HHS Public Access
Author manuscript
Ann Epidemiol. Author manuscript; available in PMC 2018 January 01.

Published in final edited form as:
Ann Epidemiol. 2017 January ; 27(1): 20–26. doi:10.1016/j.annepidem.2016.07.016.

The Utility of EMR Address Histories for Assessing Neighborhood Exposures

Amy E. Hughes[a,*] and **Sandi L. Pruitt**[a,b]

[a]Department of Clinical Sciences, University of Texas Southwestern Medical Center, Dallas, TX, USA

[b]Harold C. Simmons Comprehensive Cancer Center, Dallas, TX, USA

Abstract

Purpose—Electronic medical records (EMR) include residential address histories, which may alleviate exposure misclassification caused by exclusion of patient spatiotemporal location. EMR data are increasingly available but rarely leveraged as a measure of cumulative environmental exposure, in part due to limited understanding of the validity of EMR-derived address histories.

Methods—We compared EMR address histories to self-reported histories among 100 patients of a safety-net healthcare system completing a telephone survey. We assessed agreement and compared 7 neighborhood-level environmental exposures as assessed using both data sources.

Results—While 17.1% of respondents did not live at the most recent EMR-derived address during the survey, nearly all (98%) lived there at some point. For respondents with >1 EMR-derived address (N=64), 87.5% had once lived at the previous EMR address. Of these, 30.4% lived at ≥1 additional residences between the two most recent EMR address. For all measures, neighborhood-level environmental exposures did not differ when using EMR-derived vs. self-report addresses.

Conclusions—More recent EMR-derived addresses are more accurate and differences compared to self-reported addresses in neighborhood-level exposures are negligible. EMR-derived address histories are incomplete and likely suffer from collection bias; future research should further assess their validity and reliability.

Keywords

Residential mobility; electronic medical record; neighborhoods; geographic information systems (GIS)

*Corresponding author at: University of Texas Southwestern Medical Center, Department of Clinical Sciences, 5323 Harry Hines Blvd, E1.410D, Dallas, TX 75390- 9169, USA; Tel.: +1 214 648 3207; fax: +1 214 648 3232; Amy.Speer@UTSouthwestern.edu.

Introduction

Placed-based exposures impact a variety of health outcomes, and the effects of these exposures are often investigated through residential location. Residential mobility creates temporal variation in residential location, and decreases the individual-level precision of place-based exposure measurements. One way to address residential mobility is to explore static, areal effects of place, and this method is facilitated by the availability of US Census Bureau data. However, such measures lack information regarding cumulative environmental exposures at the individual level. Thus, reliance on these measures as proxy for cumulative environmental exposure can result in exposure misclassification [1–3].

Exposure misclassification can introduce significant bias into analyses and has direct implications for studies of neighborhood effects. Specifically, life-course neighborhood exposures—whether social, structural, or environmental—cannot be fully assessed through a single measurement in space and time, thus failure to include longitudinal information can generate exposure misclassification. Furthermore, because diverse, low-income, and older populations are both mobile [4,5], and at-risk for poor health [6–8], exposure misclassification introduced through cross-sectional measures may disproportionately affect research focused on these populations.

Epidemiologic literature supports inclusion of spatiotemporal exposure measures in place of cross-sectional measures; recent work regarding cumulative environmental exposure to over the life course [9–12] and across the activity space [13–16] incorporate time and place to derive measures of spatiotemporal exposures. Residential mobility data offer spatiotemporal records patient movement through space over time and are frequently studied in sociology [5,17–19], demography [20–24], economics [25–28], and public health [8,29,30]. A number of recent epidemiologic studies include information gleaned via address histories [31–34]. Despite current U.S. law mandating adoption of electronic medical records (EMRs), many of which contain address histories [35,36], no epidemiologic studies to date utilize this increasingly available resource. This is due, in part, to limited understanding regarding the validity of EMR-derived address histories.

Healthcare system EMR data collection procedures may lead to misclassification or other biases [2]. For example, if address collection is driven by patient attendance, then the healthcare system simultaneously fails to collect full address histories for patients in good health and excels at collecting full address histories for patients in poor health. Moreover, because EMRs are system-specific (collecting data over time as patients return to facilities within the healthcare system), address histories may be particularly incomplete for patients without a usual source of care, or who are not enrolled in integrated or health maintenance organization systems, not engaged in primary care, or under- or uninsured. Thus, understanding the validity and completeness of EMR-derived address histories is critical for at-risk populations with higher levels of residential mobility. Herein, we describe our methods to clean, process, and evaluate EMR address histories from an integrated, safety-net healthcare system. We compare EMR-derived to patient-reported address histories to better understand discrepancies between these two sources of address history data.

Methods

Study Population

Our overall cohort (N=38,410) was drawn from the Parkland-UT Southwestern Population-based Research Optimizing Screening through Personal Regimens (PROSPR) colorectal cancer (CRC) screening cohort. Cohort eligibility criteria included: male and female patients aged 50 to 64; who were Dallas County, Texas, residents; had no colectomy procedures prior to January 2012 and no prior history of CRC; and were seen at least once at Parkland Health and Hospital System (hereafter “Parkland”) primary care clinic between January 2010 and September 2012 [37]. Dallas County has few safety-net clinics for the under- and uninsured; Parkland, as the cornerstone of the county safety-net health care system, is effectively is the sole source of care for these patients. Address histories were drawn from Parkland's system-wide integrated EPIC EMR (Epic Systems Inc., Verona, WI). In addition to PROSPR cohort eligibility criteria, patients eligible for the telephone survey had at least one valid and geocoded EMR address and a valid phone number.

EMR Address Data Collection and Structure

The EMR structure resembles a web of data stored in tables. Variables of interest are stored in tables linked to address data via record identifiers, and data analysts must navigate varying connection types and collection frequencies to create rich patient-level data. For Parkland’s EMR, address data collection occurs during each visit. Patients can have multiple visits in one day, so an address can be recorded multiple times in one day. Spelling of address records is not standardized across patients. Figure 1 presents a simplified patient-level example of the EMR, with example address history.

Address Coding Procedures

Address history tables included address data in five fields: primary street, secondary street, city, state, and ZIP code. Using regular expressions, we ensured validity of each address entry by excluding entries where neither street address field contained alphanumeric data. We excluded P.O. Box addresses because they do not represent residential location. Valid residential addresses were standardized in ArcMap 10.3 [38], which increased geocoding match rates by parsing each address history entry into 9 standard address components.

Cadastral geocoding (geocoding to parcels, the legal boundaries for a plot of land) can match point locations to 68–72% of addresses [39], and limits geocoding matches only to county-confirmed addresses. To maximize the match rate, we geocoded address data to county tax parcels using a modified cadastral geocoding scheme comprising three steps: (1) attempted cadastral match, (2) attempted match within patient’s own address history, and (3) attempted text match with other patient’s geocoded data.

In the first step, addresses were geocoded to parcels using a zone-enabled address locator in ArcMap 10.3 [38] built with tax assessor data from Dallas County and the surrounding counties. Our zone-enabled address locator defined smaller "zones" of candidate match addresses by zip code, which permits differentiation between the same street address in different cities. In the second step, string distance metrics compared geocoded historical addresses with ungeocoded addresses for each patient. This step mitigates data-entry errors wherein clerks spell addresses differently from one visit to the next. In the third step, unmatched address history entries were compared to matched addresses across all patients, and data from geocoded addresses were transferred to ungeocoded addressees in the case of an exact text match.

In the case of >1 address per patient, we differentiated whether each address was unique or a duplicate of another address in the patient's history. We defined a residential change of address as a change in residential location greater than 20 feet between the patients' most recent and previous (next most recent) addresses.

Survey Procedures

We employed a stratified sampling procedure to ensure representation of patients with low/high residential mobility and low/high healthcare system visit frequency. High mobility was defined as more than 2 unique addresses; high visit frequency was defined as more than 6 healthcare system visits. We mailed introductory letters, written in both English and Spanish, to the most recent address for an equal number of patients (N=184) per strata. Following IRB policy, the letter instructed patients to call within one week to opt out of the survey. We called those who did not opt out. We attempted contact via phone up to 6 times and asked participants to complete the survey after verbal consent.

Survey questions pertained to self-reported health status, length of stay and address type for the participant's most recent address, length of stay and address type for the previous address, the number of unrecorded residential moves between addresses recorded in the EMR, and reasons for moving into and out of residences. To ascertain *accuracy of EMR-derived current residence*, survey staff read the most recent EMR address to the participant and, for that address, asked if the participant currently lived there, ever lived there, and received mail there. To ascertain *accuracy of EMR-derived past residence*, survey staff read the previous address to participants with more than one EMR address, and asked if the participant ever lived there. To determine *if EMR-derived address history was missing any prior residences*, survey staff asked participants with more than one EMR address if they lived anywhere else (and if so, how many different homes they had resided in) between the most recent and next most recent addresses.

Interviews (18 questions, approximately 20 minutes) were conducted by telephone in either English or Spanish according to participant preference. Participants received a $10 gift card via mail for completing the survey. Sample selection and analysis were performed in SAS 9.4 [40]. Surveys were conducted and documented using REDCap [41].

Analysis Procedures

We assessed the number of moves in each dataset using descriptive statistics: (1) correlation between number of moves recorded and number of self-reported moves, (2) percent agreement for recorded and self-reported moves out-of-county, and (3) the percent agreement for recorded and self-reported moves out-of-state.

Next, we examined measures of change over time for participants who moved. We augmented all incomplete EMR-derived address histories with additional addresses gleaned via self-reported histories. To assess potential for exposure misclassification within EMR address histories, we conducted difference of means tests between EMR-only data and EMR- plus self-reported data for neighborhood exposure change measures. Specifically, we implemented a series of paired t-tests. Our survey asked about missing addresses between the most recent and previous EMR addresses, so we conduct these difference of means tests for the most recent move only. We linked 2009–2013 American Community Survey (ACS) data to the EMR-only and EMR-plus addresses using ArcMap 10.3 [38]. We compared average relocation distance for the most recent move using EMR-only versus EMR-plus address histories. We also compared seven block-group level measures using EMR-only versus EMR-plus address histories: percentage of (1) black residents, (2) white residents, (3) Hispanic residents, (4) residents earning below 100% of the federal poverty level (FPL), (5) residents earning below 200% of the FPL, (6) residents earning a high school degree, and (7) residents earning a college degree.

Results

A total of 32,165 patients from the overall cohort were eligible for the survey. On average, patients had 7.4 EMR-derived addresses (SD=6.1; Range: 1–70), of which 2.1 were unique (SD=1.3; Range 1–16). For the 736 patients randomized to the survey sampling frame, staff placed more than 1,000 phone calls until a total of 101 completed the survey. Table 1 presents the survey stratification scheme with respondent frequency by recruitment strata, and Figure 2 describes the overall study enrollment. Survey respondents had on average 5.8 EMR-derived addresses (SD=4.2; Range: 1–21), of which 2.1 were unique (SD=1.1; Range: 1–6). The average time between unique addresses was 2.2 years (SD=1.7; Range=0–5.6) Table 2 presents demographics for the overall cohort (N=32,165), the sampling frame (N=736), and the survey respondents (N=101).

Accuracy of EMR-derived current and past residence

The most recent EMR address is indicative of self-reported residential location for the participant. Of participants who answered the series of questions (N=99), 100% confirmed they had lived at the most recent EMR address at some point; however, 17.1% indicated that they did not currently live at this address. Previous EMR address was also indicative of true residence. Of those queried (N=64; those with <2 addresses were not asked this question), 87.5% confirmed previously living at the previous EMR address.

Missing residences and percent agreement

In our survey sample, 45 participants were not asked missing-residence questions for two reasons: they only had one address (N=37), or did not ever reside at the previous address (N=8). Of the remaining 56 participants, 30.4% reported residence at unrecorded addresses; specifically, 16.1% resided at one unrecorded address, 5.3% at two unrecorded addresses, 3.5% at three unrecorded addresses, 3.6% resided at 5 or 6 unrecorded addresses, and 1.8% reported residing at unrecorded addresses but not knowing the exact number of unrecorded addresses. Percent agreement between the number of EMR-derived and self-reported residences was 37.6% (N=94) over the 5-year period between 2006 and 2012. Percent agreement for out-of-county and out-of-state residences was 91.7% (N=96) and 100% (N=96), respectively.

Comparing EMR mobility to self-report mobility

Table 3 presents means and standard deviations for neighborhood-level exposure measures, and for changes in these measures due to the most recent residential move. The table displays results using EMR-only data versus EMR-plus self-report data, and pertains only to the most recent residential move for participants with more than one residential address (N=67). P-values reflect the difference of means paired t-tests comparing changes in neighborhood-level exposure levels between these two sources of data. Changes in environmental exposures are similar for the unadjusted and adjusted samples. On average, survey participants who reported discrepancies in their most recent or previous EMR address did not move to significantly more or less poor, Hispanic, Black, White, high-school educated, or college-educated block groups compared to their correctly-recorded counterparts. In addition, participants with address discrepancies did not move closer or further away than their correctly-recorded counterparts.

Discussion

Our study presents novel information regarding the utility of EMR address histories for determining longitudinal environmental and neighborhood exposure, and fits with other studies that employed different methods to recover longitudinal environmental exposure information [2,42]. Within a large, mature, integrated EMR, we explored the effects of collection bias for location-based measures of environmental exposure using a survey designed to detect differences in participant's EMR-derived versus true (self-report) recent address history. While our survey sample is small, we hope to open a dialogue regarding the value of including address histories recorded in EMR systems in future geospatial health research. Our findings have a number of implications for researchers seeking to leverage EMR-derived address histories to infer place-based environmental exposures.

First, we find that more recent entries in EMR-derived address histories are more accurate. We arrive at this finding through a subset of our results: (1) the most recent address more often reflects self-reported residential addresses, (2) most respondents indicate that they did not live at another address between the most recent and previous address, but (3) percent agreement between the number of EMR-derived and self-reported addresses is low. Greater accuracy in more recent address entries may reflect recall bias wherein participants do not remember prior residences [43], respondent bias wherein participants are do not truthfully answer questions for fear of losing health services, or bias in the opportunistic address data collection process.

Second, we find that inclusion of addresses recovered through the survey does not substantially change average measures of environmental exposure. The differences between measures of environmental exposure from the two address histories sources represent a lower bound for environmental exposure misclassification; we know that true environmental exposure (measured through the survey) differs from recorded exposure (measured by the

EMR) by at least this much. There may be greater exposure misclassification among patients with varying levels of mobility or health-care system engagement, but the small cell sizes across our study strata limit such analysis. Furthermore, there may be greater exposure misclassification over a longer time period or greater number of moves but these data were not available in the current study.

Third, there is an important distinction between self-reported and EMR-derived address histories. The low percent agreement for number of EMR-derived and self-reported addresses supports this finding. Self-reported address histories capture all addresses for an individual, barring recall bias. EMR systems can only capture addresses for those who engage in the system itself, and patients engaging in healthcare differ from those who do not. For example, they are more likely to have health insurance, a usual source of care, and chronic conditions [44,45]. EMR-derived address histories are collected using opportunistic sampling that is dependent upon visit volume, and may be most accurate for those with more system visits and fewer moves. Because EMR systems likely fail to capture address histories equally for patients across all levels of mobility and health system engagement, analyses employing EMR-derived address histories may suffer from several flaws. For example, such studies face endogeneity problems (i.e., wherein documentation of residential mobility may be causally associated with healthcare utilization or vice versa), which create downward bias for estimated effect sizes.

Our study faces several limitations. We analyzed EMR data from Parkland, a safety-net provider serving a diverse, low-income, urban, and under- and uninsured population. EMR data in one healthcare system may differ from those in other systems in unknown ways. Moreover, EMR data may differ from addresses obtained from other clinical or administrative data sources (e.g., insurance claims). Homogeneity among our population (e.g., all are engaged in healthcare) may account for the small differences in environmental exposure that we observed. Respondents may be less mobile than other populations; this be a result of inclusion of patients 50 to 64 years old, who are less likely to have recently experienced the common drivers of mobility such as expanding families [17,25–27] or the need for age-dependent care.[46–48]. Misclassification bias may be greater in populations with higher rates of mobility.

Study practices limited the sampling frame to patients in the overall cohort who had a valid, geocoded EMR address for introductory letters and a valid phone number for survey calls. Both addresses and land lines may change with residential relocation, which may limit our ability to detect mobility-driven environmental exposure misclassification. Selection bias is a potential concern given eligibility criteria and survey response rates; however, demographics of the respondents, sampling frame, and overall cohort are similar which partially alleviates this concern. We did not ask respondents with one EMR address if the EMR missed a previous address. Finally, interpretation of our study is limited by its size (N=101).

Conclusions

EMR-derived address histories provide an opportunity for researchers to recover longitudinal environmental exposure information. Our findings indicate that EMR-derived address histories are most accurate for recent addresses and that collection bias exists within EMR-derived address histories. Larger studies are needed to fully determine the extent of collection bias in EMR-derived address histories and to further explore validity and reliability of these data.

Acknowledgments

Funding: This work was supported by the Cancer Prevention Research Institute of Texas (CPRIT R1208); the National Cancer Institute–funded Population-based Research Optimizing Screening through Personalized Regimens (PROSPR) Parkland-UT Southwestern Center (U54CA163308); the National Cancer Institute (5P30CA142543) for the Harold C. Simmons Cancer Center; and the National Center for Advancing Translational Sciences (UL1TR001105) for the UT Southwestern Center for Translational Medicine. Dr. Hughes was funded through a postdoctoral fellowship at the University of Texas School of Public Health Cancer Education and Career Development Program, National Cancer Institute/NIH Grant R25 CA57712.

List of Abbreviations

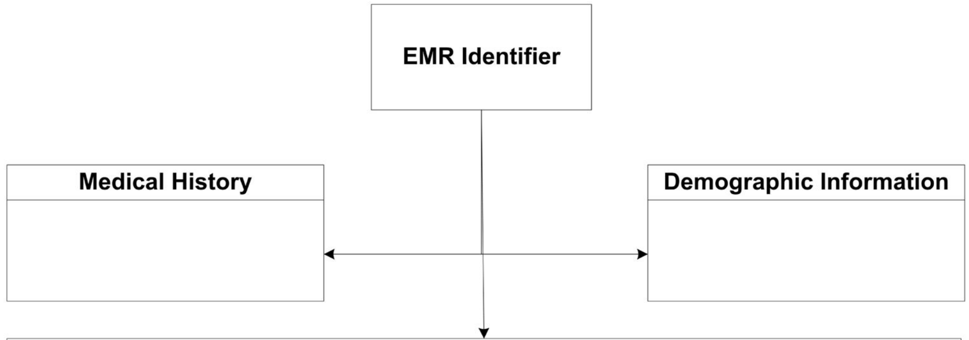
EMR	electronic medical records
GIS	geographic information systems
PROSPR	Population-based Research Optimizing Screening through Personal Regimens
CRC	colorectal cancer
FPL	federal poverty level
SD	standard deviation
N	sample size
HS	high school

References

1. Hodgson S, Lurz PWW, Shirley MDF, Bythell M, Rankin J. Exposure misclassification due to residential mobility during pregnancy. Int J Hyg Environ Health. 2015; 218:414–21. DOI: 10.1016/j.ijheh.2015.03.007 [PubMed: 25845985]
2. Bryere J, Pornet C, Dejardin O, Launay L, Guittet L, Launoy G. Correction of misclassification bias induced by the residential mobility in studies examining the link between socioeconomic environment and cancer incidence. Cancer Epidemiol. 2015; 39:256–64. DOI: 10.1016/j.canep.2014.12.008 [PubMed: 25579981]
3. Danysh HE, Mitchell LE, Zhang K, Scheurer ME, Lupo PJ. Differences in environmental exposure assignment due to residential mobility among children with a central nervous system tumor: Texas, 1995–2009. J Expo Sci Environ Epidemiol. 2015; doi: 10.1038/jes.2015.63
4. Clark, WAV. SAGE Handb Hous Stud. London; Thousand Oaks, CA: Sage Publications; 2012. Residential Mobility and the Housing Market; p. 66-83.
5. Rossi PH, Shlay AB. Residential Mobility and Public Policy Issues: “Why Families Move” Revisited. J Soc Issues. 1982; 38:21–34. DOI: 10.1111/j.1540-4560.1982.tb01768.x
6. Adler NE, Ostrove JM. Socioeconomic Status and Health: What We Know and What We Don’t. Ann N Y Acad Sci. 1999; 896:3–15. doi:10.1111/j.1749- 6632.1999.tb08101.x. [PubMed: 10681884]
7. Adler NE, Boyce T, Chesney MA, Cohen S, Folkman S, Kahn RL, et al. Socioeconomic status and health: the challenge of the gradient. Am Psychol. 1994; 49:15. [PubMed: 8122813]

8. Jelleyman T, Spencer N. Residential mobility in childhood and health outcomes: a systematic review. J Epidemiol Community Health. 2008; 62:584–92. DOI: 10.1136/jech.2007.060103 [PubMed: 18559440]
9. Clarke P, Morenoff J, Debbink M, Golberstein E, Elliott MR, Lantz PM. Cumulative Exposure to Neighborhood Context: Consequences for Health Transitions over the Adult Life Course. Res Aging. 2013; 36:115–42. DOI: 10.1177/0164027512470702 [PubMed: 24465068]
10. Kramer MR, Dunlop AL, Hogue CJ. Measuring women's cumulative neighborhood deprivation exposure using longitudinally linked vital records: a method for like course MCH research. Matern Child Health J. 2014; 18:478. [PubMed: 23430295]
11. Mishra G, Nitsch D, Black S, De Stavola B, Kuh D, Hardy R. A structured approach to modelling the effects of binary exposure variables over the life course. Int J Epidemiol. 2009; 38:528–37. DOI: 10.1093/ije/dyn229 [PubMed: 19028777]
12. Sinha G, Mark DM. Measuring similarity between geospatial lifelines in studies of environmental health. J Geogr Syst. 2005; 7:115–36. DOI: 10.1007/s10109-005-0153-8
13. Perchoux C, Chaix B, Cummins S, Kestens Y. Conceptualization and measurement of environmental exposure in epidemiology: Accounting for activity space related to daily mobility. Health Place. 2013; 21:86–93. DOI: 10.1016/j.healthplace.2013.01.005 [PubMed: 23454664]
14. Kestens Y, Lebel A, Daniel M, Theriault M, Pampalon R. Using experienced activity spaces to measure foodscape exposure. Health Place. 2010; 16:1094–1103. [PubMed: 20667762]
15. Inagami S, Cohen D, Finch BK. Non-residential neighborhood exposures suppress neighborhood effects on self-rated health. Soc Sci Med. 2007; 65:1779–91. DOI: 10.1016/j.socscimed.2007.05.051 [PubMed: 17614175]
16. Vallee J, Cadot E, Grillo F, Parizot I, Chauvin P. The combined effects of activity space and neighbourhood of residence on participation in preventive health-care activities: The case of cervical screening in the Paris metropolitan area (France). Health Place. 2010; 16:838–852. [PubMed: 20451439]
17. Rossi, PH. Why families move;: A study in the social psychology of urban residential mobility. 1. Free Press; 1955.
18. Glick PC. The Life Cycle of the Family. Marriage Fam Living. 1955; 17:3–9. DOI: 10.2307/346771
19. Glick PC. The Family Life Cycle and Social Change. Fam Relat. 1989; 38:123–9. DOI: 10.2307/583663
20. Morrill RL. Development, Diversity, and Regional Demographic Variability in the U. S Ann Assoc Am Geogr. 1993; 83:406–33. DOI: 10.1111/j.1467-8306.1993.tb01943.x
21. Morrill RL. Regional demographic structure of the united states. Prof Geogr. 1990; 42:38–53. DOI: 10.1111/j.0033-0124.1990.00038.x
22. Quillian L. Migration Patterns and the Growth of High-Poverty Neighborhoods, 1970–1990. Am J Sociol. 1999; 105:1–37. DOI: 10.1086/210266
23. Frey WH. Central City White Flight: Racial and Nonracial Causes. Am Sociol Rev. 1979; 44:425–48. DOI: 10.2307/2094885
24. Frey WH. White flight and central-city loss: application of an analytic migration framework. Environ Plan A. 1979; 11:129– 147. DOI: 10.1068/a110129 [PubMed: 12262800]
25. Mulder, CH. Migration dynamics: a life course approach. The University of Michigan: Thesis Publishers; 1993.
26. Mulder CH, Wagner M. Migration and marriage in the life course: a method for studying synchronized events. Eur J Popul. 1993; 9:55–76. DOI: 10.1007/BF01267901 [PubMed: 12344903]
27. Clark, WAV., Dieleman, FM. Households and housing: Choice and outcomes in the housing market. Transaction Publishers; 1996.
28. Abramsson M, Andersson EK. Residential Mobility Patterns of Elderly—Leaving the House for an Apartment. Hous Stud. 2012; 27:582–604. DOI: 10.1080/02673037.2012.697553
29. Jacquez GM, Shi C, Meliker JR. Local bladder cancer clusters in southeastern Michigan accounting for risk factors, covariates and residential mobility. PloS One. 2015; 10:e0124516.doi: 10.1371/journal.pone.0124516 [PubMed: 25856581]

30. Jacquez GM, Barlow J, Rommel R, Kaufmann A, Rienti M, AvRuskin G, et al. Residential Mobility and Breast Cancer in Marin County, California, USA. Int J Environ Res Public Health. 2014; 11:271–95. DOI: 10.3390/ijerph110100271
31. Manjourides J, Pagano M. Improving the power of chronic disease surveillance by incorporating residential history. Stat Med. 2011; 30:2222–33. DOI: 10.1002/sim.4272 [PubMed: 21563208]
32. Cook AJ, Gold DR, Li Y. Spatial Cluster Detection for Repeatedly Measured Outcomes while Accounting for Residential History. Biom J. 2009; 51:801–18. DOI: 10.1002/bimj.200800269 [PubMed: 19760635]
33. Hoffman K, Webster TF, Weinberg JM, Aschengrau A, Janulewicz PA, White RF, et al. Spatial analysis of learning and developmental disorders in upper Cape Cod, Massachusetts using generalized additive models. Int J Health Geogr. 2010; 9:7.doi: 10.1186/1476-072X-9-7 [PubMed: 20152039]
34. Han D, Rogerson PA, Bonner MR, Nie J, Vena JE, Muti P, et al. Assessing spatio-temporal variability of risk surfaces using residential history data in a case control study of breast cancer. Int J Health Geogr. 2005; 4:9.doi: 10.1186/1476-072X-4-9 [PubMed: 15826315]
35. U.S. 111th Congress. Excerpts from the American Recovery and Reinvestment Act of 2009. 2009; 11
36. U.S. Department of Health and Human Services O of the NC for HIT. Health IT Policymaking and Health IT Rules and Regulations. 2013.
37. Tiro JA, Kamineni A, Levin TR, Zheng Y, Schottinger JS, Rutter CM, et al. The CRC screening process in community settings: A conceptual model for the Population-based Research Optimizing Screening through Personalized Regimens Consortium. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2014; 23:1147–58. DOI: 10.1158/1055-9965.EPI-13-1217
38. ESRI. ArcGIS Desktop: Release 10.3. Redlands, CA: Environmental Systems Research Institute; 2014.
39. Dearwent SM, Jacobs RR, Halbert JB. Locational uncertainty in georeferencing public health datasets. J Expo Anal Environ Epidemiol. 2001; 11:329. [PubMed: 11571612]
40. SAS Institute, Inc. SAS System for Windows. Cary, NC, USA: SAS Institute Inc; 2013.
41. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009; 42:377–81. DOI: 10.1016/j.jbi.2008.08.010 [PubMed: 18929686]
42. Jacquez GM, Slotnick MJ, Meliker JR, AvRuskin G, Copeland G, Nriagu J. Accuracy of Commercially Available Residential Histories for Epidemiologic Studies. Am J Epidemiol. 2011; 173:236–43. DOI: 10.1093/aje/kwq350 [PubMed: 21084554]
43. Coughlin SS. Recall bias in epidemiologic studies. J Clin Epidemiol. 1990; 43:87–91. DOI: 10.1016/0895-4356(90)90060-3 [PubMed: 2319285]
44. DeVoe JE, Fryer GE, Phillips R, Green L. Receipt of preventive care among adults: insurance status and usual source of care. Am J Public Health. 2003; 93:786–791. [PubMed: 12721145]
45. Babitsch B, Gohl D, von Lengerke T. Re-revisiting Andersen's Behavioral Model of Health Services Use: a systematic review of studies from 1998–2011. GMS Psycho-Soc-Med. 2012:9.
46. Wiseman RF. Why Older People Move Theoretical Issues. Res Aging. 1980; 2:141–54. DOI: 10.1177/016402758022003
47. Meyer JW, Speare A. Distinctively Elderly Mobility: Types and Determinants. Econ Geogr. 1985; 61:79.doi: 10.2307/143676 [PubMed: 12280179]
48. Meyer JW, Cromley EK. Caregiving Environments and Elderly Residential Mobility. Prof Geogr. 1989; 41:440–50.



Address History

Visit	Street1	Street2	City	ZIP	Start	End
1	123 Main St.		Dreamland	92802	1/8/2005	3/21/2005
2	15123 Main St		Dreamland	92802	3/21/2005	8/19/2005
3		123 Mayne St	Dreamland	92802	8/19/2005	8/19/2005
4	Mayne St		Dreamland	92802	8/19/2005	12/29/2008
5	PO Box 728		DL	92802-7280	12/29/2008	1/31/2009
6	11 Parade Ave	Apt C	Roses	91103	1/31/2009	6/14/2010
7	11	Apt C	Roses	91103	6/14/2010	9/3/2011
8	11 Parde Ave	Apt C	Roses	91103	9/3/2011	

Figure 1.

Example address history for a single patient in the Epic EMR

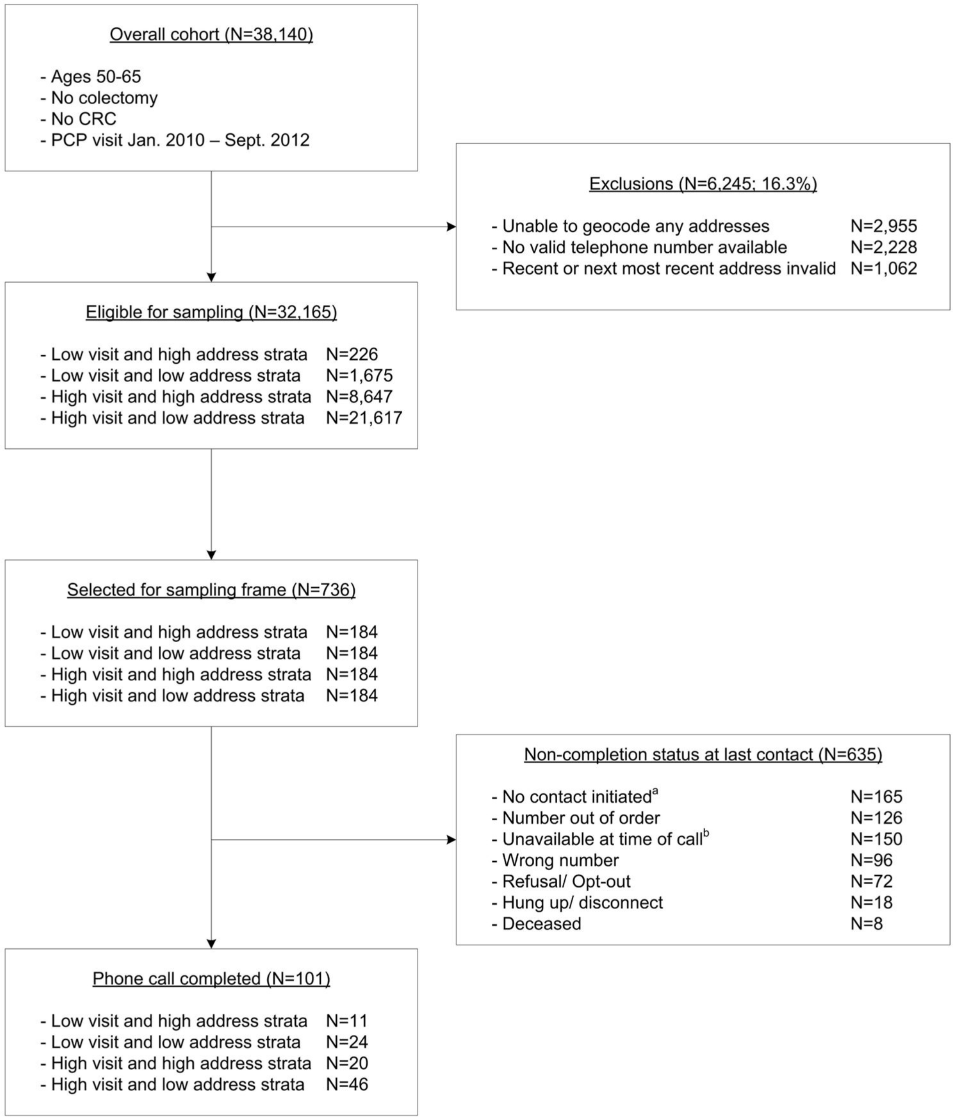


Figure 2.

Survey participant eligibility criteria and enrollment

Note: [a]In some cases (N=165), contact was not initiated because the target study enrollment was reached prior to placement of the first call to these participants. [b]Participants who were unavailable at time of last contact fell into 5 categories: answering machine picked up the call (N=99), call was answered but participant was not home (N=23), call was never answered (N=16), telephone line was busy (N=9), or call was answered by the participant who requested the survey be administered at a later date/time (N=3).

Table 1

Respondent frequency across survey recruitment strata of n=101 patients completing self-reported address history survey

Number of Unique EMR addresses	Number of healthcare visits		Total
	1–6	**>6**	
1–2 addresses	24	46	70
>2	11	20	31
Total	35	66	101

Table 2

Demographic data for respondents and participants

	Eligible Overall Cohort N=32,165		Sampling Frame N=736		Survey Respondents N=101	
	N (mean)	**% (SE)**	**N (mean)**	**% (SE)**	**N (mean)**	**% (SE)**
Age	(60.5)	(4.5)	(59.7)	(4.3)	(59.6)	(4.1)
Sex						
Male	11,971	37.2%	257	40.5%	30	29.7%
Female	20,194	62.8%	378	59.5%	70	69.3%
Race/Ethnicity						
Hispanic	12,010	37.3%	263	41.4%	41	40.6%
Black	12,358	38.4%	210	33.1%	44	43.6%
White	17,568	54.6%	385	60.6%	55	54.5%
Other	2,016	6.3%	39	6.1%	2	2.0%

Note: Race and/or ethnicity data are unavailable for patients.

Table 3

Changes in neighborhood environmental exposures associated with participants' most recent residential move, only for those with at least one change in address (n=67). Data reflect exposures calculated using addresses obtained from EMR data only (un-adjusted) and EMR data plus self-report data (adjusted).

	Unadjusted (EMR data only)						Adjusted (EMR data plus self-report data)						P-value (paired t-test)[a]
	Most recent address		Next most recent address		Change from next most recent to recent		Most recent address		Next most recent address		Change from next most recent to recent		
Measure	**Mean**	**(SD)**	**Mean**	**(SD)**	**Mean**	**(SD)**	**Mean**	**(SD)**	**Mean**	**(SD)**	**Mean**	**(SD)**	
Individual-level													
Move distance (mi)	--	--	--	--	4.452	(4.219)	--	--	--	--	5.069	(5.287)	0.113
Neighborhood-level measures -- % of block group residents who:													
Earn 100% of FPL	24.9%	(15.8%)	24.5%	(13.8%)	0.4%	(17.6%)	24.1%	(17.0%)	25.4%	(14.9%)	−1.3%	(18.5%)	0.569
Earn 200% of FPL	54.1%	(21.9%)	54.2%	(22.2%)	−0.1%	(25.5%)	51.0%	(24.1%)	55.5%	(23.4%)	−4.5%	(25.0%)	0.147
Are Hispanic	41.3%	(25.5%)	37.3%	(29.1%)	4.0%	(28.0%)	40.3%	(27.4%)	38.4%	(29.0%)	1.9%	(29.0%)	0.587
Are White	44.0%	(25.1%)	42.3%	(26.7%)	1.7%	(26.4%)	42.8%	(26.1%)	43.0%	(25.5%)	−0.2%	(28.6%)	0.959
Are Black	39.4%	(30.2%)	40.4%	(34.1%)	−1.0%	(25.9%)	35.3%	(29.9%)	38.3%	(32.2%)	−3.1%	(27.1%)	0.355
Graduated High School	68.2%	(17.6%)	71.3%	(18.1%)	−3.0%	(22.8%)	64.5%	(23.3%)	68.9%	(18.6%)	−4.4%	(29.2%)	0.222
Graduated College	13.6%	(12.2%)	16.4%	(15.4%)	−2.8%	(17.0%)	19.4%	(23.3%)	16.3%	(16.4%)	3.1%	(25.7%)	0.333

Note: EMR=electronic medical record; FPL=federal poverty level; SD=standard deviation.

[a]P-values refer to the comparison between change scores using EMR data only vs. EMR data plus self-report data.