



Natural experiments and large databases in respiratory and cardiovascular disease

Jørgen Vestbo

Affiliation: Centre for Respiratory Medicine and Allergy, The University of Manchester and University Hospital South Manchester NHS Foundation Trust, Manchester, UK.

Correspondence: Jørgen Vestbo, Centre for Respiratory Medicine and Allergy, 2nd Floor, ERC Building, University Hospital South Manchester, Southmoor Road, Manchester, M23 9LT, UK.
E-mail: jorgen.vestbo@manchester.ac.uk

ABSTRACT A number of scientific questions cannot be tested in a laboratory, clinic or clinical trial setting. In many cases, observational data can be used to test such hypotheses.

This article illustrates how epidemiology can contribute and shows the different ways of using observational data through three approaches: 1) prospective cohort study design; 2) time series analysis; and 3) a nested case-control design in pharmacoepidemiology.

In a prospective cohort study design, three cohorts were merged to study lung function decline, testing the importance of different trajectories of lung function decline for developing chronic obstructive pulmonary disease (COPD). Using these three well-described cohorts it was documented that maximally attained lung function in early adulthood is as important as excess decline in forced expiratory volume in 1 s for the development of COPD. Time series analysis is used to examine exposures and disease over time. In a recent review of cardiovascular disease some interesting associations, and not least lack of associations, were presented. Assessing effects of drugs in database studies is challenging. In a nested case-control design in a large cohort study, statins were found to reduce the risk of COPD exacerbations. These findings will be discussed.

Observational data from large databases, as well as carefully collected data in cohort studies, can be used to test hypotheses that may not be addressed in a traditional experimental setting.



@ERSpublications

Natural experiments using large databases with observational data can be of significant use in respiratory research <http://ow.ly/4mYBVx>

Introduction

“Natural experiments” is a term often used to describe studies in which the experimental setting is not determined by the researcher but where the researcher makes use of a situation. And, where a scientific question can be addressed using observational data collected in a way that makes meaningful deduction possible from association analysis. This is the case in many disease areas, not least in respiratory medicine where a significant amount of insight comes from well-conducted epidemiological studies. A number of scientific questions cannot be tested in a laboratory, clinic or clinical trial setting. The reasons are numerous but often: experiments cannot be set up to examine potentially harmful exposures in humans; certain patient groups are not available for clinical trials; the duration of required exposure and/or follow-up may exceed what is feasible in a clinical trial, *etc.* In many of these cases, observational data can be used to test such hypotheses.

Editorial comments in *Eur Respir Rev* 2016; 25: 104–107.

Received: March 25 2016 | Accepted after revision: April 21 2016

Conflict of interest: Disclosures can be found alongside the online version of this article at err.ersjournals.com

Provenance: Submitted article, peer reviewed.

Copyright ©ERS 2016. ERR articles are open access and distributed under the terms of the Creative Commons Attribution Non-Commercial Licence 4.0.

It will go beyond the scope of this article to comment on all these possibilities. The following article will, therefore, focus on three areas illustrating the potentials of this approach. They are based on three recent publications [1–3] and cover the prospective cohort study design, time series analysis, and a nested case–control design in pharmacoepidemiology with mediation analysis.

Cohort study of trajectories of chronic obstructive pulmonary disease

Cohort studies are often used to prospectively examine exposure–response relationships. As such, they are the main contributor to our knowledge on risk factors for development and/or deterioration of disease, so-called risk-factor epidemiology. Although often seen as inferior to true experiments, carefully followed cohorts have contributed to generating evidence, the most obvious example being the link between smoking and lung cancer [4, 5]. Since the seminal studies by FLETCHER and co-workers [6, 7] it has been acknowledged that decline in forced expiratory volume in 1 s (FEV₁) is a crucial determinant of risk of chronic obstructive pulmonary disease (COPD) and a number of observational studies have examined risk factors for an excess decline [8–12]. However, it is also clear that determinants of maximally achieved lung function in early adulthood are important for risk of COPD [6, 7, 13–15]; however, quantitative assessment of this impact has been lacking.

In a prospective cohort study design, three cohorts were merged to study trajectories of lung function decline, The Copenhagen City Heart Study, the Framingham Offspring Study and the Lovelace COPD Cohort Study. This enabled testing of the importance of different trajectories of lung function decline for developing COPD [1]. Cohorts with sufficient baseline data are often costly and time-consuming to set up, and follow-up requires substantial work and resources. However, a number of large cohorts are now in the public domain and can be assessed for specific purposes. Using these three well-described cohorts with spirometry at baseline in young adulthood and during 25 years of follow-up, it was possible to document that maximally attained lung function in early adulthood is probably as important as excess decline in FEV₁ for the development of COPD [1]. It also clearly demonstrated a large variability in FEV₁ decline in

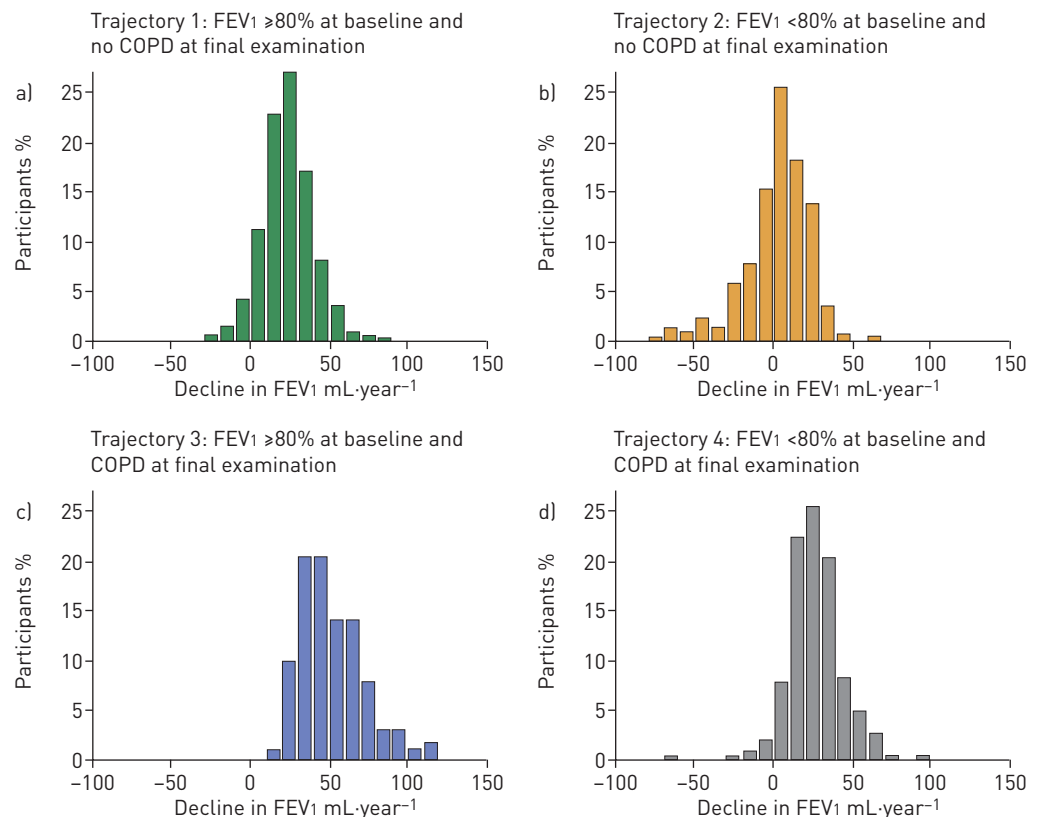


FIGURE 1 Distribution of the decline in forced expiratory volume in 1 s (FEV₁) in four lung function trajectories. Distribution of observed annual decline in forced FEV₁ among 2864 participants in the Framingham Offspring Cohort and the Copenhagen City Heart Study according to the four trajectories, defined on the basis of a normal FEV₁ (≥80% pred) or low FEV₁ (<80% pred) at baseline and the presence or absence of chronic obstructive pulmonary disease (COPD) at the final examination. Participants were considered to have COPD if they had a FEV₁/forced vital capacity ratio <0.70 and an FEV₁ <80% pred. Reproduced from [1] with permission.

subjects diagnosed with COPD, as illustrated in figure 1, and that in patients diagnosed with COPD in their 50s or 60s it is not possible to determine the trajectory they have followed based on clinical parameters. This study is an excellent example of the utility in observational cohort studies when assessing aspects of the natural history of disease.

Time series analysis in cardiovascular disease

Time series analysis is often used to examine whether variations in exposures and disease over time are linked and the analysis lets the time-dependent relationship between the variables studied inform about the strength of association while controlling for relevant confounders. This can be done by applying a number of statistical methods [16] and has been applied in many fields, including respiratory medicine [17–19].

In a recent review of cardiovascular risk factors and disease some interesting associations, and lack of associations, were presented [2]. The authors showed that cardiovascular disease death rates have been decreasing steadily both in regions with reliable trend data and globally. The declines in most European countries have been ongoing for decades without slowing. These positive trends have broadly coincided with, and benefited from, declines in smoking and physiological risk factors, such as blood pressure and serum cholesterol levels, as well as with improvements in medical care, including primary prevention, diagnosis, treatment of acute cardiovascular diseases and post-hospital care. Although not providing a true time-series analysis on the data, the study provides compelling evidence of time-associated relationships. In addition, in several former Soviet countries, increases in alcohol consumption seemed to have caused sharp increases in cardiovascular disease mortality since the early 1990s.

However, there is really nothing that can explain why the decline began when it did, or the similarities and differences in the start time and rate of decline between countries. In addition, there seems to be impact of the increase in obesity seen almost all over the world in adiposity on cardiovascular disease mortality. To this respiratory physician, divine interaction may still have played a role.

Pharmacoepidemiology

Assessing the effects of drugs in database studies is challenging and can go wrong. An example of this was the long-held view that post-menopausal oestrogen substitution reduced risk of cardiovascular events, which has been shown in several large database studies. It took two large randomised studies to show that the observed

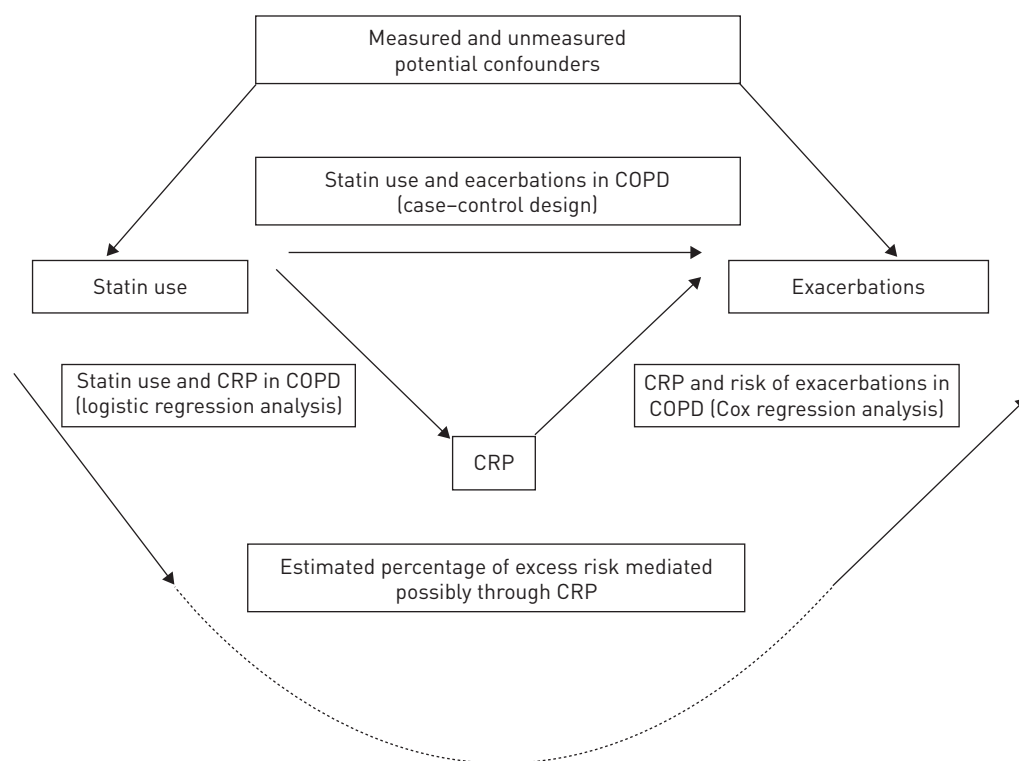


FIGURE 2 Design of a study aimed at examining if statins reduce chronic obstructive pulmonary disease (COPD) exacerbations through lowering of C-reactive protein (CRP) in a nested pharmacoepidemiological case-control register-linkage study. Reproduced and modified from [3] with permission.

beneficial effect was the result of selection bias leading to more affluent and well-educated women with lower risk of cardiovascular disease taking oestrogen [20]. The crucial thing was the lack of randomisation. A large number of other biases can distort true associations [21]; therefore, pharmacoepidemiology is strongest when searching for adverse drug effects. A good example of this is the seminal paper by SUISSA *et al.* [22] on the risk associated with not taking low doses of inhaled corticosteroids in asthma.

Despite several short comings, pharmacoepidemiology can be used to supplement information from randomised trials that benefit from a stronger study design but are often conducted in settings where only a small minority of patients will be included in trials [23]. In a recent Danish study, nested case-control design in a large cohort study was used to assess a possible beneficial effect of statins on the risk of COPD exacerbations, including a mediation analysis to test if a positive effect could be ascribed to a lowering of C-reactive protein (CRP) (figure 2) [3]. Although patients were not randomised to the treatment studied, the study was able to adjust for a number of confounders adding to the credibility of the analysis. The authors found that statins reduced the risk of COPD exacerbations by ~30% and that 15% of this effect could be ascribed to statins lowering CRP. The study came out at more or less the same time as a large randomised controlled trial demonstrating no benefit of statins on exacerbations in COPD [24]. The difference between these studies, however, was not due to differences in study design. The pharmacoepidemiological study used a broad population whereas the randomised trial could only include subjects who did not already have a cardiovascular disease indication for statins as it would be unethical to randomise them to placebo. When restricting the population in the observational study, this study did not find an effect of statins on exacerbations either [3]. If the effect is only seen in patients with or at high risk of cardiovascular disease, this indicated that lowering CRP *per se* may not be beneficial.

This actually fits well with findings from another type of natural experiment. Using Mendelian randomisation [25] it is possible to examine if an association between CRP and exacerbations is due to the effect of something increasing CRP or CRP itself; the latter examined by studying the association between genetic polymorphisms affecting CRP levels and exacerbations. This was previously done in the same setting as the observational study on statins and showed that genetically elevated CRP was not associated with COPD exacerbations [26].

Discussion

Using three examples, the value of natural experiments has hopefully been emphasised. For years, observational studies have been regarded as the lowest level of evidence provided. It seems clear that a more nuanced view is needed. As with the two first examples, observational data and careful analyses clearly provide data that cannot be obtained from an experimental setting. The third study clearly added to the data provided by a strong standard randomised trial, not least as it examined patients who for ethical reasons could not be included in a controlled trial.

Using observational data clearly provides challenges. For risk factor epidemiology, unmeasured confounders are likely to distort true associations and the increasing use of “big data” with little control over data collection may increase this risk. For pharmacoepidemiology in particular, there are biases that can be impossible to control for. In any observational setting there will be a specific reason why a specific doctor gave a specific patient a specific treatment and this cannot be measured. For this reason, there is a need for combining insights and expertise from many different areas to provide “real-life data” that are credible and as unbiased as possible [27, 28]. In addition, we need to decrease the difference between observational “real-life studies” and the randomised controlled clinical trial that we base our evidence on. One possibility is the use of randomised effectiveness trials, which uphold the randomisation but use broad inclusion criteria to represent the true patient population. Until then, careful evaluation of the combined information from trial and natural experiments is warranted. However, other areas of healthcare can be explored in epidemiology. Evaluation of overall healthcare, *e.g.* by studying post-admission mortality and comparing it between hospitals, is increasingly used for quality assurance and improvement and for detecting best practice, as well as areas of increased risk of adverse outcomes [29, 30].

In conclusion, observational data from national databases and other large databases, as well as carefully collected data as part of a cohort study, can be used to assess associations and test hypothesis that may not be addressed in a traditional laboratory or clinical experimental setting.

References

- 1 Lange P, Celli B, Agustí A, *et al.* Lung-function trajectories leading to chronic obstructive pulmonary disease. *N Engl J Med* 2015; 373: 111–122.
- 2 Ezzati M, Obermeyer Z, Tzoulaki I, *et al.* Contributions of risk factors and medical care to cardiovascular mortality trends. *Nat Rev Cardiol* 2015; 12: 508–530.
- 3 Ingebrigtsen TS, Marott JL, Nordestgaard BG, *et al.* Statin use and exacerbations in individuals with chronic obstructive pulmonary disease. *Thorax* 2015; 70: 33–40.

- 4 Doll R, Hill AB. A study of the aetiology of carcinoma of the lung. *BMJ* 1952; 2: 1271–1286.
- 5 Doll R. Proof of causality. Deduction from epidemiological observation. *Perspect Biol Med* 2002; 45: 499–515.
- 6 Fletcher CM, Peto R, Tinker CM, *et al.* The natural history of chronic bronchitis and emphysema. Oxford, Oxford University Press, 1976.
- 7 Fletcher C, Peto R. The natural history of chronic airflow obstruction. *BMJ* 1977; 1: 1645–1648.
- 8 Camilli AE, Burrows B, Knudson RJ, *et al.* Longitudinal changes in forced expiratory volume in one second in adults. *Am Rev Respir Dis* 1987; 135: 794–799.
- 9 Humerfelt S, Gulsvik A, Skjærven R, *et al.* Decline in FEV₁ and airflow limitation related to occupational exposures in men of an urban community. *Eur Respir J* 1993; 6: 1095–1103.
- 10 Vestbo J, Prescott E, Lange P, *et al.* Association of chronic mucus hypersecretion with FEV₁ decline and COPD morbidity. *Am J Respir Crit Care Med* 1996; 153: 1530–1535.
- 11 Lange P, Parner J, Vestbo J, *et al.* A 15-year follow-up of ventilatory function in adults with asthma. *N Engl J Med* 1998; 339: 1194–1200.
- 12 Ramirez-Venegas A, Sansores RH, Quintana-Carrillo RH, *et al.* FEV₁ decline in patients with chronic obstructive pulmonary disease associated with biomass exposure. *Am J Respir Crit Care Med* 2014; 190: 996–1002.
- 13 Sherrill DL, Lebowitz MD, Knudson RJ, *et al.* Smoking and symptom effects on the curves of lung function growth and decline. *Am Rev Respir Dis* 1991; 144: 17–22.
- 14 Gold DR, Wang X, Wypij D, *et al.* Effects of cigarette smoking on the pulmonary function in adolescent boys and girls. *N Engl J Med* 1996; 335: 931–937.
- 15 Svanes C, Sunyer J, Plana E, *et al.* Early life origins of chronic obstructive pulmonary disease. *Thorax* 2010; 65: 14–20.
- 16 Shumway RH, Stuffer DS. Time series analysis and its applications. New York, Springer, 2000.
- 17 Pope CA, Dockery DW, Spengler JD, *et al.* Respiratory health and PM₁₀ pollution: a daily time series analysis. *Am Rev Respir Dis* 1991; 144: 668–674.
- 18 Zmirou D, Schwartz J, Saez M, *et al.* Time-series analysis of air pollution and cause-specific mortality. *Epidemiology* 1998; 9: 495–503.
- 19 Grijalva CG, Nuorti JP, Arbogast PG, *et al.* Decline in pneumonia admissions after routine childhood immunisation with pneumococcal conjugate vaccine in the USA: a time-series analysis. *Lancet* 2007; 369: 1179–1186.
- 20 Nelson HD, Walker M, Zakher B, *et al.* Menopausal hormone therapy for the primary prevention of chronic conditions: a systematic review to update the U.S. Preventive Services Task Force recommendations. *Ann Intern Med* 2012; 157: 104–113.
- 21 Strom BL, Kimmel SE, Hennesy S, eds. Pharmacoepidemiology. 5th Edn. New York, Wiley, 2012.
- 22 Suissa S, Ernst P, Benayoun S, *et al.* Low-dose inhaled corticosteroids and the prevention of death from asthma. *N Engl J Med* 2000; 343: 332–336.
- 23 Herland K, Akselsen JP, Skjonsberg OH, *et al.* How representative are clinical study patients with asthma or COPD for a larger “real life” population of patients with obstructive lung disease? *Respir Med* 2005; 99: 11–19.
- 24 Criner GJ, Connett JE, Aaron SD, *et al.* Simvastatin for the prevention of exacerbations in moderate-to-severe COPD. *N Engl J Med* 2014; 370: 2201–2210.
- 25 Davey SG, Ebrahim S. ‘Mendelian randomization’: can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003; 32: 1–22.
- 26 Dahl M, Zacho J, Vestbo J, *et al.* C-reactive protein and chronic obstructive pulmonary disease: a Mendelian randomisation approach. *Thorax* 2011; 66: 197–204.
- 27 New JP, Bakerly ND, Leather D, *et al.* Obtaining real-world evidence: the Salford Lung Study. *Thorax* 2014; 69: 1152–1154.
- 28 Bakerly N, Woodcock A, New J, *et al.* The Salford Lung Study protocol: a pragmatic, randomised phase III real-world effectiveness trial in chronic obstructive pulmonary disease. *Respir Res* 2015; 16: 101.
- 29 Hernandez AF, Greiner MA, Fonarow GC, *et al.* Relationship between early physician follow-up and 30-day readmission among Medicare beneficiaries hospitalized for heart failure. *JAMA* 2010; 303: 1716–1722.
- 30 Agency for Healthcare Quality and Research. Measure summary. www.qualitymeasures.ahrq.gov/content.aspx?id=49189 Date last updated: September 23, 2015. Date last accessed: April 19, 2016.