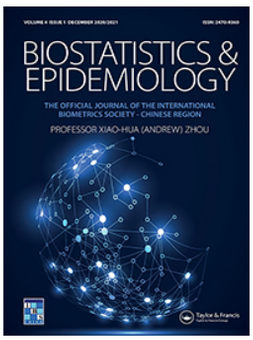

Biostatistics & Epidemiology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tbep20

Quasi-experimental design

Matthew L. Maciejewski

To cite this article: Matthew L. Maciejewski (2020) Quasi-experimental design, Biostatistics & Epidemiology, 4:1, 38-47, DOI: 10.1080/24709360.2018.1477468

To link to this article: https://doi.org/10.1080/24709360.2018.1477468

Published online: 07 Jun 2018.

BIOSTATISTICS & EPIDEMIOLOGY
2020, VOL. 4, NO. 1, 38–47
https://doi.org/10.1080/24709360.2018.1477468

Taylor & Francis
Taylor & Francis Group

Quasi-experimental design

Matthew L. Maciejewski[a,b]

[a]Center For Health Services Research in Primary Care, Durham VA Medical Center, Durham, NC, USA;
[b]Department of Population Health Sciences, Duke University Medical Center, Durham NC, USA

ABSTRACT

Quasi-experiments are similar to randomized controlled trials in many respects, but there are many challenges in designing and conducting a quasi-experiment when internal validity threats are introduced from the absence of randomization. This paper outlines design, measurement and statistical issues that must be considered prior to the conduct of a quasi-experimental evaluation. We discuss challenges for the internal validity of quasi-experimental designs, inclusion/exclusion criteria, treatment and comparator cohort definitions, and the five types of explanatory variables that one must classify prior to analysis. We discuss data collection and confidentiality, statistical power and conclude with analytic issues that one must consider.

ARTICLE HISTORY
Received 2 February 2018
Accepted 9 May 2018

KEYWORDS
Quasi-experimental design; retrospective analysis; observational study design; cohort

1. Introduction

A quasi-experiment is a prospective or retrospective study in which patients or clusters of patients self-select into (or their providers select on their behalf) one of several different treatment groups for the purpose of comparing the real-world effectiveness and safety of those non-randomized treatments. Quasi-experiments are observational studies that are similar to randomized controlled trials (RCTs) in many respects, with the primary exception being that patients self-select into different treatments instead of being randomized. Note that treatment and intervention are used interchangeably below. Given the lack of randomization in a quasi-experiment, there are many challenges in designing and conducting a quasi-experiment with strong internal validity. In this manuscript, we discuss some of these challenges and potential solutions.

For those researchers proposing to design and conduct a quasi-experiment, we highly encourage them to read the guidelines of the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement [1] and guidance manuscripts generated by International Society For Pharmacoeconomics and Outcomes Research (ISPOR) working groups [2] on retrospective [3] and prospective [4] observational studies. These documents enumerate important elements in the design, conduct, and publication of observational studies. In general, the better a quasi-experiment is designed and conducted, the easier the analysis and interpretation of the study will be at the end. The following

CONTACT Matthew L. Maciejewski matthew.maciejewski@va.gov
This article has been republished with minor changes. These changes do not impact the academic content of the article.

elements should be included in a research proposal for a quasi-experiment: background; objectives/hypotheses; study design; inclusion/exclusion criteria for patient population; treatment group definition and identification; comparator group definition and identification; collection, capture, and confidentiality of data on outcomes and explanatory variables; statistical power/sample size; and proposed statistical analysis.

Threats to internal validity should be clearly outlined in the discussion of study design, and strategies for addressing these internal validity threats via design, measurement and/or statistical methods should be discussed. Design elements relevant to internal validity threats include inclusion/exclusion criteria for overall sample, description and identification strategy of the treatment group, and description and identification strategy of the comparison group. Measurement elements relevant to internal validity threats include measurement of unobserved confounders, pre-period outcomes and outcomes that are not expected to be impacted by the treatment of interest (e.g. non-equivalent outcomes). Statistical methods relevant to internal validity threats include multivariable regression, propensity scores, instrumental variables and other methods. It may be helpful for a researcher to think about the ideal randomized trial that one might be able to conduct under the assumption of unlimited resources, access and time. Then, compare the proposed study to this idealized trial to identify the internal validity threats and ways in which these threats might be addressable.

External validity (e.g. generalizability) should also be addressed because quasi-experiments based on patients from multiple sites or a nationally representative sample can have greater external validity than a single-site RCT [5]. Challenges and solutions will be discussed below for each of these elements. The Panel on Statistics and Analytics on Veterans Health Administration (VHA) Datasets makes certain recommendations for solutions to these challenges that arise in quasi-experiments, but the investigator should consider these as guidelines that do not necessarily need to be followed in all circumstances. The discussion below is meant to engage researchers in any health care setting who are interested in evaluating whether a program, policy or intervention that is implemented on a non-randomized basis is working as intended.

2. Background

The challenge in developing a quasi-experimental evaluation is to be able to convince grant reviewers, a funding agency or journal reviewers that the study is important, that there is value to decision-makers in understanding whether the program, policy or clinical practice is working as intended in the real world (e.g. has real-world effectiveness), and that the study as designed can generate a valid estimate of the impact of a treatment on important outcomes. It is critically important to carefully address this 'So What?' question in retrospective evaluations of established programs or policies, in order to clearly and compelling state how the results of a quasi-experimental evaluation will inform any program or policy changes.

3. Objectives/hypotheses

The challenge in writing objectives/hypotheses to be tested or in describing hypotheses that have been tested is to keep them simple and understandable. We recommend one

primary objective and one hypothesis. This would be the most important of all of the objectives/hypotheses under consideration and the one on which sample size and power for the study is based. The primary objective/hypothesis should identify these four elements: patient population; intervention; control or comparison group; and outcome (PICO). Secondary objectives/hypotheses should be limited in number, and usually involve different secondary outcomes or planned subgroup analyses. For quasi-experiments, we recommend two-sided alternative hypotheses.

4. Quasi-experimental design

The experimental design of the quasi-experimental study and its rationale should be explicitly described. The most common quasi-experiment is a retrospective study of a single treatment cohort and a non-equivalent comparator cohort wherein patients self-select into (or a provider selects on behalf of a patient) either treatment or usual care and are then followed for a defined period of time. For example, a health care payer or insurer, such as Medicare in the United States, may be interested in understanding whether enrolled patients achieve better outcomes or lower costs when enrolled in an Accountable Care Organization (ACO). In this example, patients enrolled in the ACO may be compared to comparable patients not enrolled in an ACO in the same city or state.

Another common quasi-experiment is patients choosing between two treatments. A strength of this design is that it mirrors the 2-arm structure of many RCTs by enabling between-person comparisons of treatment and comparator groups and within-person changes of the same patients over time. A second advantage is that patients can potentially be followed for much longer (possibly years) than is possible in nearly all randomized trials, so patient outcomes 5 or 10 years after the initiation of treatment is possible if valid outcomes are measurable in claims or electronic health record (EHR) data and patients remain continuous users of the system of care. A third advantage is that there may be important questions that are only addressable via quasi-experiments, because a trial is too costly, is unethical or is not feasible. A fourth potential advantage is the potential greater generalizability of results from a quasi-experiment, due to the ability to examine outcomes in patients who might not otherwise participate in a randomized trial.

There are many disadvantages to quasi-experimental design studies. First, patients typically choose a treatment path from among available treatment options based on their own preferences, knowledge, past experiences, and in consultation with their health care provider. This means that patients who have chosen different treatment paths may differ in ways that would impact, or confound, their health outcomes. A second potential disadvantage of this design is that the intervention may not be well characterized, since it is unlikely under the control of the investigator as in an RCT. In addition, the intervention may be received by patients with variable (and unknown) fidelity across patients, providers or clinical sites.

Two-group cross-sectional designs are discouraged because the limitation of a single measurement for all subjects precludes any assessment of within-person changes of the same patients over time. Repeated measures on patients in both arms (as in an RCT) enable assessment of both within-person change over time and between-arm differences in a difference-in-difference analysis. Without repeated measures in a two-group

cross-sectional study design, causality is impossible to because temporal sequencing on the treatment and outcomes cannot be established [6].

The internal validity of this basic quasi-experimental design can be enhanced by other design, measurement and analytic methods that are outlined further below. In brief, the internal validity of a quasi-experiment can be greatly improved by including a matched comparator cohort with multiple pre-period assessments of the outcome and one or more non-equivalent outcomes that are not expected to change in response to the treatment of interest.

It should also be noted that prospective quasi-experiments are also valid and may provide the researcher an ability to collect data on outcomes and confounders that would not otherwise be available in a retrospective quasi-experiment that relies exclusively on data from administrative claims or EHRs. There are many other quasi-experimental designs and the reader is referred to Chapters 4–7 in the 2002 text by Shadish, Cook and Campbell [7] for more details.

4.1. Inclusion/exclusion criteria for the patient population

The challenges in identifying the target patient population for a quasi-experiment is deciding how broadly or narrowly to define the population and the extent to which (in a retrospective quasi-experiment) data exist to construct and apply the inclusion/exclusion criteria of a trial or real-world clinical practice. These challenges have implications on the internal and external validity of the study.

The patient population is defined by using inclusion and exclusion criteria, which will include geographic criteria (e.g. single site, multiple sites, nationally representative sample, entire population), as well as criteria related to demographic and clinical characteristics of patients themselves. Inclusion and exclusion criteria should be chosen to ensure that patients who self-select usual care (or an alternative treatment) are as similar as possible to patients who self-select the treatment of interest. If the population is narrowly defined in geographic and patient characteristics, with many exclusion criteria, the internal validity of the study may be enhanced, at the expense of the external validity (generalizability) of the study. If the population is broadly defined, the external validity of the trial will be enhanced, but the internal validity of the quasi-experiment might suffer. Internal validity is critical to ensure in any quasi-experimental study, particularly for research questions for which there is no experimental evidence. In such questions (e.g. effectiveness of Accountable Care Organizations or outcomes disparities due to race), the quasi-experiment will be the only means of developing an evidence base, so internal validity is paramount. For a research question for which there is experimental evidence to serve as a reference, it may be acceptable to accept a modest reduction in internal validity for a significant improvement in external validity. Ideally, researchers can conduct studies with strong internal validity and strong external validity because the two constructs can be complementary in quasi-experimental studies.

4.2. Treatment cohort definition and identification

Once the criteria for selecting the overall population (treated and untreated patients) are determined, there are several challenges in defining the self-selected intervention:

(a) patients receiving the intervention must be able to be clearly measurable from claims data, EHRs, or other data sources and (b) the intervention must be implemented fairly consistently in the real world across providers and clinics. If the intervention has multiple components or there is variation in how the intervention is implemented, we will not necessarily know which components had the best effect or which variations of implementation had the best effect, and which had little to no effect. Another issue related to a complex intervention is that there should be in the protocol a defined way to measure the fidelity of the intervention [8], i.e. the degree to which the intervention was applied in the field as envisioned in the original operational guidance or original study. For a multiple-component intervention or an intervention whose implementation can deviate from the idealized implementation, the data collection scheme should ideally have a plan to measure the intensity of use of each of the components for each patient or the ways in which the intervention deviates from the idealized implementation.

4.3. Comparator cohort definition and identification

The internal validity of an observational study is greatly weakened by the lack of a comparator (or control) cohort, so a comparator cohort should be included in the study design whenever possible. A comparator cohort for a quasi-experiment could be another active treatment chosen by patients eligible for the treatment of interest, or the comparator cohort could be patients who self-select no treatment, current standard of care, or usual care. Whatever the comparison group is defined to be, patients in the comparison group should be as similar as possible to patients who self-select the treatment of interest, which may require sample restriction to ensure that they are as comparable to the treatment group as possible in terms of eligibility and other factors [9]. Retrospective quasi-experiments relying on administrative claims or EHR data may not be able to identify all relevant inclusion and exclusion criteria, which may result in the comparator cohort including patients that would otherwise be excluded if unmeasurable exclusion criteria could be accounted for. Patients in the comparator cohort should be treated the same in all aspects of the quasi-experimental study. Restriction is particularly important when evaluating drug, surgical procedure or device interventions in which incident users may have different outcome trajectories than prevalent users [10].

4.4. Collection, capture and confidentiality of data for outcomes and explanatory variables

This section summarizes three issues: Issues with data collection in general, collection/measurement of outcomes, and collection/measurement of explanatory variables.

4.4.1. General issues in data collection

Data collection should be planned to meet the minimum needs of the protocol and publications, but should not be overly burdensome on patients and research personnel. Our experience has shown that the ideal data collection areas in a quasi-experiment should include: data to satisfy the primary and secondary objectives/hypotheses of the study; treatments that are self-selected; concomitant treatments that may confound the treatment effect of interest; variables to characterize the study sample (demographics, disease

severity); selected comorbidities that would be expected to differ between treatment and comparator cohorts to use in risk adjustment; and other confounders expected to bias the treatment effect if not adjusted for in statistical models. Methods for data capture, management, editing, and maintaining confidentiality should also be described in the proposal.

4.4.2. Collection/measurement of outcomes

Outcomes can be considered of two general types: outcomes of interest that are expected to change in response to the treatment and non-equivalent outcomes that are not expected to change in response to treatment, which have also been referred to as falsification tests or negative control outcomes. The primary outcome of interest in a quasi-experiment must be valid, reliable, relatively easy to ascertain, clinically relevant, feasible, and generalizable [11]. It must be of such importance that, if the intervention shows a treatment effect, the study has the potential to change clinical practice. Typical health outcome measures include mortality, major morbidity, clinical parameters such as blood pressure, lipid levels, etc. Typical economic outcome measures include health care utilization, health care costs, patient time costs or patient out-of-pocket costs.

Patient-centred outcomes – functional status, patient satisfaction with health care, health-related quality of life, caregiver burden – have gained greater prominence in recent years and merit consideration in the measurement set of prospective quasi-experiments [12]. These can be important outcomes, but it is also important to document studies of their reliability and validity and to describe the special analytic methods sometimes needed (e.g. to handle utilization and cost data, see section on Non-normality).

The internal validity of quasi-experiments can be improved by the inclusion of non-equivalent outcomes that are not expected to change in response to treatment [8,13]. Internal validity can also be improved by including one or more pre-period assessments of the outcome on the treatment and comparator cohorts to understand whether outcome trends in the pre-treatment period are similar between cohorts [7].

4.4.3. Collection/measurement of explanatory variables

Given the lack of randomization, the collection and measurement of explanatory variables is particularly important to ensure that an unbiased estimate of the treatment effect can be identified. Explanatory variables can be grouped into five categories (see Figure 1), based on their purpose in regression analysis: (1) explanatory variable of interest, (2) confounder, (3) control variable, (4) moderator and (5) mediator [14]. The explanatory variable of interest in a quasi-experiment is treatment or intervention that has been self-selected by patients or their providers on behalf of patients. Confounders are measures that are associated with the explanatory variable of interest and with the outcome but are not a consequence of the explanatory variable of interest. Age, gender, race and comorbid-ity measures are typically thought of as confounders in many quasi-experiments because patients choose treatments on the basis of these characteristics and these characteristics often explain variation in outcomes as well. See Vogal WB and Chen GJ, An Introduction to the Why and How of Risk Adjustment, is this issue, starting at p. 84, for a more detailed discussion. Failure to control for all important confounders will result in unobserved confounding and a biased estimate of the treatment effect.

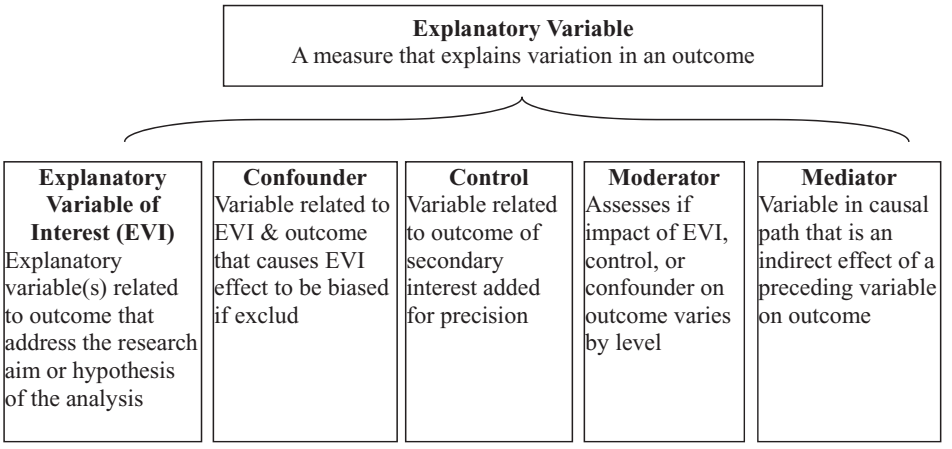

Figure 1. Types of explanatory variables.

Control variables are measures not experimentally manipulated by the scientist that are associated with the outcome and are unassociated with the explanatory variable of interest. In propensity score papers, control variables have also been referred to as risk factors. Control variables are included in a regression to improve overall model fit to the data and precision of other parameter estimates. Failure to include all the important control variables will not bias the estimate of the treatment effect, but may result in a less efficient estimate (e.g. larger confidence intervals). Moderators are measures that cause the magnitude of effect of a treatment on an outcome to differ according to the level of this explanatory variable. For example, the relationship between social support and medication adherence may be moderated by a patient's locus of control. Mediators are measures that determine the impact of another explanatory variable on the outcome, because it lies on the causal pathway between explanatory variable and the outcome. For example, dietary restraint is an important mediator of the relationship between stress and obesity [15].

To inform which variables are relevant to a particular treatment–outcome relationship, researchers should consider developing a conceptual model of all measures that are related to treatment selection and outcome determination. A conceptual model can then be used to identify the variables that are measured and the variables that are unmeasured, which can indicate the extent of unobserved confounding and suggest potential measurement opportunities. In a retrospective quasi-experiment relying exclusively upon claims data, measurement of otherwise unobserved confounding may be limited or completely precluded unless proxy measures can be identified from survey data or other administrative data-sets. A prospective quasi-experiment may provide an opportunity to do primary data collection to reduce the extent of unobserved confounding by measuring those variables that are expected *a priori* to be important (statistically significant) confounders in the treatment-outcome relationship.

4.5. Sample size/statistical power

Sample size determination and statistical power of the quasi-experiment are very important because they have major effects on the scientific validity of the study. The selection of

sample size may also impact the study's budget in terms of staff required to download, clean, construct and analyse large secondary data-sets, particularly if primary data collection is involved. In prospective quasi-experiments, study staff may also be needed to conduct primary data collection, and the cost of these efforts is greatly informed by the length of the survey instruments, the mode of data collection and the sample size. Sample size for the quasi-experiment should be based upon the primary outcome variable and in reference to the scientific literature in that area. For quasi-experiments based on very large samples, it is also important to specify a clinically meaningful result if statistical estimates are likely to be highly significant due to the large sample size. Sample size is a function of event rates assumed in the control group, variability of the outcome measure, a minimally important treatment effect to detect, desired statistical power, and the α-level of the statistical test used in the primary analysis. If data is hierarchical in structure, because patients are clustered in providers and providers are clustered in clinics or hospitals, then the size of intra-class correlation will also impact statistical power.

The method used to calculate sample size should be specified in detail in the proposal so that a reviewer can check the calculations if desired and should use the same methods as are anticipated for the final primary analysis of the trial. The treatment effect desired to be detected by the trial should be realistic. Minimum acceptable statistical power is typically 80%, although a 90%–95% power is preferable (particularly in studies with very large samples or nationally representative samples). It is useful to present in the proposal a range of sample sizes depending upon a range of assumptions about the underlying parameters and then to choose a target sample size for the study and give reasons for this choice.

4.6. Statistical analyses

The final analysis and interpretation of the study may be straightforward if researchers pay careful attention to its design and conduct to ensure internal validity threats are addressed, such that all covariates are balanced between treated and untreated cohorts. If observed covariates and confounders are imbalanced and there are important confounders that are unobserved, then the statistical analysis may require several different analyses. Internal validity threats that have not been addressed by design and measurement strategies must be addressed by statistical methods whenever possible.

Statistical issues in quasi-experiments of three kinds can arise that may require additional analysis beyond the pooled analysis of the primary outcome (and secondary outcomes): (1) imbalance in observed covariates and confounders [16], (2) imbalance in unobserved confounder, and (3) heterogeneity of treatment effects [17,18].

Imbalance in observed covariates and confounders can be addressed by regression adjustment or propensity score methods. Evaluating differences in patient characteristics by treatment and comparator cohorts can provide evidence of the extent of imbalance in observed covariates, and may suggest whether one should be concerned about imbalance in unobserved covariates and confounders. Imbalance in unobserved confounders can be addressed by instrumental variable methods [19,20]. Finally, it may be critical to assess whether the average treatment effect estimated in the pooled cohort is representative of the entire cohort or whether there is heterogeneity in the treatment effect (see paper in this issue (ADD CITATION HERE) on heterogeneity of treatment effects). Any adjusted

analyses should be pre-specified in the proposal and manuscripts, and any post-hoc analyses should be noted as such in manuscripts.

If quasi-experiments are thoughtfully designed and analysed with appropriate measurement, they can have internal validity approaching that of RCTs and greater external validity than RCTs. Internal validity threats not addressed via design or measurement can often be addressed via careful statistical analysis, which may increase the concordance between quasi-experiments and RCTs as has been shown in prior meta-analyses [21–23].

5. Conclusion

There are many clinical and policy questions that cannot be addressed via RCTs, so quasi-experiments represent the only means by which outcomes of patient undergoing treatment can be compared to patients undergoing an alternative treatment or usual care. The absence of randomization introduces many threats to internal validity that may be addressable through careful design, measurement or statistical analysis (if all else fails). Only through faithful consideration of the internal validity threats prior to getting into the data can the quality of quasi-experimental evaluations gradually improve over time.

Acknowledgments

Dr Maciejewski received support from VA HSR&D for Research Career Scientist award (RCS 10–391) and owns Amgen stock due to his spouse's employment.

Disclosure statement

No potential conflict of interest was reported by the author.

Funding

VA HSR&D for Research [grant number RCS 10-391] and HSR&D funding for SDR [grant number SDR 13-426].

Notes on contributor

Matthew L. Maciejewski is a research career scientist in the Durham VA HSR&D Center of Innovation and professor in the Department of Population Health Sciences at Duke University.

References

[1] Vandenbroucke JP, Von Elm E, Altman DG, et al. For the STROBE initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. PLoS Med. 2007;4:1628–1654.

[2] Cox E, Martin BC, Van Staa T, et al. Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: the international society for pharmacoeconomics and outcomes research good research practices for retrospective database analysis task force report–part II. Value Health. 2009;12(8):1053–1061.

[3] Johnson ML, Crown W, Martin BC, et al. Good research practices for comparative effectiveness research: analytic methods to improve causal inference from nonrandomized studies of treatment effects using secondary data sources: the ISPOR good research practices for retrospective database analysis task force report–part III. Value Health. 2009;12(8):1062–1073.
[4] Berger ML, Dreyer N, Anderson F, et al. Prospective observational studies to assess comparative effectiveness: the ISPOR good research practices task force report. Value Health. 2012; 15(2):217–230.
[5] Imai K, King G, Stuart EA. Misunderstandings between experimentalists and observationalists about causal inference. J R Stat Soc Ser A (Statistics in Society). 2008;171(2):481–502.
[6] Holland PW. Statistics and causal inference. J Am Stat Assoc. 1986;81(396):945–960.
[7] Shadish WR, Cook TD, Campbell DT. Experimental and quasi-experimental designs for generalized causal inference. Boston (MA): Houghton Mifflin Company; 2002.
[8] Prasad V, Jena AB. Prespecified falsification end points: can they validate true observational associations? JAMA. 2013;309(3):241–242.
[9] Schneeweiss S, Patrick AR, Stürmer T, et al. Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results. Med Care. 2007;45(10 Supl 2):S131–S142.
[10] Danaei G, Tavakkoli M, Hernán MA. Bias in observational studies of prevalent users: lessons for comparative effectiveness research from a meta-analysis of statins. Am J Epidemiol. 2012;175(4):250–262.
[11] Henderson WG, Demakis J, Fihn SD, et al. Cooperative studies in health services research in the department of veterans affairs. Control Clin Trials. 1998;19:134–148.
[12] Estabrooks PA, Boyle M, Emmons KM, et al. Harmonized patient-reported data elements in the electronic health record: supporting meaningful use by primary care action on health behaviors and key psychosocial factors. J Am Med Inform Assoc. 2012;19(4):575–582.
[13] Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. Epidemiology. 2010;21(3):383–388.
[14] Maciejewski ML, Weaver EM, Hebert PL. Synonyms in health services research methodology. Med Care Res Rev. 2011;68(2):156–176.
[15] Roberts C, Troop N, Connan F, et al. The effects of stress on body weight: biological and psychological predictors of change in BMI. Obesity. 2007;15:3045–3055.
[16] Austin PC. A Tutorial and case study in propensity score analysis: an application to estimating the effect of in-hospital smoking cessation counseling on mortality. Multivariate Behav Res. 2011;46(1):119–151.
[17] Kent DM, Rothwell PM, Ioannidis JP, et al. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. Trials. 2010;11:85.
[18] Lipkovich I, Dmitrienko A, D'Agostino B R, Sr. Tutorial in biostatistics: data-driven subgroup identification and analysis in clinical trials. Stat Med. 2017;36(1):136–196.
[19] Newhouse JP, McClellan M. Econometrics in outcomes research: the use of instrumental variables. Annu Rev Public Health. 1998;19:17–34.
[20] Greenland S. An introduction to instrumental variables for epidemiologists. Int J Epidemiol. 2000;29(4):722–729.
[21] Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. N Engl J Med. 2000;342(25):1878–1886.
[22] Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med. 2000;342(25):1887–1892.
[23] Furlan AD, Tomlinson G, Jadad AA, et al. Methodological quality and homogeneity influenced agreement between randomized trials and nonrandomized studies of the same intervention for back pain. J Clin Epidemiol. 2008;61(3):209–231.