

Analysis of individual drug use as a time-varying determinant of exposure in prospective population-based cohort studies

B. H. Ch. Stricker · T. Stijnen

Received: 25 February 2010 / Accepted: 17 March 2010 / Published online: 1 April 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract In pharmaco-epidemiology, the use of drugs is the determinant of interest when studying exposure-outcome associations. The increased availability of computerized information about drug use on an individual basis has greatly facilitated analyses of drug effects on a population-based scale. It seems likely that many negative findings in the early days of pharmaco-epidemiology can be explained by non-differential misclassification because of too simple (yes/no) exposure measures. In this paper, the authors discuss the importance of an adequate definition of drug exposure in pharmaco-epidemiological research and how this time-varying determinant can be analyzed in cohort studies. To reduce the risk of non-differential misclassification, a precise definition of exposure is mandatory and it is important to distinguish the complete follow-up period of a population into mutually exclusive episodes of non-use, past use and current use for each individual. By analyzing exposure to drugs as a time-dependent variable in a Cox regression model, cohort studies with complete coverage of all filled prescriptions can provide us

with valid and precise risk estimates of drug-outcome associations. However, such estimates may be biased in the presence of time-dependent confounders which are themselves affected by prior exposure.

Keywords Pharmaco-epidemiology · Drug exposure analysis · Cohort studies · Cox regression models for time-varying determinants

Introduction

In pharmaco-epidemiology, the use of drugs is the determinant of interest when studying exposure-outcome associations. The increased availability of computerized information about drug use on an individual basis has greatly facilitated analyses of drug effects on a population-based scale. This stimulated the development of pharmaco-epidemiology as a branch within epidemiology as can be seen on epidemiological congresses [1].

In the last decades of the preceding century, many pharmaco-epidemiologic studies employed a case-control design. Case-control studies of the association between diethylstilboestrol against habitual abortion and vaginal carcinoma in female offspring [2], and of salicylate-attributed Reye syndrome [3] have clearly shown the benefits of this design thanks to their valid and relevant results despite small numbers of cases in both studies. However, in hindsight this design may also have risen spurious associations such as that between use of reserpine and breast cancer [4].

Because information about drug use in these early studies had to be obtained from interview, they suffered from two important potential limitations. First, the suspicion that patients with severe diseases would tend to have a

B. H. Ch. Stricker (✉)
Department of Epidemiology, Erasmus Medical Center, PO Box 2040, 3000 CA Rotterdam, The Netherlands
e-mail: b.stricker@erasmusmc.nl

B. H. Ch. Stricker
Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands

B. H. Ch. Stricker
Department of Medical Informatics, Erasmus Medical Center, Rotterdam, The Netherlands

T. Stijnen
Department of Biostatistics, University Medical Center Leiden, Leiden, The Netherlands

better recall than non-diseased controls. This type of information bias (so-called ‘recall bias’) was suspected to be partly responsible for the association between X-ray exposure during pregnancy, and the subsequent occurrence of congenital malformations in off-spring [5]. In a subsequent similar study with hospital records, the significant risk increase of 90% found earlier was brought back to 40%, albeit still statistically significant [6]. Unfortunately, even hospital records may be incomplete when information of drug use is retrospectively gathered [7]. Second, whereas recall bias leads to differential (i.e., non-random) misclassification of exposure, also non-differential (i.e., random) misclassification may jeopardize the validity of drug-effect estimates as can be seen in Fig. 1. The argument which is often brought forward that such misclassification is less serious because it leads to a more conservative estimate, is questionable as also underestimations are non-valid and therefore unwanted. It seems likely that many negative findings in the early days of pharmaco-epidemiology can be explained by non-differential misclassification because of too simple (yes/no) exposure measures. But even with extensive scrutiny of drug history during interview, long-term exposure such as might be relevant to development of end-stage renal failure [8] will inevitably lead to both differential and non-differential exposure misclassification. It would not be fair to blame researchers for this, as laborious investigation of serious diseases should be praised and because most of them do not ignore the limitations of their data.

In this paper, we discuss the importance of an adequate definition of drug exposure in pharmaco-epidemiological research and how this time-varying determinant can be analyzed in cohort studies.

Non-differential misclassification of exposure

	Disease	No disease	
Drug A	30 (28)	70 (72)	100
	↓ 10% ↑	↓ 10% ↑	
	a	b	
	c	d	
Drug B	10 (12)	90 (88)	100

Fig. 1 Assume a prospective cohort study with 10% non-differential misclassification of exposure. Without misclassification, the true relative risk would be calculated as $[30/100]/[10/100] = 3.0$ in those treated with drug A in comparison to those on drug B. However, because in diseased as well as non-diseased 10% of individuals switches from A → B and vice versa, the number of exposed changes for each cell, i.e. cell a ($30 - 3 + 1 = 28$); cell b ($70 - 7 + 9 = 72$); cell c ($10 - 1 + 3 = 12$); cell d ($90 - 9 + 7 = 88$). Consequently, we calculate the relative risk as $[28/100]/[12/100] = 2.3$

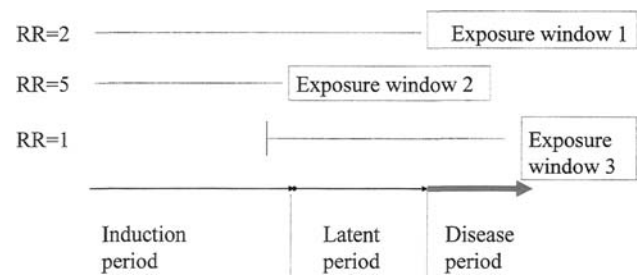


Fig. 2 Assume a prospective cohort study on the risk of cancer induction after long-term exposure to a drug A. To reduce the effect of non-differential misclassification as much as possible, cumulative exposure should only be calculated during the induction period (*exposure window 2*). If use during the latent period is included (*exposure window 1*), the measured relative risk (RR) will be diluted from $RR = 5$ towards $RR = 2$. This will be brought down even further to a $RR = 1$ if cumulative exposure is only measured during the latent and disease period (*exposure window 3*)

Drug assessment in populations

The two types of above-mentioned exposure misclassification can be effectively circumvented when drug use in populations is prospectively gathered, i.e., before onset of the event of interest and consequently unbiased. Prospective data collection prevents the occurrence of ‘recall bias’ or other types of differential exposure information bias. Non-differential misclassification, however, can still occur if the exposure definition is imprecise or if drug use falls outside the biologically plausible risk window, as can be seen in Fig. 2. Although we know that prolonged use of high doses of estrogens is a strong risk factor for liver adenoma [9], for instance, it is clear that in a woman who started using the drug 1 month before diagnosis, liver adenoma can not be caused by the estrogen. Apparently, exposure status is not merely a matter of using/non-using a drug before the event occurs but also of timing. In real life situations, the effect of exposure will depend on dose, duration of use, timing in relationship to the event, concurrent medication, and adherence to therapy.

Of course, the fundamental question is: “what is exposure?”. In an ideal situation, it tells us what the drug concentration is at the receptor site which is responsible for the biological effect. Obviously, such a situation can be achieved at best in small groups of diseased or volunteers and even there, it is a rather theoretical option and is approached nearest with the assessment of drugs and their metabolites in blood or plasma. As blood levels tell us little about, for instance, the concentration of psychotropic drugs in brain tissue, the contribution to large-scale population-based pharmaco-epidemiology is modest although it will add to interpreting the study results. Instead, large studies rely on three types of health care data which are currently becoming more and more readily available: drug exposure information from health insurance companies or sick funds

[10]; from general practitioners [11], and from pharmacies. Data from general practitioners are of great value because they consist of both drug prescriptions and information on disease, and can not only show changes in prescribing [11] but can also be pooled in large-scale collaborative research [12] to study rare events. However, a limitation is that specialist prescriptions are incompletely registered in general practice databases, and that not all of their own prescriptions are filled at pharmacies. As for health insurance companies, not all of them register the daily dose of the patient. Consequently, drug filling data from pharmacies come closest to true exposure of prescription-only drugs. This holds especially in those who use chronic medication. People who come back at regular intervals for refills of a drug, usually have a high compliance to therapy. Dividing the number of filled tablets or capsules by the daily prescribed number makes it possible to calculate when a patient is expected to come for the next vial of medicines. In this way, adherence to therapy can be assessed reliably. All mentioned data sources, however, miss over-the-counter (OTC) drugs.

Converting filled prescriptions into exposure variables

For the analysis of drug-event associations, the dose, the duration, and timing of use are highly important. Unfortunately, individual medication histories may contain a plethora of different drugs, doses, switching between drugs, and types of administration. Most western countries have some 10,000 different marketed drug products and it is an administrative and pharmacological challenge to convert this information into an analyzable dataset. Suppose, for instance, that one would be interested to study the association between long-term use of anti-inflammatory drugs (NSAIDs) and cancer. Then, use of several different pharmaceutical products over the years by one person has to be brought back to the numbers of days of use of each pharmacological entity, and to a standardized dose to facilitate comparison between products. For instance, the recommended daily dose for treatment of arthritic pain is 100 mg for diclofenac and 500 mg for naproxen. Taking the average dose without standardization would be meaningless. A well-known scheme for dose standardization is the ATC-DDD scheme of the World Health Organisation (<http://www.whooc.no>).

Drug-event analyses

Epidemiologists usually underline the importance of awareness of potential confounding in study designs to prevent non-validity [13]. For obvious reasons, appropriate

epidemiologic methods are a prerequisite for valid study results. This includes the validation of exposure measurement tools [14]. An adequate assessment of the role of drug exposure requires knowledge of the biologically relevant period during which the drug must be used to induce or modify the event of interest, and of the pharmacokinetic and pharmacodynamic properties of the drug. If we are interested in the question whether a drug may cause an event, we should only assess the exposure status during the induction period as any assessment outside this period will introduce non-differential misclassification of exposure (Fig. 2). On the other hand, if we investigate whether a drug is not a cause but modifies the disease process, we will only assess the exposure status during the latent and/or disease period. The pharmacodynamic effects of the drug should be compatible with causation, although this may not always be clear. The pharmacokinetic properties of a drug are very important for assessing the duration of exposure. For instance, a drug against cystitis such as nitrofurantoin is excreted completely within hours while the notorious carcinogenic diagnostic agent thorium dioxide which was used between the 30 and 50 s of the 20th century has a biological half-life of 400 years and a physical half-life of 5,000 years. Drugs such as the anti-arrhythmic agent amiodaron and the anxiolytic diazepam have prolonged carry-over effects because of their long half-life. Apparently, the exposure period is not merely the time during which the drug was actually taken but may have to be extended with a carry-over period of 1–2 half-lives.

Analysis of drug exposure as a time-varying determinant in prospective population-based cohort studies

As mentioned above, the prospective gathering of drug use facilitates unbiased risk estimates provided exposure is precisely defined by reference to a well-defined event with a clearly recognizable onset. Thanks to prospectively gathered and complete medication histories, exposure status can be assessed on every day of the follow-up. This is a great advantage over population-based studies where drug use is assessed on the basis of interview during repeated rounds of cross-sectional measuring. Although the analyses in this paper pertain to cohort studies, this includes nested case–control studies where the prospective exposure data come from the cohort but there are efficiency reasons to perform a case–control analysis. This may occur, for instance, when tissue samples have to be taken or when additional data gathering from medical records makes it unfeasible to perform this in the whole study cohort.

Unlike constant features such as sex and certain genetic traits, drug use is essentially a time-varying determinant. In a traditional Cox regression model [15], the hazard function in the total population at time point t is defined as:

$$\lambda(t) = \lambda_0(t) * \exp(\beta x) \quad (1)$$

in which $\lambda(t)$ represents the event rate at time t conditional on being still event free before time t . In this model the event rate is assumed to be equal to a baseline risk $\lambda_0(t)$, which is the same for everybody in the population, i.e., independent of the determinants. This baseline risk is multiplied by a term $\exp(\beta x)$, dependent on the determinants x , which are different between individuals. The parameters β quantify the effect of the determinants on the event rate. They have to be estimated from the data, together with the baseline risk $\lambda_0(t)$. There are different choices possible for the time scale t , for instance, $t = \text{age}$ (when age is strongly and exponentially associated with event occurrence), $t = \text{time since entry in the cohort}$, or $t = \text{calendar time}$. In the simplest case, x represents only one determinant x_1 , for instance sex, with $x_1 = 1$ (males) and $x_1 = 0$ (females). Then $\lambda(t)$ gives the hazard function for developing the event at time point t in males or females. For females, the hazard is $\lambda_0(t)$ and for males the hazard is $\lambda_0(t)$ multiplied by $\exp(\beta_1)$. In this model, the determinants x are not necessarily constant during follow-up, but may vary in time, such as drug use. In a study sample, the unknown hazards are then estimated from the data as:

$$h(t) = h_0(t) * \exp(b_1 x_1) \quad (2)$$

Suppose that we performed a study in which m individuals developed the event of interest during follow-up. The follow-up times at which the events (the ‘‘cases’’) occurred are denoted with t_1, \dots, t_m . (for simplicity, we assume that events do not coincide). In a Cox proportional hazards regression analysis with drug exposure as a time-varying determinant [16], the exposure status $x_{1j}[t_j]$ on the index day t_j of the case number j , is compared to the exposure status of all other cohort members on the same day of the follow-up. In this way, $j = 1, \dots, m$ strata are formed of one case each and the other cohort members who were still in the follow-up and event free at time t_j as controls. In an earlier analysis in The Rotterdam Study [17], for instance, it was investigated whether thiazide diuretics protect against hip fracture, thanks to their calcium-retaining effect [18]. An analytical matrix would look like the ones given in Table 1a and b. On the index date t_j , all cohort members have a history of thiazide use up to that time. In its simplest form, we can characterize this history as use on the index date as 1 (‘yes’) or 0 (‘no’) like in Table 1a. If i denotes the number of an arbitrary cohort member that is under follow-up at the index date t_j , the model states for this individual i

$$h(t_i) = h_0(t_j) * \exp(b_1 x_{1i}[t_j]) \quad (3)$$

in which the time-varying determinant $x_{1i}[t_j]$ has the numerical value ‘1’ (exposed) or ‘0’ (unexposed) depending on whether cohort member i is exposed or non-exposed at time point t_j . For each event time t_j , there is a set R_j (the ‘‘risk set’’) containing all individuals who were under observation at t_j . So, R_j contains case number j and its corresponding controls. Given the event at t_j , the conditional probability that out of all cohort members in R_j the cohort member with number j (the one who was observed to develop the event) will develop the event is:

$$h_0(t_j) * \exp(b_1 x_{1j}[t_j]) / \sum_{i \text{ from } R_j} \{h_0(t_j) * \exp(b_1 x_{1i}[t_j])\} \quad (4)$$

Notice that the baseline hazard rate $h_0(t_j)$ is present in numerator and denominator and cancels out. Therefore, the conditional likelihood function of all the data, defined as the product of the probabilities as given in (4) over all event times t_j , is equal to:

$$L(\beta) = \prod_{j=1}^m \{ \exp(b_1 x_{1j}[t_j]) / \sum_{i \text{ from } R_j} \exp(b_1 x_{1i}[t_j]) \}$$

However, this straightforward but simple analysis with only the status exposed/unexposed would mean that we use only a very limited part of the information that is contained in the thiazide use history and thereby introduce non-differential misclassification. As can be seen in Table 1a, knowing the numbers of days of continuous use on the index date of each cohort member facilitates calculation of more valid risk estimates as it is unlikely that only 1 day of thiazide use would already be protective while the model would consider even those people as exposed who started thiazides 1 day before the index date. There are pharmacological reasons to assume that a protective effect on hip fracture may become visible only after at least 6 weeks of calcium retention by thiazide treatment and reaches a maximum after ~ 1 year. Hence, more information, and consequently less non-differential misclassification, is obtained with the introduction of extra determinants x_2 , and x_3 , where the cumulative continuous exposure to thiazides at the index date is categorized as: $x_1 = 1$ through 42 days; $x_2 = 42$ through 365 days; $x_3 > 365$ days.

$$h(t) = h_0(t) * \exp(b_1 x_1[t] + b_2 x_2[t] + b_3 x_3[t]) \quad (5)$$

In this way, the risk for these two exposure categories is expressed in comparison to non-use and yields a more valid representation of the drug-event association than in (3) or when thiazide exposure in days would be introduced as a continuous exposure determinant.

Even more information, and therefore less non-differential misclassification, may follow from the introduction

Table 1 Apart from unique patient number, sex and age in years, the columns respectively represent: case status (1 = 'yes'; 0 = 'no'); stratum; follow-up in days; cumulative number of days of current use; number of days since last intake in past users; defined daily dose (DDD) [for hydrochlorothiazide: 25 mg and for chlorothiazide: 500 mg]; and total numbers of days of use since study entry

Patient	Sex	Age	Case	Stratum	Follow-up	Current use			
<i>(a)</i>									
4417001	V	82	1	1	961	0			
6593001	V	88	0	1	961	0			
1101001	V	93	0	1	961	0			
3000001	M	81	0	1	961	0			
5135001	V	86	0	1	961	0			
1720215	V	88	0	1	961	1			
6367517	V	86	0	1	961	0			
2191001	V	74	0	1	961	1			
1033001	V	87	0	1	961	0			
7112001	F	88	1	2	1,253	1			
1376809	M	94	0	2	1,253	0			
Patient	Sex	Age	Case	Stratum	Follow-up	Current use	Past use		
<i>(b)</i>									
4417001	V	82	1	1	961	0	0		
6593001	V	88	0	1	961	0	0		
1101001	V	93	0	1	961	0	90		
3000001	M	81	0	1	961	0	0		
5135001	V	86	0	1	961	0	0		
1720215	V	88	0	1	961	154	0		
6367517	V	86	0	1	961	0	0		
2191001	V	74	0	1	961	83	0		
1033001	V	87	0	1	961	0	0		
7112001	F	88	1	2	1,253	34	0		
1376809	M	94	0	2	1,253	0	0		
Patient	Sex	Age	Case	Stratum	Follow-up	Current use	Past use	DDD	Total use
<i>(c)</i>									
4417001	V	82	1	1	1,061	0	0	–	0
6593001	V	88	0	1	1,061	0	0	–	0
1101001	V	93	0	1	1,061	0	90	–	387
3000001	M	81	0	1	1,061	0	0	–	0
5135001	V	86	0	1	1,061	0	0	–	0
1720215	V	88	0	1	1,061	154	0	1.2	234
6367517	V	86	0	1	1,061	0	0	–	0
2191001	V	74	0	1	1,061	83	0	0.9	83
1033001	V	87	0	1	1,061	0	0	–	0
7112001	F	88	1	2	1,253	34	0	1.7	731
1376809	M	94	0	2	1,253	0	0	–	0

Note that the total use is mostly higher than current use because patients may have used thiazides intermittently

of determinants for 'past use' when one expects that a carry-over period should be taken into account. For instance, if thiazide use for more than 1 year results in a higher calcification of the hip, it will take a certain time period before discontinuation of thiazides results in returning to the

situation before starting treatment. This can be done by introducing determinants for past exposure, defined as the number of days since last intake of thiazides (Table 1b). In the previously mentioned study, additional categorical determinants were created as extra determinants x_4 , x_5 , and

x_6 where the number of days since last intake of thiazides, counting back from the index date, was categorized as: $x_4 = 1$ through 60 days; $x_5 = 61$ through 365 days; and $x_6 > 365$ days.

For obvious reasons, the determinants x_1 through x_6 should be introduced in one model. After all, it is important that the complete follow-up time of each study member is expressed in mutually exclusive episodes of non-use, past use, and current use to decrease the degree of non-differential misclassification as much as possible. Then, the full model is:

$$h(t) = h_0(t) * \exp(b_1x_1[t] + b_2x_2[t] + b_3x_3[t] + b_4x_4[t] + b_5x_5[t] + b_6x_6[t]) \quad (6)$$

This model can be extended with the inclusion of other non-time-varying determinants such as gender and baseline age $x_a, x_b, \dots, x_j, \dots, x_z$, provided usual precautions against overfitting of the model are taken into account. Adjusting for dosage may be performed by including it as a continuous determinant in mg/day or categorized, for instance by splitting current use as: current use with >1 defined daily dose (DDD); current use with ≤ 1 DDD.

The analytical matrix in Table 1c facilitates different type of analyses. For instance, would we be interested to find out whether cumulative use of nonsteroidal antiinflammatory drugs (NSAID) are associated with an increased risk of cancer, we might prefer to use the determinant 'total use'. However, if we would be interested in induction, rather than promotion, we might subtract a theoretical episode of 5 years from the index date of cancer diagnosis and calculate total use in days until that date, or in dose as cumulative DDDs. We would do this to avoid non-differential misclassification by restricting ourselves to the induction period. Would we only be interested in promotion, we would treat NSAID as an effect modifier and restrict our analysis to total use in the 5 years before cancer diagnosis because we would expect that malignant cells would already be present during that latent period.

Limitations

The method described above facilitates a clear insight into the data structure but may have some practical limitations. First, in patients who use drugs very irregularly, it may be difficult to calculate the cumulative period of continuous current use at the index date as in such patients these periods will usually be short and irregular. However, this can often be circumvented by combining current and total use. Second, because for every case the remainder of non-censored cohort members serves as a reference, huge strata

may lead to substantial computational time to run analyses. For instance, with 1000 cases in a cohort of 50,000 people, each stratum would have slightly less than 50,000 observations at the index date of that stratum, leading to a data file of $\sim 50,000,000$ records. As there are techniques to deal with such a problem, however, this may only be relevant to the less well-equipped researcher.

A methodological limitation may arise when the model is adjusted for a time-dependent co-variable which is a risk factor for the event of interest and may be influenced by the drug exposure [19]. Standard methods for estimating the effect of a time-varying exposure on survival may be biased in the presence of time-dependent confounders which are themselves affected by prior exposure. This problem can be overcome by inverse probability weighted estimation of Marginal Structural Cox Models (Cox MSM) or G-estimation of Structural Nested Cumulative Failure Time Models (SNCFTM). For this situation, the reader is referred to recent literature about such a scenario [20].

Conclusion

In pharmaco-epidemiology, the use of drugs is the determinant of interest when studying exposure-effect associations. It seems likely that many negative findings in the early days of pharmaco-epidemiology can be explained by non-differential misclassification because of too simple (yes/no) exposure measures. To reduce the risk of non-differential misclassification, a precise definition of exposure is mandatory and it is important to distinguish the complete follow-up period of a population into mutually exclusive episodes of non-use, past use and current use for each individual. By analyzing exposure to drugs as a time-dependent variable in a Cox regression model, cohort studies with complete coverage of all filled prescriptions can provide us with valid and precise risk estimates of drug-outcome associations. However, such estimates may be biased in the presence of time-dependent confounders which are themselves affected by prior exposure.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. IEA-EEF European Congress of Epidemiology 2009 Warsaw, Poland 26–29 August. *Eur J Epidemiol* 2009;24 Suppl 1.
2. Herbst AL, Ulfelder H, Poskanzer DC. Adenocarcinoma of the vagina. Association of maternal stilbestrol therapy with tumor appearance in young women. *N Engl J Med.* 1971;284:878–81.

3. Hurwitz ES, Barrett MJ, Bregman D, et al. Public health service study of Reye's syndrome and medications. Report of the main study. *JAMA*. 1987;257:1905–11.
4. Heinonen OP, Shapiro S, Tuominen L, Turunen MI. Reserpine use in relation to breast cancer. *Lancet*. 1974;2:675–7.
5. Stewart A, Webb J, Hewitt D. A survey of childhood malignancies. *Br Med J*. 1958;1:1495–508.
6. MacMahon B. Prenatal X-ray exposure and childhood cancer. *J Nat Cancer Inst*. 1962;28:1173.
7. Buring JE, Bain CJ, Hennekens CH. Alternative data sources in a case-control study of conjugated estrogens and cancer. *Am J Prev Med*. 1986;2:116–21.
8. Perneger TV, Whelton PK, Klag MJ. Risk of kidney failure associated with the use of acetaminophen, aspirin, and nonsteroidal anti-inflammatory drugs. *N Eng J Med*. 1994;331:1675–9.
9. Rooks JB, Ory HW, Ishak KG, Strauss LT, Greenspan JR, Hill AP, et al. Epidemiology of hepatocellular adenoma. The role of oral contraceptive use. *JAMA*. 1979;242:644–8.
10. Winkelmayr WC, Bucsics AE, Schautzer A, Wieninger P, Pogantsch M. Use of recommended medications after myocardial infarction in Austria. *Eur J Epidemiol*. 2008;23:153–62.
11. Hsia Y, MacLennan K. Rise in psychotropic drug prescribing in children and adolescents during 1992–2001: a population-based study in the UK. *Eur J Epidemiol*. 2009;24:211–6.
12. Thompson A. Thinking big: large-scale collaborative research in observational epidemiology. *Eur J Epidemiol*. 2009;24:727–31.
13. Groenwold RHH, Hoes AW, Hak E. Confounding in publications of observational intervention studies. *Eur J Epidemiol*. 2008;22:413–5.
14. Stang A. Appropriate epidemiologic methods as a prerequisite for valid study results. *Eur J Epidemiol*. 2008;23:761–5.
15. Cox DR. Regression models and life tables. *J R Stat Soc B*. 1972;34:187–220.
16. Clayton D, Hills M. *Statistical models in epidemiology*. New York: Oxford University Press; 1993.
17. Hofman A, Breteler MMB, van Duijn CM, Janssen HLA, Krestin GP, Kuipers EJ, et al. The Rotterdam study: 2010 objectives and design update. *Eur J Epidemiol*. 2009;24:553–72.
18. Schoofs MW, van der Klift M, Hofman A, de Laet CE, Herings RM, Stijnen T, et al. Thiazide diuretics and the risk for hip fracture. *Ann Intern Med*. 2003;139:476–82.
19. Robins JM. Estimation of the time-dependent accelerated failure time model in the presence of confounding factors. *Biometrika*. 1992;79:321–34.
20. Young JG, Hernán MA, Picciotto S, Robins JM. Relation between three classes of structural models for the effect of a time-varying exposure on survival. *Lifetime Data Anal*. 2009 Nov 6 (Epub ahead of print).