Published in final edited form as:

Ann Epidemiol. 2016 November ; 26(11): 741–745. doi:10.1016/j.annepidem.2016.10.001.

Connections between pharmacoepidemiology and cancer biology: designing biologically relevant studies of cancer outcomes

Donna R. Rivera, PharmD., MSc.¹, Katherine A. McGlynn, PhD, MPH.², and Andrew N. Freedman, PhD.¹

¹Clinical & Translational Epidemiology Branch, Epidemiology and Genomics Research, Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville, MD

²Metabolic Epidemiology Branch, Division of Cancer Epidemiology & Genetics, National Cancer Institute, Rockville, MD

Keywords

pharmacoepidemiology; cancer epidemiology; cancer biology; study design

The increasing utilization of prescription drugs among adults in the United States demonstrates the need to remain vigilant in evaluating the relationship between non-oncologic medications and the development of cancer.¹ Pharmacoepidemiology is the study of the utilization, safety, and effects of medications in a population. Well-designed pharmacoepidemiology studies can examine the relationship between medication use and the risk of either initiating or promoting cancer. Within the context of pharmacoepidemiology studies commonly used to evaluate these associations, developing a comprehensive understanding of biological mechanisms and disease pathways can provide complementary knowledge toward hypothesis development, study design, methodology, and analysis to evaluate cancer as an adverse outcome. There are several types of available data sources for observational research including, but not limited to administrative claims data, electronic health records, survey data, and disease registry data. A significant limitation of some of these data sources is that variables may not be collected for research and are typically ascertained for treatment or billing purposes. Commonly collected data elements vary by data source and can include demographic variables, medical diagnoses, medical procedures, and treatment-related details coded to standardized classifications or nomenclatures (ICD, Read, SNOMED-CT, HCPCS, CPT, NDC, RxNorm, MEDRA, LOINC, DRG). The following concepts aim to guide purposeful, connections between biologic foundations and pharmacoepidemiology studies.

Corresponding Author: Donna R. Rivera, 9609 Medical Center Drive, Rockville, MD 20850, Phone: (240)276-6809, donna.rivera@nih.gov.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Cancer Biology Overview

Carcinogenesis is the multi-stage biological process of cellular transformation through initiation, promotion, progression, invasion, and metastasis. The transitions from one stage to the next are believed to correspond to mutational (germline and somatic) and epigenetic changes associated with cellular regulatory processes including tumor suppressor genes, oncogenes, and stability genes, as well as host factors like immune response. Table 1 provides an overview of selected terms to reference for a more robust understanding of cancer biology processes. Even a basic understanding of the complex pathways (such as TNF-NF κ B or MAPK) or genomic alterations, may stimulate the development of novel biologically-driven research questions and new hypotheses that can be examined in pharmacoepidemiology studies. The process of carcinogenesis is linked to a variety of etiologic factors (genotoxic and non-genotoxic carcinogens—defined in Table 1). Medications along with numerous other endogenous and exogenous sources comprise these factors and specific examples are shown in Figure 1. Careful examination of the biological influences in both the design and analysis phases is beneficial when assessing the relationship between medications and cancer risk.

Broad Considerations for Characterizing Exposure

Understanding cancer biology may be useful in determining the risk window for exposure and the biological plausibility of an exposure association. Ascertaining the biologically relevant study period, which includes an exposure risk window, is both pivotal and challenging. The exposure risk window indicates the period of susceptibility to the effect of the exposure. It may vary on biological factors, for example immune system development may be a time of particular vulnerability to carcinogenic effects. The effect of the exposure also varies based on a myriad of factors such as, drug type, duration of exposure, intensity of exposure, and other measured and unmeasured covariates as well as varying by cancer type.

Biological plausibility suggests that there could be a potential mechanism for carcinogenesis. For example, with an immunomodulatory agent (e.g. anti-TNF drugs), it would be logical to look at cancers which are specifically related to immune regulation such as T cell and B cell lymphomas.² Similarly, if a drug alters estrogen or progesterin levels, one strategy could be to investigate whether the drug is potentially related to a hormonally induced or promoted cancer such as the case with hormone replacement therapy and the development of breast cancer.³ The Woman's Health Initiative (WHI) provides a clear example. The WHI study showed hormone replacement therapy (HRT) may increase the risk of estrogen receptor positive breast cancer. Subsequent to this finding, a significant drop in HRT prescriptions was observed in 2002. A decrease in breast cancer incidence followed from 2002 to 2005, but only among estrogen receptor-positive tumors.⁴ The timing of HRT exposure (post-menopausal) and subsequent results suggest that the outcome may be related to cancer promotion rather than initiation, as breast cancer takes time to clinically manifest. Thus, determining the relevant exposure risk window for studies of hormonal agents with similar estrogenic pathways should consider the potential influence of biology on the relevant risk window of exposure while also ensuring selection of the biologically appropriate population of interest (example: postmenopausal women).

Even taking into account these biological concepts and developing a strong biologically driven study design, supplementary methods may uncover additional insights. Uncertainties in calculating a biologically relevant exposure risk window favor the application of sensitivity analysis to examine the effects of varying drug exposure periods on the risk of outcomes. Modeling these alternative windows to define exposure attempts to demarcate important periods of exposure contribution and account for biological factors fundamental to the association. The results of testing these assumptions may allow for assessment of observed and unmeasured confounding.⁵ In particular, selecting the appropriate exposure definition can minimize measurement biases such as misclassification, which is essential for strong internal validity. The use of sensitivity analysis may also be useful when scientific questions have limited previous research or unknown biological mechanisms, relying on related clinical or methodological knowledge for study design.

Quantifying Drug Exposure

Alongside identification and definition of exposure lies the principle of appropriate measurement of an exposure. The dose-response relationship is one pharmacologically relevant basis for beginning to assess various adverse events associated with medication use. Exposure (dose) can be evaluated in a myriad of ways such as: medication strength, cumulative therapeutic duration, defined daily dose (DDD), area under the concentration time curve (AUC), calculation of adherence (MPR or PDC), or calculation of relative dose intensity (RDI, commonly used for chemotherapy). Dose measures should be selected based on appropriateness for the study question and can be calculated using various algorithms when using administrative claims or other health data. Because cancer may often be observed only after considerable time has passed since the relevant exposure, dosing and adherence measures are relevant.

There are several variations of dose-related exposure to consider. Acute high dose drug exposure and chronic low dose drug exposure have differential dose intensity effects on organ systems during metabolism (exp. renal and hepatic) which may induce varying outcomes (exp. targeted damage) for evaluation. Even the timing of exposure can play a role, as medication use during phases of pediatric cellular development and growth might pose a differential risk of adverse outcome development. Immune system impairment during drug exposure might precipitate the development of an outcome such as cancer. Incorporating drug specific properties (mechanism of action, pharmacodynamics and pharmacokinetics, half-life, toxicology, or duration of use) and drug–disease interactions, affords an enhanced understanding of drug exposure and its interaction with the measured outcome. These pharmacological concepts guide drug discovery and applying this knowledge to pharmacoepidemiology studies may yield innovative connections for evaluating or preventing outcomes.

Consideration of Latency

Along with defining the relevant exposure risk window, understanding the latency period is integral to measuring cancer as an outcome. Latency is defined for this purpose as the time period between the initiation of the biologically relevant exposure and the development of

cancer, or the ability to adequately recognize or detect the cancer in the context of database studies. The induction period (the pathogenesis of the disease cascade from biological onset to initiation) is beyond the reach of this discussion.

Cancer biology may offer insight on variable selection and study design when accounting for latency periods in pharmacoepidemiology studies. For example, Nadler and Zurbenko modeled minimum latency times ranging from less than six months for lymphoproliferative cancers to four years for solid tumors and eleven years for mesothelioma.⁶ Applying these latency periods to a study by Vogelstein et al. indicates the number of non-synonymous somatic tumor mutations correlates proportionally with the length of latency period.^{6, 7} For example, Acute Lymphoblastic Leukemia has less than 25 somatic mutations and a reported minimum latency of less than 6 months, whereas a solid tumor such as lung cancer could have approximately 200 somatic mutations and a minimum latency of around 11 years.⁷

The availability of long-term data to explore latency is fundamental to assess the temporal relationship between the exposure and the appearance of an outcome; longitudinal data is particularly advantageous for use in chronic disease outcomes with a long latency period whereas a shorter latency period may require less follow-up to evaluate more proximate exposures. For example, in a prospective cohort study of lung cancer selecting a one-year study period from exposure to outcome may yield few cancer cases, and those that developed may not have any association with the exposure (e.g. previously dormant or missed diagnoses). This provides additional support to the concept that conventional pre-clinical research and clinical trials detecting a solid tumor are more easily able to identify medications that act as promoters rather than initiators because the exposure and outcome are more likely to occur during the relatively short follow-up period. Alternatively, clinical trials with a short follow-up period may be better able to detect lymphoproliferative cancers associated with the study medication due to the shorter disease latency. In cancer pharmacoepidemiology, these challenges highlight the essential role of supplementary longitudinal observational research and Phase IV studies for postmarketing surveillance and pharmacovigilance.

Defining the Cancer Outcome

In addition to selecting the relevant exposure risk window and accounting for the latency period, the outcome should be appropriately defined. There are several ways to identify a cancer outcome including linkage to a cancer registry or through the use of diagnostic codes. In the development of an administrative database algorithm to identify esophageal cancer, Walker et al. note the inherent challenges associated with identifying the true clinical onset of disease in the absence of clinical or cancer registry data.⁸ This is due to the fact that many patients eventually diagnosed with esophageal cancer had a history of non-specific symptom diagnosis codes preceding the cancer diagnosis, which may have been directly related to the diagnosis of the cancer. From a data perspective, it is difficult to determine the time period before a diagnosis that symptoms are relevant enough to translate into an outcome definition algorithm. From a clinical perspective, the outcome can be equally difficult to define because of the vague or absent symptomology that can precede the clinical diagnosis of cancer (relevant if complementary clinical data is available). From a biological perspective,

it is important to consider that medication acting as an initiator may take years to eventually cause the outcome, whereas its action as a promoter may lead to a shorter time to diagnosis. This reflects a sizeable concern in the pharmacoepidemiologic analysis of most cancers because of the uncertainty surrounding biologically relevant study components including exposure and outcome windows, as drug exposure, presentation of initial symptoms, and diagnosis date might be years apart. Further methods development around cancer case identification algorithms and subsequent algorithm validation would improve study quality, signal assessment, and exposure risk window accuracy.⁹

Another decision in constructing the outcome algorithm is whether to measure cancer as a single (specific cancer) or composite outcome (any cancer). Creating a composite outcome in this instance can make risk attribution difficult and increase the risk of clinically uninterpretable findings. Similarly, the use of all or any cancer as a composite endpoint implies that all outcomes comprising the composite endpoint have similar underlying treatment effects, importance, and frequency.^{10, 11}

Methodological Considerations for Enhancing Analyses

There is wide variation in the availability of supplementary data including genomic, therapeutic, testing, comorbidities, or patient reported health behavior data. Biologically relevant covariates should be included in analyses when available to assess alternative causes, reduce biases, and improve specificity. Limitations of administrative databases suggest that complementary clinical data from electronic medical records (EMR) can provide enhanced biological information consistent with disease pathophysiology (including cancer stage, histology, or type) and genomic expression (including mutations and biomarkers); this integrated data may be useful to detect clinical variation and for more robust analysis to explore potential confounding, effect modification (such as patient subgroup differences), and advanced modeling techniques.

The use of cancer biomarker data offers one example of potential effect modification. Information on biomarkers may be available from genomic testing results or may be deduced via the particular therapy reported. Early sun exposure is a known environmental risk factor in the development of melanoma. The ongoing process of chronic sun exposure and subsequent mutations (NRAS, KIT) can cause melanoma or the presence of a BRAF mutation can push the initiated melanocyte to melanoma in the absence of chronic sun exposure.¹² In a population based study, the effect of an exposure might differ between the BRAF mutated and wild type melanoma (cancer outcome) requiring subgroup examination. Along with biomarkers, cancer subtypes can be an important aspect of biological exploration. For example, breast cancer is a heterogeneous disease comprised of four molecular subtypes (Luminal A, Luminal B, Triple Negative (basal), and HER2) that can show differences in survival.¹³ Effect modification may also be derived from the presence of clinically relevant SNPs. VEGFR1 expression is linked with poorer outcomes in bevacizumab treated patients and presence of the M420del hOCT1 SNP is linked to differential outcomes of imatinib treatment.^{14,15} Greater synergy between pharmacoepidemiology and pharmacogenomics can examine the clinical importance of these variations on outcomes.

Conclusion

Pharmacoepidemiology studies can be enhanced by obtaining insights from cancer biology. Studies should be carefully designed based on an understanding of carcinogenesis, relevant mechanisms of the study medication, and time varying disease factors. The biological contributions highlighted provide thought provoking considerations for design, interpretation, and discussion. Figure 2 summarizes the applications of this piece to three study components where biologically relevant questioning can specifically improve pharmacoepidemiology studies. Hypothesis generation makes use of understanding the overarching association between mechanisms and outcomes, ensuring an appropriate rationale when available. Study design involves the more technical elements of defining biologically relevant exposure risk windows and outcomes as well as estimating the effects of drug exposure. The third component, methods, is the structural backbone of every study which assesses and addresses epidemiology tenets that ensure the association has been examined appropriately to reflect the suggested results. Application of biological mechanisms to pharmacoepidemiology informs ongoing discussions which seek to enrich the quality and design of observational studies while aiming to ensure the safest medication use for all patients.

References

1. Kantor ED, Rehm CD, Haas JS, Chan AT, Giovannucci EL. Trends in Prescription Drug Use Among Adults in the United States From 1999–2012. *JAMA*. 2015; 314(17):1818–1830. [PubMed: 26529160]
2. Shiels MS, Katki HA, Hildesheim A, Pfeiffer RM, Engels EA, Williams M, Kemp TJ, Caporaso NE, Pinto LA, Chaturvedi AK. Circulating Inflammation Markers, Risk of Lung Cancer, and Utility for Risk Stratification. *Journal of the National Cancer Institute*. 2015; 107(10):199.
3. Ravdin PM, Cronin KA, Howlader N, Berg CD, Chlebowski RT, Feuer EJ, Edwards BK, Berry DA. The decrease in breast-cancer incidence in 2003 in the United States. *New England Journal of Medicine*. 2007; 356(16):1670–1674. [PubMed: 17442911]
4. Cronin KA, Ravdin PM, Edwards BK. Sustained lower rates of breast cancer in the United States. *Breast Cancer Research and Treatment*. 2009; (117):223–224.
5. Delaney, JAC., Seeger, JD. Sensitivity Analysis. In: Velentgas, P, Dreyer, NA, Nourjah, P., et al., editors. *Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide*. Vol. Chapter 11. Rockville (MD): Agency for Healthcare Research and Quality (US); 2013. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK126178/>
6. Nadler DL, Zurbenko IG. Developing a weibull model extension to estimate cancer latency. *ISRN Epidemiology*. 2013 2012.
7. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer Genome Landscapes. *Science*. 2013; 339(6127):1546–1558. [PubMed: 23539594]
8. Walker AM. Identification of esophageal cancer in the General Practice Research Database. *Pharmacoepidemiology and drug safety*. 2011; 20(11):1159–1167. [PubMed: 21928362]
9. Lanes S, Brown JS, Haynes K, Pollack MF, Walker AM. Identifying health outcomes in healthcare databases. *Pharmacoepidemiology and drug safety*. 2015; 24(10):1009–1016. [PubMed: 26282185]
10. Montori, Victor M., Permyer-Miralda, Gaietà, Ferreira-González, Ignacio, Busse, Jason W., Pacheco-Huergo, Valeria, Bryant, Dianne, Alonso, Jordi, et al. Validity of composite end points in clinical trials. *BMJ*. 2005; 330:594–596. [PubMed: 15761002]
11. Tomlinson G, Detsky AS. Composite End Points in Randomized Trials: There Is No Free Lunch. *JAMA*. 2010; 303(3):267–268. [PubMed: 20085955]

12. Walker, G., Hacker, E. Ultraviolet Light as a Modulator of Melanoma Development. INTECH Open Access Publisher; 2011.
13. Haque R, Ahmed SA, Inzhakova G, Shi J, Avila C, Polikoff J, Bernstein L, Enger SM, Press MF. Impact of breast cancer subtypes and treatment on survival: an analysis spanning two decades. *Cancer Epidemiol Biomarkers Prev.* 2012 Oct; 21(10):1848–55. [PubMed: 22989461]
14. Lambrechts D, Claes B, Delmar P, Reumers J, Mazzone M, Yesilyurt BT, Devlieger R, Verslype C, Tejpar S, Wildiers H, de Haas S. VEGF pathway genetic variants as biomarkers of treatment outcome with bevacizumab: an analysis of data from the AViTA and AVOREN randomised trials. *The Lancet Oncology.* 2012; 13(7):724–733. [PubMed: 22608783]
15. Giannoudis A, Wang L, Jorgensen AL, Xinarianos G, Davies A, Pushpakom S, Liloglou T, Zhang JE, Austin G, Holyoake TL, Foroni L. The hOCT1 SNPs M420del and M408V alter imatinib uptake and M420del modifies clinical outcome in imatinib-treated chronic myeloid leukemia. *Blood.* 2013; 121(4):628–637. [PubMed: 23223357]

Figure 1.
Examples of Potential Etiologic Causes of Cancer

Figure 2.
Recommended questions to consider at key steps of the study

Table 1

Selected Cancer Biology Terms

Terminology	Definition
Carcinogenesis	The complex interplay between genetics, environmental exposure, and external factors which induces conditions that are conducive to the development and progression of cancer. The steps of carcinogenesis Initiation, Promotion, and Progression (Invasion, Metastasis)
Driver Gene Mutation	Mutation that confers a selective growth advantage
Epigenetic	Changes unrelated to DNA sequencing that affect gene expression or phenotype
Genotoxic	Agents that induce direct DNA damage in their carcinogenic action
Germline	In the genome from conception, inherited. Example: A germline mutation is inherited.
Non Genotoxic	Agents that induce cancer, but not directly through DNA damage such as through tumor promotion, endocrine modification, immunosuppression, tissue-specific toxicity, and inflammatory responses
Oncogene	Proto-oncogenes normally function in the promotion of cell growth and cell division, however mutations cause uncontrolled division from usual upregulation mechanisms which force dominant or gain of function mutations. Activation of these genes to oncogenes is linked to cancer.
Passenger Gene Mutation	Mutation that has no effect on neoplastic process and occurs during general proliferation (successive clonal expansions). These mutations account for >99.9% of the alterations in tumors
Somatic	In a non-germ cell, is not inherited. Example: A somatic mutation is not inherited
Tumor Suppressor Gene	Normal function of tumor suppressor genes is to prevent or suppress cancer by keeping cell division regulated, which maintains the system to avoid cancer by down regulation of cell proliferation. However, loss of function mutations force inactivation of these mechanisms. Inactivation of these genes is linked to cancer.

* partially adapted from Vogelstein, B., Papadopoulos, N., Velculescu, et al. Cancer Genome Landscapes. *Science*. 2013; 339(6127),pp.1546–1558.