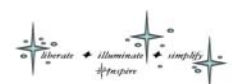




Problems with the Use of Observational Studies to Draw Cause and Effect Conclusions About Interventions

<p>Definition</p>	<p>Observational Study Defined</p> <p>Epidemiological study in which observations are made, but investigators do not control the exposure or intervention and other factors. Changes or differences in one characteristic are studied in relation to changes or differences in others, without the intervention of the investigator. Observational studies are highly prone to selection and observation bias and confounding.</p> <p>Tip: If an intervention is "assigned" through the research, it is an experiment. If it is chosen, then the study type is observational.</p>
<p>Key Points</p>	<p>While observational studies have been used by many for drawing cause and effect conclusions for interventions, frequently the conclusions have been wrong, such as with hormone replacement studies as treatment for preventing subsequent cardiac events in women with previous myocardial infarction. This is due to special opportunities for bias, confounding and chance which are reduced or eliminated by an experimental design, especially if randomization is used.</p> <p>If evidence regarding efficacy from valid, relevant randomized controlled trials is not available, evidence-based decisions are rarely possible. Exceptions may occur in situations where specific criteria have been met (see last entry below), giving preference to study designs or information in the following order: 1) experiments, 2) cohort designs, 3) case/control studies, 4) database studies, and 5) case series.</p> <p>Bottom Line: The positive predictive value for true results with a high quality RCT is approximately 85%; the positive predictive value for a high quality observational study is approximately 20%. Ioannidis JPA. Why Most Published Research Findings are False. <i>PLoS Med</i> 2005; 2(8):696-701.</p>
<p>Key Problems</p>	<p>Baseline differences in groups, choice and lack of control create special challenges in observational studies and are likely to affect the observed outcomes. These problems are amplified below (Epi-speak included.)</p>
<p>Bias Choice Lack of Control Chance</p>	<p>There are numerous instances where observational studies have been in agreement with RCTs or have provided helpful insights and solutions to health care problems. However, there are also numerous instances in which observational studies have not been in agreement with RCTs and have misled end users. For example, based on numerous observational studies reporting benefits, clinicians and patients accepted hormone replacement therapy (HRT) as an effective therapy for prevention of second cardiac events in women who had experienced a myocardial infarction. However RCTs demonstrated that subsequent cardiac events were not prevented by HRT; in fact RCTs revealed significant adverse events associated with HRT. Similarly, observational studies reported decreased mortality with vitamin A, beta-carotene and vitamin E, but high quality RCTs demonstrated increased mortality with these antioxidants. In a review of 18 meta-analyses (1211 clinical trials), outcomes of non-randomized trials varied from outcomes of RCTs (8 studies) from 76% underestimation to 160% overestimation. (Kunz R, Oxman A. <i>BMJ</i> 1998. Vol317:1185-90.)</p> <p>Differing Baseline Characteristics, Choices and Lack of Control</p> <p>You need two equal groups treated in the same way (except for the intervention being studied and its comparator) to understand if the intervention—and not some baseline difference or difference in experience or treatment between the groups—caused the difference in outcomes. In observational studies, investigators attempt to minimize baseline differences between groups through statistical adjustments for the differences in baseline variables such as sex, age, disease severity, co-morbidities, etc.; however, confounders (both known and unknown) are not adequately managed through this strategy. The actual factor responsible for the outcomes would have to be</p>



Problems with the Use of Observational Studies to Draw Cause and Effect Conclusions About Interventions

adjusted, which would require prior knowledge of the answer to the question that the research is designed to answer. Differing baseline characteristics (selection bias) remain a major threat to validity in all observational studies because adjustment models cannot reliably “equalize” differences between groups.

In addition to baseline differences between groups, observational studies are flawed by the many different exposures and choices likely to be made by subjects (or their health care clinicians). Differences in experiences (e.g., co-interventions) are likely to represent important confounders which, in some cases, represent the true reason for the differences in outcomes. These differences in exposures and choices could be due to —

- Health status of patients
- Provider skills
- Provision of care (e.g., affordability which can link to socioeconomic factors or patient demand, etc. which could drive provision of care and link to other confounding factors)
- Patient perceptions (e.g., risk aversion) and personal characteristics (e.g., more well-read, healthy-user effect)
- Other unknown confounders (e.g., genetic issues, environmental exposures, risk factors)
- It should be emphasized that in observational studies baseline differences, interventions, co-interventions and other exposures cannot be adequately controlled. This lack of control is likely to result in selection bias, performance bias, differences in follow-up (attrition bias) and differences in how outcomes are evaluated (assessment bias).

Examples of How Baseline Differences in Patients and “Choice” May Affect Outcomes

- Physician directs patients who are more likely to have favorable outcomes (i.e., those with better prognostic characteristics) to one treatment over another
- Physician favors one treatment over another due to training or skill level
- A more demanding patient may receive more monitoring or additional care
- A patient who is more educated, intelligent, information-seeking or affluent might engage in other behaviors that contribute to the outcomes
- Risk aversion may be associated with other personal factors which may be confounders

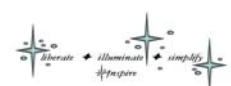
Examples of Methods Used in RCTs to Control or Minimize The Above-Mentioned Confounders Found in Observational Studies

Factors that may be controlled in experiments, but not in the “natural world” of care or observational studies (where patients choose their interventions) include —

- Achieving comparability of baseline risks through randomization
- Blinding
- Controlling uniformity in procedures, follow-up length and treatment methods such as dosing or specifications on which and how procedures are done
- Controlling allowed and disallowed medications and wash-out
- Adherence measurements and concomitant medication monitoring
- Prevention of contamination or “migration” to other arm of the study
- Quality control for standardization such as in training, measurement and reporting
- Just-in-time monitoring for pre-specified definitions of treatment success or failure or other endpoints

Examples of How “Lack of Control” in Observational Studies May Affect Outcomes

- Adherence is more likely to be poorer outside of a controlled trial
- Patients may be taking other medications (there are no “disallowed” medications in observational studies—and you really can’t be sure what a patient is doing)
- Other procedures may be used which could be the actual explanation for the observed results
- Monitoring, measurement and reporting of outcomes are likely to be driven by the choice of treatment and are likely to vary between individuals
- People tend to root for the intervention—you will notice what you are primed to notice—and there is no blinding in observational studies



Problems with the Use of Observational Studies to Draw Cause and Effect Conclusions About Interventions

The Play of Chance in Observational Studies:

- When multiple exposures are examined in observational studies (e.g., in case-control studies evaluating associations between exposures (independent variables) such as diet, weight, socioeconomic status, exercise, etc. and cancer (dependent variable) there is an increased probability of finding a statistically significant differences in outcomes between groups in some variables because of chance.

You Can't Blind in an Observational Study

- Blinding matters. This has been established in RCTs by comparing results of blinded RCTs with results of non-blinded RCTs. Lack of blinding in RCTs has been reported to exaggerate benefit by up to a relative 48% (Schulz, PMID: 7823387, Kjaergard PMID: 11730399, Moher PMID: 9746022)
- In observational studies patients choose their interventions, care-givers are not blinded and outcomes assessors are frequently not blinded.

Even if the Observational Study Leads us in the Right Direction...

With rare exceptions (see below) the only way to know that results from an observational study are likely to be true is to know that a valid RCT has reported similar results (confirmed the findings of an observational study). Furthermore, even if the results of an observational study suggest benefit from an intervention, observational studies are more prone to bias and are thus likely to falsely inflate the results favoring the intervention.

Theorized as
Possibly
Reliable
Indicators of
Efficacy in
Observational
Studies (Grade
U [Uncertain]
Evidence)

We have mentioned that there are rare exceptions to the points made above. There is general agreement that causality *MAY* be present when there is dramatic change following application of an intervention or technology that is unlikely to be due to confounding (i.e., close to all-or-none results: example, before treatment all died and following treatment, high survival rate). Examples of all-or-none results are antibiotics for meningitis and electrical cardioversion for ventricular fibrillation

Public health interventions represent another area where valid RCTs may not be available, yet efficacy of interventions has frequently been accepted even when outcomes are less dramatic than all-or-none results. For example, purification of water, the use of smoke alarms and seatbelts, many immunizations and some screening initiatives have been accepted by most as providing beneficial effects on health outcomes. In some instances studies of public health initiatives may allow cause and effect conclusions. Confounding while present, may be more limited in some public health interventions than in many observational studies because of the extremely broad application of the public health interventions.

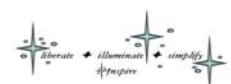
Below we list what we believe to be other key considerations for concluding that the effects reported in an observational study are due to the interventions, (i.e., that causality may exist).

Primary Considerations for Determining Efficacy in Observational Studies:

1. There is extremely low likelihood of improvement without some intervention; and,
2. Any effects are highly likely to be attributable to the intervention (e.g., a single intervention or technology was utilized and the likelihood of patients utilizing co-interventions is low or co-interventions were equivalent in compared groups [note: equivalence should take into account considerations such as administration, duration, dosing, etc.]); and,
3. Convincing sustained improvement is documented following use of the intervention or technology.

OR

1. Illness or condition has a predictable course or outcome (e.g., symptoms, diagnostic signs, progression or resolution of the condition are predictable); and,
2. Any effects in this population are likely to be attributable to the intervention (e.g., plausible confounders have been considered, a single intervention or technology was utilized and the likelihood of patients utilizing co-interventions is low or co-interventions were equivalent in groups being compared [note: equivalence should take into account considerations such as administration, duration, dosing, etc.]); and,
3. Convincing sustained improvement in this particular population is documented following use of the intervention or technology with results demonstrably and significantly earlier than what is predicted based on reliable



Problems with the Use of Observational Studies to Draw Cause and Effect Conclusions About Interventions

knowledge of natural history.

Secondary Considerations Lending Support for Concluding Efficacy:

Factoring in the following conditions may decrease uncertainty:

- Immediacy of effect which may increase confidence in the effectiveness of the intervention or technology not only because of temporal proximity, but also because a delay in effect provides more opportunity for application of a co-intervention which could be a confounder. When considering immediacy, caution is urged because the effect could be due to potential of confounding.
- Dramatic effects (e.g., approximately 80% or greater relative risk reduction). When considering dramatic effects, caution is urged because the effects could be due to the strength of a confounder. For example, The Cochrane Handbook states that observational evidence with large effect sizes where a dose-response gradient exists and all plausible confounding would most likely demonstrate a reduced effect size, may be of moderate or even high quality. Higgins JPT, Green S (2008) Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.0 [updated February 2008]. The Cochrane Collaboration. Available from www.cochrane-handbook.org accessed 08/07/2008. Section 12.2.3]. However, we remain cautious about this conclusion as confounding may result in large differences between groups. Example: in one study of adolescents with migraine headache the placebo effect was ~ 60% (Winner P, Linder SL, Lipton RB, Almas M, Parsons B et al. (2007) Eletriptan for the acute treatment of migraine in adolescents: results of a double-blind, placebo-controlled trial. *Headache* 47:511-8).
- Dose/response relationship. When considering dose/response relationship, caution is urged because patient expectation bias may result in increased perception of benefit or increased placebo effect if patient is aware of the greater dosing through increase in side effects, for example.
- Consistent changes in multiple outcomes. When considering consistent change, caution is urged because the change in multiple outcomes may be due to confounding (e.g., a causal association between an unknown confounder and each of the outcomes).

Uses of Observational Studies

Observational Study Types

- Cohort (best overall observational study type)
- Cross-sectional (useful in diagnostic testing, though RCTs used also)
- Case-Control (weakest design)

Observations are best used for—

- Generating hypotheses and ideas for study topics
- Understanding the natural history of diseases and prognosis
- Studies of diagnostic testing
- Studies of quality of life
- Satisfaction (caution as this frequently confounded)
- Understanding the effects of clinical improvement activities via process measurement and success of implementation, etc.

Some other uses include the following (special thanks to Jeff White, PharmD, WellPoint Rx)—

- Provide supportive information beyond efficacy of the generalizability of an RCT as it is risky to assume that efficacy evidence in a RCT will result in similar benefits to patients in a "real-world" environment. For example, what percent of members appear to benefit from an intervention given the circumstances of the real-world environment, recognizing that there may explanations other than causality for the observed outcomes (e.g., prognostic variables, biases and chance).
- Provide insight into real-world practice patterns, given the impact of non-compliance, switching medications, comorbidities, marketing, patient preferences, etc.
- Understand the degree of off-label use by linking individual drug to an ICD-9 diagnosis code, for example.
- An assessment of the total cost of care (pharmacy and medical) for members on various therapies recognizing that the cost estimate may not apply to other locations or populations and that the cost estimate may have been distorted by prognostic variables, bias and chance present in the data.
- Safety assessment recognizing that bias, confounding and chance may explain or distort the findings.

