Published in final edited form as:

Ann Epidemiol. 2016 November ; 26(11): 735–740. doi:10.1016/j.annepidem.2016.04.012.

Challenges in evaluating cancer as a clinical outcome in postapproval studies of drug safety

Simone P. Pinheiro, ScD, MSc[a,*], **Donna R. Rivera, PharmD, MSc**[b], **David J. Graham, MD, MPH**[a], **Andrew N. Freedman, PhD**[b], **Jacqueline M. Major, PhD, MS**[a], **Lynne Penberthy, MD, PhD**[b], **Mark Levenson, PhD**[c], **Marie C. Bradley, PhD, MPH, MPharm**[a,b], **Hui-Lee Wong, PhD, MSc**[a], and **Rita Ouellet-Hellstrom, PhD, MPH**[a]

[a]Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, MD

[b]Clinical & Translational Epidemiology Branch, Epidemiology and Genomics Research, Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville, MD

[c]Office of Biostatistics, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, MD

Abstract

Pharmaceuticals approved in the United States are largely not known human carcinogens. However, cancer signals associated with pharmaceuticals may be hypothesized or arise after product approval. There are many study designs that can be used to evaluate cancer as an outcome in the postapproval setting. Because prospective systematic collection of cancer outcomes from a large number of individuals may be lengthy, expensive, and challenging, leveraging data from large existing databases are an integral approach. Such studies have the capability to evaluate the clinical experience of a large number of individuals, yet there are unique methodological challenges involved in their use to evaluate cancer outcomes. To discuss methodological challenges and potential solutions, the Food and Drug Administration and the National Cancer Institute convened a two-day public meeting in 2014. This commentary summarizes the most salient issues discussed at the meeting.

Keywords

Cancer epidemiology; Pharmacoepidemiology; Drug safety

Introduction

The approval process of pharmaceuticals involves rigorous evaluation of carcinogenic potential. Cancer signals (drug-cancer associations) may be hypothesized based on drug class or pharmacological properties or may arise from preapproval nonclinical and clinical trial data, studies conducted following a drug's approval or in some instances from adverse

[*]Corresponding author. U.S. Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, MD, 20993. Tel.: +1-301-796-4951; fax: +1-301-796-9837. simone.pinheiro@fda.hhs.gov (S.P. Pinheiro).

event reporting. If identified before approval (e.g., animal studies, human clinical trials), depending on the strength of the evidence, the information on the signal may be included in the product label, its approved use potentially restricted, and/or product approval may carry a post-marketing requirement for further evaluation by the drug sponsor. Preapproval studies have important limitations. For example, clinical trials frequently have highly selected study participants due to strict inclusion criteria, limited number of participants, and short duration of follow-up time.

Human postapproval studies, frequently observational, may better reflect real-world-use patterns and capture the clinical experience for a larger and more representative sample of participants with longer follow-up time. These studies may be better suited to evaluate cancer signals arising before and/or after drug approval. However, there are important methodological challenges in the evaluation of cancer as outcomes in the postapproval setting, which require careful attention to ensure high-quality research design and appropriate interpretation of results.

To discuss these methodological challenges and identify opportunities to improve the available resources and methods, the U.S. Food and Drug Administration (FDA) and the National Cancer Institute (NCI) convened a two-day public meeting in September of 2014 [1]. Discussions focused on harnessing available resources, along with challenges and strategies to improve design, conduct, and implementation of studies that evaluate the potential risk of cancer associated with use of nononcological drugs and biological products. Epidemiology methodology was not the focus of this meeting unless directly related to the evaluation of cancer as an unintended effect of regulated products. During the meeting, three common, overlapping themes emerged: (1) the relation between cancer biology and epidemiology; (2) data or population sources; and (3) methodological challenges. Main points of discussions, challenges, and potential solutions arising from meeting discussions are summarized in Table 1.

Theme 1: The relation between cancer biology and epidemiology

Cancer biology encompasses the complex system of molecular signaling pathways and biochemical changes intertwined with regulatory proteins, immunomodulation, and epithelial–mesenchymal transition. Cancer genetics intersects cancer biology to elucidate the genomic and epigenetic mutations that underlay these changes. The fundamental multistage process of carcinogenesis includes initiation, promotion, progression, invasion, and metastasis [2]. Each of these stages corresponds to specific mutational or epigenetic changes that correspond with biological changes that can provide insight into targeted mechanisms and help inform study elements including risk windows [3–6]. Etiologic factors can be broadly categorized as genotoxic (direct DNA damage) or nongenotoxic (alternative mechanisms such as immune suppression, inflammatory response, or endocrine modification) [7,8]. These mechanisms are important when considering the plausibility of any agent to cause or promote cancer. Medicines can impact and possibly contribute to cancer development through various mechanisms at various biological phases. However, the ability to detect this progression is challenging. Drugs acting as initiators may take years to eventually manifest a cancer, whereas their action as promoters in the inherited

susceptibilities or suitable clinical conditions may have a shorter time to detection and diagnosis. Examples of known promoters include hormones such as estrogen and drugs such as diethylstilbestrol. Cancer biology in the context of study design has the ability to inform and generate hypotheses, contribute biologically relevant mechanisms and pathways, and provide insights for key study elements.

Theme 2: Data or population sources

The quality of any study is heavily reliant on the quality and extent of data capture.

Because prospective systematic data collection for cancer outcomes may be long-term, expensive, and challenging, discussions at the meeting focused on opportunities to leverage existing resources. Because existing data sources are generally collected for purposes other than research, they may have inadequate or incomplete information on the clinical experience of patients.

Electronic health care data sources

Existing data sources frequently used to evaluate cancer following drug exposure include administrative claims, electronic medical records, registries, and health maintenance organization (HMO) databases. Administrative claims data are typically collected for insurance reimbursement purposes, and electronic medical records data are collected for routine clinical care provision. The purpose of data collection can affect the availability, quality, and completeness of data to evaluate a specific drug exposure and adverse outcome association. General guidelines for the use of electronic health care data in pharmacoepidemiology are found in the literature [9–11].

Several postapproval observational study examples were discussed at the meeting. These include studies using existing electronic health care data to evaluate the association between exposure to insulin glargine and risk of cancer using these data [12], and those leveraging cancer registry data to identify cases for interview when evaluating the risk of osteosarcoma associated with use of teriparatide [13]. Use of existing databases can provide large sample sizes for timely and cost-effective analyses. Some databases (e.g., specific HMOs) can even link to disease or cancer registries and vital status databases. However, these databases also have important limitations. Even with the vast amounts of available data on a large number of individuals, when restricting to new users of pharmaceutical products, excluding patients with preexisting conditions or imposing other inclusion/exclusion criteria, the loss in the number of study participants available for analyses is impressive, reaching at times 90%, for example, the glargine/cancer study.

Continuity of coverage can be another issue with regard to enrollment and disenrollment, transfers, dual, or even overlapping insurance coverage limiting long-term follow-up of individual patients. The available time to follow patients in most electronic health care databases averages approximately 2 to 5 years, which may not be sufficient to identify cancers with a long latency period. Follow-up in HMO networks is generally longer, but they usually have restrictive formularies which can affect uptake of certain pharmaceuticals. An exception is the U.S. Medicare database that allows long-term clinical follow-up for patients over the age of 65 years or younger if eligible. With the availability of Medicare part D

drug-coverage data beginning in 2006, once a patient enters the U.S. Medicare system, long-term clinical follow-up is possible. The picture vastly differs for younger individuals (<65 years). Although commercial databases capture information on younger individuals, these data sources lack comprehensive long-term information due to high rates of annual turnover. Although no immediate solutions were identified to achieve longer term follow-up of younger patients, focusing on patients at higher risk of cancer, in whom cancers may develop sooner, may be a promising strategy in the evaluation of cancer risk in younger patients.

Cancer registries

Access to cancer registry data is considered the gold standard for identifying and characterizing cases. Cancer registry data may include detailed quality-controlled information on the time of diagnoses, date of diagnosis, including stage, prognostic indicators, histologic grade, demographic information about the patient, and information on the initial course of therapy with limited information on chemotherapy and orally administered treatments. Although cancer registries can provide a reliable source of cancer information, they often do not contain data on noncancer drug therapy or important capture of cancer risk factors such as smoking and BMI before diagnosis or detailed outpatient cancer treatment following diagnosis. Also, many cancer registries are neither nationally representative nor do they capture information on all reportable cancers in the United States.

There are different types of cancer registries: hospital, state, regional, national, and international. In the United States, well-recognized national cancer registries are associated with the NCI's Surveillance, Epidemiology, and End Results (SEER) Program [14]. Better known regional registries are associated with specific health plans (e.g., Kaiser Permanente), and the state cancer registries include all 50 states, the District of Columbia, Puerto Rico, and U.S.-affllliated Pacific Islands (e.g., Guam, American Samoa). The state or central cancer registries have a legal mandate that requires all health care providers to report certain information on a cancer case to the registry, and this reporting is HIPAA exempt. Patient identifiers are maintained to enable long-term follow-up of each patient.

The focus of cancer registries is to collect information on cancer from time of diagnosis to initial treatment and death; therefore, registries lack data on patients' precancer exposures including drug use. To assess the risk of exposures, cancer registry reports could be linked to other linkable databases that capture drug information. Although many registries link to a variety of data sources to support their mission of cancer surveillance, the ability to contact patients for interviews requires obtaining institutional review board approval, and the mechanisms to do so vary according to the state laws governing the central registry. Only state cancer registries together could provide national cancer information. However, because each state is governed separately, there are approximately 50 different requirements for review and approval of each research protocol before any linkage can occur, entailing a time-consuming and inefficient process. This is one of the biggest challenges to their collective use for national research.

The North American Association of Central Cancer Registries (NAACCR) [15], a nongovernmental umbrella organization, represents and assists state cancer registry

stakeholders and facilitates cancer registration and data quality. Presentations and discussions at the meeting suggested that NAACCR could assist in streamlining the approval process by the states and facilitate data linkages. This has already been demonstrated effectively in a pilot of a cohort study linking four different registries (Virtual Pooled Cancer Registry Project) [1].

Although complete and timely data capture in cancer registries involves a time lag averaging 2 years, real-time case ascertainment based on near real-time acquisition of electronic pathology reports (e-path reports) offers an opportunity to capture more recent data. However, the e-path report is limited to information that is available at the time of initial surgery such as histological characterization of prognostic markers, date of diagnosis, and state where diagnosed.

Theme 3: Methodological challenges

The third theme materializing from the meeting involved methodological challenges, including challenges related to identifying cancers, defining the risk period, estimating risk, evaluating dose response, identifying comparison groups, and evaluating rare cancers resulting from rare exposures.

Identification of cancer

When clinical studies or linkages to cancer registries are not possible, cancer identification in large databases relies on computer-directed algorithms. These algorithms often require validation with a gold standard (e.g., medical records). The validity of most algorithms to identify cancer is poor, and investigators are trying to improve identification by developing more complex, inclusive algorithms. An important effort currently underway at the Food and Drug Administration is the development of validated computer algorithms using the SEER–Medicare linked database. The computer algorithms are subsequently applied to the remaining U.S. Medicare database for testing and validation.

Identifying the date of incident cancer is also challenging in databases. When using electronic data, cancer history is identified during a predefined look-back period that is usually too short to identify chronic cancer experiences. Identifying an accurate diagnostic date based on procedure and other clinical diagnostic codes from a history of benign symptomology also presents a challenge because some cancers are insidious and may develop long before the cancer is symptomatic or before the diagnosis is made. Thus, determining whether the drug exposure actually preceded the development of cancer can be problematic [16].

Considerations related to evaluation of multiple cancers

Even if specific cancers are identified in preapproval studies, the evidence for a cancer signal may not be strong. Such signals raise hypotheses, but it remains unclear whether only one or biologically similar cancers may be associated with drug treatment in humans. Therefore, when considering cancer, the question frequently arises on whether investigators should consider all cancers in aggregate, individual cancers separately, or consider only the specific cancer(s) that generated the signal.

Experts at the meeting generally agreed that aggregating all cancers into a single composite outcome should be discouraged because cancers express heterogeneity, each with specific biological mechanisms. The use of a rational, biologically driven approach to evaluating specific cancers is preferred to appropriately understand risk attribution. For example, if a drug influences hormone levels, the drug may potentially be related to a hormonally influenced cancer. Looking at subtypes of cancers is also likely informative, as common subtypes across different cancers may provide information on biological plausibility in terms of mechanisms of action or specificity. When using large, linkable databases, information on cancer types and tumor markers is usually available from some cancer registry records (e.g., KRAS) and can be used for this purpose.

Another related issue is whether, for exploratory analyses, investigations should look beyond the prespecified cancer type(s) and consider other cancers given availability of data. Although no consensus was reached among experts, all agreed on the importance of interpreting the results of exploratory analyses of multiple cancers cautiously and considering the role of chance findings if doing multiple evaluations.

Defining risk window and latency

In this document, the term risk period is used interchangeably with risk window to refer to the period(s) when drug exposure exerts its effect on cancer development. The risk period includes the concepts of induction (time from exposure to disease manifestation) and latency (time from manifestation to disease detection) [17]. Because induction and latency periods cannot accurately be differentiated in observational studies, the term latency is used to refer to the time from exposure to ascertainment of outcome.

Contrary to adverse events which are expected to occur shortly after drug exposure, most cancers are generally thought to develop over some period of time after initial exposure to a causative agent (e.g., smoking and lung cancer). However, the notion of long latency in the development of cancers may not necessarily generalize to novel exposures such as new pharmaceuticals or new dosages. This concept will require a paradigm shift in the evaluation of drug-related cancer signals, some of which arise from imbalances in trials of relatively short duration. Information that generated the cancer signal, along with understanding of cancer biology, may help inform the risk period assumptions necessary in the evaluation of signals associated with use of pharmaceuticals.

In general, the actual risk period relevant for drug-associated cancers is poorly understood, and it is possible that the length of the latency period varies by chemical product, dose, type of cancer, and individual genetic susceptibility [7,18]. Understanding the potential biologically relevant period of risk based on observed latency or lag between the symptom, the diagnosis, and treatment requires a combination of statistical, clinical, and molecular knowledge.

Some analytical approaches have been considered in an attempt to minimize misclassification of the period of risk. One involves evaluating multiple risk windows and lagging a cumulative exposure metric by fixed time intervals. Another involves the use of methods such as maximum likelihood or cubic splines to empirically model and identify the

best fitting risk window [19–24]. Results from the latter are population specific and may not generalize or be comparable to all populations. Importantly, given the nature of most drug-related cancer signals, there is rarely sufficient data to allow for reliable estimations of empirical risk functions, regardless of analytical methodology used.

Considerations in risk estimation

Several issues impact the ability to accurately estimate risk in drug-related cancer studies. As discussed, the biologically relevant risk window is typically uncharacterized, yet assumptions about the risk window will influence the accuracy of risk estimates. Furthermore, cancer risk is not expected to be proportional over time. Thus, average estimates (e.g., hazard ratio averaged across follow-up time) are likely not able to properly capture the risk profile over time. Modeling strategies able to accommodate the time-varying nature of cancer risk are needed. In general, some experts agreed that the use of flexible modeling techniques including splines to estimate the cumulative incidence risk functions using the entire available data is preferred in the evaluation of cancer outcomes, particularly in the setting of unknown risk windows.

In addition to considering the time-varying nature of cancer risk, it is also important to consider that patients may not be exposed to the drug during the entire period of follow-up. In general, ignoring adherence in studies of safety outcomes tends to dilute the drug effect.

Analyses that take into account patient exposure patterns are preferred. Because factors that influence drug adherence may be related to both the subsequent development of cancer and future treatment and are themselves influenced by treatment history, methods that estimate effects of time-varying treatment in the presence of time-varying confounders should be considered.

Evaluating dose response

The dose response relationship can be related to graded guidelines for dealing with levels of toxicity in the case of pharmaceuticals such as digoxin or warfarin in some settings. Although evaluation of a dose response may be important, these evaluations are often limited by the approved drug dosages, which typically reflect the lowest dose that demonstrates efficacy and safety in preclinical studies. Because of the chronic nature of cancer and restrictions sometimes imposed on drug use (e.g., labeled use), there may not be a clear response or dose response.

Considering several dimensions of drug exposure and time may be most informative. These dimensions might include parameters describing the shape of the exposure response, time since exposure, and intensity (e.g., level of exposure accrued during each time point) of exposure.

Identifying the comparison group

When comparing an observed to an expected (i.e., background) cancer risk, the expected cancer risk should be that of an internal or within-study reference group. This is the preferred comparison to minimize differences in study populations across different data

sources and study periods. However, finding an appropriate reference group in drug-cancer studies, even within the same disease indication, is challenging for several reasons. First, there may be specific reasons why patients and prescribers consider initiating one treatment versus another, even for patients with the same underlying condition. For example, there may be important differences in disease activity and patient-specific factors between patients initiating a first-line compared to a second-line or a third-line treatment and between patients initiating treatment with a newly marketed versus an established product. Patients initiating treatment with a second-line biological disease-modifying antirheumatic drug (bDMARDs) are more likely to have active disease than patients initiating treatment with first line nonbiological [(nb) DMARD], such as methotrexate.

Investigators should consider factors influencing cancer risk such as disease duration, comorbidities, and medication adherence when selecting comparison groups.

Evaluating rare exposure and rare cancers

Evaluating cancer outcomes in settings of rare or restricted use of the suspected drug exposure is difficult with existing resources. One example discussed at the meeting identified the challenges of evaluating the risk of osteosarcoma, a rare bone tumor in humans, in association with teriparatide use, a second-line injectable product used for the treatment of postmenopausal women with uncontrolled osteoporosis [1], also a rare exposure. Existing data sources cannot identify specific rare cancers without linking to cancer registry information, the challenges of which were already described. Over the years, several study designs (case series and a registry) were implemented to circumvent these difficulties and evaluate the risk. These studies have yielded scant data given the challenges when attempting to identify and interview cancer patients through multilayered requirements imposed by state cancer registries. Experts agreed that studies evaluating rare cancers subsequent to rare exposures may only be able to ascertain high levels of risk given the resources available to date.

Future directions

The most immediate next step identified at the meeting likely to facilitate the identification and characterization of cancers was proposed by NCI/SEER and NCI/NAACCR, leveraging work they have ongoing.

NCI is supporting NAACCR in the development of the virtual pooled registry project [1]. This is a feasibility project, using a distributed network format that attempts to create a nationally centered matching capacity between population-based state cancer registry records and cancer investigators, to facilitate the identification, location, and duplication of cancer records. The current demonstration project includes only seven states with the hope of expanding to all states. A recent survey of central registries that represents over 65% of the U.S. population reported their willingness to participate. Additional pilot studies are underway to support the development of standardized processes for data submission and matching on a larger scale.

Conclusions

Based on the limitations of currently available tools and resources, the ability to provide solutions will arise from collaborations between multidisciplinary scientists. It may not be important, or possible, to conduct studies to evaluate the effects of all cumulative drug exposure patterns on cancer risk. Instead, it may be more relevant, and feasible, to conduct studies that evaluate the effects of a drug on cancer risk under certain specific scenarios, including those most relevant to logical populations, comorbidities, and/or labeled use. Observational studies can provide complementary evidence of medication use and augment gaps in settings where clinical trials are less feasible. Harnessing the power of robust data is vital to producing high-quality pharmacoepidemiology studies evaluating nononcology treatments and the development of cancer. Improving data quality and database resources, advancing methodological options, and evaluating biological mechanisms in the study design are essential efforts in improving ability to evaluate the drug-related cancer risk in the postapproval space.

Acknowledgments

The authors are indebted to the members of the panel of this meeting, who provided invaluable expertise and contributions to meeting discussions. Speakers, panelists, and moderators included, in alphabetical order by last name, Drs. Elizabeth Andrews, Laurent Azoulay, Marie Bradley, Howard Fingert, Andrew Freedman, David J. Graham, Laurel Habel, Miguel Hernan, Solomon Iyasu, John Leighton, Stephanie Leuenroth-Quinn, Mark Levenson, Jay Lubin, Jill MacKinnon, Jacqueline Major, Katherine McGlynn, Rita Ouellet-Hellstrom, Simone P. Pinheiro, Marsha E. Reichman, David Richardson, Tracy Salaam, Judy Staffa, Samy Suissa, Alexander Walker, Cunlin Wang, and Hui-Lee Wong.

References

1. US Food and Drug Administration. [accessed 20.12.2014] Methodological considerations in evaluation of cancer as an adverse outcome associated with use of non-biological drugs and biological products in the post-approval setting; public meeting. Presentations, transcripts, agenda background Mater available. Available at: http://www.fda.gov/Drugs/NewsEvents/ucm401452.htm
2. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011; 144(5):646–74. [PubMed: 21376230]
3. Potter JD. Colorectal cancer: molecules and populations. J Natl Cancer Inst. 1999; 91(11):916–32. [PubMed: 10359544]
4. Stratton MR. Journeys into the genome of cancer cells. EMBO Mol Med. 2013; 5(2):169–72. [PubMed: 23339072]
5. Chia WK, Ali R, Toh HC. Aspirin as adjuvant therapy for colorectal cancer—reinterpreting paradigms. Nat Rev Clin Oncol. 2012; 9(10):561–70. [PubMed: 22910681]
6. Ember I, Nemeth A, Varga C, Perjesi P, Arany I, Feher K, et al. Molecular epidemiologic markers: a new concept in the preventive medicine with special attention to the prevention of cancer. Cent Eur J Occup Environ Med. 2005; 11(1):3–15.
7. Hernandez LG, van SH, Luijten M, van BJ. Mechanisms of non-genotoxic carcinogens and importance of a weight of evidence approach. Mutat Res. 2009; 682(2–3):94–109. [PubMed: 19631282]
8. Hoeijmakers JH. Genome maintenance mechanisms for preventing cancer. Nature. 2001; 411(6835): 366–74. [PubMed: 11357144]
9. 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):1–10.

10. US Food and Drug Administration. [accessed 20.12.2014] Best practices for conducting and reporting pharmacoepidemiology safety studies using electronic healthcare data: guidance for industry and FDA staff. 2013. Available at: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM243537.pdf
11. Strom, BL., Kimmel, SE., Hennessy, S. Pharmacoepidemiology. West Sussex. UK: John Wiley & Sons Ltd; 2012.
12. Habel, L. [accessed 9.11.2014] Use of Administrative databases to evaluate drug-related cancer outcomes: the insulin glargine example. Presentation. available at: http://www.fda.gov/Drugs/NewsEvents/ucm401452.htm
13. Andrews EB, Gilsenan AW, Midkiff K, Sherrill B, Wu Y, Mann BH, et al. The US postmarketing surveillance study of adult osteosarcoma and teriparatide: study design and findings from the first 7 years. J Bone Miner Res. 2012; 27(12):2429–37. [PubMed: 22991313]
14. National Cancer Institute. [accessed 20.12.2014] Surveillance, Epidemiology and End Results Program. 2015. Available at: http://seer.cancer.gov
15. North American Association of Central Cancer Registries (NAACCR). [accessed 20.12.2014] 2015. Available at: http://www.naaccr.org
16. Walker AM. Identification of esophageal cancer in the General Practice Research Database. Pharmacoepidemiol Drug Saf. 2011; 20(11):1159–67. [PubMed: 21928362]
17. Rothman, KJ., Greenland, S., Lash, TL. Modern Epidemiology. Philadelphia, PA: Lippincott Williams & Wilkins; 2012.
18. Rothman KJ. Induction and latent periods. Am J Epidemiol. 1981; 114(2):253–9. [PubMed: 7304560]
19. Silverman DT, Samanic CM, Lubin JH, Blair AE, Stewart PA, Vermeulen R, et al. The Diesel Exhaust in Miners study: a nested case-control study of lung cancer and diesel exhaust. J Natl Cancer Inst. 2012; 104(11):855–68. [PubMed: 22393209]
20. Richardson DB, Terschuren C, Pohlabeln H, Jockel KH, Hoffmann W. Temporal patterns of association between cigarette smoking and leukemia risk. Cancer Causes Control. 2008; 19(1):43–50. [PubMed: 17906957]
21. Langholz B, Thomas D, Xiang A, Stram D. Latency analysis in epidemiologic studies of occupational exposures: application to the Colorado Plateau uranium miners cohort. Am J Ind Med. 1999; 35(3):246–56. [PubMed: 9987557]
22. Checkoway H, Pearce N, Hickey JL, Dement JM. Latency analysis in occupational epidemiology. Arch Environ Health. 1990; 45(2):95–100. [PubMed: 2334237]
23. Rappazzo KM, Daniels JL, Messer LC, Poole C, Lobdell DT. Exposure to fine particulate matter during pregnancy and risk of preterm birth among women in New Jersey, Ohio, and Pennsylvania, 2000–2005. Environ Health Perspect. 2014; 122(9):992–7. [PubMed: 24879653]
24. Checkoway H, Lundin JI, Costello S, Ray R, Li W, Eisen EA, et al. Possible procarcinogenic association of endotoxin on lung cancer among Shanghai women textile workers. Br J Cancer. 2014; 111(3):603–7. [PubMed: 24921918]

Table 1

Important challenges and potential solutions according to specific topics of interest

Issues of interest	Relevant discussions/potential solutions
Theme 1: Cancer biology/pharmacoepidemiology relationship	
Cancer biology	• Generate hypotheses—informs consideration of – Mechanism(s) of action and biological plausibility of a pharmaceutical agent in the causation of cancer – Biological pathways and signaling cascades. Examples include medicines affecting hormone levels and potential relationship with hormonally inducted or promoted cancer, immunomodulatory agents, and cancers with a strong link to the immune system • Study design elements—informs – Specificity and plausibility of exposure to clinical presentation of cancer – Biologically relevant cancer types, cancer stage, histology, and molecular signature (e.g., EGFR in lung cancer) • Study analysis—informs – Coherence and temporality of study results – Sensitivity analyses to evaluate latency and exposure definitions • Identification of confounders, for example, comorbid conditions that affect the same mechanisms/relevant pathways
Theme 2: Data or population sources	
Prospective cohorts (new and ongoing)	• Advantages – Ability to collect baseline information before outcome occurrence – May collect biological specimens – May link to other databases including cancer registries and state vital status databases • Limitations – Expensive – Long-term follow-up of patients is logistically challenging – Subject to complex HIPAA regulations • Solution – Harness use of ongoing cohort studies when possible and appropriate
Use of existing data or population sources: exposure &cancer outcome • Electronic claims databases and HMOs • Electronic medical records	A promising approach with unique methodological challenges • Advantages – Large number of study participants, no recall bias, some may link to other databases • Limitations – Data collected for other purposes so not all needed information is available (e.g., BMI, smoking, family history) – High turnover rate resulting in short duration of follow-up (some with long-term ability to follow-up; e.g., Medicare for patients 65+ and younger if disabled)

Issues of interest	Relevant discussions/potential solutions
	– Limited options to link to national data: for example, state cancer registries
Cancer identification—cancer registries	• Strengths
	– Provides detailed cancer diagnosis data, date of diagnosis, stage, prognostic indicators, histologic grade, demographic information
	– Some registries provide cancer types, tumor markers, and histology
	• Limitations
	– Lack data on noncancer drug therapy, outpatient cancer treatment, behavioral risk factors
	– Some cancer registries are neither nationally representative nor have information on all reportable cancers in the United States
	– Long data lag (averaging 2 years)
	• Solutions
	– Facilitate ability to link data to state cancer registries across the country
	– Explore use of e-path cancer reports to obtain timely cancer data
Theme 3: Methodological challenges	
Cancer identification—computer algorithms	When linkage to cancer registries is not possible, algorithms have been considered to identify occurrence of cancer outcomes
	• Strength
	– Developed using existing data
	• Limitations
	– Low PPV for many cancers
	– Not applicable to all data sources
	– Requires validation against gold standard
	– Difficult to identify incident cases because nonspecific symptoms may occur long before diagnoses
	• Solutions
	– Efforts exploring development of algorithms using SEER–Medicare linkage (ongoing at the FDA)
Selecting the cancer outcome	• Points of discussion
• Aggregated	– Aggregating all cancers is discouraged
• Targeted specific cancers or only	– Use of biologically driven approach is preferred
• Signaling cancer?	– Exploratory analyses should be clearly specified and interpreted
	– Role of chance needs to be carefully considered
Risk window and latency	• Problem
	– Biologically relevant period of risk is unknown
	• Solutions
	– Evaluate multiple windows; for example, lagging analyses
	– Using statistical methods to model and identify the best fitting window; for example, splines
	– Estimate cumulative risk function using entire data

Issues of interest	Relevant discussions/potential solutions	
	–	Consideration: available data may not be sufficient to allow for reliable empirical estimations of risk functions
Time-varying nature of cancer risk	•	Problem
	–	Risk of cancer can vary over time so use of person-years and hazards ratio may not be ideal
	–	Labeling and restrictions limit use and ability to evaluate response/dose-response clearly
	•	Solutions
	–	Estimate the cumulative incidence risk function using the entire data
Comparison group	•	Problem
	–	The underlying disease may influence cancer risk; for example, disease duration
	–	There may be specific reasons why patients receive certain treatments, even if for the same indication; for example, first-line versus second-line or third-line treatments
	•	Solution
	–	Consider disease activity/duration and type of treatment when selecting comparator group
Rare cancers and rare drug exposure	•	Problem
	–	What study design(s) can be considered to evaluate rare cancers associated with rare exposures?
	•	Solutions: No good solution proposed but the following were suggested
	–	Studies evaluating the association between rare exposures and rare cancers may only be able to ascertain high levels of risk
	–	Facilitate the ability to link exposure databases to cancer registries (e.g., Virtual Pooled Cancer Registry Project)