BCPT Basic & Clinical Pharmacology & Toxicology

Basic & Clinical Pharmacology & Toxicology Doi: 10.1111/bcpt.12946

MiniReview

Considerations for Pharmacoepidemiological Studies of Drug–Cancer Associations

Anton Pottegård[1], Søren Friis[2,3], Til Stürmer[4], Jesper Hallas[1,5] and Shahram Bahmanyar[6,7]

[1]Clinical Pharmacology and Pharmacy, Department of Public Health, University of Southern Denmark, Odense, Denmark, [2]Danish Cancer Society Research Center, Danish Cancer Society, Copenhagen, Denmark, [3]Department of Public Health, University of Copenhagen, Copenhagen, Denmark, [4]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, [5]Department of Clinical Biochemistry and Pharmacology, Odense University Hospital, Odense, Denmark, [6]Department of Medicine, Centre for Pharmacoepidemiology, Solna, Karolinska Institutet, Stockholm, Sweden and [7]Clinical Epidemiology Unit, Department of Medicine, Solna, Karolinska Institutet, Stockholm, Sweden

(Received 12 September 2017; Accepted 11 December 2017)

Abstract: In this MiniReview, we provide general considerations for the planning and conduct of pharmacoepidemiological studies of associations between drug use and cancer development. We address data sources, study design, assessment of drug exposure, ascertainment of cancer outcomes, confounder adjustment and future perspectives. Aspects of data sources include assessment of complete history of drug use and data on dose and duration of drug use, allowing estimates of cumulative exposure. Outcome data from formal cancer registries are preferable, but cancer data from other sources, for example, patient or pathology registries, medical records or claims are also suitable. The two principal designs for observational studies evaluating drug–cancer associations are the cohort and case–control designs. A key challenge in studies of drug–cancer associations is the exposure assessment due to the typically long period of cancer development. We present methods to examine early and late effects of drug use on cancer development and discuss the need for employing 'lag-time' in order to avoid reverse causation. We emphasize that a new-user study design should always be considered. We also underline the need for 'dose–response' analyses, as drug–cancer associations are likely to be dose-dependent. Generally, studies of drug–cancer associations should explore risk of site-specific cancer, rather than cancer overall. Additional differentiation may also be crucial for organ-specific cancer with various distinct histological subtypes (e.g., lung or ovary cancer). We also highlight the influence of confounding factors and discuss various methods to address confounding, while emphasizing that the choices of methods depend on the design and specific objectives of the individual study. In some studies, use of active comparator(s) may be preferable. Pharmacoepidemiological studies of drug–cancer associations are expected to evolve considerably in the coming years, due to the increasing availability of long-term data on drug exposures and cancer outcomes, the increasing conduct of multinational studies, allowing studies of rare cancers and subtypes of cancer, and methodological improvements specifically addressing cancer and other long-term outcomes.

Use of prescription and over-the-counter drugs represents exogenous exposures that may result in either increase or reduction in cancer risk. Clear associations have been established for a number of drugs, for example, the preventive effect of aspirin use against colorectal cancer [1,2] or the increased risk of renal cancer with use of phenacetin [3,4]. Further, new hypotheses often arise, such as the recent concerns about carcinogenic effects of lithium [5–7] and pioglitazone [8–12]. A significant challenge in the elucidation of drug effects on cancer development is that the effects typically first become manifest several years after drug initiation. The long period of cancer development and the relatively low incidence of most individual cancer types impede the ability of traditional pharmacovigilance systems, notably spontaneous reporting of adverse effects, to identify drug–cancer associations. Consequently, analyses based on large-scale healthcare data sets are essential to provide solid data on potential drug effects on cancer incidence.

The public health importance of identifying carcinogenic effects of drugs is apparent, as even small carcinogenic effects of widely used drugs will result in many additional cancer cases. Only about a dozen drugs have been established as '*definitely carcinogenic to humans*' (Group 1) by the International Agency for Research on Cancer (IARC) [13,14]. However, about 50 drugs are currently classified as '*possible or probably carcinogenic*' (Group 2A/2B), and additional studies are necessary to confirm or disprove these suspicions [15]. Importantly, and often ignored, pharmacoepidemiological studies may also substantiate lack of carcinogenicity for specific drugs. This has considerable value by reassuring prescribers and patients of the safety of drugs and thus promoting their appropriate use, which may be compromised by preliminary reports of carcinogenicity from a variety of sources, including case series, adverse event reporting systems or animal experiments. Finally, identification

Author for correspondence: Anton Pottegård, Clinical Pharmacology and Pharmacy, Department of Public Health, University of Southern Denmark, J.B. Winsløws Vej 19, 2nd, 5000 Odense C, Denmark (e-mail apottegaard@health.sdu.dk).

of potential beneficial effects of drug use on cancer risk or prognosis may provide important evidence that can be pursued further in experimental and intervention studies, as well as provide clues to the development of new agents for medical cancer prevention and treatment.

In the present MiniReview, we provide general considerations on the conduct of pharmacoepidemiological studies of drug–cancer associations. Importantly, while we are aiming to describe aspects that should be considered when planning such studies, our suggestions should be viewed as guidance rather than as mandatory requirements. Specifically, we address choices of (i) data sources and (ii) study design, (iii) assessment of drug exposure, (iv) ascertainment of cancer outcomes, (v) confounder adjustment and (vi) future perspectives.

Considerations Regarding Data Sources

In pharmacoepidemiological studies, the exposure is usually not occurring as a single episode, but prolonged and variable over time, during one or more exposure periods and thus needs to be handled in a time-dependent manner. Therefore, the most detailed drug use history should be sought – such as claims data on dispensed prescriptions or population-based prescription data from drug registries. Apart from dates of prescription or dispensing, data on dose or duration are also necessary to estimate cumulative exposure. In designing studies to evaluate effects of a particular treatment on cancer incidence, it should be considered that individual cancer types have various induction and latent periods [16,17]. The relevant exposure periods for different drugs thus vary in relation to the time from initiation to manifest malignancy of specific cancers (induction period) or from initiation to diagnosis (latent period). While these induction and latent periods are usually unknown, long-term data on drug exposure and long follow-up allow us to make different assumptions about the relevant exposure periods.

Another challenge in the assessment of drug–cancer associations is the low incidence of many cancer types, thus requiring very large sample sizes. One way to overcome this challenge is to perform multi-site, multi-database or multinational studies. However, such initiatives involve several practical challenges, including legal issues and the lack of established common data models.

Minimum requirement for outcome assessment is to have individual-level data on cancer incidence. Although data from population-based cancer registries are preferred, information from other sources such as diagnostic or treatment records may be sufficient. It is also important to obtain data on reasons of loss to follow-up – including migration or insurance gaps. Information on death and cause of death are helpful to deal with competing risks. It is also essential to have information on potentially important confounding factors. Some factors such as sex and age are necessary in all studies, while other factors depend on the specific study objectives and design.

As an example of registries useful for studies on drug–cancer associations, the Nordic countries (i.e., Denmark, Finland, Norway, Iceland and Sweden) have nationwide registries containing continuously updated, complete, individual-level data on, for example, prescription drugs [18,19] and incident

Table 1.
The authors' top 10 'suggested reading' relevant to pharmacoepidemiological studies of drug–cancer associations (in non-prioritized order)

#	Authors	Title	Topic	Citation
01	Walker	For drug-induced carcinogenesis, the observations are the hypothesis. Invited editorial for the Mini-Symposium on Cancer Pharmacoepidemiology	Basic concepts	Ann Epidemiol 2016; 26(11):749-750
02	Pinheiro *et al.*	Challenges in evaluating cancer as a clinical outcome in postapproval studies of drug safety	Basic concepts	Ann Epidemiol 2016; 26(11):735-740
03	Friis *et al.*	European Code against Cancer 4th edition: Medical exposures, including hormone therapy, and cancer	Basic concepts	Cancer Epidemiol 2015; 29 Suppl 1:S107-19
04	Rivera *et al.*	Connections between pharmacoepidemiology and cancer biology: designing biologically relevant studies of cancer outcomes	Exposure ascertainment	Ann Epidemiol 2016; 26(11):741-745
05	Pottegård & Hallas	New use of prescription drugs prior to a cancer diagnosis	Reverse causation	Pharmacoepidemiol Drug Saf 2017;26(2):223-227
06	Umar *et al.*	Future directions in cancer prevention	Cancer biology	Nat Rev Cancer 2012; 12 (12): 835–848
07	Rothman	Induction and latent periods	Induction/latency	Am J Epidemiol 1981; 114(2):253-9
08	Pottegård *et al.*	Identification of associations between prescribed medications and cancer: a nationwide screening study	Hypothesis generation	EBioMedicine 2016;7:73-9
09	Lund *et al.*	The active comparator, new-user study design in pharmacoepidemiology: historical foundations and contemporary application	The active comparator, new-user design	Curr Epidemiol Rep 2015; 2(4):221-228
10	Stürmer *et al.*	Adjustments for unmeasured confounders in pharmacoepidemiological database studies using external information	Advanced confounder adjustment	Med Care 2007; 45(10 Supl 2):S158-98

cancers [20]. Demographic (including socio-economic) data and hospital diagnoses are also available and can be used for confounder adjustment. The unique personal identification number, assigned to all residents in each Nordic country at birth or immigration, allows for individual-level linkage of registers and databases continuously over time. Key prescription data include the Anatomical Therapeutic Chemical (ATC) classification [21], formulation, package size and number, number of defined daily doses (DDD) per package and date of purchase. Further, in all Nordic countries, reporting of all incident cancer cases to cancer registries is mandatory by healthcare providers. The cancer data include the primary site of the cancer, time of diagnosis, diagnosis code according to the International Classification of Diseases (ICD), currently version 10 (ICD-10) and histology and stage of the malignancy.

In the United States, the introduction of drug coverage for older adults, Medicare Part D, in 2006 has facilitated use of the Medicare fee-for-service data for pharmacoepidemiological studies on cancer incidence [22,23]. The Medicare data cover inpatient (Part A) and outpatient (Part B) diagnoses and procedures, as well as prescription data at the pharmacy level (Part D). The advantages of the US Medicare data include population-based and nationwide coverage, size (over 20 million older adults with fee-for-service data within Parts A, B and D coverage), age range (generally high morbidity, drug use and cancer incidence), low turnaround (the great majority stay in these plans until death) and the linkage to US death data. Limitations include the lack of information before age 65, issues related to changing co-payments during the calendar year (the 'doughnut hole') and time to release (~18 months). Medicare data can be linked to other data sources (e.g., the National Death Index, cancer registries) but at considerable investment (both time and money).

Considerations for the Choice of Study Design

The two principal observational designs relevant for studying drug–cancer associations are the cohort design and the case–control design. In the cohort design, drug users are compared with non-users or users of comparator drug(s) and followed over time with respect to the outcome of interest. In the case–control design, persons with the outcome of interest, that is, cancer, are compared with persons without cancer with respect to their history of drug use. The cohort and case–control designs are observational; that is, the researchers do not interfere with the drug use of study participants but only retrieve data on the drugs they acquire on their own initiative, the outcome(s) of interest and potential confounding factors.

Cohort versus *case–control design.*

Although the case–control and cohort designs at first glance appear to be exact opposites, the underlying concepts of the two designs are similar, and the case–control study is best understood as an efficient sampling of the experience underlying a cohort study [24]. Specifically, the controls can be viewed as a sample of the exposure distribution in the source population that has not (yet) experienced the outcome of interest. In a well-designed case–control study, the estimated odds ratio provides unbiased estimates of the incidence rate ratio that would have emerged from a cohort study in the same source population [25]. If the case–control study is carefully nested within the source population, incidence rates, incidence rate differences and attributable proportions can also be estimated. The cohort and case–control designs are equally vulnerable to the influence of unknown or imprecisely measured confounding. While we aim to highlight the differences in the applicability of the two designs below, we believe both designs have merit in studies of drug–cancer associations.

There are circumstances where a cohort approach would be most efficient. One such scenario is if the exposure is rare and the cancer outcome is not rare among users (e.g., risk of skin cancer among immunocompromised patients [26]). Also, if outcomes other than cancer are of interest in the same study, establishing a data set structured for cohort analyses will be more efficient. Further, the cohort study is generally easier to communicate to non-epidemiological professionals and laymen readers than the case–control study. This not only stems from the seemingly more straight-forward 'prospective' way of assessing exposure and outcomes, but also from the cohort study's ability to display baseline characteristics that can even be balanced by the use of propensity score methods [27]. A similar description of characteristics among study subjects in a case–control study (i.e., cases *versus* controls) will, as it should, show a higher prevalence of established risk factors among the cases, which may be erroneously perceived as a problematic imbalance in the study [28]. Lastly, the cohort design more readily facilitates the implementation of a new-user design (see below) and the use of active comparators, which can be of great value in handling of confounding within some therapeutic areas.

In some situations, it would be more convenient to choose a case–control rather than a cohort approach. Firstly, cancer is usually a rare outcome. All things equal, this renders a case–control study markedly more efficient in terms of required size of the analytic dataset and computations than a cohort study. It should be emphasized, though, that when using already collected data (e.g., claims data), this gain in efficiency is purely computational, not statistical, and given current computing power often moot, even in very large databases. Secondly, a cumulative dose–response association is important for establishing plausibility of a causal effect. In a cohort design, this would imply that an exposed person should have his/her exposure level re-classified each time he/she crosses the boundary of an exposure level, which may be computationally demanding, if not difficult. In the case–control design, the exposure level of cases and controls could be computed once, that is, on their index date, albeit under the assumption of no time-varying confounding. Lastly, the establishment of one dataset that enables investigation of various drug exposures in relation to one (or more) cancer types is more easily performed using a case–control approach, which is valuable in settings where extraction of raw data is taxing in terms of time and expenditure.

New-user versus *prevalent eser designs.*

When assessing drug effects, including on cancer outcomes, drug initiation is the most principled starting point [29,30]. Studying prevalent rather than new (incident) users of drugs potentially violates several principles of causal inference, including the absence of an underlying hypothetical intervention. As such, using a 'new-user' design, that is, restricting the study population to new users of the drug under study, should be considered [30]. The new-user design also reduces misclassification of drug use due to left truncation of the time period used for exposure ascertainment in prevalent users. Such misclassification is of particular importance when a carcinogenic effect of limited exposure needs to be assessed. Conversely, when only distant or very long-term drug use is anticipated to influence cancer risk, restriction to new users may hinder a meaningful evaluation [7].

Although the new-user design is most readily implemented in cohort studies, it may also be applied in case–control studies, by nesting these within a new-user cohorts [31]. Overall, the potential value of employing a new-user study design, either as the main analysis or as a sensitivity analysis, should always be considered when conducting drug–cancer association studies [32].

Other study designs.

The advantage of clinical trials over observational studies is that randomization effectively addresses confounding, as covariate balance between drug initiators and non-initiators is guaranteed at baseline. Importantly, this balance extends to covariates that are unmeasured or unknown at the time of the study. The downsides of randomized clinical trials are the prohibitive resource requirements, potential ethical and logistic challenges, and with respect to cancer outcomes, the typically relatively small number of outcomes. For obvious ethical reasons, no randomized clinical trials have been launched with the purpose of demonstrating a carcinogenic effect of a drug. There are, however, some examples of trials aimed at establishing cancer-preventive drug effects, for example, of 5α-reductase inhibitors [33,34], aspirin [35,36] and selective COX2 inhibitors [37–39]. In addition, secondary analyses are increasingly being conducted in clinical trials with primary intervention outcomes other than cancer, for example, cardiovascular trials of aspirin use [40–43]. As the original study materials were based on randomization, these secondary analyses preserve some benefits of a clinical trial. Randomization only removes baseline confounding, however, and treatment changes during follow-up should generally not be ignored when assessing cancer outcomes [44–46].

Several observational self-controlled designs have emerged since the 1990s. These designs share the common feature of comparing the occurrence of outcomes between exposed and unexposed follow-up within the same individual [47]. Thereby, confounders that are stable over time are eliminated by design. However, as they all focus on acute and transient effects of drugs, they have little if any relevance for studying drug–cancer associations.

Finally, ecological, or macro, designs should be mentioned. Here, the unit of analysis is a population rather than an individual. An ecological design would for example compare the use of a drug and a specific cancer incidence rate in one country with another country (or region). Although often surprisingly persuasive for lay readers, epidemiologists generally consider ecological designs as weak. It requires a high population attributable proportion for an ecological design to be effective, that is, a substantial proportion of the cancer occurrence in the population should be attributable to use of the specific drug. To our knowledge, there are only few good examples of this approach within drug–cancer epidemiology [48–50].

Considerations Regarding Assessment of Drug Exposure

For most suspected drug–cancer associations, the relevant exposure window during carcinogenesis is unknown. If the drug exposure influences carcinogenesis (early stage), then the risk period would only start long after initial drug exposure. If the drug has influence on later stages in cancer development, the risk period would start earlier. The period between the first occurrence of a cluster of cancer cells to a stage detectable by screening measures or clinical symptoms is typically long, for some cancer types up to 20–30 years [16]. Hence, it is almost inconceivable that manifest clinical cancer can be an immediate effect of drug exposure due to the long period of development for most cancer types. There are a few examples of relatively rapid cancer development induced by exposure to drug use, for example, the association between use of systemic immunosuppressants and increased risk of non-Hodgkin's lymphoma and skin cancer in organ transplant recipients [51–53]. An example of very long cancer development – about 20 years – is the association between use of diethylstilbestrol and adenocarcinoma of the vagina and cervix [54,55].

Variability in latency of carcinogenic drug effects is often described in terms of the drug being either a cancer 'initiator' or 'promotor'. Conceptually, a cancer initiator is one that is a contributory cause of the first clone of malignant cells, whereas a promotor does not possess this property but acts a contributory factor in accelerating cancer growth of an already existing neoplastic lesion [17,56]. As such, cancer initiators are thought as having a longer latency than promotors. However, in most drug–cancer studies, the mechanisms underlying a potential carcinogenic or anti-neoplastic effect are unknown, at least in terms of what stage of carcinogenesis is affected. Consequently, a conceptual framework of potential promotors and initiators typically provides little guidance when designing a specific study. Nevertheless, as drugs may influence later stages of carcinogenesis, studies monitoring drug–cancer associations relatively shortly after initiation of drug therapy also hold scientific value.

Lag-time.

The typically long period of cancer development and latency of any carcinogenic or anti-neoplastic drug effects supports implementation of a 'lag-time period'. In practical terms, this

means that cancer outcomes diagnosed shortly after drug initiation should not be regarded as occurring during 'exposed time', as these outcomes cannot meaningfully be ascribed to the drug exposure. Correspondingly, a certain period after drug discontinuation should be considered as time at risk. As the lag period after drug initiation covers both induction and latency of the specific cancer, and the lag period after drug discontinuation primarily covers the latent period (plus any potential carry-over effect), a longer lag period after initiation than after discontinuation should be considered.

A second reason for lagging drug effects is the possibility of protopathic bias (reverse causation). Consider a middle-aged man who consults his general practitioner (GP) because of obstructive urinary symptoms. Initially, the GP would be likely to interpret the underlying condition as benign prostate hyperplasia and prescribe a drug relieving the obstructive urologic symptoms. However, additional diagnostic work-up may reveal a prostate cancer as the underlying condition. As the use of symptomatic drugs preceded the cancer diagnosis, an apparent association between these drugs and prostate cancer would emerge. In most cases, however, such reverse causation has a time frame of <6 months [57].

The typically long induction and latent periods, the possibility of reverse causation and the likely higher probability to detect prevalent pre-clinical cancer due to medicalization around treatment initiation are the main reasons for employing a lag-time in drug–cancer analyses. Reverse causation justifies a lag-time in the order of minimum a few months [57]. The optimal time frame for addressing the long induction and latent periods is unknown and depends on the specific drug–cancer association being studied. However, for most drug–cancer associations, it is likely considerably longer than the period necessary to address reverse causation. Acknowledging the uncertainty surrounding the mechanisms of single drug–cancer associations, and thus the optimal (biologically relevant) lag-time, studies should generally evaluate various lag periods in sensitivity analyses [58].

Dose–response association.
Most known drug–cancer associations are dependent on cumulative amount, that is, stronger associations with higher dosages and longer-term use. This may be expressed either by cumulative duration of drug use, by cumulative quantity of the drug or by the prescribed daily dose (coupled to duration of use). The choice is often driven by the available data, but mechanistic considerations are also relevant. If it is believed that the dose intensity, that is, the daily dose, is less important, it would be most appropriate to perform analyses according to the cumulative duration. If the dose intensity does matter, then the cumulative quantity or if feasible a direct measure of intensity over time would be more appropriate. Regardless of the measure used, dose–response or duration of use analyses should always be carried out.

When interpreting dose–response analyses, special attention should be paid to trivial, low-level exposure. If very low exposure shows an association with cancer, there is a chance that the association is explained by confounding [59,60]. While a dose–response pattern is generally considered supportive of causality, caution is advised as some confounders may act in a graded fashion. As an example, progressively heavier use of bronchodilators will likely correlate to smoking history and thereby show a confounded but dose-dependent association with lung cancer risk. Lastly, another good reason for analysing cumulative dose–response effects is that associations that are explained by reverse causation often have an inverse cumulative dose/duration–response association and thereby can be distinguished from causal effects.

Defining treatment episodes.
Constructing treatment episodes for study subjects from prescription registries may be challenging. If prescription data do not include information on the prescribed dose and duration of a specific treatment, treatment episodes need to be estimated on the basis of purchasing dates and quantities of drug prescribed. In the attempt to construct treatment episodes, investigators encounter temporal gaps and overlaps among prescriptions, and different methods accounting for various prescription patterns may lead to different estimates of drug effects [61,62]. As there is typically long latency between the relevant drug exposure and cancer diagnosis, such considerations of the precise period of drug intake may be less important in cancer pharmacoepidemiology than in studies of acute or semi-acute outcomes. Nevertheless, varying prescription patterns without specific knowledge of dose or duration of treatment introduces an additional source of uncertainty regarding exposure ascertainment, and researchers are, for this and the aforementioned reasons, encouraged to apply different exposure measures in individual studies.

Considerations Regarding Cancer Outcome Ascertainment

Important objectives of cancer classification and registration are to assemble and compare cancer incidence data across populations and countries, and to provide valid cancer diagnoses for specific research purposes in analytical studies. During the last two decades, official cancer registration and establishment of cancer registries have increased steadily worldwide, substantially facilitating epidemiological and other cancer research [63]. Valuable sources of international cancer incidence data include the United States' SEER program [64]; the NORDCAN program covering Scandinavian countries [20]; and the GLOBOCAN program estimating cancer incidence and mortality data on a global scale [65].

Traditionally, the most important diagnostic and clinical modalities used to establish cancer diagnoses are reports from pathology departments (providing histological diagnoses), hospital records and death certificates, although the proportion of cancer cases identified solely by death certification is low today in most countries [63]. Together, clinical and histological diagnoses provide a high level of precision in the diagnosis. However, access to records from radiotherapeutic and oncology departments, medical records, imaging measures and

haematology laboratories can also provide important information, for example, for diagnosis of brain and other tumours of the central nervous system and for haematological malignancies. In general, completeness and validity of cancer registration increase with the number of modalities used to ascertain the diagnoses; and significant omissions may raise concerns that case finding is incomplete. Use of multiple sources of cancer ascertainment demands efficient registration procedures to ensure that all records pertaining to the same case are combined into a single registration [66].

In the absence of data from cancer registries, claims-based algorithms can be developed to identify cancer incidence. These algorithms usually cover a combination of diagnostic codes, procedures and take relative timing of these into account. They need to be validated against a gold standard (usually cancer registry data), ideally in a similar population to the one studied. Examples include the commonly used algorithms published by Setoguchi *et al.* [67] that showed high specificity of a definition using two specific cancer codes within 2 months for a variety of cancers. These algorithms are relatively crude, however, and more work is needed to refine and extend them to specific populations.

Using 'cancer overall' as outcome should be discouraged, as this essentially goes against our understanding of cancer as a heterogeneous disease [56]. It is inconceivable that a pharmaceutical agent should act as a universal carcinogen, as not even strong carcinogens, such as tobacco smoking and radiation, are universally carcinogenic [13]. An analysis of cancer overall is essentially driven by the effect of the drug on the cancers with the highest incidence. Currently, cancer diagnoses are most often classified according to organ site, using ICD diagnostic codes. However, in continuation of aspects of specificity, it is unlikely that a given drug would induce all types of cancers, even within the same organ. Thus, if feasible, differentiation according to histological subtypes should be employed in studies of drug–cancer associations. Consider immunosuppressants that are known to induce skin cancer; currently, there is only firm evidence that the association pertains to non-melanoma skin cancer, not melanomas [14]. In some organs, for example, the prostate or colorectum, one histological type (adenocarcinoma) comprises most cancer cases, and as such restriction to adenocarcinomas makes little difference, although it should still be carried out, if feasible. However, at many cancer sites, for example, oesophagus, lung, breast, ovary, testis and kidney, differentiation into distinct histologic subtypes is important, as an analysis lumping together all histological cancer types would be driven by the effect of the drug on the most prevalent subtype.

Increasing knowledge of the epidemiology, histopathology and today also molecular profiles of cancer diseases will facilitate even more detailed classifications categorizing subtypes of cancer both within and across organ sites. Identification of more 'refined' cancer subtypes may thereby, in the years to come, increase the specificity of risk associations and predictions. While beyond the scope of the present MiniReview, readers interested in the application of molecular biology in cancer classification are referred to reviews specifically addressing this issue [68–70].

Considerations Regarding Confounder Adjustment

Adjustment for confounding in drug–cancer association studies should generally follow the same principles as in any other pharmacoepidemiological study. As one noticeable exception, the assessment of confounder variables should take into account the induction/latency period of cancer outcomes, as discussed above. If, for example, no lag-time is considered in the assessment of baseline or time-varying covariates, early symptoms of a cancer or other effects of the increased medical attention preceding a cancer diagnosis may inadvertently be included in the adjustment model. This is especially important in case–control studies, where confounders are often defined on the index date. Further, it is important to emphasize that while many databases used for drug–cancer studies lack information on important and well-established cancer risk factors, for example, family history, many of these will have no apparent effect on the prescribing of the majority of drugs and thus will confer no confounding effect in drug–cancer association studies.

Active comparator.

Important unmeasured covariates, like for example body mass index (BMI), can sometimes be balanced by study design rather than by statistical control. Take as an example a study on the effects of insulin glargine on cancer outcomes (including colorectal cancer, for which BMI is a risk factor). In this study, BMI would be an obvious confounder if we compared patients initiating insulin glargine with non-initiators, because a high BMI is one of the main reasons to initiate insulin therapy in patients with type 2 diabetes. However, using an active comparator, human NPH insulin, the confounding potential of BMI is mitigated because BMI does not affect the choice of insulin after the indication for adding insulin has been made [71]. This highlights the value of the active comparator new-user cohort study design to reduce the potential for unmeasured confounding. Subtle differences in the indication of comparator drugs might still exist and would need to be controlled for by standard analytical techniques. It is important to point out that with an active comparator, the scientific question is changed from absolute effects (e.g., safety) to effects relative to a clinical alternative. The above example of a study with active comparator does not answer the question whether adding insulin in patients with type 2 diabetes increases the risk of colorectal cancer, but the question whether the effect of insulin glargine on colorectal cancer is different from the effect of NPH insulin. As such, if the active comparator has the same potential to cause cancer, for example, through a shared pharmacological action, a carcinogenic effect might be missed. Lastly, the idea of an active comparator should not be pursued at all costs, but only if a suitable active comparator can be identified, that is, a drug with

comparable indications. If the drug under study is the first-line treatment for mild disease, using an active comparator that is only used as a second-line treatment or for more severe disease might introduce more bias than it removes.

Time-varying confounding.

While confounding at drug initiation can often be limited by study design (and remaining confounding removed during analysis), the time-varying confounding affecting drug persistence and changes in treatment over time is related to (lack of) effectiveness and side effects. Unfortunately, many registries that contain necessary data on drug prescribing do not contain much information (if any) on (immediate) effectiveness and (subtle) side effects. For example, many databases do not include measures of glucose control or other markers of diabetes severity that may affect persistence on antidiabetic drugs and risk of some cancers [72]. Confounding from disease severity can bias a study in either direction. On one hand, more severe disease may be associated with longer duration or higher intensity of treatment. Conversely, severe disease may also be associated with switching and thus shorter duration of first-line treatments [73]. If good data on the drivers of adherence/persistence are available in a database, methods that allow for control for time-varying confounders affected by prior treatment should be used to reduce bias [74].

Future Perspectives

The study of drug–cancer associations is in many respects still in its infancy, and the field will likely evolve considerably over the years to come. Three drivers of this evolution deserve to be highlighted. Firstly, the accumulation of long-term data on drug use and cancer incidence is of particular value in studies of cancer aetiology, keeping in mind the potentially long induction periods of cancer development. For example, the Swedish Prescribed Drug Registry, covering the entire population of Sweden (approximately 10 million inhabitants), was established in 2005 [75], and in the United States, drug coverage (part D) for the Medicare population was introduced in 2006. These databases have thus only recently achieved an age where they begin to contribute meaningfully to the field of drug–cancer studies. Secondly, as the pharmacoepidemiological community accumulates experience in conducting multinational studies, the door is open for the study of rare types of cancers (or drug exposures). Lastly, methodological work, specifically addressing studies of long-term outcomes, is both needed and ongoing. For example, further research on the implementation of advanced methods, such as marginal structural models or structural nested failure time models, is needed to properly adjust for time-varying factors [76,77]. These three drivers, coupled with other emerging opportunities, such as incorporation of genetic data, clearly indicate that we will be able to perform more comprehensive drug–cancer association studies in the future, to secure the safe use of drugs in the population and potentially identify new therapeutic avenues for cancer prevention and treatment.

References

1 Drew DA, Cao Y, Chan AT. Aspirin and colorectal cancer: the promise of precision chemoprevention. Nat Rev Cancer 2016;**16**:173–86.

2 Chubak J, Whitlock EP, Williams SB, Kamineni A, Burda BU, Buist DS *et al.* Aspirin for the prevention of cancer incidence and mortality: systematic evidence reviews for the U.S. Preventive Services Task Force. Ann Intern Med 2016;**164**:814–25.

3 Jensen OM, Knudsen JB, Tomasson H, Sørensen BL. The Copenhagen case–control study of renal pelvis and ureter cancer: role of analgesics. Int J Cancer 1989;**44**:965–8.

4 Dubach UC, Rosner B, Stürmer T. An epidemiologic study of abuse of analgesic drugs. Effects of phenacetin and salicylate on mortality and cardiovascular morbidity (1968 to 1987). N Engl J Med 1991;**324**:155–60.

5 Zaidan M, Stucker F, Stengel B, Vasiliu V, Hummel A, Landais P *et al.* Increased risk of solid renal tumors in lithium-treated patients. Kidney Int 2014;**86**:184–90.

6 Licht RW, Grabenhenrich LB, Nielsen RE, Berghöfer A, International Group for the Study of Lithium (IGSLi). Lithium and renal tumors: a critical comment to the report by Zaidan et al.. Kidney Int 2014;**86**:857.

7 Pottegård A, Hallas J, Jensen BL, Madsen K, Friis S. Long-term lithium use, risk of renal, upper urinary tract cancers. J Am Soc Nephrol 2016;**27**:249–55.

8 Lewis JD, Ferrara A, Peng T, Hedderson M, Bilker WB, Quesenberry CP Jr *et al.* Risk of bladder cancer among diabetic patients treated with pioglitazone: interim report of a longitudinal cohort study. Diabetes Care 2011;**34**:916–22.

9 Lewis JD, Habel LA, Quesenberry CP, Strom BL, Peng T, Hedderson MM *et al.* Pioglitazone use and risk of bladder cancer and other common cancers in persons with diabetes. JAMA 2015;**314**:265–77.

10 Sharfstein JM, Kesselheim AS. The safety of prescription drugs. JAMA 2015;**314**:233–4.

11 Korhonen P, Heintjes EM, Williams R, Hoti F, Christopher S, Majak M *et al.* Pioglitazone use and risk of bladder cancer in patients with type 2 diabetes: retrospective cohort study using datasets from four European countries. BMJ 2016;**354**:i3903.

12 Garry EM, Buse JB, Lund JL, Pate V, Stürmer T. Comparative safety of pioglitazone versus clinically meaningful treatment alternatives concerning the risk of bladder cancer in older US adults with type 2 diabetes. Diabetes Obes Metab 2018;**20**:129–40.

13 Friis S, Kesminiene A, Espina C, Auvinen A, Straif K, Schüz J. European code against cancer 4th edition: medical exposures, including hormone therapy, and cancer. Cancer Epidemiol 2015;**39** (Suppl 1):S107–19.

14 International Agency for Research on Cancer. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. A Review of Human Carcinogens. Volume 100 A: Pharmaceuticals. International Agency for Research on Cancer, Lyon, 2012.

15 IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, Meeting. Some drugs and herbal products. 2016. http://www.ncbi.nlm.nih.gov/books/NBK350406/ (last accessed on 29 June 2016).

16 Umar A, Dunn BK, Greenwald P. Future directions in cancer prevention. Nat Rev Cancer 2012;**12**:835–48.

17 Rothman KJ. Induction and latent periods. Am J Epidemiol 1981;**114**:253–9.

18 Wettermark B, Zoëga H, Furu K, Korhonen M, Hallas J, Nørgaard M *et al.* The Nordic prescription databases as a resource for pharmacoepidemiological research–a literature review. Pharmacoepidemiol Drug Saf 2013;**22**:691–9.

19 Pottegård A, Schmidt SAJ, Wallach-Kildemoes H, Sørensen HT, Hallas J, Schmidt M. Data resource profile: the Danish national prescription registry. Int J Epidemiol 2017;**46**:798–798f.

20 Engholm G, Ferlay J, Christensen N, Bray F, Gjerstorff ML, Klint A *et al.* NORDCAN–a Nordic tool for cancer information, planning, quality control and research. Acta Oncol 2010;**49**:725–36.
21 WHO Collaborating Centre for Drug Statistics Methodology. Guidelines for ATC Classification and DDD Assignment 2016. WHO Collaborating Centre for Drug Statistics Methodology, Oslo, 2016.
22 Hong J-L, Jonsson Funk M, Buse JB, Henderson LM, Lund JL, Pate V *et al.* Comparative effect of initiating metformin versus sulfonylureas on breast cancer risk in older women. Epidemiol Camb Mass 2017;**28**:446–54.
23 Gokhale M, Buse JB, Gray CL, Pate V, Marquis MA, Stürmer T. Dipeptidyl-peptidase-4 inhibitors and pancreatic cancer: a cohort study. Diabetes Obes Metab 2014;**16**:1247–56.
24 Miettinen OS. The 'case–control' study: valid selection of subjects. J Chronic Dis 1985;**38**:543–8.
25 Rothman KJ, Greenland S, Lash TL. Modern Epidemiology, 3rd edn. Wolters Kluwer Health, Lippincott Williams & Wilkins, Philadelphia, PA, 2008.
26 Perez HC, Benavides X, Perez JS, Pabon MA, Tschen J, Maradei-Anaya SJ *et al.* Basic aspects of the pathogenesis and prevention of non-melanoma skin cancer in solid organ transplant recipients: a review. Int J Dermatol 2017;**56**:370–8.
27 Stürmer T, Wyss R, Glynn RJ, Brookhart MA. Propensity scores for confounder adjustment when assessing the effects of medical interventions using nonexperimental study designs. J Intern Med 2014;**275**:570–80.
28 Suissa S. The Quasi-cohort approach in pharmacoepidemiology: upgrading the nested case–control. Epidemiology 2015;**26**: 242–6.
29 Kramer MS, Lane DA, Hutchinson TA. Analgesic use, blood dyscrasias, and case-control pharmacoepidemiology. A critique of the International Agranulocytosis and Aplastic Anemia Study. J Chronic Dis 1987;**40**:1073–85.
30 Lund JL, Richardson DB, Stürmer T. The active comparator, new user study design in pharmacoepidemiology: historical foundations and contemporary application. Curr Epidemiol Rep 2015;**2**:221–8.
31 Pottegård A, Friis S, Andersen M, Hallas J. Use of benzodiazepines or benzodiazepine related drugs and the risk of cancer: a population-based case–control study. Br J Clin Pharmacol 2013;**75**:1356–64.
32 Johnson ES, Bartman BA, Briesacher BA, Fleming NS, Gerhard T, Kornegay CJ *et al.* The incident user design in comparative effectiveness research. Pharmacoepidemiol Drug Saf 2013;**22**:1–6.
33 Thompson IM, Goodman PJ, Tangen CM, Lucia MS, Miller GJ, Ford LG *et al.* The influence of finasteride on the development of prostate cancer. N Engl J Med 2003;**349**:215–24.
34 Andriole GL, Bostwick DG, Brawley OW, Gomella LG, Marberger M, Montorsi F *et al.* Effect of dutasteride on the risk of prostate cancer. N Engl J Med 2010;**362**:1192–202.
35 Baron JA, Cole BF, Sandler RS, Haile RW, Ahnen D, Bresalier R *et al.* A randomized trial of aspirin to prevent colorectal adenomas. N Engl J Med 2003;**348**:891–9.
36 Cook NR, Lee I-M, Gaziano JM, Gordon D, Ridker PM, Manson JE *et al.* Low-dose aspirin in the primary prevention of cancer: the Women's Health Study: a randomized controlled trial. JAMA 2005;**294**:47–55.
37 Bertagnolli MM, Eagle CJ, Zauber AG, Redston M, Solomon SD, Kim K *et al.* Celecoxib for the prevention of sporadic colorectal adenomas. N Engl J Med 2006;**355**:873–84.
38 Arber N, Eagle CJ, Spicak J, Rácz I, Dite P, Hajer J *et al.* Celecoxib for the prevention of colorectal adenomatous polyps. N Engl J Med 2006;**355**:885–95.
39 Baron JA, Sandler RS, Bresalier RS, Quan H, Riddell R, Lanas A *et al.* A randomized trial of rofecoxib for the chemoprevention of colorectal adenomas. Gastroenterology 2006;**131**:1674–82.
40 Rothwell PM, Wilson M, Elwin C-E, Norrving B, Algra A, Warlow CP *et al.* Long-term effect of aspirin on colorectal cancer incidence and mortality: 20-year follow-up of five randomised trials. Lancet 2010;**376**:1741–50.
41 Rothwell PM, Price JF, Fowkes FGR, Zanchetti A, Roncaglioni MC, Tognoni G *et al.* Short-term effects of daily aspirin on cancer incidence, mortality, and non-vascular death: analysis of the time course of risks and benefits in 51 randomised controlled trials. Lancet 2012;**379**:1602–12.
42 Kim S, Baron JA, Mott LA, Burke CA, Church TR, McKeown-Eyssen GE *et al.* Aspirin may be more effective in preventing colorectal adenomas in patients with higher BMI (United States). Cancer Causes Control 2006;**17**:1299–304.
43 Stürmer T, Glynn RJ, Lee IM, Manson JE, Buring JE, Hennekens CH. Aspirin use and colorectal cancer: post-trial follow-up data from the Physicians' Health Study. Ann Intern Med 1998;**128**: 713–20.
44 Hernán MA, Hernández-Díaz S, Robins JM. Randomized trials analyzed as observational studies. Ann Intern Med 2013;**159**:560–2.
45 Murray EJ, Hernán MA. Adherence adjustment in the Coronary Drug Project: A call for better per-protocol effect estimates in randomized trials. Clin Trials Lond Engl 2016;**13**:372–8.
46 Mansournia MA, Higgins JPT, Sterne JAC, Hernán MA. Biases in randomized trials: a conversation between trialists and epidemiologists. Epidemiol Camb Mass 2017;**28**:54–9.
47 Hallas J, Pottegård A. Use of self-controlled designs in pharmacoepidemiology. J Intern Med 2014;**275**:581–9.
48 Linden G, Henderson BE. Genital-tract cancers in adolescents and young adults. N Engl J Med 1972;**286**:760–1.
49 Banks E, Canfell K. Recent declines in breast cancer incidence: mounting evidence that reduced use of menopausal hormones is largely responsible. Breast Cancer Res 2010;**12**:103.
50 Chlebowski RT, Kuller LH, Prentice RL, Stefanick ML, Manson JE, Gass M *et al.* Breast cancer after use of estrogen plus progestin in postmenopausal women. N Engl J Med 2009;**360**:573–87.
51 Euvrard S, Kanitakis J, Claudy A. Skin cancers after organ transplantation. N Engl J Med 2003;**348**:1681–91.
52 Rama I, Grinyó JM. Malignancy after renal transplantation: the role of immunosuppression. Nat Rev Nephrol 2010;**6**:511–9.
53 Engels EA, Pfeiffer RM, Fraumeni JF Jr, Kasiske BL, Israni AK, Snyder JJ *et al.* Spectrum of cancer risk among US solid organ transplant recipients. JAMA 2011;**306**:1891–901.
54 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.
55 Herbst AL, Anderson D. Clear cell adenocarcinoma of the vagina and cervix secondary to intrauterine exposure to diethylstilbestrol. Semin Surg Oncol 1990;**6**:343–6.
56 Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;**144**:646–74.
57 Pottegård A, Hallas J. New use of prescription drugs prior to a cancer diagnosis. Pharmacoepidemiol Drug Saf 2017;**26**:223–7.
58 Walker AM. For drug-induced carcinogenesis, the observations are the hypothesis: invited editorial for the Mini-Symposium on Cancer Pharmacoepidemiology. Ann Epidemiol 2016;**26**:749–50.
59 Pottegård A, Schmidt SAJ, Olesen AB, Achacoso N, Van Den Eeden SK, Hallas J *et al.* Use of sildenafil or other phosphodiesterase inhibitors and risk of melanoma. Br J Cancer 2016; **115**:895–900.
60 Loeb S, Folkvaljon Y, Lambe M, Robinson D, Garmo H, Ingvar C *et al.* Use of phosphodiesterase type 5 inhibitors for erectile dysfunction and risk of malignant melanoma. JAMA 2015;**313**:2449–55.
61 Gardarsdottir H, Egberts TCG, Heerdink ER. The association between patient-reported drug taking and gaps and overlaps in

antidepressant drug dispensing. Ann Pharmacother 2010;**44**:1755–61.
62 Gardarsdottir H, Souverein PC, Egberts TCG, Heerdink ER. Construction of drug treatment episodes from drug-dispensing histories is influenced by the gap length. J Clin Epidemiol 2010;**63**:422–7.
63 Forman D, Bray F, Brewster DH, Gombe Mbalawa C, Kohler B, Piñeros M *et al.* (eds). Cancer Incidence in Five Continents Vol. X. IARC Scientific Publications No.164. International Agency for Research on Cancer, Lyon, 2014.
64 Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. Cancer J Clin 2017;**67**:7–30.
65 Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin 2015;**65**:87–108.
66 Parkin DM, Chen VW, Ferlay J, Galceran J, Storm HH, Whelan SL. Comparability and Quality Control in Cancer Registration. IARC Technical Report No. 19. IARC Press, Lyon, 1994, revised and reprinted 1996, reprinted 1998.
67 Setoguchi S, Solomon DH, Glynn RJ, Cook EF, Levin R, Schneeweiss S. Agreement of diagnosis and its date for hematologic malignancies and solid tumors between medicare claims and cancer registry data. Cancer Causes Control 2007;**18**:561–9.
68 Sherman ME, Howatt W, Blows FM, Pharoah P, Hewitt SM, Garcia-Closas M. Molecular pathology in epidemiologic studies: a primer on key considerations. Cancer Epidemiol Biomarkers Prev 2010;**19**:966–72.
69 Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. Nature 2013;**501**:338–45.
70 Lochhead P, Chan AT, Nishihara R, Fuchs CS, Beck AH, Giovannucci E *et al.* Etiologic field effect: reappraisal of the field effect concept in cancer predisposition and progression. Mod Pathol 2015;**28**:14–29.
71 Stürmer T, Marquis MA, Zhou H, Meigs JB, Lim S, Blonde L *et al.* Cancer incidence among those initiating insulin therapy with glargine versus human NPH insulin. Diabetes Care 2013;**36**:3517–25.
72 Stürmer T, Buring JE, Lee I-M, Gaziano JM, Glynn RJ. Metabolic abnormalities and risk for colorectal cancer in the physicians' health study. Cancer Epidemiol Biomarkers Prev 2006;**15**:2391–7.
73 Hernán MA, Robins JM. Per-protocol analyses of pragmatic trials. N Engl J Med 2017;**377**:1391–8.
74 Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology 2000;**11**:550–60.
75 Wallerstedt SM, Wettermark B, Hoffmann M. The first decade with the Swedish Prescribed Drug Register – a systematic review of the output in the scientific literature. Basic Clin Pharmacol Toxicol 2016;**119**:464–9.
76 Picciotto S, Hernán MA, Page JH, Young JG, Robins JM. Structural nested cumulative failure time models to estimate the effects of interventions. J Am Stat Assoc 2012;**107** (499) https://doi.org/10.1080/01621459.2012.682532
77 Petersen M, Schwab J, Gruber S, Blaser N, Schomaker M, van der Laan M. Targeted maximum likelihood estimation for dynamic and static longitudinal marginal structural working models. J Causal Inference 2014;**2**:147–85.