BMJ Open

A Simple Design for Assessing Comorbidity Patterns in Disease Natural History

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-012105
Article Type:	Research
Date Submitted by the Author:	01-Apr-2016
Complete List of Authors:	Kiri, Victor; University of Port Harcourt, Pharmaceutical Sciences; FV&JK Consulting Ltd, Epidemiology
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Research methods, Respiratory medicine
Keywords:	Cohort studies, Epidemiological methods, Epidemiology of chronic diseases, Longitudinal Studies, Research Design in Epidemiology

A Simple Design for Assessing Comorbidity Patterns in Disease Natural History

Victor A Kiri, MPH, PhD, CStat, CSci[1-2]

[1] Faculty of Pharmaceutical Sciences, University of Port Harcourt, Choba, Nigeria
[2] FV&JK Consulting Ltd, Guildford, Surrey, UK

Running Title: Comorbidity Patterns in Disease Natural History

Research Institution: GlaxosmithKline R&D (Author was an employee during the research)

Correspondence:
Professor Victor A. Kiri (Visiting Scholar)
Faculty of Pharmaceutical Sciences, University of Port Harcourt, Choba, Port Harcourt, Nigeria
Tel: +44 1483 838717
Fax: +44 1483 838717
E-mail: victor.kiri@fvjkconsult.com

Word count: 2,992 words

Ethics, funding, data sharing: The results used to illustrate the design have been previously presented at conferences by the American Thoracic Society (ATS) and the International Society of PharmacoEpidemiology (ISPE) with the approval of GlaxosmithKline R&D- the sponsor

Abstract

Background: Patients with a chronic disease often suffer from other diseases called comorbidities which can be important factors in the assessment of risks associated with the disease and its management. However, comorbidities can pose important methodological issues because factors such as time, age, duration and the disease can influence their impact on the risk of interest.

Methods: To identify comorbidities of a chronic disease, it is common practice to construct two separate cohorts of patients- a set with the disease and another as a random sample of patients free of the disease- and compare the event rates for each candidate comorbidity over a specific period between the two, whilst accounting for factors which may confound the results. We describe an incidence-based alternative approach that exploits the longitudinal properties of observational databases to track incident event rates along the natural history of the chronic disease. We illustrate it in a retrospective cohort of chronic obstructive pulmonary disease (COPD) patients aged 50 and over- each COPD patient matched to another without COPD on certain confounding factors.

Results: We obtained 24,079 matched pairs. Smoking related chronic conditions in particular such as lung cancer, asthma, other respiratory diseases, fracture and osteoporosis were more common in COPD patients. We also found evidence of time-varying associations.

Conclusion: Our findings in COPD suggest time is an important factor and comorbidity studies which are based on information in a single fixed period (such as first year post diagnosis of COPD) are more likely to report spurious associations.

Keywords: Cohort studies, Epidemiological methods, Epidemiology of chronic diseases, Longitudinal Studies, Research Design in Epidemiology

Article Summary

Strengths

- Explored the longitudinal properties of the data to obtain comparable estimates of incident event rates in each cohort
- Tracked the trend in incident events along the natural history of the disease
- Reduced likelihood of spurious associations compared with the traditional single-point estimation approach which is based on a single observation window

Limitations

- The lack of control for the likely effect of smoking on the results due to the limited scope of information
- The underlying attendance patterns of the patients could affect the probability of diagnosis of the comorbid conditions of interest

Introduction

Comorbidity is defined as any disease which coexists with a chronic disease of interest and the level of comorbid disorders may depend on the chronic disease type. Comorbidities are important for several reasons. Firstly, the safety profile and the potential for adverse effects associated with a given therapy may depend on the extent and severity of pre-existing comorbidities in the particular patient population. Secondly, the effectiveness of the therapy may vary among the patients because its benefits may be affected by the types of pre-existing comorbidities. For instance, there is evidence asthmatic patients, particularly those who also have chronic obstructive pulmonary disease (COPD) have an increased risk of death from causes other than COPD.[1] In such situations, it is clearly clinically relevant to know whether the increased risk is related to the severity of the primary disease, its treatment, or the comorbidity. In general, comorbidity remains an unresolved issue in both the morbidity and mortality of patients living with chronic diseases.

Clinical trial data are generally inadequate for assessing the incidence or prevalence of comorbidities in a particular chronic disease population or for gaining an understanding of their possible implications on the safety and effectiveness of the therapies involved. Since comorbidities may occur more frequently in patients with a particular chronic disease than in those of similar demographic characteristics who are free of the disease, information on the common comorbidities associated with the chronic disease such as background incidence rates can enhance pharmacovigilance and risk management activities, especially for events which may otherwise be falsely classified as safety signals associated with the drug.

Of course, information about comorbidity is also important in clinical practice. In a given chronic disease, such information can influence the quality of life of the patient as well as decisions on treatment.[2-4] There are many examples in pharmacoepidemiological studies where lack of adequate control of the possible influence of comorbidity has resulted in effect estimates confounded by disease severity and other forms of bias.[5-9] Observational databases with rich longitudinal information such as the UK Clinical

Practice Research Datalink (CPRD) and many of the US claims databases as well as those in some EU countries can serve as useful resources for obtaining the incidence and prevalence rates of medical events in patients with a particular chronic disease. In such studies, it is standard practice to compare the estimated rates with those obtained from a control population which often is a random sample of the population that is free of the chronic disease. In most situations, matching on factors such as age and gender which are generally known to influence the type, proportion and impact of comorbidities is often used to facilitate comparability between the two populations as cohorts[10-12]. However, the use of an unmatched control population is not uncommon, despite the risk that by so doing, we may lose the ability to adequately control for confounding factors in our assessment of the association between comorbid events and the chronic disease.

Matching on the propensity scores is a popular approach for handling confounding factors in the assessment of the safety or effectiveness of an intervention in observational studies. However, the methodology may not be appropriate for comorbidity studies of the kind under description as these do not involve any intervention. In this setting, the propensity score becomes the probability of a patient being diagnosed with the chronic disease of interest and as such, in any matched pair, both the patient diagnosed with the disease and his/her counterpart would have the same chance of experiencing events which are associated with the disease. Comorbidities are factors associated with the chronic disease. Thus making adoption of the propensity scores methodology in such studies an avoidable error[12]. Instead, it may be more sensible to use an appropriate sampling strategy to match each patient with the disease to another patient free of the disease on one or two factors identified as potential key confounders such as age and gender in this setting[10-13].

Another important dimension to comorbidity assessment is the role of time which often plays a major role in disease severity. We think its influence can also be assessed by studying the natural history of the disease. Thus, to assess whether a particular comorbid condition is a risk factor for the chronic disease of interest, it may be useful to consider the pattern of the event in relation to the natural history of the chronic disease.

In practice, this can be done by estimating the relevant event rates (i.e. ideally as incident rates) over time such that spans the periods prior to diagnosis of the chronic disease and afterwards. Indeed, the use of incident events in preference over prevalent cases may provide a more incisive insight to the nature of the relationship between the comorbid condition and the chronic disease although the effectiveness of this approach may depend on the number of years for which reliable historical data are available.

In this paper, we will recap the conventional approach for identifying comorbidities which may be associated with any particular chronic disease. We will then describe an innovative incidence-based methodology for identifying patterns of associations between comorbidities and the chronic disease along its natural history which we consider as a more viable alternative. By way of illustration, we will also reproduce some of the results reported elsewhere in a previous application of the new approach in chronic obstructive pulmonary disease (COPD) based on the UK CPRD population (formerly, the GPRD).[14]

Methods

Conventional Approach: usually involves distinct patient populations in a matched cohort design in the following format-

1. One set of patients who have a record of diagnosis or consultation for chronic disease X in an a priori specified calendar year of interest and a random sample of patients who according to their medical records, are free of the disease
2. Both sets are from the same database population with each member also satisfying certain pre-specified inclusion/exclusion study criteria.
3. The date of the diagnosis/consultation for disease X in the specified calendar year- regardless of whether it is a pre-existing or new condition- is taken as the index date and this is also assigned to the matched control so as to ensure same start of follow-up for each pair.
4. Matching is usually on important measurable variables (i.e. likely confounding factors) identified as key to facilitating comparability between the two cohorts[10,12].

Age and gender are the most commonly used factors in this regard. The two cohorts may also be matched on other variables such as the duration of historical records at index-date. Indeed, depending on the primary purpose of the study, the pool of eligible controls for each case may be restricted to only those whose last records span for at least as long as that of the case so as to minimize the impact of between-pair differences in loss to follow-up[12].

Incidence-Based Trend Analytical Approach: This involves a pre-specified study period that spans over a reasonable number of years (i.e. d), instead of the conventional method which either uses a single calendar year to identify patients with chronic disease X or assesses event rates only in the post-diagnostic period. In this sense, the new approach is also different from the incidence-based methodology described elsewhere.[13]

1. The study period consists of two separate phases: an earlier period of duration d1 years for the identification of incident cases of X and a subsequent period of d2 years post diagnosis. The total period for trend analysis is thus d=d1 + d2.
2. Cohort X consists exclusively of patients newly diagnosed with condition X over the study period (i.e. incident diagnosis) and the incident diagnosis date is defined as index date. Patients with any record of diagnosis/consultation for disease X outside of the study period are excluded.
3. Each member of this cohort is then matched to a patient from a random sample of those in the database who are free of disease X in their entire medical history (i.e. X=0). The matched control is assigned the same index date.
4. As in the conventional approach, the matching variables include age and gender.
5. However, unlike the former approach, each case is additionally matched to its control on total completed years of medical records pre and post index date to ensure that the control is followed-up for as long as the case- each having the same duration for the trend analysis.

Indeed, an aspect of the incidence-based approach has been successfully applied to assess the risk of cataract among idiopathic thrombocytopenic purpura patients in the CPRD.[15]

Data analysis: For each year i relative to the index date (i=1, 2, …, d, with i=1 for the earliest observed year) and for each candidate comorbid event k, we estimate the incidence rate per 1,000 person-years (IR_{ik}) for each cohort as well as the corresponding 95% confidence interval in a conditional logistic regression model involving relevant individual characteristic measures as explanatory covariates.[16] We also estimate the rates ratio and its corresponding 95% confidence interval using the conditional logistic regression approach to account for the matching variable, often ignored at some cost in the analysis of matched cohort data.[17-18]

To assess trends in rates ratios along the natural history of X, we fit a linear regression to the annual rate ratios on a logarithmic scale for the candidate comorbid event k and estimate the average annual percentage change over the periods prior to and post index date and separately also for the overall period of evaluation (i.e. d years). The resulting slope of each regression line is assessed for statistical significance.

Application: By way illustration of the new methodology, we have reproduced the details of a previous application in the UK CPRD over a ten-year period in which we evaluated the incident patterns of medical events from a list of candidates thought to have possible associations with COPD.[14] Thus this illustration does not constitute a study of COPD.

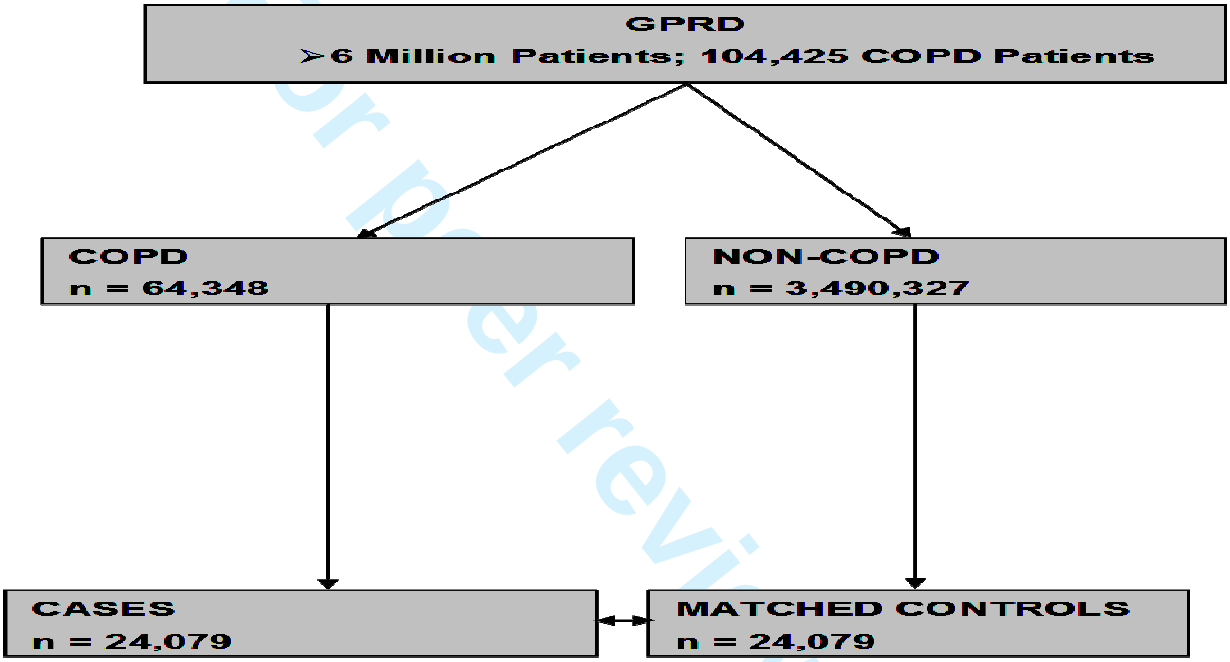
We used a retrospective cohort of patients aged 50+ with a diagnosis of COPD. Each COPD patient was matched to another patient without COPD on year of birth, gender, general practice and completed years of medical records up to at least a year after the index date for COPD between 1990 and 1998, the index date of the COPD patient having been assigned to the matched non-COPD counterpart. We then estimated the annual incidence rates per 1,000 person-years for each event in each cohort over the

ten-year period as well as the corresponding annual rates ratios (RRs) and their 95% confidence intervals such that RR>1 indicates a higher rate in COPD.

Results

A total of 24,079 COPD patients were each matched to a non-COPD patient (Figure 1).

Figure 1: Study design



The annual event rates in COPD and the corresponding annual rates ratios are as shown in Tables 1 and 2 correspondingly.

According to these results, the incidences of many of the smoking related chronic conditions were more common in COPD patients than those free of the disease.[19-20] They were consistently at higher risk of suffering from conditions such as lung cancer, asthma, other respiratory diseases, fracture, osteoporosis, thoracic, mediastinal, cardiac, nervous system and psychiatric disorders as early as several years before diagnosis of COPD. However, we found no evidence of association between COPD and conditions such as pneumonia, glaucoma, ear and labyrinth disorders, reproductive system, breast disorders and vascular diseases other than angina and cardiac

disorders, although there was apparent sign of annual elevation in risk over time for some of the conditions. The pattern for angina was particularly inconsistent in terms of statistical significance- levels were significantly higher in the COPD patients only for the immediate 1-year periods before and after COPD diagnosis- thus highlighting the unreliable nature of methods which rely solely on events in the first year of diagnosis of COPD.[13]

Table 1: Annual incidence rates of certain conditions per 1000 person-years in COPD patients

Events	PRIOR TO COPD DIAGNOSIS					POST COPD DIAGNOSIS				
	Year -5	Year -4	Year -3	Year -2	Year -1	Year 1	Year 2	Year 3	Year 4	Year 5
Lung Cancer	0.51	0.29	0.91	0.51	1.42	4.38	7.34	6.83	6.21	5.18
Asthma	40.15	51.39	64.46	76.25	110.19	118.19	58.44	41.35	41.87	36.83
Pneumonia	3.18	4.78	6.35	7.34	18.54	16.75	22.63	23.65	22.34	23.29
Respiratory Infections	3.21	3.50	2.74	2.37	3.72	4.38	4.82	5.84	6.24	7.30
Other respiratory, thoracic and mediastinal disorder	61.50	91.83	117.82	169.80	289.85	199.11	147.24	130.74	127.90	105.45
Angina	19.71	23.21	24.35	26.86	31.72	31.90	19.53	19.24	15.70	19.86
Cardiac disorders	35.00	48.95	70.88	107.16	250.61	187.28	125.12	113.59	115.89	115.34
Other vascular disorders	36.57	45.04	51.47	57.49	56.79	63.62	52.41	51.43	52.82	48.03
Cataract	10.48	11.68	12.05	14.38	15.07	16.24	18.40	18.18	20.44	16.28
Glaucoma	4.93	5.22	5.29	4.85	5.77	5.58	4.42	5.07	3.80	4.85
Fracture	13.83	12.99	15.62	16.86	15.48	20.59	19.16	20.18	21.64	18.18
Osteoporosis	3.39	4.60	5.95	5.91	6.64	10.18	8.18	10.26	11.46	11.17
Skin Bruises	4.64	4.02	3.91	4.85	4.64	5.91	5.22	6.35	6.53	5.91
Other skin and subcutaneous tissue disorders	52.23	66.72	83.69	98.22	98.00	99.97	93.51	92.71	87.86	83.80
Ear and labyrinth disorders	40.15	45.88	49.57	53.95	54.93	49.68	47.60	45.15	48.55	50.33
Nervous system disorders	42.41	53.18	60.59	70.59	80.37	84.57	79.90	80.23	73.95	76.07
Psychiatric disorders	33.47	39.57	46.25	48.25	53.36	59.31	50.44	42.74	45.08	42.12
Reproductive system and breast disorders	19.35	21.83	18.98	19.27	18.14	16.35	13.72	13.87	13.03	14.24
Social circumstances	7.88	7.12	5.51	5.69	7.41	8.91	9.02	11.86	10.95	14.38

Indeed, we also found evidence of time-varying associations. For example, the annual levels for skin-related events were significantly and consistently higher among COPD patients only after the chronic disease had been diagnosed- thus suggesting possible association with either treatment or severity of COPD or both. It is worthy of note that an assessment based strictly on data in the post COPD diagnosis period would have offered a single conclusion, namely an association between the condition and COPD regardless of severity and treatment.

Table 2: Annual incidence rates ratios of certain conditions per 1000 person-years in COPD and non-COPD patients

Events	PRIOR TO COPD DIAGNOSIS					POST COPD DIAGNOSIS					Annual %change: 5-year Prior	Annual %change: Entire Period
	Year -5	Year -4	Year -3	Year -2	Year -1	Year 1	Year 2	Year 3	Year 4	Year 5		
Lung cancer	4.7	3.9	5.3*	10.7*	16.9*	52.2*	14.3*	10.2*	6.6*	8.2*	42.8#	27.4#
Asthma	3.7*	4.6*	6.7*	8.1*	14.0*	18.9*	12.3*	8.5*	9.7*	7.1*	38.1#	25.0#
Pneumonia	3.8	2.9	3.1	3.2	7.5	7.4	7.4	5.6	8.1	6.1	16.2	21.4#
Respiratory Infections	1.1	1.9*	1.6	1.1	1.4	1.9*	1.8*	1.4	1.2	1.5	-0.2	3.7
Other respiratory, thoracic and mediastinal disorder	1.4*	1.6*	1.8*	2.3*	3.7*	2.8*	2.1*	2.0*	1.9*	1.6*	25.0#	6.6#
Angina	1.2	1.1	1.2*	1.2*	1.6*	1.9*	1.1	1.2*	1.0	1.6*	6.9	2.8
Cardiac disorders	1.2*	1.5*	1.7*	2.2*	4.7*	4.0*	2.6*	2.2*	2.4*	2.3*	35.9#	10.7#
Other vascular disorders	0.9	1.0	1.0	1.0	1.0	1.3*	1.0	1.0	1.0	0.9	1.2	0.3
Cataract	1.2	1.1	1.0	1.3*	1.1	1.2*	1.2	1.1	1.5*	1.3	1.1	2.5#
Glaucoma	1.1	0.7	1.0	0.8	0.9	1.1	1.1	1.2	0.9	1.2	-3.5	1.6
Fracture	1.2	1.1	1.2	1.3*	1.2*	1.4*	1.5*	1.6*	1.5*	1.2	1.5	4.1#
Osteoporosis	1.2	1.7*	1.5*	1.6*	1.8*	2.3*	1.7*	2.4*	2.0*	2.2*	8.0	8.4#
Skin Bruises	1.2	1.1	1.1	1.3	1.1	1.7*	1.2	1.8*	1.9*	1.5	1.1	5.6#
Other skin and subcutaneous tissue	1.0	1.0	1.1*	1.1	1.0	1.2*	1.2*	1.3*	1.2*	1.2*	0.4	2.3#
Ear and labyrinth disorders	1.2*	1.1	1.1	1.0	1.0	1.1	1.1	1.0	1.1	1.1	-4.0	0.3
Nervous system disorders	1.1	1.2*	1.1*	1.1*	1.2*	1.3*	1.3*	1.3*	1.2*	1.3*	1.2	3.1#
Psychiatric disorders	1.1	1.2*	1.3*	1.3*	1.4*	1.8*	1.5*	1.4*	1.6*	1.2	4.8#	3.9#
Reproductive system and breast disorders	0.9	1.0	0.9	1.0	0.9	1.0	1.0	1.0	1.0	1.2	1.0	0.9
Social circumstances	1.1	1.2	1.5*	1.2	1.8*	1.3*	1.2	1.4*	1.2	1.1	11.3	2.0

* Significantly higher rate in COPD patients

# $p < 0.05$ and hence the annual change was significantly different from zero

Discussion

In this paper, we have described the features of an incidence-based methodology for identifying potential comorbid conditions for any particular chronic disease. The methodology exploits the longitudinal properties of observational databases to track incident event rates along the natural history of the chronic disease, as it involves the periods prior to its formal diagnosis and beyond. The results of its application in COPD, as previously described in detail elsewhere, revealed significant time-dependent associations between the chronic disease and certain conditions. We found evidence that in COPD patients, the likelihood of diagnosis of certain comorbid events were highest in the immediate 1-year periods before and after diagnosis of the chronic disease, perhaps due to the diagnostic-related activities experienced by these patients.

If true, then a methodology which relies solely on data in the first year post diagnosis of COPD is much more likely to suggest associations which may be spurious than our approach.

These findings may have interpretational implications on the results of comorbidity studies which are based exclusively on data in the immediate year post diagnosis of any chronic disease of interest. Our results also suggest the trends approach which maintains the longitudinal quality of the data in the assessment of comorbidity associations with a chronic disease, may be more reliable than the traditional single estimate approach. Indeed, the new approach offers a facility for enhancing our understanding of the natural history of the chronic disease in relation to the burden of comorbidity in the management of patients living with the condition. With appropriate data, the method may also be useful to pharmacovigilance activities for any particular of interest, as it offers longitudinal results which may be used to put information from spontaneous reports into an appropriate context. We can do this be done by assessing the incident patterns of the event in two separately matched cohorts of the (1) exposed versus unexposed persons in one and (2) the chronic disease patients versus those free of the disease in the other.

We acknowledge the existence of alternative methods for obtaining matched cohorts in disease natural history studies and we have provided our reasons for excluding the propensity score approach for consideration. In the setting of exploration of possible associations between a chronic disease and comorbidities, we believe the propensity score is exactly the same as the disease risk score- a probability estimate of a patient's likelihood of disease occurrence which has never been used for such disease natural history studies.[21-23] Outside of this setting, we think propensity score matched cohorts could be useful for assessing factors associated with actual clinical practice in a chronic disease- such as the management of such patients in terms of resource utilization independent of other sources of resource use (i.e. confounding factors including comorbidities, among others).

A potential limitation of the new methodology, though common in disease natural studies conducted in general practice databases, is the possibility that the underlying behavior and attendance patterns of the patients at the practices could affect the probability of diagnosis of the events. For example, as COPD patients may have higher rates of doctor consultations than those without COPD (i.e. for routine checks, treatment of acute exacerbations as recommended in guidelines, among many other disease-related reasons), some events may have a higher likelihood of diagnosis in the COPD group.[24] Clearly a notable limitation of the COPD illustration was the lack of control for the likely effect of smoking status which was due to the limited scope of information on smoking in the CPRD at the time of the study. Thus, smoking could indeed account in part for the observed differences between the two groups. Furthermore, the requirement of having at least one year follow-up might also introduce some bias in event estimates because of the possibility of significant differences between the two original cohorts in the proportion of patients with the comorbidities of interest over that period.[13]

The strengths of our methodology include the provision for exploiting the longitudinal properties of observational databases to obtain comparable estimates of event rate ratios as well as the provision for estimating the incidence patterns of such events over time which may facilitate a much clearer understanding of the nature of their associations with the disease.

Acknowledgments:
The author is grateful to Dr Kourtney Davis for her encouragement and is indebted to GlaxoSmithKline R&D for the *excellence in research* award granted him for the development of the incidence-based design whilst he was its employee. Finally, this work is humbly dedicated to the memory of Dr George Visick, a former research colleague at GlaxoSmithKline R&D whose untimely death remains hard to bear.

What is already known on this subject?
The risk of adverse effects associated with a treatment and its effectiveness may depend on the extent and severity of pre-existing comorbidities in the particular disease

To identify comorbidities, it is common practice to compare single point estimates of the rate for each candidate over a specified period between patients with the disease and a random sample of disease-free patients

What this study adds?
Introduced and illustrated a new methodology which tracks the trends of incident events along the natural history of the disease, thereby exploiting the longitudinal properties of observational data in contrast to the conventional approach of single point estimates

The approach facilitates a clearer understanding of the nature of the associations

The findings in COPD suggest the increased likelihood of spurious associations by the single point estimation approach which is based on a single observation window

Contributorship statement: Author conducted the research, including data analysis and writing of the manuscript

Competing interests: Author consults for the pharmaceutical industry on epidemiological methods

Funding: Author was an employee of GlaxoSmithKline Research & Development during the period of the research

Data sharing statement: No additional data available

References

1. Lange P Ulrick CS, Vesbo J. Mortality in adults with self-reported asthma. Copenhagen City Heart Study Group. Lancet. 1996 May 11;347(9011):1285-9
2. Wijnhoven HA, Kriegsman DM, Hesselink AE, de Haan M, Schellevis FG. The influence of co-morbidity on health-related quality of life in asthma and COPD patients. Respir Med 2003; 97: 468-475
3. Eisner MD, Yelin EH, Trupin L, Blanc PD. The influence of chronic respiratory conditions on health status and work disability. Am J Public Health 2002; 92: 1506-1513

4. Iribarren C, Tolstykh IV, Eisner MD. Are patients with asthma at increased risk of coronary heart disease?. Int J Epidemiol 2004;33:743-748
5. Blais L, Ernst P, Suissa A. Confounding by indication and channeling over time: the risks of p -agonists. Am J Epidemiol 1996;144:1161-1169
6. Spitzer WO, Buist AS. Case-control study of prescribed fenoterol and death from asthma in New Zealand. Thorax 1990;45:645-646
7. Moser M, Gifford R. Diuretic therapy and the risk of cardiac arrest. N Engl J Med 1994;331:1235-1236
8. Psaty BM, Heckbert SR, Koepsell TD, et al. The risk of myocardial infarction associated with antihypertensive drug therapies. JAMA 1995;274:620-625
9. Hoes AW, Grobbee DE, Lubsen J, et al. Diuretics, betablockers, and the risk for sudden cardiac death in hypertensive patients. Ann Intern Med 1995;123:481-487
10. Liao P, Yegneswaran B, Vairavanathan S, Zilberman P, Chung F. Postoperative complications in patients with obstructive sleep apnea: a retrospective matched cohort study. Can J Anesth/J Can Anesth 2009; 56:819–828. DOI 10.1007/s12630-009-9190-y
11. Skillrud DM, Offord KP, Miller RD. Higher risk of lung cancer in chronic obstructive pulmonary disease. A prospective, matched, controlled study. Ann Intern Med 1986;105:503-507
12. Hansell AL, Lam KA, Richardson S, et al. Medical event profiling of COPD patients. Pharmacoepidemiol Drug Saf 2004;13:547–555
13. Soriano JB, Visick GT, Muellerova H, Payvandi N, Hansell AL. Patterns of comorbidities in newly diagnosed COPD and asthma in the primary care. Chest 2005; 128:2099–2107
14. Kiri VA, Muellerova H, Visick G. Comorbidity Profiling of COPD patients in the UK Primary Care using an incidence based approach to detect associations with the disease. Am J Respir Crit Care Med 2005; 2:A851
15. Feudjo-Tepie MA, Susan A . Hall SA, John Logie J, Robinson NJ. Risk of cataract among idiopathic thrombocytopenic purpura patients in the UK general

practice research database. pharmacoepidemiol drug safe 2009;18:380–385. DOI: 10.1002/pds.1723

16. Breslow NE, Day NE. Statistical Methods in Cancer Research. Volume 1- The Analysis of Case-Control Studies. IARC 1980;32
17. Sjölander A, Greenland S. Ignoring the matching variables in cohort studies - when is it valid and why? Stat Med. 2013 Jun 12:0. doi: 10.1002/sim.5879
18. Rothman, K., Greenland, S., Lash, T.: Modern Epidemiology. Lippincott Williams & Wilkins 2008;Chpt 11
19. British Thoracic Society Guidelines - COPD. 2007. 9-10-2007. Ref Type: Internet Communication
20. Soriano JB, Maier WC, Egger P, Visick G, Thakrar B, Sykes J, et al. Recent trends in physician diagnosed COPD in women and men in the UK. Thorax 2000;55:789-94
21. Miettinen OS. Stratification by a multivariate confounder score. American Journal of Epidemiology 1976; 104: 609–20
22. Arbogast PG, Ray WA. Use of disease risk scores in pharmacoepidemiologic studies. Stat Methods Med Res 2009; 18; 67 originally published online Jun 18, 2008; DOI: 10.1177/0962280208092347
23. Wyss R, Ellis AR, Brookhart A, Funk MJ, Girman CJ, Simpson RJ Jr and Stürmer T. Matching on the disease risk score in comparative effectiveness research of new treatments. 2015. DOI: 10.1002/pds.3810
24. Fabbri LM, Rabe KF. From COPD to chronic systemic inflammatory syndrome. Lancet 2007;370:797-9

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract [**PAGE 1**]
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found [**PAGE 2**]
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported [**PAGE 3**]
Objectives	3	State specific objectives, including any prespecified hypotheses [**PAGE 6**]
Methods		
Study design	4	Present key elements of study design early in the paper [**PAGE 6**]
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection [**PAGE 7**]
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up [**PAGE 7 as applicable for an illustration**] *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed [**PAGE 7**] *Case-control study*—For matched studies, give matching criteria and the number of controls per case
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable [**PAGE 8 as applicable for an illustration**]
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group [**PAGE 8 as applicable for an illustration**]
Bias	9	Describe any efforts to address potential sources of bias [**PAGE 7 as applicable for an illustration**]
Study size	10	Explain how the study size was arrived at [**Not applicable**]
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why [**PAGE 8 as applicable for an illustration**]

Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding [**PAGE 8 as applicable for an illustration**]
		(*b*) Describe any methods used to examine subgroups and interactions [**Not applicable**]
		(*c*) Explain how missing data were addressed [**Not applicable**]
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed [**PAGE 8 as applicable for an illustration**] *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy
		(*e*) Describe any sensitivity analyses

Continued on next page

Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed [**PAGES 9-11 as applicable for an illustration**]
		(b) Give reasons for non-participation at each stage
		(c) Consider use of a flow diagram [**PAGE 9**]
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders
		(b) Indicate number of participants with missing data for each variable of interest
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount) [**Not applicable for an illustration**]
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time [**PAGES 10-11**]
		Case-control study—Report numbers in each exposure category, or summary measures of exposure
		Cross-sectional study—Report numbers of outcome events or summary measures
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included [**PAGES 10-11**]
		(*b*) Report category boundaries when continuous variables were categorized
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses
Discussion		
Key results	18	Summarise key results with reference to study objectives [**PAGES 12**]
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias [**PAGES 13**]
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence [**PAGES 11-12**]
Generalisability	21	Discuss the generalisability (external validity) of the study results
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based [**PAGE 14**]

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

A Simple Design for Assessing Comorbidity Patterns in Disease Natural History

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-012105.R1
Article Type:	Research
Date Submitted by the Author:	07-Jun-2016
Complete List of Authors:	Kiri, Victor; University of Port Harcourt, Pharmaceutical Sciences; FV&JK Consulting Ltd, Epidemiology
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Research methods, Respiratory medicine
Keywords:	Cohort studies, Epidemiological methods, Epidemiology of chronic diseases, Longitudinal Studies, Research Design in Epidemiology

A Simple Design for Assessing Comorbidity Patterns in Disease Natural History

Victor A Kiri, MPH, PhD, CStat, CSci[1-2]

[1] Faculty of Pharmaceutical Sciences, University of Port Harcourt, Choba, Nigeria
[2] FV&JK Consulting Ltd, Guildford, Surrey, UK

Running Title: Comorbidity Patterns in Disease Natural History

Research Institution: GlaxosmithKline R&D (Author was an employee during the research)

Correspondence:
Professor Victor A. Kiri (Visiting Scholar)
Faculty of Pharmaceutical Sciences, University of Port Harcourt, Choba, Port Harcourt, Nigeria
Tel: +44 1483 838717
Fax: +44 1483 838717
E-mail: victor.kiri@fvjkconsult.com

Word count: 2978 words

Ethics, funding, data sharing: The results used to illustrate the design have been previously presented at conferences by the American Thoracic Society (ATS) and the International Society of PharmacoEpidemiology (ISPE) with the approval of GlaxosmithKline R&D- the sponsor

Abstract

Background: Patients with a chronic disease often suffer from other diseases called comorbidities which can be important factors in the assessment of risks associated with the disease and its management. However, comorbidities can pose important methodological issues because factors such as time, age, duration and the disease can influence their impact on the risk of interest.

Methods: To identify comorbidities of a chronic disease, it is common practice to construct two separate cohorts of patients- a set with the disease and another as a random sample of patients free of the disease- and compare the event rates for each candidate comorbidity over a specific period between the two, whilst accounting for factors which may confound the results. We describe an incidence-based alternative approach that exploits the longitudinal properties of observational databases to track incident event rates along the natural history of the chronic disease. We illustrate it in a retrospective cohort of chronic obstructive pulmonary disease (COPD) patients aged 50 and over- each COPD patient matched to another without COPD on certain confounding factors.

Results: We obtained 24,079 matched pairs. We found that chronic conditions such as lung cancer, asthma, fracture and osteoporosis were more common in COPD patients. We also found evidence of time-varying associations.

Conclusion: Our findings in COPD suggest time is an important factor and comorbidity studies which are based on information in a single fixed period (such as first year post diagnosis of COPD) are more likely to report spurious associations.

Keywords: Cohort studies, Epidemiological methods, Epidemiology of chronic diseases, Longitudinal Studies, Research Design in Epidemiology

Article Summary

Strengths

- Explored the longitudinal properties of the data to obtain comparable estimates of incident event rates in each cohort
- Tracked the trend in incident events along the natural history of the disease
- Reduced likelihood of spurious associations compared with the traditional single-point estimation approach which is based on a single observation window

Limitations

- The lack of control for the likely effect of smoking on the results due to the limited scope of information
- The underlying attendance patterns of the patients could affect the probability of diagnosis of the comorbid conditions of interest

Introduction

Comorbidity is defined as any disease which coexists with a chronic disease of interest and the level of comorbid disorders may depend on the chronic disease type. Comorbidities are important for several reasons. Firstly, the safety profile and the potential for adverse effects associated with a given therapy may depend on the extent and severity of pre-existing comorbidities in the particular patient population. Secondly, the effectiveness of the therapy may vary among the patients because its benefits may be affected by the types of pre-existing comorbidities. For instance, there is evidence asthmatic patients, particularly those who also have chronic obstructive pulmonary disease (COPD) have an increased risk of death from causes other than COPD.[1] In such situations, it is clearly clinically relevant to know whether the increased risk is related to the severity of the primary disease, its treatment, or the comorbidity. In general, comorbidity remains an unresolved issue in both the morbidity and mortality of patients living with chronic diseases.

Since comorbidities may occur more frequently in patients with a particular chronic disease than in those of similar demographic characteristics who are free of the disease, information on the common comorbidities associated with the chronic disease such as background incidence rates can enhance pharmacovigilance and risk management activities, especially for events which may otherwise be falsely classified as safety signals associated with the drug.

Of course, information about comorbidity is also important in clinical practice. In a given chronic disease, such information can influence the quality of life of the patient as well as decisions on treatment.[2-4] There are many examples in pharmacoepidemiological studies where lack of adequate control of the possible influence of comorbidity has resulted in effect estimates confounded by disease severity and other forms of bias.[5-9] Observational databases with rich longitudinal information such as the UK Clinical Practice Research Datalink (CPRD) and many of the US claims databases as well as those in some EU countries can serve as useful resources for obtaining the incidence and prevalence rates of medical events in patients with a particular chronic disease. In

such studies, it is standard practice to compare the estimated rates with those obtained from a control population which often is a random sample of the population that is free of the chronic disease. In most situations, matching on factors such as age and gender which are generally known to influence the type, proportion and impact of comorbidities is often used to facilitate comparability between the two populations as cohorts[10-12]. However, the use of an unmatched control population is not uncommon, despite the risk that by so doing, we may lose the ability to adequately control for confounding factors in our assessment of the association between comorbid events and the chronic disease.

Matching on the propensity scores is a popular approach for handling confounding factors in the assessment of the safety or effectiveness of an intervention in observational studies. However, the methodology may not be appropriate for comorbidity studies of the kind under description as these do not involve any intervention. In this setting, the propensity score becomes the probability of a patient being diagnosed with the chronic disease of interest and as such, in any matched pair, both the patient diagnosed with the disease and his/her counterpart would have the same chance of experiencing events which are associated with the disease. Comorbidities are factors associated with the chronic disease. Thus making adoption of the propensity scores methodology in such studies an avoidable error[12]. Instead, it may be more sensible to use an appropriate sampling strategy to match each patient with the disease to another patient free of the disease on one or two factors identified as potential key confounders such as age and gender in this setting[10-13].

Another important dimension to comorbidity assessment is the role of time which often plays a major role in disease severity. We think its influence can also be assessed by studying the natural history of the disease. Thus, to assess whether a particular comorbid condition is a risk factor for the chronic disease of interest, it may be useful to consider the pattern of the event in relation to the natural history of the chronic disease. In practice, this can be done by estimating the relevant event rates (i.e. ideally as incident rates) over time such that spans the periods prior to diagnosis of the chronic disease and afterwards. Indeed, the use of incident events in preference over prevalent

cases may provide a more incisive insight to the nature of the relationship between the comorbid condition and the chronic disease although the effectiveness of this approach may depend on the number of years for which reliable historical data are available.

In this paper, we will recap the conventional approach for identifying comorbidities which may be associated with any particular chronic disease. We will then describe an innovative incidence-based methodology for identifying patterns of associations between comorbidities and the chronic disease along its natural history which we consider as a more viable alternative. By way of illustration, we will also reproduce some of the results reported elsewhere in a previous application of the new approach in chronic obstructive pulmonary disease (COPD) based on the UK CPRD population (formerly, the GPRD).[14]

Methods

Conventional Approach: usually involves distinct patient populations in a matched cohort design in the following format-

1. One set of patients who have a record of diagnosis or consultation for chronic disease X in an a priori specified calendar year of interest and a random sample of patients who according to their medical records, are free of the disease
2. Both sets are from the same database population with each member also satisfying certain pre-specified inclusion/exclusion study criteria.
3. The date of the diagnosis/consultation for disease X in the specified calendar year- regardless of whether it is a pre-existing or new condition- is taken as the index date and this is also assigned to the matched control so as to ensure same start of follow-up for each pair.
4. Matching is usually on important measurable variables (i.e. likely confounding factors) identified as key to facilitating comparability between the two cohorts[10,12].

Age and gender are the most commonly used factors in this regard. The two cohorts may also be matched on other variables such as the duration of historical records at index-date. Indeed, depending on the primary purpose of the study, the pool of eligible

controls for each case may be restricted to only those whose last records span for at least as long as that of the case so as to minimize the impact of between-pair differences in loss to follow-up[12].

Incidence-Based Trend Analytical Approach: This involves a pre-specified study period that spans over a reasonable number of years (i.e. d), instead of the conventional method which either uses a single calendar year to identify patients with chronic disease X or assesses event rates only in the post-diagnostic period. In this sense, the new approach is also different from the incidence-based methodology described elsewhere.[13]

1. The study period consists of two separate phases: an earlier period of duration d1 years for the identification of incident cases of X and a subsequent period of d2 years post diagnosis. The total period for trend analysis is thus d=d1 + d2.
2. Cohort X consists exclusively of patients newly diagnosed with condition X over the study period (i.e. incident diagnosis) and the incident diagnosis date is defined as index date. Patients with any record of diagnosis/consultation for disease X outside of the study period are excluded.
3. Each member of this cohort is then matched to a patient from a random sample of those in the database who are free of disease X in their entire medical history (i.e. X=0). The matched control is assigned the same index date.
4. As in the conventional approach, the matching variables include age and gender.
5. However, unlike the former approach, each case is additionally matched to its control on total completed years of medical records pre and post index date to ensure that the control is followed-up for as long as the case- each having the same duration for the trend analysis.

Indeed, an aspect of the incidence-based approach has been successfully applied to assess the risk of cataract among idiopathic thrombocytopenic purpura patients in the CPRD.[15]

Data analysis: For each year i relative to the index date (i=1, 2, …, d, with i=1 for the earliest observed year) and for each candidate comorbid event k, we estimate the incidence rate per 1,000 person-years (IR_{ik}) for each cohort as well as the corresponding 95% confidence interval in a conditional logistic regression model involving relevant individual characteristic measures as explanatory covariates.[16] We also estimate the rates ratio and its corresponding 95% confidence interval using the conditional logistic regression approach to account for the matching variable, often ignored at some cost in the analysis of matched cohort data.[17-18]

To assess trends in rates ratios along the natural history of X, we fit a linear regression to the annual rate ratios on a logarithmic scale for the candidate comorbid event k and estimate the average annual percentage change over the periods prior to and post index date and separately also for the overall period of evaluation (i.e. d years). The resulting slope of each regression line is assessed for statistical significance.

Application: By way illustration of the new methodology, we have reproduced the details of a previous application in the UK CPRD over a ten-year period in which we evaluated the incident patterns of medical events from a list of candidates thought to have possible associations with COPD.[12-14] Thus this illustration does not constitute a study of COPD.

We used a retrospective cohort of patients aged 50+ with a diagnosis of COPD. Each COPD patient was matched to another patient without COPD on year of birth, gender, general practice and completed years of medical records up to at least a year after the index date for COPD between 1990 and 1998, the index date of the COPD patient having been assigned to the matched non-COPD counterpart. We then estimated the annual incidence rates per 1,000 person-years for each event in each cohort over the ten-year period as well as the corresponding annual rates ratios (RRs) and their 95% confidence intervals such that RR>1 indicates a higher rate in COPD. The age group is same as in the previously reported COPD studies conducted on the database.[12-14]

Results

A total of 24,079 COPD patients were each matched to a non-COPD patient (Figure 1).

The annual event rates in COPD and the corresponding annual rates ratios are as shown in Tables 1 and 2 correspondingly.

According to these results, the incidences of many of the smoking related chronic conditions were more common in COPD patients than those free of the disease.[19-20] They were consistently at higher risk of suffering from conditions such as lung cancer, asthma, other respiratory diseases, fracture, osteoporosis, thoracic, mediastinal, cardiac, nervous system and psychiatric disorders as early as several years before diagnosis of COPD. However, we found no evidence of association between COPD and conditions such as pneumonia, glaucoma, ear and labyrinth disorders, reproductive system, breast disorders and vascular diseases other than angina and cardiac disorders, although there was apparent sign of annual elevation in risk over time for some of the conditions. The pattern for angina was particularly inconsistent in terms of statistical significance- levels were significantly higher in the COPD patients only for the immediate 1-year periods before and after COPD diagnosis- thus highlighting the unreliable nature of methods which rely solely on events in the first year of diagnosis of COPD.[13]

Table 1. Annual incidence rates of certain conditions per 1000 person-years in COPD patients*

CANDIDATE CONDITIONS	PRIOR TO COPD DIAGNOSIS					POST COPD DIAGNOSIS				
	Year -5	Year -4	Year - 3	Year - 2	Year - 1	Year 1	Year 2	Year 3	Year 4	Year 5
Lung cancer	0.51	0.29	0.91	0.51	1.42	4.38	7.34	6.83	6.21	5.18
Asthma	40.15	51.39	64.46	76.25	110.19	118.19	58.44	41.35	41.87	36.83
Pneumonia	3.18	4.78	6.35	7.34	18.54	16.75	22.63	23.65	22.34	23.29
Respiratory Infections	3.21	3.50	2.74	2.37	3.72	4.38	4.82	5.84	6.24	7.30
Other respiratory, thoracic and mediastinal disorder	61.50	91.83	117.82	169.80	289.85	199.11	147.24	130.74	127.90	105.45
Angina	19.71	23.21	24.35	26.86	31.72	31.90	19.53	19.24	15.70	19.86
Cardiac disorders	35.00	48.95	70.88	107.16	250.61	187.28	125.12	113.59	115.89	115.34
Other vascular disorders	36.57	45.04	51.47	57.49	56.79	63.62	52.41	51.43	52.82	48.03
Cataract	10.48	11.68	12.05	14.38	15.07	16.24	18.40	18.18	20.44	16.28
Glaucoma	4.93	5.22	5.29	4.85	5.77	5.58	4.42	5.07	3.80	4.85
Fracture	13.83	12.99	15.62	16.86	15.48	20.59	19.16	20.18	21.64	18.18
Osteoporosis	3.39	4.60	5.95	5.91	6.64	10.18	8.18	10.26	11.46	11.17
Skin Bruises	4.64	4.02	3.91	4.85	4.64	5.91	5.22	6.35	6.53	5.91
Other skin and subcutaneous tissue disorders	52.23	66.72	83.69	98.22	98.00	99.97	93.51	92.71	87.86	83.80
Ear and labyrinth disorders	40.15	45.88	49.57	53.95	54.93	49.68	47.60	45.15	48.55	50.33
Nervous system disorders	42.41	53.18	60.59	70.59	80.37	84.57	79.90	80.23	73.95	76.07
Psychiatric disorders	33.47	39.57	46.25	48.25	53.36	59.31	50.44	42.74	45.08	42.12
Reproductive system and breast disorders	19.35	21.83	18.98	19.27	18.14	16.35	13.72	13.87	13.03	14.24
Social circumstances	7.88	7.12	5.51	5.69	7.41	8.91	9.02	11.86	10.95	14.38

* Reproduced from Kiri et al. (2005); See Hansell et al. (2005) and Soriano et al. (2005) for details of the events that make up the candidate conditions

Indeed, we also found evidence of time-varying associations. For example, the annual levels for skin-related events were significantly and consistently higher among COPD patients only after the chronic disease had been diagnosed- thus suggesting possible association with either treatment or severity of COPD or both. It is worthy of note that an assessment based strictly on data in the post COPD diagnosis period would have

offered a single conclusion, namely an association between the condition and COPD regardless of severity and treatment.

Table 2. Annual incidence rates ratios of certain conditions per 1000 person-years in COPD and non-COPD patients^

CANDIDATE CONDITIONS	PRIOR TO COPD DIAGNOSIS					POST COPD DIAGNOSIS						
	Year -5	Year -4	Year -3	Year -2	Year -1	Year 1	Year 2	Year 3	Year 4	Year 5	Annual %change: 5-year Prior	Annual %change: Entire Period
Lung cancer	4.7	3.9	5.3*	10.7*	16.9*	52.2*	14.3*	10.2*	6.6*	8.2*	42.8#	27.4#
Asthma	3.7*	4.6*	6.7*	8.1*	14.0*	18.9*	12.3*	8.5*	9.7*	7.1*	38.1#	25.0#
Pneumonia	3.8	2.9	3.1	3.2	7.5	7.4	7.4	5.6	8.1	6.1	16.2	21.4#
Respiratory Infections	1.1	1.9*	1.6	1.1	1.4	1.9*	1.8*	1.4	1.2	1.5	-0.2	3.7
Other respiratory, thoracic and mediastinal disorder	1.4*	1.6*	1.8*	2.3*	3.7*	2.8*	2.1*	2.0*	1.9*	1.6*	25.0#	6.6#
Angina	1.2	1.1	1.2*	1.2*	1.6*	1.9*	1.1	1.2*	1.0	1.6*	6.9	2.8
Cardiac disorders	1.2*	1.5*	1.7*	2.2*	4.7*	4.0*	2.6*	2.2*	2.4*	2.3*	35.9#	10.7#
Other vascular disorders	0.9	1.0	1.0	1.0	1.0	1.3*	1.0	1.0	1.0	0.9	1.2	0.3
Cataract	1.2	1.1	1.0	1.3*	1.1	1.2*	1.2	1.1	1.5*	1.3	1.1	2.5#
Glaucoma	1.1	0.7	1.0	0.8	0.9	1.1	1.1	1.2	0.9	1.2	-3.5	1.6
Fracture	1.2	1.1	1.2	1.3*	1.2*	1.4*	1.5*	1.6*	1.5*	1.2	1.5	4.1#
Osteoporosis	1.2	1.7*	1.5*	1.6*	1.8*	2.3*	1.7*	2.4*	2.0*	2.2*	8.0	8.4#
Skin Bruises	1.2	1.1	1.1	1.3	1.1	1.7*	1.2	1.8*	1.9*	1.5	1.1	5.6#
Other skin and subcutaneous tissue disorders	1.0	1.0	1.1*	1.1	1.0	1.2*	1.2*	1.3*	1.2*	1.2*	0.4	2.3#
Ear and labyrinth disorders	1.2*	1.1	1.1	1.0	1.0	1.1	1.1	1.0	1.1	1.1	-4.0	0.3
Nervous system disorders	1.1	1.2*	1.1*	1.1*	1.2*	1.3*	1.3*	1.3*	1.2*	1.3*	1.2	3.1#
Psychiatric disorders	1.1	1.2*	1.3*	1.3*	1.4*	1.8*	1.5*	1.4*	1.6*	1.2	4.8#	3.9#
Reproductive system and breast disorders	0.9	1.0	0.9	1.0	0.9	1.0	1.0	1.0	1.0	1.2	1.0	0.9
Social circumstances	1.1	1.2	1.5*	1.2	1.8*	1.3*	1.2	1.4*	1.2	1.1	11.3	2.0

* Significantly higher rate in COPD patients
# $p < 0.05$ and hence the annual change was significantly different from zero
^ Reproduced from Kiri et al. (2005)

Discussion

In this paper, we have described the features of an incidence-based methodology for identifying potential comorbid conditions for any particular chronic disease. The methodology exploits the longitudinal properties of observational databases to track incident event rates along the natural history of the chronic disease, as it involves the periods prior to its formal diagnosis and beyond. The results of its application in COPD, as previously described in detail elsewhere, revealed significant time-dependent associations between the chronic disease and certain conditions. We found evidence that in COPD patients, the likelihood of diagnosis of certain comorbid events were highest in the immediate 1-year periods before and after diagnosis of the chronic disease, perhaps due to the diagnostic-related activities experienced by these patients. If true, then a methodology which relies solely on data in the first year post diagnosis of COPD is much more likely to suggest associations which may be spurious than our approach.

These findings may have interpretational implications on the results of comorbidity studies which are based exclusively on data in the immediate year post diagnosis of any chronic disease of interest. Our results also suggest the trends approach which maintains the longitudinal quality of the data in the assessment of comorbidity associations with a chronic disease, may be more reliable than the traditional single estimate approach. Indeed, the new approach offers a facility for enhancing our understanding of the natural history of the chronic disease in relation to the burden of comorbidity in the management of patients living with the condition. With appropriate data, the method may also be useful to pharmacovigilance activities for any particular of interest, as it offers longitudinal results which may be used to put information from spontaneous reports into an appropriate context. We can do this be done by assessing the incident patterns of the event in two separately matched cohorts of the (1) exposed versus unexposed persons in one and (2) the chronic disease patients versus those free of the disease in the other.

We acknowledge the existence of alternative methods for obtaining matched cohorts in disease natural history studies and we have provided our reasons for excluding the propensity score approach for consideration. In the setting of exploration of possible associations between a chronic disease and comorbidities, we believe the propensity score is exactly the same as the disease risk score- a probability estimate of a patient's likelihood of disease occurrence which has never been used for such disease natural history studies.[21-23] Outside of this setting, we think propensity score matched cohorts could be useful for assessing factors associated with actual clinical practice in a chronic disease- such as the management of such patients in terms of resource utilization independent of other sources of resource use (i.e. confounding factors including comorbidities, among others).

A potential limitation of the new methodology, though common in disease natural studies conducted in general practice databases, is the possibility that the underlying behavior and attendance patterns of the patients at the practices could affect the probability of diagnosis of the events. For example, as COPD patients may have higher rates of doctor consultations than those without COPD (i.e. for routine checks, treatment of acute exacerbations as recommended in guidelines, among many other disease-related reasons), some events may have a higher likelihood of diagnosis in the COPD group.[24] Clearly a notable limitation of the COPD illustration was the lack of control for the likely effect of smoking status which was due to the limited scope of information on smoking in the CPRD at the time of the study. Thus, smoking could indeed account in part for the observed differences between the two groups. Furthermore, the requirement of having at least one year follow-up might also introduce some bias in event estimates because of the possibility of significant differences between the two original cohorts in the proportion of patients with the comorbidities of interest over that period.[13]

The strengths of our methodology include the provision for exploiting the longitudinal properties of observational databases to obtain comparable estimates of event rate ratios as well as the provision for estimating the incidence patterns of such events over

time which may facilitate a much clearer understanding of the nature of their associations with the disease.

Acknowledgments:

The author is grateful to Dr Kourtney Davis for her encouragement and is indebted to GlaxoSmithKline R&D for the *excellence in research* award granted him for the development of the incidence-based design whilst he was its employee. The author is also grateful to the reviewers for their useful suggestions. Finally, this work is humbly dedicated to the memory of Dr George Visick, a former research colleague at GlaxoSmithKline R&D whose untimely death remains hard to bear.

What is already known on this subject?
The risk of adverse effects associated with a treatment and its effectiveness may depend on the extent and severity of pre-existing comorbidities in the particular disease

To identify comorbidities, it is common practice to compare single point estimates of the rate for each candidate over a specified period between patients with the disease and a random sample of disease-free patients

What this study adds?
Introduced and illustrated a new methodology which tracks the trends of incident events along the natural history of the disease, thereby exploiting the longitudinal properties of observational data in contrast to the conventional approach of single point estimates

The approach facilitates a clearer understanding of the nature of the associations

The findings in COPD suggest the increased likelihood of spurious associations by the single point estimation approach which is based on a single observation window

Contributorship statement: Author conducted the research, including data analysis and writing of the manuscript

Competing interests: Author consults for the pharmaceutical industry on epidemiological methods

Funding: Author was an employee of GlaxoSmithKline Research & Development during the period of the research

Data sharing statement: No additional data available

References

1. Lange P Ulrick CS, Vesbo J. Mortality in adults with self-reported asthma. Copenhagen City Heart Study Group. Lancet. 1996 May 11;347(9011):1285-9
2. Wijnhoven HA, Kriegsman DM, Hesselink AE, de Haan M, Schellevis FG. The influence of co-morbidity on health-related quality of life in asthma and COPD patients. Respir Med 2003; 97: 468-475
3. Eisner MD, Yelin EH, Trupin L, Blanc PD. The influence of chronic respiratory conditions on health status and work disability. Am J Public Health 2002; 92: 1506-1513

4. Iribarren C, Tolstykh IV, Eisner MD. Are patients with asthma at increased risk of coronary heart disease?. Int J Epidemiol 2004;33:743-748
5. Blais L, Ernst P, Suissa A. Confounding by indication and channeling over time: the risks of p -agonists. Am J Epidemiol 1996;144:1161-1169
6. Spitzer WO, Buist AS. Case-control study of prescribed fenoterol and death from asthma in New Zealand. Thorax 1990;45:645-646
7. Moser M, Gifford R. Diuretic therapy and the risk of cardiac arrest. N Engl J Med 1994;331:1235-1236
8. Psaty BM, Heckbert SR, Koepsell TD, et al. The risk of myocardial infarction associated with antihypertensive drug therapies. JAMA 1995;274:620-625
9. Hoes AW, Grobbee DE, Lubsen J, et al. Diuretics, betablockers, and the risk for sudden cardiac death in hypertensive patients. Ann Intern Med 1995;123:481-487
10. Liao P, Yegneswaran B, Vairavanathan S, Zilberman P, Chung F. Postoperative complications in patients with obstructive sleep apnea: a retrospective matched cohort study. Can J Anesth/J Can Anesth 2009; 56:819–828. DOI 10.1007/s12630-009-9190-y
11. Skillrud DM, Offord KP, Miller RD. Higher risk of lung cancer in chronic obstructive pulmonary disease. A prospective, matched, controlled study. Ann Intern Med 1986;105:503-507
12. Hansell AL, Lam KA, Richardson S, et al. Medical event profiling of COPD patients. Pharmacoepidemiol Drug Saf 2004;13:547–555
13. Soriano JB, Visick GT, Muellerova H, Payvandi N, Hansell AL. Patterns of comorbidities in newly diagnosed COPD and asthma in the primary care. Chest 2005; 128:2099–2107
14. Kiri VA, Muellerova H, Visick G. Comorbidity Profiling of COPD patients in the UK Primary Care using an incidence based approach to detect associations with the disease. Am J Respir Crit Care Med 2005; 2:A851
15. Feudjo-Tepie MA, Susan A . Hall SA, John Logie J, Robinson NJ. Risk of cataract among idiopathic thrombocytopenic purpura patients in the UK general

practice research database. pharmacoepidemiol drug safe 2009;18:380–385. DOI: 10.1002/pds.1723

16. Breslow NE, Day NE. Statistical Methods in Cancer Research. Volume 1- The Analysis of Case-Control Studies. IARC 1980;32
17. Sjölander A, Greenland S. Ignoring the matching variables in cohort studies - when is it valid and why? Stat Med. 2013 Jun 12:0. doi: 10.1002/sim.5879
18. Rothman, K., Greenland, S., Lash, T.: Modern Epidemiology. Lippincott Williams & Wilkins 2008;Chpt 11
19. British Thoracic Society Guidelines - COPD. 2007. 9-10-2007. Ref Type: Internet Communication
20. Soriano JB, Maier WC, Egger P, Visick G, Thakrar B, Sykes J, et al. Recent trends in physician diagnosed COPD in women and men in the UK. Thorax 2000;55:789-94
21. Miettinen OS. Stratification by a multivariate confounder score. American Journal of Epidemiology 1976; 104: 609–20
22. Arbogast PG, Ray WA. Use of disease risk scores in pharmacoepidemiologic studies. Stat Methods Med Res 2009; 18; 67 originally published online Jun 18, 2008; DOI: 10.1177/0962280208092347
23. Wyss R, Ellis AR, Brookhart A, Funk MJ, Girman CJ, Simpson RJ Jr and Stürmer T. Matching on the disease risk score in comparative effectiveness research of new treatments. 2015. DOI: 10.1002/pds.3810
24. Fabbri LM, Rabe KF. From COPD to chronic systemic inflammatory syndrome. Lancet 2007;370:797-9

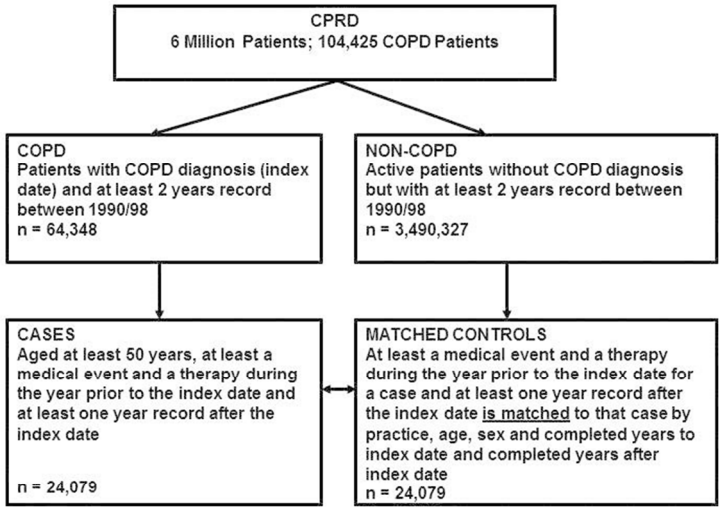


Figure 1. Selection of COPD incident cases and controls from the Clinical Practice Research Database (CPRD)

81x60mm (300 x 300 DPI)

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract [**PAGE 1**]
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found [**PAGE 2**]
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported [**PAGE 3**]
Objectives	3	State specific objectives, including any prespecified hypotheses [**PAGE 6**]
Methods		
Study design	4	Present key elements of study design early in the paper [**PAGE 6**]
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection [**PAGE 7**]
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up [**PAGE 7 as applicable for an illustration**] *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed [**PAGE 7**] *Case-control study*—For matched studies, give matching criteria and the number of controls per case
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable [**PAGE 8 as applicable for an illustration**]
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group [**PAGE 8 as applicable for an illustration**]
Bias	9	Describe any efforts to address potential sources of bias [**PAGE 7 as applicable for an illustration**]
Study size	10	Explain how the study size was arrived at [**Not applicable**]
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why [**PAGE 8 as applicable for an illustration**]

Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding [**PAGE 8 as applicable for an illustration**]
		(*b*) Describe any methods used to examine subgroups and interactions [**Not applicable**]
		(*c*) Explain how missing data were addressed [**Not applicable**]
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed [**PAGE 8 as applicable for an illustration**] *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy
		(*e*) Describe any sensitivity analyses

Continued on next page

Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed [**PAGES 9-11 as applicable for an illustration**]
		(b) Give reasons for non-participation at each stage
		(c) Consider use of a flow diagram [**PAGE 9**]
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders
		(b) Indicate number of participants with missing data for each variable of interest
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount) [**Not applicable for an illustration**]
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time [**PAGES 10-11**]
		Case-control study—Report numbers in each exposure category, or summary measures of exposure
		Cross-sectional study—Report numbers of outcome events or summary measures
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included [**PAGES 10-11**]
		(*b*) Report category boundaries when continuous variables were categorized
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses
Discussion		
Key results	18	Summarise key results with reference to study objectives [**PAGES 12**]
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias [**PAGES 13**]
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence [**PAGES 11-12**]
Generalisability	21	Discuss the generalisability (external validity) of the study results
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based [**PAGE 14**]

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

A Simple Design for Assessing Comorbidity Patterns in Disease Natural History

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-012105.R2
Article Type:	Research
Date Submitted by the Author:	16-Jun-2016
Complete List of Authors:	Kiri, Victor; University of Port Harcourt, Pharmaceutical Sciences; FV&JK Consulting Ltd, Epidemiology
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Research methods, Respiratory medicine
Keywords:	Cohort studies, Epidemiological methods, Epidemiology of chronic diseases, Longitudinal Studies, Research Design in Epidemiology

A Simple Design for Assessing Comorbidity Patterns in Disease Natural History

Victor A Kiri, MPH, PhD, CStat, CSci[1-2]

[1] Faculty of Pharmaceutical Sciences, University of Port Harcourt, Choba, Nigeria
[2] FV&JK Consulting Ltd, Guildford, Surrey, UK

Running Title: Comorbidity Patterns in Disease Natural History

Research Institution: GlaxosmithKline R&D (Author was an employee during the research)

Correspondence:
Professor Victor A. Kiri (Visiting Scholar)
Faculty of Pharmaceutical Sciences, University of Port Harcourt, Choba, Port Harcourt, Nigeria
Tel: +44 1483 838717
Fax: +44 1483 838717
E-mail: victor.kiri@fvjkconsult.com

Word count: 2969 words

Ethics, funding, data sharing: The results used to illustrate the design have been previously presented at conferences by the American Thoracic Society (ATS) and the International Society of PharmacoEpidemiology (ISPE) with the approval of GlaxosmithKline R&D- the sponsor

Abstract

Background: Patients with a chronic disease often suffer from other diseases called comorbidities which can be important factors in the assessment of risks associated with the disease and its management. However, comorbidities can pose important methodological issues because factors such as time, age, duration and the disease can influence their impact on the risk of interest.

Methods: To identify comorbidities of a chronic disease, it is common practice to construct two separate cohorts of patients- a set with the disease and another as a random sample of patients free of the disease- and compare the event rates for each candidate comorbidity over a specific period between the two, whilst accounting for factors which may confound the results. We describe an incidence-based alternative approach that exploits the longitudinal properties of observational databases to track incident event rates along the natural history of the chronic disease. We illustrate it in a retrospective cohort of chronic obstructive pulmonary disease (COPD) patients aged 50 and over- each COPD patient matched to another without COPD on certain confounding factors.

Results: We obtained 24,079 matched pairs. We found that chronic conditions such as lung cancer, asthma, fracture and osteoporosis were more common in COPD patients. We also found evidence of time-varying associations.

Conclusion: Our findings in COPD suggest time is an important factor and comorbidity studies which are based on information in a single fixed period (such as first year post diagnosis of COPD) are more likely to report spurious associations.

Keywords: Cohort studies, Epidemiological methods, Epidemiology of chronic diseases, Longitudinal Studies, Research Design in Epidemiology

Article Summary

Strengths

- Explored the longitudinal properties of the data to obtain comparable estimates of incident event rates in each cohort
- Tracked the trend in incident events along the natural history of the disease
- Reduced likelihood of spurious associations compared with the traditional single-point estimation approach which is based on a single observation window

Limitations

- The lack of control for the likely effect of smoking on the results due to the limited scope of information
- The underlying attendance patterns of the patients could affect the probability of diagnosis of the comorbid conditions of interest

Introduction

Comorbidity is defined as any disease which coexists with a chronic disease of interest and the level of comorbid disorders may depend on the chronic disease type. Comorbidities are important for several reasons. Firstly, the safety profile and the potential for adverse effects associated with a given therapy may depend on the extent and severity of pre-existing comorbidities in the particular patient population. Secondly, the effectiveness of the therapy may vary among the patients because its benefits may be affected by the types of pre-existing comorbidities. For instance, there is evidence asthmatic patients, particularly those who also have chronic obstructive pulmonary disease (COPD) have an increased risk of death from causes other than COPD.[1] In such situations, it is clearly clinically relevant to know whether the increased risk is related to the severity of the primary disease, its treatment, or the comorbidity. In general, comorbidity remains an unresolved issue in both the morbidity and mortality of patients living with chronic diseases.

Since comorbidities may occur more frequently in patients with a particular chronic disease than in those of similar demographic characteristics who are free of the disease, information on the common comorbidities associated with the chronic disease such as background incidence rates can enhance pharmacovigilance and risk management activities, especially for events which may otherwise be falsely classified as safety signals associated with the drug.

Of course, information about comorbidity is also important in clinical practice. In a given chronic disease, such information can influence the quality of life of the patient as well as decisions on treatment.[2-4] There are many examples in pharmacoepidemiological studies where lack of adequate control of the possible influence of comorbidity has resulted in effect estimates confounded by disease severity and other forms of bias.[5-9] Observational databases with rich longitudinal information such as the UK Clinical Practice Research Datalink (CPRD) and many of the US claims databases as well as those in some EU countries can serve as useful resources for obtaining the incidence and prevalence rates of medical events in patients with a particular chronic disease. In

such studies, it is standard practice to compare the estimated rates with those obtained from a control population which often is a random sample of the population that is free of the chronic disease. In most situations, matching on factors such as age and gender which are generally known to influence the type, proportion and impact of comorbidities is often used to facilitate comparability between the two populations as cohorts[10-12]. However, the use of an unmatched control population is not uncommon, despite the risk that by so doing, we may lose the ability to adequately control for confounding factors in our assessment of the association between comorbid events and the chronic disease.

Matching on the propensity scores is a popular approach for handling confounding factors in the assessment of the safety or effectiveness of an intervention in observational studies. However, the methodology may not be appropriate for comorbidity studies of the kind under description as these do not involve any intervention. In this setting, the propensity score becomes the probability of a patient being diagnosed with the chronic disease of interest and as such, in any matched pair, both the patient diagnosed with the disease and his/her counterpart would have the same chance of experiencing events which are associated with the disease. Comorbidities are factors associated with the chronic disease- making adoption of the propensity scores methodology in such studies an avoidable error[12]. Instead, it may be more sensible to use an appropriate sampling strategy to match each patient with the disease to another patient free of the disease on one or two factors identified as potential key confounders such as age and gender in this setting[10-13].

Another important dimension to comorbidity assessment is the role of time which often plays a major role in disease severity. We think its influence can also be assessed by studying the natural history of the disease. Thus, to assess whether a particular comorbid condition is a risk factor for the chronic disease of interest, it may be useful to consider the pattern of the event in relation to the natural history of the chronic disease. In practice, this can be done by estimating the relevant event rates (i.e. ideally as incident rates) over time such that spans the periods prior to diagnosis of the chronic disease and afterwards. Indeed, the use of incident events in preference over prevalent

cases may provide a more incisive insight to the nature of the relationship between the comorbid condition and the chronic disease although the effectiveness of this approach may depend on the number of years for which reliable historical data are available.

In this paper, we will recap the conventional approach for identifying comorbidities which may be associated with any particular chronic disease. We will then describe an innovative incidence-based methodology for identifying patterns of associations between comorbidities and the chronic disease along its natural history which we consider as a more viable alternative. By way of illustration, we will also reproduce some of the results reported elsewhere in a previous application of the new approach in COPD based on the UK CPRD population (formerly, the GPRD).[14]

Methods

Conventional Approach: usually involves distinct patient populations in a matched cohort design in the following format-

1. One set of patients who have a record of diagnosis or consultation for chronic disease X in an a priori specified calendar year of interest and a random sample of patients who according to their medical records, are free of the disease
2. Both sets are from the same database population with each member also satisfying certain pre-specified inclusion/exclusion study criteria.
3. The date of the diagnosis/consultation for disease X in the specified calendar year- regardless of whether it is a pre-existing or new condition- is taken as the index date and this is also assigned to the matched control so as to ensure same start of follow-up for each pair.
4. Matching is usually on important measurable variables (i.e. likely confounding factors) identified as key to facilitating comparability between the two cohorts[10,12].

Age and gender are the most commonly used factors in this regard. The two cohorts may also be matched on other variables such as the duration of historical records at index-date. Indeed, depending on the primary purpose of the study, the pool of eligible controls for each case may be restricted to only those whose last records span for at

least as long as that of the case so as to minimize the impact of between-pair differences in loss to follow-up[12].

Incidence-Based Trend Analytical Approach: This involves a pre-specified study period that spans over a reasonable number of years (i.e. d), instead of the conventional method which either uses a single calendar year to identify patients with chronic disease X or assesses event rates only in the post-diagnostic period. In this sense, the new approach is also different from the incidence-based methodology described elsewhere.[13]

1. The study period consists of two separate phases: an earlier period of duration d1 years for the identification of incident cases of X and a subsequent period of d2 years post diagnosis. The total period for trend analysis is thus d=d1 + d2.
2. Cohort X consists exclusively of patients newly diagnosed with condition X over the study period (i.e. incident diagnosis) and the incident diagnosis date is defined as index date. Patients with any record of diagnosis/consultation for disease X outside of the study period are excluded.
3. Each member of this cohort is then matched to a patient from a random sample of those in the database who are free of disease X in their entire medical history (i.e. X=0). The matched control is assigned the same index date.
4. As in the conventional approach, the matching variables include age and gender.
5. However, unlike the former approach, each case is additionally matched to its control on total completed years of medical records pre and post index date to ensure that the control is followed-up for as long as the case- each having the same duration for the trend analysis.

Indeed, an aspect of the incidence-based approach has been successfully applied to assess the risk of cataract among idiopathic thrombocytopenic purpura patients in the CPRD.[15]

Data analysis: For each year i relative to the index date (i=1, 2, …, d, with i=1 for the earliest observed year) and for each candidate comorbid event k, we estimate the

incidence rate per 1,000 person-years (IR_{ik}) for each cohort as well as the corresponding 95% confidence interval in a conditional logistic regression model involving relevant individual characteristic measures as explanatory covariates.[16] We also estimate the rates ratio and its corresponding 95% confidence interval using the conditional logistic regression approach to account for the matching variable, often ignored at some cost in the analysis of matched cohort data.[17-18]

To assess trends in rates ratios along the natural history of X, we fit a linear regression to the annual rate ratios on a logarithmic scale for the candidate comorbid event k and estimate the average annual percentage change over the periods prior to and post index date and separately also for the overall period of evaluation (i.e. d years). The resulting slope of each regression line is assessed for statistical significance.

Application: By way of illustration of the new methodology, we have reproduced the details of a previous application in the UK CPRD over a ten-year period in which we evaluated the incident patterns of medical events from a list of candidates of a priori interest, thought to have possible associations with COPD.[12-14] Comorbidity was defined as any event resulting from any consultation with a general practitioner which is significantly more common in COPD patients. Thus this illustration does not constitute a study of COPD.

We used a retrospective cohort of patients aged 50+ with a diagnosis of COPD. Each COPD patient was matched to another patient without COPD on year of birth, gender, general practice and completed years of medical records up to at least a year after the index date for COPD between 1990 and 1998, the index date of the COPD patient having been assigned to the matched non-COPD counterpart. We then estimated the annual incidence rates per 1,000 person-years for each event in each cohort over the ten-year period as well as the corresponding annual rates ratios (RRs) and their 95% confidence intervals such that RR>1 indicates a higher rate in COPD. The age group is same as in the previously reported COPD studies conducted on the database.[12-14]

Results

A total of 24,079 COPD patients were each matched to a non-COPD patient (Figure 1).

The annual event rates in COPD and the corresponding annual rates ratios are as shown in Tables 1 and 2 correspondingly.

According to these results, the incidences of many of the smoking related chronic conditions were more common in COPD patients than those free of the disease.[19-20] They were consistently at higher risk of suffering from conditions such as lung cancer, asthma, other respiratory diseases, fracture, osteoporosis, thoracic, mediastinal, cardiac, nervous system and psychiatric disorders as early as several years before diagnosis of COPD. However, we found no evidence of association between COPD and conditions such as pneumonia, glaucoma, ear and labyrinth disorders, reproductive system, breast disorders and vascular diseases other than angina and cardiac disorders, although there was apparent sign of annual elevation in risk over time for some of the conditions. The pattern for angina was particularly inconsistent in terms of statistical significance- levels were significantly higher in the COPD patients only for the immediate 1-year periods before and after COPD diagnosis- thus highlighting the unreliable nature of methods which rely solely on events in the first year of diagnosis of COPD.[13]

Table 1. Annual incidence rates of certain conditions per 1000 person-years in COPD patients*

CANDIDATE CONDITIONS	PRIOR TO COPD DIAGNOSIS					POST COPD DIAGNOSIS				
	Year -5	Year -4	Year - 3	Year - 2	Year - 1	Year 1	Year 2	Year 3	Year 4	Year 5
Lung cancer	0.51	0.29	0.91	0.51	1.42	4.38	7.34	6.83	6.21	5.18
Asthma	40.15	51.39	64.46	76.25	110.19	118.19	58.44	41.35	41.87	36.83
Pneumonia	3.18	4.78	6.35	7.34	18.54	16.75	22.63	23.65	22.34	23.29
Respiratory Infections	3.21	3.50	2.74	2.37	3.72	4.38	4.82	5.84	6.24	7.30
Other respiratory, thoracic and mediastinal disorder	61.50	91.83	117.82	169.80	289.85	199.11	147.24	130.74	127.90	105.45
Angina	19.71	23.21	24.35	26.86	31.72	31.90	19.53	19.24	15.70	19.86
Cardiac disorders	35.00	48.95	70.88	107.16	250.61	187.28	125.12	113.59	115.89	115.34
Other vascular disorders	36.57	45.04	51.47	57.49	56.79	63.62	52.41	51.43	52.82	48.03
Cataract	10.48	11.68	12.05	14.38	15.07	16.24	18.40	18.18	20.44	16.28
Glaucoma	4.93	5.22	5.29	4.85	5.77	5.58	4.42	5.07	3.80	4.85
Fracture	13.83	12.99	15.62	16.86	15.48	20.59	19.16	20.18	21.64	18.18
Osteoporosis	3.39	4.60	5.95	5.91	6.64	10.18	8.18	10.26	11.46	11.17
Skin Bruises	4.64	4.02	3.91	4.85	4.64	5.91	5.22	6.35	6.53	5.91
Other skin and subcutaneous tissue disorders	52.23	66.72	83.69	98.22	98.00	99.97	93.51	92.71	87.86	83.80
Ear and labyrinth disorders	40.15	45.88	49.57	53.95	54.93	49.68	47.60	45.15	48.55	50.33
Nervous system disorders	42.41	53.18	60.59	70.59	80.37	84.57	79.90	80.23	73.95	76.07
Psychiatric disorders	33.47	39.57	46.25	48.25	53.36	59.31	50.44	42.74	45.08	42.12
Reproductive system and breast disorders	19.35	21.83	18.98	19.27	18.14	16.35	13.72	13.87	13.03	14.24
Social circumstances	7.88	7.12	5.51	5.69	7.41	8.91	9.02	11.86	10.95	14.38

* Reproduced from Kiri et al. (2005); See Hansell et al. (2005) and Soriano et al. (2005) for details of the events that make up the candidate conditions

Indeed, we also found evidence of time-varying associations. For example, the annual levels for skin-related events were significantly and consistently higher among COPD patients only after the chronic disease had been diagnosed- thus suggesting possible association with either treatment or severity of COPD or both. It is worthy of note that an assessment based strictly on data in the post COPD diagnosis period would have

offered a single conclusion, namely an association between the condition and COPD regardless of severity and treatment.

Table 2. Annual incidence rates ratios of certain conditions per 1000 person-years in COPD and non-COPD patients^

CANDIDATE CONDITIONS	PRIOR TO COPD DIAGNOSIS					POST COPD DIAGNOSIS						
	Year -5	Year -4	Year -3	Year -2	Year -1	Year 1	Year 2	Year 3	Year 4	Year 5	Annual %change: 5-year Prior	Annual %change: Entire Period
Lung cancer	4.7	3.9	5.3*	10.7*	16.9*	52.2*	14.3*	10.2*	6.6*	8.2*	42.8#	27.4#
Asthma	3.7*	4.6*	6.7*	8.1*	14.0*	18.9*	12.3*	8.5*	9.7*	7.1*	38.1#	25.0#
Pneumonia	3.8	2.9	3.1	3.2	7.5	7.4	7.4	5.6	8.1	6.1	16.2	21.4#
Respiratory Infections	1.1	1.9*	1.6	1.1	1.4	1.9*	1.8*	1.4	1.2	1.5	-0.2	3.7
Other respiratory, thoracic and mediastinal disorder	1.4*	1.6*	1.8*	2.3*	3.7*	2.8*	2.1*	2.0*	1.9*	1.6*	25.0#	6.6#
Angina	1.2	1.1	1.2*	1.2*	1.6*	1.9*	1.1	1.2*	1.0	1.6*	6.9	2.8
Cardiac disorders	1.2*	1.5*	1.7*	2.2*	4.7*	4.0*	2.6*	2.2*	2.4*	2.3*	35.9#	10.7#
Other vascular disorders	0.9	1.0	1.0	1.0	1.0	1.3*	1.0	1.0	1.0	0.9	1.2	0.3
Cataract	1.2	1.1	1.0	1.3*	1.1	1.2*	1.2	1.1	1.5*	1.3	1.1	2.5#
Glaucoma	1.1	0.7	1.0	0.8	0.9	1.1	1.1	1.2	0.9	1.2	-3.5	1.6
Fracture	1.2	1.1	1.2	1.3*	1.2*	1.4*	1.5*	1.6*	1.5*	1.2	1.5	4.1#
Osteoporosis	1.2	1.7*	1.5*	1.6*	1.8*	2.3*	1.7*	2.4*	2.0*	2.2*	8.0	8.4#
Skin Bruises	1.2	1.1	1.1	1.3	1.1	1.7*	1.2	1.8*	1.9*	1.5	1.1	5.6#
Other skin and subcutaneous tissue disorders	1.0	1.0	1.1*	1.1	1.0	1.2*	1.2*	1.3*	1.2*	1.2*	0.4	2.3#
Ear and labyrinth disorders	1.2*	1.1	1.1	1.0	1.0	1.1	1.1	1.0	1.1	1.1	-4.0	0.3
Nervous system disorders	1.1	1.2*	1.1*	1.1*	1.2*	1.3*	1.3*	1.3*	1.2*	1.3*	1.2	3.1#
Psychiatric disorders	1.1	1.2*	1.3*	1.3*	1.4*	1.8*	1.5*	1.4*	1.6*	1.2	4.8#	3.9#
Reproductive system and breast disorders	0.9	1.0	0.9	1.0	0.9	1.0	1.0	1.0	1.0	1.2	1.0	0.9
Social circumstances	1.1	1.2	1.5*	1.2	1.8*	1.3*	1.2	1.4*	1.2	1.1	11.3	2.0

* Significantly higher rate in COPD patients
# $p < 0.05$ and hence the annual change was significantly different from zero
^ Reproduced from Kiri et al. (2005)

Discussion

In this paper, we have described the features of an incidence-based methodology for identifying potential comorbid conditions for any particular chronic disease. The methodology exploits the longitudinal properties of observational databases to track incident event rates along the natural history of the chronic disease, as it involves the periods prior to its formal diagnosis and beyond. The results of its application in COPD, as previously described in detail elsewhere, revealed significant time-dependent associations between the chronic disease and certain conditions. We found evidence that in COPD patients, the likelihood of diagnosis of certain comorbid events were highest in the immediate 1-year periods before and after diagnosis of the chronic disease, perhaps due to the diagnostic-related activities experienced by these patients. If true, then a methodology which relies solely on data in the first year post diagnosis of COPD is much more likely to suggest associations which may be spurious than our approach.

These findings may have interpretational implications on the results of comorbidity studies which are based exclusively on data in the immediate year post diagnosis of any chronic disease of interest. Our results also suggest the trends approach which maintains the longitudinal quality of the data in the assessment of comorbidity associations with a chronic disease, may be more reliable than the traditional single estimate approach. Indeed, the new approach offers a facility for enhancing our understanding of the natural history of the chronic disease in relation to the burden of comorbidity in the management of patients living with the condition. With appropriate data, the method may also be useful to pharmacovigilance activities for any particular of interest, as it offers longitudinal results which may be used to put information from spontaneous reports into an appropriate context. We can do this be done by assessing the incident patterns of the event in two separately matched cohorts of the (1) exposed versus unexposed persons in one and (2) the chronic disease patients versus those free of the disease in the other.

We acknowledge the existence of alternative methods for obtaining matched cohorts in disease natural history studies and we have provided our reasons for excluding the propensity score approach for consideration. In the setting of exploration of possible associations between a chronic disease and comorbidities, we believe the propensity score is exactly the same as the disease risk score- a probability estimate of a patient's likelihood of disease occurrence which has never been used for such disease natural history studies.[21-23] Outside of this setting, we think propensity score matched cohorts could be useful for assessing factors associated with actual clinical practice in a chronic disease- such as the management of such patients in terms of resource utilization independent of other sources of resource use (i.e. confounding factors including comorbidities, among others).

A potential limitation of the new methodology, though common in disease natural studies conducted in general practice databases, is the possibility that the underlying behavior and attendance patterns of the patients at the practices could affect the probability of diagnosis of the events. For example, as COPD patients may have higher rates of doctor consultations than those without COPD (i.e. for routine checks, treatment of acute exacerbations as recommended in guidelines, among many other disease-related reasons), some events may have a higher likelihood of diagnosis in the COPD group.[24] Clearly a notable limitation of the COPD illustration was the lack of control for the likely effect of smoking status which was due to the limited scope of information on smoking in the CPRD at the time of the study. Thus, smoking could indeed account in part for the observed differences between the two groups. Furthermore, the requirement of having at least one year follow-up might also introduce some bias in event estimates because of the possibility of significant differences between the two original cohorts in the proportion of patients with the comorbidities of interest over that period.[13]

The strengths of our methodology include the provision for exploiting the longitudinal properties of observational databases to obtain comparable estimates of event rate ratios as well as the provision for estimating the incidence patterns of such events over

time which may facilitate a much clearer understanding of the nature of their associations with the disease.

Acknowledgments:

The author is grateful to Dr Kourtney Davis for her encouragement and is indebted to GlaxoSmithKline R&D for the *excellence in research* award granted him for the development of the incidence-based design whilst he was its employee. The author is also grateful to the reviewers for their useful suggestions. Finally, this work is humbly dedicated to the memory of Dr George Visick, a former research colleague at GlaxoSmithKline R&D whose untimely death remains hard to bear.

What is already known on this subject?
The risk of adverse effects associated with a treatment and its effectiveness may depend on the extent and severity of pre-existing comorbidities in the particular disease

To identify comorbidities, it is common practice to compare single point estimates of the rate for each candidate over a specified period between patients with the disease and a random sample of disease-free patients

What this study adds?
Introduced and illustrated a new methodology which tracks the trends of incident events along the natural history of the disease, thereby exploiting the longitudinal properties of observational data in contrast to the conventional approach of single point estimates

The approach facilitates a clearer understanding of the nature of the associations

The findings in COPD suggest the increased likelihood of spurious associations by the single point estimation approach which is based on a single observation window

Contributorship statement: Author conducted the research, including data analysis and writing of the manuscript

Competing interests: Author consults for the pharmaceutical industry on epidemiological methods

Funding: Author was an employee of GlaxoSmithKline Research & Development during the period of the research

Data sharing statement: No additional data available

References

1. Lange P Ulrick CS, Vesbo J. Mortality in adults with self-reported asthma. Copenhagen City Heart Study Group. Lancet. 1996 May 11;347(9011):1285-9
2. Wijnhoven HA, Kriegsman DM, Hesselink AE, de Haan M, Schellevis FG. The influence of co-morbidity on health-related quality of life in asthma and COPD patients. Respir Med 2003; 97: 468-475
3. Eisner MD, Yelin EH, Trupin L, Blanc PD. The influence of chronic respiratory conditions on health status and work disability. Am J Public Health 2002; 92: 1506-1513

4. Iribarren C, Tolstykh IV, Eisner MD. Are patients with asthma at increased risk of coronary heart disease?. Int J Epidemiol 2004;33:743-748
5. Blais L, Ernst P, Suissa A. Confounding by indication and channeling over time: the risks of p -agonists. Am J Epidemiol 1996;144:1161-1169
6. Spitzer WO, Buist AS. Case-control study of prescribed fenoterol and death from asthma in New Zealand. Thorax 1990;45:645-646
7. Moser M, Gifford R. Diuretic therapy and the risk of cardiac arrest. N Engl J Med 1994;331:1235-1236
8. Psaty BM, Heckbert SR, Koepsell TD, et al. The risk of myocardial infarction associated with antihypertensive drug therapies. JAMA 1995;274:620-625
9. Hoes AW, Grobbee DE, Lubsen J, et al. Diuretics, betablockers, and the risk for sudden cardiac death in hypertensive patients. Ann Intern Med 1995;123:481-487
10. Liao P, Yegneswaran B, Vairavanathan S, Zilberman P, Chung F. Postoperative complications in patients with obstructive sleep apnea: a retrospective matched cohort study. Can J Anesth/J Can Anesth 2009; 56:819–828. DOI 10.1007/s12630-009-9190-y
11. Skillrud DM, Offord KP, Miller RD. Higher risk of lung cancer in chronic obstructive pulmonary disease. A prospective, matched, controlled study. Ann Intern Med 1986;105:503-507
12. Hansell AL, Lam KA, Richardson S, et al. Medical event profiling of COPD patients. Pharmacoepidemiol Drug Saf 2004;13:547–555
13. Soriano JB, Visick GT, Muellerova H, Payvandi N, Hansell AL. Patterns of comorbidities in newly diagnosed COPD and asthma in the primary care. Chest 2005; 128:2099–2107
14. Kiri VA, Muellerova H, Visick G. Comorbidity Profiling of COPD patients in the UK Primary Care using an incidence based approach to detect associations with the disease. Am J Respir Crit Care Med 2005; 2:A851
15. Feudjo-Tepie MA, Susan A . Hall SA, John Logie J, Robinson NJ. Risk of cataract among idiopathic thrombocytopenic purpura patients in the UK general

practice research database. pharmacoepidemiol drug safe 2009;18:380–385. DOI: 10.1002/pds.1723

16. Breslow NE, Day NE. Statistical Methods in Cancer Research. Volume 1- The Analysis of Case-Control Studies. IARC 1980;32
17. Sjölander A, Greenland S. Ignoring the matching variables in cohort studies - when is it valid and why? Stat Med. 2013 Jun 12:0. doi: 10.1002/sim.5879
18. Rothman, K., Greenland, S., Lash, T.: Modern Epidemiology. Lippincott Williams & Wilkins 2008;Chpt 11
19. British Thoracic Society Guidelines - COPD. 2007. 9-10-2007. Ref Type: Internet Communication
20. Soriano JB, Maier WC, Egger P, Visick G, Thakrar B, Sykes J, et al. Recent trends in physician diagnosed COPD in women and men in the UK. Thorax 2000;55:789-94
21. Miettinen OS. Stratification by a multivariate confounder score. American Journal of Epidemiology 1976; 104: 609–20
22. Arbogast PG, Ray WA. Use of disease risk scores in pharmacoepidemiologic studies. Stat Methods Med Res 2009; 18; 67 originally published online Jun 18, 2008; DOI: 10.1177/0962280208092347
23. Wyss R, Ellis AR, Brookhart A, Funk MJ, Girman CJ, Simpson RJ Jr and Stürmer T. Matching on the disease risk score in comparative effectiveness research of new treatments. 2015. DOI: 10.1002/pds.3810
24. Fabbri LM, Rabe KF. From COPD to chronic systemic inflammatory syndrome. Lancet 2007;370:797-9

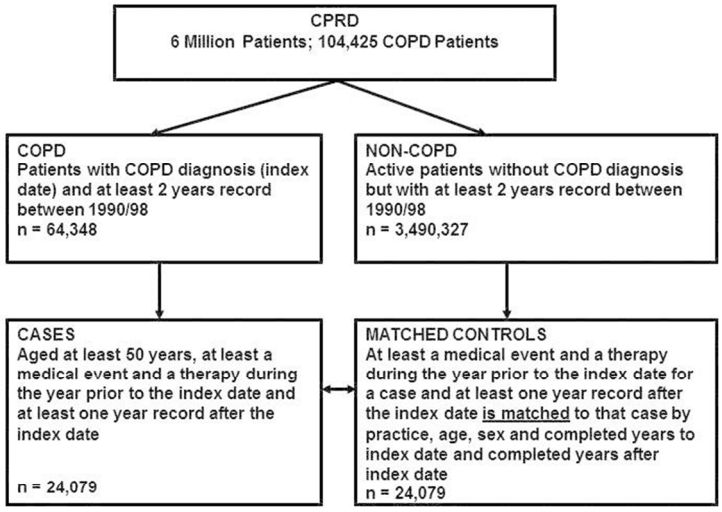


Figure 1. Selection of COPD incident cases and controls from the Clinical Practice Research Database (CPRD)

81x60mm (300 x 300 DPI)

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract [**PAGE 1**]
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found [**PAGE 2**]
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported [**PAGE 3**]
Objectives	3	State specific objectives, including any prespecified hypotheses [**PAGE 6**]
Methods		
Study design	4	Present key elements of study design early in the paper [**PAGE 6**]
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection [**PAGE 7**]
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up [**PAGE 7 as applicable for an illustration**] *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed [**PAGE 7**] *Case-control study*—For matched studies, give matching criteria and the number of controls per case
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable [**PAGE 8 as applicable for an illustration**]
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group [**PAGE 8 as applicable for an illustration**]
Bias	9	Describe any efforts to address potential sources of bias [**PAGE 7 as applicable for an illustration**]
Study size	10	Explain how the study size was arrived at [**Not applicable**]
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why [**PAGE 8 as applicable for an illustration**]

Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding [**PAGE 8 as applicable for an illustration**]
		(*b*) Describe any methods used to examine subgroups and interactions [**Not applicable**]
		(*c*) Explain how missing data were addressed [**Not applicable**]
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed [**PAGE 8 as applicable for an illustration**] *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy
		(*e*) Describe any sensitivity analyses

Continued on next page

Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed [**PAGES 9-11 as applicable for an illustration**]
		(b) Give reasons for non-participation at each stage
		(c) Consider use of a flow diagram [**PAGE 9**]
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders
		(b) Indicate number of participants with missing data for each variable of interest
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount) [**Not applicable for an illustration**]
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time [**PAGES 10-11**]
		Case-control study—Report numbers in each exposure category, or summary measures of exposure
		Cross-sectional study—Report numbers of outcome events or summary measures
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included [**PAGES 10-11**]
		(*b*) Report category boundaries when continuous variables were categorized
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses
Discussion		
Key results	18	Summarise key results with reference to study objectives [**PAGES 12**]
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias [**PAGES 13**]
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence [**PAGES 11-12**]
Generalisability	21	Discuss the generalisability (external validity) of the study results
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based [**PAGE 14**]

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.