American Journal of Epidemiology

Vol. 173, No. 2
DOI: 10.1093/aje/kwq350
Advance Access publication:
November 17, 2010

Practice of Epidemiology

Accuracy of Commercially Available Residential Histories for Epidemiologic Studies

Geoffrey M. Jacquez*, Melissa J. Slotnick, Jaymie R. Meliker, Gillian AvRuskin, Glenn Copeland, and Jerome Nriagu

* Correspondence to Dr. Geoffrey M. Jacquez, Biomedware, 3526 W. Liberty Road, Suite 100, Ann Arbor, MI 48103 (e-mail: jacquez@biomedware.com).

Initially submitted May 27, 2010; accepted for publication September 16, 2010.

A key problem facing epidemiologists who wish to account for residential mobility in their analyses is the cost and difficulty of obtaining residential histories. Commercial residential history data of acceptable accuracy, cost, and coverage would be of great value. The present research evaluated the accuracy of residential histories from LexisNexis, Inc. The authors chose LexisNexis because the Michigan Cancer Registry has considered using their data, they have excellent procedures for privacy protection, and they make available residential histories at 25 cents per person. Only first and last name and address at last-known residence are required to access the residential history. The authors compared lifetime residential histories collected through the use of written surveys in a case-control study of bladder cancer in Michigan to the 3 residential addresses routinely available in the address history from LexisNexis. The LexisNexis address matches, as a whole, accounted for 71.5% of participants' lifetime addresses. These results provided a level of accuracy that indicates routine use of residential histories from commercial vendors is feasible. More detailed residential histories are available at a higher cost but were not analyzed in this study. Although higher accuracy is desirable, LexisNexis data are a vast improvement over the assumption of immobile individuals currently used in many spatial and spatiotemporal studies.

data collection; residential mobility; validation studies

In the present study, we assessed the availability of residential mobility data for routine use in cancer surveillance and conducted an accuracy assessment of commercially supplied residential histories compared with those collected by written survey. Although our focus was on cancer, the results pertain in general to epidemiologic studies of longer-term health outcomes, correlates, and predictors that act over an individual's life course. There is a growing recognition that residential mobility must be accounted for in epidemiologic studies of cancer (1–9), for evaluation of clusters (4, 10–22), for reconstructing exposures (1, 23–33), and as a source of exposure misclassification (34–38). Whereas residential histories are routinely recorded for many European countries, researchers in the United States typically obtain residential histories through interviews, a time-consuming and expensive task that is subject to recall error. Furthermore, the address geocoding process can introduce substantial positional errors into the resulting x, y coordinates (39–46), and improved mechanisms are needed for obtaining the address records themselves and for validating addresses obtained using traditional survey instruments (1, 2, 23, 25, 39–41, 47–56). We thus decided to use a representative case-control study of bladder cancer in Michigan (3) for which residential histories had been collected by survey. We compared these with residential histories obtained from LexisNexis (http://www.LexisNexis.com), using each participant's reported first and last name and last street address. When this study was conducted, the data provider was known as ChoicePoint. In 2008, Reed Elsevier completed its purchase of ChoicePoint and announced that the company would be combined with the LexisNexis Risk and Information Analytics Group. The ChoicePoint product originally used in this study is now available from LexisNexis as LexisNexis Risk Solutions. Should commercial residential history data be of acceptable accuracy, their low cost and broad coverage would be of great value to all researchers requiring residential history data. To our knowledge, the present study is the first to compare and contrast

Table 1. Sample Structure of Survey and LexisNexis Data Sets, 2008–2009

Random Identification No.	First Name	Last Name	Address	City	State	Zip Code	Date In	Date Out
001234	John	Doe	456 State Street	Ann Arbor	MI	48103	1/1/1960	1/1/1965
001234	John	Doe	Grand River	Lansing	MI	48906	1/1/1965	5/1980
001234	John	Doe	123 Main Street	Royal Oak	MI	48067	5/1980	8/3/2004

commercially available residential history data with those obtained from a survey.

MATERIALS AND METHODS

Study participants and data sources

Residential history data compiled from a bladder cancer case-control study conducted in 11 counties of southeastern Michigan were compared with addresses for the participants provided by LexisNexis. For the case-control study, bladder cancer patients $\leq$80 years of age upon diagnosis were recruited from the Michigan State Cancer Surveillance Program, and controls were selected from an age-weighted list using a random digit dialing procedure. Controls were frequency-matched to cases based on age ($\pm$5 years), race, and gender. Recruitment was limited to individuals who had lived anywhere in 1 of the 11 counties in the study area (Genesee, Huron, Ingham, Jackson, Lapeer, Livingston, Oakland, Sanilac, Shiawassee, Tuscola, and Washtenaw) for $\geq$5 consecutive years before being contacted. Participants with a prior cancer diagnosis were excluded, with the exception of those with nonmelanoma skin cancers. All participants were assigned a random identification number to maintain confidentiality. The bladder cancer study was approved by the University of Michigan Institutional Review Board-Health Committee. Further details on the study design have been published elsewhere (57).

One aspect of the case-control study involved collection of detailed information on where participants lived throughout their lifetime. Each participant was asked to complete a written residential history form detailing each address at which they had lived for >1 year since birth. The residential history forms were mailed to participants and participants were asked to complete the forms at home; the surveys were later collected and reviewed by researchers during a home visit. The street name, street number, city, state, and zip code were requested for each residence in as much detail as possible. If participants were unable to recall the exact address, they were asked to provide the closest cross streets. For each address, participants were also asked to recall the dates on which they had moved into and out of the residence. These data were double-entered into a database and cross-checked for any discrepancies; if discrepancies were found, the original written document was consulted. The residential histories collected by survey were not subjected to any external validation beyond that described above.

A data set containing 3 addresses for each participant was purchased from LexisNexis. The LexisNexis data contained the name, address, state, zip code, and dates for which residence at a particular address commenced and stopped. We used the 3 most recent addresses because these are routinely available in the LexisNexis ''Batch 411'' product, providing the histories at 25 cents per person while masking protected data such as Social Security numbers. LexisNexis data were matched to study participants using first and last name, street address, city, state, and zip code. These parameters were used to mimic the type of data that might be routinely available to cancer researchers and cancer registries.

Survey data for 994 cases and controls were available for use in the study described herein. LexisNexis data were not available for 40 participants who provided addresses from the survey, and LexisNexis provided addresses for 8 participants with missing survey data. Therefore, address accuracy assessment was conducted for 946 individuals.

Database manipulation and structure

Both the LexisNexis and survey data sets were compared and manipulated using Microsoft Excel. To allow for comparison between the 2 datasets, random identification numbers associated with each individual from the bladder cancer study were assigned to the names and addresses listed in the LexisNexis data. Each database was then structured so that each row in the database contained a different address (Table 1). It is important to note that temporal mismatch existed between the LexisNexis data and the survey data; this issue is quantified and discussed in more detail below. In addition, the LexisNexis data contained up to 3 addresses per participant, whereas the survey data contained lifetime residential history data.

Quantification of accuracy assessment

Five different metrics were created to quantify the accuracy of the LexisNexis data in matching various aspects of the survey data. As recall bias existed in the survey data, we could not treat the residential histories from the surveys as error-free. Hence, lack of matching between the survey and LexisNexis records may be explained by inaccuracies in one or both of the data sources. These limitations are addressed below. The 5 metrics used to assess accuracy of the LexisNexis data are also listed below (Table 2).

Metrics 1, 2, and 3 built on each other and assessed the basic ability of the LexisNexis data to accurately reflect results reported by study participants. Metric 4 involved a calculation of the number of years represented at each address, both for the LexisNexis data and as reported in the survey by study participants. Years spent at each address were graphed against each other using Microsoft Excel

Table 2. Metrics Calculated to Assess LexisNexis Data Accuracy, 2008–2009

Metric	Title	Description
1	City match	LexisNexis matches survey city only
2	Street match	LexisNexis matches survey city and street
3	Detailed match	LexisNexis matches survey city, street, and address number
4	Years at address	Distribution of years listed for LexisNexis and survey addresses
5	Survey years	Survey years correctly accounted for by metrics 2 and 3

(Microsoft, Seattle, Washington), and a Pearson correlation coefficient was calculated. Lastly, metric 5 involved an assessment of the number of years a participant reported spending at an address that had a successful match for metrics 2 and 3.

A match-point value of 1.0 was assigned for metrics in which the match criteria were completely met. For a small number of participants, a street name or number was listed for the LexisNexis data but was not recalled by the participant. If this was the case, a score of 0.5 was assigned, based on the assumption that if a participant had been able to accurately recall the address, it would have matched the LexisNexis data. The accuracy of this assumption was tested by selecting LexisNexis addresses with a 0.5 match score for metric 2 (n = 122). These addresses were then cross-referenced with the survey data to determine how many participants reported cross-streets. For those providing cross-streets (n = 85), the address provided by LexisNexis was mapped using Google maps. The map was checked for the reported cross-streets, and the number of LexisNexis addresses for which the cross-streets verified the location was recorded.

Lastly, the levels of completeness of both the survey data and the LexisNexis data were assessed. The numbers of missing addresses for each data set were tallied, and basic statistics regarding completion of the survey data were calculated.

RESULTS

Addresses provided by LexisNexis matched the addresses provided by participants in the survey data for both street number and street name (detailed match) for 53% of addresses (Table 3). Even more addresses were successfully matched if only street name or city was specified in the match criteria (Table 3). Furthermore, the majority of street matches (85%) were also detailed matches.

Table 3. Percentage of LexisNexis Addresses (n = 2,388) Matched to Survey Data, 2008–2009

Metric	Metric Title	Total Match Count[a]	Percent Match
1	City match	1,701	71
2	Street match	1,475	62
3	Detailed match	1,259	53

[a] Sum of points, not total count (partial matches = 0.5, complete matches = 1.0 point).

Metric 4 involved calculation of the number of years at each address as reported in the LexisNexis data and the survey data. For the LexisNexis data, descriptive statistics were calculated for both years at each address (raw addresses) and the sum of years at all 3 addresses (aggregated addresses) (Table 4). This metric was created to assess the ability of the LexisNexis data to roughly reflect the time stamp reported by participants at each address. We also reported the statistics on the number of years at each residence for all of the survey data (raw survey data (all residences)), the last 3 addresses for the survey data (raw survey data (last 3 residences)), and the LexisNexis data when time at place of residence was >1 year (raw addresses (>1 year)). This facilitated comparisons between the LexisNexis and survey data, as the survey requested all addresses with occupancy >1 year and we used only the 3 most recent addresses from LexisNexis. These results suggest that, when compared with the raw survey data, the LexisNexis data underestimated time spent at each address (Table 4). Specifically, participants reported spending an average of 9.4 years at each address, whereas the LexisNexis data listed an average of 4.3 years at each address (Table 4). However, the LexisNexis data set has a much greater proportion of addresses in which a time stamp of <1 year is documented (46.4%) (Table 4). Participants in the bladder cancer study were specifically asked to report only addresses at which they had lived for >1 year. This detail of the study design may have resulted in a greater discrepancy when comparing the 2 data sets than actually existed. When we compared only the 3 most recent addresses of ≥1-year duration from both data sets, the average number of years at each address remained higher for the survey data (13.5 years compared with 8.5 years). However, it is worth noting that the sample size was markedly different for the 2 data sets when this comparison was made: 2,800 for the sample data vs. 1,099 for LexisNexis data. The reduction in sample size for the LexisNexis data arises for 2 reasons: either 1 or both time stamps were not provided (e.g., the ‘‘from’’ and ‘‘to’’ dates), or the time interval reported by LexisNexis was <1 year. Most of the matches (Table 4) included addresses listed as one of the 3 most recent by the survey participants. It is possible that recall was likely better for these and also that the LexisNexis data were better, as the addresses in the database were more recent (i.e., technological advances may have improved the database in recent years). This is a question for future study.

Results assessing the difference in years reported at matched addresses by each data set (Table 5) indicated that, on average, the LexisNexis data largely underestimated the amount of time spent at each address. Participants reported spending at each address an average of 12.9 years in excess of the number of years reported in the LexisNexis data (Table 5). These results indicate that there is a degree of temporal mismatch between the LexisNexis data and the survey data. Limiting the comparison to the last 3 addresses reported by participants versus all LexisNexis data and to LexisNexis addresses at which the participants lived for >1

Table 4. Years Spent at Each Address as Reported by LexisNexis and Survey Data (Metric 4), 2008–2009

No. of Years at Residence	LexisNexis Data[a] (3 Residences)			Raw Survey Data (All Residences)	Raw Survey Data (Last 3 Addresses)
	Raw Addresses	Raw Addresses (>1 Year)	Aggregated Addresses		
Average	4.3	8.5	9.9	9.4	13.5
Median	0.9	7.9	9.2	5	9
90th percentile	12.8	14.5	17.1	23	33
75th percentile	7.9	12.1	13.4	13	20
Minimum	<1	1.0	<1	<1	<1
Maximum	34.4	34.4	36.0	74.0	74.0
% of addresses with <1 year residence time	46.4	0	n/a	1.6	0.7
Total no. of addresses with years reported	2,207[b]	1,099	n/a	6,739	2,800[b]

[a] All LexisNexis data are limited to what the vendor determined as the most recent last 3 residences.
[b] Total addresses are not in agreement, as some study participants did not report 3 addresses, and years spent at each address were missing for some LexisNexis addresses.

year versus all survey data did not change the results substantially (Table 5).

The majority (96.8%) of the matched LexisNexis addresses were also in the 3 most recent addresses reported in the survey. This finding is supportive of use of the commercial data from a temporal standpoint, although further analyses are needed to address the ability of the LexisNexis data to accurately reflect time periods spent at each address. For the 49 addresses that were not among the 3 most recent survey addresses reported, a mismatch indicated that the LexisNexis data are not capturing complete mobility for all participants. Therefore, it is of interest to repeat these analyses for a population with increased residential mobility.

The extent of this temporal mismatch is represented in Figure 1. This figure demonstrates that although there is a general positive trend between the number of years reported in the 2 data sets for each address ($r = 0.40$), differences remain, particularly for LexisNexis when the number of years at a given residence is between 7.8 and 8.0 (Figure 1).

To assess the degree to which the LexisNexis data could be used for life-course assessments, the proportion of an individual's lifetime residential history correctly accounted for by the LexisNexis data was also calculated (Table 6). This approach would be comparable to a researcher's using only the LexisNexis addresses with positive match scores for street or detailed match criteria to account for geographical placement throughout the lifetime, irrespective of temporal mobility patterns. These results indicated that, when considering the 3 most recent addresses reported in the survey, the LexisNexis address matches, as a whole, accounted for 71.5% of participants' lifetime addresses (Table 6). Because the LexisNexis data only included 3 addresses per participant, the statistic comparing only 3 addresses from the survey data seemed more logical. In fact, because of the temporal mismatch discussed above, the reported 71.5% is likely an underestimation of the accuracy of the LexisNexis data. Results comparing all survey addresses are also presented in Table 6, and the LexisNexis address matches accounted for 42.6% of participants' lifetime addresses, a considerable decrease from the 71.5% match reported

Table 5. Differences[a] in Number of Years Reported at Each Address[b] Between Survey Data and LexisNexis Data, 2008–2009

Descriptive Statistic	Difference, years		
	All Survey and LexisNexis Data	Last 3 Survey and All LexisNexis Data	All Survey and LexisNexis >1 Year
Mean	12.9	13	12.4
Median	8.7	8.8	8.7
75th percentile	21.1	21.3	21.2
90th percentile	32	32	30
Total no. of matched addresses with years reported in both data sets	1,398	1,360	896

[a] Survey minus LexisNexis years.
[b] Addresses assessed were those matched according to metric 1 or metric 2.

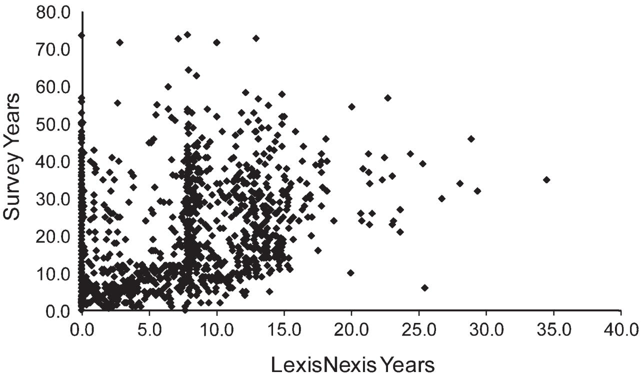

Figure 1. Relation between years reported at each address by participants and years reported at the same address by the LexisNexis data ($r = 0.40$). It includes only addresses with a detailed or street match score of 0.5 or 1.0, for which number of years at residence were provided by both the LexisNexis and survey data sets.

above. LexisNexis does provide longer address histories for an additional cost, but we did not use them in this study.

Partial matches, or matches for which a score of 0.5 was assigned because of incompleteness of the survey data, accounted for 122 (7.9%) of the 1,536 street matches and 40 (3.1%) of the 1,279 detailed matches. Of the 122 street matches with a 0.5 score, participants gave cross-streets for 85. Cross-streets helped to show that the LexisNexis address was correct for 62 (72.9%) of these 85 addresses.

Lastly, to further quantify agreement between the 2 data sets, the number of missing addresses was quantified for both the LexisNexis data and the survey data. These results indicated that the percentage of missing addresses was similar when comparing the 3 most recent addresses from the survey data with the LexisNexis data (Table 7). When all addresses from the survey data were compared, however, the percentage of missing addresses increased (Table 7). This is likely an indication of difficulty on the part of the participant in recalling past addresses.

Table 6. Total Number of Lifetime Years Spent at Each Address and Percentage of Lifetime History Reported by Participant, as Accounted for by Matched LexisNexis Data[a] (Metric 5), 2008–2009

	All Survey Addresses	3 Most Recent Survey Addresses
Total years at address, reported by participant		
Detailed match only	23,610.2	23,610.2
Street or detailed match	26,938.3	26,938.3
Total years of survey data	63,305.8	37,695.7
Percentage of lifetime history years accounted for by:		
Detailed match only	37.30	62.63
Street or detailed match	42.55	71.46

[a] LexisNexis addresses with a score of 0.5 or 1.0 for metric 2 and/or metric 3 were used.

Table 7. Quantification of Missing Addresses for Both Survey[a] and LexisNexis[b] Data, 2008–2009

Data Set	No. of Missing Addresses	Total No. of Addresses Expected	% of Addresses Missing
Survey, all data	1,974	6,754	29.2
Survey, last 3 addresses	458	2,801	16.4
LexisNexis data	450	2,838	15.9

[a] Missing addresses are defined as those for which a street address is missing but a city and/or state is provided.

[b] Missing addresses are defined as blank addresses in the LexisNexis data.

Finally, the percentage of years with missing addresses from the survey data was calculated. Of 37,701.5 total years for all study participants, addresses were missing for 3,130.1 (8.3%) of these years. If considering the use of LexisNexis data to supplement last 3 residences in this data set, it is estimated that 79.8% (8.3% + 71.46%) of total lifetime years are accounted for by either missing addresses (8.3%) or a street or detailed match with the LexisNexis data (71.46%).

DISCUSSION

The results discussed herein suggest that the LexisNexis data may be promising in supplementing collection of residential history data in epidemiologic investigations. In particular, the fact that positive matches (either detailed or street matches) between the LexisNexis data and the survey data accounted for approximately 70% of participants' 3 most recent residences suggests that the addresses provided by the LexisNexis data are fairly accurate.

The LexisNexis data were successful in matching the survey data in 71% of addresses when the match criterion was city alone. Such a match may be adequate for assessing some exposures that may occur by city geography, such as exposure to contaminants in public drinking water supplies. The LexisNexis data also successfully matched the survey data by metric 2 (street match) and metric 3 (detailed match) 62% and 53% of the time, respectively. Furthermore, results from a small pilot study using this data set presented similar findings, suggesting that it may be feasible for researchers to validate use of LexisNexis data with a small sample size before large-scale implementation. The pilot study randomly selected 25 participants to compare the 3 most recent addresses from the survey data with the LexisNexis data. Results were similar to those presented above, with a street match and detailed match correctly accounting for 64% and 52%, respectively.

One caveat that should be made explicit is that stating 71.5% accuracy over lifetimes implicitly assumes that the older residences that were not represented in the 3 most recent LexisNexis address samples would be available and would have a similar accuracy to the first 3 residences. This is not necessarily the case. Additional studies are needed using participants' entire residential histories to evaluate whether accuracy is sensitive to how long ago that residence was occupied by the participant.

The temporal mismatch between the survey and LexisNexis data likely presents a concern for researchers wishing to substitute collection of residential history data with LexisNexis data. The LexisNexis data could be useful in this capacity if researchers were willing to accept some degree of exposure misclassification by dividing the LexisNexis addresses into equal time periods or by relying on the dates listed by the LexisNexis data. It is also likely that participant recall of years spent at past residences is in error; however, the consistent underestimation from LexisNexis for duration at each residence is a cause for concern in the temporal accuracy.

LexisNexis data may be most helpful to researchers wishing to supplement collection of complete residential histories. Researchers may wish to provide participants with the list of addresses from LexisNexis and ask them to simply fill in the number of years at that address or to make any corrections to the existing data. In this manner, the LexisNexis data could be helpful in reducing recall error. Similarly, the LexisNexis data could be used subsequent to data collection to supplement missing or incomplete addresses from the survey data.

Exposure misclassification resulting from biases or error in recall is a concern in epidemiologic studies. In geography-based exposure assessments, such error is partly due to difficulties in recalling residential mobility. Recall error may be present in the address provided or the amount of time a participant reports spending at that address. For these reasons, there are limitations to using the survey data as a basis for comparison when assessing reliability of the LexisNexis data. Specifically, as mentioned previously, study participants were only asked to report residences at which they had lived for >1 year. Therefore, there may have been addresses captured by the LexisNexis data that were not even considered in the survey data. In addition, completion of the forms varied differentially for each individual, in that some participants provided very detailed accounts of the addresses lived at over their lifetimes, whereas others provided primarily street names, cities, or cross-streets. In this study, the LexisNexis data were successful in supplementing gaps in the survey data nearly 72% of the time for the 3 most recent addresses. Because some participants listed multiple unknown addresses in 1 city and no cross-street was provided for some of those blank addresses, this percentage is likely an underestimation of the true accuracy of the LexisNexis data. This result suggests that the LexisNexis data could be extremely useful in augmenting residential history data collected in epidemiologic studies. Anecdotally, one of the authors recorded his residential history to the best of his recollection and then compared it with that obtained from LexisNexis. Several of the dates from LexisNexis proved more accurate than those obtained by recall; LexisNexis provided street addresses that the author no longer remembered, and 1 actual place of residence that was unintentionally omitted by that author was accurately provided by LexisNexis.

These findings raise the question of how much of the lack of exact matching between residential histories collected by LexisNexis and those collected by written survey is attributable to errors and omissions by LexisNexis and how much is attributable to participant recall error and researcher recording error. This question was not addressed within the scope of the present study, but it is an important one for future research.

As noted above, this study used only the last 3 addresses for reasons of convenience and because these are routinely available in the LexisNexis "Batch 411" product, providing the histories at 25 cents per person while masking protected data such as Social Security numbers. Residential histories of longer record are available at a higher expense (e.g., several dollars per person as opposed to 25 cents). When compared with the cost of obtaining residential history information via survey instruments, this still will usually be a great savings. In practice, the length of the mobility history researchers will wish to obtain should be chosen to reflect the health outcome and/or exposure they are studying, as well as the stage in the life course thought to be impacted. There are some characteristics of the survey data set that may also have influenced the results presented herein. Specifically, this population is a geographically stable population, with only about 20% of the total person-years being spent outside of southeastern Michigan (57). In addition, the average age of this population is 65 years. It would be useful to repeat these analyses with a younger, more mobile population, a task we leave for later study. It is possible that the age distribution of the population may have influenced the ability to recall residences occupied early in life; on the other hand, it is possible that this age distribution may be a primary contributing factor to the stability of the population. For the above reasons, it is critical to repeat this type of accuracy assessment with additional data sets for different types of populations.

It also would be useful to apply this study method to compare residential histories from different commercial vendors. LexisNexis is not the only commercial source for residential history data, and an accuracy evaluation of residential histories from alternative commercial vendors is needed.

ACKNOWLEDGMENTS

Author affiliations: BioMedware, Ann Arbor, Michigan (Geoffrey M. Jacquez, Melissa J. Slotnick, Gillian AvRuskin); Graduate Program in Public Health, Department of Preventive Medicine, Stony Brook University (SUNY), New York, New York (Jaymie R. Meliker); School of Public Health, University of Michigan, Ann Arbor, Michigan (Geoffrey M. Jacquez, Jerome Nriagu); and Michigan Cancer Registry, Michigan Department of Community Health, Lansing, Michigan (Glenn Copeland).

This research was funded in part by grant 2R44CA135818 from the National Cancer Institute to BioMedware. Collection of the survey data was funded by National Cancer Institute grant RO-1 CA96002-10.

The authors thank Stacey Fedewa, Zorimar Rivera, Caitlyn Meservey, Danielle Movsky, and Lisa Bailey for assistance with survey data collection.

The perspectives stated in this manuscript are those of the authors and do not necessarily reflect the official views of the National Cancer Institute. Several of the authors are employed by BioMedware, a research and development company based in Michigan. BioMedware does not have any business relationship with LexisNexis beyond purchasing data to conduct the study.

Conflict of interest: none declared.

REFERENCES

1. Gallagher LG, Webster TF, Aschengrau A, et al. Using residential history and groundwater modeling to examine drinking water exposure and breast cancer. *Environ Health Perspect*. 2010;118(6):749–755.
2. Nuckols J, Airola M, Colt J, et al. The impact of residential mobility on exposure assessment in cancer epidemiology. *Epidemiology*. 2009;20(6):S259–S260.
3. Meliker JR, Slotnick MJ, AvRuskin GA, et al. Lifetime exposure to arsenic in drinking water and bladder cancer: a population-based case-control study in Michigan, USA. *Cancer Causes Control*. 2010;21(5):745–757.
4. Jacquez GM, Meliker JR. Case-control clustering for mobile populations. In: Fotheringham S, Rogerson P, eds. *The SAGE Handbook of Spatial Analysis*. Los Angeles, CA: Sage Publications; 2009:321–329.
5. Kingsley BS, Schmeichel KL, Rubin CH. An update on cancer cluster activities at the Centers for Disease Control and Prevention. *Environ Health Perspect*. 2007;115(1):165–171.
6. Rushton G, Armstrong MP, Gittler J, et al. Geocoding in cancer research: a review. *Am J Prev Med*. 2006;30(suppl 2): S16–S24.
7. Palmer JR, Wise LA, Hatch EE, et al. Prenatal diethylstilbestrol exposure and risk of breast cancer. *Cancer Epidemiol Biomarkers Prev*. 2006;15(8):1509–1514.
8. Vieira V, Webster T, Weinberg J, et al. Spatial analysis of lung, colorectal, and breast cancer on Cape Cod: an application of generalized additive models to case-control data. *Environ Health*. 2005;4:11. (doi: 10.1186/1476-069X-4-11).
9. Reynolds P, Hurley SE, Quach AT, et al. Regional variations in breast cancer incidence among California women, 1988–1997. *Cancer Causes Control*. 2005;16(2):139–150.
10. Jepsen MR, Simonsen J, Ethelberg S. Spatio-temporal cluster analysis of the incidence of *Campylobacter* cases and patients with general diarrhea in a Danish county, 1995–2004. *Int J Health Geogr*. 2009;8:11. (doi: 10.1186/1476-072X-8-11).
11. Goovaerts P. Medical geography: a promising field of application for geostatistics. *Math Geol*. 2009;41:243–264.
12. Tango T. A class of multiplicity adjusted tests for spatial clustering based on case-control point data. *Biometrics*. 2007;63(1):119–127.
13. Meliker JR, Jacquez GM. Space-time clustering of case-control data with residential histories: insights into empirical induction periods, age-specific susceptibility, and calendar year-specific effects. *Stoch Environ Res Risk Assess*. 2007; 21(5):625–634.
14. Jacquez GM, Meliker J, Kaufmann A. In search of induction and latency periods: space-time interaction accounting for residential mobility, risk factors and covariates. *Int J Health Geogr*. 2007;6(1):35. (doi: 10.1186/1476-072X-6-35).
15. McNally RJ, Pearce MS, Parker L. Space-time clustering analyses of testicular cancer amongst 15–24-year-olds in Northern England. *Eur J Epidemiol*. 2006;21(2):139–144.
16. Lix LM, Hinds A, DeVerteuil G, et al. Residential mobility and severe mental illness: a population-based analysis. *Adm Policy Ment Health*. 2006;33(2):160–171.
17. Jacquez GM, Meliker JR, AvRuskin GA, et al. Case-control geographic clustering for residential histories accounting for risk factors and covariates. *Int J Health Geogr*. 2006;5:32. (doi: 10.1186/1476-072X-5-32).
18. Bell BS, Hoskins RE, Pickle LW, et al. Current practices in spatial analysis of cancer data: mapping health statistics to inform policymakers and the public. *Int J Health Geogr*. 2006;5:49. (doi: 10.1186/1476-072X-5-49).
19. Song C, Kulldorff M. Tango's maximized excess events test with different weights. *Int J Health Geogr*. 2005;4:32. (doi: 10.1186/1476-072X-4-32).
20. Ozonoff A, Webster T, Vieira V, et al. Cluster detection methods applied to the upper Cape Cod cancer data. *Environ Health*. 2005;4:19. (doi: 10.1186/1476-069X-4-19).
21. Jacquez GM, Kaufmann A, Meliker J, et al. Global, local and focused geographic clustering for case-control data with residential histories. *Environ Health*. 2005;4:4. (doi: 10.1186/1476-069X-4-4).
22. Hurley SE, Reynolds P, Goldberg DE, et al. Residential mobility in the California Teachers Study: implications for geographic differences in disease rates. *Soc Sci Med*. 2005;60(7): 1547–1555.
23. Spiegelman D. Approaches to uncertainty in exposure assessment in environmental epidemiology. *Annu Rev Public Health*. 2010;31:149–163.
24. Maxwell SK, Meliker JR, Goovaerts P. Use of land surface remotely sensed satellite and airborne data for environmental exposure assessment in cancer research. *J Expo Sci Environ Epidemiol*. 2010;20(2):176–185.
25. Jacquez GM, Meliker JR. Exposure reconstruction using space-time information technology. In: *Encyclopedia of Environmental Health*. Philadelphia, PA: Elsevier. In press.
26. Pujol JM, Eisenberg JE, Haas CN, et al. The effect of ongoing exposure dynamics in dose response relationships. *PLoS Comput Biol*. 2009;5(6):e1000399. (doi:10.1371/journal.pcbi.1000399).
27. Gerharz LE, Krüger A, Klemm O. Applying indoor and outdoor modeling techniques to estimate individual exposure to PM2.5 from personal GPS profiles and diaries: a pilot study. *Sci Total Environ*. 2009;407(18):5184–5193.
28. Meliker JR. Reconstructing individual-level exposure to environmental contaminants using time-GIS. In: Hornsby K, Yuan M, eds. *Understanding Dynamics of Geographic Domains*. Boca Raton, FL: Taylor & Francis; 2008:75–91.
29. AvRuskin GA, Jacquez GM, Meliker JR. Using satellite derived land cover information for a multi–temporal study of self-reported recall of proximity to farmland. *J Expo Sci Environ Epidemiol*. 2008;18(4):381–391.
30. Zandbergen PA. Influence of geocoding quality on environmental exposure assessment of children living near high traffic roads. *BMC Public Health*. 2007;7:37. (doi: 10.1186/1471-2458-7-37).
31. Meliker JR. *Lifetime Exposure to Arsenic in Drinking Water in Southeastern Michigan: Application to a Bladder Cancer Case-Control Study* [dissertation]. Ann Arbor, MI: The University of Michigan; 2006.
32. Reynolds P, Hurley SE, Gunier RB, et al. Residential proximity to agricultural pesticide use and incidence of breast cancer in California, 1988–1997. *Environ Health Perspect*. 2005;113(8):993–1000.
33. Meliker JR, Slotnic M, AvRuskin G, et al. Improving exposure assessment for environmental epidemiology: applications of

a space-time information system. *J Geogr Syst*. 2005;7(1): 49–66.

34. Canfield MA, Ramadhani TA, Langlois PH, et al. Residential mobility patterns and exposure misclassification in epidemiologic studies of birth defects. *J Expo Sci Environ Epidemiol*. 2006;16(6):538–543.
35. Fell DB, Dodds L, King WD. Residential mobility during pregnancy. *Paediatr Perinat Epidemiol*. 2004;18(6):408–414.
36. Schulman J, Selvin S, Shaw GM, et al. Exposure misclassification due to residential mobility during pregnancy in epidemiologic investigations of congenital malformations. *Arch Environ Health*. 1993;48(2):114–119.
37. Khoury MJ, Stewart W, Weinstein A, et al. Residential mobility during pregnancy: implications for environmental teratogenesis. *J Clin Epidemiol*. 1988;41(1):15–20.
38. Rothman KJ. Induction and latent periods. *Am J Epidemiol*. 1981;114(2):253–259.
39. Zandbergen PA. Positional accuracy of spatial data: non-normal distributions and a critique of the national standard for spatial data accuracy. *Trans GIS*. 2008;12(1):103–130.
40. Casaca J, Fonseca AM. Modelling positional errors with isotropic random vector fields. In: Caetano M, Painho M, eds. *Proceedings of the 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, 5 – 7 July 2006*. Lisboa, Portugal: Instituto Geográfico Português; 2008:258–263.
41. Zimmerman DL, Fang X, Mazumdar S, et al. Modeling the probability distribution of positional errors incurred by residential address geocoding. *Int J Health Geogr*. 2007;6:1. (doi: 10.1186/1476-072X-6-1).
42. Griffith DA, Millones M, Vincent M, et al. Impacts of positional error on spatial regression analysis: a case study of address locations in Syracuse, New York. *Trans GIS*. 2007; 11(5):655–679.
43. Ward MH, Nuckols JR, Giglierano J, et al. Positional accuracy of two methods of geocoding. *Epidemiology*. 2005;16(4):542–547.
44. Cayo MR, Talbot TO. Positional error in automated geocoding of residential addresses. *Int J Health Geogr*. 2003;2:10. (doi: 10.1186/1476-072X-2-10).
45. Bonner MR, Han D, Nie J, et al. Positional accuracy of geocoded addresses in epidemiologic research. *Epidemiology*. 2003;14(4):408–412.
46. Dunn R, Harrison A, White J. Positional accuracy and measurement error in digital databases of land use: an empirical study. *Int J Geogr Inf Sci*. 1990;4(4):385–398.
47. Zandbergen PA, Hart TC. Geocoding accuracy considerations in determining residency restrictions for sex offenders. *Crim Justice Policy Rev*. 2009;20(1):62–90.
48. Jacquez GM, Rommel R. Local indicators of geocoding accuracy (LIGA): theory and application. *Int J Health Geogr*. 2009;8:60. (doi: 10.1186/1476-072X-8-60).
49. Zimmerman DL, Fang X, Mazumdar S. Spatial clustering of the failure to geocode and its implications for the detection of disease clustering. *Stat Med*. 2008;27(21): 4254–4266.
50. Zimmerman D. Statistical methods for incompletely and incorrectly geocoded cancer data. In: Rushton G, Armstrong M, Gittler J, et al, eds. *Geocoding Health Data*. Boca Raton, FL: CRC Press; 2008.
51. Zandbergen PA. A comparison of address point, parcel and street geocoding techniques. *Comput Environ Urban Syst*. 2008;32(3):214–232.
52. Mazumdar S, Rushton G, Smith BJ, et al. Geocoding accuracy and the recovery of relationships between environmental exposures and health. *Int J Health Geogr*. 2008;7:13. (doi: 10.1186/1476-072X-7-13).
53. Henry KA, Boscoe FP. Estimating the accuracy of geographical imputation. *Int J Health Geogr*. 2008;7:3. (doi: 10.1186/1476-072X-7-3).
54. Goldberg DW, Wilson JP, Knoblock CA, et al. An effective and efficient approach for manually improving geocoded data. *Int J Health Geogr*. 2008;7:60. (doi: 10.1186/1476-072X-7-60).
55. Goldberg D. *A Geocoding Best Practices Guide*. Springfield, IL: North American Association of Central Cancer Registries; 2008.
56. Abe T, Stinchcomb D. Geocoding practices in cancer registries. In: Rushton G, Armstrong MP, Gittler J, et al, eds. *Geocoding Health Data*. Boca Raton, FL: CRC Press; 2008: 111–125.
57. Meliker JR, Slotnick MJ, AvRuskin GA, et al. Individual lifetime exposure to inorganic arsenic using a space-time information system. *Int Arch Occup Environ Health*. 2007;80(3): 184–197.