1	2	1	0	6	4	0	4	0	4	0	8
W	F	0	2	0	0	0	3	3	6	7	7	5	3	3	3	3
W	M	0	1	2	1	5	2	0	0	7	7	6	3	3	3	3
W	Ar	0	2	1	1	0	2	2	3	8	9	3	3	3	3	3
W	My	2	0	5	3	1	0	1	3	8	5	9	2	2	2	2
W	0n	2	0	0	0	3	4	3	5	8	1	6	8	8	8	8
W	Jl	2	2	0	3	3	1	6	3	2	9	4	5	5	5	5
W	Ag	0	2	1	2	0	5	2	2	0	2	2	4	4	4	4
W	Sp	0	1	0	1	1	1	3	5	3	9	1	1	1	1	1
W	0c	0	0	2	2	4	1	4	4	8	1	2	0	0	0	0
W	1	1	1	1	2	2	0	3	8	6	5	1	1	1	1	1
W	2	1	1	1	0	6	5	3	4	7	0	0	0	0	0	0
W		3	4	8	8	8	0	5	9	6	0	0	0	0	0	0

The risk of re-identification is large for simple variables such as time of last incarceration and particularly when added to other variables such as military service. Similar types of analyses could be done for other variables including receipt of TANF/Medicaid.

#### IV Recommendations

##### Baseline Mothers Data Dictionary

- 1) Baby Birth date Variables
  - a) **Remove specific information about baby birth year and month.** Could transform these variables into an indicator of the age of baby. If specific birth date information is removed/transformed, then the baby weight variables could remain as is.
  - b) To reduce disclosure risk, **the baby weight variables could be transformed into range variables.** For example, 7-8lbs, 9-10lbs, 11-12lbs, etc.
- 2) Earnings  
**Transform any variables identifying specific earnings/assistance from welfare, etc., into ranges.** For example, \$11,000-\$14,999; \$15,000-\$19,999. If made into ranges disclosure risk is reduced because specific amounts cannot be used to reference and/or identify individuals by matching to other data sources

##### One Year Mother's Questionnaire

- 1) Birth date/Marriage Date Variables  
**Remove specific information about baby birth and marriage month and year information.**
- 2) Assistance  
**Transform any variables identifying specific information about government assistance and Head Start/Early Head Start into ranges.** For example, \$11,000-\$14,999; \$15,000-\$19,999. If made into ranges disclosure risk is reduced because specific amounts cannot be used to reference and/or identify individuals by matching to other data sources
- 3) Father Incarceration Variables
  - a) **Specific information about when father went to jail and was released from jail should be transformed into an indication of length of time in jail/incarcerated.** If this type of transformation is done, variables that specify why father was in jail can remain as is
  - b) **The variables that identify reasons for father's incarceration could be transformed into categories grouped by type.** For example, combine Theft, Theft by deception, and Crime related to stolen property into one variable, and variables about drug use (4) into one.
- 4) Welfare related variables
  - a) **Transform any variables identifying specific earnings/assistance from welfare, etc., into ranges.** For example, \$11,000-\$14,999; \$15,000-\$19,999. If made into ranges disclosure risk is reduced because specific amounts cannot be used to reference and/or identify individuals by matching to other data sources

##### Baseline Fathers Data Dictionary

- 1) Income  
**Transform any variables identifying specific earnings/assistance from welfare, etc., into ranges.** For example, \$11,000-\$14,999; \$15,000-\$19,999. If made into ranges disclosure risk is reduced because specific amounts cannot be used to reference and/or identify individuals by matching to other data sources

##### One-Year Fathers Data Dictionary

- 1) Baby birthdate, baby weight, marriage data

- a) **Remove specific information about baby birth year and month.** Could transform these variables into an indicator of the age of baby. If specific birthdate information is removed/transformed, then the baby weight variables could remain as is.
  - b) To reduce disclosure risk, the **baby weight variables could be transformed into range variables.** For example, 7-8lbs, 9-10lbs, 11-12lbs, etc.
  - c) Remove specific information about marriage month and year. Could transform these variables, if possible, into length of time married.
- 2) Childcare and Support  
**Remove specific information about month and date childcare agreement was reached.** Could transform this variable into length/duration of agreement. If month and date information is removed, then the information about support/month and total amount paid can remain as is.
- 3) Public Assistance  
**Remove specific month and date information about how long respondent has been receiving welfare/TANF.**
- 4) Jail/Incarceration
- a) **Specific information about when respondent went to jail and was released from jail should be transformed into an indication of length of time in jail/incarcerated.** If this type of transformation is done, variables that specify why respondent was in jail can remain as is
  - b) The variables that identify reasons for **respondent's incarceration could be transformed into categories grouped by type.** For example, combine Theft, Theft by deception, and Crime related to stolen property into one variable, and variables about drug use (4) into one.
- 5) Income
- a) **Transform any variables identifying specific earnings/assistance from welfare, etc., into ranges.** For example, \$11,000-\$14,999; \$15,000-\$19,999. If made into ranges disclosure risk is reduced because specific amounts cannot be used to reference and/or identify individuals by matching to other data sources.

## Appendix I: List of all variables and mnemonics for variables that pose some disclosure risk

### Baseline Mother's Questionnaire

- 1) Baby/Birth date Questions
  - a) Questions about baby birth date/year and birth weight have a substantial disclosure risk given public health records.
  - b) M1btrmon – What date was your baby born, (month)?
  - c) M1brtyr – Year of baby's birth
  - d) M1a13a – In which month of pregnancy did you 1st see doctor/other health care provider?
  - e) M1a14a1 – How much did the 1st baby weigh at birth (lbs)?
  - f) M1a14a2 – How much did the 1st baby weight at birth (oz)?
  - g) M1a14b1 – How much did the 2nd baby weigh at birth (lbs)?
  - h) M1a14b2 – How much did the 2nd baby weight at birth (oz)?
  - i) M1a14c1 – How much did the 3rd baby weigh at birth (lbs)?
  - j) M1a14c2 - How much did the 3rd baby weight at birth (oz)?
- 2) Abortion questions
  - a) M1b27 – When you found out you were pregnant, did you think about having an abortion?
  - b) M1b28 – Did BF suggest that you haven an abortion?
- 3) Neighborhood questions
  - a) These questions have a disclosure risk if combined with other information in the dataset (e.g., public housing)
    - b) M1f1a – How long have you lived in neighborhood – Years?
    - c) M1f1b – How long have you lived in neighborhood – months?
    - d) M1f1c – How long have you lived in neighborhood – all my life?

- e) Yes = 323
- f) M1f3 – Do you live in public housing project?
- i) Yes = 508 (10.0%)
- g) M1f4 – Is the fed/state/local government helping to pay for your rent?
- i) Yes = 600 (12.3%)
- 4) Citizenship/Ethnicity
  - a) Citizenship question, with only 17% indicating non-citizenship pose disclosure risk if combined with other variables in the dataset
  - b) M1h2 – Were you born in the U.S.?
  - c) No = 832 (17.0%)
  - d) M1h3b – Are you Mexican, Puerto Rican, Cuban or other Hispanic?
- 5) Earnings
  - a) Earnings is a confidentiality issue, and variables indicating public assistance if combined with external data present disclosure risk.
  - b) M1i2b – About how much did you earn?
  - c) M1j1b – In last yr did you have income from public assistance/welfare/food stamps?
  - i) Yes = 1738 (35.7%)
  - d) M1j2a – In last 12 mo, about how much money did you receive from earnings?
  - e) M1j2b – In 12 mo, about how much money did you receive from pub. Assis/welfare?
  - f) M1j2c – In last 12 mo, about how much money did you receive from unemployment/SS/welfare?
  - g) M1j3 – What was your total household income before taxes in the past 12 months?

#### One Year Mother's Questionnaire

- 1) Birthdates/Marriage Dates. Specific birth date and marriage date information pose a substantive disclosure risk because such information can be use to link to external data to identify individuals.
  - a) M2a2a1 – What is child's birth month?
  - b) M2a2a3 – what is child's birth year?
  - c) M2a8a1a – When did you and father get married? (Month)
  - d) M2a8a1b – When did you and father get married? (Year)
- 2) Assistance. Specific information about amount and time period of government or social assistance poses a disclosure risk given a possibility that external data could be linked and identities deduced.
  - a) M2b31c – How much money does a government agency give you? (\$)
  - b) M2b31cp – Per what time period?
  - c) M2b31f – How much money does head start/early head start give you?(\$)
  - d) M2b31fp – Per what time period?
- 3) Father/BF questions. Specific information about date of entry into and exit from incarceration along with information about the type of crime the resulted in jail time poses a significant disclosure risk given such information is recorded elsewhere and if combined, identities of respondents could be revealed.
  - a) M2c10a1 – When was the legal agreement first reached? – (Month)
  - b) M2c10a2 – When was the legal agreement first reached? – (Year)
  - c) M2c12 – How much are the payments supposed to be per month? – (\$Amount)
  - d) M2c27c – How much does he pay each month? – (\$)
  - e) M2c33d1 – When did father go to jail/prison? – (Month)
  - f) M2c33d2 – When did father go to jail/prison? – (Year)
  - g) M2c33e1 – When will father be released from prison? – (Month)
  - h) M2c33e2 – When will father be released from prison? – (Year)
  - i) M2c36a1 – What was father in jail for? – Assault
  - j) M2c36a2 – What was father in jail for? – Robbery
  - k) M2c36a3 – What was father in jail for? – Theft

- l) M2c36a4 – What was father in jail for? – Theft by deception
- m) M2c36a5 – What was father in jail for? – Crime related to stolen property
- n) M2c36a6 – What was father in jail for? – Destruction of property
- o) M2c36a7 – What was father in jail for? – Other property offense
- p) M2c36a8 – What was father in jail for? – Gambling
- q) M2c36a9 – What was father in jail for? – Commercial vice
- r) M2c36a10 – What was father in jail for? – Possession/use marijuana/hashish
- s) M2c36a11 – What was father in jail for? – Sale marijuana/hashish
- t) M2c36a12 – What was father in jail for? – Possession of other illicit drugs
- u) M2c36a13 – What was father in jail for? – Sale/manufacture other illicit drugs
- v) M2c36a14 – What was father in jail for? – Major traffic offense
- w) M2c36a15 – What was father in jail for? – Underage drinking
- x) M2c36a16 – What was father in jail for? – Status offense
- y) M2c36a17 – What was father in jail for? – Non-payment of child support
- z) M2c36a18 – What was father in jail for? – Probation violation
- aa) M2c36a19 – What was father in jail for? – Other
- bb) M2c37a1 – How long did he spend in jail? – (Yrs)
- cc) M2c37a2 – How long did he spend in jail? – (Months)
- dd) M2c37a3 – How long did he spend in jail? – (Weeks)
- ee) M2c37a4 – How long did he spend in jail? – (Days)
- ff) M2c38 – How much money does father normally earn per year? – (\$Amount)
- gg) M2g1a – In what country/territory was your father born?
- hh) M2g1b – In what country/territory was your mother born?
- ii) M2g1d – Are you a US citizen?
- 4) Welfare related questions. Specific information about amount and time period of government, welfare, or other social assistance poses a disclosure risk given a possibility that external data could be linked and identities deduced.
  - a) M2h5 – Is this home in a public housing project?
  - b) M2h8a – Since child’s birth, have you received help from – Child supp. collection agency?
  - c) M2h8b – Since child’s birth, have you received help from – Visiting nurse/health start?
  - d) M2h8c – Since child’s birth, have you received help from – Head start/Early Head Start?
  - e) M2h8d – Since child’s birth, have you received help from – Childcare referral agency?
  - f) M2h8e – Since child’s birth, have you received help from – WIC Program?
  - g) M2h9a1 – In past 12 months have you received help from Welfare or TANF?
  - h) M2h9b1 – How many months did you receive help from welfare/TANF?
  - i) M2h9c1 – How much money did you receive last month from welfare/TANF?
  - j) M2h9a2 – In past 12 months, have you received help from food stamps?
  - k) M2h9b2 – How many months did you receive help from food stamps?
  - l) M2h9c2 – How much money did you receive last month from food stamps?
  - m) M2h10a – Int Chk: Did mother receive welfare/TANF in last 12 months?
  - n) M2h10b – Is respondent currently receiving welfare or TANF?
  - o) M2h10c1 – For how long have you been receiving welfare/TANF? – (In years)
  - p) M2h10c2 – For how long have you been receiving welfare/TANF? – (In months)
  - q) M2h11 – Have you ever received welfare/TANF?
  - r) M2h11a1 – When did you last receive welfare/TANF? – (Month)
  - s) M2h11a2 – When did you last receive welfare/TANF? – (Year)
  - t) M2h12 – Do you or child receive SSI?
  - u) M2h12a – Who receives SSI
  - v) M2h12b – How many months in the last 12 did you/child receive SSI?
  - w) M2h13 – Int Chk: Was respondent on welfare/TANF in the past 12 months?
  - x) M2h17 – Int Chk: Has respondent received food stamps in the past 12 months?

- y) M2j3 – Are you/children covered by Medicaid or other assistance pgms or HMO?

#### Baseline Father's Questionnaire

- 1) Neighborhood variables
  - a) F1f1a – Number of yrs lived in your neighborhood?
  - b) F1f1b – Number of months lived in your neighborhood?
  - c) F1f1c – Lived in your neighborhood, all your life?
  - d) F1f3 – Do you live in a public housing projects?
  - e) F1f4 – Is the fed/state/local government helping you pay your rent?
- 2) Citizenship/Location
  - a) F1h2 – Were you born in US?
  - b) F1h2b – Geographical region of father's birth place?
  - c) F1h2c – In what year did you first come to the US to live?
- 3) Income
  - a) F1j8 – About how much did you earn?
  - b) F1j10 – How much did you earn from all of your jobs in the last year?
  - c) F1j16a – In last yr how much did you earn from off the books work?
  - d) F1j16b – In last yr, how much did you earn from your own business?
  - e) F1j16c – In the last yr how much did you earn from hustles?
  - f) F1j16d – In last yr how much did you earn from other activities?
  - g) F1k2a – How much money did you receive from Public assistance/welfare etc.?
  - h) F1k2b – How much money did you receive from Unemployment/SS/Disability?
  - i) F1k3 – In last yr, what was your total hh income before taxes?

#### One-Year Fathers Data Dictionary

- 1) Marriage/Baby dates. Specific birth date and marriage date information pose a substantive disclosure risk because such information can be use to link to external data to identify individuals.
  - a) F2a7a1a – When did you and mother get married? (Month)
  - b) F2a7a1b – When did you and mother get married (Year)
  - c) F2b5a1 – When was the child last weighed at doctor's office? (Month)
  - d) F2b5a2 – When was child last weighed at doctor's office (Year)
  - e) F2b5b1 – How much did child weigh on that day? - (Pounds)
  - f) F2b5b2 – How much did child weigh on that day? – (Ounces)
- 2) Childcare/Support. Specific information related to legal agreements (e.g., child support, etc.) poses a disclosure risk given local, state, and federal governments keep records of such payment and if linked with this data, it could be possible to deduce respondent identities.
  - a) F2b21a12 – What care arrangement (s) are you using now? – Head start/Early head start
  - b) F2b26f – Who provides financial assistance for childcare? – Head Start or Early Head Start
  - c) F2b27f – How much money does head/early Head Start give you? (\$)
  - d) F2b27fp – Per what time period do you receive this amount?
  - e) F2c9 – Do you have legal agreement requiring you to provide financial support to child?
  - f) F2c9a1 – When was the legal agreement first reached? – (Month)
  - g) F2c9a2 – When was the legal agreement first reached? – (Year)
  - h) F2c11 – How much are the payments supposed to be per month? – (\$Amount)
  - i) Could leave this in if agreement month and date are removed
  - j) F2c11p – Per what time period is that sum paid?
  - k) Same as i
  - l) F2c13a – How much of the agreed on child support have you paid since (Date in C9A)?
  - m) F2c13b – Could you please give me a range? – (Dollar ranges)
  - n) F2c13c1 – What is the amount of arrears? – (\$Amount)
  - o) F2c13c2 – Could you give me a range of amount of arrears? – (Dollar range)

- n) F2fc19a – Do you have a legal agreement or an informal agreement?
- o) F2fc19b – How much does she give you each month? – (\$Amount)
- p) F2fc19bp – Per what time period did she agree to pay this sum?
- q) F2fc21c – About how much child support did you give these children in the past 12 mths?
- r) F2fc21c1 – Could you please give me a range of child support paid in past 12 months?
- 3) Citizenship
  - a) F2g1a – In what country/territory was your father born?
  - b) F2g1b – In what country/territory was your mother born?
  - c) F2g1c – Int Chk: Were both parents born in the U.S.?
  - d) F2g1d – Are you a U.S. citizen?
  - e) Fx2g1 – TCO – which of your bio. parents were born in the U.S.?
- 4) Public Assistance. Specific information about amount and time period of government, welfare, or other social assistance poses a disclosure risk given a possibility that external data could be linked and identities deduced.
  - a) F2h5 – Is this home in a public housing project?
  - b) F2h6 – Is fed, state, or local government helping to pay rent?
  - c) F2h7a – Since child’s birth, have you received help from – employment/welfare agency?
  - d) F2h8a1 – In past 12 months have you received help from Welfare or TANF?
  - e) F2h8b1 – How many months did you receive help from welfare/TANF in last 12 months?
  - f) F2h8c1 – How much help did you receive from welfare in the last month?
  - g) F2h8a2 – In past 12 months, have you received help from food stamps?
  - h) F2h8b2 – How many months did you receive help from food stamps in last 12 months?
  - i) F2h8c2 – How much help did you receive from food stamps in the last month?
  - j) F2h8a3 – In past 12 months, have you received unemployment/workman’s comp?
  - k) F2h9 – Int Chk: Did father receive welfare or TANF in the last 12 months?
  - l) F2h9a – Are you currently receiving welfare or TANF support?
  - m) F2h9b1 – For how long have you been receiving welfare/TANF? – (Year)
  - n) F2h9b2 – For how long have you been receiving welfare/TANF? – (month)
  - o) F2h9c – Have you ever received welfare/TANF?
  - p) F2h9d1 – When did you last receive welfare/TANF? – (Month)
    - i) High disclosure
  - q) F2h9d2 – When did you last receive welfare/TANF? – (Year)
    - i) High disclosure
  - r) F2h10 – Do you or child receive SSI?
  - s) F2h10a – Who receives SSI?
  - t) F2h10b – How many months in the last 12 did you/child receive SSI?
  - u) F2h10c – Approximately how much did you receive each month?
  - v) F2h11 – Int Chk: Was respondent on welfare or TANF in the last 12 months?
  - w) F2h15 – Int Chk: Has respondent received food stamps in the past 12 months?
- 5) Jail/Incarceration. Specific information about date of entry into and exit from incarceration along with information about the type of crime the resulted in jail time poses a significant disclosure risk given such information is recorded elsewhere and if combined, identities of respondents could be revealed.
  - a) F2fh18 – Have you been stopped by police but not arrested (not for traffic)?
  - b) F2fh19 – Have you ever been booked or charged w/breaking law?
  - c) F2fh20 – Do you have any charges pending against you?
  - d) F2fh20a – How many charges do you have pending against you?
  - e) F2h20b1 – What charges do you have pending? – Assault?
  - f) F2h20b2 – What charges do you have pending? – Robbery
  - g) F2h20b3 – What charges do you have pending? – Theft
  - h) F2h20b4 – What charges do you have pending? – Theft by deception
  - i) F2h20b5 – What charges do you have pending? – Crime related to stolen property

- j) F2h20b6 – What charges do you have pending? – Destruction of property
- k) F2h20b7 – What charges do you have pending? – Other property offense
- l) F2h20b8 – What charges do you have pending? – Gambling
- m) F2h20b9 – What charges do you have pending? – Commercial vice
- n) F2h20b10 – What charges do you have pending? – Possession/use of marijuana/hashish
- o) F2h20b11 – What charges do you have pending? – Sale of marijuana or hashish
- p) F2h20b12 – What charges do you have pending? – Possession of other illicit drugs
- q) F2h20b13 – What charges do you have pending? – Sale/manufacture of other illicit drugs
- r) F2h20b14 – What charges do you have pending? – Major traffic offense
- s) F2h20b15 – What charges do you have pending? – Underage drinking/purchase of alcohol
- t) F2h20b16 – What charges do you have pending? – Status offense
- u) F2h20b17 – What charges do you have pending? – Non payment of child support
- v) F2h20b18 – What charges do you have pending? – Parole or probation violation
- w) F2h20b19 – What charges do you have pending? – Other
- x) F2fh21 – How you ever been convicted of any charges (not minor traffic)?
- y) F2fh21a – How many times have you been convicted of something?
- z) F2fh21c1 – When was your most recent conviction? – (Month)
- aa) F2fh21c2 – When was your most recent conviction? – (Year)
- bb) F2h21d1 – What charges were you convicted of? – Assault
- cc) F2h21d2 – What charges were you convicted of? – Robbery
- dd) F2h21d3 – What charges were you convicted of? – Theft
- ee) F2h21d4 – What charges were you convicted of? – Theft by deception
- ff) F2h21d5 – What charges were you convicted of? – Crime related to stolen property
- gg) F2h21d6 – What charges were you convicted of? – Destruction of property
- hh) F2h21d7 – What charges were you convicted of? – Other property offense
- ii) F2h21d8 – What charges were you convicted of? – Gambling
- jj) F2h21d9 – What charges were you convicted of? – Commercial vice
- kk) F2h21d10 – What charges were you convicted of? – Possession/use of marijuana/hashish
- ll) F2h21d11 – What charges were you convicted of? – Sale of marijuana/hashish
- mm) F2h21d12 – What charges were you convicted of? – Possession of other illicit drugs
- nn) F2h21d13 – What charges were you convicted of? – Sale/manufacture of other illicit drugs
- oo) F2h21d14 – What charges were you convicted of? – Major traffic offense
- pp) F2h21d15 – What charges were you convicted of? – Underage drinking/purchase of alcohol
- qq) F2h21d16 – What charges were you convicted of? – Status offense
- rr) F2h21d17 – What charges were you convicted of? – Non payment of child support
- ss) F2h21d18 – What charges were you convicted of? – Parole or probation violation
- tt) F2fh22 – Have you ever spent time in correctional institution?
- uu) F2fh22a – How many times were you sent to a correctional institution?
- vv) F2fh25a1 – When were you most recently incarcerated? –(Month)
- i) High disclosure risk
- ww) F2fh25a2 – When were you most recently incarcerated? – (Year)
- i) High disclosure risk
- xx) F2fh26a – When were you most recently released from incarceration? – (Month)
- i) High disclosure risk
- yy) F2fh26b – When were you most recently released from incarceration? – (Year)
- b) Income
- a) F2k15ad – About how much (do/did) you usually earn in (this/that) job, before taxes? (dollar)
- b) F2k15ac – About how much (do/did) you usually earn in (this/that) job, before taxes? (cents)
- c) F2k15ap – Per what time period do you usually earn this amount?
- d) F2k20 – How much did you earn total from all your regular jobs in past 12 months?
- e) F2k20a – Could you please give me the range of your earnings in last 12 months?

- f) F2k23e – How much did you earn in past 12-month from your business?
- g) F2k23e1 – How much did you receive in past 12 months from your business? – (\$Range)
- h) F211 – What was your total household income for the past year before taxes?
- i) F211a – Could you please give a range for your total household income in last year?

## Appendix II

Below is a list of examples of various municipal offices that may provide information for sites in the data set.

Court Records  
Suffolk County  
City Hall  
Room 601  
Boston, MA 02101-1706  
(617) 635-4600

Bexar County District Clerk  
Bexar County Courthouse  
100 Dolorosa Street, 1st Floor  
San Antonio, Texas 78205-3002  
(210) 335-2113

Marriage licenses  
Wayne County Marriage Licenses  
2 Woodward Ave Rm 201  
Detroit, MI  
(313) 224-5514

Essex County Clerk  
Newark, NJ  
(973) 621-4920

One City Hall Square  
Room 213  
Boston, MA 02201  
(617) 635-4175

Birth Records  
One City Hall Square  
Room 213  
Boston, MA 02201  
(617) 635-4175

The Shops at Willow Lawn  
1601 Willow Lawn Drive  
Suite 275  
Richmond, VA 23230  
(804) 662-6200

## Appendix III

### Citations related to data confidentiality

American Association of Public Opinion Research. Best practices for survey and public opinion research [online]. AAPOR Web site (<http://www.aapor.org/>). <http://www.aapor.org/>. Accessed September 2004.

Bellach, B., and D. Schon. 1996. Legislation to protect individual confidentiality: The case of cancer registration in Germany. *Science of the Total Environment* 184, no. 1-2: 33-36.

Bennett, Colin J. 1997. Arguments for the standardization of privacy protection policy: Canadian initiatives and American and international responses. *Government Information Quarterly* 14, no. 4: 351-362.

Berman, Jules J. 2002. Confidentiality issues for medical data miners. *Artificial Intelligence in Medicine* 26, no. 1-2: 25-36.

Boruch, Robert F., and Joe S. Cecil. 1979. *Assuring the confidentiality of social research data*. Philadelphia: University of Pennsylvania Press.

Boulos, Maged N. Kamel. 2004. Towards evidence-based, GIS-driven national spatial health information infrastructure and surveillance services in the United Kingdom. *International Journal of Health Geographics* 3, no. 1.

Burridge, Jim. 2003. Information preserving statistical obfuscation. *Statistics and Computing* 13, no. 4: 321-327.

Carney, P. A., B. M. Geller, H. Moffett, M. Ganger, M. Sewell, W. E. Barlow, N. Stalnaker, S. H. Teplin, C. Sisk, V. L. Ernster, H. A. Wilkie, B. Yankaskas, S. P. Poplack, N. Urban, M. M. West, R. D. Rosenberg, S. Michael, T. D. Mercurio, and R. Ballard-Barbash. 2000. Current medicolegal and confidentiality issues in large, multicenter research programs. *American Journal of Epidemiology* 152, no. 4: 371-378.

Cecil, Joe S. 1993. Confidentiality legislation and the United States federal statistical system. *Journal of Official Statistics* 9, no. 2: 519-535.

Chamberlayne, R., B. Green, M. L. Barer, C. Hertzman, W. J. Lawrence, and S. B. Sheps. 1998. Creating a population-based linked health database: A new resource for health services research. *Canadian Journal of Public Health-Revue Canadienne De Sante Publique* 89, no. 4: 270-273.

Chiang, Yu-Cheng, Tsan-sheng Hsu, Sun Kuo, Churn-Jung Liao, and Da-Wei Wang. 2003. Preserving confidentiality when sharing medical database with the Cellsecu system. *International Journal of Medical Informatics* 71, no. 1: 17-23.

Choldin, Harvey M. 1988. Government statistics: The conflict between research and privacy. *Demography* 25, no. 1: 145-154.

Cios, Krzysztof J., and G. William Moore. 2002. Uniqueness of medical data mining. *Artificial Intelligence in Medicine* 26, no. 1-2: 1-24.

Committee on the Role of Institutional Review Boards in Health Services Research Data Privacy Protection. 2000. *Institutional review boards and health services research data privacy: A workshop summary*. The Institute of Medicine (IOM): Shaping the Future for Health [conference]. Washington, DC: National Academy Press.

Cox, Lawrence H. 1991. [Enhancing access to microdata while protecting confidentiality: Prospects for the future]: Comment. *Statistical Science* 6, no. 3: 232-234.

Dale, A., R. Creeser, B. Dodgeon, S. Gleave, and H. Filakti. 1993. An introduction to the OPCS Longitudinal Study. *Environment and Planning A* 25, no. 10: 1387-1398.

Defays, D., and P. Nanopoulos. 1992. Panels of enterprises and confidentiality: The small aggregates method. *Proceedings of the 1992 Symposium on Design and Analysis of Longitudinal Surveys*. Ottawa, Canada: Statistics Canada.

de Guchteneire, Paul, and Ekkehard Mochmann. . 1990. Data protection and data access: Reports from ten countries on data protection and data access in social research, with an annotated international bibliography. Amsterdam: North-Holland for SWIDOC/IFDO.

Dobson, Jerry. Is GIS a privacy threat? geoworld. Adams Business Media. Accessed 2004 September 28.

Domingo-Ferrer, Joseph, ed. 2002. Inference control in statistical databases. LNCS: Lecture notes in computer science. G. Goos, J. Hartmanis, and J. van Leeuwen, eds. 2316. Heidelberg: Springer.

Domingo-Ferrer, Joseph. 2002. Advances in inference control: An overview. Inference control in statistical databases: From theory to practice . Josep Domingo-Ferrer, ed. 1-7. Vol. 2316 / 2002. Heidelberg : Springer-Verlag .

Duncan, G. T. 2001. Confidentiality and statistical disclosure limitation. International Encyclopedia of the Social & Behavioral Sciences. 2521-2525. Oxford: Elsevier Science Ltd.

Duncan, G. T., and S. Kaufman. 1996. Who should manage information and privacy conflicts?: Institutional design for third-party mechanisms. International Journal of Conflict Management 7, no. 1: 21-44.

Duncan, George, and Diane Lambert. 1989. The risk of disclosure for microdata. Journal of Business & Economic Statistics 7, no. 2: 207-217.

Duncan, George T., Virginia A. de Wolf, Thomas A. Jabine, and Miron L. Straf. 1993. Report on the panel on confidentiality and data access. Journal of official statistics 9, no. 2: 271-274.

Duncan, George T., Thomas Jabine, and Virginia A. de Wolf. , eds. 1993. Private lives and public policies: Confidentiality and accessibility of government statistics. Washington, D.C.: National Academy Press.

Dunn, Christopher S., and Erik W. Austin. 1998. Protecting confidentiality in archival data resources. ICPSR Bulletin XIX, no. 1: 1-8.

Elliot, Mark. 2001. Disclosure risk assessment. Confidentiality, disclosure, and data access. Pat Doyle, Julia I. Lane, Jules Theeuwes, and Laura M. Zayatz, eds. 75-90. Amsterdam: North-Holland.

Fienberg, S. E., and J. McIntyre. 2004. Data swapping: Variations on a theme by Dalenius and Reiss. Annals of the New York Academy of Sciences. 14-29. Berlin: Springer-Verlag.

Fienberg, Stephen E. 1997. "Confidentiality and disclosure limitation methodology: Challenges for national statistics and statistical research." Carnegie Mellon Department of Statistics Technical Report, Working Paper No. 668. Carnegie Mellon Department of Statistics. Pittsburgh, Pennsylvania.

Fienberg, Stephen E. 1994. Conflicts between the needs for access to statistical information and demands for confidentiality. Journal of Official Statistics 10, no. 2: 115-132.

Fienberg, Stephen E., Margaret E. Martin, and Miron L. Straf. , eds. 1985. Sharing research data. Washington, D.C.: National Academy Press.

Frank, Ove. 1978. Application of information-theory to problem of statistical disclosure. Journal of Statistical Planning and Inference 2, no. 2: 143-152.

Greenia, N. H. 2004. Developing adoptable disclosure protection techniques: Lessons learned from a US experience. 343-352. Annals of the New York Academy of Sciences. Berlin: Springer-Verlag.

Gunn, P. P., A. M. Fremont, M. Bottrell, L. R. Shugarman, J. Galegher, and T. Bikson. 2004. The Health Insurance Portability and Accountability Act privacy rule -- a practical guide for researchers. Medical Care 42, no. 4: 321-327.

Harries, Keith. 1999. Mapping crime: Principle and practice. Washington, DC: The National Institute of Justice, Crime Mapping Research Center.

Kruse, Robin L., Bernard G. Ewigman, and George C. Tremblay. 2001. The zipper: A method for using personal identifiers to link data while preserving confidentiality. Child Abuse & Neglect 25, no. 9: 1241-1248.

Langenderfer, Jeff, and Don Lloyd Cook. 2004. Oh, what a tangled web we weave: The state of privacy protection in the information economy and recommendations for governance. Journal of Business Research 57, no. 7: 734-747.

- Martin, D., and G. Higgs. 1997. Population georeferencing in England and Wales: Basic spatial units reconsidered. *Environment and Planning A* 29, no. 2: 333-347.
- McGovern, T. F. 1998. Vulnerability: Reflection on its ethical implications for the protection of participants in SAMHSA programs. *Ethics & Behavior* 8, no. 4: 293-304.
- McMillen, D. 2004. Privacy, confidentiality, and data sharing: Issues and distinctions. *Government Information Quarterly* 21, no. 3: 359-382.
- Moreno, Jonathan D. 2001. Goodbye to all that: The end of moderate protectionism in human subjects research. *Hastings Center Report* 31, no. 3: 9-17.
- Murray, Shoon. 1992. Turning an elite cross-sectional survey into a panel study while protecting anonymity. *The Journal of Conflict Resolution* 36, no. 3: 586-595.
- National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. 1979. *The Belmont report: Ethical principles and guidelines for the protection of human subjects of research*. Department of Health, Education, and Welfare. Washington, DC:
- National Research Council (U.S.). 2003. *Improving access to and confidentiality of research data: A report of a workshop*. Washington, D.C.: National Academy Press.
- O'Dwyer, L. A., and D. L. Burton. 1998. Potential meets reality: GIS and public health research in Australia. *Australian and New Zealand Journal of Public Health* 22, no. 7: 819-823.
- Ohno-Machado, L., P. S. P. Silveira, and S. Vinterbo. 2004. Protecting patient privacy by quantifiable control of disclosures in disseminated databases. *International Journal of Medical Informatics* 73, no. 7-8: 599-606.
- Onsrud, Harlan J., Johnson, Jeff P., and Xavier Lopez. 1994. Protecting personal privacy in using geographic information systems. *Photogrammetric Engineering and Remote Sensing* 60, no. 9: 1083-1095.
- Potok, Nancy. 2000. Data stewardship and accountability at the U. S. Census Bureau. [presented at] Panel on "Benefits and stewardship of linked survey and administrative data," Federal Committee on Statistical Methodology Statistical Policy Seminar. Washington, DC: United States Census Bureau.
- Presser, Stanley, Eleanor Singer, and John Van Hoewyk. 2000. Knowing versus feeling as factors in willingness to provide information to the Census. *Social Science Research* 29, no. 1: 140-147.
- Robbin, A. 2001. The loss of personal privacy and its consequences for social research. *Journal of Government Information* 28, no. 5: 493-527.
- Robbin, A., and H. Koball. 2001. Seeking explanation in theory: Reflections on the social practices of organizations that distribute public use microdata files for research purposes. *Journal of the American Society for Information Science and Technology* 52, no. 13: 1169-1189.
- Robbin, Alice, Heather Koball, and Thomas Jabine. 1999. A survey of statistical disclosure limitation (SDL) practices of organizations that distribute public use microdata. *Workshop on Confidentiality of and Access to Research Data Files*, Committee on National Statistics, National Academy of Sciences.
- Robling, M. R., K. Hood, H. Houston, R. Pill, J. Fay, and H. M. Evans. 2004. Public attitudes towards the use of primary care patient record data in medical research without consent: A qualitative study. *Journal of Medical Ethics* 30, no. 1: 104-109.
- Sande, G. 2002. Exact and approximate methods for data directed microaggregation in one or more dimensions. *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems* 10, no. 5: 459-476.
- Saxby, Stephen. 1986. Data Protection Act 1984: Security and accuracy of data. *Computer Law & Security Report* 2, no. 2: 14.
- Schouten, B., and M. Cigrang. 2003. Remote access systems for statistical analysis of microdata. *Statistics and Computing* 13, no. 4: 381-389.
- Simon, G. E., J. Unutzer, B. E. Young, and H. A. Pincus. 2000. Large medical databases, population-based research, and patient confidentiality. *American Journal of Psychiatry* 157, no. 11: 1731-1737.

Singer, Eleanor, and Felice J. Levine. 2003. Research synthesis: Protection of human subjects of research: Recent developments and future prospects for the social sciences. *Public Opinion Quarterly* 67: 148-164.

Singer, Eleanor, Nancy A. Mathiowetz, and Mick P. Couper. 1993. The impact of privacy and confidentiality concerns on survey. *Public Opinion Quarterly* 57, no. 4: 465.

Sweeney, L. 2002. K-anonymity: A model for protecting privacy. *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems* 10, no. 5: 557-570.

Sweeney, L. 2002. Achieving k-anonymity privacy protection using generalization and suppression. *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems* 10, no. 5: 571-588.

Sweeney, Latanya. 2001. Information explosion. Confidentiality, disclosure, and data access. Pat Doyle, Julia I. Lane, Jules Theeuwes, and Laura M. Zayatz, eds. 43-74. Amsterdam: North-Holland.

Tendick, Patrick. 1992. Assessing the effectiveness of the noise addition method of preserving confidentiality in the multivariate normal case. *Journal of Statistical Planning and Inference* 31, no. 3: 273-282.

Thompson, E. J. 1995. The 1991 Census of Population in England and Wales. *Journal of the Royal Statistical Society Series A-Statistics in Society* 158: 203-240.

United Nations Economic Commission for Europe, and Statistics Sweden. 2003. Statistical confidentiality and access to microdata: Proceedings of the seminar session of the 2003 Conference of European Statisticians. Conference of European Statisticians (CES). New York, Geneva: United Nations.