Developing a framework for incorporating real-world evidence into drug funding decisions:

CanREValue Collaboration Methods Working Group Progress Report on Real World Survival Data

CanREValue Collaboration

Overview of the CanREValue Collaboration

The Canadian Real-world Evidence for Value in Cancer (CanREValue) Collaboration is a pan-Canadian, multi-stakeholder initiative established in 2017 under a Canadian Institutes of Health Research Partnerships for Health System Improvement Grant. The central project is titled "Developing a framework for the incorporation of real-world evidence (RWE) into cancer drug funding decisions in Canada". The goal of the project is to develop and test a framework for the generation and use of RWE for cancer drugs to facilitate:

I. Reassessment of cancer drugs by recommendation-makers; and
II. Refinement of funding decisions, renegotiation of drug prices, or disinvestment as appropriate by decision-makers/payers across Canada.

Once developed, the framework could potentially be used to support evidence-based policy reform, pricing, and reallocation of funding from low- to high-value settings. In addition, the framework could facilitate the accountability and sustainability of the cancer system if used by the provinces and by other players in the healthcare system.

Working Groups

As part of developing the framework, five working groups (WGs) have been established to develop the framework (Figure 1). To ensure that the framework can support the needs of various stakeholders, the CanREValue Collaboration brought together a broad range of stakeholders from across different organizations and agencies (Figure 2). The WGs and their key deliverables are listed below.

- ***RWE planning and drug selection WG****:* Recommend criteria to identify and prioritize potential drug candidates for real-world evidence studies and advise on the necessary provincial infrastructure needed for the conduct of RWE.
- ***RWE Data WG:*** Recommend strategies for data access and provide advice on harmonization of data elements relevant for RWE studies across provinces.
- ***RWE Methods WG:*** Recommend methods to analyze real-world data with methodological rigor.
- ***RWE Reassessment and Uptake WG:*** Develop a process to incorporate RWE for HTA reassessment and advise on strategies to incorporate RWE results into policy-making.
- ***RWE Engagement WG:*** Establish mechanisms to ensure that stakeholders from across all relevant jurisdictions can provide feedback and input into proposed recommendations from each WGs.

.

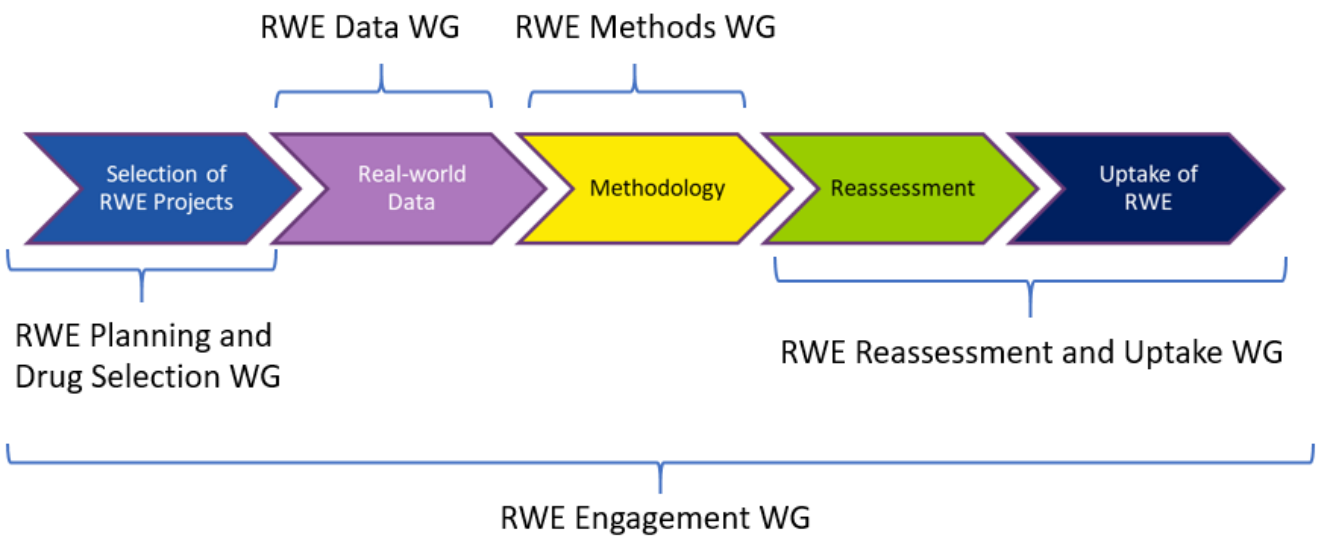

Figure 1: **CanREValue Collaboration Working Group (WG) Structure**

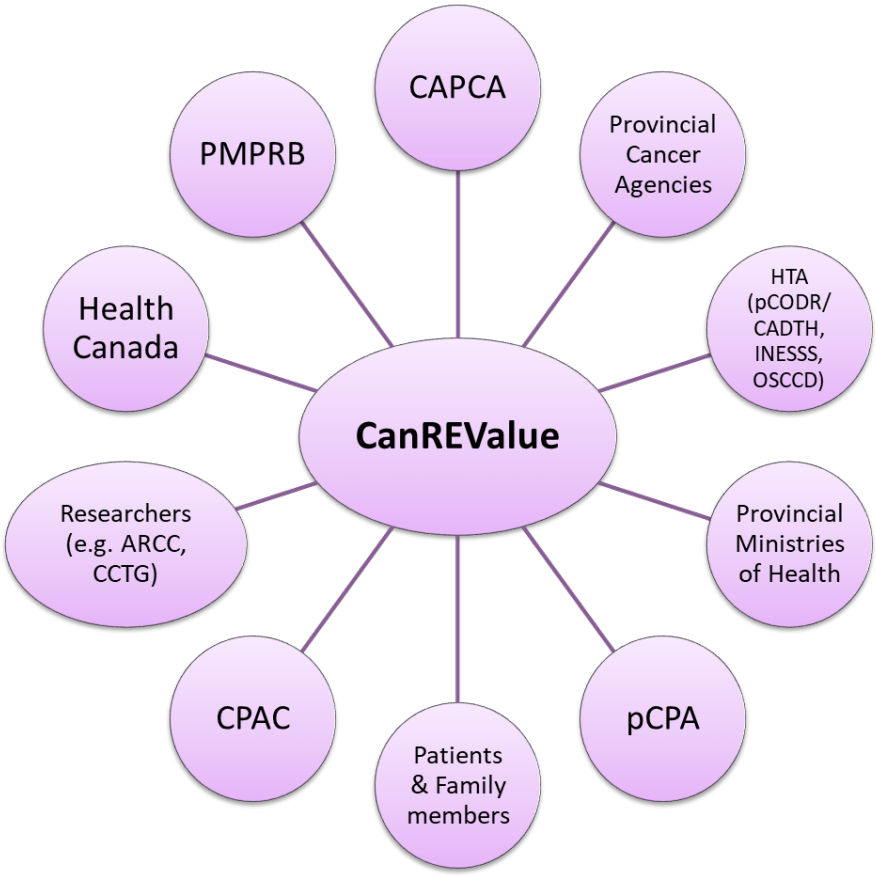

Figure 2: **CanREValue Collaboration Stakeholders**

CanREValue Collaboration Methods Working Group Progress Report on Real World Survival Data

Authors

Kelvin Chan (PI of the CanREValue grant)
Sunnybrook Research Institute
Canadian Centre for Applied Research in Cancer Control

Jeffrey S. Hoch (Methods Working Group Chair)
University of California, Davis

Jaclyn Beca
Ontario Health (Cancer Care Ontario)
Canadian Centre for Applied Research in Cancer Control

Lisa Currie
Health Canada

Wei Fang Dai
Ontario Health (Cancer Care Ontario)
Canadian Centre for Applied Research in Cancer Control

David Greiss
pan-Canadian Pharmaceutical Alliance

Richard Grieve
London School of Hygiene & Tropical Medicine

Miguel Hernan
Harvard University

Luke Keele
University of Pennsylvania

Nick Latimer
University of Sheffield

Rebecca Mercer
Ontario Health (Cancer Care Ontario)
Canadian Centre for Applied Research in Cancer Control

Ambica Parmar
Sunnybrook Research Institute

Petros Pechlivanoglou
Hospital for Sick Children

Eleanor Pullenayegum
Hospital for Sick Children

Michael Raphael
Sunnybrook Research Institute

Michelle Ross
Health Canada

Erin Strumpf
McGill University

Rinku Sutradhar
ICES
University of Toronto

Mina Tadrous
Women's College Hospital

Wanrudee Isaranuwatchai
Unity Health Toronto
Canadian Centre for Applied Research in Cancer Control

METHODS WORKING GROUP PROGRESS REPORT

PROGRESS REPORT

The CanREValue Methods Working Group was formed to recommend methods to analyze real-world data (RWD) with methodological rigor. The process we are undertaking to reach that goal include

1) Summarizing existing methods to consider for main outcomes;
2) Testing the identified methods using real-world data; and
3) Providing recommendations on which methods to consider in which context.

Currently, we are completing the first step, using "survival" as a specific outcome. Survival is one of the most commonly used and relevant outcomes in the cancer drug decision-making processes. It will be important in future work to examine other outcomes such as safety, cost-effectiveness, and patient reported outcomes, and appropriate methods will be determined and presented as that work is being undertaken. The following section provides a brief overview of the context and a summary of existing methods to consider when analysing real-world survival data.

CONTEXT

Real-world evidence (RWE) can make a valuable contribution to the evidence used by a broad range of stakeholders in decision-making and resource allocation processes. RWE can also help to answer comparative effectiveness questions efficiently, using fewer resources and covering larger, more representative populations than clinical trials. The need to assess RWE is driven by the fact that the clinical gains observed in trials are often modest, the trial results are frequently immature at the time decisions are required, and the cost of new therapies is great. In addition, there can be important differences in the performance of a new drug in controlled trial settings vs real-world settings. There are a number of reasons why clinical trials may yield different results from those generated in real world settings, such as higher adherence due to study protocol and follow-up requirements, or healthier subjects selected for inclusion in the trial population (1, 2). It is beneficial to assess the effectiveness of a new drug or regimen in the full population who would be seen in general practice. From a policy perspective, such evidence can be useful to reassess a drug for continued public reimbursement decisions.

RWE comes with its own set of strengths and limitations. Focusing on methods, the nature of study design (including the lack of randomization) to generate RWE comparing treatments requires methods to adjust for selection bias and the effects of confounding (3-5). While there are other important topics in the discussion of RWE in policy (including data availability and quality, as well as operationalization, implementation, and uptake of RWE), the specific objective of the Methods Working Group in this report is to provide information on the methodologies to generate RWE from the available data.

Choosing survival as an outcome

The Methods Working Group of the CanREValue Collaboration chose to first focus on the assessment of survival, as a clinically important and policy-relevant effectiveness outcome for the reassessment of cancer drugs using retrospective data from administrative databases. Overall survival is a reliable and

objective endpoint for assessing clinical effectiveness, ascertainable in administrative data, and an important outcome to all stakeholders.

Objective

This summary presents a review of methods to analyze and adjust for biases when assessing survival outcomes for comparative effectiveness of treatments in real-world, population-based, retrospective analyses using administrative data for post-funding reassessment and decision-making in cancer.

Methods

We conducted a narrative review, using published and grey literature, to identify potential methods to adjust for biases when conducting analysis of real-world survival outcome data. We briefly describe main methodological challenges in RWE and the selected methods that we believe are most relevant for real-world retrospective analysis of survival outcomes in the post-funding decision making phase.

SUMMARY OF FINDINGS

Main challenges of real-world survival data

Analysis of survival data can involve descriptive analyses in the form of Kaplan Meier curves, as well as statistical tests to assess differences between Kaplan Meier curves using log-rank tests. Methods to assess the magnitude of relative treatment effects on survival outcomes often involve Cox proportional hazards analysis, a regression-based analysis that estimates a hazard ratio between treatments. These methods are used in both clinical trial and real-world studies. These methods can assess treatment differences in randomized clinical trials without any further adjustments, since randomization of treatment group assignment minimizes systematic differences between groups.

Though there are a number of potential limitations from non-randomized, observational data, two main challenges are highlighted as particularly common and important to address when conducting real-world, retrospective analysis to compare survival outcomes in cancer. These challenges are common to observational data analysis, and they include:

- Baseline confounding by indication (e.g., people who receive Drug A may be sicker than people who receive Drug B), and
- Time-varying confounders (e.g., an important demographic or clinical characteristic might change through time).

The issue of baseline confounding occurs because patients in real-world settings are not assigned randomly to receive a specific treatment, as is the case in a randomized controlled trial. In other words, there may be some clinical factors that are associated both the choice of treatment and with the outcome of interest. Such clinical factors could include patient's performance status, disease severity, comorbidity, prior treatments received, lab tests or biochemical markers/status, as examples. To address this issue, imbalances in baseline characteristics between treatment groups must be accounted for in order to obtain an unbiased estimate of the relationship of interest i.e., the effect of the treatment on the outcome (survival).

The issue of time-varying confounders occurs because patients may experience changes over time. For example, patients may receive additional treatments after progression, which is also likely affect the outcome (survival). The use of subsequent treatments may differ between groups, which is where confounding may occur. Importantly, treatment sequence and the impact of reimbursement policies must be considered in any design. For example, if comparing patients receiving a new drug in the current period to a historical control group, there may be subsequent treatments available in the more recent period that patients in the control group could not access. Randomized controlled trials often address this issue by blinding treatment assignment, so that subsequent treatment choice will not be affected by the initial assigned treatment, or by restricting the treatments allowed after progression in the trial protocol. Crossover-adjustment methods are an example of methods that have occasionally been used to estimate survival in the presence of a specific type of time-varying confounding involving cross-over between the two treatment assignments in randomized clinical trials when patients in the control group receive the intervention after progression, though this is a narrower challenge than what is seen in real-world data.

Other limitations present when using retrospective observational data to conduct RWE of survival outcomes that may be relevant in certain contexts include:

- Time-varying primary exposure (e.g., similar to what occurs in a randomized controlled trial, person 1 may take Drug A for a different amount of time than person 2 takes Drug A);
- Time-varying effect of primary exposure (e.g., the drug's effect may vary through time in a particular person);
- Immortal time bias (e.g., when the period of immortal time is either incorrectly attributed to the treated group through a time fixed analysis or excluded from the analysis because the start of follow-up for the treated group is defined by the start of treatment and is, by design, later than that for the untreated group);
- Unmeasured patient-level characteristics (such as weight, genetic variation)
- Left truncation (delayed entry to a cohort);
- Informative censoring (which occurs when the event times are not independent of the censoring times); and
- Interval censoring (e.g., if actual survival time is not known); and
- Competing risks (e.g., when looking at time to disease progression, death is a competing risk).

From the above list of challenges, some items are more relevant than others in the context of cancer drug evaluation. Specifically, competing risks may be less relevant of an issue given the common use in oncology of progression-free survival (PFS) as an endpoint. PFS is defined to include both progression and death, such that there is no competing risk. Moreover, progression is often not defined in health administrative databases, limiting the ability to evaluate this endpoint. Immortal time bias and left truncation are commonly addressed in RWE studies for decision-making on drug therapies by defining a common start time based on initiation of treatment. Lastly, when the outcome is death, dates are usually accurately recorded such that interval censoring is not a challenge. It is important to assess for potential clustering of death dates or any patterns in censoring, and apply correction methods where appropriate.

In order to apply survival analysis methods to non-randomized, retrospective real-world data to assess comparative effectiveness of treatments, the above primary limitations related to confounding must be

addressed with additional methods to reduce bias in the estimates. The remainder of the report discusses methods that may be able to address biases due to confounding in real-world data.

Potential methods to consider

Methods available to address the abovementioned limitations related to confounding can be organized into the following four clusters; all of which have both advantages and disadvantages to consider. Please note that the following section is not meant to be a comprehensive list of methods. We welcome comments on other methods to consider.

1) ***Propensity scores (PS):*** *Propensity score matching (PSM) and inverse probability of treatment weighting (IPTW) (6-8)*

- Propensity score methods are commonly used methods to balance baseline characteristics. These methods summarize all baseline characteristics into a single score representing an individual's probability of receiving a particular treatment. Patients who have similar propensity scores have a similar distribution of the included baseline characteristics.
- Propensity scores are extensively used and described in the literature. Use of this method should align with established best-practices to ensure optimal use. Specifically, when matching, assessing balance, weighting, and reporting score performance (7-9).
- With propensity score matching (PSM), the outcome is compared between matched pairs who have similar characteristics. PSM estimates the effect of the study drug amongst the subset of patients receiving the study drug. This is relevant when only a subset of eligible patients receives the study drug, especially if there are clinical reasons for which a patient may be most suitable to receive the study drug. However, because a suitable match with similar characteristics has to be identified for each patient, the matched study cohort may not include all patients who are eligible for the study. As such, this method may not be suitable when studying drugs with a relatively smaller sample size.
- With inverse probability of treatment weighting (IPTW), each individual is given a weight based on their propensity score to create a theoretical population with all patients that are balanced on the observed characteristics. The outcome can be evaluated in the weighted cohorts, and represents the average change if a population of untreated patients were all given the study drug. IPTW can be extended to balancing time-varying confounders.
- Other methods (e.g., high-dimensional propensity score) using similar concepts are currently being evaluated (9).

2) ***Other matching methods****: Genetic matching (GM) (10-14) and disease risk score (DRS) (15-22)*

- Genetic matching (GM) is a generalization of PSM, which optimises covariate balance by using an automated iterative search algorithm. By matching on individual covariates, as well as the PS, GM can reduce bias and gain efficiency in the estimation of treatment effects, even when the PS is mis-specified.
- Disease risk score (DRS) is the prognostic analogue to PS. It is a confounder summary score derived from the probability of a disease outcome rather than treatment exposure. DRS is advantageous over other confounding adjustment techniques when exposure is

rare, when one wants to study multiple exposures, or if one wants to explore study effect modification by outcome risk. The DRS methods are limited in situations of rare outcomes and where a potential confounder is highly correlated with an exposure.

3) ***G methods***

- Under a collection of G methods, there are a number of models, which include the g formula (parametric and non-parametric), g estimation of a structural nested model, and inverse probability weighted marginal structural model. In brief, these methods allow estimation of the impact of treatment on outcome of interest without the requirement of strict assumptions as in standard regression models (23). These methods are appropriate for assessing both baseline and time-varying confounding.
- G formulas assess the outcome considering each level of each exposure/treatment and covariates. When the potential number of values of each covariate become large, assuming parametric distributions and simulation help to assess the potential outcomes over the probability distributions for the covariates (24).
- Inverse probability weighted marginal structural models involve reweighting of observations so that the levels of confounding variables become equally balanced between treated and untreated individuals. Reweighting can be conducted at each time point to address time-varying confounders.

4) ***Other Methods****: Covariate adjustment and instrumental variable (IV) (25-31)*

- Covariate adjustment can be conducted in any regression-based method to account for differences between groups. However, including a large number of variables (particularly when the sample size or number of events is small), leads to a problem of overfitting. This method is one of the more commonly used methods to address baseline confounding but other methods may be more effective in reducing biases in real-world data.
- Instrumental variables (IV) are used to balance treatment groups on baseline characteristics by way of an instrument, which is a variable associated with treatment choice but otherwise has no effect on the outcome. The treatment differences caused by the instrument are not confounded. Since the balancing process is not conditioned on all the observed baseline characteristics, (as they are in PS methods), the balance is not limited to observed confounders. This means that instrumental variable methods also address unobserved confounding when assumptions are met. However, it can be difficult to identify a suitable instrument, which can make this method difficult to apply.
- Of note, methods to address biases as a result of baseline and time-varying confounding in observational studies may also be applied to analyses integrating RWD with clinical trial data for match-adjusted indirect comparison and network meta-analysis studies. While beyond the scope of this report, many of the issues and methods above are relevant for the use of RWE to inform funding decision making. Further work is needed to address the application of methods for the purpose of informing comparison to clinical trial data.

SUMMARY

This document shares the progress being made by the CanREValue Methods Working Group in identifying methods to undertake and adjust for biases when assessing survival outcomes in RWE studies. In this report, we have presented a brief review of the existing methods for addressing confounding in the conduct of real-world, population-based, retrospective comparative effectiveness analysis of cancer drug treatments in which overall survival is the outcome of interest. We have also presented some of the limitations of these various methods.

We look forward to more interaction with the community in the coming months.

The CanREValue Methods Working Group is also working on a manuscript to provide an overview with more details on methods to analyze real-world survival data. Please contact us at canrevalue@cc-arcc.ca if you would like more information about this work.

Keywords

Cancer; Real-world evidence; Survival; Observational studies; Statistical analysis

REFERENCES

1. Rothwell PM. External validity of randomised controlled trials:"to whom do the results of this trial apply?". The Lancet. 2005;365(9453):82-93.
2. Horton R. Common sense and figures: the rhetoric of validity in medicine (Bradford Hill Memorial Lecture 1999). Statistics in medicine. 2000;19(23):3149-64.
3. Rochon P, Gurwitz J, Sykora K, Mamdani M, Streiner D, Garfinkel S, et al. Reader's guide to critical appraisal of cohort studies: 1. Role and design. British Medical Journal. 2005;330:895-7.
4. Normand S, Sykora K, Li P, Mamdani M, Rochon P, Anderson G. Reader's guide to critical appraisal of cohort studies: 3. Analytical strategies to reduce confounding. British Medical Journal. 2005;330:1021-3.
5. Mamdani M, Sykora K, Li P, Normand S, Streiner D, Austin P, et al. Reader's guide to critical appraisal of cohort studies: 2. Assessing for potential confounding. British Medical Journal. 2005;330:960-2.
6. Aalen OO, Cook RJ, Røysland K. Does Cox analysis of a randomized survival study yield a causal treatment effect? Lifetime data analysis. 2015;21(4):579-93.
7. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Statistics in medicine. 2015;34(28):3661-79.
8. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate behavioral research. 2011;46(3):399-424.
9. Austin PC, Wu CF, Lee DS, Tu JV. Comparing the high-dimensional propensity score for use with administrative data with propensity scores derived from high-quality clinical data. Statistical methods in medical research. 2019:0962280219842362.
10. Kreif N, Grieve R, Radice R, Sekhon JS. Regression-adjusted matching and double-robust methods for estimating average treatment effects in health economic evaluation. Health Services and Outcomes Research Methodology. 2013;13(2-4):174-202.
11. Abadie A, Imbens GW. Bias-corrected matching estimators for average treatment effects. Journal of Business & Economic Statistics. 2011;29(1):1-11.
12. van der Laan MJ, Gruber S. Collaborative double robust targeted maximum likelihood estimation. The international journal of biostatistics. 2010;6(1).
13. Kang JD, Schafer JL. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. Statistical science. 2007;22(4):523-39.
14. Hirano K, Imbens GW. Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. Health Services and Outcomes research methodology. 2001;2(3-4):259-78.
15. MIETTINEN OS. Stratification by a multivariate confounder score. American journal of epidemiology. 1976;104(6):609-20.
16. Arbogast PG, Ray WA. Use of disease risk scores in pharmacoepidemiologic studies. Statistical methods in medical research. 2009;18(1):67-80.

17. Arbogast PG, Ray WA. Performance of disease risk scores, propensity scores, and traditional multivariable outcome regression in the presence of multiple confounders. American journal of epidemiology. 2011;174(5):613-20.
18. Glynn RJ, Gagne JJ, Schneeweiss S. Role of disease risk scores in comparative effectiveness research with emerging therapies. Pharmacoepidemiology and drug safety. 2012;21:138-47.
19. Cadarette SM, Ban JK, Consiglio GP, Black CD, Dubins D, Marin A, et al. Diffusion of Innovations model helps interpret the comparative uptake of two methodological innovations: co-authorship network analysis and recommendations for the integration of novel methods in practice. Journal of clinical epidemiology. 2017;84:150-60.
20. Tadrous M, Mamdani MM, Juurlink DN, Krahn MD, Lévesque LE, Cadarette SM. Performance of the disease risk score in a cohort study with policy-induced selection bias. Journal of comparative effectiveness research. 2015;4(6):607-14.
21. Tadrous M, Gagne JJ, Stürmer T, Cadarette SM. Disease risk score as a confounder summary method: systematic review and recommendations. Pharmacoepidemiology and drug safety. 2013;22(2):122-9.
22. Xu S, Shetterly S, Cook AJ, Raebel MA, Goonesekera S, Shoaibi A, et al. Evaluation of propensity scores, disease risk scores, and regression in confounder adjustment for the safety of emerging treatment with group sequential monitoring. Pharmacoepidemiology and drug safety. 2016;25(4):453-61.
23. Naimi AI, Cole SR, Kennedy EH. An introduction to g methods. International journal of epidemiology. 2017;46(2):756-62.
24. Keil AP, Edwards JK, Richardson DR, Naimi AI, Cole SR. The parametric g-formula for time-to-event data: towards intuition with a worked example. Epidemiology (Cambridge, Mass). 2014;25(6):889.
25. Wunsch H, Linde-Zwirble WT, Angus DC. Methods to adjust for bias and confounding in critical care health services research involving observational data. Journal of critical care. 2006;21(1):1-7.
26. Newhouse JP, McClellan M. Econometrics in outcomes research: the use of instrumental variables. Annual review of public health. 1998;19(1):17-34.
27. Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH. Instrumental variables: application and limitations. Epidemiology. 2006:260-7.
28. Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? Epidemiology. 2006:360-72.
29. Greenland S. An introduction to instrumental variables for epidemiologists. International journal of epidemiology. 2000;29(4):722-9.
30. Angrist JD, Krueger AB. Instrumental variables and the search for identification: From supply and demand to natural experiments. Journal of Economic perspectives. 2001;15(4):69-85.
31. Penrod JD, Goldstein NE, Deb P. When and how to use instrumental variables in palliative care research. Journal of palliative medicine. 2009;12(5):471-4.