



Bed-sharing and childhood asthma: from associations to causal modelling



CrossMark

Alban Ramette and Claudia Elisabeth Kuehni

Affiliation: Paediatric Respiratory Epidemiology Research Group, Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland.

Correspondence: Claudia E. Kuehni, Institute of Social and Preventive Medicine, University of Bern, Finkenhubelweg 11, CH-3012, Bern, Switzerland. E-mail: claudia.kuehni@ispm.unibe.ch



@ERSpublications

We need to explore causal associations and apply novel statistical approaches in respiratory epidemiology <http://ow.ly/GWOYT>

The paper by *LUIJK et al.* [1] in this issue of the *European Respiratory Journal* is a highlight for two reasons: it is the first to investigate parent–child bed-sharing practices and risk of childhood asthma; and it uses novel statistical approaches to explore causal relationships.

One can hypothesise diverse influences of bed-sharing practices on respiratory disease. First, the close contact between bed-sharing parents and their offspring could lead to increased transmission of infectious agents and, thus, more virus-induced wheeze in the children. Conversely, the hygiene hypothesis suggests that contact with common pathogens in early childhood might protect from later allergic disease [2]. Lastly, parents of children with nocturnal asthma could tend to bed-share as a way to monitor their children. *LUIJK et al.* [1] took the challenge of investigating these associations by studying the early-life trajectories of 6160 children from the Generation R study, a population-based cohort from the Netherlands. They assessed bed-sharing twice, at ages 2 and 24 months, and wheezing at ages 1, 2, 3, 4 and 6 years. Data were analysed using generalised estimating equation (GEE) models, which allow analysis of longitudinal outcome data with several measurements per subject. The authors found no association between bed-sharing in infancy (at age 2 months) and later wheeze. In contrast, bed-sharing in toddlerhood (at age 2 years) was associated with more subsequent wheeze. The authors then used cross-lagged modelling to investigate the directions of the association. This approach compares regression coefficients between variables measured at subsequent time-points and estimates the strength of the causal effects between them. It confirmed that bed-sharing in infancy was not associated with subsequent wheezing but showed that bed-sharing in toddlers was, and the association became stronger for wheeze occurring at later ages. It found no evidence suggesting that wheezing in infancy leads to bed-sharing at age 2 years.

The study gives no final answer as to whether the relationship found between bed-sharing in toddlerhood and wheezing at ages 3–6 years is causal, or what the mechanisms could be. It certainly lends no support to a protective effect suggested by the hygiene hypothesis, nor does it suggest that parents tend to share beds with wheezing toddlers to monitor their breathing. This leaves the third option, that bed-sharing leads to more infections, which in turn trigger wheezing episodes or lead to airway remodelling, but the support for this explanation was not strong. For instance, the associations were strongest for wheeze at age 6 years, 4 years after the assessment of sleeping practices, rather than simultaneously or shortly later. In addition, adjusting the model for respiratory tract infections did not attenuate the association between bed-sharing and wheeze, suggesting that the association is not explained by infectious mechanisms. In fact,

Received: Dec 22 2014 | Accepted: Dec 23 2014

Support statement: The work presented in this editorial was funded by the Swiss National Science Foundation (grant number 32003B-144068). Funding information for this article has been deposited with FundRef.

Conflict of interest: None declared.

Copyright ©ERS 2015

the effect sizes were moderate to small (OR <2) and were further attenuated in the fully adjusted model. Therefore, residual confounding by factors not measured in the study remains possible. Despite the remaining uncertainties, we warmly compliment the authors for their efforts to shed light on the underlying mechanisms.

The “chicken and egg” causality dilemma in respiratory epidemiology

Why don't we see more of this kind of analysis in respiratory epidemiology? A number of techniques have been developed in analytical epidemiology to determine whether suspected risk factors are associated with an outcome of interest, how strong the association is, and whether it might be causal or not [3]. If we see evidence for an association, *i.e.* if the suspected risk factor is more common in diseased persons, we must first explore the influence of random sampling error (*e.g.* by performing significance tests or, preferably, calculating confidence intervals), of systematic error (bias) and of confounding (by a factor that is related to both exposure and outcome of interest). Only then can we consider a causal relationship between the observed associations. The Bradford Hill criteria for a causal association, published in 1965, still provide a useful background framework for thoughts on causality. These criteria include the temporality of an association, its strength, biological plausibility, dose response, reversibility, specificity, consistency between studies, analogy to well-known processes and coherence with other findings [4, 5].

Common diseases, such as asthma and COPD, have a multifactorial pathogenesis, with multiple genes and environmental factors interacting, complex causal paths and diverse phenotypes [6, 7]. The effects of single risk factors are often small and the direction of causal pathways questionable, such as in the study by LUIJK *et al.* [1]. Standard statistical methods are largely not adapted to detect them and this may be the cause of many contradicting results in the clinical literature [8]. Risk factors can act at different levels, including biological, behavioural and group levels, and the interrelation between them often includes dynamic feedback and changes over time [9]. These networks cannot be analysed using standard methods [10].

Multivariate statistical methods, where several outcomes (dependent variables) are conjointly analysed as a function of one or more risk factors, are promising tools for respiratory epidemiology. Those methods are inherently more complex than simple regression modelling because they deal with many variables that display dependencies that cannot be described by simpler modelling approaches. For instance, path analysis (and its generalisation, structural equation modelling) can be used to estimate known causal effects or for testing causal models. If causal relationships are difficult to see in a complex data set, then Bayesian networks [11] may be used to generate new hypotheses about causal pathways from a statistical viewpoint, without *a priori* categorisation of the variables as outcomes or exposures. The role of the researchers is then to scrutinise the biological or clinical plausibility of the relationships generated from the data-centred models and to make sense of them. Such an approach may generate new ideas for experimental studies. Other fields of research such as computer science, biology and ecology [12], which also deal with large, noisy data sets, have successfully embraced such techniques that deal with covariation among responses and complex relationships with predictor variables.

Use of multivariate and causal models in European Respiratory Society publications

We wanted to know how representative the methods used by LUIJK *et al.* [1] are of current epidemiological studies investigating causal relationships. To this end, we undertook a quick systematic review of all papers published by the four regular European Respiratory Society (ERS) publications (*European Respiratory Journal*, *European Respiratory Review*, *ERS Monograph* and *Breathe*) since 1988. We searched for the occurrence of typical statistical methods in the full text of articles.

The analysis method most often encountered (in 2220 articles) was “logistic regression” (fig. 1). Many authors (908 (41%) out of 2220 articles) wrongly called it “multivariate” when they included several exposures in their model. The proper term for this would be “multivariable” or “multiple” regression [13]. Truly multivariate methods (*i.e.* considering multiple outcome variables) were rarely used. Among these, we mainly found descriptive multivariate statistics: principal component analysis (97 articles), cluster analysis (180 articles), correspondence analysis (16 articles) and multivariate ANOVA (17 articles). Only a handful of papers had used approaches that explicitly deal with causation, such as structural equation modelling (12 articles) or path analysis (nine articles). None had used Bayesian networks to discover new causal structures.

In a second step, we paid specific attention to articles published between 2010 and 2013 in the *European Respiratory Journal* only (table 1). We found 1196 original articles, 236 with the term “epidemiology” in one or more record field. We identified 155 articles focusing on risk factors and outcomes, retrieved the full texts and extracted information on analytic methods.

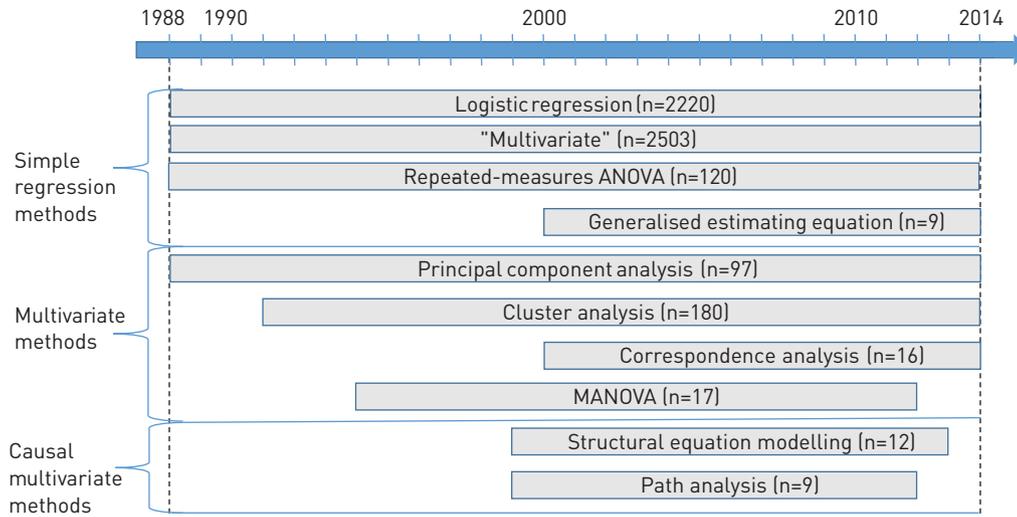


FIGURE 1 Overview of statistical methodology usage in the full text content of the four regular European Respiratory Society (ERS) publications from 1988 to November 2014. Publications searched were: *European Respiratory Journal*, *European Respiratory Review*, *ERS Monograph* and *Breathe*. The search terms were used with quotation marks on the ERS search interface (<http://ersjournals.com/search>) and the total number of returned results for each query is shown. MANOVA: multivariate ANOVA.

We found that nearly all studies (144 (93%) out of 155) had used statistical significance tests and most (133 (86%) out of 155) had employed simple regression models (logistic, linear, Poisson or Cox regression, GEEs, or linear mixed models). Among these, only 10 had used GEEs and 10 linear mixed models, methods that allow analysis of repeated outcomes. Only about half of the studies (71 (46%) out of 155) had adjusted for confounders in multivariable regressions. Again, many studies mislabelled this as “multivariate regression”. Truly multivariate descriptive methods had been used in seven (5%) studies:

TABLE 1 Use of analysis methods in original articles on risk factors and health outcomes published in the *European Respiratory Journal*, 2010–2013

Analysis methods used	Details	Publication year				
		2010	2011	2012	2013	Total
Total number of studies retrieved	Epidemiological original studies looking at risk factor–disease associations	49 (100)	40 (100)	44 (100)	22 (100)	155 (100)
Simple methods						
Statistical significance tests	Chi-squared test, Fisher’s exact test, Student’s t-test, ANOVA	47 (96)	37 (93)	40 (91)	20 (91)	144 (93)
Simple regression methods	Logistic, linear, Poisson and Cox regressions	33 (67)	27 (68)	34 (77)	18 (82)	112 (72)
Regression methods for repeated outcomes	GEEs, linear mixed models	7 (14)	5 (13)	8 (18)	1 (5)	21 (14)
Regressions adjusted for confounders		22 (45)	15 (38)	20 (45)	14 (64)	71 (46)
Multivariate methods	Principal component analysis, cluster analysis, factor analysis, co-inertia analysis	4 (8)	2 (5)	1 (2)	0 (0)	7 (5)
Causal multivariate methods	Cross-lag modelling, path analysis, Bayesian networks	0 (0)	0 (0)	1 (2)	0 (0)	1 (1)

Data are presented as n (%) studies investigating risk factor–outcome associations. We performed an OVID Medline search, which targeted only the *European Respiratory Journal*, and only original research articles from observational epidemiological studies (i.e. we excluded meta-analyses, reviews, historical articles, consensus development conferences, guidelines, randomised controlled trials, clinical trials, comment, biographies and editorial). The term “epidemiolo*” was added as research term across all fields. We carefully examined the methods section of each article, removed duplicates, and retained only studies that looked at associations between risk factor(s) and outcome(s). GEE: generalised estimating equation.

principal component analysis (one article), cluster analysis (four articles), factor analysis (one article) and co-inertia analysis (one article). Multivariate methods for investigating causal relationships had only been used by one (1%) study. There, the authors had used path analysis to describe effects of breastfeeding on infant weight gain and lung function at different ages [14]. No study had used a Bayesian network. We found no evidence for increasing use of multivariate and causal methods over time.

Our search was encumbered by the fact that many abstracts lacked appropriate terms to identify them as epidemiological studies looking at risk factors. Therefore, our search probably missed out some studies. We do not think this has strongly distorted our findings. If anything, the studies missed because their abstracts did not describe their methodology will have used less, rather than more, novel causal methods. In summary, our two searches therefore suggest that causal models, such as used by LUIJK *et al.* [1], represent a rare exception in epidemiological articles in ERS publications.

Lessons learned

Because the respiratory system serves as a major interface between the host and the external environment, many factors and scales of analyses must be considered to progress in our understanding of the multiple aetiologies and comorbidities associated with asthma and other respiratory diseases [15]. If epidemiological research is to inform clinical practice and population health interventions, then more complex approaches, which describe the interdependencies of multiple factors on health outcome, are needed. We can begin to disentangle the intricate relationships between and among multiple outcomes and exposures by using exploratory or hypothesis-driven multivariate models. Large data sets, computing power and methodological advancements are at our fingertips. Looking at the publications of the past years, we think that improvements are possible at three levels.

1) Study design: good analysis needs good data. Biases (systematic errors in design, conduct or analysis of a study leading to mistaken results) must be foreseen and minimised at the planning stage of a study, as they can hardly ever be dealt with properly at the analysis stage. Problems with inaccurately measured variables arise often with large routine datasets, collected for other purposes [16]. In our search, we found many cohort studies that had collected information on exposures and outcomes only at one time-point and were analysed cross-sectionally. Thus, despite using cohort datasets, these studies could not investigate the temporality of the exposure–disease relationship.

2) Statistical methodologies: there will always be the need for straightforward studies that test a clear hypothesis by a simple regression model. However, we think that methods developed for causal modelling, which allow the analysis of interdependencies between multiple exposures and outcomes and possible reverse causation, should be employed more often than in 1% of published studies. Epidemiological simulation models are another interesting but underemployed tool to test causal relationships [17, 18]. To make full use of such methods, well-trained statisticians must be fully integrated into planning and analysing studies, and epidemiologists need appropriate training [19]. Equally important is that clinicians and basic scientists critically think through new hypotheses suggested by data-driven models, to make sure the analyses and findings have clinical or public health relevance.

3) Review and publishing: in theory, the review process should help to get rid of inappropriate or confusing use of terms (*e.g.* multivariate *versus* multivariable) and encourage the use of alternative statistical options where appropriate. This seems not to have worked in quite some studies. Furthermore, we had a hard time retrieving epidemiological studies from the databases and evaluating relevant information from the original papers because of a lack of standardisation in reporting, for instance, of study design and statistical methodologies. This may be improved by adhering to guidelines for reporting observational studies [20], and by a careful choice of keywords describing study design and methods.

The paper by LUIJK *et al.* [1] is therefore a step in the right direction, and we hope it will generate increased interest in exploring causal associations and applying novel statistical approaches in respiratory epidemiology.

Acknowledgements

The authors sincerely thank their colleagues of the Institute of Social and Preventive Medicine (University of Bern, Bern, Switzerland): Ben Spycher for comments on the manuscript, and Myrona Goutaki, Maja Jurca, Jingying Wang and Florian Halbeisen for their help with data extraction and literature search.

References

- 1 Luijk MPCM, Sonnenschein-van der Voort AMM, Mileva-Seitz VR, *et al.* Is parent–child bed-sharing a risk for wheezing and asthma in early childhood? *Eur Respir J* 2015; 45: 661–669.
- 2 Brooks C, Pearce N, Douwes J. The hygiene hypothesis in allergy and asthma: an update. *Curr Opin Allergy Clin Immunol* 2013; 13: 70–77.
- 3 Gordis L. *Epidemiology*. 5th Edn. Philadelphia, Saunders, 2013.

- 4 Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965; 58: 295–300.
- 5 Lucas RM, McMichael AJ. Association or causation: evaluating links between “environment and disease”. *Bull World Health Organ* 2005; 83: 792–795.
- 6 Barnes KC. Genetic studies of the etiology of asthma. *Proc Am Thorac Soc* 2011; 8: 143–148.
- 7 Spycher BD, Silverman M, Brooke AM, *et al.* Distinguishing phenotypes of childhood wheeze and cough using latent class analysis. *Eur Respir J* 2008; 31: 974–981.
- 8 Buchanan AV, Weiss KM, Fullerton SM. Dissecting complex disease: the quest for the Philosopher’s Stone? *Int J Epidemiol* 2006; 35: 562–571.
- 9 Galea S, Riddle M, Kaplan GA. Causal thinking and complex system approaches in epidemiology. *Int J Epidemiol* 2010; 39: 97–106.
- 10 Flegal KM, Graubard BI, Williamson DF, *et al.* Reverse causation and illness-related weight loss in observational studies of body weight and mortality. *Am J Epidemiol* 2011; 173: 1–9.
- 11 Lewis FI, Ward MP. Improving epidemiologic data analyses through multivariate regression modelling. *Emerg Themes Epidemiol* 2013; 10: 4.
- 12 Ramette A. Multivariate analyses in microbial ecology. *FEMS Microbiol Ecol* 2007; 62: 142–160.
- 13 Hidalgo B, Goodman M. Multivariate or multivariable regression? *Am J Public Health* 2013; 103: 39–40.
- 14 Soto-Ramirez N, Alexander M, Karmaus W, *et al.* Breastfeeding is associated with increased lung function at 18 years of age: a cohort study. *Eur Respir J* 2012; 39: 985–991.
- 15 Martinez FD. Genes, environments, development and asthma: a reappraisal. *Eur Respir J* 2007; 29: 179–184.
- 16 Lash TL, Abrams B, Bodnar LM. Comparison of bias analysis strategies applied to a large data set. *Epidemiol* 2014; 25: 576–582.
- 17 Smith BT, Smith PM, Harper S, *et al.* Reducing social inequalities in health: the role of simulation modelling in chronic disease epidemiology to evaluate the impact of population health interventions. *J Epidemiol Community Health* 2014; 68: 384–389.
- 18 Eubank S, Guclu H, Kumar VS, *et al.* Modelling disease outbreaks in realistic urban social networks. *Nature* 2004; 429: 180–184.
- 19 Spitz MR, Lam TK, Schully SD, *et al.* The next generation of large-scale epidemiologic research: implications for training cancer epidemiologists. *Am J Epidemiol* 2014; 180: 964–967.
- 20 von Elm E, Altman DG, Egger M, *et al.* The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med* 2007; 4: e296.