Biostatistics (2019) **20**, 1, *pp.* 1–16
doi:10.1093/biostatistics/kxx058
Advance Access publication on November 9, 2017

Using ecological propensity score to adjust for missing confounders in small area studies

YINGBO WANG

Novartis Pharma AG, WSJ-103.1, Novartis Campus, Basel 4002, Switzerland

MONICA PIRANI, ANNA L. HANSELL

MRC-PHE Centre for Environment and Health, Imperial College London, St Mary's Campus, Norfolk Place, London W2 1PG, UK

SYLVIA RICHARDSON

MRC Biostatistics Unit, Cambridge Institute of Public Health, Forvie Site, Robinson Way, Cambridge Biomedical Campus, Cambridge CB2 0SR, UK

MARTA BLANGIARDO*

MRC-PHE Centre for Environment and Health, Imperial College London, St Mary's Campus, Norfolk Place, London W2 1PG, UK

m.blangiardo@imperial.ac.uk

SUMMARY

Small area ecological studies are commonly used in epidemiology to assess the impact of area level risk factors on health outcomes when data are only available in an aggregated form. However, the resulting estimates are often biased due to unmeasured confounders, which typically are not available from the standard administrative registries used for these studies. Extra information on confounders can be provided through external data sets such as surveys or cohorts, where the data are available at the individual level rather than at the area level; however, such data typically lack the geographical coverage of administrative registries. We develop a framework of analysis which combines ecological and individual level data from different sources to provide an adjusted estimate of area level risk factors which is less biased. Our method (i) summarizes all available individual level confounders into an area level scalar variable, which we call *ecological propensity score* (EPS), (ii) implements a hierarchical structured approach to impute the values of EPS whenever they are missing, and (iii) includes the estimated and imputed EPS into the ecological regression linking the risk factors to the health outcome. Through a simulation study, we show that integrating individual level data into small area analyses via EPS is a promising method to reduce the bias intrinsic in ecological studies due to unmeasured confounders; we also apply the method to a real case study to evaluate the effect of air pollution on coronary heart disease hospital admissions in Greater London.

Keywords: Environmental epidemiology; Hierarchical model; Missing data; Observational study; Propensity score; Spatial statistics.

*To whom correspondence should be addressed.

© The Author 2017. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

Small area studies are commonly used in epidemiology to investigate the spatial variation of a health condition across the population or to evaluate the geographic patterns of diseases in relation to environmental, demographic, and socio-economic factors.

These studies are based on administrative registries, which are characterized by good spatial coverage for large populations, but usually only record a very limited set of information (typically age, sex, and spatial location) and thus miss important confounders (e.g. data on lung cancer or respiratory diseases from Hospital Episode Statistics (HES) databases do not include information about smoking), potentially leading to biased estimates of the effects of risk factors (e.g. air pollution).

In this article, we integrate data from administrative registries with cohorts/surveys, which contain detailed information on participants. Through this, the inferences drawn at the ecological level will take advantage of the population representativeness and of the statistical power from the administrative registries, but at the same time will allow to adjust the effect estimates for all the potential confounders available from the cohorts/surveys.

Registries contain data on each individual in the target population, while cohorts/surveys typically cover only a subset of individuals; for this reason confounders obtained from the latter source will be typically partially measured and available only on a subset of areas, leading to a missing data issue which needs to be tackled. Multiple imputation (MI), pioneered by Rubin (1987) is probably the most common strategy to deal with this issue and consists of a Monte Carlo simulation to replace the missing values with a relatively small number of simulated versions. The two main approaches to implement MI are joint modeling and chained equations. As Hughes *and others* (2014) showed that the difference in the results from the two approaches are negligible, chained equations are preferred as they are based on conditional distributions, enabling fast computations. As an example, the Multiple Imputation using Chained Equation (MICE) proposed by Buuren and Oudshoorn (1999) fixes initial values for all the missing variables and regresses each of these against the remaining ones cyclically a number of times (see White *and others* 2011 for a thorough review of the method). Then, in a separate step, the imputed variables are included in the substantive model, which evaluates the link between the exposure/risk factor and health outcome. Such approach can suffer from lack of "congeniality": Meng (1994) states that for a model to be congenial the imputation model needs to include the same variables (including the dependent variable) as those of the substantive model, in order to avoid estimates biased toward the null; this might be non-trivial for non-linear relationships between the outcome and the exposure/risk factors, which is the case when the outcome is available in the form of aggregated counts as in small area studies. Furthermore, such method is not suited to add spatial dependency in the imputation model, which may lead to bias if the missing values are geographically related, as it is often the case in a small area study.

In a Bayesian perspective, missing data imputation is generally considered via the integration of the imputation and the analysis models in a coherent global framework (see for instance Molitor *and others* 2009, Daniels and Hogan 2008). This approach benefits from extreme flexibility, but entails heavy computational burden if there are more than a few missing covariates, *de facto* requiring oversimplifications of the epidemiological problem, as in reality the number of potential partially measured confounders is typically large.

To avoid high-dimensional imputation, a viable alternative summarizes the partially measured confounders from the cohorts/surveys through a composite index, so that only one variable needs to be imputed. In this perspective, the propensity score (PS) has been suggested, in the form of a calibration model proposed by Sturmer *and others* (2005) as well as in a Bayesian framework developed by McCandless *and others* (2012). The first approach estimates a gold standard PS on the units with information on the partially measured confounders, while an "error-prone" PS is estimated on all the units. Then Sturmer *and others* (2005) specify a regression model to estimate the relationship between the gold standard PS and

the error-prone PS in order to predict the gold standard where missing. Note that the proposed calibration model assumes a linear relationship between the gold standard PS and the error-prone PS; moreover, the outcome variable is not included in the regression (in contrast to what is recommended in the missing data literature). The second method is a Bayesian Propensity Score formulated to include information from a cohort with fully observed confounders in an individual level study based on administrative data. The PS is built on the cohort data to summarize the confounders. On the individuals who do not have information on the confounders, the approach then imputes the PS from its empirical distribution. This strategy can be applied effectively regardless of the number of unmeasured confounders.

In this article, we develop a novel Bayesian modeling framework to fit the PS on small area studies and ecological covariates, which we call *ecological propensity score* (EPS). In particular, (i) we propose an imputation model for areas with a missing EPS; this accounts for the spatial structure of the data and can easily accommodate non-linearity in the relationship with other variables; (ii) we include the estimated/imputed EPS in the analysis model in a flexible way to provide effective confounder adjustment when evaluating the effect of a risk factor on a health outcome; (iii) we discuss and account for the different sources of feedback across the overall modeling framework.

The structure of this article is as follows: Section 2 introduces our proposed EPS framework for small area studies, and Section 3 presents an extensive simulation study to evaluate the performance of the developed approach and to compare it with the commonly used MICE for imputing missing data; Section 4 applies this method to assess the link between airborne particle pollution and coronary heart disease CHD hospital admissions in Greater London; finally we include a discussion and concluding remarks.

2. The ecological propensity score (EPS)

Before starting with the description of the EPS framework, we set the notation used throughout the paper: (i) $i \in S$ identifies the set of areas with coverage from the survey or cohort; (ii) $i \in I$ identifies the set of areas without coverage from the survey or cohort; (iii) O_i, E_i are the number of health outcome occurrences, which are observed and expected (using standardized rates); they are available on all the areas ($i \in I \cup S$); d) X_i is the exposure or risk factor of interest observed on all the areas ($i \in I \cup S$); e) $\boldsymbol{C}_i$ is the set of P area level confounders observed on all the areas ($i \in I \cup S$); f) $\boldsymbol{M}_i$ is the set of Q area level confounders which are unobserved on all the areas ($i \in I \cup S$), but for which individual level data $\boldsymbol{m}_{qi} = \{m_{qi1}, \ldots, m_{qin}\}$ are available on $i \in S$ from the survey or cohort.

To develop the modeling framework, we consider three steps: (i) **EPS estimation**—in the areas with available individual level information on the full set of confounders ($i \in S$) the information is aggregated through small area estimation models and the EPS is estimated using these ecological confounders. (ii) **EPS imputation**—in the areas without survey coverage ($i \in I$) a Bayesian imputation model is set up to impute EPS based (1) on the relationship between the estimated EPS and other variables in the S survey areas and (2) on the spatial structure of the data. (iii) **EPS adjustment**—the EPS (estimated or imputed) for all the areas is then included in the analysis model providing confounder adjustment to the ecological relationship of interest between X and O.

As our approach is formulated in a Bayesian framework a joint model could be specified for the three components; however, when using PS approaches to adjust for confounder effects, may be necessary to prevent feedback in estimating some parameters to avoid biases (Zigler *and others*, 2013). In the remainder of this section, we describe the model in details and discuss the issues around feedback.

2.1. *EPS estimate*

The first step of the EPS framework consists in using the individual level confounding information to estimate corresponding (latent) area level confounders on the survey areas; as the survey samples are

typically small and at the same time they are often spatially distributed, we carry out this estimation by using a hierarchical spatial model for all the confounders allowing for each survey area to be embedded in a larger spatial unit d_i, e.g. used for the survey stratification, to ensure complete spatial coverage. From the survey (or cohort), n values $\{m_{qi1}, \ldots, m_{qin}\}$ are observed on the i-th area ($i \in S$), for the q-th confounder variable; such variables are typically dichotomous (e.g. smoking status, so that $m_{qij} \sim \text{Bernoulli}(M_{qi})$), discrete (e.g. number of cigarettes smoked, so that $m_{qij} \sim \text{Poisson}(M_{qi})$), or continuous (e.g. BMI, so that $m_{qij} \sim \text{Normal}(M_{qi}, \sigma_q^2)$), hence to estimate M_{qi}, we specify a linear model through the appropriate link function (logit, log, identity) as follows:

$$g(M_{qi}) = \alpha_q + \psi_{1qd_i} + \psi_{2qd_i} + v_{qi}, \tag{2.1}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_Q)$ is the intercept for each confounder, and $\boldsymbol{\psi}_{qd_i}$ model the spatially structured and unstructured variability for the larger spatial units ($d = 1, \ldots, D$); a multivariate version of the BYM prior (MVBYM) proposed by Besag *and others* (1991) is specified to account for spatial dependency for each confounder as well as for correlation between these. An intrinsic autocorrelation structure is assumed on ψ_{1qd} governed by the variance $\sigma^2_{\psi_{1q}}$ (Besag and Kooperberg, 1995), while each ψ_{2qd} follows a normal distribution with a common variance $\sigma^2_{\psi_{2q}}$. For the qth confounder, a local smoothing is provided by ψ_{1qd} based on the values of the set of neighbors $\mathcal{N}(d)$, and a global smoothing is included through ψ_{2qd} based on the values of all the units. In addition, a structure of the correlation between the confounders should be included to allow borrowing of strength, as some are observed on more individuals than others (e.g. some confounders might have been collected for several years while others might not). Thus, we extend $\sigma^2_{\psi_{1q}}$ to the multivariate Σ_{ψ_1} (where the diagonal of Σ_{ψ_1} is $\sigma^2_{\psi_{1q},\ q=1,\ldots,Q}$ and the off diagonal is the covariance, $\rho_{\psi'_{1qq}}\sigma_{\psi_{1q}}\sigma_{\psi_{1q'},\ \ q'\neq q}$); similarly the between confounder non-spatial correlation $\rho_{\psi'_{2qq}}$ is taken into account by replacing $\sigma^2_{\psi_{2q}}$ with Σ_{ψ_2}. This specification leads to the multivariate BYM model (MVBYM). See Gamerman *and others* (2003) for details and applications of MVBYM (see also Section 1 of supplementary material available at *Biostatistics* online for more details on the MVBYM prior). In addition, $v_{qi} \sim \text{Normal}(0, \omega_v^2)$ captures the residual over-dispersion at the inference unit level.
EPS is then derived following the specification of McCandless *and others* (2012):

$$\text{logit}(P(X_i = 1|\boldsymbol{C}_i, \boldsymbol{M}_i)) = \theta_1 + \boldsymbol{C}_i^T\boldsymbol{\theta}_C + \boldsymbol{M}_i^T\boldsymbol{\theta}_M; \qquad \text{EPS}_i = \boldsymbol{M}_i^T\boldsymbol{\theta}_M, \tag{2.2}$$

where X needs to be dichotomized and $\boldsymbol{\theta}_C = (\theta_{C_1}, \ldots, \theta_{C_P})$, $\boldsymbol{\theta}_M = (\theta_{M_1}, \ldots, \theta_{M_Q})$, while $\boldsymbol{M}_i$ are the area level estimates for the Q confounders obtained in (2.1). Note that through this EPS specification we summarize only the Q confounders that can be estimated in the survey areas S. This is a crucial point as the approach allows the imputation of the EPS on the I areas where the Q confounders are missing (see next section for the details), while at the same time the P confounders available on all the areas are included directly in the analysis. We expand on this point in Section 5.

Here feedback from X to $\boldsymbol{M}$ in (2.2) needs to be cut as it would distort the estimates of $\boldsymbol{M}$ in (2.1). Nevertheless uncertainty on $\boldsymbol{M}$ is fed forward from (2.1) to the EPS estimate in (2.2), which thus has a mixing distribution over the uncertainty in $\boldsymbol{M}$. A summary of that mixing distribution will then be used in the imputation model as detailed below.

2.2. *EPS imputation*

The EPS estimation presented in (2.2) can be used only on the survey areas; when incomplete areas are present in the study region an additional step is needed to impute EPS due to the fact that the individual

level covariates and consequently $\boldsymbol{M}_i$ are missing. Thus for each area i ($i \in S \cup I$) we specify:

$$\text{EPS}_i \sim \text{Normal}(\eta_1 + f(\boldsymbol{C}_i) + \gamma X_i + \phi_{d_i},\ \ \sigma^2_{\text{EPS}}), \tag{2.3}$$

where EPS_i is the posterior mean from the estimation in (2.2) for $i \in S$ and is missing for $i \in I$. In (2.3), the link between X and $\boldsymbol{M}$ as written in (2.2) changes and now specifies a relationship between the posterior mean for EPS and (i) the dichotomous X and (ii) the confounders $\boldsymbol{C}$ through the function $f(.)$. This function should meet three criteria: (i) be data-driven, as (2.3) does not have an epidemiological interpretation, (ii) be able to cope with a high dimensional $\boldsymbol{C} = (\boldsymbol{C}_1, \ldots, \boldsymbol{C}_P)$, c) be easy to scale up to over thousands of observational units if needed. In order to satisfy these three requirements, we choose a second-order random walk (RW(2)) as functional form for $f()$ to link each $\boldsymbol{C}$ to (2.3), which was initially proposed by Fahrmeir and Lang (2001) to adjust for non-linearity. The specification of the RW2 is presented in Section 2 of supplementary material available at *Biostatistics* online.

An additional random effect $\boldsymbol{\phi}$ is included to account for residual variation using the same spatial resolution as in (2.1) to ensure full spatial coverage. As it is likely that these residuals will exhibit a spatial structure, a conditional autoregressive model is specified on $\boldsymbol{\phi}$ based on neighborhood similarities, through the univariate formulation of the BYM included in (2.1).

The EPS_i for $i \in I$ is predicted using the relationship with $\boldsymbol{C}$, X, and the spatial structure estimated on $i \in S$.

2.3. *EPS adjustment*

The analysis model is specified in a small area framework to evaluate the direct (area level) effect of a risk factor (or exposure) X on a health end point O after adjustment for observed confounders $\boldsymbol{C}$ and unobserved confounders $\boldsymbol{M}$ using EPS. In its typical formulation, for each area ($i \in S \cup I$) the number of cases of the health outcome O_i is modeled as follows:

$$\begin{aligned} O_i &\sim \text{Poisson}(E_i \lambda_i) \\ \log(\lambda_i) &= \beta_1 + \beta_2 X_i + \boldsymbol{C}_i^T \boldsymbol{\beta}_C + h(\text{EPS}_i) + \xi_{1i} + \xi_{2i}, \end{aligned} \tag{2.4}$$

where E_i represents the expected number of cases obtained from standardized rates. The parameter ξ_{1i} accounts for over-dispersion, while ξ_{2i} accounts for spatially structured variation and is typically modeled through a conditional autoregressive structure similarly to (2.1) and (2.3).

The EPS_i feeds into (2.4) as:

- $\boldsymbol{M}_i \boldsymbol{\theta}_M$ estimated for $i \in S$ from (2.2);
- the posterior predictive distribution from (2.3) for $i \in I$.

As EPS is an index, it is not interpretable *per se*, thus it needs to be included in the analysis model through a non-linear flexible function $h(\cdot)$; similarly to the imputation model specification, we assign a RW(2) to link the EPS into the analysis model.

Note that (2.3) and (2.4) are jointly estimated, so that the outcome is allowed to influence the missing EPS imputation via its feedback, as recommended in the missing data literature (see for instance Kenward and Carpenter, 2007); at the same time uncertainty from the EPS imputation is carried forward into the analysis model.

A graphical representation of the EPS framework including where feedback is allowed/cut is visualized in Figure 1.

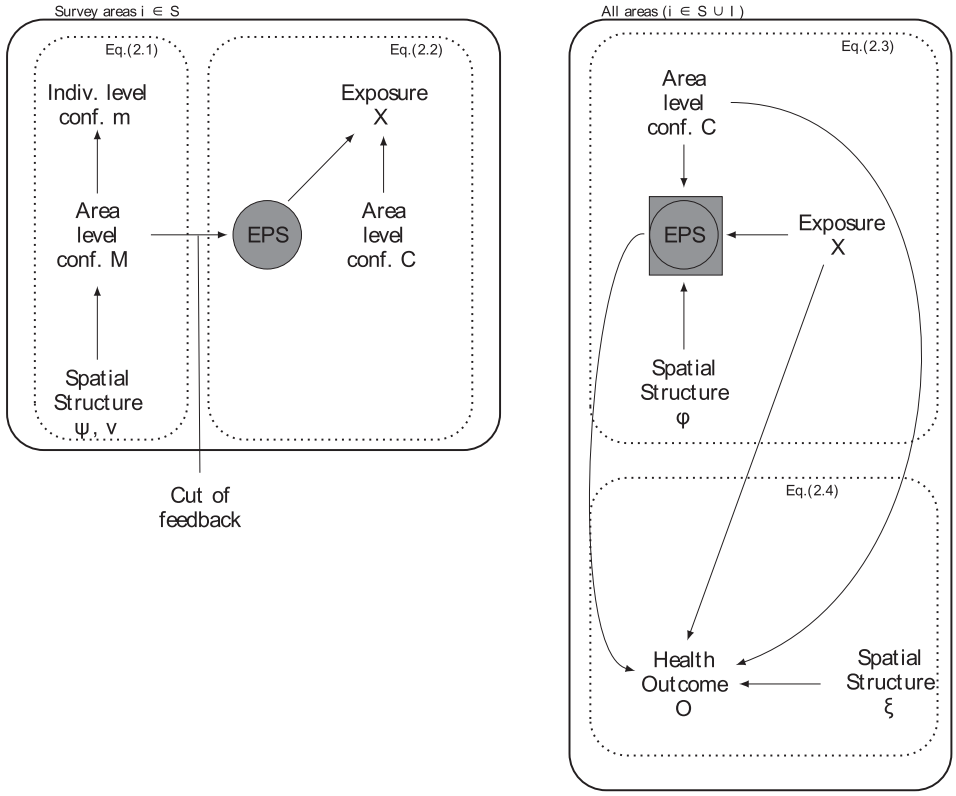

Fig. 1. The figure represents the developed EPS framework. The left hand side corresponds to (2.1) and (2.2): the area level confounders $\boldsymbol{M}$ are estimated from the individual level confounders $\boldsymbol{m}$ (2.1) and the EPS is estimated (2.2). Note that this model is only specified on the $i \in S$ areas. EPS is represented by a circle as a latent quantity (not observed). The right hand side presents the EPS imputation and adjustment and is specified on $i \in I \cup S$: at the top the EPS is either obtained from (2.2) for $i \in S$ and included as an observed quantity or missing for $i \in I$, imputed through the relationship with $\boldsymbol{C}$, X, and the spatial structure (2.3) and included as a latent quantity. Thus the EPS is represented here as both a square and a circle as the combination of observed and latent quantities. At the bottom, the estimated and imputed EPS is included in the analysis model (corresponding to (2.4)) to provide confounder adjustment when assessing the effect of X on O. Note that the right hand side is separated from the left hand side as the EPS estimation is carried out disjointly from the imputation and adjustment (no feedback from the latter to the former). At the same time the imputation and adjustment are jointly carried out so that feedback is allowed from O to influence the EPS imputation.

2.4. *Specifying priors*

To complete the model specification, prior distributions need to be assigned. For MVBYM in Model 1, the priors for $\boldsymbol{\Sigma}_{\psi_1}^{-1}$ or $\boldsymbol{\Sigma}_{\psi_2}^{-1}$ are chosen to be Wishart$(\nu, \mathbf{W})$ centred around the empirical variance $\hat{\boldsymbol{\Sigma}}$ by specifying $\mathbf{W}_{\psi_1} = \nu_{\psi_1}\hat{\boldsymbol{\Sigma}}$ or $\mathbf{W}_{\psi_2} = \nu_{\psi_2}\hat{\boldsymbol{\Sigma}}$. In general, an empirical variance $\hat{\boldsymbol{\Sigma}}$ is considered to be an approximate estimator of the true variance $\mathrm{E}(\boldsymbol{\Sigma})$. This approximation is acceptable because the degree of freedom ν is typically chosen to be the minimum to ensure that the specified Wishart is a diffused prior, i.e. $\nu_{\psi_1} = \nu_{\psi_2} = Q$ where Q is the number of confounders as presented in Section 2.1 (Lunn *and others*, 2012).

The rest of priors are chosen to be minimally informative, i.e. the regression coefficients are modeled as Normal$(0, 10^5)$ and the standard deviation of the random effects are modeled as Uniform$(0, 10^3)$.

All the models were implemented in the BUGS language (Lunn *and others*, 2012).

3. Simulation

In this section, we present a simulation study to evaluate the performance of the EPS framework and to compare it with the commonly used MICE approach (Buuren and Oudshoorn 1999), which we implemented through the corresponding `miceR` package.

We consider three scenarios: (i) the missing confounders $\boldsymbol{M}_i = (M_{1i}, \ldots, M_{Qi})$ are available in all areas ($i \in S \cup I$), and this is the benchmark (Scenario 1); (ii) the missing confounders $\boldsymbol{M}_i$ are assumed to be available in some areas ($i \in S$, Scenario 2); (iii) the missing confounders $\boldsymbol{M}_i$ are not directly available, and these confounders might be estimated through individual level data from surveys/cohorts (only on the survey areas, $i \in S$). To assess the impact of the sample size of individual level data on the exposure estimation, Scenario 3 includes different numbers of individuals (5, 10, 20, 100) sampled from the survey/cohort in each area.

The variables are simulated in the following sequence on 300 areas: (i) two confounders $\boldsymbol{C}$ and four confounders $\boldsymbol{M}$ are generated from the inverse logit transformation of multivariate normal distributions, (ii) X is simulated based on $\boldsymbol{C}$ and $\boldsymbol{M}$ (using (2.2)), (iii) O is simulated from a Poisson with $E = 100$ (to mimic the real case we are illustrating in the next section), X and confounders $\boldsymbol{C}$ and $\boldsymbol{M}$, (iv) for each area n individual data $\boldsymbol{m}_{qi} = (m_{qi1}, \ldots, m_{qin})$ are simulated from a Bernoulli distribution based on the proportion $\boldsymbol{M}_{qi}$. To mimic the limited survey coverage, Missing At Random (MAR) criterion is applied to remove $\boldsymbol{M}$ (for Scenario 2) or $\mathbf{m}$ (for Scenario 3) from around 50% of the areas. The detailed simulation process is described in Section 3 of supplementary material available at *Biostatistics* online. This process is repeated 100 times to create 100 paired ecological level and individual level data sets.

As MICE cannot account for spatial dependency, we did not include spatial random effects in this simulation. Additionally at present MICE cannot be linked to a multilevel model to estimate $\boldsymbol{M}$ as in (2.1), thus we extract the posterior means of $\boldsymbol{M}$ provided by (2.1), and then plug these into MICE. The outcome is included into the imputation model by adding the Standardized Mortality Ratio (i.e. SMR $= O/E$) as an extra predictor. The missing $\boldsymbol{M}$ are imputed ten times and included in the analysis model to estimate the exposure effect parameter β_2 using the same formulation as (2.4).

Then the ten estimated β_2 are summarized through Rubin's combination rule (Rubin, 1987).

Bias, root mean squared error (RMSE), width and coverage of the 95% credibility interval (CI) are used to compare the performance of the simulation scenarios; the true β_2 is chosen to be 0.5, which corresponds to a 64.9% ($(\exp(0.5) - 1) \times 100$) increment in the health risk for high vs low exposure. The BUGS code and simulated data sets are provided at https://github.com/martablangiardo/Wangetal_Biostatistics.

Note that an additional simulation set to evaluate the performance of our model for spatially structured variables and non-linearity in the confounder–outcome relationship is included in Section 5 of supplementary material available at *Biostatistics* online.

3.1. *Results*

Table 1 presents the results of the simulation study when the true value of $\beta_2 = 0.5$. The benchmark (Scenario 1) assumes the availability of confounders $\boldsymbol{M}$ in all areas, which allows to evaluate the impact of using EPS as a summary index instead of including each confounder separately into the analysis model. The estimation from EPS adjustment is almost identical to the regression approach which directly includes $\boldsymbol{M}$ as covariates. The CI width of β_2 from the EPS adjustment model is slightly wider than that of the regression adjustment, and this agrees with Senn *and others* (2007) which showed that the estimation from PS stratification (very close to PS adjustment) is more conservative than that from the direct regression analysis.

Ignoring the information from the confounders $\boldsymbol{M}$ allows us to evaluate what size of bias we are potentially dealing with (naïve case). Since $\boldsymbol{M}$ is simulated to be a confounder for the relationship between

Table 1. *EPS performance and comparison with MICE on the simulation study (true $\beta_2 = 0.5$, 100 simulated data sets)*

	Adjustment/ imputation	Posterior mean for β_2	Bias	RMSE	CI95 width	CI95 coverage
Scenario 1: **M** are available in all areas						
	Direct adj	0.50	0.00	0.02	0.054	0.90
	EPS adj	0.50	0.00	0.02	0.062	0.95
Naïve case: Ignoring **M**						
	NA	0.78	0.28	0.28	0.049	0.00
Scenario 2: **M** are only available in some areas						
Case 2.1: Analysis on $i \in S$	EPS adj	0.51	0.01	0.03	0.087	0.89
Case 2.2 : Analysis on	MICE	0.54	0.04	0.05	0.107	0.65
$i \in S \cup I$	EPS imput	0.50	0.00	0.03	0.084	0.90
Scenario 3: **M** are NOT directly available, but ***m*** are available in some areas						
extbfSample size $n = 5$						
Case 3.1.1: Analysis on $i \in S$	EPS adj	0.59	0.08	0.09	0.094	0.16
Case 3.2.1: Analysis on	MICE	0.62	0.12	0.14	0.132	0.20
$i \in S \cup I$	EPS imput	0.57	0.07	0.08	0.087	0.32
Sample size $n = 10$						
Case 3.1.2: Analysis on $i \in S$	EPS adj	0.59	0.09	0.10	0.094	0.47
Case 3.2.2: Analysis on	MICE	0.62	0.12	0.14	0.129	0.51
$i \in S \cup I$	EPS imput	0.57	0.07	0.08	0.086	0.60
Sample size $n = 20$						
Case 3.1.3: Analysis on $i \in S$	EPS adj	0.59	0.09	0.10	0.094	0.52
Case 3.2.3: Analysis on	MICE	0.60	0.10	0.11	0.128	0.53
$i \in S \cup I$	EPS imput	0.57	0.07	0.08	0.085	0.65
Sample size $n = 100$						
Case 3.1.4: Analysis on $i \in S$	EPS adj	0.52	0.02	0.04	0.088	0.78
Case 3.2.4: Analysis on	MICE	0.56	0.06	0.06	0.119	0.56
$i \in S \cup I$	EPS imput	0.51	0.01	0.03	0.083	0.85

outcome and exposure, ignoring it leads to a serious bias and RMSE in estimating β_2 (both equal to 0.28), as well as to a small uncertainty around it (CI width equal to 0.049, which is smaller than in any other case).

Scenario 2 and 3 assume the partial availability of $\boldsymbol{M}$ and two missing data approaches are adopted to include the areas with missing $\boldsymbol{M}$: MICE and EPS imputation. In Scenario 2, $\boldsymbol{M}$ is available directly at the ecological level, whereas Scenario 3 mimics the real situation where $\boldsymbol{M}$ is not available directly at the ecological level, but may be estimated through the individual level information.

We find that considering only the survey areas (which can be interpreted as a complete case analysis) provides slightly more biased results and at the same time greater uncertainty than when all the areas are considered and EPS is imputed whenever missing; this is expected as the sample size is smaller ($i \in S$ as opposed to $i \in S \cup I$). In addition using MICE leads to more biased results than using the survey areas only (e.g. bias ranging from 0.04 to 0.10) and is also characterized by a much larger uncertainty (CI width ranging from 0.107 to 0.132).

Furthermore, Scenario 3 allows us to assess the impact of individual level sample size (n = 5, 10, 20, 100) on the risk estimate for EPS imputation. As expected the bias decreases and the coverage increases as the number of sampled individuals increases, while the uncertainty tends to decrease slightly; the latter might seem counter-intuitive, but it can be explained by the fact that the number of incomplete areas is

kept the same across cases 3.2.1–3.2.4 in the table, to be able to clearly evaluate the gain in accuracy and precision when more individuals are surveyed. In the next section, we will show how considering a larger number of surveyed individuals leads to an increase in the proportion of incomplete areas and how this impacts on the β_2 point estimates and uncertainty.

Table 1 in supplementary material available at *Biostatistics* online shows that the same conclusion can be drawn in the absence of an exposure effect ($\beta_2 = 0$).

4. Illustrative example: air pollution and health in Greater London

We apply the proposed methodology to investigate the link between air pollution exposure and CHD (a subset of cardiovascular diseases, CVD) hospital admissions in Greater London. Many studies have provided evidence that short-term exposure (hours or weeks) is associated with CVD incidence (COMEAP 2006) and large cohort studies in the US (Puett *and others*, 2009; Miller *and others*, 2007; Lipsett *and others*, 2011) and in the Netherlands (Hoek *and others*, 2002) have focused on the association of long-term exposure to outdoor particulate matter (PM, defined as a mixture of particles smaller than a specific size, e.g. PM_{10} includes particles smaller than 10 μm) and CVD related death. However the epidemiological evidence is more limited regarding the association between long-term exposure to air pollution and CVD morbidity, for instance non-fatal outcomes, such as hospital admissions due to CVD (COMEAP 2006).

We consider the HES, an administrative registry which provides information on admissions at population level and available to us through the Small Area Health Statistics Unit (SAHSU) at Imperial College London. The analysis is conducted for the year 2001 at the electoral ward level, with an average population of about 9600 persons. To identify the cases, we used the International Classification of Disease version 10 (ICD-10) codes I20 to I25. The number of observed cases in each ward is denoted as O_i (the total number of cases across Greater London was 52 358 in 2001). The total number of residents aged 16 and above in the same region and same year is 5 723 855.

As exposure X, we consider the annual mean level of PM_{10} in 2001 at the ward level published by Vienneau *and others* (2010) and obtained through a land used regression model combining monitoring data from the national air quality networks (UK) and additional predictors related to traffic, population, land use, and topography. As a binary exposure variable is needed to estimate the EPS as presented in (2.2), for the purpose of our paper we dichotomize X using the median of the PM_{10} concentration as cut-off (25μg/m^3).

The confounders $\boldsymbol{C}$ available for every ward at the ecological level are deprivation (measured through the Carstairs index, issued by the Census 2001) and ethnicity (defined as the proportion of non-white and issued by the Census). In addition, the age–sex stratified population count in 2001 census is used to calculate the expected number of hospital admissions E_i in each ward, and thus age and sex are included indirectly in the outcome-exposure analysis.

Following the standard approach in small area studies, we first run an ecological regression model which assumes a Poisson distribution on the number of hospitalizations, $O_i \sim \text{Poisson}(E_i\lambda_i)$, and a linear regression on λ_i:

$$\log(\lambda_i) = \beta_1 + \beta_2 X_i + \boldsymbol{C}_i^T \boldsymbol{\beta}_C + \xi_{1i} + \xi_{2i},$$

where X is the exposure, $\boldsymbol{C}$ are the confounders and ξ_{1i} and ξ_{2i} are spatially unstructured and structured random effects as described in (2.4). The results show a posterior mean estimate (95% CI) for $\exp(\beta_2)$ [relative risk (RR)of CHD hospital admissions for high vs low air pollution exposure] equal to 0.89 (0.84 − 0.94). This result is counter-intuitive as it suggests that the risk of hospital admissions is reduced by 11% for the wards with PM_{10} higher than 25 μg/m^3 compared with the wards with the PM_{10} level lower than that threshold. Note that the "protective" effect of air pollution on hospital admissions persists if

other categorizations (e.g. quintiles) or a continuous exposure is considered. Residual confounding is the likely cause of this result as the confounders included in the analysis are far from exhaustive given the limited information collected from administrative data sets like HES and Census.

To overcome this issue we integrate additional sources of data providing information on potential confounders via the EPS. We consider the Health Survey for England (HSfE) which is an on-going survey on the health of the individuals living in England and includes each year around 8000 subjects across the country. Through this, we collect information on the following individual level confounders for the years 1994–2001: Education, Smoking, Passive smoking, Drinking, Obesity, Mental illness, Regular exercise, Diabetes, High blood pressure, Vitamin taken, High cholesterol, and Table salt intake.

The combined HSfE surveys covered 87.1% of London wards. From the simulation in Section 3.1, it emerged that the number of individual measurements (i.e. sample size n per area) in each area can affect the accuracy of the risk estimates. It might be possible to set a threshold defining the minimum number of subjects per area: above this threshold the areas are included in the EPS calculation, whereas below they are pooled with the incomplete areas (with missing $\boldsymbol{M}$) and then imputed through (2.3). However, this needs to be traded-off against the information lost when areas with some surveyed individuals (albeit a small number) are considered missing. Figure 2(a) shows the histogram of the number of HSfE samples in the wards of London, and Figure 2(b) displays the trade-off between the number of subjects per area and the geographical coverage. If there is at least one subject sampled in any ward, 87.1% of the areas will be used in the EPS estimation, while the remaining 12.9% will be included through the imputation model. Setting the threshold to 5 subjects leads to a drop in coverage to 75.4%, while considering 10 and 20 subjects per ward as threshold lead to 62.6% and 32.0% of the survey coverage, respectively, as shown in Figure 2(b). Due to the low geographical coverage, we did not consider the threshold higher than 20 subjects per ward.

We checked for differences in the values of the measured confounders, outcome, and exposure between wards above and below the threshold. For instance, the wards with at least 5 subjects are compared with the ones with <5 subjects through density plots in Figure 3, showing the distribution of common shared

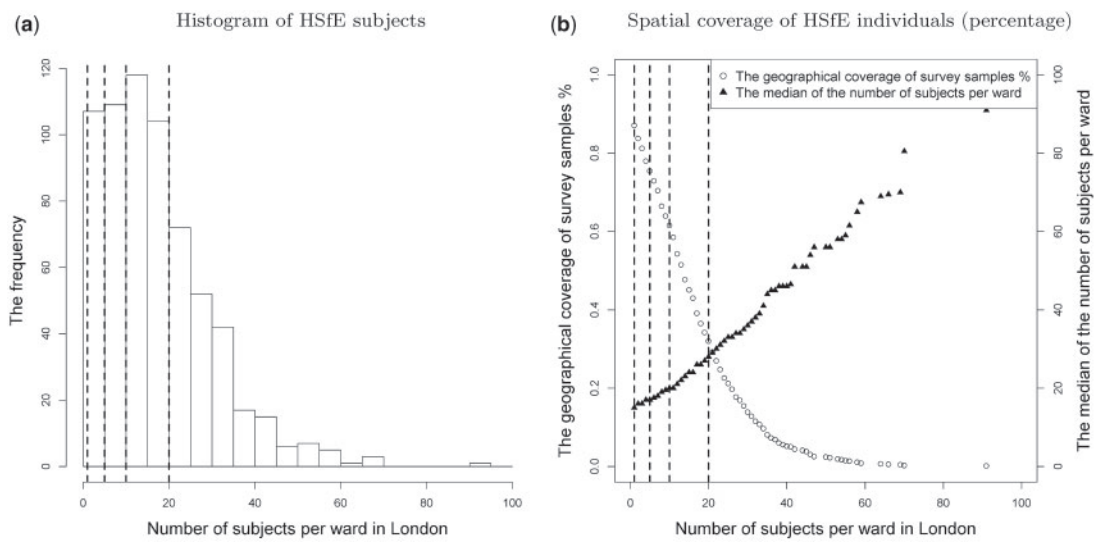

Fig. 2. HSfE distribution and coverage in Greater London: (a) the number of surveyed subjects per ward; (b) the trade-off between spatial coverage and number of sampled individuals. The four dashed lines represent the thresholds considered for estimating/imputing EPS (1, 5, 10, and 20 subjects per ward).

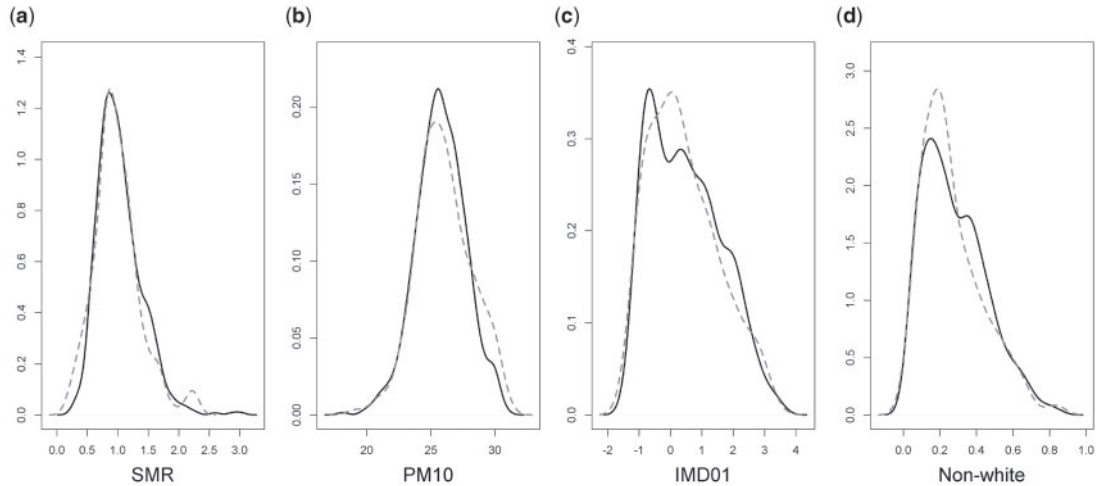

Fig. 3. Comparison of wards covered/not covered by HSfE. The four density plots represent the distribution of the following variables across wards (from left to right): (a) SMR=O/E; (b) PM_{10}, included as a continuous variable ($\mu g/m^3$); (c) IMD01, the Carstairs index at 2001; (d) Proportion of non-white individuals. The solid lines are the distributions in the wards with ≥5 individuals sampled by HSfE, while the dashed line the distributions for the wards with <5 individuals sampled by HSfE or not covered by HSfE. A similar behavior can be seen for all the variables across wards.

Table 2. *RR of hospital admission: PM_{10} higher than 25 $\mu g/m^3$ (chosen cut point) in Greater London. The main analysis with ≥5 subjects per wards is presented in bold, while the others are used as sensitivity analyses*

Areas	Data used	≥1 subject (12.9% missing)		≥5 subjects (24.6% missing)		≥10 subjects (37.4% missing)		≥20 subjects (68.0% missing)	
		RR (95% CI)	95% CI width	RR (95% CI)	95% CI width	RR (95% CI)	95% CI width	RR (95% CI)	95% CI width
$i \in S$	X,***C***	0.91 (0.86–0.96)	0.10	**0.94 (0.89–1.00)**	**0.11**	0.96 (0.90–1.03)	0.13	0.96 (0.88–1.04)	0.16
	X,***C***,EPS	1.04 (0.94–1.14)	0.10	**1.06 (1.01–1.12)**	**0.12**	1.07 (1.01–1.15)	0.14	1.05 (0.95–1.16)	0.21
$i \in S \cup I$	X,***C***	0.89 (0.84–0.94)	0.10	**0.89 (0.84–0.94)**	**0.10**	0.89 (0.84–0.94)	0.10	0.89 (0.84–0.94)	0.10
	X,***C***,EPS	1.05 (0.99–1.12)	0.13	**1.02 (0.96–1.09)**	**0.13**	1.03 (0.97–1.10)	0.13	1.06 (0.99–1.13)	0.14
	X,***C***,***M*** (MICE)	0.98 (0.93–1.04)	0.11	**0.93 (0.88–0.99)**	**0.11**	0.92 (0.87–0.98)	0.11	0.89 (0.84–0.94)	0.10

variables are comparable for these two groups. Based on this we use ≥5 subjects as our main analysis while ≥1, 10, and 20 subjects per ward serve as the sensitivity analysis (see Table 2).

The final data set includes 757 wards and 11 364 individuals from the HSfE. The EPS framework ran on a desktop with a Intel i7 3.6 GHz processor and 16Gb of RAM. It took 90 min to run two Markov Chain Monte Carlo chains, each for 50 000 iterations (25 000 discarded as burnin and 25 000 retained to estimate the posterior distribution). In Figure 1 of supplementary material available at *Biostatistics* online, we provide

convergence checks through the $\hat{R}$ statistics, as suggested by Gelman and Hill (2006). The BUGS code is provided at https://github.com/martablangiardo/Wangetal_Biostatistics.

First, we present the results of the analysis on the survey areas, i.e. including only the wards with at least five subjects from HSfE ($i \in S$). Table 2 shows that in a standard small area framework, the naïve approach ignoring the confounders $\boldsymbol{M}$ leads to a protective effect of air pollution (posterior mean of the RR equal to 0.914 with 95% CI (0.89–1.00)). However, after the inclusion of $\boldsymbol{M}$ through EPS adjustment, it changes to 1.08 (1.01–1.14). This implies that residual confounding is still substantial in the outcome–exposure relationship in Greater London if only adjusting for the few confounding factors available in all areas (deprivation, ethnicity, age, and sex), while the protective effect disappears when including the confounders $\boldsymbol{M}$ as well. By including all the wards in the analysis via EPS imputation, the results are similar with an estimate of 1.03 and a slightly narrower CI (0.97–1.09). On the other hand MICE produces a relative risk estimate equal to 0.93 (0.88–0.99), suggesting the presence of some bias in the MICE imputation framework. We list potential reasons behind this bias in the Discussion.

Changing the threshold on the minimum number of subjects surveyed in each area does not impact substantially on the results: the point estimates remain stable across the different analyses, which points towards robustness of the estimated relationship between air pollution and CHD. However, when considering only $i \in S$ uncertainty of the estimates increases from 0.13 to 0.21 for the model including the EPS. This is expected as the sample size decreases substantially when the number of surveyed individuals increases. In comparison, the model on $i \in S \cup I$ which includes the EPS imputation does not show an important change in uncertainty as the threshold change, which is stable around 0.12–0.14; this suggests that what is gained in precision increasing the number of surveyed individuals is counterbalanced by the increase in uncertainty from the imputation step. In comparison, MICE shows biased results towards the naïve approach, consistently across the different thresholds.

In summary, the result of EPS in Table 2 shows that the wards exposed to a high level of PM_{10} (over 25 μg/m^3 annually) have on average a 2–4% higher risk of hospital admission than the wards with PM_{10} below that threshold in Greater London, albeit the estimate is characterized by a certain degree of uncertainty (95% CI – 3%–12%). Note that the results are consistent for different thresholds on the PM_{10} scale (see Section 6 of supplementary material available at *Biostatistics* online for the results of the model using a threshold of 25% or 75% on the PM_{10} distribution). This is somewhat consistent with the individual level study by Cesaroni *and others* (2014) which showed a small but positive effect of PM_{10} on coronary events across 11 cohorts in Europe.

5. DISCUSSION AND CONCLUSION

In this article, we developed a novel small area modeling framework based on a propensity score index to estimate the effect of an exposure on a health outcome when there are only a limited number of confounders available at the ecological level. We estimate EPS which synthesizes additional individual level confounders from external sources into an area level scalar and define a strategy to impute EPS for the areas where the individual records are missing.

Our EPS is built on the PS defined by McCandless *and others* (2012) which considers only the partially measured confounders $\boldsymbol{M}$; this differs from the original definition proposed by Rosenbaum and Rubin (1983) that is $\mathrm{PS}_i = \theta_0 + \boldsymbol{C}_i^T \boldsymbol{\theta}_C + \boldsymbol{M}_i^T \boldsymbol{\theta}_M$, which would include also the ecological confounders fully observed on all the areas. Our choice is justified by the fact that, being $\boldsymbol{C}$ fully observed, they should be directly included in the adjustment model. Hence, our interest is to link only $\boldsymbol{M}$ to the outcome–exposure analysis via EPS and at the same time to model the missingness of $\boldsymbol{M}$. An imputation model can be assigned to the partial PS to model the missingness of $\boldsymbol{M}$, but this would not be possible if the original PS definition were used.

We build on from the partial PS and we (i) propose to estimate $\boldsymbol{M}$ from the individual data through a hierarchical model with a spatial structure, (ii) specify a structured imputation model to predict the EPS when missing which accommodates a non-linear relationship between the EPS and the $\boldsymbol{C}$ as well as a spatial structure, (iii) block the feedback from the outcome O to the EPS estimation, and (iv) allow the feedback from O to the EPS imputation. Points (iii) and (iv) are crucial: in McCandless *and others* (2012) the feedback from the outcome influences the estimate of the PS estimation model since this is jointly estimated with the adjustment model. This joint framework has been criticized by Rubin (2007) and recently by Zigler *and others* (2013), which in a simulation study showed that the feedback from the outcome leads to a biased estimation of the PS estimates. As we separate the EPS estimate from the analysis model, we are not affected by this issue. In light of this, it is crucial to stress that we are not framed in a fully Bayesian perspective, but we acknowledge recent contributions in this direction (see for instance Saarela *and others*, 2016, who presented a fully Bayesian doubly robust regression for causal inference, focusing on inverse probability of treatment weighting as opposed to PS covariate adjustment). On the other hand in line with the missing data literature (e.g. Kenward and Carpenter, 2007) the EPS imputation should allow for the input of the outcome O which we ensure through the joint specification of the EPS imputation and of the analysis model.

We note that an alternative formulation of the imputation model presented in (2.3) would consist of including the exposure variable X as a response: $\text{logit}(P(X_i = 1|\boldsymbol{C}_i, \text{EPS}_i)) = \theta_1 + \boldsymbol{C}_i^T \boldsymbol{\theta}_C + \text{EPS}_i$ and consequently $\text{EPS}_i = \eta_1 + f(\boldsymbol{C}_i) + \phi_{d_i}$, as originally presented in McCandless *and others* (2012). This alternative formulation would be the most natural way of defining the relationship between EPS and $\boldsymbol{C}$ as it is framed in the same perspective of the estimation model in (2.2). Through a simulation study (presented in the Section 4 of supplementary material available at *Biostatistics* online), we have concluded that the two formulations are equivalent in terms of model performance; at the same time the alternative formulation requires an additional logistic regression, being potentially less computational efficient, thus for the main simulation and for the illustrative example we have presented the specification in (2.3).

We run an extensive simulation study to evaluate the performance of our method and to compare it with MICE, a well-designed (and commonly used) method to deal with missing data; note that this was the natural model to compare against as a standard Bayesian imputation model would not have been computational viable when considering a relatively large number of confounders. We found that the results from MICE are biased towards the naïve model (i.e. ignoring the confounders $\boldsymbol{M}$) and consider that this might be related to the following two issues: (i) MICE suffers from the uncongeniality meaning that the functional form of Y in the imputation model is different from its feedback in the analysis model; for instance, in a Poisson model, the outcome has to be included through the *SMR* to predict EPS in the imputation model, which is different from the feedback function of the observed and expected count in a Poisson likelihood; (ii) in MICE, a missing confounder will be predicted based on other confounders in the same area which are also missing, thus little can be gained from the iterative prediction with the missing confounders as predictors, and this results in increased uncertainty in the estimates. In addition, as MICE is not designed to take into account the spatial correlation in the imputation stage, to ensure a fair comparison we did not include any spatial structure in the simulation study; however ecological variables are in general spatially correlated, and it is important to account for such correlation in the prediction of missing confounders.

The issues mentioned above can be mitigated by using EPS. First, EPS incorporates the imputation model as a sub-model via a full Bayesian framework, which guarantees the feedback from the outcome to the missing variable is congenial, thus eliminating the bias arising from imputing a missing covariate in a Poisson likelihood.

Second, EPS compresses all missing covariates into one scalar variable, and the single missing EPS can be dealt with more effectively by the imputation model, thus avoiding the issue of using variables with missing data as predictors (as in MICE); as a result, the EPS framework is characterized by less uncertainty

than considering the survey areas only. Third, EPS provides the foundation to extend the complexity of the imputation model via the Bayesian framework: we showed through an additional simulation (in Section 5 of supplementary material available at *Biostatistics* online) that the proposed method works when spatially structured variables are considered as well as when there is a degree of non-linearity between the confounders and the health outcome.

We believe that for a adequate performance of our EPS strategy, the predictive strength of the imputation model in (2.3) is crucial and should be carefully investigated in each case study. Indeed, adequate predictive strength is needed in order to balance the influence of the outcome on the EPS imputation. This is a delicate and understudies issue which will be explored in further work.

By adopting RW(2) as the link function in the PS adjustment, the EPS framework is not only highly flexible in dealing with non-linearity, but can also be implemented to analyze a large ecological data set due to its much faster speed than the splines based functions. Moreover, RW(2) can be extended to multiple covariates like $\boldsymbol{C}$, and we also considered RW(2) as the ideal function to accommodate the potential non-linearity in the imputation model.

Through the simulation and the application, we showed that EPS is a promising way to incorporate individual level data sets like surveys or cohorts to adjust for unmeasured confounders in small area studies. Future work will extend the framework to deal with categorical instead of binary exposures and to several correlated exposure variables which are particularly relevant aspects in the field of environmental epidemiology aiming at feeding back policy makers on the epidemiological risk of environmental hazards.

SUPPLEMENTARY MATERIAL

Supplementary material is available at http://biostatistics.oxfordjournals.org.

ACKNOWLEDGMENTS

The authors thank Nicola Best for fruitful discussion on the developed methodology, Gianluca Baio and Alexina Mason for providing comments on the draft of the manuscript. *Conflict of Interest*: None declared.

FUNDING

The work of the UK Small Area Health Statistics Unit is funded by Public Health England as part of the MRC-PHE Centre for Environment and Health, funded also by the UK Medical Research Council. This paper does not necessarily reflect the views of Public Health England or the Department of Health. SAHSU holds approvals from the National Research Ethics Service - reference 12/LO/0566 and 12/LO/0567 - and from the Health Research Authority Confidentially Advisory Group (HRA-CAG) for Section 251 support (HRA - 14/CAG/1039). Marta Blangiardo, Monica Pirani, Anna Hansell and Sylvia Richardson acknowledge support from the MRC Methodology grant MR/M025195/1. Sylvia Richardson acknowledges support from the MRC programme grant MC_UP_0801/1.

REFERENCES

BESAG, J. AND KOOPERBERG, C. (1995). On conditional and intrinsic autoregressions. *Biometrika* **82**, 733–746.

BESAG, J., YORK, J., AND MOLLI'E, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* **43**, 1–20.

BUUREN, S. V. AND OUDSHOORN, K. (1999). *Flexible Multivariate Imputation by MICE*. Leiden, The Netherlands: TNO Prevention Center.

CESARONI, G., FORASTIERE, F., AND STAFOGGIA, M. E. A. (2014). Long term ex posure to ambient air pollution and incidence of acute coronary events: prospective cohort study and meta-analysis in 11 European cohorts from the escape project. *BMJ* **348**, 1–16.

COMEAP (2006). COMEAP: cardiovascular disease and air pollution - Publications - GOV.UK.

DANIELS, M. AND HOGAN, J. (2008). *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. Boca Raton, US: CRC Press.

FAHRMEIR, L. AND LANG, S. (2001). Bayesian inference for generalized additive mixed models based on Markov random field priors. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **50**, 201–220.

GAMERMAN, D., MOREIRA, A. R. B., AND RUE, H. (2003). Space-varying regression models: specifications and simulation. *Computational Statistics & Data Analysis* **42**, 513–533.

GELMAN, A. AND HILL, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Modelling*. New York, US: Cambridge University Press.

HOEK, G., BRUNEKREEF, B., GOLDBOHM, S., FISCHER, P., AND VAN DEN BRANDT, P. A. (2002). Association between mortality and indicators of traffic-related air pollution in the Netherlands: a cohort study. *The Lancet* **360**, 1203–1209.

HUGHES, R. A., WHITE, I. R., SEAMAN, S. R., CARPENTER, J. R., TILLING, K., AND STERNE, J. A. (2014). Joint modelling rationale for chained equations. *BMC Medical Research Methodology* **14**, 28.

KENWARD, M. G. AND CARPENTER, J. (2007). Multiple imputation: current perspectives. *Statistical Methods in Medical Research* **16**, 199–218.

LIPSETT, M. J., OSTRO, B. D., REYNOLDS, P., GOLDBERG, D., HERTZ, A., JERRETT, M., SMITH, D. F., GARCIA, C., CHANG, E. T., and BERNSTEIN, L. (2011). Long-term exposure to air pollution and cardiorespiratory disease in the California Teachers Study Cohort. *American Journal of Respiratory and Critical Care Medicine* **184**, 828–835.

LUNN, D., JACKSON, C., BEST, N., THOMAS, A., AND SPIEGELHALTER, D. (2012). *The BUGS Book: A Practical Introduction to Bayesian Analysis*. Boca Raton, US: CRC Press.

MCCANDLESS, L. C., RICHARDSON, S., AND BEST, N. (2012). Adjustment for missing confounders using external validation data and propensity scores. *Journal of the American Statistical Association* **107**, 40–51.

MENG, X.-L. (1994). Multiple-imputation inferences with uncongenial sources of input. *Statistical Science* **9**, 538–558.

MILLER, K. A., SISCOVICK, D. S., SHEPPARD, L., SHEPHERD, K., SULLIVAN, J. H., ANDERSON, G. L., AND KAUFMAN, J. D. (2007). Long-term exposure to air pollution and incidence of cardiovascular events in women. *New England Journal of Medicine* **356**, 447–458.

MOLITOR, N.-T., BEST, N., JACKSON, C., AND RICHARDSON, S. (2009). Using Bayesian graphical models to model biases in observational studies and to combine multiple sources of data: application to low birth weight and water disinfection by-products. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **172**, 615–637.

PUETT, R. C., HART, J. E., YANOSKY, J. D., PACIOREK, C., SCHWARTZ, J., SUH, H., SPEIZER, F. E., AND LADEN, F. (2009). Chronic fine and coarse particulate exposure, mortality, and coronary heart disease in the Nurses Health Study. *Environmental Health Perspectives*, **117**, 1697–1701.

ROSENBAUM, P. R. AND RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55.

RUBIN, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York, US: John Wiley & Sons.

RUBIN, D. B. (2007). The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. *Statistics in Medicine* **26**, 20–36.

SAARELA, O., BELZILE, L., AND STEPHENS, D. (2016). A Bayesian view of doubly robust causal inference. *Biometrika*, **103**, 667–681.

SENN, S., GRAF, E., AND CAPUTO, A. (2007). Stratification for the propensity score compared with linear regression techniques to assess the effect of treatment or exposure. *Statistics in Medicine* **26**, 5529–5544.

STURMER, T., SCHNEEWEISS, S., AVORN, J., and GLYNN, R. J. (2005). Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. *American Journal of Epidemiology*, **162**, 279–289.

VIENNEAU, D., DE HOOGH, K., BEELEN, R., FISCHER, P., HOEK, G., AND BRIGGS, D. (2010). Comparison of land-use regression models between Great Britain and the Netherlands. *Atmospheric Environment* **44**, 688–696.

WHITE, I. R., ROYSTON, P., AND WOOD, A. M. (2011). Multiple imputation using chained equations: issues and guidance for practice. *Statistics in Medicine* **30**, 377–399.

ZIGLER, C. M., WATTS, K., YEH, R. W., WANG, Y., COULL, B. A., AND DOMINICI, F. (2013). Model feedback in Bayesian propensity score estimation. *Biometrics* **69**, 263–273.

[*Received January 18, 2017; revised September 26, 2017; accepted for publication October 13, 2017*]