NERDCAT-Obs:

A clinician's guide to appraising observational studies (cohort & case-control studies)

Table of Contents

Note: This critical appraisal tool builds from the fundamentals learned from reading and appraising the more straightforward randomized controlled trial (RCT). For more on fundamental and advanced topics in clinical trials, please see NERDCAT-RCT at www.nerdlmps.wordpress.com

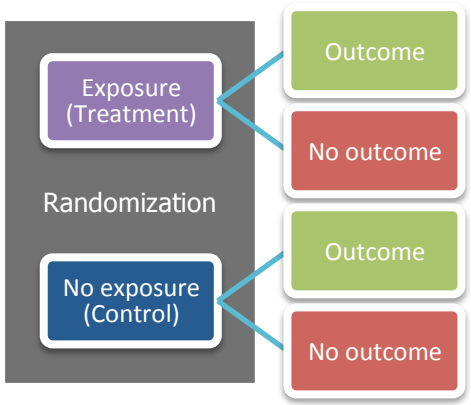
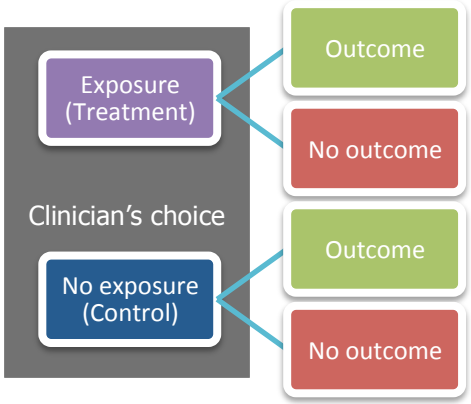
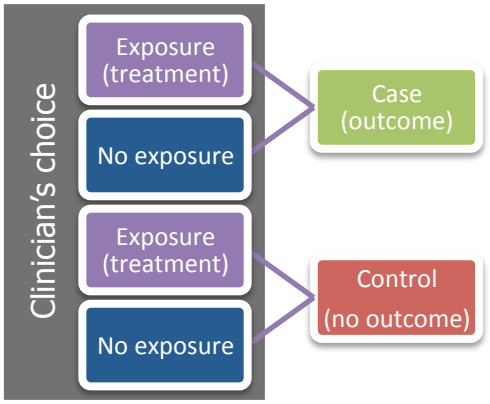
Study title: _____

CLINICAL QUESTION & STUDY FLOW		Data source	Definition
D	Study characteristics: • Study type: • Setting: • Year of study:		
P	Average patient: • Age: • Sex: • Race: • Pathology, stage/severity of disease: • Comorbidities: • Previous interventions:		
I/C	Drug, dose, route, duration Co-interventions:		
O			

GENERALIZABILITY	
1. Does my practice setting differ significantly from that in the trials? Some questions to consider: • Era: Same/similar diagnostic criteria used for disease/outcome? • Setting: 1^{o}, 2^{o}, or 3^{o} care? • Single centre or multicentre study? • Country: Were there Canadians in the study?	
2. Were differences between study participants & my patients (i.e. SCRAPP characteristics) clinically significant?	
3. Are the interventions evaluated in the study similar to those available in my practice?	
4. Can I readily measure the outcomes evaluated in the study?	

INTERNAL VALIDITY – Are the results reliable? Answering slightly different questions based on design: Cohort: Aside from the exposure of interest, did the exposed & non-exposed groups start & finish with a similar risk for the outcome? Case-control: Did the cases & controls have a similar risk for being exposed in the past?	
5. Confounding (including 'confounding by indication') Cohort: Do exposed/non-exposed have the same likelihood & severity of prognostic factors known to be associated with the outcome? Case-control: Did the cases & controls have a similar enough risk for being exposed in the past?	Known risk factors for outcome (circle those accounted-for in study):
6. Reverse causation (a.k.a. protopathic bias)	
7. Misclassification & detection bias (including recall bias) Cohort: Were similar methods used to detect the outcome? Did similar circumstances lead to investigation for the outcome? Case-control: Were the circumstances/methods for determining exposure similar for cases & controls?	
8. Attrition bias Cohort: Was follow-up sufficiently complete in both exposed and non-exposed participants?	
9. Immortal-time bias Cohort: Was there a differential lag time between start of follow-up and eligibility for the exposure or non-exposure group?	
10. Were methods used to minimize above issues with confounding/bias sufficient (determine for each issue identified)? Methods to minimize confounding/bias in observational studies: 1. Design: Restriction, matching (± propensity score, instrumental variable) 2. Analysis: Stratification, multivariable regression (± propensity score)	

RESULTS	
11. Strength of association How large is the magnitude of association? • <2: Small (may be result of residual bias or confounding from study design) • 2-5: Moderate • >5-10: Large (more likely to represent a true association between exposure and outcome) Is there a dose-response gradient (i.e. increasing HR/OR/RR with larger doses or longer duration of exposure)?	
12. Precision Information obtained from the confidence interval: • Are the results statistically significant? • Does the magnitude of association differ between the lower and upper bound of the confidence interval? • Are results clinically important at both bounds of the confidence interval?	

	RCT	Cohort	Case-Control
Flow		 *<u>Prospective:</u> As with RCTs, follows patient forward in time from baseline to end of follow-up.* *<u>Retrospective:</u> Use historical data (e.g. from a registry) to reconstruct patient characteristics, EXPOSURE status & OUTCOME.*	 *Patients are identified based on whether or not they experienced an OUTCOME of interest. Then, we look to see if they were exposed to the drugs/toxins/etc. of interest prior to experiencing this outcome.*

Pros	• Lowest susceptibility to bias when done right (high internal validity)
	• Can establish, between an intervention and an outcome:
	  ○ Cause-and-effect relationship;
	  ○ Dose-response relationship;
	  ○ Temporal relationship
		• Good generalizability
		  ○ "Real-world" population is well represented
		  ○ Can get a realistic and accurate incidence of OUTCOMES of interest
		• Measures effect of EXPOSURE on >1 OUTCOMES
		• Can explore dose-response & temporal relationships between an EXPOSURE & OUTCOME
		• Not susceptible to recall bias
			• Measures association of rare OUTCOMES
			  ○ Would otherwise require massive RCT/cohort
			• Can evaluate association of multiple EXPOSURES/variables on a single OUTCOME
			• Able to evaluate association between OUTCOMES that take a long time to develop (e.g. cancers) and EXPOSURES

Cons	• Often poor generalizability (low external validity)
	• Sample size usually limits detection of rare outcomes
		• Susceptible to **confounding**, **selection/allocation bias**, performance bias, detection bias
		• High loss-to-follow-up makes it difficult to find associations between an exposure with outcomes that take a long time to develop (e.g. cancer)
		• <u>Retrospective:</u>
		  ○ May not have measured patient characteristics of interest (not able to adjust for residual confounding)
		  ○ May have high incompletion of data
			• Often too unreliable enough to determine causation
			• **Susceptible to confounding and same biases as cohorts** plus, in some instances, additional biases (e.g. recall bias)

INTERNAL VALIDITY: Although observational studies include some unique biases, the most important ones are the same as with RCTs, but with their own twists and ways of handling them.		
Bias/confounder	Explanation	Example
Confounding by indication (checklist question #5)	• A form of selection bias; • The most common threat to validity in observational studies; • Occurs when an individual is more likely to receive a treatment because of an underlying indication that also happens to be a risk factor for the outcome under study.	Consider a situation where a patient with rheumatoid arthritis develops an upper GI bleed on naproxen. Despite this unpleasant experience, she insists on continuing some sort of NSAID for ongoing, disabling pain in both upper limbs. Her physician prescribes celecoxib and a PPI in order to minimize her risk of recurrent upper GI bleed, but she unfortunately dies from gastric perforation 1 year later. • If, in practice, celecoxib is mostly prescribed to patients at highest risk of a recurrent bleed, as above, a cohort study that does not account for (or cannot acquire information about) previous upper GI bleeds may falsely conclude that celecoxib has a similar or greater risk of upper GI bleed compared to non-selective NSAIDs.
Reverse causation (a.k.a. protopathic bias) (checklist question #6)	Occurs when a precursor to the outcome of interest or a subclinical version of it leads to the exposure of interest. The full-blown version of the outcome of interest presents itself after outcome, which may then lead to the false belief that the exposure caused the outcome.	Consider a situation where an obese patient with undiagnosed coronary artery disease (CAD) is experiencing intermittent retrosternal chest pain (which is actually stable angina). He goes to the walk-in clinic, is diagnosed with GERD and given a PPI. He takes the PPI for 1 week and believes he experiences relief, but wakes up one morning with intractable chest pain. He calls 9-1-1, gets taken to the local hospital and is diagnosed with NSTEMI; • In other words, subclinical manifestation of the outcome of interest (stable angina from CAD) led to exposure (PPI use), which was followed by an outcome (MI); • If this situation is prevalent enough, a cohort study looking at the association between PPI use and MI may falsely conclude that PPIs cause MIs.
Misclassification (checklist question #7)	Can occur in 2 ways: • Random (i.e. non-differential): Subjects in both groups have equal opportunity to be misclassified. This type of misclassification leads to imprecision in the measure of association, which reduces the study's power and ability to find a difference (false-negative); • Biased (i.e. differential): Subjects in 1 group are more likely to be misclassified; ○ Leads to exaggeration or attenuation of the estimate of association	Random misclassification: Consider that hospital physicians frequently omit acute kidney injury (AKI) from their documentation, which in turn leads to omission from administrative databases for up to 85% of patients.[Am J Kidney 2011;57:29-43] • A cohort study using administrative data examining the association of AKI and use of antivirals such as valacyclovir would not necessarily be at risk of any bias from this misclassification, but would require a very large sample size in order to offset the resulting imprecision and risk for a false-negative result. Biased misclassification: Consider the same issue with AKI as above, but for an administrative cohort study examining the association between NSAID use and AKI. Since NSAIDs are known nephrotoxins, physicians may be more prone to document the presence of AKI during hospital stay in NSAID users compared to non-users. The result would be greater magnitude of underreporting of AKI in NSAID non-users, creating biased misclassification.

Recall bias (checklist question #7)	• Form of biased misclassification that can occur in retrospective cohort and case-control studies that involve participant interview after-the-fact to inquire about previous exposures; • Occurs because individuals who develop a disease or experience an outcome tend to have better recall of potentially dangerous exposures that could have led to the outcome (because it is human nature to try to determine the cause of bad fortune).	Consider a child that develops a syndrome of unclear etiology/pathophysiology at 3 years of age. The parents of this child, searching for the cause of this mysterious illness, seeks out and becomes attuned to every exposure, from chemical to environmental, prior to development of the syndrome. On the other hand, the parent of a perfectly healthy, average 3 year-old would have no reason to obsess over and memorize these facts. • An observational study using a survey of parents to identify the association between frequency of acetaminophen use before age 3 and development of this syndrome may falsely identify such an association.

Immortal-time bias
 	
(checklist question #9)
	• Unique to cohort studies;
	• Form of biased misclassification;
	• Occurs when there is a "lag time" between beginning of follow-up & when individual becomes eligible for one of the study groups (usually the EXPOSURE group);
	  ○ During the lag time, patients in the EXPOSURE group cannot die or experience the study outcome, giving them an apparent survival advantage;
		Consider the case of the impact of statin use on development of type 2 diabetes. We know from meta-analyses of RCTs that statins modestly increase blood glucose in a dose-dependent manner, leading to an increase in diagnosis of diabetes. J Investig Med 2009;57:495-9, Lancet 2010;375:735-42, JAMA 2011;305:2556-64
		• Prior to this knowledge, however, a cohort study suggested an association between statin use and decreased progression of diabetes to insulin dependence.[Diabet Med 2004;21:962-7]
		• A subsequent re-analysis of this study, [BMJ 2010;340:b5087] however, exemplified why immortal-time bias completely explained this apparent beneficial association, due to the following:
		  1. The definition of "statin use" required ≥1 year of filling a prescription for a statin;
		  2. Any patient who filled prescriptions for a statin for ≤364 days was included in the group of "NON-users";
		  3. Any patient who started insulin within the 1st year of taking statins was included in the group of "NON-users".
		  ○ As a result of the above, many patients exposed to statins and experiencing the study outcome were misclassified as "non-users".

INTERNAL VALIDITY: Methods used to minimize confounding & bias in observational studies		
	Explanation	Issues that are minimized
Design		
High-quality data source	• Standardized, adjudicated non-differential recording of all exposure, covariables and outcome data minimizes bias; • Hierarchy of data quality: Patient records > codes from clinical database > codes from administrative database ≥ patient self-report.	All
Clear definition of exposure & outcome (or cases & controls)		• Misclassification • Immortal-time bias
Restriction	Restricting eligibility for a study with inclusion/exclusion criteria can minimize expected baseline (& subsequent treatment) differences between groups.	• Confounding • Allocation & performance bias • Attrition bias (restricted to patients with complete follow-up)
Matching	• Investigators select key characteristics likely to have a confounding effect on the association under study (generally age, sex, & 1-2 other factors such as a comorbidity score). They then match a patient in the exposure group to ≥1 non-exposure patients (or cases & controls for case-control studies), minimizing baseline differences; • May cause "overmatching" (if factor is not a confounder, decreases study power) & residual confounding (if no other methods used to minimize); • Can also match based on propensity score or instrumental variable	• Confounding • Allocation bias
Analysis		
Stratification	A type of subgroup analysis based on 1+ characteristics to determine whether the association between exposure and outcome is confounded by the subgroup characteristic.	• Confounding • Allocation bias
Multivariable regression (i.e. statistical adjustment)	Statistical adjustment that accounts for multiple covariables that could affect the outcome & confound or bias the association.	• Confounding • Allocation & performance bias
Sensitivity analyses	• Active control(s): Define control group as individuals receiving similar intervention for indication under study (e.g. different antibiotic for UTI, H2RA instead of PPI, etc) • Tracer: Repeat analysis replacing exposure with similar drug not expected to cause outcome	• Confounding • Allocation bias
Misc.		
Propensity score	• Multivariable regression is used to calculate a probability (or *propensity*) *score* (0-100%) for any individual study participant to be exposed based on their baseline characteristics • This score can either be used as a characteristic for matching, or as a covariable in regression	• Confounding (especially by indication) • Allocation bias

RESULTS: Additional caveats & considerations	
Choice of measure of association	Results of **cohort** studies can be presented as HR, OR or RR, & NNTs/NNHs can be calculated; Results of **case-control** studies CANNOT be presented as HR or RR; only ORs can be calculated. • Absolute risk and NNTs cannot generally be calculated from case-control studies because the underlying population incidence is unknown; • Exception: **Nested** case-control studies, in which cases and controls are identified from a defined patient cohort (from a cohort study or a RCT), such as the Framingham Heart Study or the GUSTO-1 RCT sample population.
Subgroup analyses in observational studies	Can be handled in the same way as in RCTs (see NERDCAT-RCT), except with *even more caution and skepticism (i.e. they're often not worth your time)*. • Exception: Stratification, which is similar to subgroup analysis, is used to minimize bias and confounding in observational studies and is acceptable as long as it is used to bolster the study's primary analysis rather than create additional conclusions.
"Trends towards statistical significance"	Don't even *think* about "trends to statistical significance" in observational studies! It is far more probable that unknown confounders account for a "trend towards benefit/harm" than the actual intervention. Apply the same logic as with RCTs in terms of looking at the confidence interval to see whether it rules out a clinically important difference.