NON-EXPERIMENTAL STUDIES ABOUT BENEFIT, HARM OR CAUSATION – cohort and cross-sectional studies examining the benefits and harm of exposures, including therapy and other interventions. (Relevant JAMA Users' Guide Number IV: references (7))

Introduction:

- The structure of a non-experimental study investigating benefit, harm or causation is very similar to a controlled trial. The occurrence of health outcomes (prevalence or incidence) is measured in subgroups defined by specific exposures and comparisons, which may be interventions in some instances.
- The major difference between experimental and non-experimental studies is that the investigator controls the allocation of the exposures (or treatment) to participants in an experiment, whereas the investigator in a non-experimental study categorises participants into exposure and comparison subgroups after measuring factors (i.e. the exposures) that the study participants are exposed to. In other words, the non-experimental study investigator observes a "natural" experiment rather than conducting one.
- The main weakness of non-experimental studies is the potential for confounding. As exposure allocation is not controlled by the investigator, it is common to find differences between exposure and comparison subgroups other than the main exposures of interest, that also influence health outcomes. These differences are known as confounding factors causing a mixing of effects. For questions about the benefits or harm of therapy, experimental studies (particularly randomised controlled trials) are usually superior to non-experimental studies because of the large potential for confounding in the latter.

Longitudinal (or cohort) studies:

- Cohort studies are basically non-experimental versions of controlled trials and are undertaken to investigate the effects (both benefits and harms) of exposures. In a cohort study, participants are recruited into the study population, exposures are measured, and then participants are followed up over time to measure outcomes.
- As mentioned above, cohort studies are not the most appropriate study design for examining the effects of interventions, because the potential for confounding is typically greatest when people are selected or self-select the exposures of interest (particularly therapies or exposures requiring a conscious decision by the participant, such as taking leisure time physical activity). Nevertheless, cohort studies are often used to

investigate the effects of therapy because of other the shortcomings of experimental studies.

- Cohort studies can often be conducted in situations where controlled trials are not possible. In some situations a trial would be unethical (e.g. investigating the adverse effects of a dangerous exposure such as electromagnetic radiation or cigarette smoking). In addition, trials are often not feasible when the effect of exposure (e.g. cigarette smoking) takes many years to cause an outcome (e.g. lung cancer) or when the outcomes of interest are uncommon (e.g. asthma death) and very large numbers of study participants are required to identify sufficient outcomes.
- Non-experimental longitudinal studies are also the most appropriate design for investigating prognosis.

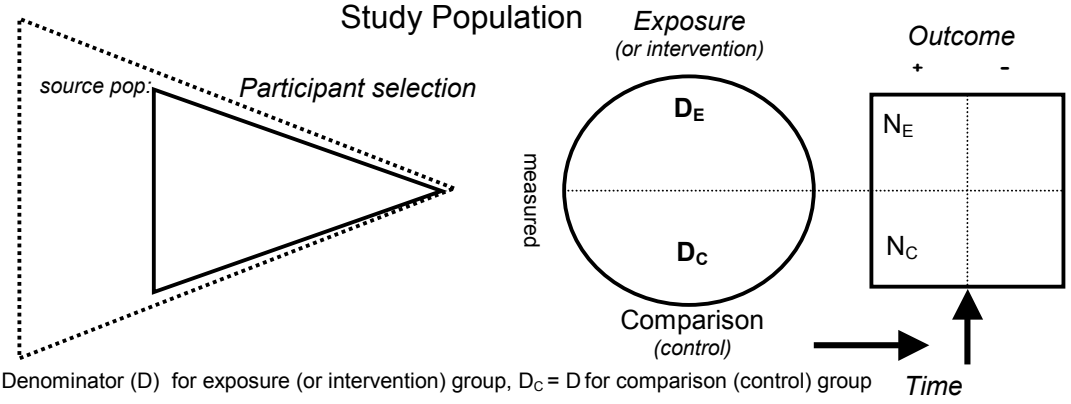
Cross-sectional studies:

- The cross-sectional study has an identical structure to the cohort study except that the exposures and outcomes are measured at the same time (i.e. cross-sectionally), whereas in a cohort study outcomes are typically measured after the exposure/s has been measured (i.e. longitudinally).
- Cross-sectional studies are the design of choice for assessing the prevalence of health-related outcomes in a target population. In such studies it is very important that the study population is representative of the target or source population of interest (i.e. the findings in the study population must be generalisable to the target population).
- Cross-sectional studies are also the design of choice for comparing diagnostic tests with a reference standard.
- Cross-sectional studies may be undertaken to investigate causal associations between exposures and outcomes, although they are not ideally suited for this purpose; especially if the outcomes of interest are acute events. As outcome measurements are made at one point in time in cross-sectional studies, many acute outcomes would be missed, particularly if they are either fatal (e.g. coronary death) or recovery occurs quickly and there are no lasting signs or symptoms of the event (e.g. asthma attacks).
- If the outcome of interest can affect the exposure of interest (e.g. a myocardial infarction may lower blood pressure), then it is not possible to validly investigate the association in a cross-sectional study, because the outcome (myocardial infarction) may be measured before the exposure (blood pressure) has been measured.
- It is therefore important to document whether the exposure was measured before the outcome occurred (i.e. check if the association is temporally correct).

- As cohort studies and most cross-sectional studies are simply longitudinal and cross-sectional versions of the same study design, they are considered together in one appraisal guide.

GATE Checklist for Cohort & Cross-sectional Studies (causation or intervention, benefit or harm)

Study author, title, publication reference	*Key 5 part study question (PECOT). Was it focussed?*



D_E = Denominator (D) for exposure (or intervention) group, D_C = D for comparison (control) group

N_E = Numerator (N) for exposure group, N_C = N for comparison group

SECTION 1: STUDY VALIDITY		Appraised by:	
Evaluation criterion (NAXS = not applicable for cross-sectional studies)		*How well was this criterion addressed?*	Quality ✓ ? x
Participants	What were the key selection (inclusion & exclusion) criteria? Were they well defined? Were they replicable?		
	Were inclusion & exclusion criteria appropriate given study question?		
Exposures & Comparison	What were the exposures (or interventions) & comparison? Well defined? Replicable?		
	Was measurement of variables similar & valid in all groups?		
	Were exposure & comparison groups similar at start of study except for study exposures?		
	If not, were differences stratified / adjusted for in analyses?		
	Were all participants analysed in groups to which initially assigned?		
	Were participants, health workers, researchers blind to exposures?		
	Apart from study exposures, were groups treated equally?		
	Were exposures remeasured during follow-up & were there important changes? (NAXS)		
Outcomes	What outcome measures were used? Well defined? Replicable?		
	How complete was follow up? Was it sufficient? How many dropouts? (NAXS)		
	Was outcome assessment blind?		
Time	Was follow up time sufficiently long to detect important effects on outcomes of interest? (NAXS)		
QUALITY OF STUDY DESIGN: *How successfully do you think the study minimised bias? Very well = +, okay = ∅, poorly = -*			

SECTION 2: STUDY RESULTS: MAGNITUDE & PRECISION OF EFFECTS	
What measures of occurrence (incidence / prevalence) & exposure effects (RR /RD /NNTs) were reported?	
What measures of precision of effects were reported (CIs, p-values)?	

THE NUMBERS TABLE: OCCURRENCE, EFFECT ESTIMATES & PRECISION

Outcomes* & Time (T)	**Exposure event rate ($EER=N_E/D_E/T$) or mean***	**Comparison event rate ($CER=N_C/D_C/T$) or mean***	**Relative Risk* (RR = EER/CER) ± (95% CI)**	**Risk difference or mean difference (RD = CER-EER) ± (95% CI)**	**Number Needed to Treat* (NNT = 1/RD) ± (95% CI)**

* *if outcomes continuous, can calculate means, mean differences, but not NNTs (don't usually calculate relative means)*
D_E = Denominator (D) for exposure (intervention) group(s), D_C = D for comparison (control) group
N_E = Numerator (N) for exposure group(s), N_C = N for comparison group

	Quality ✓ ? x
Could useful effect estimates (e.g. RR, RDs or mean differences, NNTs) be calculated? For benefits & harm?	
What was the magnitude and direction of the effect estimates?	
Was the precision of the effect estimates sufficient?	
If no statistically significant effects detected, was there sufficient power?	
QUALITY OF STUDY RESULTS: *Useful, precise +/or sufficient power? Very good = +, okay = ∅, poor = -*	

SECTION 3: STUDY APPLICABILITY			
Participants	Was the source population for participants well described?		
	Were participants representative of source population?		
	Can the relevance / similarity of the participants to a specific target group(s) be determined?		
Exposures & Comparison	Were the characteristics of the study setting well described? *e.g. rural, urban, inpatient, primary care*		
	Can the applicability/relevance of exposures be determined?		
	Can the relevance of the comparison group be determined?		
Outcomes	Were all important outcomes considered: benefits? harms? costs?		
	Are likely benefits greater than potential harms & costs (or vice versa)? In what target group(s)?		
QUALITY OF STUDY APPLICABILITY: (a) Was it possible to determine applicability? *Very well = +, okay = ∅, poorly = - (b) Are findings applicable in your practice/setting? Very well = +, okay = ∅, poorly = -*			