Published in final edited form as:

Epidemiology. 2016 January ; 27(1): 51–56. doi:10.1097/EDE.0000000000000404.

A new technique for evaluating land use regression models and their impact on health effect estimates

Meng Wang^{1,2}, Bert Brunekreef^{1,3}, Ulrike Gehring¹, Adam Szpiro⁴, Gerard Hoek¹, and Rob Beelen¹

¹Institute for Risk Assessment Sciences, Utrecht University, Utrecht, the Netherlands

²Department of Environmental and Occupational Health Sciences, University of Washington, Seattle, Washington, USA ³Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands ⁴Department of Biostatistics, University of Washington, Seattle, Washington, USA

Abstract

BACKGROUND—Leave-one-out cross-validation that fails to account for variable selection does not properly reflect prediction accuracy when the number of training sites is small. The impact on health effect estimates has rarely been studied.

OBJECTIVES—Develop an improved validation procedure for land-use regression models with variable selection and investigate health effect estimates in relation to land-use regression model performance.

METHODS—We randomly generated ten training and test sets for nitrogen dioxide and particulate matter. For each training set we developed models and evaluated them using a cross-holdout validation approach. Cross-holdout validation develops new models for each evaluation compared to refitting the model without variable selection, as in standard leave-one-out cross-validation. We also implemented holdout validation, which evaluates model predictions using independent test sets. We evaluated the relationship between cross-holdout validation and holdout validation R^2 and estimates of the association between air pollution and forced vital capacity in the Dutch birth cohort.

RESULTS—Cross-holdout validation R^2 s were generally identical to holdout validation R^2 s, but were notably smaller than leave-one-out cross-validation R^2 s. Decreases in forced vital capacity in relation to air pollution exposure were larger for land-use regression models that had larger holdout validation and cross-holdout validation R^2 s rather than leave-one-out cross-validation R^2 .

Conclusion—Cross-holdout validation accurately reflects predictive ability of land-use regression models and is a useful validation approach for small datasets. Land-use regression predictive ability in terms of hold-out validation and cross-holdout validation rather than leave-

Corresponding author: Meng Wang, Department of Environmental and Occupational Health Sciences, School of Public Health, University of Washington, Seattle, WA, USA, tel: +1 (206) 839-7477, wang0109@u.washington.edu.

The authors report no conflicts of interest.

one-out cross-validation was associated with the magnitude of health effect estimates in a case study.

Introduction

Long-term exposure to air pollution has been associated with adverse health outcomes.¹ Recent epidemiologic studies increasingly relied on modeling techniques for estimation of individual air pollution exposure. Land-use regression modeling which uses land use, geographic, and traffic characteristics to explain spatial variations of air pollution concentrations measured at multiple sites in a study area is one of the most popular approaches.

Model evaluation is essential especially when land-use regression models are based on relatively small numbers of training sites.^{2,3} A commonly used evaluation approach for land-use regression modeling is leave-one-out cross-validation: a model is developed using N training sites, this model is refitted N times without variable selections using $N-1$ sites, the N refitted models are used to predict the concentrations at the left-out sites, and the correlation between these N model predictions and the measured concentrations at these sites are calculated. Previous studies suggested that this method may overestimate the predictive ability of land-use regression models at truly independent sites because the predicted sites were not completely independent from model development: the N models based on $N-1$ sites are refitted using the selected predictors from the original model, and were not rebuilt with each iteration.^{2,3} This statistical issue has been described previously and it has been concluded that true validation (which we call holdout validation) must be applied to the entire sequence of modeling steps including variable selection for any set of training and test datasets.⁴ By doing so, the land-use regression model is developed using a training dataset and the model is validated on a completely independent test dataset. The holdout validation approach likely better reflects the predictive power of the land-use regression model at addresses of subjects where no measurements were taken, assuming that the validation sites are representative of the distributions of subjects' addresses. Examples are discussed in several studies.⁵⁻⁸ However, this is only feasible when sufficient numbers of measurement sites are available for separation into two independent datasets of sufficient size.

Few studies have systematically analyzed the extent to which effect estimates in epidemiologic studies are affected by the prediction error associated with the application of land-use regression models for exposure assessment. A previous study has suggested that the variance of such effect estimates could be substantial when a small number of sites is used for land-use regression modeling.⁹ Hence, it is necessary to investigate the variability of effect estimates associated with land-use regression model prediction errors.

In this study, we developed a new technique for evaluation of land-use regression models, combining leave-one-out cross validation and hold-out validation. We investigated whether the use of our method can represent the predictive ability of a land-use regression model at independent sites. Secondly, we explored the relationships between land-use regression model prediction errors and the magnitude of effect estimates using forced vital capacity

data from the Dutch Prevention and Incidence of Asthma and Mite Allergy birth cohort study.

Methods

Model development

We used measured annual average concentrations of nitrogen dioxide (NO₂), particulate matter with diameters <2.5µm (PM_{2.5}), and PM_{2.5} absorbance (an index of black carbon) from the European Study of Cohorts for Air Pollution Effects in the study area that covered the Netherlands and Belgium. The measurement sites for NO₂ (80 sites) and PM (40 sites) were spread over the Netherlands and part of Belgium and were measured between February 2009 and February 2010, allowing us to split the data for validation. We developed land-use regression models were developed for the three air pollutants using a supervised stepwise linear regression method. Predictor variables included European-wide common variables such as road length, residential density and land use as well as local specific traffic intensity and population density variables (eTable 1). A detailed description of the measurement and land-use regression model development procedures have been published elsewhere.¹⁰⁻¹³

Model evaluation: Combining leave-one-out cross validation and hold-out validation

We created a flowchart to illustrate the main evaluation procedures and the terms used in the analysis (Figure 1). We divided our sampling sites into a training set and a test set, each of which contained 50% of all sites. We performed ten random stratified selections of sites according to site type (urban background, rural background, and near street sites) and generated ten training sets and ten test data sets. Then we developed models for each of the training sets.

We used three approaches to evaluate the models: leave-one-out cross-validation, holdout validation (representing the true predictive ability of the land-use regression models at the external test sites), and cross-holdout validation. Cross-holdout validation is a combination of the cross-validation and hold-out validation approach, which requires variable selection during the validation process. Unlike the leave-one-out cross-validation which excludes one site and refits the already developed model with the remaining N-1 sites (model structure is fixed, only coefficients change), we successively built N new evaluation models based on the N-1 sites until each of the sites had been removed and predicted by the evaluation models once. Therefore, each of the test sites was completely independent from the model building. Cross-holdout validation is a surrogate for holdout validation and is calculated from N evaluation models with N-1 sites. We reported the output of leave-one-out cross-validation, cross-holdout validation, and holdout validation by the regression (R²) of predictions and observations at the left-out sites to be comparable with previous studies.

Comparison of R²s between cross-holdout validation and holdout validation is indirect because of different models derived from distinct data sets for validation (cross-holdout validation: training sets; holdout validation: test sets). Therefore, we additionally calculated an intermediate holdout validation R² based on N evaluation model predictions to the same test sets in order to link the above validation approaches. The leave-one-out cross-validation

and cross-holdout validation were conducted within the training sets while the holdout validation applied the land-use regression models from the training sets to predict the concentrations of air pollutants at the test sets. Regression based R^2 values were calculated and the entire simulation processes were repeated for all ten sets of training and test datasets.

Model evaluation: Variance in health effect estimates

To evaluate the variability in health effect estimates due to choice of a specific land-use regression model, we selected forced vital capacity measured at age 8 years, from the Dutch Prevention and Incidence of Asthma and Mite Allergy birth cohort study, as the health outcome for our case study. We previously showed negative associations between forced vital capacity and air pollutants (NO_2 , $\text{PM}_{2.5}$, and $\text{PM}_{2.5}$ absorbance).¹⁴ Ethics approval to perform the study was obtained from the local authorized institutional review boards, and written informed consent was obtained from the parents or legal guardians of all participants. More information about the Prevention and Incidence of Asthma and Mite Allergy study can be found elsewhere.^{15,16}

For each of the ten land-use regression modeling training data sets, exposures to NO_2 , $\text{PM}_{2.5}$, and $\text{PM}_{2.5}$ absorbance were estimated using the default model with N sites ($N=40$ for NO_2 and $N=20$ for PM) and N evaluation models with N-1 sites. Then the estimates of each model were linked to health data individually. We used linear regression analyses with natural log (ln)-transformed forced vital capacity as dependent variable to analyze the effects of exposure to each air pollutant on forced vital capacity at the current addresses as described elsewhere.¹³ We specified our confounder models for each pollutant with adjustments for individual level variables: sex, ln(age), ln(weight), ln(height), ethnicity; parental allergies; parental education; breastfeeding; maternal smoking during pregnancy; smoking, mold/dampness, and furry pets in the child's home; and recent respiratory infections.¹⁴ The health effect estimates from the N land-use regression evaluation models with N-1 sites in each training set were compared to the health effect estimates from the default land-use regression model with N sites. Finally, we investigated for each pollutant the association between the magnitude of the estimated effect on forced vital capacity and the predictive performances of the default land-use regression models (using leave-one-out cross-validation, holdout validation and cross-holdout validation R^2). We hypothesized that effect estimates would be larger when using exposure models with better predictive performance. Effect estimates were presented as the percent-change in forced vital capacity, with 95% CIs, for a given increase in exposure ($10\mu\text{g}/\text{m}^3$ for NO_2 , $5\mu\text{g}/\text{m}^3$ for $\text{PM}_{2.5}$ and $1\ 10^{-5}/\text{m}$ for $\text{PM}_{2.5}$ absorbance). All analyses have been done with SAS 9.3.

Results

Model evaluation

Figure 2 shows the performances of the default models based on the 10 training datasets and the evaluation models that consecutively excluded one site from the default model. We found generally identical mean values between the holdout validation R^2 and the cross-holdout validation R^2 which combines cross and hold-out validation as explained in the methods section. The mean leave-one-out cross-validation R^2 s ($R^2:0.59-0.91$) were higher than the

mean cross-holdout validation R^2 s (R^2 :0.45-0.75) (~12%) for all the pollutants and were slightly lower (~4%) than the mean model R^2 s (R^2 : 0.69-0.93). In contrast, the magnitude of the cross-holdout validation R^2 s was very similar to those of the holdout validation R^2 (0.52-0.79) and the intermediate holdout validation R^2 (0.55-0.78). The holdout validation R^2 from models based on N or N-1 observations were very similar. Model R^2 s and the leave-one-out cross-validation R^2 s were both high for all the pollutants, but overestimated the true predictive ability of the models at the external locations as indicated by holdout validation R^2 s of the same pollutants. The variability of the cross-holdout validation R^2 is higher than the holdout validation R^2 across models, which is more apparent for $PM_{2.5}$ than for the NO_2 and $PM_{2.5}$ absorbance. Traffic and population variables were dominant in all model structures for all pollutants. Variables in the $PM_{2.5}$ models were more diverse than the variables in the NO_2 and $PM_{2.5}$ absorbance models (eFigure 1).

Variance in health effect estimates

The present analysis included 1036 participants from this cohort with successful lung function measurements at age 8; complete information on sex, age, height, and weight at the time of lung function measurement (Table 1). Mean forced vital capacity is 2.0 ± 0.30 L.

Figure 3 shows the correlations between the performance factors (i.e. leave-one-out cross-validation, cross-holdout validation and holdout validation R^2 s) for NO_2 , $PM_{2.5}$ and $PM_{2.5}$ absorbance and the forced vital capacity effect estimates of the default models in the ten simulations. We found negative correlations between holdout validation R^2 s and forced vital capacity effect estimates for all three pollutants (Pearson R: -0.58~-0.79), indicating larger (more negative) effect estimates when holdout validation R^2 s were larger. Correlations with forced vital capacity effect estimates were slightly weaker for cross-holdout validation R^2 (Pearson R: -0.57~-0.58) than for holdout validation R^2 s. In contrast, the correlations of the leave-one-out cross-validation R^2 s of the NO_2 , $PM_{2.5}$ and $PM_{2.5}$ absorbance models with the forced vital capacity effect estimates were much weaker (R: -0.23~-0.35) than the correlations of the holdout validation R^2 s or cross-holdout validation R^2 s with the forced vital capacity effect estimates.

In order to investigate the stability of the effect estimates of forced vital capacity due to selection of a specific land-use regression model, we plotted all 400 (NO_2) or 200 ($PM_{2.5}$ and $PM_{2.5}$ absorbance) effect estimates of evaluation models in eFigures 2-4. Effect estimates were generally robust for NO_2 and $PM_{2.5}$ absorbance but more variable for $PM_{2.5}$. This is in line with the results shown in Figure 3. Moreover, exposure estimates were generally similar for the evaluation models for NO_2 and $PM_{2.5}$ absorbance but differed significantly among the evaluation models for $PM_{2.5}$ (eTable 2). Correlations between model predictions for $PM_{2.5}$ (Pearson R median: 0.75, range: -0.27 to 0.99) were lower than those for $PM_{2.5}$ absorbance (median:0.94, range: 0.36 to 0.99) and NO_2 (median: 0.98, range: 0.58 to 0.99) (eTable 2).

Discussion

Our study showed that the cross-holdout validation evaluation approach with variable selection produced results equivalent to using a holdout validation approach that reflects the

predictive ability of the land-use regression models we developed. Forced vital capacity effect estimates were fairly robust with different model specifications and decreases in forced vital capacity in relation to air pollution exposure were larger when using land-use regression models for exposure assignment, which had larger holdout validation R^2 and cross-holdout validation R^2 for all the pollutants.

Model evaluation

Our results support findings from the statistics literature showing that, with a multistep modeling procedure, cross-validation must be applied to the entire sequence of modeling steps.⁴ The cross-holdout validation approach makes use of the principles of leave-one-out cross-validation and holdout validation and has advantages especially when applying to land-use regression models based on small numbers of training sites. Compared to holdout validation, which requires splitting data into training and test datasets, the cross-holdout validation approach may be more useful in practice, especially for the model with small number of sampling sites, as it allows using all the available data. The leave-one-out cross-validation method, as demonstrated before, results in overestimation of the predictive ability of models, especially when the number of training sites is small.^{2,3} The principal difference between the leave-one-out cross-validation and the cross-holdout validation approach is that we do not develop a single model using all sites which is then refitted N times to $N-1$ sites (leave-one-out cross-validation). Instead, N models are developed which will be different in structure, not just coefficients. Although cross-holdout validation is based on the performance of N different models and not of a single one, cross-holdout validation provides a good estimation of the predictive ability of the default model, hence may be a useful validation approach for small datasets. We cannot choose from these N models which one(s) to use for exposure assignment in an epidemiologic study. The consequence would be that instead of one effect estimate, we need to produce N effect estimates. While this implies a computational burden, the resulting distribution of effect estimates may provide a better picture of the exposure model-related uncertainty than the 95% confidence interval of a single effect estimate based on a single exposure model. Within European Study of Cohorts for Air Pollution Effects, we have generally not found that the model leave-one-out cross-validation R^2 s found in different study areas were related to the magnitude of the effect estimates.^{14,17,18} It would be of interest to investigate whether this would change when calculating the cross-holdout validation R^2 s in the different study regions. Based on our results obtained with a single effect estimate in a single cohort, we hypothesize that there could also be an association between land-use regression model cross-holdout validation R^2 s and the magnitude of effect estimates across endpoints and cohorts.

Variance in health effect estimates

We observed larger decreases in forced vital capacity in relation to NO_2 , $\text{PM}_{2.5}$, and $\text{PM}_{2.5}$ absorbance for models with higher holdout validation R^2 values. This could be explained by underestimation of the 'true' health effect due to larger exposure misclassification when the models exhibited poorer predictions at the cohort addresses, i.e. lower model holdout validation R^2 s. This is what one would expect to observe when exposure misclassification is primarily classical, not Berkson.¹⁹ The classical-type misclassification is the one that can introduce bias in the health-effect estimate. The magnitude of the cross-holdout validation

R^2 shows a similar tendency. However, better leave-one-out cross-validation R^2 does not indicate a clear association with health effect estimates and thus indicates again that cross-holdout validation may be a better approach than leave-one-out cross-validation when applying to health analysis. Given the use of cross-validation in many papers of land-use regression models of air pollution, our study provides an additional note of caution, together with previous statistical literature, for using cross-validation in epidemiologic studies.

We found small variability of the effect estimates for forced vital capacity among the evaluation models for NO_2 (400 models based on 39 sites) and $\text{PM}_{2.5}$ absorbance (200 models based on 19 sites). The estimated changes in forced vital capacity in the default land-use regression model (40 sites for NO_2 and 20 sites for $\text{PM}_{2.5}$ absorbance) agreed well with the majority of those in the evaluation. This suggests that health effect estimates are robust regardless of selection of a specific land-use regression model. In contrast, we observed a relatively larger variation of forced vital capacity effect estimates between the different $\text{PM}_{2.5}$ evaluation models. This could be explained by different exposure estimates due to substantial differences between the model structures. In our study, $\text{PM}_{2.5}$ estimates differed among the evaluation models and the correlations between model predictions for $\text{PM}_{2.5}$ were lower than those for $\text{PM}_{2.5}$ absorbance and NO_2 (eTable 2). Moreover, traffic and population variables dominated the NO_2 and $\text{PM}_{2.5}$ absorbance models while $\text{PM}_{2.5}$ models also frequently included other variables such as natural/urban green, industrial, and port areas. In the previous study, Gehring et al. (2013)¹⁴ reported a robust association between forced vital capacity and $\text{PM}_{2.5}$ using data from the same population. The magnitude of the health effect estimate (i.e. forced vital capacity) was higher than the average value in our study. The discrepancy with Gehring et al. (2013)¹⁴ could be attributed to larger number of sampling sites ($N=40$) as input in the land-use regression model compared with $N=20$ in our exercises. As indicated in Basagana et al. (2013)⁹, health effect estimates tend to show smaller variance and less bias when the number of sampling sites was increased for land-use regression model.

Even though our study reveals variability of health effect estimate due to selection of exposure models, we did not attempt to correct for measurement error in exposure-health models. However, this could be done in several ways^{20,21}, including a two-stage correction approach using predicted exposures in the exposure-health models with bootstrap resampling to correct for bias and uncertainty²⁰. Moreover, it is also worth noting that better prediction accuracy does not always lead to better health effect inference.¹⁹

Conclusion

Cross-holdout validation reflects the true predictive ability of land-use regression models and may be a useful validation approach for small datasets. Land-use regression predictive ability indicated by holdout validation and cross-holdout validation R^2 s rather than leave-one-out cross-validation R^2 was associated with the magnitude of health effect estimates in a case study.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding sources:

The research leading to these results has received funding from the European Community's Seventh Framework Program (FP7/2007-2011): ESCAPE (grant agreement number: 211250). The PIAMA study is supported by The Netherlands Organization for Health Research and Development; The Netherlands Organization for Scientific Research; The Netherlands Asthma Fund; The Netherlands Ministry of Spatial Planning, Housing, and the Environment; and The Netherlands Ministry of Health, Welfare, and Sport. Additional supports were provided by U.S. EPA grant RD-83479601 and National Institute Health grants R01-ES009411 and R01-ES020871.

References

1. Hoek G, Krishnan RM, Beelen R, et al. Long-term air pollution exposure and cardio- respiratory mortality: a review. *Environ Health*. 2013; 12(1):43. [PubMed: 23714370]
2. Basagana X, Rivera M, Aguilera I, et al. Effect of the number of measurement sites on land use regression models in estimating local air pollution. *Atmos Environ*. 2012; 54:634–642.
3. Wang M, Beelen R, Eeftens M, et al. Systematic Evaluation of Land Use Regression Models for NO₂. *Environ Sci Technol*. 2012; 46(8):4481–4489. [PubMed: 22435498]
4. Hastie, T.; Tibshirani, R.; Friedman, J. Cross-Validation. 2. Springer-Verlag; 2009. The Elements of Statistical Learning; p. 241-247.
5. Beckerman BS, Jerrett M, Serre M, et al. A Hybrid Approach to Estimating National Scale Spatiotemporal Variability of PM_{2.5} in the Contiguous United States. *Environ Sci Technol*. 2013; 47(13):7233–7241. [PubMed: 23701364]
6. Bergen S, Sheppard L, Sampson PD, et al. A National Prediction Model for PM Component Exposures and Measurement Error-Corrected Health Effect Inference. *Environ Health Perspect*. 2013; 121(9):1017–1025. [PubMed: 23757600]
7. Vienneau D, de Hoogh K, Bechle MJ, et al. Western European land use regression incorporating satellite- and ground-based measurements of NO₂ and PM₁₀. *Environ Sci Technol*. 2013; 47(23): 13555–13564. [PubMed: 24156783]
8. Wang M, Beelen R, Bellander T, et al. Performance of Multi-City Land Use Regression Models for Nitrogen Dioxide and Fine Particles. *Environ Health Perspect*. 2014; 122(8):843–849. [PubMed: 24787034]
9. Basagana X, Aguilera I, Rivera M, et al. Measurement error in epidemiologic studies of air pollution based on land-use regression models. *Am J Epidemiol*. 2013; 178(8):1342–1346. [PubMed: 24105967]
10. Beelen R, Hoek G, Vienneau D, et al. Development of NO₂ 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.
11. Cyrus J, Eeftens M, Heinrich J, et al. Variation of NO₂ and NO_x concentrations between and within 36 European study areas: Results from the ESCAPE study. *Atmos Environ*. 2012; 62:374–390.
12. Eeftens M, Tsai M, Ampe C, et al. Spatial variation of PM_{2.5}, PM₁₀, PM_{2.5} absorbance and PM_{coarse} concentrations between and within 20 European study areas and the relationship with NO₂ - Results of the ESCAPE project. *Atmos Environ*. 2012; 62:303–317.
13. Eeftens M, Beelen R, de Hoogh K, et al. Development of Land Use Regression Models for PM_{2.5}, PM_{2.5} Absorbance, PM₁₀ and PM_{coarse} in 20 European Study Areas; Results of the ESCAPE Project. *Environ Sci Technol*. 2012; 46(20):11195–11205. [PubMed: 22963366]
14. Gehring U, Gruzdeva O, Agius RM, et al. Air pollution exposure and lung function in children: the ESCAPE project. *Environ Health Perspect*. 2013; 121(11-12):1357–1364. [PubMed: 24076757]

15. Brunekreef B, Smit J, de Jongste J, et al. The prevention and incidence of asthma and mite allergy (PIAMA) birth cohort study: design and first results. *Pediatr Allergy Immunol*. 2002; 15(13 Suppl):55–60.
16. Wijga AH, Kerkhof M, Gehring U, et al. Cohort profile: The Prevention and Incidence of Asthma and Mite Allergy (PIAMA) birth cohort. *Int J Epidemiol*. 2013; 43(2):527–35. [PubMed: 23315435]
17. Beelen R, Raaschou-Nielsen O, Stafoggia M, et al. Effects of long-term exposure to air pollution on natural-cause mortality: an analysis of 22 European cohorts within the multicentre ESCAPE project. *Lancet*. 2014; 383(9919):785–795. [PubMed: 24332274]
18. Raaschou-Nielsen O, Andersen ZJ, Beelen R, et al. Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE). *Lancet Oncol*. 2013; 14(9):813–822. [PubMed: 23849838]
19. Szpiro AA, Paciorek CJ, Sheppard L. Does more accurate exposure prediction necessarily improve health effect estimates? *Epidemiology*. 2011; 22(5):680–685. [PubMed: 21716114]
20. Szpiro AA, Paciorek CJ. Measurement error in two-stage analyses, with application to air pollution epidemiology. *Environmetrics*. 2013; 24(8):501–517. [PubMed: 24764691]
21. Gryparis A, Paciorek CJ. Measurement error caused by spatial misalignment in environmental epidemiology. *Biostatistics*. 2009; 10(2):258–274. [PubMed: 18927119]

Figure 1.

A flowchart illustrating the main evaluation procedures and terms used in this study.

Note: all the LOOCV and CHV are done in training sets, and HV is done in test sets

Figure 2.

Model performances (mean $\pm$ standard deviation) of NO_2 , $\text{PM}_{2.5}$ and $\text{PM}_{2.5}$ absorbance for the default models based on N sites (PM: N=20; NO_2 : N=40; black dots: Model R^2 and HV R^2) and for the evaluation models based on N-1 sites (grey dots: Model R^2 , LOOCV R^2 , CHV R^2 and HV_{intermediate} R^2) measured from the European Study of Cohorts for Air Pollution Effects in the Netherlands. LOOCV: leave-one-out cross-validation; HV: hold-out validation; CHV: cross-hold-out validation.

Figure 3.

Correlations (Pearson correlation coefficients) between estimated changes in forced vital capacity (%) per $10\mu\text{g}/\text{m}^3$ for NO_2 , $5\mu\text{g}/\text{m}^3$ for $\text{PM}_{2.5}$ and 1 unit for $\text{PM}_{2.5}$ ABS (absorbance) in exposure from the Dutch Prevention and Incidence of Asthma and Mite Allergy study and R^2 s of holdout validation (HV), cross-holdout validation (CHV) and leave-one-out cross-validation (LOOCV) for $\text{PM}_{2.5}$, $\text{PM}_{2.5}$ absorbance and NO_2 default models in ten times simulations. The vertical bars show the 95% confidence interval of the changes in forced vital capacity (%) assessed by each of the default exposure models.

Table 1

Description of the study population and lung function measurements (N = 1036)

Variable	N	Percent or Mean±SD
Female sex	1036	50
Respiratory infections	1032	24
Allergic mother	1036	66
Allergic father	1033	33
Dutch ethnicity ^a	1022	96
High maternal SES ^b	1033	39
High paternal SES ^b	1021	44
Breastfeeding	1036	53
Mother smoked during pregnancy	1022	15
Smoking at child's home ^c	968	16
Mold/dampness in child's home ^c	963	29
Furry pets in home ^c	948	50
Height (cm)	1036	132.9±5.60
Weight (kg)	1036	28.9±4.80
Age (years)	1036	8.1±0.30

^a Ethnicity: Dutch;^b SES: Socioeconomic status;^c At the age of the lung function measurement.