Journal of Exposure Science and Environmental Epidemiology (2011) 21, 42–48

www.nature.com/jes

The impact of daily mobility on exposure to traffic-related air pollution and health effect estimates

ELEANOR SETTON[a], JULIAN D. MARSHALL[b], MICHAEL BRAUER[c], KATHRYN R. LUNDQUIST[b], PERRY HYSTAD[d], PETER KELLER[a] AND DENISE CLOUTIER-FISHER[a]

[a]*Geography Department, University of Victoria, Victoria, British Columbia, Canada*
[b]*Department of Civil Engineering, University of Minnesota, Minneapolis, Minnesota, USA*
[c]*School of Environmental Health, University of British Columbia, Vancouver, British Columbia, Canada*
[d]*School of Population and Public Health, University of British Columbia, Vancouver, British Columbia, Canada*

Epidemiological studies of traffic-related air pollution typically estimate exposures at residential locations only; however, if study subjects spend time away from home, exposure measurement error, and therefore bias, may be introduced into epidemiological analyses. For two study areas (Vancouver, British Columbia, and Southern California), we use paired residence- and mobility-based estimates of individual exposure to ambient nitrogen dioxide, and apply error theory to calculate bias for scenarios when mobility is not considered. In Vancouver, the mean bias was 0.84 (range: 0.79–0.89; SD: 0.01), indicating potential bias of an effect estimate toward the null by $\sim 16\%$ when using residence-based exposure estimates. Bias was more strongly negative (mean: 0.70, range: 0.63–0.77, SD: 0.02) when the underlying pollution estimates had higher spatial variation (land-use regression *versus* monitor interpolation). In Southern California, bias was seen to become more strongly negative with increasing time and distance spent away from home (e.g., 0.99 for 0–2 h spent at least 10 km away, 0.66 for ≥ 10 h spent at least 40 km away). Our results suggest that ignoring daily mobility patterns can contribute to bias toward the null hypothesis in epidemiological studies using individual-level exposure estimates.
Journal of Exposure Science and Environmental Epidemiology (2011) **21,** 42–48; doi:10.1038/jes.2010.14; published online 30 June 2010

Keywords: *measurement error, bias, air pollution, space-time modeling.*

Introduction

Accurately measuring or estimating exposure to outdoor air pollution can be a challenge for epidemiological studies of large populations, especially given peoples' tendency to move about over time through changing pollution concentrations. When logistics and costs preclude personal monitoring, surrogates of personal exposure are used, which may introduce error into epidemiological analyses.

The magnitude of error in health effect estimates that is caused by the use of a surrogate exposure measure depends in part on study design. Time-series studies of acute effects are relatively insensitive to exposure error (Zeger et al., 2000; Sheppard et al., 2005). Similarly, cohort studies that include multiple communities, assign exposure based on community-average pollution concentration, and consider long-term health effects are relatively unaffected by a lack of personal exposure measures (Berhane et al., 2004). In both cases, the error introduced by using the surrogate measure is of Berkson type (i.e., true exposures vary about the assigned mean exposure for the group), which has been shown to widen the confidence interval, but not alter the magnitude or direction of the effect estimate (Armstrong, 1998). However, when variability in pollution concentrations (and exposures) is higher within community than between community, there is potential for exposure misclassification and underestimation of health effects (Navidi and Lurmann, 1995; Jerrett et al., 2005; Wilson et al., 2005; Miller et al., 2007).

A growing number of studies of long-term effects of exposure to air pollutants use model-based estimates of ambient (outdoor) pollution to better capture within-community variability in exposure (Hoek et al., 2008; Nordling et al., 2008). For this approach, exposure is assumed to equal the outdoor concentration at subjects' residential locations. Often, for reasons of confidentiality, only the postal codes or census areas of subjects' residences are available. Hereafter, we use the term "residence" to indicate any of these potential geographic identifiers. Recent literature using this approach emphasizes increasingly high-resolution estimates to better characterize concentrations at residential locations. However, if study subjects spend time away from home — for example, at work, school, or shopping — this approach

1. Address all correspondence to: Dr. Eleanor Setton, Geography Department, University of Victoria, PO Box 3060 STN CSC, Victoria, BC V8W 3R4, Canada.
Tel.: +250 472 4624. Fax: +250 721 6216.
E-mail: elsetton@uvic.ca
Received 1 November 2009; accepted 12 February 2010; published online 30 June 2010

may lead to errors in individual exposure estimates. In this study, we argue that these errors are classical (i.e., the surrogate measure varies around the true exposure), rather than Berkson (true exposures vary about the assigned mean), and therefore have the potential to bias health effect estimates toward the null (Armstrong, 1998). Little or no extant research investigates bias caused by daily mobility. Our study explores the potential magnitude of this bias.

Evidence for the effect of mobility on exposure levels can be found in empirical studies that compare personal monitoring with ambient monitoring at subjects' residences. A study of nitrogen dioxide (NO_2) in Portage, Wisconsin, correlated personal exposure with outdoor NO_2 at residence, full-time worker status, and commute distance (Quackenboss et al., 1986). In Helsinki, Kousa et al. (2001) found that work locations downtown were one of the strongest predictors of personal exposure to NO_2; other predictors included outdoor concentrations at residential or central locations. Nethery et al. (2008) compared personal exposure monitoring of NO_2 with land-use regression (LUR) estimates of outdoor concentrations developed by Henderson et al. (2007). The weak but statistically significant correlation between personal monitoring and LUR estimates at the residential postal code (Pearson's $r = 0.18$) was improved when LUR concentrations at subjects' work locations were included (Pearson's $r = 0.28$). The preceding examples focus on NO_2, which shows relatively high spatial variability within communities because of its association with vehicle emissions. As expected, daily mobility is of less concern for exposure estimation when pollutant concentrations are spatially homogeneous. For example, Strand et al. (2006) report that for sulfate, children's personal exposure was highly correlated with home indoor ($r = 0.94$) and with school outdoor ($r = 0.92$) concentrations. Sulfate has no major indoor sources and typically is not expected to vary significantly at scales $<$100–1000 km (Gilliland et al., 2005).

Exposure simulation studies provide an opportunity to explore how mobility affects exposure estimates. Setton et al. (2008) used census-based work flow estimates and high-resolution (LUR) estimates of annual average NO_2 concentrations developed by Henderson et al. (2007) to simulate census tract-specific exposure distributions for working people in Metro Vancouver. They show that although time spent at residential locations contributes most to exposure differences among census tracts, time spent at work locations contributes most to within-census tract variability in exposures. Marshall et al. (2006) combined geocoded origin–destination survey records ($n = 25{,}064$) for California's South Coast Air Basin (SoCAB), time-varying outdoor concentrations estimates from the CAMx air dispersion model, and a mobility-based exposure model that accounts for time spent indoors, outdoors, and in vehicles. Their results show that average inhalation intake rates are higher for mobility-based estimates than for residence-only estimates. Increases were higher for butadiene and particulate hexavalent chromium (30 and 27%, respectively) than for ozone, benzene, and fine particulate matter emitted by diesel engines (2, 5, and 8%, respectively).

Using results obtained from the study by Setton et al. (2008) for Metro Vancouver and from Marshall et al. (2006) and Marshall (2008) for Southern California, we investigate the potential bias in relative risk estimates associated with using outdoor pollution levels at the residential address only as exposure measurements *versus* estimates that incorporate time spent away from homes. We examine three hypotheses specific to mobile populations: (1) ignoring daily mobility (e.g., using residence-only exposure estimates) will contribute to negative bias in effect estimates; (2) increasing spatial variation in pollution estimates will lead to stronger negative bias; and (3) negative bias will be stronger as distance and time spent away from residence increases.

Methods

Metro Vancouver

Metro Vancouver (population: 2.1 million; study area: $\sim 60 \times 60\,km^2$) is a relatively low-pollution urban area in southwestern British Columbia. The exposure data set was produced to investigate spatial variability and mobility effects on exposures; see Setton et al. (2008). For each of 382 census tracts, microenvironment simulation produced two distributions ($n = 10{,}000$ each) of time-weighted, seasonally adjusted estimates of annual exposure to NO_2: one distribution assumes that all time is spent at the residential location; the other distribution incorporates time spent away from home and associated changes in pollution levels. Time–activity patterns from the CHAPS (Canadian Human Activity Pattern Survey) with weekday work were used to simulate exposures for "workers" in each census tract. Work flow data obtained from the 2001 census reporting were used to identify work destination census tracts; records reporting the same census tract for work and residence were defined as "non-mobile" and were omitted from the simulation. Time spent neither at work nor at home was assumed to be within 5 km of the residential census tract centroid. Although incorporated in the algorithm to estimate exposure, record-specific information on the maximum distance and time away from home is not retained in the output; therefore, analysis of the impact of increasing time and distance away from home cannot be conducted with the Metro Vancouver data set. We used two approaches for estimating spatial variability in annual average NO_2 concentrations: LUR (Figure 1a) (Henderson et al., 2007) and inverse-distance weighted (IDW) interpolation of monitoring station data (Figure 1b) (Setton et al., 2008). For each pollution estimate (IDW and LUR), we created 10,000 population samples (each with $n = 382$), by randomly selecting a single estimate from each

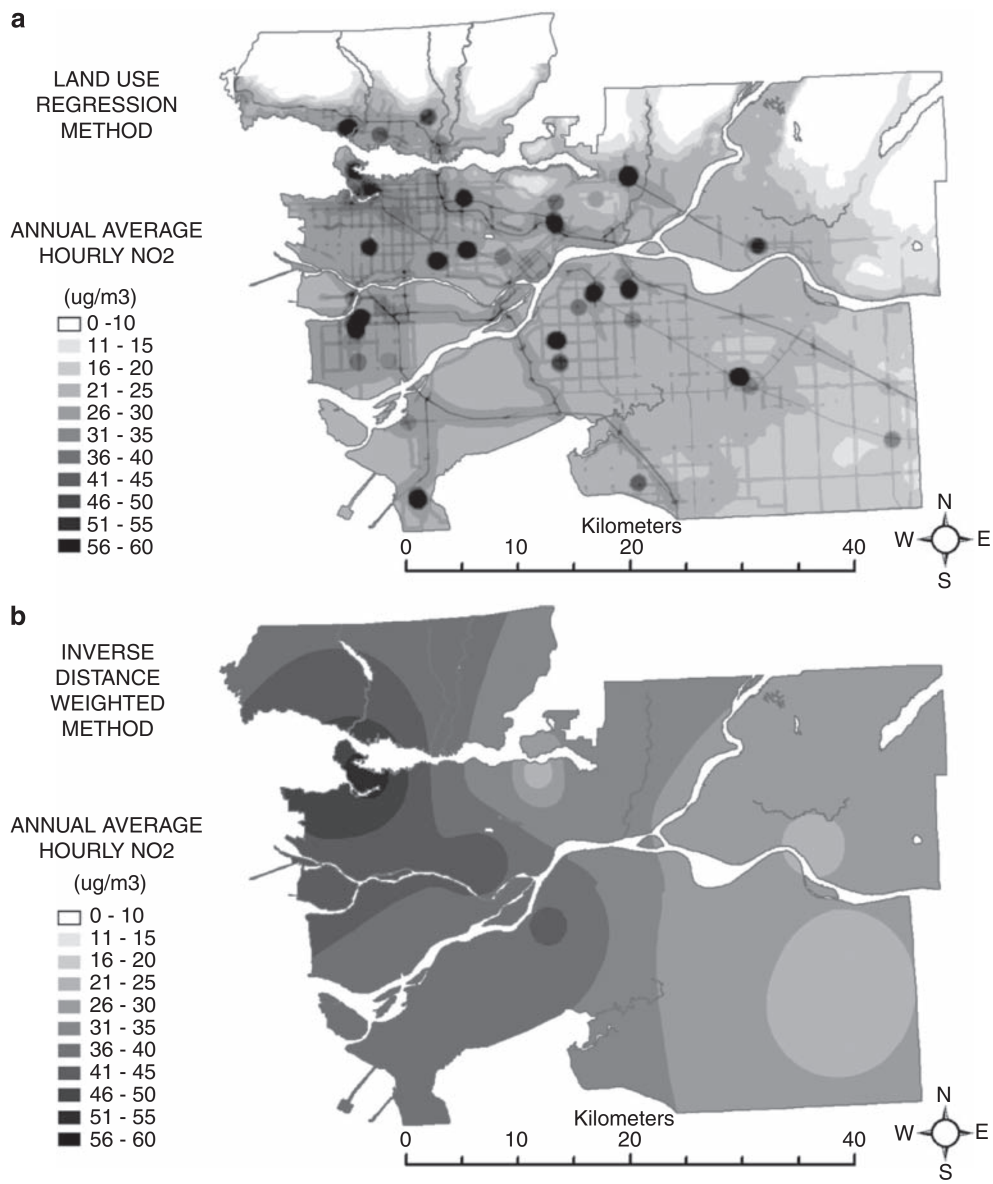

Figure 1. Annual average nitrogen dioxide concentrations, Metro Vancouver: (a) land-use regression, (b) inverse distance weighting of the 16 monitoring stations.

of the census tract distributions, then calculating the associated bias.

Southern California

The SoCAB (population: 15.9 million; study area: $\sim 70 \times 120\,km^2$) is an urban environment in Southern California well known for its poor air quality. The exposure data set consists of transportation survey records for 25,064 individuals residing in the SoCAB who participated in the Southern California Association of Governments (SCAG) year-2000 transportation survey (SCAG — Southern California Association of Government, 2003). Survey respondents recorded travel activities (such as type of activity, start and end time, location for trips) for 24-h periods; SCAG geocoded all travel activities (the latitude and longitude of trip origins and destinations). Each individual record and trip was spatially and temporally linked to hourly concentration estimates (grid cells: $2 \times 2\,km^2$) from the commercially available CAMx dispersion model. As with the Vancouver data set, two exposure estimates were generated per person: one incorporating mobility and the other assuming that subjects spent all time at their residential locations. Annual average NO_2 concentrations produced by the CAMx model are shown in Figure 2.

The SCAG database contains a 1-day record for most (85%) individuals and a 2-day record for the remaining

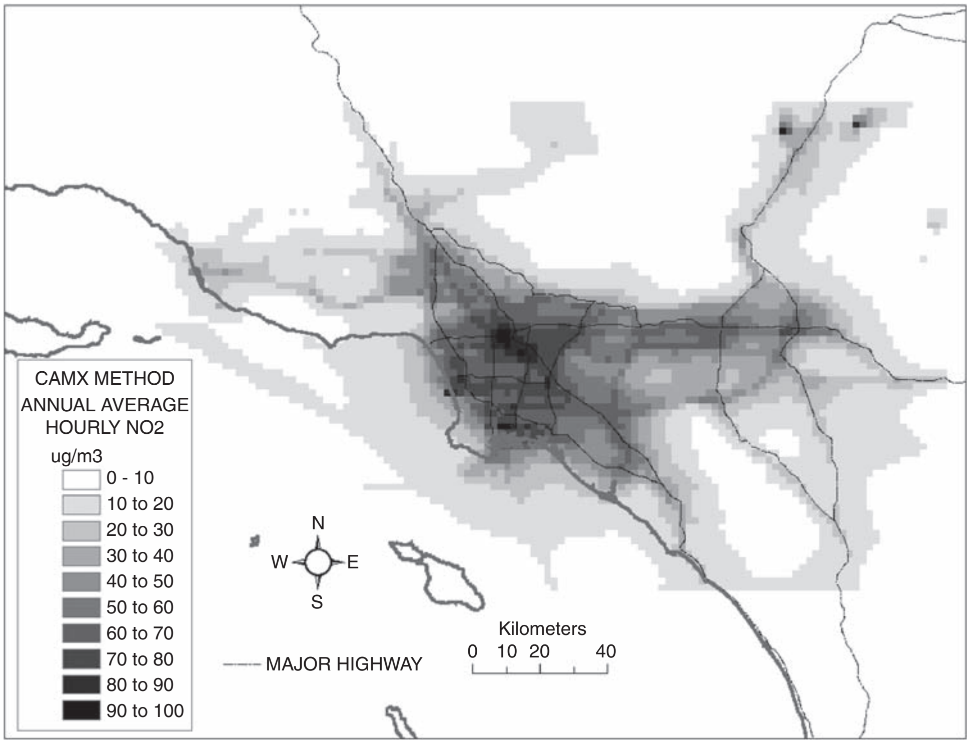

Figure 2. Annual average nitrogen dioxide concentrations, South Coast Air Basin.

(15%) individuals. To avoid overrepresentation for some individuals, we deleted all second-day records. Records reporting zero for maximum distance traveled, but with larger than zero values in time spent away from home (148 records or 0.6%) were deleted as incomplete. Given our interest in exploring bias associated with mobile populations, records indicating no time spent away from home (6369 records or 25%) were considered as non-mobile and deleted, leaving 18,547 mobility records for our analysis. We created 10,000 samples (each with $n = 3600$, approximating a 20% sample of the full data set) and calculated the bias associated with each. In addition, because the SoCAB data set includes time spent away from home and distance from home, we explored the effect on bias of increasing time and distance away from home locations. To do so, we calculated bias for subsamples of records showing specific ranges of time and distance from home (see below).

Comparison of the Two Data sets

The data sets used in this study have been developed for different purposes and by different methods, and so are dissimilar in the following ways: (1) the SoCAB data represent actual individuals, with travel information identifying real trip origins, destinations, and time spent away from home for all mobile individuals. The Metro Vancouver data set uses a Monte Carlo approach to simulate representative work-related time–activity patterns; results represent the distribution of possible exposures per census tract, rather than any specific individuals, for workers. (2) The SoCAB data provide a 1-day estimate of travel patterns and pollutant exposures per person, whereas the Vancouver estimates are annual averages. The two methods differ; we do not expect the results to be directly comparable. Rather, both approaches provide distinct yet meaningful information about bias factors attributable to mobility.

Calculation of a Bias Factor

General classical error theory suggests that:

$$Z = X + E \qquad (1)$$

where Z is the value of the surrogate (or observed) measure, X the true value, and E the error in measuring X (Armstrong, 1990). For our calculations, we assume that residence-only-based estimates of exposure are analogous to Z (the typically used surrogate measure) and that the mobility-based estimates are closer to X (true measures, such as might be collected using personal monitoring). Therefore, error E for each individual equals Z (the residence-only estimate) minus X (the mobility estimate).

For the classical error model, E is assumed to be independent of X, but in our data sets, E shows varying degrees of positive correlation with X. Therefore, we use the following equation, provided in the study by Wacholder (1995), to calculate the bias factor expected in regression coefficients of simple linear models in the presence of the correlation between E and X:

$$\mathrm{BIAS} = \frac{\sigma^2 + \phi}{\sigma^2 + 2\phi + \omega^2} \tag{2}$$

where, σ^2 is the variance of X, ϕ the covariance of (X, E), and ω^2 the variance of E (Wacholder, 1995). Equation (2) provides a multiplier that would apply to the relative risk estimate produced using the surrogate (residence-only) exposure estimates. For example, if Eq. (2) yields the value 0.75, then the bias due to the use of the surrogate (residence-only) measure is negative and the relative risk is being underestimated by 25%.

Results

For Metro Vancouver, the mean bias associated with using residence-only NO_2 estimates is 0.70 (range: 0.63–0.77; SD: 0.02) for the LUR approach and 0.84 (range: 0.79–0.89; SD: 0.01) for the IDW approach (Table 1).

Biases for subsamples of the SoCAB data set with selected combinations of time and distance away from home are shown in Figure 3. Bias is negative (Eq. (2) values are <1.0) and increases in magnitude (Eq. (2) values are increasingly <1.0) with increasing time and distance spent away from home locations, from 0.99 for 0–2 h spent at least 10 km away to 0.61 for ≥ 10 h spent at least 40 km away. For the entire SoCAB data set ($n = 18{,}547$), the overall bias is 0.93 (range: 0.91–0.95; SD: 0.005).

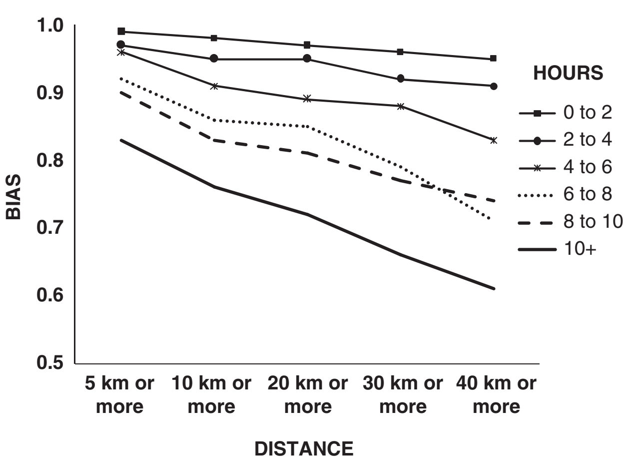

Figure 3. Calculated bias and its dependence on time and distance spent away from home, South Coast Air Basin.

Table 1. Descriptive statistics for bias calculation results

Sample	Mean	Median	SD	Min	Max	Range
Metro Vancouver, LUR	0.70	0.70	0.02	0.62	0.78	0.16
Metro Vancouver, IDW	0.84	0.85	0.013	0.80	0.90	0.10

Our results support all three hypotheses: ignoring daily mobility (e.g., using residence-only exposure estimates) contributes to negative bias in effect estimates; increasing spatial variation in pollution estimates leads to stronger negative bias; and negative bias is stronger as distance and time spent away from residence increases.

Discussion

Ignoring geographic mobility through the use of residence-only exposure estimates produced negative bias in all of our data sets, a condition which occurs when the variance of the "true" (mobility-based) estimate is smaller than that of the surrogate (residence-only) estimate, given the classical error model (Armstrong, 1990). In both data sets, residential locations of the subjects (real or simulated) are well dispersed geographically; therefore, the variance of the residential-only exposure estimates reflects the spatial variability of the pollution map. The mobility-based exposures are based on time-weighted averages of concentrations from various locations in the study area, and therefore represent spatially averaged concentrations, resulting in mobility-based exposure data sets with lower variation than the associated residence-only data sets.

In Metro Vancouver, bias was more strongly negative for the LUR approach than for the IDW monitoring-based approach; given that spatial variability in NO_2 concentrations is greater for LUR than for the monitoring approach; this result supports our intuitive hypothesis that stronger negative bias will exist when pollution variability is high. Although this result is derived in this study from different methods to map one pollutant (NO_2), we expect that the same result would apply when comparing pollutants with different levels of spatial variability.

Finally, we expected that negative bias will be stronger with increasing distance and time spent away from home. That hypothesis is well supported by the analysis of the SoCAB data set. This finding illustrates the importance of understanding the mobility characteristics of a population used for epidemiological studies of outdoor air pollution impacts on population health.

To our knowledge, the analyses presented in this study are the first to explore the effect of using residence-only-based estimates of ambient air pollution concentrations as exposure measures for large cohort-based epidemiological studies, instead of mobility-based measures. In a somewhat analogous scenario based on empirical data (concentrations measured at school-only *versus* personal monitors), Van Roosbroeck et al. (2008) report the unadjusted (exposure measured as the ambient concentration of NO_2 and soot at school locations) and adjusted (based on regression calibration with personal monitoring data) prevalence ratios for four respiratory conditions in school children in the Netherlands. For NO_2, the bias factor apparent in the prevalence ratios ranges from 0.33 to 0.54. This suggests a stronger negative bias than our results. Our results are limited to ambient exposures only; the impact of capturing exposure to indoor sources of NO_2 at the children's home locations may contribute to the increased bias detected in their study.

Our study is limited by the use of data sets that were developed for other purposes, and so we cannot fully explore the hypotheses put forth. We note that the overall bias calculated for the SoCAB data ($n = 18,547$) is relatively weak in comparison with that of the Metro Vancouver data set. We identify this as an opportunity for further investigation, in the event that the data sets can be made more directly comparable. However, as a first step, the analyses presented in this study provide useful information and suggest future research avenues that may be fruitful in developing a deeper understanding of when it is most advisable to incorporate mobility-based estimates in epidemiological analyses.

Although these results support our hypothesis that ignoring mobility results in negative bias, they also suggest more complex relationships that depend on the spatial distribution of the sample population in relation to the spatial distribution of pollution. In both study areas, the sample populations are well dispersed spatially. This may not always be the case in epidemiology studies. For example, a sample population that is clustered near a single point source of pollution might have similar exposures based on residential locations, but more variable mobility-based exposures when considered as a group. In this case, the Berkson error may dominate (“true” measures vary around residence-only measures), in which case, no negative bias would occur. Another scenario could reflect a spatially dispersed study population, but a pollution map that is dominated by a large point source in one area. Only those subjects traveling away from or into the “hot spot” would have mobility-based exposures much different than their residence-based estimates, with likely unpredictable effects on the variability of the residence-only and mobility-based exposures. These hypotheses are presented in this study as potential avenues for future simulation studies.

In conclusion, our study illustrates that negative bias due to ignoring mobility can occur in simple linear models of exposure to outdoor air pollution and health effects, and that bias may be stronger as the spatial variability of pollution concentrations or of peoples' locations increase. Gilliland et al. (2005) suggest that high spatial variability (≤ 50 m to 4 km) is expected for NO_2 and nitric oxide, elemental carbon, organics including polycyclic aromatic hydrocarbons, and metals, such as hexavalent chromium, cadmium, lead, beryllium, nickel, arsenic, iron, and manganese. Epidemiological studies of these pollutants can show the greatest benefits from incorporating information on mobility in exposure estimates. Bias may also become more strongly negative as the geographic mobility (time and distance away from home) of the study population increases; however, the spatial pattern of pollution may modify this effect, especially if a few discrete “hot spots” of pollution are present in the study area. Future research on how different patterns of pollution modify bias due to geographic mobility is warranted, as is research on the potential bias related to ignoring mobility in more complex health effects models.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

We thank Dr. Ben Armstrong for his guidance during the conceptual development of this paper. Partial funding was provided by Health Canada and by a grant from the US Environmental Protection Agency's Science to Achieve Results (STAR) program. Although the research described in the article has been funded in part by the US Environmental Protection Agency's STAR program through Grant RD-83362401-0, it has not been subjected to any EPA review and therefore does not necessarily reflect the views of the Agency, and no official endorsement should be inferred.

References

Armstrong B.G. The effects of measurement errors on relative risk regressions. *Am J Epidemiol* 1990: 132(6): 1176–1184.

Armstrong B.G. Effect of measurement error on epidemiological studies of environmental and occupational exposures. *Occup Environ Med* 1998: 55(10): 651–656.

Berhane K., Gauderman W.J., Stram D.O., and Thomas D.C. Statistical issues in studies of the long-term effects of air pollution: the Southern California Children's Health Study. *Stat Sci* 2004: 19(3): 414–434.

Gilliland F., Avol E., Kinney P., Jerrett M., Dvonch T., and Lurmann F., et al. Air pollution exposure assessment for epidemiologic studies of pregnant women and children: lessons learned from the Centers for Children's Environmental Health and Disease Prevention Research. *Environ Health Perspect* 2005: 113(10): 1447–1454.

Henderson S.B., Beckerman B., Jerrett M., and Brauer M. Application of land use regression to estimate long-term concentrations of traffic-related nitrogen oxides and fine particulate matter. *Environ Sci Technol* 2007: 41(7): 2422–2428.

Hoek G., Beelen R., de Hoogh K., Vienneau D., Gulliver J., and Fischer P., et al. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmos Environ* 2008: 42(33): 7561–7578.

Jerrett M., Burnett R.T., Ma R.J., Pope C.A., Krewski D., and Newbold K.B., et al. Spatial analysis of air pollution and mortality in Los Angeles. *Epidemiology* 2005: 16(6): 727–736.

Kousa A., Monn C., Rotko T., Alm S., Oglesby L., and Jantunen M.J. Personal exposures to NO2 in the EXPOLIS-study: relation to residential indoor, outdoor and workplace concentrations in Basel, Helsinki and Prague. *Atmos Environ* 2001: 35(20): 3405–3412.

Marshall J.D. Environmental inequality: air pollution exposures in California's South Coast Air Basin. *Atmos Environ* 2008: 42(21): 5499–5503.

Marshall J.D., Granvold P.W., Hoats A.S., McKone T.E., Deakin E., and Nazaroff W.W. Inhalation intake of ambient air pollution in California's South Coast Air Basin. *Atmos Environ* 2006: 40(23): 4381–4392.

Miller K.A., Siscovick D.S., Sheppard L., Shepherd K., Sullivan J.H., and Anderson G.L., et al. Long-term exposure to air pollution and incidence of cardiovascular events in women. *N Engl J Med* 2007: 356(5): 447–458.

Navidi W., and Lurmann F. Measurement error in air-pollution exposure assessment. *J Expo Anal Environ Epidemiol* 1995: 5(2): 111–124.

Nethery E., Leckie S.E., Teschke K., and Brauer M. From measures to models: an evaluation of air pollution exposure assessment for epidemiological studies of pregnant women. *Occup Environ Med* 2008: 65(9): 579–586.

Nordling E., Berglind N., Melen E., Emenius G., Hallberg J., and Nyberg F., et al. Traffic-related air pollution and childhood respiratory symptoms, function and allergies. *Epidemiology* 2008: 19(3): 401–408.

Quackenboss J.J., Spengler J.D., Kanarek M.S., Letz R., and Duffy C.P. Personal exposure to nitrogen-dioxide — relationship to indoor outdoor air-quality and activity patterns. *Environ Sci Technol* 1986: 20(8): 775–783.

SCAG — Southern California Association of Government. Year 2000 Post-Census Regional Travel Survey, Final Report of Survey Results, Southern California Association of Government, Austin, TX, 2003.

Setton E.M., Keller C.P., Cloutier-Fisher D., and Hystad P.W. Spatial variations in estimated chronic exposure to traffic-related air pollution in working populations: a simulation. *Int J Health Geogr* 2008: 7.

Sheppard L., Slaughter J.C., Schildcrout J., Liu L.J.S., and Lumley T. Exposure and measurement contributions to estimates of acute air pollution effects. *J Expo Anal Environ Epidemiol* 2005: 15(4): 366–376.

Strand M., Vedal S., Rodes C., Dutton S.J., Gelfand E.W., and Rabinovitch N. Estimating effects of ambient PM2.5 exposure on health using PM2.5 component measurements and regression calibration (vol 16, pg 30, 2006). *J Expo Sci Environ Epidemiol* 2006: 16(5): 471.

Van Roosbroeck S., Li R.F., Hoek G., Lebret E., Brunekreef B., and Spiegelman D. Traffic-related outdoor air pollution and respiratory symptoms in children - the impact of adjustment for exposure measurement error. *Epidemiology* 2008: 19(3): 409–416.

Wacholder S. When measurement errors correlate with truth — surprising effects of nondifferential misclassification. *Epidemiology* 1995: 6(2): 157–161.

Wilson J.G., Kingham S., Pearce J., and Sturman A.P. A review of intraurban variations in particulate air pollution: implications for epidemiological research. *Atmos Environ* 2005: 39(34): 6444–6462.

Zeger S.L., Thomas D., Dominici F., Samet J.M., Schwartz J., and Dockery D., et al. Exposure measurement error in time-series studies of air pollution: concepts and consequences. *Environ Health Perspect* 2000: 108(5): 419–426.