JJCO Japanese Journal of Clinical Oncology

Japanese Journal of Clinical Oncology, 2015, 45(4) 323–327
doi: 10.1093/jjco/hyu220
Advance Access Publication Date: 14 January 2015
Review Article

OXFORD

Review Article

Methodological issues in observational studies and non-randomized controlled trials in oncology in the era of big data

Shiro Tanaka*, Sachiko Tanaka, and Koji Kawakami

Department of Pharmacoepidemiology, Graduate School of Medicine and Public Health, Kyoto University, Kyoto, Japan

*For reprints and all correspondence: Shiro Tanaka, Department of Pharmacoepidemiology, Graduate School of Medicine and Public Health, Kyoto University, Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan. E-mail: tanaka.shiro.8n@kyoto-u.ac.jp

Received 14 July 2014; Accepted 21 December 2014

Abstract

Non-randomized controlled trials, cohort studies and database studies are appealing study designs when there are urgent needs for safety data, outcomes of interest are rare, generalizability is a matter of concern, or randomization is not feasible. This paper reviews four typical case studies from methodological viewpoints and clarifies how to minimize bias in observational studies in oncology. In summary, researchers planning observational studies should be cautious of selection of appropriate databases, validity of algorithms for identifying outcomes, comparison with incident users or self-control, rigorous collection of information on potential confounders and reporting details of subject selection. Further, a careful study protocol and statistical analysis plan are also necessary.

Key words: bias, confounding, database, epidemiology, propensity score

Introduction

The gold standard for confirmatory clinical research is randomized clinical trials (RCT) (1). This most scientifically rigorous study design is, however, not always the best solution to real clinical problems. For example, in the case of an immediate safety study for a new anticancer medicine after its release, randomization may not be ethical when there are important identified risks of the drug. An alternative approach is to survey the frequency of adverse events during its use (i.e. All-Case Surveillance) (2), but it does not provide a definitive conclusion on the causal relationship with a specific adverse event because of lack of a control group. Another criticism is generalizability of the results of RCTs—patients who are with poor performance status or comorbidity, elderly or women of childbearing potential are usually excluded in RCTs to ensure safety of the study participants, making it difficult to extrapolate study results to such special populations. Moreover, some RCTs for rare cancers set their statistical power at 70%, but is it sensible to select a study design which leads to an erroneous conclusion so frequently? Researchers, pharmaceutical companies and regulatory agencies including the US Food and Drug Administration (FDA) and Pharmaceuticals and Medical Devices Agency (PMDA), who have promoted RCTs, are now searching for alternative approaches, such as observational studies (3).

Despite their several non-negligible advantages over RCTs, some researchers are skeptical of evidence from non-randomized controlled trials and observational studies. Indeed, a systematic review reported that non-randomized studies tended to show treatment effects larger than those from RCTs (4). Epidemiologists are concerned with bias in observational studies, namely selection bias, information bias and confounding (5), and have developed rigorous methodologies against these. Can emerging methodologies such as propensity score (6) or database studies (3) overcome such biases? This paper aims to contrast non-randomized controlled trials and observational studies with RCTs and to clarify how to minimize bias in studies with non-randomized designs in oncology through review of four typical case studies.

Case studies

Post-marketing safety

The first example is a post-marketing study of gefitinib (7). After the approval of gefitinib for treatment for advanced non-small-cell lung cancer (NSCLC) in July 2002, an increase in spontaneous reports of interstitial lung disease (ILD) indicated the need for clinical studies on the causal relationship between gefitinib and ILD. Therefore, an independent academic team together with scientists from AstraZeneca planned this study as a 'post-marketing clinical trial' defined in the Good Post-marketing Study Practice (GPMSP). The eligibility criteria were advanced or recurring NSCLC who had received at least one chemotherapy regimen. Patients and their physicians could select the most appropriate treatment c50 mg gefitinib or chemotherapy after the registration of the cohort, and the patients were followed for up to 12 weeks after treatment initiation and assessed for the primary outcome, incidence of ILD. Diagnosis of ILD made at each participating institute was adjudicated by a blinded independent case review board based on computed tomography (CT) scans. A total of 3166 patients were recruited, and ILD developed in 122 patients. This study revealed by multivariate logistic regression and sensitivity analysis using propensity score that the odds ratio for gefitinib versus chemotherapy was 3.2 [95% confidence interval (CI), 1.9–5.4], with an elevated risk during the first 4 weeks by an odds ratio of 3.8 (95% CI, 1.9–7.7).

Table 1 summarizes the potential and degree of selection bias, information bias and confounding in non-randomized controlled trials, cohort studies and database studies. The current case study is classified as a cohort study in which patients are registered prospectively, diagnosis of outcome is relatively accurate, and confounding is a major potential source of bias. These methodological features of each study design will be discussed further in Section 3.

A distinction of this particular study was that data were collected and managed in accordance with the GPMSP, a counterpart of the good clinical practice (GCP) for post-marketing setting, implying that its quality control and quality assurance meets the requirements for GCP trials. Together with accurate diagnosis of ILD, this case study shows that cohort design is a particularly appealing option if there is a rare, important identified risk of a drug. In other words, pharmaceutical companies should consider comparative cohort studies rather than 'All-Case Surveillance' (2), which lack a control group but routinely used in post-marketing surveillance in Japan, in cases of a rare, important identified risk.

Confirmation of efficacy

Discussion on confounding has also been raised in a study that is being planned by the Japan Clinical Oncology Group (JCOG) (8). The aim of the study was to compare overall survival of proton beam therapy for resectable hepatocellular carcinoma with that of the standard care, surgery and this study was designed as a multicenter, non-randomized controlled trial rather than a RCT or an observational study. The differential safety profile and cost of proton beam therapy (uninsured by public health insurance) render it infeasible to randomly allocate patients to these two treatment options. Therefore, this trial is designed as a non-randomized 'confirmatory' trial with a total planned sample size of 270 patients and follow-up of 5 years. Propensity score analysis (3), rather than stratified Cox regression which is a default of the JCOG, is used in this trial.

As summarized in Table 1, lack of randomization inherently results in a potential of confounding. Study design in this case study is valid only if information on all potential confounders is collected and used for bias adjustment through propensity score analysis.

Rare events

The third example is a database study of varenicline, a newly approved drug prescribed for smoking cessation, by the Mini-Sentinel program (9). In June 2011, the FDA issued a Drug Safety Communication indicating that varenicline may increase cardiovascular diseases based on safety review of a placebo-controlled registration trial (10). The FDA also called for systematic review of all RCTs of varenicline to

Table 1. Potential and degree of selection bias, information bias and confounding in non-randomized controlled trials, cohort studies and database studies

	Non-randomized controlled trial	Cohort study	Database study
Description	Clinical trials in which each patient or investigator may select the patient's treatment group.	Follow-up studies in which subsets of a defined population are identified by treatment of interest.	Prescription for a specific indication is identified in an administrative database and outcomes of each regimen are compared.
Sample size	100–1000	100–10 000	More than 10 000
Selection bias	As in RCTs	Prospective registration with eligibility criteria usually broader than RCTs.	Selection from a possibly exhaustive population.
Information bias			
Diagnosis	As in RCTs	As in RCTs	Based on indication and disease names covered by insurance.
Assessment			
	Assessed by CTCAE	Not assessed by CTCAE	Not assessed by CTCAE
Acute toxicity			
Late toxicity	Limited	Limited	Possibly identifiable by records
Progression/ recurrence	Available	Not as accurate as RCTs	Difficult to identify
Death	Available	Available	Available
Follow-up length	As in RCTs	Possibly longer than RCTs	Within source healthcare system (e.g. health insurance)
Accuracy of timing	Accurate	Accurate	Possibly rounded-off to month or day
Confounding	Present	Present	Present

RCT, randomized clinical trial; CTCAE, Common Terminology Criteria for Adverse Events.

determine the causal relationship with cardiovascular risk. A total of four systematic reviews have been conducted by Pfizer and academic researchers, and although there was an overlap of data, they reached conflicting conclusions. The relative risks for cardiovascular adverse events of varenicline when compared with placebo estimated by the first (11), second (12), third (13) and fourth reviews (14) were 1.95 (95% CI, 0.79–4.82; 15 trials; 7002 patients; 19 events), 1.72 (95% CI, 1.09–2.71; 14 trials; 8316 patients; 79 events), 1.58 (95% CI, 0.90–2.76; 22 trials; 9232 patients; 52 events) and 1.30 (95% CI, 0.79–2.23; network meta-analysis of 63 trials), respectively. The observed cardiovascular adverse events in premarketing RCTs were apparently too small to provide sufficient statistical power for a definitive conclusion, and the FDA also requested that the Mini-Sentinel program (15) perform a rapid safety assessment using its large-scale administrative databases.

In this retrospective analysis of administrative databases (9), subjects were identified as individuals who filled a first prescription for varenicline or the comparator drug, bupropion hydrochloride, between 1 January 2006 and 5 July 2011. The populations analyzed were further restricted to adults who were continuously enrolled in the health plan with medical and drug coverage and had a diagnosis code for tobacco use disorder (ICD-9-CM code, 305.1) without any diagnosis code for cardiovascular disease. The outcome was defined as a composite of acute myocardial infarction (410.xx), intermediate coronary syndrome or unstable angina (411.1), and acute coronary occlusion without myocardial infarction (414.0x) recorded as the primary diagnosis in an inpatient or emergency department setting after the index date, which is defined as the date of the first dispensing. The analysis population with tobacco use disorder consisted of 89 519 varenicline users (56 cardiovascular events) and 113 378 bupropion users (118 cardiovascular events), and the incidence rate ratio adjusted for age, sex and data partner was 1.02 (95% CI, 0.71–1.47).

An apparent strength of database studies is their relatively large sample size (Table 1). In this specific example, the width of 95% CI [0.71–1.47 (9)] was much narrower than those from the systematic reviews (e.g. 0.79–2.23 in 14). However, it may not be plausible that regression adjustment by age, sex and data partner can completely exclude bias due to confounding.

Drug utilization

Administrative databases also provide an important measure to describe trends and utilization of cancer treatment. Cancer care in Japan has rapidly changed from in-hospital care to outpatient care and from in-hospital prescription to external prescription at pharmacies, so a drug utilization study (16) aimed to describe the use of oral anticancer medicines in insurance pharmacies. This study analyzed databases of dispensings in 489 pharmacies provided by two major pharmacy chains in Japan. A total of 31 628 patients who received oral anticancer medicines between 1 June 2011 and 31 May 2012 with 156 904 dispensings were identified in the databases. The patients received hormone therapy ($n = 19\,899$; 62.9%), anti-metabolic medicines ($n = 9002$; 28.5%), molecularly targeted medicines ($n = 1716$; 5.4%), alkylating compound medicines ($n = 839$; 2.7%), microtubular inhibitors ($n = 148$; 0.5%) and immune-suppressing medicines ($n = 24$; 0.1%). These findings suggest not only an increasing use of oral anticancer medicines in insurance pharmacies, but also the importance of pharmacy-clinic cooperation in clinical practice. Populations covered by pharmacy dispensing databases are unique since they are limited to outpatients, but may include uninsured dispensings unlike claims databases (Table 2).

Statistical issues

Selection bias

Selection bias can occur when entry or participation of patients into a study is related to exposures or outcomes of interest. Examples of selection bias include Berkson's bias, healthy worker effects and immortal time bias (5,17). Retrospective studies are often criticized for higher susceptibility to selection bias than non-randomized controlled trials and cohort studies, which register subjects prospectively. This would be true if subjects were retrospectively identified in medical records, but retrospective analysis of a database could be an approximately exhaustive survey in which selection bias does not occur if the database covered a relevant target population completely (e.g. beneficiaries).

Administrative databases can be classified by their data holders, i.e. hospitals and clinics, pharmacies and health insurance societies or health maintenance organizations (Table 2). As illustrated in the third case study (9), the Mini-Sentinel program initiated by the FDA has developed a pharmacovigilance system that uses data from electronic medical records and claims maintained by collaborating data partners separately. As of 22 September 2013, its database included ~150 million people and 4 billion pharmaceutical dispensings (15). In Japan, the Medical Information for Risk Assessment Initiative (MIHARI) Project by the Pharmaceuticals and Medical Devices

Table 2. Characteristics of administrative databases according to data holder

	Electronic medical records database	Pharmacy dispensing database	Claims database
Source	Medical records, ordering system	Dispensings at a pharmacy	Health insurance claims/receipts
Population	Patients in a medical institute	Outpatients	Beneficiaries
Contents	In- and outpatient information, laboratory measurements	Information on dispensings (e.g. dosage, administration)	Medical and dispensing claims, diagnosis procedure combination, health check-up
Identification of disease	Medical records, indication	Indication	Disease name covered by insurance
Follow-up	Within institute	Within pharmacy chain	Within health insurance
Potential measures for linkage	Insurance certificate number, medical records ID	Insurance certificate number	Insurance certificate number, Hash function
Advantages	Amount of information, linkage to disease registries	Coverage of uninsured dispensings	Potential for coverage of all beneficiaries in Japan
Example	MIHARI Project, Platform for CISA	Pharmacy chain stores	JMDC (18), MDV (19), NDB

MIHARI, Medical Information for Risk Assessment Initiative; CISA, Clinical Information Statistical Analysis; JMDC, Japan Medical Data Center; MDV, Medical Data Vision; NDB, National Data Base by Ministry of Health, Labour and Welfare.

Agency has been creating an electronic medical records database by extracting data from 10 participating hospitals in SS-MIX format. National Data Base (NDB)by Ministry of Health, Labour and Welfare and domestic database venders such as the Japan Medical Data Center (JMDC) Co. (18) and Medical Data Vision (MDV) Co. (19) also provide researchers with claims databases. Several chains of insurance pharmacies in Japan also maintain administrative databases which include information on dispensings and prescription formulas. As illustrated in Table 2, databases vary widely in terms of population covered, data source, purpose of data collection, follow-up and variables included. Thus, record linkage would be the most valuable future work as the infrastructure for methodologically sound database studies. In practice, a check list may be helpful for researchers selecting databases (20).

Consideration of selection bias is important not only for choosing databases, but also for statistical analysis and reporting. Immortal time bias can arise when the period between entry into a database and date of first exposure to a drug of interest, during which death has not occurred, is either misclassified or simply excluded (17). Because immortal time bias is frequent in database studies that compare against non-users, incident user cohort design, which restricts analyses to individuals under observation at the start of the current course of treatment, is regarded as a basic study design (21). An alternative approach is comparison with self-control. The self-controlled case series design can be used for examining associations between acute outcomes and transient exposures using only data on specific cases, that is, on individuals who have experienced the outcome of interest (22,23). For transparent reporting of results of observational studies, the STROBE statement recommends description of the setting, location, eligibility criteria, relevant date, and a flow diagram of the number of individuals at each stage of the study (24).

Information bias

Some drawbacks of pharmacy dispensing and claims databases are lack of diagnostic information and the problem of disease name for reimbursement (Tables 1 and 2). Subjects in the database study of varenicline (9) were identified by disease names on claims, which may be determined just for reimbursement, and the utilization study of oral anticancer medicines (16) could not describe drug utilization among patients with specific types of cancer since diagnostic information other than indication of drugs was not available. For the same reasons, information on acute toxicity during chemotherapy or progression and recurrence in databases of dispensings or claims would not be as accurate as in RCTs. Length of follow-up is also important for long-term outcomes in particular (Table 2).

Information bias can result in misclassification of outcomes, exposures and confounders. If misclassification of an outcome or exposure is non-differential or random, the exposure–outcome association would be biased toward the null (i.e. underestimated), but the direction of bias is not predictable in the case of differential misclassification or non-differential misclassification of confounders (5). However, the degree of bias can be quantified if the degree of misclassification, which is usually expressed as sensitivity and specificity, is known (5). Therefore, it is important to understand the sensitivity and specificity of algorithms for identifying outcomes, and a database study often accompanies a validation study for the algorithms. Algorithms to identify cerebrovascular accidents, transient ischemic attacks, congestive heart failure, deep vein thrombosis, pulmonary embolism, angioedema and total hip arthroplasty revision are considered to perform well (25), but there has not been any validation study for algorithms for cancer except those for lymphoma (26). With regard to death, algorithms based on claims have moderate sensitivity (~60%), high specificity (99.99%) and high positive predictive values (94.8%), indicating presence of 'zombie' claims presumably due to the delay of reimbursement processes (27). The Japanese Epidemiological Association has proposed use of vital statistics as a Japanese version of the National Death Index to improve the accuracy of mortality information.

Confounding

RCTs and observational studies have two different principles of statistics. RCTs essentially ignore the patients' characteristics and just compare survival curves of randomly allocated treatment groups. The role of statisticians in RCTs is to maximize power while keeping α error rate under a given level. In contrast, comparison of two survival curves may be distorted in observational studies. Here, the major goal of statistical analysis is to minimize bias due to confounding (Table 1). In the first case study (7), the suspected relationship between gefitinib and ILD was already recognized before initiation of the study, and consequently, gefitinib tended to be used for patients who were with a performance status of 2–3 or adenocarcinoma, women or non-smokers more frequently than chemotherapy. To adjust for apparent confounding by indication, data on age, performance status, duration of lung cancer, concurrent cardiac disease, severity of pre-existing pulmonary emphysema, smoking status, extent of normal lung on CT and pre-existing ILD were collected and included in the multivariate logistic regression. The second case study, a non-randomized confirmatory trial of proton beam therapy for resectable hepatocellular carcinoma, plans to adjust for confounding through the propensity score analysis (8). The confounders listed in the protocol concept are UICC TNM classification, Child-Pugh classification, α-fetoprotein, sex, age, size of tumor, ICGR15, use of private health insurance, income and occupation status. Influence of confounding was also examined in the third case study, but the confounders used were only age, sex and data partner (9).

It is impossible to exclude the possibility of bias due to confounding completely, but the more information available on confounders, the more we can reduce bias through appropriate confounder selection and adjustment (e.g. propensity score 6). The identification of confounders requires expert substantive knowledge about the causal network, which consists in part of exposure and outcome (e.g. pathophysiological and clinical knowledge) (5,28). Specifically, three criteria for identifying confounders have been suggested (5): (i) a confounder must be associated with the exposure under study in the source populations; (ii) a confounder must be a risk factor for the outcome (i.e. it must predict who will develop disease), though it need not actually cause the outcome; (iii) a confounder must not be affected by the exposure or the outcome. Graphical screening using causal diagrams (29) and criteria for confounder selection derived from formal theory concerning causal diagrams (30) are also helpful. In principle, no statistical method can remove bias due to confounding completely if there are unmeasured confounders. In other words, the crucial part of observational studies is rigorous collection of information on potential confounders, rather than statistical techniques for confounder adjustment.

Conclusion

This brief review suggests that researchers planning observational studies should be cautious of selection of appropriate databases,

validity of algorithms for identifying outcomes, comparison with incident users or self-control, rigorous collection of information on potential confounders and reporting details of subject selection. Guidelines from regulatory agencies (31,32), an excellent text book (5) and checklists for a study protocol, selection of databases and reporting (20,24,32) would be helpful resources for researchers. There is a considerable body of literature on study protocols and statistical analysis plans for clinical trials, but very little addressing those for observational studies. To overcome bias inherent in observational studies, a careful study protocol and statistical analysis plan are also necessary (31,32).

Acknowledgements

We appreciate that Radiation Therapy Study Group and Hepato-biliary and Pancreatic Cancer Study Group in the Japan Clinical Oncology Group permitted us to publish part of design of PC 1315 (8).

Funding

This work was supported in part by the National Cancer Center Research and Development Fund (26-A-4).

Conflict of interest statement

None declared.

References

1. Hahn OM, Schilsky RL. Randomized controlled trials and comparative effectiveness research. *J Clin Oncol* 2012;30:4194–201.
2. Pharmaceuticals and Medical Devices Agency. Q & A on use-results surveys conducted as All-Case Surveillance and Early Post-Marketing Phase Vigilance (EPPV) for Prescription Drugs. http://www.pmda.go.jp/kijunsakusei/guideline.html (1 July 2014, date last accessed).
3. Hershman DL, Wright JD. Comparative effectiveness research in oncology methodology: observational data. *J Clin Oncol* 2012;30:4215–22.
4. Ioannidis JP, Haidich AB, Pappa M, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA* 2001;286:821–30.
5. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*, 3rd edn. Philadelphia: Lippincott, Williams & Wilkins 2008.
6. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;70:41–55.
7. Kudoh S, Kato H, Nishiwaki Y, et al. Interstitial lung disease in Japanese patients with lung cancer: a cohort and nested case-control study. *Am J Respir Crit Care Med* 2008;177:1348–57.
8. Radiation Therapy Study Group and Hepato-biliary and Pancreatic Cancer Study Group in the Japan Clinical Oncology Group. PC1315 (unpublished protocol concept), 2014.
9. Toh S, Baker MA, Brown JS, Kornegay C, Platt R; Mini-Sentinel Investigators. Rapid assessment of cardiovascular risk among users of smoking cessation drugs within the US Food and Drug Administration's Mini-Sentinel program. *JAMA Intern Med* 2013;173:817–9.
10. The US Food and Drug Administration. FDA Drug Safety Communication: Chantix (varenicline) may increase the risk of certain cardiovascular adverse events in patients with cardiovascular disease. Published June 16, 2011. http://www.fda.gov/Drugs/DrugSafety/ucm259161.htm (1 July 2014, date last accessed).
11. The US Food and Drug Administration. FDA Drug Safety Communication: Safety review update of Chantix (varenicline) and risk of cardiovascular adverse events. Published December 12, 2012. http://www.fda.gov/drugs/drugsafety/ucm330367.htm (1 July 2014, date last accessed).
12. Singh S, Loke YK, Spangler JG, Furberg CD. Risk of serious adverse cardiovascular events associated with varenicline: a systematic review and meta-analysis. *CMAJ* 2011;183:1359–66.
13. Prochaska JJ, Hilton JF. Risk of cardiovascular serious adverse events associated with varenicline use for tobacco cessation: systematic review and meta-analysis. *BMJ* 2012;344:e2856.
14. Mills EJ, Thorlund K, Eapen S, Wu P, Prochaska JJ. Cardiovascular events associated with smoking cessation pharmacotherapies: a network meta-analysis. *Circulation* 2014;129:28–41.
15. Psaty BM, Breckenridge AM. Mini-Sentinel and regulatory science-big data rendered fit and functional. *N Engl J Med* 2014;370:2165–7.
16. Takizawa O, Urushihara H, Tanaka S, Kawakami K. Utilization trend and management of oral anticancer medicines using nationwide pharmacy database and questionnaire survey at pharmacies in Japan. In: Proceedings of the 30th International Conference on Pharmacoepidemiology & Therapeutic Risk Management 2014. Taipei, Taiwan.
17. Suissa S. Immortal time bias in observational studies of drug effects. *Pharmacoepidemiol Drug Saf* 2007;16:241–9.
18. Kimura S, Sato T, Ikeda S, Noda M, Nakayama T. Development of a database of health insurance claims: standardization of disease classifications and anonymous record linkage. *J Epidemiol* 2010;20:413–9.
19. Hashikata H, Harada KH, Kagimura T, Nakamura M, Koizumi A. Usefulness of a large automated health records database in pharmacoepidemiology. *Environ Health Prev Med* 2011;16:313–9.
20. Hall GC, Sauer B, Bourke A, Brown JS, Reynolds MW, LoCasale R. Guidelines for good database selection and use in pharmacoepidemiology research. *Pharmacoepidemiol Drug Saf* 2012;21:1–10.
21. Schneeweiss S. A basic study design for expedited safety signal evaluation based on electronic healthcare data. *Pharmacoepidemiol Drug Saf* 2010;19:858–68.
22. Whitaker HJ, Farrington CP, Spiessens B, Musonda P. Tutorial in biostatistics: the self-controlled case series method. *Stat Med* 2006;25:1768–97.
23. Ueyama H, Hinotsu S, Tanaka S, et al. Application of a self-controlled case series study to a database study in children. *Drug Saf* 2014;37:259–68.
24. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Epidemiology* 2007;18:805–35.
25. Carnahan RM. Mini-Sentinel's systematic reviews of validated methods for identifying health outcomes using administrative data: summary of findings and suggestions for future research. *Pharmacoepidemiol Drug Saf* 2012;21 (Suppl 1):90–9.
26. Herman RA, Gilchrist B, Link B, Carnahan RM. A systematic review of validated methods for identifying lymphoma using administrative data. *Pharmacoepidemiol Drug Saf* 2012;21(Suppl 1):203–12.
27. Ooba N, Setoguchi S, Ando T, et al. Claims-based definition of death in Japanese claims database: validity and implications. *PLoS ONE* 2013;8:e66116.
28. Greenland S. Invited commentary: variable selection versus shrinkage in the control of multiple confounders. *Am J Epidemiol* 2008;167:523–9.
29. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology* 1999;10:37–48.
30. VanderWeele TJ, Shpitser I. A new criterion for confounder selection. *Biometrics* 2011;67:1406–13.
31. Pharmaceuticals and Medical Devices Agency. Guidelines for pharmacoepidemiological studies using databases of medical or other information in pharmacovigilance. http://www.info.pmda.go.jp/kyoten_iyaku/db_bunseki.html (1 July 2014, date last accessed).
32. The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance. Guide on Methodological Standards in Pharmacoepidemiology. http://www.encepp.eu (1 July 2014, date last accessed).