The Matched Multiple Birth File

Developed by Joyce A. Martin, M.P.H.; Sally A. Curtin, M.A.; Margaret L. Saulnier, M.A.; Jaleh Mousavi, B.A.; and Melissa M. Park

Population of multiple birth/fetal death records

This data set, "The Matched Multiple Birth File," was developed to allow for analysis of characteristics of sets of births and fetal deaths in multiple deliveries. Such analysis is not possible using the traditional NCHS Live Birth and Fetal Death Files because these files contain individual records of births and deaths in multiple deliveries, but do not identify set members. Thus, characteristics specific to the multiple set (e.g., gender combination of the set, outcome of the set, birthweight differences among set mates) are not available.

To match the members of multiple deliveries we first identified all records reported as twins, triplets, and other higher order multiple births in the restricted-use U.S. Live Birth and Fetal Death files for 1995-98. All live birth and fetal death records with reported pluralities of 2 or greater, indicating that the event occurred in a multiple delivery, were selected for matching. (Records with a "not stated" plurality are imputed with a plurality of "1" in these files and were, therefore, not included.) For confidentiality reasons, only births and fetal deaths in twin, triplet, and quadruplet deliveries are included on the public-release file.

Note: Fetal deaths of less than 20 completed weeks of gestation which are part of a matched set are included in this file (see tables 1 and 2). However, as these events are reported by only a small number of reporting areas and are considered less completely reported than fetal deaths of

longer gestations, all statistics presented in the following discussion exclude these events and considerable caution should be used in analysis of them.

Matching algorithm

We developed an algorithm consisting of variables from live birth and fetal death records to match members of twin, triplet, and quadruplet sets. The most obvious variables on which to match, the name and address of the mother, are not collected by NCHS due to confidentiality considerations. The algorithm used in the first stage of matching was based on variables we considered the most uniquely identifying to the pregnancy and which were reported with a reasonable amount of completeness and accuracy: plurality, State and county of occurrence of delivery, mother's date of birth, and the date of delivery (within 2 sequential days). We then identified live birth/fetal death records which had identical values for these five variables. If the number of records with identical information equaled the reported plurality (e.g., three records reported as triplets) these records were considered members of the same twin, triplet, or quadruplet set and assigned a unique set identification number. We also identified instances where the number of records with identical data exceeded the reported plurality of the records. These records were later visually reviewed by the authors and matched where appropriate. Approximately 93 percent of all records were matched in the first stage. All other records were considered unmatched and included in the next stage of the process.

Stage two included all of the unmatched records from stage one except where the number of records with identical data exceeded the reported plurality of the records (see above). In this

stage, we added the date of the reported last menstrual period (LMP) and allowed a one digit difference in the mother's date of birth, or the in the date of delivery to allow for obvious miscoding of these items (e.g. for mother's date of birth a one digit difference would be 04/19/64 for one record and 04/19/44 for the other). Records with not stated mother's date of birth were also matched at this stage if data for the other items were identical (1 digit miscodes were not allowed). Stage two resulted in approximately 5 percent of additional records matched, for a cumulative total of over 98 percent of all records matched.

In stage three of the matching process we developed several more complicated algorithms which included additional items from the birth/fetal death record (e.g.: total number of prenatal visits, weight gain during pregnancy, education of the mother). This cycle also allowed for slightly more leniency in the matching of records with obvious miscodes. Most of the stage three process was done by hand, that is, unmatched records were visually reviewed to identify matches. (A detailed description of this final process is available upon request). Following this, the final stage of the matching process, the cumulative total of twin and triplet records matched was 98.8 percent.

NOTE: Although the proportion of unmatched records is very low for this file, it is important to note that there are important differences in characteristics between matched and unmatched records. For example, unmatched records are more likely to include infant (12.6 percent) and fetal deaths (19.2 percent) than matched records (3.1 and 1.9 percent respectively).

Verification of algorithm

To verify our algorithm we surveyed State vital statistics offices and were able to identify only one, Missouri, which maintained a State database of matched sets of multiple records. We then compared our results for the State of Missouri to that of the Missouri office for data year 1995. Matches on the Missouri file were based primarily on the mother's name and address, which would tend to make them highly reliable and thus we considered the Missouri file the "gold standard". In comparing the two files, we found no "false positives"; that is, we did not inadvertently match any records which were not truly members of sets of multiples births. We found only **6** "false negatives"; that is, records that were unmatched on our file but were matched on the Missouri file. These records comprised only .05% of the more than 2,000 Missouri records on our file. The reason our algorithm failed to match these records was due to poor data quality of the items in our algorithm. For example, the date of delivery for some records was miscoded to such an extent that it did not meet our matching criteria and was not considered a match. Missouri was able to match the records because they had the more identifying information -- the mother's name and address. While not perfect, we felt that the comparison with the Missouri file validated our algorithm as a reliable method for identifying sets of multiples births.

Inclusion of infant deaths

In order to analyze multiple delivery sets for the full perinatal period, information on infant deaths are included on the Matched Multiple Birth File. The Linked Live Birth/Infant Death Cohort Data Sets for 1995-98 were used to identify infant deaths of up to one year of age

which correspond to the live twin and triplet birth records. Information pertaining to the death (e.g. date of death, cause of death) was appended to the end of the appropriate live birth record. Thus, the final file contains all live births and fetal deaths in twin, triplet, and quadruplet deliveries, plus infant deaths.

Imputation of plurality

Plurality (of records originally coded as "2" or higher) was imputed (changed) for a small number of records (63), or .01 percent of all records on the Matched Multiple Birth File. The imputation was performed to correct what appeared to be obvious miscoding of plurality. In general, plurality was imputed where information for records matched but 1 or more of the potential set member records indicated a different plurality. For example, if three live birth records matched and two of these records indicated a plurality of 3 but the remaining record had a plurality of 2, then the plurality for the later record was changed to a 3. Plurality was imputed for all members of a set only where information on all records matched and certificate numbers were perfectly consecutive (for example, if three live birth records with pluralities of "2" matched and certificate numbers were consecutive, the plurality for all three records was changed to "3." All of the records in which plurality was imputed are flagged on the data set.

Additional data fields

A "Birth ID" field was created to indicate whether a record was a live birth which survived the first year of life, a live birth which did not survive the first year, or a fetal death. Records that were not matched were retained on the file to allow comparison of their data with

those of matched records. A “set completeness” flag was generated to distinguish between records that were members of complete sets (2 of 2 twins, or 3 of 3 triplets, or 4 of 4 quadruplets), records that were members of incomplete sets (2 of 3 triplets; or 2 or 3 of 4 quadruplets), or records that were unmatched (1 of 2 twins, or 1 of 3 triplets, or 1 of 4 quadruplets).

A “set birth order” field was developed to capture the order of births born within the twin, triplet, or quadruplet set, an item not included in the traditional live birth data set. The set birth order field is derived from the total birth order (live births and fetal deaths). Where the total birth order of all records is stated, the set birth order follows the chronology of the total birth order except that the set birth order for the first born of the set is coded as “1", the second born as “2", the third born (for triplets) as “3", and the fourth born (for quadruplets) as “4". For example, if the total birth order for one record of a twin set is “5" and for the other record, “6", the set birth order for the first record is “1" and the second record is “2".

The set birth order was imputed where possible--for twins, one record had a total birth order of 1 and the total birth order of the second record was “not stated”. In this case we know that the first record was the firstborn of the set because the total birth order was “1". Thus, the set order of the record with the “not stated” total birth order would be imputed to “2"; which occurred in 5.1 percent of twin records (204 twin records). Similarly, for triplets, if only one record had a “not stated” total birth order and the other two records were stated and one of these records had a total birth order of “1", then the set birth order of the remaining member could be imputed; only 3 triplet records, and 1 quadruplet record were imputed in this manner.

A limitation of this derived variable is the high percent not stated (11.6 percent for

matched records). In general, there is confusion in reporting total birth order for plural sets which often results in inconsistent birth orders between set members (e.g. both set members have the same total birth order). In situations where the total birth order is inconsistent between set members, the set birth order was coded as "not stated".

Acknowledgments

This file was developed and prepared under the general direction of Stephanie Ventura, Chief of the Reproductive Statistics Branch (RSB). Melissa Park also of RSB provided technical review. The Systems, Programming, and Statistical Resources Branch (SPSRB) lent programming support. Staff of the Data Acquisition and Evaluation Branch carried out quality evaluation and acceptance procedures for the State data files on which this report is based. The Registration Methods staff of the Division of Vital Statistics consulted with State vital statistics offices regarding the collection of birth certificate data.

Questions on the file should be directed to the Joyce Martin (301) 458-4362, or Melissa Park (301) 458-4418 of the Reproductive Statistics Branch, Division of Vital Statistics, NCHS, 6525 Belcrest Road, Room 810, Hyattsville, Maryland 20782.