Identification of Iowa Live Births in the Agricultural Health Study

Paul A Romitti[1], **Shinobu Watanabe-Galloway**[1], **William T. Budelier**[1], **Charles F. Lynch**[1], **Soman Puzhankara**[1], **Donna Wong-Gibbons**[1], **Jane A. Hoppin**[2], and **Michael C.R. Alavanja**[3]

[1] University of Iowa, College of Publichealth

[2] Environmental Epidemiology Branch, National Institutes for Environmental Health

[3] Division of Cancer Epidemiology and Genetics, National Cancer Institute

Abstract

In the Agricultural Health Study, information on participant live births was largely provided by female partners of male private applicators. At the Iowa site, such information was available for 13,599 (42.9%) of 31,707 applicators. To improve identification of live births among Iowa participants, we used a probabilistic and deterministic approach to link available demographic data from 31,707 households and information on live births from 13,599 households with 1,014,916 Iowa birth certificates. Record linkage identified 16,611 (93.7%) of 17,719 reported live births and 17,883 additional live births, most (14,411 or 80.6%) not identified due to non-participation by female partners. This record linkage produced an expanded cohort of live born children among Iowa participants, which will facilitate improved study of the effects of agricultural exposures, including pesticides, on selected birth outcomes and childhood disease.

Keywords

children; occupational exposures; pesticides; pregnancy; prospective cohort

---

Adverse pregnancy outcomes have been associated with periconceptional occupational exposure[1] and, also, residential proximity to areas of pesticide application.[2] Even with avoidance of periconceptional or perinatal exposure, children are vulnerable to both an increased risk and an increased susceptibility to pesticides, because of their relative proximity to the ground, immature immune systems, pica behavior, and relative intake to body mass ratio. [3] Among children of farmers and farm workers, such exposures are intensified through pesticide application and related activities.[4,5]

Recent U.S. estimates show that over 500 million kilograms of pesticides are applied annually. [6] Published studies of the relation between early life exposure to pesticides and potential health and developmental effects have often been constrained by use of cross-sectional or retrospective approaches for exposure assessment. Given the limitations in such approaches, [7] prospective, periodic assessment of pesticide exposure,[8] particularly at different life stages of development,[9] might provide a clearer understanding of the effects of these chemicals on birth outcomes and childhood disease. Establishment of such longitudinal studies poses many challenges inherent in cohort studies, including subject recruitment and retention, comprehensive exposure assessment, morbidity and mortality follow-up, and enumeration of sufficient sample size to examine rare outcomes, such as birth defects and fetal deaths. Attempts at conducting such studies to date[10,11,12,13] have varied in their ability to meet these challenges.

The Agricultural Health Study (AHS) is a large, prospective cohort study being conducted in Iowa and North Carolina to examine the health of pesticide applicators, primarily farmers, and their partners.[14] In Iowa, pesticide applicators were recruited during attendance at mandatory

training or testing sessions for state certification to apply restricted-use pesticides. Applicators that consented and their partners were asked to complete several self-administered questionnaires; female partners of male applicators or female applicators were asked to report on their pregnancy history and provide additional detail on live born children. Although about 75% of partners of married applicators provided such information, this totaled less than one-half of all enrolled households. Further, even among female respondents, the number of response foils included to record pregnancies and live born children may not have been sufficient to capture information on all previous pregnancies or live born children.

To address this gap in enumeration of all live births among applicators and partners enrolled in the AHS, we linked AHS questionnaire data and Iowa birth certificate (BC) records to identify pregnancies among AHS participants for a 25-year period. Record linkage was conducted at the Iowa AHS site only, as this site produced the majority of children (81%) reported in the AHS questionnaires[15] Over the 25-year period, the BC was revised; thus, available demographic data for record linkage varied among this birth period (Appendix). Also, differences in entries of names of children (e.g., nicknames vs. full names), missing paternal information in BC records, and changes in maternal last names due to divorce or remarriage were encountered. Because of these data issues, use of a simple exact match would have been insufficient to achieve our goal to maximize correct matches and minimize incorrect ones. Our linkage method, therefore, relied on a combined probabilistic and deterministic approach in which the strength of match between an AHS subject and BC record was based both on the number of data items available and the observed probability of values for each item. This approach is similar to those reported previously to construct reproductive histories by linking live births and fetal death records[16], link Medicaid claims data to state vital statistics records for low birth weight infants[17], and link clients of a maternal and child health project to live and fetal death certificates.[18] The current report details the record linkage methods used and the success of those methods to identify live born children of Iowa AHS participants. This effort provides a more comprehensive dataset of children of pesticide applicators for longitudinal study of the effects of agricultural exposures on selected birth outcomes and childhood disease.

MATERIALS AND METHODS

Agricultural Health Study

The AHS is a collaborative effort of the National Cancer Institute, the National Institute of Environmental Health Sciences, the U.S. Environmental Protection Agency, and the National Institute for Occupational Safety and Health (http://aghealth.nci.nih.gov/). Details on enrollment data collection procedures for the AHS have been described.[14] Briefly, in Iowa, enrollment of private and commercial applicators extended from December 1993 to December 1997. Private applicators included farmers, and commercial applicators included persons employed either by crop protection companies or businesses that used pesticides but whose primary function was not pesticide application (e.g., grain millers). After participation in certification activities at designated sites, applicators were requested to complete the AHS enrollment questionnaire (EQ), which asked about pesticide exposure, lifestyle behaviors, diet, medical conditions, and demographics. Following completion of the EQ, private applicators were given three take-home questionnaires, applicator (AQ), spouse (SQ), and female and family health (FFHQ), and commercial applicators up to two, AQ, and if female, the FFHQ (http://aghealth.nci.nih.gov/questionnaires.html). The SQ was not requested from commercial applicators, because their work was generally not proximate to their home, which reduced the potential for pesticide exposure for their spouse. All data collection and linkage activities in Iowa for the AHS were approved by the University of Iowa Institutional Review Board.

Pregnancy History and Child Data

Reports of children were identified from FFHQ items on pregnancy history and information on live born children. Women provided information on up to 6 pregnancies, but only 4 live births that occurred since 1975. Based on an initial enrollment year of 1993, live births that occurred pre-1975 were ≥ 18 years of age. Maternal reports for these births would have required the offspring's consent; thus, pregnancy outcomes were classified as pre-1975 or 1975-enrollment. Depending upon the number and outcome of pregnancies a woman reported from 1975 through enrollment, a pregnancy may have been included in the pregnancy history section, in the information about children section, in both sections, or in neither section.

Based on questionnaire responses provided, demographic data on an AHS family available for record linkage may have included information from the parent(s), each reported pregnancy, and each reported child. Available data were used to create three input files: one containing parent data (father only, mother only or both); one containing an individual record for each reported pregnancy along with parent data; and one containing an individual record for each reported child along with parent data; thus, a married male applicator (1 record) whose spouse reported 6 pregnancies (6 records) and 4 live births (4 records) would generate a total of 11 'subject' records to be linked to BC data (Figure 1A). For a child reported both in the pregnancy history section and the information on children section, the duplicate record generated from the pregnancy history report was removed in a subsequent programming step, as it contained fewer variables than the child report for record linkage.

Birth Certificate Data

Electronic BCs for deliveries by Iowa residents from January 1, 1975, through December 31, 1999, were provided by the Iowa Department of Public Health. The start year of 1975 was chosen to coincide with FFHQ items for child information on live births delivered from January 1, 1975, to enrollment. BC records through 1999 were used to include all potential pregnancies, including those at time of completion of the FFHQ, for all women enrolled in the study.

Demographic data extracted from BCs included: child (first, middle and last names and birth date); mother (first, middle, last and maiden names, birth date or age and county, city and zip code at child's birth); and father (first, middle and last names and birth date or age). Over the 25-year period, the format of the BC was revised; thus, demographic data available for record linkage varied across years (Appendix). Because of this, a combined probabilistic and deterministic approach was developed in which the strength of match between an AHS 'subject' and BC (henceforth referred to as a subject-BC record pair) was based on both the number of items available and on the observed frequency (probability) of values for each item.

Probabilistic and Deterministic Approach

Population-based probabilities for child, mother, and father demographic items were constructed using the full 25-year BC data set. The frequency of each observed data value for each demographic item extracted from the BCs was calculated. Counts were taken of each name, SoundEx name (established algorithm to compare phonetically-similar names), initial, full (month/day/year) or partial (two of month, day and year) birth date, city, county, and zip code and divided by the total number of observations for that item to obtain the probability of each observed data value for each demographic item.

A computer program was developed in FORTRAN to construct each subject-BC record pair. Because an AHS family could be composed of different individuals due to family structure, the sex of the applicator, or participation of a spouse (e.g., applicator only, applicator-spouse), a household rather than an individual applicator was used as the unit of linkage. Households that included one enrolled private applicator (male or female) were termed one-private

applicator households, and those with two enrolled private applicators (male and female) were termed two-private applicator households. The record linkage program was pilot-tested on 3,000 (approximately 10%) of the AHS families proportional to the distribution of one- and two-private applicator households enrolled.

Record linkage was conducted by reading one BC record at a time and comparing each demographic item extracted with the corresponding item for an AHS 'subject' (applicator, spouse, or child) using a deterministic approach (child first name to child first name, child first initial to child first initial, and so forth); because some vital records included only parent age or year of birth, age was compared to that calculated from the birth date in the EQ or SQ. An item comparison was also performed for 'family name', in which the last names of the child, mother, and father were compared in all possible combinations (e.g., mother last to child last, mother maiden to child last, and so forth) except for mother maiden to father last, as those names were not expected to match.

Each item comparison was assigned a match level based on the highest degree of an item match. For name items, match levels included (in rank order): an exact match 'M' (e.g., Shaun vs. Shaun); SoundEx match 's' (e.g., Shaun vs. Shawn); initial match 'i' (e.g., S vs. Shaun); non-match 'x'; or missing data '.' from either AHS or BC records. Match levels for delivery or birth date items included exact match; match on two of month, day or year; age match; non-match; or missing data. For address items, these levels included zip code, city, county, non-match, or missing data. To tally the results of item comparisons and match level for an AHS 'subject', a 14-character summary code was generated. This code consisted of four characters – one each for first name, middle name, last name, and birth date – for the child (positions 1–4), mother (5–8), and father (9–12), plus one character each for family name match (13) and combined address match (14). Each character represented whether the respective item for an AHS 'subject' matched to that in a vital record and at what level. For example, in Figure 1B, the summary code of xiMM MiMM MisM M M for Edmund indicates, in order, mismatch on child first name (Ed vs. Edmund), child middle initial match, exact match on child last name, child full birth date match, mother first name exact match, mother middle initial match, mother last name exact match (on maiden name), mother full birth date match, father first name exact match, father middle initial match, father last name SoundEx match (Farmer vs. Frmer entered in BC data file), father full birth date match, family name match, and address match on zip code.

Combining the population-based probabilities generated for each BC data value with the summary code generated by the deterministic item linkage permitted calculation of the probability that a subject-BC record pair matched by chance (i.e., false positive). To calculate the probability of a chance match, the multiplicative product of the previously calculated item probabilities for data values for each data item represented in the summary code was calculated (record match probability). Items with missing data for one or both record sources were excluded from the calculation of the record match probability; thus, as mentioned, the record match probability was based on the number of items available for comparison and the probabilities of the data values for each available item. A best possible probability for the subject-BC record pair was also calculated assuming that every available data item for an AHS 'subject' exactly matched the corresponding field in a BC record (best match probability). The ratio of the logarithm of the record match probability to the logarithm of the best match probability was used as an indicator of the strength of the matched items relative to the total items available. This value is equivalent to the percentage of available items which matched, weighted by the frequency of the items, and served as a check of the goodness of fit of a particular comparison. Taking the ratio of the logarithms rather than the ratio of the probabilities allowed us to account for both the strength of the items available for a match (a 'stronger' item has a lower probability of a false positive match), and the strength of the match

(how many items of the record pair actually matched). For example, consider a record pair with record match probability = 10^{-6} and best match probability = 10^{-12}, and another record pair with record match probability = 10^{-2} and best match probability = 10^{-8}. The ratio of probabilities is 10^{-6} for both record pairs; however, the ratio of the logarithms of the probabilities is 0.5 for the first pair and 0.25 for the second pair, reflecting the higher strength of the first record pair. Additionally, taking the logarithm is moved probabilities to a range (0–15) that facilitated manual review. Probabilities were calculated for the family as a whole and for the individual child, mother, and father.

Continuing with the example in Figure 1B, information for the eldest child, Gregory, was reported in the pregnancy history, but not in the information on children section, due to the restriction of child information to the most recent child and three additional live born children. The year of birth for Gregory was 1985, which corresponds to the BC period with the largest number of missing data fields (Appendix). As such, the individual probabilities for child, mother, and father were quite low for this record match. In contrast, the record match for Peter, during the same BC period as Gregory, produced a larger overall probability score due to the reported child information for Peter. Similarly, the record matches for the other reported children, Susan, Diane, and Edmund, produced large overall probability scores. The magnitude of these scores, which were highest for all matches identified for this family, reflected the available information in the most recent BC period that permitted matching on more items. Lastly, the record match for Lucy demonstrated the ability of the linkage program to find unreported children.

Automated and Manual Record Review

The probability score for a subject-BC record pair was reviewed automatically and manually against a set of criteria to decide whether the pair was a definite match. Criteria for accepting or rejecting a probability score for a subject-BC record pair as a definite match were empirically determined by manual review of the pilot study matches. Criteria required sufficient data based on the best match probability value and a sufficient ratio of record match probability to best match probability. A low criteria level was used to assign a record pair as a potential match. For example, when comparing child data, a record pair was classified as a potential match if ((child_best>6) AND (child_ratio>0.5)). To resolve potential matches, data for the match were linked to available Iowa Department of Motor Vehicle files using BC entries of names and birth dates of the child (14 years of age or older), mother, and father. Addresses obtained from motor vehicle files for a child, mother, and/or father within a household were compared. Matches to siblings and duplicate matches between pregnancy and child information entries were removed by a second program.

Using the ratio of the record match probability to the best match probability, a subject-BC record pair was classified as: an automatic definite match; an automatic non-match; undetermined due to insufficient data; or needs manual review. Separate criteria were used for child, mother, father, and family matches. For example, when comparing child data, a record pair was classified as an automatic definite match if ((child_best>8) AND (child_ratio=1)); thus there must have been sufficient information on the child's name to make a determination (best probability must be above a threshold), and there must also be an exact match on the child's name (actual probability must equal best probability for the ratio to be unity). Conversely, a record pair was classified as an automatic non-match if ((child_prob<5.5) OR (child_ratio<0.55)) meaning that there was either insufficient information on the child's name (probability below a threshold) or not enough elements of the child's name matched.

A subject-BC record pair that could not automatically be classified as a definite match or non-match was flagged for manual review. The record pairs with probability scores just above the 'definite' threshold or just below the 'non-match' threshold were also manually reviewed (to

ensure that our criteria did not misclassify record matches). For each AHS 'subject', data items from AHS questionnaires and BC records, as well as corresponding data identified from motor vehicle files, were placed side by side in an electronic form, manually reviewed by two of us (S.W-G, W.T.B.), and assigned a status of definite match, non-match, or undetermined match; discrepancies in manual review were resolved by consensus. Returning to the record match for Gregory in Figure 1B, although the ratios of the logarithm of the record match probability to the logarithm of the best match probability were mostly 1, the small number of fields containing information meant that the negative logs of the probabilities were low, and this potential match had to be manually reviewed before being classified as a definite match. By comparison, scores for the child best match and the ratios identified for the remaining children, except for Edmund, met the criteria for automatic definite matches. The match for Edmund, which showed a mismatch only for child first name, could have also represented a multiple birth in that same pregnancy; therefore, these types of matches were manually reviewed.

RESULTS

The Iowa AHS site enrolled 31,877 of 38,807 (82.1%) private applicators and 4,916 of 10,510 (46.8%) commercial applicators. Of the 31,877 private applicators recruited, 31,434 were males and 443 were females; corresponding totals for commercial applicators were 4,712 and 204. Given the modest proportion of commercial applicators (4,916/36,793; 13.4%), the even smaller proportion of female commercial applicators (204/36,793; 0.6%), and that spouses of commercial applicators were not enrolled, record linkage focused only on private applicators.

The 31,877 private applicators enrolled in the AHS corresponded to 31,537 one-private applicator (includes 7 households in which the spouse was a commercial applicator) and 170 two-private applicator households. Using applicator sex and type of questionnaire(s) completed, each household was classified as: Father-Mother-Child (FMC), Father-Mother (FM), Father-Child (FC); Father (F); Mother-Child (MC); or Mother (M) family unit (Figure 2). To illustrate, a household, in which a male applicator completed the EQ and his spouse completed the SQ and FFHQ, was classified as FMC family unit.

A review of FFHQ responses requested for pregnancy information and information on live born children showed that AHS female participants provided reports of 40,946 pregnancies (all birth years) and 17,675 children (birth years 1975–1999). Reports of pregnancies contained month and year of delivery or termination (n=39,755; 97.1%), year only (n=460; 1.1%), month only (n=617; 1.5%), or neither month nor year (n=114; 0.3%). Reports of children contained month, day, and year of delivery (n=16,658; 94.2%), month and year (n=235; 1.3%), month and day (n=684; 3.9%), month only (n=89; 0.5%), day only (n=1; <0.1%), year only (n=4; <0.1%), or neither month, day nor year (n=4; <0.1%).

Among reported pregnancies with a year of delivery (n=40,215), eligible pregnancies for record linkage were limited to live births delivered from 1975–1999. This subset of pregnancies (n=16,789) included 16,444 singleton births, 242 multiple births, and 103 with a missing multiple birth status that were classified as singletons. These pregnancies and those without a birth year (n=731) were compared to reported live born children delivered from 1975 to 1999 (n=17,675). Comparison of the 242 pregnancies reported as multiple births showed that 235 were twins (n=470 births) and 7 pregnancies were triplets (n=21 births). Comparison of pregnancies without a reported year of delivery (n=731) produced matches for 166 pregnancies (150 singletons; 16 twins) reported as live births and 66 pregnancies (62 singletons; 4 twins) reported as stillbirths but determined from the comparison to child information to be live births. Also, of the 778 live born children without a reported year of delivery, comparison identified a year of birth for 732 (94.1%), all singleton births. In addition, there were 429 reported live born children with an eligible birth year who were not matched to a pregnancy. Together, these

eligible pregnancies (17,188 singletons; 255 twins; 7 triplets) represented 17,719 live births (Table 1).

Linkage of FFHQ data items for each of the 31,707 households to similar items on each of 1,014,916 BC records for the years 1975 to 1999 identified 34,494 record matches. These matches were classified as either pre-enrollment matches (children born before the date that the applicator enrolled in the AHS) or post-enrollment matches (children born after the date that the applicator enrolled in the AHS and, therefore, would not be reported in the FFHQ). Of the 17,719 live born deliveries reported by mothers in the FFHQ, 16,611 (93.7%) were identified through BC record linkage (Table 1), 15,549 as automatic definite matches, and 1,062 as definite matches following manual review. Of these 16,611 matches, 16,578 (99.8%) were determined to be live births, and 33 were determined to be stillbirths. Examination of linkage results across the different BC periods showed that most of the unidentified births occurred during the earliest two periods (1975–82 and 1983–88) when less demographic data were available for record linkage.

Linkage of EQ and SQ data items with BC records also identified 17,883 live births not reported by applicators or their spouses. Of these, 14,411 (80.6%) were pre-enrollment deliveries. Although most of the non-reported births were for households who did not provide a completed FFHQ (n=14,199; 98.5%), some included households who did provide an FFHQ but omitted inclusion of the birth due to limited response slots for pregnancy history and child information. The remaining 3,472 births identified were for post-enrollment deliveries. Of the 34,494 births identified by the record linkage, 33 stillbirths were excluded, 12 third generation children (children of applicators who were themselves children of applicators) were excluded, and one report for each of 14 children who were reported by two families and found linked to each family's report was excluded (i.e., only one set of these 14 children were included). The resulting total included data for 34,435 live births.

DISCUSSION

Questionnaire data from the AHS and Iowa BC records permitted linkage to identify live born children of Iowa AHS participants from January 1, 1975, through December 31, 1999. The linkage method developed successfully matched nearly 94% of the reported pregnancies among AHS participants. Those pregnancies not matched were more commonly from the earlier time periods of the BC (Table 1) when less demographic information for both mothers and fathers was included (Appendix). Because of this, the linkage method relied on a combined probabilistic and deterministic approach in which the strength of match between subject-BC record pair was based on both the number of items available and the observed probability of values for each item. Overall, this combined approach permitted identification of almost 94% of reported pregnancies, including nearly 93% during birth periods with BCs with limited demographic information. The linkage method also matched pre-enrollment deliveries, which were unreported largely due to a household not returning a completed FFHQ. In addition, it was successful in identifying a number of post-enrollment deliveries providing a mechanism for periodic updates of the reproductive histories of Iowa AHS participants without a completed FFHQ.

Although nearly 94% of the reported pregnancies among AHS participants were confirmed, a single record electronic search of BC data for reported pregnancies not identified by the linkage method failed to verify any such pregnancies. With the lack of parental BC data, particularly for fathers, it is possible that unreported pregnancies by AHS participants, either through questionnaire response instructions or through non-participation of female spouses, were also under-ascertained during the earlier time periods. Another possible limitation is the restriction of the pregnancy outcomes to live births only. In Iowa, only fetal deaths with a gestational age

of 20 weeks or greater and/or a weight of 350 grams or greater are issued a fetal death certificate. Because of this, we were unable to use vital records either to verify reported pregnancies that ended prior to 20 weeks gestation or to identify similar unreported pregnancies; however, comparison of maternal interview and BC data for previous pregnancy history among mothers who participated in a Iowa case-control study of orofacial clefts showed a high concordance; thus included BC data for pregnancy history is likely a reliable source of information for previous pregnancies that did not end in a live birth.[19]

Even with the high potential for periconceptional and childhood exposure among children of agricultural workers,[5] previous studies of farm workers and their families have produced discordant results which might be partly attributed to the challenges in collecting comparable data across studies.[20] These authors also cite the challenges imposed in defining the agricultural worker population, particularly for migrant and seasonal workers; communicating with workers, gaining access to their homes and workplaces and in the organization (i.e., seasonality) of agricultural work. Our available data sources and record linkage approach are most applicable to farm families.

Establishment of the current data set offers numerous advantages. Data collection includes detailed assessment of parental agricultural exposures,[14] morbidity and mortality follow-up, [15] and biologic specimen collection from applicators and spouses.[21] Also, sub-studies in the AHS have evaluated pesticide use practices,[22] quality of exposure reports,[23,24] household exposures,[25] and evaluation of non-respondents for collection of questionnaire data,[26] and biologic samples.[21] Additionally, the AHS is working with farmers, a more stable population than farm workers; farmers have better knowledge of their personal pesticide use and may also have better access to health care. Further, a comparison of characteristics of Iowa AHS participants enrolled as private applicators with the Iowa Census of Agriculture for 1992 and 1997[27] found that AHS participants likely represent Iowa farmers at the higher end of pesticide usage.

In summary, a data set of children of pesticide applicators was created to address the gap in enumeration of all live births among applicators and partners enrolled by the Iowa AHS site. This effort provides a more comprehensive dataset of children of pesticide applicators for longitudinal study of the effects of agricultural exposures on selected birth outcomes and childhood disease. For example, with necessary human subjects approval, these data can be these data can be linked to Iowa death certificates and to the National Death Index to track mortality and to Iowa disease surveillance systems to expand on previously published analyses, such as pediatric cancer,[15] and, where sample size permits, to examine the occurrence of other outcomes, such as birth defects and fetal deaths.

References

1. Pierik FH, Burdorf A, Deddens JA, et al. Maternal and paternal risk factors for cryptorchidism and hypospadias: a case-control study in newborn boys. Environ Health Perspect 2004;112:1570–1576.10.1289/ehp.7243 [PubMed: 15531444]

2. Rull RP, Ritz B, Shaw GM. Neural tube defects and maternal residential proximity to agricultural pesticide applications. Am J Epidemiol 2006;163:743–753.10.1093/aje/kwj101 [PubMed: 16495467]

3. Mills PK, Zahm SH. Organophosphate pesticide residues in urine of farmworkers and their children in Fresno County, California. Am J Ind Med 2001;40:571–577.10.1002/ajim.10007 [PubMed: 11675626]

4. Alavanja MC, Sandler DP, McDonnell CJ, et al. Characteristics of pesticide use in a pesticide applicator cohort: The Agricultural Health Study. Environ Res 1999;80:172–179.10.1006/enrs.1998.3888 [PubMed: 10092410]

5. Gladen BC, Sandler DP, Zahm SH, et al. Exposure opportunities of families of farmer pesticide applicators. Am J Ind Med 1998;34:581–587.10.1002/(SICI)1097-0274(199812)34:6<581::AID-AJIM6>3.0.CO;2-U [PubMed: 9816416]

6. Kiely, T.; Donaldson, D.; Grube, A. Pesticides Industry Sales and Usage. U.S. Environmental Protection Agency; 2004.

7. McCauley LA, Anger WK, Keifer M, et al. Studying health outcomes in farmworker populations exposed to pesticides. Environ Health Perspect 2006;114:953–960. 10.1289/ehp.8526 [PubMed: 16760000]

8. Blair A, Zahm SH. Patterns of pesticide use among farmers: Implications for epidemiologic research. Epidemiology 1993;4:55–62. [PubMed: 8420582]

9. Landrigan PJ, Kimmel CA, Correa A, Eskenazi B. Children's health and the environment: public health issues and challenges for risk assessment. Environ Health Perspect 2004;112:257–265. 10.1289/ehp. 6115 [PubMed: 14754581]

10. Arbuckle TE, Zhiqiu L, Mery LS. An exploratory analysis of the effect of pesticide exposure on the risk of spontaneous abortion of an Ontario farm population. Environ Health Perspect 2001;109:851–857. [PubMed: 11564623]

11. Berkowitz GS, Obel J, Deych E, et al. Exposure to indoor pesticides during pregnancy in a multiethnic, urban cohort. Environ Health Perspect 2003;111:79–84. [PubMed: 12515682]

12. Boisen KA, Kaleva M, Main KM, et al. Difference in prevalence of congenital cryptorchidism in infants between two Nordic countries. Lancet 2004;363:1264–1269.10.1016/S0140-6736(04) 15998–9 [PubMed: 15094270]

13. Savitz DA, Arbuckle T, Kaczor D, Curtis KM. Male pesticide exposure and pregnancy outcome. Am J Epidemiol 1997;146:1025–1036. [PubMed: 9420527]

14. Alavanja MC, Sandler DP, McMaster SB, et al. The Agriculture Health Study. Environ Health Perspect 1996;104:362–369. [PubMed: 8732939]

15. Flower KB, Hoppin JA, Lynch CF, et al. Cancer risk and parental pesticide application in children of Agricultural Health Study participants. Environ Health Perspect 2004;112:631–635.10.1289/ehp. 6586 [PubMed: 15064173]

16. Adams MM, Wilson HG, Casto DL, et al. Constructing reproductive histories by linking vital records. Am J Epidemiol 1997;145:339–348. [PubMed: 9054238]

17. Bell RM, Keesey J, Richards T. The urge to merge: linking vital statistics records and Medicaid claims. Med Care 1994;32:1004–1018. [PubMed: 7934268]

18. Holian J. Client and birth record linkage: a method, biases, and lessons. Eval Pract 1996;17:227–35. [PubMed: 12293100]

19. Romitti, PA.; Hanson, JW.; Munger, RG.; Burns, TL. Assessing the validity of birth certificate information. Presented at: 120th Annual Meeting of the American Public Health Association; November 1992; Washington D.C.

20. Arcury TA, Quandt SA, Barr DB, et al. Farmworker exposure to pesticides: methodologic issues for the collection of comparable data. Environ Health Perspect 2006;114:923–928. [PubMed: 16759996]

21. Engel LS, Rothman N, Knott C, et al. Factors associated with refusal to provide a buccal cell sample in the Agricultural Health Study. Cancer Epidemiol Biomarkers Prev 2002;11:493–496. [PubMed: 12010865]

22. Curwin B, Sanderson W, Reynolds S, et al. Pesticide use and practices in an Iowa farm family pesticide exposure study. J Agric Saf Health 2002;8:423–433. [PubMed: 12549246]

23. Blair A, Tarone R, Sandler D, et al. Reliability of reporting on life-style and agricultural factors by a sample of participants in the Agricultural Health Study from Iowa. Epidemiology 2002;13:94–99. [PubMed: 11805592]

24. Hoppin JA, Yucel F, Dosemeci M, Sandler DP. Accuracy of self-reported pesticide use duration information from licensed pesticide applicators in the Agricultural Health Study. J Expo Anal Environ Epidemiol 2002;12:313–316. [PubMed: 12198579]

25. Curwin B, Hein M, Sanderson W, et al. Urinary pesticide concentrations among children, mothers and fathers living in farm and non-farm households in Iowa. Ann Occup Hyg 2007;51:53–65. [PubMed: 16984946]

26. Tarone RE, Alavanja MCR, Zahm SH, et al. The Agricultural Health Study: Factors affecting completion and return of self-administered questionnaires in a large prospective cohort study of pesticide applicators. Am J Ind Med 1997;31:233–242.10.1002/(SICI)1097-0274(199702) 31:2<233::AID-AJIM13>3.0.CO;2-2 [PubMed: 9028440]

27. Lynch CF, Sprince NL, Heywood E, et al. Comparison of farmers in the agricultural health study to the 1992 and the 1997 censuses of agriculture. Agricultural Health and Safety: Recent Advances 2005;10:13–22. 10.1300/j096v10n01_03

A.

Applicator Shaun M. Farmer and spouse Mary A. Farmer (all names are fictitious) reported 6 pregnancies and 4 children in the questionnaire.

Pregnancy End Dates:		Child Names and Delivery Dates:	
1.	02 1985	1.	Ed S. 08 08 1993
2.	08 1987	2.	Susan T. 04 22 1992
3.	10 1990	3.	Diane R. 10 16 1990
4.	02 1991	4.	Peter C. 08 27 1987
5.	04 1992		
6.	08 1993		

This family thus produced 11 'subject' records as input to the matching program: applicator and spouse (record 1); applicator and spouse and each reported pregnancy (records 2-7); applicator and spouse and each reported child (records 8-11).

B.

All 4 reported children (Ed, Susan, Diane, and Peter) and 1 additional pregnancy (1985) were matched to birth certificate (BC) records; the other reported pregnancy was a miscarriage. In addition, an unreported child, born in 1997 (Lucy) after the questionnaire was completed, was matched using applicator and spouse information. The six children identified are listed below.

	Child	Mother	Father	Family	Address	Match Status
Gregory P. Farmer	...i	...a	...a	M	x	
Log record match prob.	2.5	1.36	1.48	8.92	8.92	
Log best match prob.	2.5	1.36	1.48	8.92	11.42	
Ratio	1	1	1	1	0.78	Definite match by manual review
Peter C. Farmer	MiMM	...a	...a	M	.	
Log record match prob.	11.72	1.2	1.22	12.18	12.18	
Log best match prob.	11.72	1.2	1.22	12.18	12.18	
Ratio	1	1	1	1	1	Automatic definite match
Diane R. Farmer	MiMM	MiMM	MiMM	M	M	
Log record match prob.	13.82	10.29	11.33	25.55	29.13	
Log best match prob.	13.82	10.29	11.33	25.55	29.13	
Ratio	1	1	1	1	1	Automatic definite match
Susan T. Farmer	MiMM	MiMM	MiMM	M	M	
Log record match prob.	13.22	10.29	11.33	24.95	28.53	
Log best match prob.	13.22	10.29	11.33	24.95	28.53	
Ratio	1	1	1	1	1	Automatic definite match
Edmund S. Farmer	xiMM	MiMM	MisM	M	M	
Log record match prob.	9.26	10.29	8.76	20.99	24.58	
Log best match prob.	13.27	10.29	9.73	25.01	28.59	
Ratio	0.70	1	0.90	0.84	0.86	Definite match by manual review
Lucy E. Farmer	....	MiMM	MiMM	M	M	
Log record match prob.	0	10.29	11.33	17.91	21.5	
Log best match prob.	0	10.29	11.33	17.91	21.5	
Ratio		1	1	1	1	Automatic definite match

Figure 1.

Sample Agricultural Health Study subject-birth certificate data linkage
Summary codes: 'M'=name, date, or address zip code exact match; 's'=name SoundEx match; 'i'=name initial match; 'a'=age match; 'x'=non-match; '.'=missing data
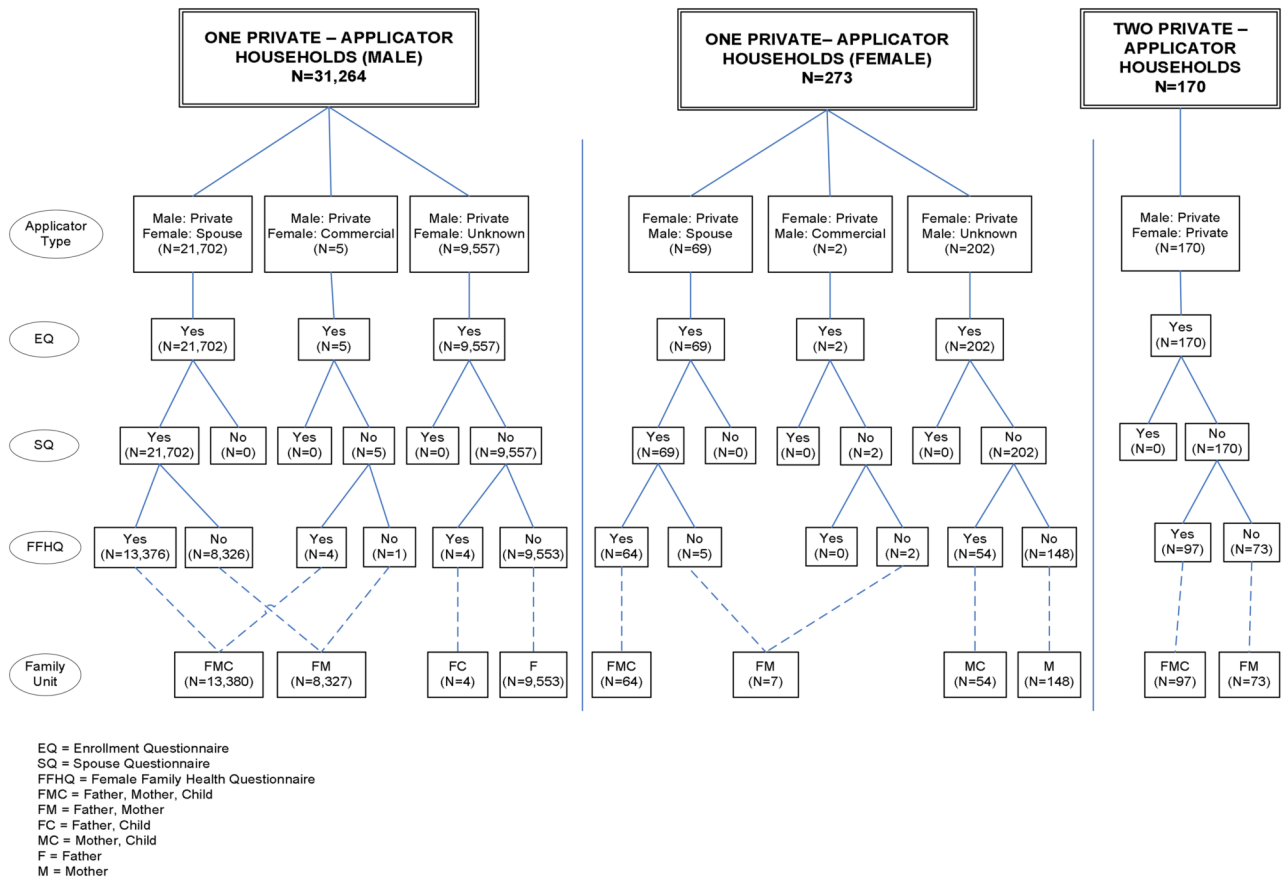
EQ = Enrollment Questionnaire
SQ = Spouse Questionnaire
FFHQ = Female Family Health Questionnaire
FMC = Father, Mother, Child
FM = Father, Mother
FC = Father, Child
MC = Mother, Child
F = Father
M = Mother

Figure 2.
Distribution of households identified by linkage of Iowa Agricultural Health Study
questionnaire data and Iowa birth certificate records.

Table 1

Percentage of reported children of Iowa Agricultural Health Study (AHS) participants found by record linkage program by birth certificate period

BC period	Children reported in AHS surveys	Children found by linkage program	Children not found by linkage program	% Children found by linkage program
1975–82	7,926	7,341	585	92.6 %
1983–88	5,385	4,985	400	92.6 %
1989–94	3,895	3,787	108	97.2 %
1995–99	513	498	15	97.1 %
Total	17,719	16,611	1,108	93.7 %

Appendix

Demographic data available for Iowa birth certificates by time period

Demographic Data[*]	Time Period			
	1975–1982	1983–1988	1989–1994	1995–1999
Child				
Last name				
First name				
Middle name				
Birth or delivery date				
Mother				
Last name				
Maiden name				
First name				
Middle name				
Birth date				
Age				
Address				
Father				
Last name				
First name				
Middle name				
Birth date				
Age				

[*] Shaded bar represents available data items during specified time period.