



HHS Public Access

Author manuscript

J Expo Sci Environ Epidemiol. Author manuscript; available in PMC 2015 September 23.

Published in final edited form as:

J Expo Sci Environ Epidemiol. 2008 July ; 18(4): 381–391. doi:10.1038/sj.jes.7500610.

Using satellite derived land cover information for a multi-temporal study of self-reported recall of proximity to farmland

Gillian A. Avruskin^a, Jaymie R. Meliker^a, and Geoffrey M. Jacquez^{a,b}

^aBioMedware Inc., Ann Arbor, Michigan, USA

^bDepartment of Environmental Health Sciences, School of Public Health, University of Michigan, Ann Arbor, Michigan, USA

Abstract

Exposure misclassification is a major concern in epidemiologic studies. The potential for misclassification becomes even more problematic when participants are asked to recall historical information. Yet, historical information is important in cancer studies, where latency is long and causative exposures may have occurred years or even decades prior to diagnosis. Even though self-reported proximity to farmland is a commonly used exposure measure, the accuracy of recall is seldom, if ever validated. Geographic Information Systems (GISs) and land cover information derived from satellite imagery can allow researchers to assess the accuracy of this exposure measure, and to quantify the extent and importance of exposure misclassification. As part of a bladder cancer case–control study in Michigan, participants were asked whether they lived on a farm, or within a distance of 1/4, 1/4–1, 1–5, or >5 miles from farmland for each residence over their lifespan. Responses from 531 participants over two time periods — 1978 and 2001 — were investigated. Self reported proximity to farmland was compared to a “gold standard” derived from Michigan land cover files for the same time periods. Logistic regression and other statistical measures including sensitivity, specificity, and percentage matching were evaluated. In comparing self-reported and land cover-derived proximity to farmland, cases exhibited better agreement than controls in 2001 (adjusted OR = 1.74; 95% CI = 1.01, 2.99) and worse agreement in 1978, although not significantly (adjusted OR = 0.74; 95% CI = 0.47, 1.16). When comparing 2001 with 1978, both cases and controls showed better agreement in 2001, but only cases showed a significant difference (adjusted OR=2.36; 95% CI = 1.33, 4.18). These differences in agreement may be influenced by differences in educational attainment between cases and controls, although adjustment for education did not diminish the association. Gender, age, number of years at residence, and geocoding accuracy did not influence agreement between the proximity approaches. This study suggests that proximity measures taken from satellite-derived land cover imagery may be useful for assessing proximity to farmland, and it raises some concerns about the use of self-reported proximity to farmland in exposure assessments.

Keywords

GIS; proximity; exposure assessment; participant recall

¹Address all correspondence to: Dr. G. A. Avruskin, BioMedware Inc., 516 N State St, Ann Arbor, MI 48104-1236, USA. Tel.: +1 734 913 1098. Fax: + 1 734 913 2201. avruskin@biomedware.com.

Introduction

A growing number of studies in environmental epidemiology use geographic information systems (GISs) and global positioning systems (GPSs) to assist in exposure assessment by measuring proximity of individuals to different contaminant sources (Royster et al., 2002; Nuckols et al., 2004). Examples include proximity to toxic releases from industries (Burra et al., 2006; Choi et al., 2006), as well as proximity to non-point sources of aqueous nitrates, petrochemicals (Yu et al., 2006) and other environmental contaminants (Wickre et al., 2004). Proximity measures prove especially valuable when estimating exposures that occurred many years ago, and when present-day biomarkers are not appropriate measures of past exposures.

Collecting residential histories using surveys with self-reported proximity to farmland and industrial locations is increasingly common in epidemiologic studies of chronic diseases with long latency periods, such as cancer (Reynolds et al., 2004). Proximity to agricultural crops is often used to assess exposure to herbicides, fertilizers, and pesticides (Brody et al., 2004; Lu et al., 2006; Meyer et al., 2006; Ward et al., 2006). GIS and satellite imagery have been applied repeatedly to create historical crop maps that support estimation of exposures to agricultural pesticides based on residential proximity to different crop types (Ward et al., 2000; Xiang et al., 2000). There is wide agreement that individuals residing in close proximity to crops or farmed land are at increased risk of pesticide exposure (Ward et al., 2000; Lu et al., 2006). Drift from pesticide applications can extend from 500 to beyond 1000 m (Frost and Ware, 1970; Byass and Lake, 1977) increasing exposure risk for people living within this distance. And drift from other types of pesticide applications has been demonstrated at distances of 300–800 m from the application area leaving an intermediate distance of about 500 m or 0.3 miles (Ward et al., 2000).

To date and to our knowledge, residential histories and self reported proximity to farmland have yet to be adequately validated against a “gold standard”. There has been only one validation study that compared recall of proximity to agricultural crops up to 5 years prior to interview. Land use survey maps served as the “gold standard” to which participant responses were compared (Rull et al., 2006). In that study of neural tube defects, differences in recall just after conception were identified between cases and controls, indicating potential recall bias. How well participants remembered their proximity or accuracy of recall decades prior to interview, however, was not assessed.

With increasing use of residential histories in case–control studies, it is critical for the accuracy and reliability of recall to be quantified and for recall error to be nondifferential between healthy and diseased participants. In this paper we compare how well participants remember proximity to farmland (accuracy of recall) in 1978 and 2001 with farmland maps generated from satellite-derived imagery. This recall accuracy is investigated among cases and controls in a cancer case–control study, with consideration of the influence of age, gender, educational attainment, geocoding accuracy, and number of years at a residence.

Methods

Study Population

The study population comes from a population-based case– control study of bladder cancer being conducted in southeastern Michigan. Cases for the parent study were recruited from the Michigan State Cancer Registry and controls were frequency matched to cases by age (75 years), race, and gender and were recruited using random digit dialing (Avruskin et al., 2004; Meliker et al., 2005). Participants must have lived in an 11 county study area of southeastern Michigan (Figure 1) for at least 5 years prior to recruitment and had no prior history of other cancers (with the exception of non-melanoma skin cancer). Participants completed a phone interview and answered questions about demographic and socioeconomic characteristics including current age and education. At the time of writing, data were available for 220 cases and 440 controls in the ongoing study. But due to the temporal nature of this particular investigation, the study population was restricted to those who lived in the study area in both 1978 and 2001 leaving a total of 531 participants, 184 diagnosed with bladder cancer and 347 controls. All cases completed the interview between 2003 and 2004. About 27% of the 70 cases diagnosed in 2000 completed the interview in 2003 and 11% diagnosed in 2000 completed the interview in 2004. The remaining cases were diagnosed in 2001 and completed the interview in 2004.

Within a few months of completing the telephone interview, participants also filled out a written questionnaire describing their residential mobility history. They were asked to include each address where they lived for at least 1 year. If they did not remember a complete address (street, number, and city), they were asked for major cross streets. Each residence in the study area was geocoded using the Michigan Geographic Framework roads file (generated from maps created during the 1990s — with some county maps prepared in 1991 while others were completed in 1995; see <http://www.mcgi.state.mi.us/mgdl/?action=meta>, under the heading geographic framework and then county specific road files) and assigned a geographic coordinate in ArcGIS. Spelling sensitivity equal to 75, minimum candidate score equal to 10, and a minimum geocoded score of 60 were used in the geocoding process. Addresses that were not automatically geocoded were manually matched using cross streets with the assistance of internet mapping services. If cross streets were not provided, best informed guess placed the address somewhere along the road or, as a last resort, the residence was placed at the town centroid.

For each place of residence, participants reported proximity to the nearest farm and whether their residence was a farm (using a “Yes” and “No” check box; a definition of “farm” was not provided). If “No” was chosen, they were asked “If this was not a farm approximately how close was the nearest farm?” They were then asked to check the appropriate box of “Less than one-quarter of a mile,” “One-quarter to 1 mile,” “1 mile to 5 miles,” or “Greater than 5 miles.” If participants left this category blank, their nonresponse was treated as missing data. In 1978, 29 participants did not report a proximity measure; in 2001, 20 participants did not fill in a response.

Land Use Data

Land cover, land use data files were selected for 1978 and 2001 from the Michigan Center for Geographic Information (<http://www.mcgi.state.mi.us/mgdl>). The 1978 MIRIS (Michigan Resource Information System) data originate from the Michigan Department of Natural Resources (MDNR) and represent a compilation of data from county and regional planning commissions and their subcontractors. The original 1978 CAD (computer aided design) files were converted into GIS-compatible files by the Michigan State University Center for Remote Sensing and GIS and by MDNR. Files were formatted for the entire state in the Michigan Georef coordinate system. Variations in quality and consistency between counties exist due to the nature of the data collection and compilation; therefore, horizontal accuracy is measured at ± 80 ft for this 1978 file.

The 1978 file was imported into ArcGIS 9.0 (ESRI, Redlands, CA, USA), with the geocoded residential information overlaying it. Four land use classes were selected from the 1978 file to represent farmland (Appendix 1). These were (a) cropland, rotation, and permanent pasture (covering 54% of the study area), (b) orchards, vineyards and ornamental (covering less than 1% of the study area), (c) other agricultural land (covering $<1/10$ of 1% of the study area), and (d) Christmas tree plantations (also covering $<1/10$ of 1% of the study area). These land use categories were merged into one polygonal file to represent all farmland types.

The 2001 land cover data file comprises data for the Southern Lower Peninsula of Michigan derived from classification of landsat thematic mapper (TM) imagery. The data are stored in a raster format, a data structure representing a rectangular grid of pixels, where each pixel or cell has a resolution of 30 m. The data were downloaded as a TIFF (tagged image file format) in the Michigan GeoRef coordinate system. The classification system used to produce this file is very similar to the system used to produce the 1978 MIRIS file. However, the 2001 data include five agricultural groups: forage crops/non-tilled herbaceous agriculture (vegetation used for fodder production such as alfalfa and hay, covering almost 30% of the study area); row crops (annual, crops planted in rows such as corn and soybeans, covering 20% of the study area); Non-vegetated agriculture (land area tilled for crop production with less than 25% currently vegetated, covering less than 1% of the study area); orchards/vineyards/nursery (excluding woody trees not grown for Christmas trees, covering less than 1% of the study area); and Christmas tree plantations (none in study area) (Appendix 2). These five land cover categories were saved as a vector layer using the raster to vector dialogue of ArcGIS's Spatial Analyst.

To measure proximity to farmland, buffers were created around each geocoded residence in each time period at a distance of <5 , <1 , and $<1/4$ miles, and contained within the farmland (Figure 2). These categories match those presented to the participants on the residential history questionnaire. Each residence was then assigned to one of the above categories for both 1978 and 2001.

Statistical Analysis

Self-reported proximity to farm land was compared with the distance calculated from each residence to farmland using the Michigan satellite-derived land cover maps for 1978 and 2001 to assess accuracy of self reported proximity. Using the five categories of proximity to farmland already described, percentage matching and a weighted Kappa statistic were calculated. This statistic measures the amount of agreement between two measures beyond that expected by chance (Szklo and Nieto, 2000). Full weight (1.00) was assigned for perfect agreement between categories for reported and GIS-derived values. A weight of 0.75 was assigned for disagreement between adjacent categories, a weight of 0.5 for disagreement across two categories, and a weight of 0 for disagreement across three or more categories. Spearman correlation coefficients were also computed as a measure of agreement between self-reported proximity and satellite-derived proximity.

In addition, in order to compare our results directly with the validation analysis of Rull et al. (2006), the five categories were collapsed into two categories — 1/4 mile from farmland and >1/4 mile from farmland. Studies have shown pesticide drift from 300 m to over 1 km depending on the type of application and environmental conditions. The 1/4 mile threshold was chosen because this was the closest intermediate distance presented to our participants for the range of drift from pesticides, and this approximate distance has also been used in other studies (Ward et al., 2000; Rull et al., 2006). Sensitivity, specificity, and percentage matching were calculated and compared between 1978 and 2001, and between cases and controls using Fisher's exact test and two-tailed *P*-values. Among those participants classified by the land use file as living >1/4 miles from farmland, specificity is defined as the proportion of those self-reporting to be >1/4 miles from farmland. On the other hand, among those participants classified by the land use file as living 1/4 miles from farmland, sensitivity refers to the proportion of those self-reporting to be 1/4 miles from farmland.

Logistic regression analyses were also conducted by comparing those whose self-reported proximity agreed with the proximity classification of the satellite-derived data, with those whose self-reported proximity did not agree with the satellite-derived proximity classification (this measure was treated as the dependent variable). Factors influencing agreement between these classification approaches were examined. First, influence of case-control status and year of residence (1978 or 2001) were investigated in univariate analyses, as well as in multivariate analyses, adjusting for gender, education (at least some college education versus not having any college education), age, number of years spent at residence, geocoding accuracy (automatically geocoded versus those geocoded using a cross street or city/town center), and whether or not a participant lived at the same residence for both years. Next, data were stratified into four groups: case residences in 1978, case residences in 2001, control residences in 1978, and control residences in 2001. Agreement between classification approaches was again treated as the dependent variable in logistic regression analyses for each of these strata. The following variables were examined for their influence on agreement between classification approaches: gender, education, age, number of years spent at residence, geocoding accuracy, and whether or not a participant lived at the same residence for both years. All analyses were conducted in SAS version 9.1 (SAS Institute, Cary, NC, USA); odds ratios (ORs) and 95% confidence intervals (CIs) were calculated.

Results

Demographic and residential characteristics of the 531 cases and controls are presented in Table 1. Over 75% of both the cases and controls are male and just over half of the participants are older than 70. The controls have attained higher education than the cases (69.7% versus 53.8%). In 1978 similar proportions of cases and controls lived in an urban environment, with only 31% and 35% respectively, living in a rural area. In 2001, slightly higher percentages of cases (40%) and controls (46%) lived in rural areas.

Approximately 55% of the participants were automatically matched, 40% were matched using cross streets, and 5% matched using town centroid in 1978. For residences occupied in 2001, 67% were automatically matched and 33% were matched using cross streets; none required matching to town centroid. Among the cases, 62% were automatically geocoded in 1978 and 71% in 2001 (Table 1). Among the controls, 55% were automatically geocoded in 1978 and 65% in 2001. Approximately half of the cases and half of the controls lived in the same residence in 1978 and 2001.

Self-reported proximity was first compared with satellite-derived proximity using the five categories of proximity to farmland recorded in the questionnaire (on a farm, <1/4 mile, 1/4–1 mile, 1–5 miles, and >5 miles). Similar levels of association were observed for cases and controls in 1978 and 2001 (Table 2). In 1978, the percentage match between self-report and land cover information was 19% for cases and 20% for controls. In 2001, the match was 26% for cases and 27% for controls. These similarities between cases and controls, and the slight improvements in 2001 compared with 1978 are also reflected in the values of the weighted Kappa statistic and the Spearman correlation coefficient (Table 2).

Comparisons were also examined using two categories of proximity to farmland (1/4 mile from farmland and >1/4 mile from farmland) for cases and controls in 1978 and 2001. Overall, specificity results were between 0.95 and 0.99, sensitivity between 0.39 and 0.61, and percentage match between 74% and 87% (Table 3a). Similar results were observed among controls comparing 1978 with 2001. However, higher sensitivity and percentage match were observed in cases in 2001 compared with those in 1978. In comparing cases with controls, a lower sensitivity was observed for cases in 1978 and a higher percentage match in 2001. These results are consistent with those from logistic regression analyses (Table 3b and c) in which cases exhibit better agreement between proximity approaches than controls for 2001 (adjusted OR=1.74; 95% CI=1.01, 2.99) and worse agreement for 1978, although not significantly (adjusted OR=0.74; 95% CI=0.47, 1.16). When comparing 2001 with 1978, cases showed better agreement in 2001 (adjusted OR = 2.36; 95% CI=1.33, 4.18); no difference was seen among controls (adjusted OR = 0.97; 95% CI=0.66, 1.42). Differences between unadjusted and adjusted analyses were minor, signifying limited influence of the factors including gender, education, age, years spent at residence, geocoding accuracy, and whether or not lived at the same residence in 1978 and 2001.

The influence of gender, education, age, years spent at residence, and geocoding accuracy on agreement between proximity approaches was further examined among smaller strata composed of cases for 1978, cases for 2001, controls for 1978, and controls for 2001 (Table

4). Overall, gender, age, and geocoding accuracy did not influence recall or agreement between self-reported proximity and satellite-derived proximity approaches in any of the strata. College education improved agreement between proximity approaches among controls for 1978 (compared with no college education; OR =2.06; 95% CI = 1.18, 3.59); no effect of college education was detected in the other strata. Among controls reporting proximity to farmland in 2001, individuals with greater number of years spent at a residence showed better recall than individuals with fewer years at a residence (OR=1.02; 95% CI=1.00, 1.05). This relationship was also significant among controls for 2001 using 45 years at residence *versus* 5 as categorical data (OR = 2.72; 95% CI = 1.17, 6.29). No effect of duration at a residence was seen in the other strata.

Discussion

This study examined differences between cases and controls in two calendar years — 1978 and 2001 — and assessed their ability to self report proximity to farmland compared against a “gold standard” derived from satellite data. Using the five categories of proximity to farmland recorded in the questionnaire (on a farm, <1/4 mile, 1/4–1 mile, 1–5 miles, 45 miles), percentage match between self-reported and satellite-derived proximity measures ranged from 19% to 27%, with Spearman correlation coefficients from 0.52 to 0.70, indicating substantial misclassification. Differences were not observed between cases and controls; however, better agreement was reported in 2001 compared with 1978. Data were collapsed into two categories (< 1/4 mile from farmland and >1/4 mile from farmland) to compare in a different manner the classification errors that existed in the five category analyses. Approximately one-third of the participants classified themselves as living >5 miles from farmland; however, the satellite-derived data shows all participants in both years living <5 miles from farmland. In collapsing the data into two categories, percentage match increased substantially from 74% to 87%, indicating less exposure misclassification using this categorization system. We also found that cases exhibited significantly better recall than controls in 2001 and worse recall in 1978 (although not significantly), indicating potential for recall bias. When comparing cases to cases and controls to controls over time, cases showed better agreement in 2001 compared to 1978, while there was no reporting difference in controls between these two time periods. Differences between unadjusted and adjusted analyses were minor, suggesting that gender, education, age, years spent at residence, geocoding accuracy, and whether or not lived at the same residence (in 1978 and 2001) did not help to explain differences in agreement between 2001 and 1978 or between cases and controls. Our findings of case–control differences in the two category analyses raise some concerns about the use of self-reported proximity to farmland in exposure assessments.

Rull et al. (2006) conducted a similar study in California (although it was atemporal) and observed some differences within strata of case and control mothers. Sensitivity was poor and tended to be higher for cases; conversely, specificity was high and similar in magnitude among cases and controls. In general, case mothers were more likely to accurately report residential proximity to crops than control mothers. Our results using the two category classification also indicate some case–control differences. While specificity was high for cases and controls for both years (0.95–0.98), sensitivity was higher for cases than controls in 2001 and lower for cases in 1978; similar trends were observed in the logistic regression

analyses. Given the high specificity and lower sensitivity values, if exposure is rare, then misclassification would not be a major concern. However, considering differential recall between cases and controls or between different years in addition to exposure misclassification raises slight concern about the use of self-reported proximity to farmland over time. Our findings are also comparable to Rull et al. (2006) in that we both investigated whether particular demographic characteristics contributed disproportionately to differences between cases and controls and their ability to remember proximity to farmland. Our overall findings were not well explained by demographic factors such as age and gender although education played a small role in controls for 1978. Recall among controls for 1978 was better if controls had some college education, compared with no college education. Rull et al. (2006) also found differential reporting in education along with differential reporting in other demographic characteristics, including geographic regions, urban and rural residents, and across levels of maternal employment. We found it surprising that educational attainment influenced historical recall among controls but not among cases and feel that this deserves further investigation.

As with any survey-based study, certain limitations exist. Participant errors contributing to misclassification of farm status might include inaccurate estimates of distance to a farm or inaccurate classification of a farm. In a few instances, the 92 cases and 178 controls who lived at the same residence for both time periods (1978, 2001) may have changed farm status from living on or near farmland in 1978 to living further away from farmland in 2001. The questionnaire only asks proximity to farmland once for each residence, and therefore participants may not have thought about changing farm status over the course of this 23-year period. Since about half of the cases and half of the controls live at the same residence for both years, we expected no differential recall. However, within the control group we did observe better recall in 2001 among participants who spent greater number of years at their residence compared with those controls in 2001 who spent fewer years at their residence. It is not surprising that participants living at a residence for a longer period of time give more reliable proximity responses when there is no change in farmland proximity; however, it is surprising that recall among cases was also not influenced by duration of time at a residence.

Errors in address geocoding can also contribute to misclassification of proximity to farmland, and therefore affect proximity measures. One such error occurs when a participant cannot remember an exact street number. Other ambiguities exist when using typical geocoding procedures, especially in rural areas. Actual housing structures may lie up to 200 ft away from a road or be obstructed by a crop field or vineyard (Rull et al. 2006; Ward et al. 2006); the geocoding procedure, however, cannot automatically detect this and places the housing structure on the road or somewhere near the crop field or vineyard. This results in misclassification error. We believe though that these types of errors should occur equally for both cases and controls especially given their similarly proportioned urban–rural status (Table 1). We also adjusted for geocoding status in our analysis and found that it was not significant for cases and controls nor did it play a role on recall over the two time periods.

We referred earlier to the satellite-derived land cover maps as a “gold standard”; however, imperfections exist in these data. The maps for 1978 and 2001 were not produced in the same manner. The 2001 file is derived from a Landsat TIFF image, whereas the 1978 data

are a combination of CAD files from planning offices and their subcontractors. In 1978, 54% of the agricultural land in the study is classified as cropland, rotation, and permanent pasture. Where as in 2001, agricultural land decreased to 50% of the study area. Even though many crops or orchards may enlarge, change location, or shrink during years between surveys (Rull and Ritz, 2003), a decrease of 5%–10% of agricultural land from 1978 to 2001 seems reasonable given the conversion of farmland to subdivisions in the suburban parts of the study area.

Our analysis would also have been improved if the category of farmland was specific to different types of crops and if the classification system for the land use maps was more consistent across time. The 1978 file contained four classifications for cropland and the 2001 file contained five such classifications (the appendix lists all land use categories in both years). To compare these files, we were forced to group different types of agricultural crops such as orchards, vineyards, row crops, or nonpermanent crops into one category and refer to it as “farmland”. Results might vary if the definition of farmland is narrowed to a specific type of crop such as row crop, vineyard, corn or soybeans and the land use file follows this categorization. For instance, Rull et al. (2006) compared proximity to any agricultural crop (similar to our proximity to farmland), but also evaluated proximity to three specific crops — any nonpermanent crops, any orchards, and any vineyards. Their findings, however, did not suggest an association between neural tube defects and general proximity to any crops or specific crop types, except for vineyards. Comparing proximity in our study to only the orchards and vineyards classification of the land use files does not improve agreement between self-reported and land use-derived proximity measures (data not shown). It is important to note here that without actual data regarding pesticide use or type of pesticides, crop or plant rotation practices, and type of farm (i.e. organic) — it is impossible to describe the intensity of exposure, and therefore conclusions drawn from this study about pesticide use as a proxy for exposure are limited. Instead, this approach is helpful for analyzing self-reported proximity to farmland and quantifying misclassification of proximity to farmland when compared to a satellite-derived data.

This study examined differences between cases and controls in two calendar years — 1978 and 2001 — and assessed their ability to recall proximity to farmland compared against a “gold standard” derived from satellite data. Cases exhibited better agreement between proximity approaches than controls in 2001. When comparing 2001 with 1978, both cases and controls showed better agreement in 2001, but only cases showed a significant difference. Demographic characteristics were also investigated and only education seemed to play a small role for controls in 1978. Limitations do exist in this type of investigation, yet we found that using GIS technology and a “gold standard” of satellite-derived data is useful for evaluating the reliability of self-reported proximity to exposure source. This study suggests that proximity measures taken from satellite-derived land cover imagery may be useful for assessing proximity to farmland and it raises some concerns about the use of self-reported proximity to farmland in exposure assessments.

Acknowledgements

We thank the participants of this study for taking part in this research. We also thank Jerome Nriagu, principal investigator of the case-control study, for providing access to this data set, and Melissa Slotnick, Stacey Fedewa, Nicholas Mank, Caitlyn Meservey, and Taylor Builee for valuable assistance with data collection and data entry. We are grateful to the Michigan State Cancer Registry and the Michigan Public Health Institute for assisting with participant recruitment. This research was funded by the National Cancer Institute, Grant R01 CA96002-10.

References

- AvRuskin GA, Jacquez GM, Meliker JR, Slotnick MJ, Kaufmann AM, Nriagu JO. Visualization and exploratory analysis of epidemiologic data using a novel space time information system. *Int J Health Geogr.* 2004; 3(26):1–10. [PubMed: 14748927]
- Brody JG, Aschengrau A, McKelvey W, Rudel RA, Swartz CH, Kennedy T. Breast cancer risk and historical exposure to pesticides from wide-area applications assessed with GIS. *Environ Health Perspect.* 2004; 112(8):889–897. [PubMed: 15175178]
- Burra TA, Elliott SJ, Eyles JD, Kanaroglou PS, Wainman BC, Muggah H. Effects of residential exposure to steel mills and coking works on birth weight and preterm births among residents of Sydney, Nova Scotia. *Can Geogr.* 2006; 50(2):242–255.
- Byass JB, Lake JR. Spray drift from a tractor-powered field sprayer. *Pestic Sci.* 1977; 8(2):117–126.
- Choi HS, Shim YK, Kaye WE, Ryan PB. Potential residential exposure to toxics release inventory chemicals during pregnancy and childhood brain cancer. *Environ Health Perspect.* 2006; 114(7): 1113–1118. [PubMed: 16835067]
- Frost KR, Ware GW. Pesticide drift from aerial and ground applications. *Agric Eng.* 1970; 51(8):460–464.
- Lu CS, Fenske RA, Simcox NJ, Kalman D. Pesticide exposure of children in an agricultural community: evidence of household proximity to farmland and take home exposure pathways. *Environ Res.* 2000; 84(3):290–302. [PubMed: 11097803]
- Meliker JR, Slotnick MJ, AvRuskin GA, Kaufmann AM, Jacquez GM, Nriagu JO. Improving exposure assessment in environmental epidemiology: application of spatio-temporal visualization tools. *J Geogr Syst.* 2005; 7(1):49–66.
- Meyer KJ, Reif JS, Veeramachaneni DNR, Luben TJ, Mosley BS, Nuckols JR. Agricultural pesticide use and hypospadias in eastern Arkansas. *Environ Health Perspect.* 2006; 114(10):1589–1595. [PubMed: 17035148]
- Nuckols JR, Ward MH, Jarup L. Using geographic information systems for exposure assessment in environmental epidemiology studies. *Environ Health Perspect.* 2004; 112(9):1007–1015. [PubMed: 15198921]
- Reynolds P, Hurley SE, Goldberg DE, Yerabati S, Gunier RB, Hertz A, et al. Residential proximity to agricultural pesticide use and incidence of breast cancer in the California Teachers Study cohort. *Environ Res.* 2004; 96(2):206–218. [PubMed: 15325881]
- Royster MO, Hilborn ED, Barr D, Carty CL, Rhoney S, Walsh D. A pilot study of global positioning system/geographical information system measurement of residential proximity to agricultural fields and urinary organophosphate metabolite concentrations in toddlers. *J Expo Anal Environ Epidemiol.* 2002; 12(6):433–440. [PubMed: 12415492]
- Rull RP, Ritz B, Shaw GM. Validation of self-reported proximity to agricultural crops in a case-control study of neural tube defects. *J Expo Sci Environ Epidemiol.* 2006; 16(2):147–155. [PubMed: 16047039]
- Rull RR, Ritz B. Historical pesticide exposure in California using pesticide use reports and land-use surveys: an assessment of misclassification error and bias. *Environ Health Perspect.* 2003; 111(13):1582–1589. [PubMed: 14527836]
- Szklo, M.; Nieto, FJ. *Epidemiology: Beyond the Basics.* Gaithersburg, Maryland: Aspen Publishers; 2000.
- Ward MH, Lubin J, Giglierano J, Colt JS, Wolter C, Bekiroglu N, et al. Proximity to crops and residential exposure to agricultural herbicides in Iowa. *Environ Health Perspect.* 2006; 114(6): 893–897. [PubMed: 16759991]

- Ward MH, Nuckols JR, Weigel SJ, Maxwell SK, Cantor KP, Miller RS. Identifying populations potentially exposed to agricultural pesticides using remote sensing and a Geographic Information System. *Environ Health Perspect.* 2000; 108(1):5–12. [PubMed: 10622770]
- Wickre JB, Karagas MR, Folt CL, Sturup S. Environmental exposure and fingernail analysis of arsenic and mercury in children and adults in a Nicaraguan gold mining community. *Arch Environ Health.* 2004; 59(8):400–409. [PubMed: 16268116]
- Xiang H, Nuckols JR, Stallones L. A geographic information assessment of birth weight and crop production patterns around mother's residence. *Environ Res.* 2000; 82(2):160–167. [PubMed: 10662530]
- Yu CL, Wang SF, Pan PC, Wu MT, Ho CK, Smith TJ, et al. Residential exposure to petrochemicals and the risk of leukemia: using geographic information system tools to estimate individual-level residential exposure. *Am J Epidemiol.* 2006; 64(3):200–207. [PubMed: 16754633]

Appendix 1

1978 Land use categories sorted by percentage of study area covered.

Class name 1978	% Study area covered
Cropland, rotation, and permanent pasture ^a	53.29
Single family, duplex	8.33
Central hardwood	7.77
Herbaceous rangeland	6.99
Lowland hardwood	6.06
Shrub rangeland	4.67
Shrub/scrub wetland	2.68
Lakes	1.24
Wooded wetland	1.03
Emergent wetland	0.90
Permanent pasture ^a	0.76
Outdoor recreation	0.71
Institutional	0.60
Open pit	0.60
Industrial	0.51
Pine	0.45
Neighborhood business	0.43
Orchards, vineyards, and ornamental ^a	0.42
Road transportation	0.41
Multi-family-low rise	0.25
Aspen, birch	0.23
Reservoirs	0.20
Aquatic bed wetland	0.15
Industrial park	0.14
Mobile home park	0.13
Other agricultural land ^a	0.13
Lowland conifer	0.13
Utilities, waste disposal	0.12
Cemeteries	0.10

Class name 1978	% Study area covered
Air transportation	0.09
Commercial, services, and institutional	0.07
Central business district	0.07
Other upland conifer	0.06
Shopping center, mall	0.05
Confined feeding operations	0.05
Streams and waterways	0.04
Open and other	0.04
Christmas tree plantation ^a	0.04
Rail transportation	0.02
Communication facilities	0.01
Multi-family-medium to high rise	0.01
Flats	0.01
Northern hardwood	0.01
Underground extractive	0.00
Beaches and riverbanks	0.00
Wetlands	0.00
Wells	0.00
Transportation, communication, and utilities	0.00
Forested land	0.00
Barren	0.00
Strip commercial	0.00
Water transportation	0.00
Coniferous forest	0.00
Broadleaved forest (generally deciduous)	0.00
Sand other than beaches	0.00

^aLand use categories used to describe a “farm” for this study.

Appendix 2

2001 Land use categories sorted by percentage of study area covered.

Class name 2001	% Study area covered
Forage crops ^a	29.16
Row crops ^a	20.46
Herbaceous openland	7.27
Mixed upland deciduous	5.69
Roads/pavement	4.39
Lowland deciduous forest	4.33
Aspen type	3.64
Northern hardwoods	3.09
Emergent wetland	2.93

Class name 2001	% Study area covered
Low-intensity urban (residential)	2.84
Oak type	2.65
Lowland shrub	2.17
High-intensity urban	1.93
Mixed non-forest wetland	1.83
Pines	1.69
Water	1.23
Upland shrub	1.13
Upland mixed forest	1.01
Floating aquatic	0.70
Parks, golf courses	0.52
Non-vegetated agriculture ^a	0.26
Lowland coniferous forest	0.22
Orchards/vineyards/nursery ^a	0.21
Other upland deciduous	0.17
Other conifers	0.15
Mud flats	0.10
Other bare\sparsely vegetated	0.10
Sand, soil	0.07
Lowland mixed forest	0.02
Airports	0.02
Exposed rock	0.00
Low-density trees	0.00
Mixed upland conifers	0.00
Non-stocked forest	0.00
Christmas tree plantation ^a	0.00

^aLand use categories used to describe a “farm” for this study.



Figure 1.
Map of study area.



Figure 2.
Buffer around geocoded residence of four participants in 2001.

Table 1

Characteristics of cases and controls.

	Cases (N = 184; no. (%))		Controls (N = 347; no. (%))	
<i>Demographic characteristics</i>				
Male	143	77.7	309	89.0
Female	41	22.3	38	11.0
Age >70	103	56.0	175	50.4
Age 70	81	44.0	172	49.6
Some college education	99	53.8	242	69.7
No college education	85	46.2	105	30.3
Lived at same residence in 1978 and 2001	92	50.0	178	51.3
<i>1978 residential characteristics</i>				
Urban ^a	127	69.0	225	64.8
Rural ^a	57	31.0	122	35.2
Automatic geocode	114	62.0	190	54.8
Geocode with assistance	70	38.0	157	45.2
>5 years at current residence	167	90.8	313	90.2
5 years at current residence	17	9.2	34	9.8
<i>2001 residential characteristics</i>				
Urban ^a	110	59.8	187	53.9
Rural ^a	74	40.2	160	46.1
Automatic Geocode	131	71.2	224	64.6
Geocode with assistance	53	28.8	123	35.4
>5 years at current residence	175	95.1	321	92.5
5 years at current residence	9	4.9	26	7.5
No change in proximity to farm from 1978 to 2001 ^b	130	70.7	237	68.3
Change in self-reported proximity to farm from 1978 to 2001	54	29.4	110	31.7

^aBased on 1990 Census urbanized areas.^bIncludes participants who maintained the same residence in 1978 and 2001.

Table 2
 Comparison of participant recall proximity to farmland and satellite-derived proximity to farmland: five categories of proximity to farmland by case-control status and by calendar years of residence (1978/2001).

	Participant recall: proximity to farmland					% match	K_{weighted}	Spearman
	On a farm	<1/4 mile	1/4-1 mile	1-5 miles	>5 miles			
Using satellite-derived land cover information: proximity to farmland Cases, 1978 (N = 184)								
On a farm	5	5	1	1	0	19	0.38*	0.52*
<1/4 mile	4	13	10	15	16			
1/4-1 mile	0	2	6	21	31			
1-5 miles	0	0	2	11	32			
>5 miles	0	0	0	0	0			
Cases, 2001 (N = 184)								
On a farm	9	7	2	1	1	26	0.50*	0.70*
<1/4 mile	2	15	10	6	1			
1/4-1 mile	0	0	8	26	10			
1-5 miles	0	1	3	15	58			
>5 miles	0	0	0	0	0			
Controls, 1978 (N = 347)								
On a farm	12	15	6	1	0	20	0.40*	0.55*
<1/4 mile	10	28	16	18	15			
1/4-1 mile	3	4	12	46	60			
1-5 miles	0	3	0	17	66			
>5 miles	0	0	0	0	0			
Controls 2001 (N = 347)								
On a farm	20	9	8	2	0	27	0.50*	0.68*
<1/4 mile	6	34	24	21	2			
1/4-1 mile	1	8	8	35	30			
1-5 miles	0	2	0	33	93			
>5 miles	0	0	0	0	0			

* $P < 0.0001$.

Table 3

a. Validity of recall of proximity to farmland stratified by case-control status and by calendar years of residence (1978/2001): sensitivity, specificity, and percentage match

	Controls (N = 347)			Cases (N = 184)			1978			2001		
	1978	Comparing 2001, 1978 (P-value)	1978	2001	Comparing 2001, 1978 (P-value)	1978	Controls (N = 347)	Cases (N = 184)	Comparing controls, cases (P-value)	Controls (N = 347)	Cases (N = 184)	Comparing controls, cases (P-value)
Specificity	0.95	0.83	0.98	0.99	0.60	0.60	0.95	0.98	0.35	0.95	0.99	0.06
Sensitivity	0.55	0.90	0.39	0.61	0.02*	0.02*	0.54	0.39	0.05*	0.55	0.61	0.51
% Match	80	0.92	74	87	<0.01*	<0.01*	80	74	0.14	80	87	0.04*

b. Validity of recall of proximity to farmland, influence of case-control status for 1978 and 2001: odds ratios

	2001			1978		
	Unadjusted OR	95% CI	Multivariate-adjusted OR*	Unadjusted OR	95% CI	Multivariate-adjusted OR*
Controls (N = 347)	1.00		1.00			
Cases (N = 184)	1.77	1.05, 2.97	1.74	1.01, 2.99	0.72	0.47, 1.11
						0.74
						0.47, 1.16

c. Validity of recall of proximity to farmland, influence of years of residence (1978/2001) among cases and controls: odds ratios

	Cases (N = 184)			Controls (N = 347)		
	Unadjusted OR	95% CI	Multivariate-adjusted OR*	Unadjusted OR	95% CI	Multivariate-adjusted OR*
1978	1.00		1.00			
2001	2.41	1.37, 4.22	2.36	1.33, 4.18	0.67, 1.43	0.97
						0.66, 1.42

Specificity: among those participants classified by the satellite-derived land use file as living 40.25 miles from farmland, proportion of those self-reporting to be >0.25 miles from farmland.

Sensitivity: among those participants classified by the satellite-derived land use file as living 0.25 miles from farmland, the proportion of those self-reporting to be 0.25 miles from farmland.

* P<0.05.

CI, confidence interval; OR, odds ratio.

* Adjusted for gender, education, age, years at residence, geocoding accuracy, whether or not same residence for both years.

CI, confidence interval; OR, odds ratio.

* Adjusted for gender, education, age, years at residence, geocoding accuracy, whether or not same residence for both years.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4
Influence of demographic, residential, and geocoding characteristics in predicting recall of proximity to farmland.

	Among cases for 1978			Among cases for 2001			Among controls for 1978			Among controls for 2001		
	N	OR	CI	N	OR	CI	N	OR	CI	N	OR	CI
No college education	85	1.00		85	1.00		105	1.00		105	1.00	
Some college education	99	0.89	0.45, 1.77	99	0.98	0.40, 2.40	242	2.06	1.18, 3.59	242	1.22	0.69, 2.14
Female	41	1.00		41	1.00		38	1.00		38	1.00	
Male	143	1.01	0.45, 2.27	143	0.76	0.28, 2.10	309	1.61	0.60, 4.31	309	1.71	0.64, 4.56
Age ^a	184	1.02	0.98, 1.05	184	1.03	0.98, 1.08	347	1.00	0.97, 1.03	347	1.01	0.99, 1.04
Years spent at residence ^a	184	1.01	0.98, 1.03	184	1.02	0.99, 1.06	347	1.00	0.98, 1.02	347	1.02	1.00, 1.05
Not automatically geocoded	70	1.00		53	1.00		157	1.00		123	1.00	
Automatically geocoded	114	0.88	0.44, 1.76	131	0.87	0.33, 2.27	190	0.97	0.57, 1.67	224	0.96	0.55, 1.67

CI, confidence interval; OR, odds ratio.

^a Age and years spent at residence are considered continuous variables and therefore there is no reference group.