A New Approach to Effect Heterogeneity for Binary Outcomes

Anders Huitfeldt

(Based on working papers written in collaboration with Andrew Goldstein, Sonja Swanson, Mats Julius Stensrud and Etsuji Suzuki)

August 31st, 2017

Crocker's Rules

- I declare myself to be operating by *Crocker's Rules*
- This means that you are allowed to optimize your feedback for clarity of information, not for being nice to me.
- If I'm offended by your question or comment, it is my own fault.
- Crocker's Rules do not imply reciprocity. This is done for myself, to maximize information received.

Background

- Randomized trials are often conducted in populations that differ systematically from the populations in which the results will be used to inform clinical decisions.
- Treatment effects often differ between populations
- Several different statistical methods have been proposed to standardize findings from the experimental study population *s* to a different target population *t*
- Less attention has been given to how one should *reason about* which covariates *V* need to be standardized over

Notation and Setup

- This presentation is motivated by the following problem:
  - We have experimental evidence for the causal effect of treatment with drug A on binary outcome Y in the study population ($P = s$)
  - We wish to predict the effect of introducing the treatment in the target population ($P = t$), in which we can only collect observational data.
  - The drug is not currently available in the target population

Notation and Setup

- Because we have a randomized trial in population s, the baseline risk $\Pr(Y^{a=0} = 1|P = s)$, and the risk under treatment $\Pr(Y^{a=1} = 1|P = s)$, are identified from the data
- Since treatment is currently not available in population t, everyone in that population is currently untreated, and the baseline risk is therefore identified from the data: $\Pr(Y^{a=0} = 1|P = t) = \Pr(Y = 1|P = t)$.
- Our goal is to use this information, in combination with subject matter knowledge, to predict $\Pr(Y^{a=1} = 1|P = t)$.
  - Subject matter knowledge = Homogeneity assumption?

Approaches to Effect Homogeneity

- Any attempt to extrapolate the findings from population *s* to population *t* will depend on a belief that *something* - for example a conditional effect parameter - in population *t* is equal to the corresponding parameter in population *s*
- Our conclusions depend heavily on what parameter we assume is equal between the populations - that is, on how we operationalize effect homogeneity.

Approaches to Effect Homogeneity

The following definitions of effect homogeneity have been proposed:

- Effect Homogeneity in Measure
  - $RD_s = RD_t$
  - $RR_s = RR_t$
  - $OR_s = OR_t$
- Effect Homogeneity in Distribution
  - $Y^a \perp\!\!\!\perp P \mid V = v$ ("S-ignorability")
  - $Y^a \perp\!\!\!\perp P^a \mid V^a = v$ ("S-admissibility")
- Homogeneity of COST Parameters
  - $Y^{a=1} \perp\!\!\!\perp P \mid Y^{a=0}, V = v$

Approaches to Effect Homogeneity

- The goal of this presentation is to provide a framework for understanding what assumptions the different options for operationalizing effect homogeneity make about the underlying biology.
- This will enable investigators to reason about which set of conditions is most closely approximated in the specific context of their own study.

Outline of Presentation

- We first review the shortcomings of traditional definitions based on conditional homogeneity of effect measures
- We then discuss approaches based on effect homogeneity in distribution, with a particular emphasis on Bareinboim and Pearl's graphical models for transportability, and show how these graphs make strong assumptions that are often violated in realistic settings.

Outline of Presentation

- We then propose a new approach based on Counteractual Outcome State Transition parameters, which links the choice of effect measure to a counterfactual causal model.
- We show how these parameters can be used to encode background beliefs about the underlying biological processes.
- If the COST parameters are equal between population, there are important implications for model choice, meta-analysis and research generalization.

Part 1

Shortcomings of Standard Measures of Effect

No Biological Interpretation

- No biologically plausible model has been proposed that would guarantee (conditional) homogeneity of either the risk difference, risk ratio or odds ratio.

Logically Invalid Predictions

- The risk ratio and risk difference (but not the odds ratio) may make predictions outside the range of logically valid probabilities

Zero-bounds

- The odds ratio has a "zero bound" if the baseline incidence in the target population is 0 or 1: Regardless of the data from the trial, the investigator is doomed to conclude that treatment has no effect in population t
- The risk ratio has one such zero bound, at $T_0 = 0$.

Non-collapsibility

- The odds ratio is non-collapsible.
- In other words, the marginal value of the odds ratio may not be a weighted average of the stratum-specific odds ratios under any weighting scheme, even in the absence of confounding or other forms of structural bias.

Baseline Risk Dependence

- If the risk difference, the risk ratio or the odds ratio is equal across the populations, then the proportion of the population that responds to treatment is required to be a function of the baseline risk.

Asymmetry

- If we use a risk ratio model, our empirical predictions are not invariant to how the outcome variable is encoded in the database: The conclusions depend strongly on whether we count the living or the dead.
- This asymmetry can equivalently be conceptualized in terms of two separate risk ratio models:

$$RR(-) = \frac{\Pr(Y^{a=1} = 1)}{\Pr(Y^{a=0} = 1)}$$

$$RR(+) = \frac{1 - \Pr(Y^{a=1} = 1)}{1 - \Pr(Y^{a=0} = 1)}$$

Standard Measures of Effect

Table: Different effect measures may result in different predictions based on the same data

	RR(-)	*RR(+)*	*RD*	*OR*
Baseline risk in trial	2%	2%	2%	2%
Treated risk in trial	3%	3%	3%	3%
Effect	RR(-)=1.5	RR(+)=0.99	RD=0.01	OR=1.515
Baseline risk in target population	10%	10%	10%	10%
Predicted risk in target population	15%	10.9%	11%	14.4%

Previous suggestions

- Previous suggestions in the literature:
  - MC Sheps suggested using RR(-) if treatment reduces incidence and RR(+) if treatment increases incidence (NEJM 1956)
  - Jon Deeks provided empirical evidence for the same idea (Statistics in Medicine, 2002)
  - Glasziou and Irwig suggested considering "relative benefits and absolute harms" (BMJ, 1995)
- We agree with all these approaches (which we show are closely related).
- Yet most investigators continue to always use RR(-)
- In these papers, we provide a causal model that allows us to represent the biological background knowledge which led to these suggestions in formal counterfactual notation.

Part 2

Effect Homogeneity in Distribution

Effect Homogeneity in Distribution

- One potential response to these shortcomings might be to abandon effect measures altogether, and reason about the counterfactual distributions f($Y^{a=0}$) and f($Y^{a=1}$) separately
- For example, we can define effect homogeneity as "S-Ignorability":

$$Y^a \perp\!\!\!\perp P \mid V = v$$

- Bareinboim and Pearl's causal diagrams for transportability are an example of this approach
- This approach is elegant, complete and mathematically sophisticated

Effect Homogeneity in Distribution

- However, as with any approach that relies on effect homogeneity in distribution, the transportation diagrams rely on strong assumptions:
- Unless the investigator has accounted for all causes of the outcome *Y* that differ between the study population and the target population, the model is not an accurate approximation of the data generating mechanism.
- This differs from approaches based on effect measures, where it may be sufficient to account for all variables that are associated with the effect of *A* on *Y*.

Effect Homogeneity in Distribution

- Suppose we have data from a randomized trial on the effect of placebo vs standard of care on coronary heart disease in men, and we have concluded that there is no effect.
- Now suppose we wish to make predictions about the effects of placebo in women.
- If we believe there are causes of CHD that are differently distributed between men and women, then if we use an approach based on effect homogeneity in distribution, we are forced to conclude that we can make no predictions about the effect in women.
- In contrast, if we use an approach based on effect homogeneity in measure, we can instead try to account for all variables that are associated with the effect of placebo.

Effect Homogeneity in Distribution

- The assumption of conditional effect homogeneity in distribution is strong, testable and often empirically violated:
  - Any regression model that justifies the absence of an interaction term $beta_3 \times A \times P$ by invoking conditional effect homogeneity in distribution is known to be misspecified if the main effect of P is not equal to zero.
  - If the approaches based on the risk difference, risk ratio and odds ratio result in different predictions, we can falsify Pearl's model.
  - If the conditional baseline risk in the two populations differ, we can also falsify Pearl's model.

Effect Homogeneity in Distribution

- An approach based on effect homogeneity in distribution suggests doing meta-analysis in the control arm separately from meta-analysis in the active arm.
- This approach arguably throws away randomization(?)

Part 3

Introducing Counterfactual Outcome State Transition parameters

Basic idea

- In the following, we will link *beliefs about the relevant biology* to a *counterfactual causal model*.
- This causal model can in some settings be used to determine the choice effect measure, such that we don't have to go all the way to effect homogeneity in distribution.
- For example: If an adverse reaction associated with drug A is determined by unmeasured gene X, and the distribution of this gene is equal between two populations, then RR(+), and not the standard risk ratio RR(-), will be equal between those two populations.
- Our model formalizes the counterfactual theory that underlies this argument, thereby clarifying the scope and limits of the line of reasoning.

COST Parameters

- *Counterfactual Outcome State Transition parameters* are effect measures based on the probability of switching outcome state in response to treatment.

COST Parameters

- G is defined as the probability of being a case under treatment, among those who would have been cases under no treatment.

$$G = \Pr(Y^{a=1} = 1 | Y^{a=0} = 1)$$

- In a deterministic model, this can be interpreted as the fraction who are ‘Doomed”, among those who are either “Doomed” or “Preventative”

COST Parameters

- *H* is defined as the probability of not being a case under treatment, among those who would not have been cases under no treatment.

$$H = \Pr(Y^{a=1} = 0 | Y^{a=0} = 0)$$

- In a deterministic model, this can be interpreted as the proportion who are “Immune”, among those who are either “Immune” or “Causal”

COST Parameters

- The effect of introducing treatment in population t is said to be equal to the effect of introducing treatment in population s if and only if $G_t = G_s$ and $H_t = H_s$.
- Note that this can equivalently be written as $Y^{a=1} \perp\!\!\!\perp P \mid Y^{a=0}$, similar to the notation used by Gechter (Working paper, 2016)

COST Parameters

- This definition of effect equality resolves all major shortcomings of standard effect measures: The underlying parameters are symmetric, collapsible, have no zero constraints, do not make predictions outside valid probabilities, and are not baseline risk dependent.
- The definition does however have a major drawback: The COST parameters are not identified from the data without further assumptions

Identification of COST Parameters

- The key condition that is necessary for identification is monotonicity.
- If treatment monotonically reduces incidence, $H = 1$ whereas if treatment monotonically increases incidence, $G = 1$.
- The plausibility of the monotonicity condition varies depending on the specific scientific context. For example, it is often a plausible approximation in the case of certain side effects of drugs.

Identification of COST Parameters

- If treatment monotonically reduces the incidence of the outcome, G is identified from the data of a randomized trial and is equal to the standard risk ratio, $RR(-)$
- If treatment monotonically reduces the incidence of the outcome and the effects are equal in the sense defined in this paper, $RR(-)_s = RR(-)_t$
- If the effects are equal and treatment reduces the incidence of the outcome but not monotonically so, $RR(-)_s$ is a biased estimate of $RR(-)_t$. We prove results on the direction and magnitude of the bias, as a function of the extent of non-monotonicity and of the differences in baseline risks in the two populations.

Identification of COST Parameters

- If treatment monotonically increases the incidence of the outcome, H is identified from the data of a randomized trial and is equal to the recoded risk ratio, $RR(+)$
- If treatment monotonically increases the incidence of the outcome and the effects are equal in the sense defined in this paper, $RR(+)_s = RR(+)_t$
- If the effects are equal and treatment increases the incidence of the outcome but not monotonically so, $RR(+)_s$ is a biased estimate of $RR(+)_t$.

Asymmetry of COST Parameters

- Unfortunately, COST parameters are not symmetric to the coding of the exposure variables
- In other words, instead of using the definition $Y^{a=1} \perp\!\!\!\perp P \mid Y^{a=0}$, we could have assumed $Y^{a=0} \perp\!\!\!\perp P \mid Y^{a=1}$
- This would have led to results that are reversed from those discussed on the last slide.

Asymmetry of COST Parameters

- We will refer to the condition $Y^{a=1} \perp\!\!\!\perp P \mid Y^{a=0}$, as "Equality of the Effect of Introducing Treatment"
- Similarly, we wil refer to the condition ($Y^{a=0} \perp\!\!\!\perp P \mid Y^{a=1}$ as "Equality of the Effect of Removing Treatment"
- We next proceed to show that it is possible to reason, based on biological a priori knowledge, about which effect measure is more likely to be constant across populations.

Asymmetry of COST Parameters

- Consider a situation where the effect of treatment with A is fully explained by a variable X
- For example, if A is an antibiotic, X may be a bacterial gene
- Beliefs about these biological processes can be encoded as restrictions on the distribution of the counterfactuals $Y^{a,x}$.

Asymmetry of COST Parameters

- We will assume that A has no effect in the absence of X, that X is equally distributed in the two populations, and that X is independent of the baseline risk.
- If we further believe that X has no effect in the absence of exposure with A, but prevents the outcome in the presence of A, then we expect equality of the effect of introducing treatment.
- If we instead assume that X has no effect in the presence of A, we get equality of the effect of removing treatment.

Asymmetry of COST Parameters

- In many medical applications, such as treatment with antibiotics or adverse reactions to drugs, arguments can be made that the first type of effect equality is more likely than the second
- Evolutionary arguments also support equality of the effect of introducing the drugs over the alternative.

Empirical predictions

- If there is equality of the effects of introducing a drug, meta-analysis based on RR(-) will be more homogenuous for drugs that decrease the incidence of *Y*, whereas meta-analysis based on RR(+) will be more homogenuous for exposures that increase the incidence
- This was shown empirically by Deeks in 2002
- However, there are conceptual problems related to differential power of standard tests for heterogeneity on RR(-) and RR(+) scales.
- A new statistical test for heterogeneity may need to be developed
- In contrast to Cochran's Q and the I-statistic, this can not be based on the absolute deviations from the overall meta-analytic effect estimate.

Overview of Working Papers

- "The Choice of Effect Measure for Binary Outcomes - Introducing Counterfactual Outcome State Transition parameters" introduces COST parameters, and argues that they solve several shortcomings associated with standard effect measures.
- "Effect Heterogeneity and Variable Selection for Standardizing Experimental Findings" introduces standardization formulas for COST parameters, and describes how this approach relates to "transportation formulas" derived from Bareinboim and Pearl's selection diagrams.
- "On the Collapsibility of Causal Effect Measures" is a short report on different definitions of collapsibility, and how this relates to weights used for standardization of effect measures.

Future Work

- Extensions to time-to-event data?