Residential mobility impacts exposure assessment and community socioeconomic characteristics in longitudinal epidemiology studies

Cole Brokamp¹, Grace K. LeMasters¹,² and Patrick H. Ryan¹,³

Epidemiologic studies commonly use residential locations to estimate environmental exposures or community-level characteristics. The impact of residential mobility on these characteristics, however, is rarely considered. The objective of this analysis was to examine the effect of residential mobility on estimates of traffic-related air pollution (TRAP), greenspace, and community-level characteristics. All residential addresses were reported from birth through age seven for children enrolled in the Cincinnati Childhood Allergy and Air Pollution Study. Exposure to TRAP at each address was estimated using a land use model. Greenspace was estimated using satellite imagery. Indices of neighborhood deprivation and race were created based on socioeconomic-census tract measures. Exposure estimates using the birth record address, the last known address, and the annual address history were used to determine exposure estimation error and bias in the association with asthma at age seven. Overall, 54% of the cohort moved at least once prior to age seven. Each move was separated by a median of 4 miles and associated with a median decrease of 4.4% in TRAP exposure, a 5.3% increase in greenspace, an improved deprivation index, and no change in the race index. Using the birth record address or the last known address instead of the annual address history resulted in exposure misclassification leading to a bias toward the null when associating the exposures with asthma. Using a single address to estimate environmental exposures and community-level characteristics over a time period may result in differential assessment error.

INTRODUCTION
Residential addresses are commonly used in epidemiologic studies to estimate exposure to environmental pollutants or chemicals that vary geographically including polycyclic aromatic hydrocarbons,¹ nitrogen oxides,² volatile organic compounds,³ greenspace,⁴ both the total mass and elemental components of particulate matter,⁵,⁶ as well as community socioeconomic characteristics.⁷ Accurate exposure assessment is especially critical for longitudinal environmental health studies where subjects can change their residence (residential mobility) during the course of the study. Although it is clear that other influences such as indoor sources and personal activities affect exposure,⁸,⁹ exposure misclassification may also occur when using address per subject. Nevertheless, studies often use only either the enrollment address or the last known address in place of a full address history.¹⁰,¹¹

There have been some studies that found stronger associations between health outcomes and environmental pollutants after accounting for residential mobility.¹²,¹³ However, other studies have found the difference between inclusion address and the full address history for exposure assessment and association with health outcomes to be small and non-differential.¹⁴–¹⁶ Many of these studies have been conducted in rural areas with exposure models having limited spatial resolution. However, though residential changes, particularly in urban areas, may result in a relatively small change in distance, they could result in large changes in exposure levels depending on distance from the original source, terrain, and proximity to other point or mobile sources.¹⁷ These changes may only become apparent when using a high resolution exposure assessment model. Therefore, it is important to consider residential mobility in an urban setting where exposures and community characteristics are highly variable at a local scale.

Accurately characterizing early life environmental exposures is particularly important as this represents a period of enhanced susceptibility during child development. Studies using the birth record address may assume a constant exposure throughout the follow-up period, which may be violated by residential mobility. Moreover, increased residential changes in families with children have been associated with increased poverty,¹⁸,¹⁹ unemployment,²⁰ single parenting,²¹ and various health outcomes including increased illegal drug use, increased levels of teenage pregnancy, and higher levels of behavioral and emotional problems.²² The relationship of residential mobility with important epidemiological covariates suggests the exposure misclassification may be higher among certain groups and this could bias epidemiologic studies in a differential manner. The timing of residential mobility also plays an important role in exposure misclassification, with earlier moves likely resulting in greater misclassification.

¹Department of Environmental Health, University of Cincinnati, Cincinnati, Ohio, USA; ²Division of Asthma Research, Cincinnati Children’s Hospital Medical Center, Cincinnati, Ohio, USA and ³Division of Biostatistics and Epidemiology, Cincinnati Children’s Hospital Medical Center, Cincinnati, Ohio, USA. Correspondence: Dr. Patrick H. Ryan, Cincinnati Children’s Hospital Medical Center, 3333 Burnet Avenue, ML 5041, Cincinnati, OH 45229, USA.
Tel.: +5138034704. Fax: +5136367509.
E-mail: patrick.ryan@cchmc.org
Received 20 October 2015; accepted 29 December 2015; published online 9 March 2016
The aim of our study was to characterize the residential mobility patterns of a cohort of children from birth to age seven and to characterize how these moves impacted common epidemiologic study metrics including traffic related air pollution (TRAP), greenspace, and socioeconomic status. These specific metrics were chosen also because of their well validated exposure assessment methods that have a high spatial resolution. This is necessary for exploring the changes in an urban area, where moves often cover a short distance and pollutant concentration gradients vary highly at the local scale. We tested the hypothesis that residential mobility during early childhood will lead to significant misclassification bias for both the estimated exposure levels and the association with asthma diagnosis at age seven. This hypothesis was tested by comparing the use of birth record and age seven (or last known) residential address to assess exposure levels in comparison with utilizing the full yearly residential history. In addition, we examined how the exposure estimates and community characteristics influenced the timing and frequency of residential moves.

METHODS

Study Cohort

The Cincinnati Childhood Allergy and Air Pollution Study (CCAAPS) is an ongoing prospective birth cohort of high-risk atopic children.23,24 Children born between October 2001 and July 2003 in the Greater Cincinnati and Northern Kentucky region were screened by birth record and enrolled if they lived less than 400 m or more than 1,500 m from the nearest major road.24 Furthermore, each study participant must have had at least one parent with symptoms of rhinitis, asthma, or eczema and allergic sensitization by a positive skin prick test result to one of 17 aeroallergens. Informed consent was obtained and the study was approved by the University of Cincinnati Institutional Review Board. Residential address history from birth to age seven for each participant was obtained from birth records and during study visits conducted at enrollment, ages one, two, three, four, and seven.25

Exposure Assessment

Traffic related air pollution (TRAP), greenspace, and area socioeconomic status were all estimated using the geocoded residential addresses for each participant from birth through age seven. See the Supplementary Digital Information for more details on each exposure assessment method. TRAP was derived using a previously described and validated LUR model.26 Greenspace was estimated at each address using satellite-derived normalized difference vegetation index (NDVI) images. A raster image of the Cincinnati area was obtained from the United States Forest Service and the average NDVI within 400 m of each address was extracted. NDVI ranges from −1 to 1 and a higher value represents more surrounding greenspace. Area socioeconomic status indices were based on eight census tract level variables (fraction that graduated high school, fraction of households in poverty, median household income, fraction of population receiving public assisted income, fraction of houses that are vacant, median home value, white fraction of population, and black fraction of population) that were collected from the Census 2010 5-year American Community Survey for all census tracts in the counties where CCAAPS subjects resided. A principal component analysis was applied and the first component was called the “deprivation index” with a higher value representing a census tract with increased deprivation. The second principal component was named the “race index”, with a higher value representing a census tract with a higher fraction of black population and lower fraction of white population. Both components were normalized to a range of (0,1) by subtracting the minimum and dividing by difference of the resulting range.

Quantifying Moving Behavior

Because of the longitudinal nature of the study and the reliance upon parental report, yearly addresses did not exactly match even if a family had truly not moved due to abbreviations and other notational conventions. Therefore a fuzzy string matching was implemented to determine if a child had changed home addresses from the previous year. If the street address number and name differed by more than 30% of the total character length according to the generalized Levenshtein edit distance, these were classified as having moved. A 30% change was chosen after assessing the performance by individual inspection of different cutoff percentages on random subsamples (n = 75) of the data. Of 762 total subjects, 149 were lost to follow-up during the study period due to refusal to participate or inability to contact. For these subjects, the address history was right censored at the year in which they were lost. Four moves were outside of the study area and over 60 miles in distance. Because the LUR model only covered the study region, these subjects were classified as moved and right censored at that time. The latitude and longitude coordinate for each address was determined by geocoding using the “EZ-Locate” software from Tele Atlas (Netherlands). These locations were used to assess exposure for TRAP, greenspace, and the socioeconomic characteristics at each address.

Statistical Analysis

Moving distances were calculated based on the geocoded latitude and longitude points using the Vincenty great circle distance method based on the WGS84 ellipsoid. Boxplots were used to compare the percentage changes in exposure assessments at the first, second, third, and all combined moves. Scatter plots were used to further assess the change in exposure assessments for all moves. These plots were overlaid with a local second degree polynomial regression fit line (and 95% confidence interval) spanning over 75% of the data points. Exposure estimates for the total seven year period for all children were calculated based on three methods: (1) using the birth address, (2) using the last known address, and (3) the mean of yearly exposures, calculated every year based on the entire annual address history. These estimates were visually compared using scatter plots and again overlaid with local polynomial regression fits. Agreement between the exposure estimate methods were calculated using the intraclass correlation coefficient based on a one-way model with only the subjects considered as random effects.

Bias in the estimated association with asthma was assessed by calculating the unadjusted odds ratios for the development of asthma at age seven based on separate logistic regression models using the three exposure assessment methods. See the Supplementary Digital Information for the methodological details on asthma diagnosis. For the outcome misclassification analysis, only children with an asthma diagnosis and valid exposure estimate were used in each logistic regression, with the sample size ranging from 581 to 589 because of missing information. A Cox proportional hazards model was used to determine the effects of all exposures on the expected time to a residential change. All exposures were encoded as time-dependent covariates and the proportional hazards assumption was verified by testing the correlation between the transformed time to event and scaled Schoenfeld residuals. Multiple residential changes were present for some subjects and the variance-covariance matrix was adjusted for correlated responses by using the Huber sandwich estimator. To identify if any of the exposures interacted with one another along with the time to a moving event, all pairwise interactions were tested. To illustrate the interaction term in the model, hazard ratios were calculated using example covariate values compared with the overall mean of all exposure covariates. The example covariate values used the overall mean of greenspace, a continuous range of the deprivation index from 0 to 1 and the 10th, 25th, 50th, 75th and 90th percentiles of TRAP. All statistical analyses were carried out in R. All geographic calculations were done in R using the raster,27 geos,28 and rgdal29 packages. Code used for analysis is available upon request.

RESULTS

Cohort and Geographic Characteristics

The total cohort is composed of 762 children with 415 (54%) males, 172 (23%) African-Americans, and 693 (93%) with a mother that had a high school level of education or higher. All children resided in the greater Cincinnati area, which is mostly urban.

Characterization of Residential Mobility

Of the 762 children enrolled in CCAAPS, complete address records were available for 613 through age seven. Since 149 children had missing observations for home address due to loss of follow-up and were right censored, it is likely that these results underestimated the amount of total moves. A majority of the
children in the cohort (54%) moved at least once during the follow-up period and 351 children (46%) did not move. Figure 1a presents a histogram of the number of moves, where 262 children (34%) moved once, 92 children (12%) moved twice, and 57 children (7%) moved three or more times (only eight children moved four times and five children moved five times). Timing of residential changes is important to note because an earlier move is likely to result in greater exposure misclassification due to an increased number of total years in which the exposure classification is incorrect. The frequency of the moves differed by age (Figure 1b), with the majority occurring at ages 3 and 4 (19% and 20% of all moves, respectively). The least amount of moves occurred at age 1 (only 7% of all moves). The median distance for each move was 3.98 miles, with 14% between 0 and 1 miles, 15% between 1 and 2 miles, and 12% between 2 and 3 miles. Figure 1c shows a histogram of the moving distances for all moves, where each bar represents a 1 mile interval. In total, there were 636 total moves and of these, 482 (76%) resulted in a change in distance for each move (each bar represents a 1 mile interval).

Change in Exposures Due to Residential Mobility
Figure 2a shows the percent change in the deprivation index, race index, TRAP, and greenspace measures for each move. For all moves overall, TRAP decreased by a median of 4.4%, greenspace increased by a median of 5.3%, the deprivation index decreased by a median of 7.2%, and the race index did not change. These changes were in the same direction but with a larger magnitude for a subject’s first move. However, the amount of changes decreased with a subject’s second move and even more with their third move. The change in individual variables for all moves used to calculate both TRAP and the deprivation index are summarized in the Supplementary Digital Information. Most of the individual variable changes reflected the same trends as seen overall, with no specific variables particularly responsible for the overall changes in TRAP or the deprivation index.

Figure 2b contains plots of the old versus new exposure variables for each move. Children with lower TRAP values of about 0.5 μg/m³ or less moved on average to new locations with roughly the same amount of TRAP exposure. However, children with existing higher levels of TRAP (about 0.5 μg/m³ and higher) moved on average to a location with lower TRAP. Children with current greenspace values less than an NDVI of 0.5 moved on average to new locations with an average greenspace value of 0.5 and children with higher greenspace values (about 0.5 and higher) moved on average to locations with a lower greenspace. As shown in Figure 2b, children who live in census tracts with deprivation indices < 0.5, in general, move to census tracts with similar deprivation indices. However, children that live in census tracts with deprivation indices > 0.5 moved on average to a location with a decreased deprivation index. Children with race indices above or below 0.6 moved on average to locations with a lower and higher race index, respectively. Overall, the new value of greenspace and the race index did not depend on the old values and likely represent a regression to the mean. However, the mean new value for TRAP and the deprivation index differed depending on the old value, with subjects seeing decreases in TRAP and the deprivation index only if their old exposures were high. Subjects with lower (about 0.5 and below) values of TRAP and deprivation index tended to move to locations with similar exposure values.

Exposure Assessment Error
The intraclass correlation coefficient between each of the three exposure address metrics (first, last, and mean) was calculated (Table 1). The intraclass correlations between the birth exposure estimates and the mean exposure estimates are similar for all four exposures, ranging from 0.85 to 0.93. Examination of the plots between individual exposure estimates with an overlaid locally fit polynomial regression line (and its 95% confidence interval) reveals more nuanced patterns. Figure 3a compares the birth record address estimates with the mean exposure estimates for all subjects. When using the birth address, both the deprivation index and the TRAP exposure were slightly overestimated for subjects with higher exposure estimates. However, the birth record address estimate of greenspace and the race index did not show the same differential pattern.

Exposures were calculated and plotted similarly, but using the last known address instead of the birth record address (Figure 3b). The intraclass correlations for all exposures based on the last address were slightly lower than those using the birth address, ranging from 0.74 to 0.90. Examining the scatter plots (Figure 3b) shows that in this case, using the last address to estimate exposure strongly underestimated the true deprivation indices and TRAP exposures only for those subjects with high exposure values. Again, this differential pattern was not present for the greenspace exposure and race index.

In order to identify the effects of this exposure misclassification bias, we associated each exposure with asthma, diagnosed at the end of the study period (age 7). Table 2 shows the unadjusted odds ratios for each of the separate exposures, each calculated using the birth record address, the last known address, or the mean exposure. When associating asthma with TRAP exposure the
odds ratio using the mean exposure estimate was 3.76, but using the birth record or last known address estimates both biased the estimate towards the null (OR: 2.76 and 2.61, respectively), that is, equivalent to underestimating the true odds ratio by roughly 30% in both cases. Greenspace did not suffer a significant amount of bias, as the mean estimate odds ratio was 0.15 and the birth and last estimate odds ratios were 0.14 and 0.18, respectively. The deprivation index association with asthma was also biased towards the null by roughly 35% when using the birth exposure estimate (OR: 9.28) or the last exposure estimate (OR: 9.50) instead of the mean exposure estimate (OR: 14.65). Finally, the race index showed the largest bias, with a bias towards the null of about 100%, changing from 0.31 for the mean exposure estimate to 0.62 for the birth exposure estimate and 0.69 for the last exposure estimate.

Influence of Exposure Variables on Timing of Move
Greenspace, TRAP exposure, and the deprivation index (but not the race index) were univariately associated with the time to residential change. Therefore, a multivariate cox proportional hazards model was developed by incorporating these three exposures as time-varying covariates. Each hazard ratio (HR) in Table 3 is calculated for a 0.1 unit increase in each covariate as most covariates ranged from zero to one. A 0.1 unit increase in greenspace NDVI was associated with an increased average time until moving (HR: 0.83, 95% CI: (0.76,0.91)), while both increased TRAP (0.1 μg/m³) and deprivation (index of 0.1) were associated with an decreased average time until moving (HR: 1.31, 95% CI: (1.08,1.58) and HR: 1.53, 95% CI: (1.30,1.81), respectively). An interaction between the deprivation index and exposure to TRAP was also observed. Although increased levels of TRAP and deprivation were both associated with a shorter time to a residential change, this effect was reduced when both were high (interaction HR: 0.95, 95% CI: (0.91,0.98)). All other pairwise interactions were not significant and not included in the final model. In order to illustrate this interaction, we calculated the hazard ratios and the accompanying 95% confidence intervals for a range of deprivation indices and TRAP concentrations (Figure 4). Because the cox proportional hazards model is one which can only identify the relative hazard for a given set of covariates, the hazard ratios in Figure 4 were generated by comparison to a subject with covariate values set to the overall means (TRAP: 0.38 μg/m³, deprivation index: 0.41, greenspace: 0.54). As an example, a subject with the highest deprivation index and TRAP exposure at the 10th percentile is about 5.5 more times likely to move at any given time as compared to the average subject. Although increased TRAP exposure on its own increases a subjects hazard to move, its interaction with increased deprivation has a much bigger impact on the hazard. A subject also with the highest deprivation, but with the 90th percentile of TRAP exposure is only 2.25 more times likely to move at any given time as compared to the average subject. Since increasing levels of TRAP exposure and deprivation are associated with a higher risk of moving, exposure misclassification for these subjects will be exaggerated as they are likely to move earlier and their exposure will be assessed using the incorrect address for longer periods of time.

DISCUSSION
This study on a cohort of children residing in an urban location found that residential moves are common and result in changing exposures throughout early childhood, a particularly important time period of enhanced susceptibility to environmental exposures during development. On average, residential changes were associated with decreased exposure to TRAP, increased greenspace, and improved neighborhood deprivation indices. Examining the changes per move suggests that children that move one time are moving to locations with lower deprivation,
lower TRAP, and higher greenspace. However, children that move often throughout childhood do not see as great an increase in the quality of their surroundings. Failing to account for these residential changes by using either the birth or last address caused differential exposure misclassification and bias in estimating the association with asthma. Our finding that children with increased TRAP and deprivation moved at earlier ages further exacerbates the potential for exposure misclassification.

There have been few analyses of the impact of residential moving on exposure assessment in environmental health studies. Two such studies that have been conducted specifically examined maternal residential mobility during pregnancy. The first, conducted in Texas using a case-control cohort from a study of birth defects, found that older age, higher income, hispanic ethnicity, and higher parity were all indicators of lower mobility during pregnancy. Our finding that children with increased TRAP and deprivation moved at earlier ages is consistent with these findings. Furthermore, both our study and the previous study report that exposure misclassification due to residential change is differential. The second study, conducted state-wide in New York using a cohort from a birth defect study, found that there was not a large discrepancy of ozone and PM$_{10}$ concentrations between the beginning of pregnancy and birth. Both of these pollutants, however, have low spatial variability compared to TRAP. In addition, the Chen et al study assessed exposure based on EPA air monitoring data by using buffers, of which the smallest was 247 miles and the average move was 10 miles; hence their finding may be due to the resolution of their exposure assessment model rather than to a true unchanging exposure between moves.

Two other studies have specifically found that accounting for residential mobility in their exposure models lead to stronger associations with health outcomes. In the first, residential proximity to highways was associated with coronary heart disease and accounting for moving away from the highway during the study period resulted in decreased risk while those moving closer to highways had an increased risk. Another cohort study implemented a land use model to assess NO$_2$ exposure and found that the association with diabetes was stronger after calculating exposures using annual address history instead of birth record address. Both of these studies agree with our results that accounting for residential mobility is important to avoid a bias towards the null when associating TRAP, deprivation, and greenspace with asthma.
In contrast to our findings, the ESCAPE project, using a LUR model to assess NO2 exposure, found that the differences in estimated NO2 exposure between the inclusion address and the annual address history concentration were small and non-differential. However, that study was completed in the Swedish county of Vasterbotten, a predominantly rural area, decreasing the impact that a move may have on exposure assessment. In contrast, the CCAAPS cohort was conducted in an urban area and was designed to assess exposure to TRAP with high spatial variability. Thus, relatively small changes in residential location may result in large differences in exposure as most traffic-related pollutants, including EC, ultrafine particles, VOCs, and PAHs are elevated near their source and decay rapidly within 200–300 m.

There are some limitations to our study which should be considered. First, we are unable to account for the time spent indoors and personal activities that could influence exposure to greenspace and TRAP. However, this is usually always the case in cohort studies because accurate and precise location history for a cohort is costly and time consuming. Without a true gold standard of exposure, it is difficult to assess the other biases due to personal activities and time spent indoors. We have recently begun using personal samples for children that detect ultrafine particles and PM, and future research could determine if using more than one address as a location estimate would be useful when estimating exposure based on address history. Here we focused strictly on material deprivation, rather than social deprivation, because although social deprivation is related to respiratory health on an individual level, it is not related to a community level. Another weakness is that this study focuses only on the mobility of families with children and it is unclear if this will generalize to other populations including elderly or people without children. Nonetheless, children are an important population to study and this study will be valuable for future geospatial exposure studies. Specifically, we could not examine the effect of residential mobility during pregnancy because of a lack of location information during this period. It is known that exposures occurring during fetal development are especially important and although not investigated here, exposure misclassification is very important during this time period. Lastly, for the CCAAPS cohort, asthma diagnosis was only available at age seven and although the timing of asthma diagnosis could have allowed for a more in depth analysis of the relationship of the exposures and asthma, that is beyond the scope of this study.

A major strength of this study is the available detailed address history for the CCAAPS cohort. This information is rare in large cohort studies, and it permitted an assessment of the impact of using the birth address or last known address to assess exposure. Furthermore, the high resolution of our TRAP and greenspace models allowed us to identify exposure estimation changes resulting from small moves that could have been overlooked if using a less precise model. Greenspace has recently become a focus of epidemiologic studies, having been positively associated with an increase in perceived health, decreased anxiety and depression, and decreased behavioral and emotional problems in children. This study is the first to examine the influence of residential mobility on greenspace exposure and will be important for future studies given that the association between greenspace and mental health is dependent on age.

In conclusion, using birth record addresses tend to overestimate true exposures because these estimates do not account for residential changes, which usually result in decreased deprivation and TRAP exposure while using the last available address record tends to underestimate true exposures because these estimates do not account for early life exposures that were higher. These exposure misclassifications occur more severely with subjects that have a high deprivation index and TRAP exposure. Furthermore, using the birth record or last known address biases the odds ratio towards the null and underestimates the true association of TRAP exposure and the deprivation index with asthma diagnosis as compared to using the mean exposure estimate. This finding suggests that using only one address as opposed to the annual address history to estimate these characteristics and exposures can lead to misclassification error and bias in longitudinal epidemiologic studies.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

ACKNOWLEDGEMENTS
The authors thank the study and clinic staff for their efforts in study coordination, subject recruitment, data management, and data collection. They also thank the participating CCAAPS families for their time and effort. This work was supported by grants from the National Institute of Environmental Health Sciences (5R01ES011170 and R01ES019890).

REFERENCES
1 Aguilera I, Sunyer J, Fernández-Patier R, Hoek G, Aguirre-Allaro A, Melefsté K et al. Estimation of outdoor NOx, NO2, and BTEX exposure in a cohort of pregnant women using land use regression modeling. Environ Sci Technol 2007; 42: 815–821.
2 Beelen R, Hoek G, Vienneau D, Eeftens M, Dimakopoulou K, Pedeli X et al. Development of NO 2 and NO x land use regression models for estimating air pollution exposure in 36 study areas in Europe—the ESCAPE project. Atmos Environ 2013; 72: 10–23.
Supplementary Information accompanies the paper on the Journal of Exposure Science and Environmental Epidemiology website (http://www.nature.com/jes)