UNIVERSITY OF CALIFORNIA

Los Angeles

Transporting Average Causal Effects across Observational Settings

A thesis submitted in partial satisfaction

of the requirements for the degree Master of Science

in Epidemiology

by

Warren Coons

2024

ABSTRACT OF THE THESIS

Transporting Average Causal Effects across Observational Settings

by

Warren Coons

Master of Science in Epidemiology

University of California, Los Angeles, 2024

Professor Onyebuchi A. Arah. Chair

This thesis investigates and demonstrates *observational transportability*, a causal inference approach that combines observational information from a study population and observational covariate data from a target population to generate potential causal insights in the target population. Inspired by transportability in the setting of a randomized control trial (RCT), a set of identifiability assumptions for observational transportability is provided. Then, a general formula is obtained for the average causal effect in the target population (TACE) under this framework. The concept is then used to extrapolate the average causal effect of blood lead on hypertension from a study population represented by the National Health and Nutrition Examination Survey (NHANES) to a target population represented by the Behavioral Risk Factor Surveillance System (BRFSS), thereby computing the average causal effect in the target

population. Limitations of observational transportability in generating unbiased causal effects in target populations are discussed. The study concludes by considering observational transportability in practice, including an approach enabled by extracting models published in other studies.

The thesis of Warren Coons is approved.

Beate R. Ritz

Roch A. Nianogo

Onyebuchi A. Arah, Committee Chair

University of California, Los Angeles

2024

Dedication

I dedicate this work to my family, friends, colleagues, and professors who have supported me throughout my journey. Many people have come and gone, but those who stuck around have always pushed me to be the best I can be. I am not the same person I am today without their support and their guidance, and I am forever grateful for them. While I am excited to move on to the next chapter of my life, the memories will always be with me.

Table of Contents

List of Figures

List of Tables

List of Symbols

In order of appearance:

Variables

X – a binary exposure or treatment
Y – the study outcome
Y_x – the potential outcome of Y under X = x
S – selection; S = 1 indicates the study population; S = 0 indicates the target population
W, W1, W2, W_x, $W1_x$, $W2_x$ – placeholder variables and variable sets used to illustrate counterfactual equivalence
D – a placeholder in Definition 1; a set S-admissible in distribution
D_x – a placeholder in Definition 1; a set where each variable corresponds to its counterfactual variable in D under an intervention setting X = x
T – a placeholder in Definition 2; a set S-admissible on the risk difference scale
T_x – a placeholder in Definition 2; a set where each variable corresponds to its counterfactual variable in T under an intervention setting X = x
V – a set that is S-admissible on the risk difference scale
C – a set defined such that Y_x, conditional on C ∪ V, is exchangeable over exposure X in the presence of selection (S)
C1, C2, V1, V2 – covariates used for illustrative purposes in Figures 1, 2, and 3
U – a placeholder used for defining the function g; an arbitrary set of covariates
i – an index variable referring to an individual either in the study or the target population
V_x – a set where each variable corresponds to its counterfactual variable in V under an intervention setting X = x
C_x – a set where each variable corresponds to its counterfactual variable in C under an intervention setting X = x

Functions

E – the expectation function
f_A – the joint probability density function for a given set A with respect to S
g – a stratum-specific risk difference function
P – the probability function

Assumptions

A1 through A5 delineate the necessary assumptions to calculate the average causal effect in the target population (TACE) when transporting from a randomized control trial (RCT). B1 through B5 delineate the necessary assumptions for identifying the TACE when transporting from an observational study.

List of Acronyms

In order of appearance:

RCT – randomized control trial
TACE – target average causal effect, or the average causal effect in the target population
BRFSS – Behavioral Risk Factor Surveillance System
NHANES – National Health and Nutrition Examination Survey
EMM – effect-measure modifier(s) (implied to be over the risk difference scale)
SACE – study average causal effect, or the average causal effect in the study population
ACE – average causal effect
RD_{xy} – difference of potential outcomes $Y_{x=1}$ and $Y_{x=0}$
S-DAG – selection directed acyclic graph
$TACE_{RCT}$ – TACE under the RCT transportability framework
$TACE_{OBS}$ – TACE under the observational transportability framework
CDC – Centers for Disease Control and Prevention
BMI – body mass index
p – *p*-value for a given test statistic
CI – confidence interval
N – variable indicating sample size for the study and target population demographics
SD – column variable indicating standard deviations for the study and target demographics
PL – indicator that a standard probability law was used for a particular step in a proof

Background

Scientific studies conducted in one setting are often undertaken with the premise that they will generalize to the population in which the sample is drawn or extend to other populations. However, there is always the risk that differences between the settings will render this generalization unwarranted. This paper concerns the question of transportability, a concept that specifies the assumptions necessary and means to draw causal inferences for a population of interest utilizing measurements in a separate population. By incorporating information about differences between the two populations, transportability can be wielded to make causal inferences in situations where otherwise simple extrapolations would be invalid.

Bareinboim, working with Pearl, was the first to formalize transportability mathematically [1]. Bareinboim and Pearl outlined fundamental concepts and definitions in a series of seminal papers in the field [1, 2, 3]. This paper utilizes the concepts and definitions and frames them in an epidemiological context. In epidemiology, results are generated from *study populations* (taken from a *source population*) and are extended to *target populations*. *Generalizability* refers to making inferences about a target population when the study population is a subset, while transportability applies to situations when the target population does not completely subsume the study population [4]. Transportability effectively utilizes information from a study population and a non-overlapping target population to draw inferences about the target population. The act of drawing inferences in this manner is known as *transporting*.

The method of transporting inferences from observational data is known as *observational transportability* [1]. While there has been a steadily growing body of transportability literature focused on transporting effects from randomized control trials (RCTs) [5, 6, 7], there is a

conspicuous lack of studies on observational transportability. This is an area of much-needed research, as many causal studies in epidemiology and related fields can only be done with observational data. When feasible, observational transportability could circumvent the need to conduct additional epidemiological studies to obtain desired causal quantities, providing major benefits in reducing the time and cost of research and informing policies and interventions.

This study aims to derive a methodology for transporting effect measurements from observational data under a strict set of assumptions and demonstrate how one could apply the method in a real-world scenario. First, the average causal effect in the target population (TACE) under the observational transportability framework is derived, and the necessary assumptions to estimate a causal effect without bias are clarified in terms of observable quantities from the source and target populations. The sets of measured variables needed in both the study and target populations to enable such observational transportability on the risk difference scale are outlined. Lastly, observational transportability is applied by obtaining the TACE of blood lead levels on the risk of hypertension in a cross-sectional dataset from the Behavioral Risk Factor Surveillance System (BRFSS) using data from the National Health and Nutrition Examination Survey (NHANES). The TACE is compared with the SACE, or the average causal effect in the study population. Due to the cross-sectional study design of our study population, the application does not yield true causal insights but is useful as an exercise.

Methods

Notation and Definitions

X is the exposure or intervention, and Y is the study outcome. In an experimental setting, X = 1 refers to the treated group, and X = 0 refers to the untreated group. In an observational setting, X = 1 indicates the exposed group, and X = 0 indicates the unexposed group. S represents selection; S=1 indicates the study population and S=0 indicates the target population. Conditioning on S=s in a probabilistic expression means that the other variables are measured in population s. For example, E(Y|X,S=1) designates the expectation of Y conditional on X within the study population. While the exposure is treated as a binary variable in the derivations, findings can be translated to settings where the exposure is a continuous variable. The Application section provides an example of this translation.

The average causal effect (ACE) is defined as the expectation of the contrast between $Y_{x=1}$ and $Y_{x=0}$, where Y_x is the potential outcome under the intervention X=x. Under the observational transportability framework, the TACE, the ACE of X on Y in the target population, is derived on the risk difference scale, which is equivalent to $E(Y_{x=1} - Y_{x=0}|S=0)$. For simplicity, all variables are assumed to be discrete in the derivations.

Situations where other downstream factors of X besides Y, such as mediators, are dependent on selection are ignored. As such, variables besides Y that are expressed in counterfactual values under treatment or exposure will be equivalent to their observed value. This notion of *counterfactual equivalence* can be extended to sets. For instance, if W is a set of variables W1 and W2, then W_x is defined as the set of variables in W converted to their counterfactual values under an intervention setting X = x, which can be denoted $W1_x$ and $W2_x$.

A concept that has been used extensively in the transportability literature is defined and functions as a necessary assumption:

Definition 1: S-admissibility in Distribution

Let D be a set, and D_x be a set where each variable corresponds to its counterfactual variable in D under an intervention setting X = x. D_x is S-admissible in distribution if it satisfies the relation ($Y_x \perp\!\!\!\perp S_x | D_x$), where S_x designates selection under an intervention setting X = x.

S-admissibility was a concept introduced by Bareinboim in terms of the do-calculus [1] and adapted using counterfactual language by Pearl [8], whose framing of S-admissibility is borrowed in the definition. Under counterfactual equivalence, S_x and D_x are equivalent to their observed counterparts, S and D. In this situation, D is S-admissible in distribution if ($Y_x \perp\!\!\!\perp S|D$).

The definition of S-admissibility is extended to look at a causal contrast:

Definition 2: S-admissibility on the Risk Difference Scale

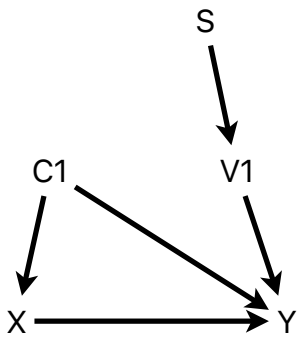
Let T be a set, and T_x be a set where each variable corresponds to its counterfactual variable in T under an intervention setting X = x. T_x is S-admissible on the risk difference scale if it satisfies the relation ($RD_{xy} \perp\!\!\!\perp S_x | T_x$), where S_x designates selection under an intervention setting X = x, and RD_{xy} is equivalent to the risk difference $Y_{x=1} - Y_{x=0}$.

Similarly, T is S-admissible on the risk difference if ($RD_{xy} \perp\!\!\!\perp S|T$) in the absence of variables affected by X other than Y that are dependent on selection. The assumption that a set T is S-admissible on the risk difference scale is mathematically weaker than the corresponding distribution assumption, as the latter guarantees the former. Only the weaker assumption is necessary for the derivations. Furthermore, a set that is S-admissible in distribution contains variables that are effect-measure modifiers (EMM) over at least one scale, such as over the risk difference scale. The subset of all EMM over the risk difference scale is S-admissible on the

corresponding scale. This follows from the fact that the non-effect modifiers can be removed from the conditioning set without modifying the risk difference.

Selection directed acyclic graphs (S-DAGs) provide a useful illustration for observational transportability by depicting non-parametric causal relationships between variables in the data-generating process. Figure 1 provides an S-DAG where C1 is a confounder, and V1 is a cause of Y that differs between the study and target population, i.e., that depends on selection. V1 provides S-admissibility in distribution and, therefore, on the risk difference scale (regardless of its effect-measure modification properties). Additionally, the union of C1 and V1 is conditionally exchangeable over exposure in the presence of selection.

Figure 1:



Selection directed acyclic graph (S-DAG) where i) X represents exposure; ii) Y represents the outcome; iii) V1 differs in distribution between the study (S=1) and target (S=0) populations; iv) C1 is a confounder.

The function f_A is used to denote the joint probability density function for a given set or union of sets A with respect to S. In addition, stratum-specific risk differences within the study population are frequently referenced in the assumptions and derivations. The stratum-specific risk difference can be denoted by the function g, where

$$g(U) = E(Y|X=1,U,S=1) - E(Y|X=0,U,S=1)$$

given a set of covariates U.

Identifiability Assumptions and Formula for Transporting from an RCT

In this section, the identifiability assumptions and formula for identifying the TACE are highlighted with the study population consisting of data from an RCT. Examining the conditions and derivation in an RCT setting provides a point of inspiration for observational transportability. The identifiability assumptions for estimating the TACE in a trial setting were previously listed out by Dahabreh [5]. They have been reworked in the context of continuous treatment.

RCT Transportability Identifiability Assumptions

Consider a set of covariates V. The necessary assumptions for identifiability are as follows:

A1: Consistency of Potential Outcomes: *If $X^i = x$, then $Y_x^i = Y^i$ for every individual i in S = 1 and S = 0.*

This assumes that individuals do not modify their behavior in the study or target population because they are aware that they are being studied (no Hawthorne effects).

A2: Conditional Exchangeability Over Treatment in the Trial: *$E(Y_x|X=x,V=v,S=1) = E(Y_x|V=v,S=1)$ for a treatment group X=x and each v with $f_V(v,S=1) > 0$.*

This is expected to hold through randomized treatment assignment.

A3: Positivity of Treatment Assignment Probability in S=1: *For a treatment group X=x,*

$$P(X = x|V=v,S=1) > 0$$

for every v with $f_V(v,S=1) > 0$.

This is also expected to hold through randomization.

A4: Conditional Exchangeability in Measure Between the Trial and the Target Population:

For the treatment groups X = 1 and X = 0, $E(Y_{x=1} - Y_{x=0}|V=v,S=0) = E(Y_{x=1} - Y_{x=0}|V=v,S=1)$ for every v with $f_V(v,S=0) > 0$.

Equivalently, V is S-admissible on the risk difference scale. V can be reduced to a set of EMM.

A5: Positivity of the Probability of Participation in the Trial: *$P(S=1|V=v) > 0$ for every v with $f_V(v,S=0) > 0$.*

Each covariate pattern within the target population should have a non-zero probability of occurring in the trial.

Assumptions A1 through A5 are necessary for identification of the TACE when the study population consists of trial data.

RCT Transportability Formula

Under A1 through A5,

$$TACE_{RCT} = \sum_V g(V)P(V|S=0)$$

The proof of the formula is in the Appendix. The result demonstrates that the following information is sufficient for the TACE to be calculable in an RCT setting:

- A list of variables that is S-admissible on the risk difference scale, denoted by V.
- Experimental data on V within the study population.

- Observational data on V within the target population.
- Treatment and outcome data in the study population.

New Result: Extending the Assumptions and Formula for Observational Transportability

Under the observational transportability framework, the identifiability assumptions and formula for the TACE on the risk difference scale are illuminated. Since the study population consists of observational data, the properties outlined by Dahabreh [5] cannot be directly taken advantage of. Instead, identifiability conditions are adjusted such that conditional exchangeability over exposure is ensured not by design but through a conditioning set that deconfounds the association of X and Y in the presence of selection.

Observational Transportability Identifiability Assumptions

Consider non-overlapping sets C and V. The necessary assumptions for identifiability are as follows:

B1: Consistency of Potential Outcomes: *If $X^i = x$, then $Y_x^i = Y^i$ for every individual i in S = 1 and S = 0.*

This assumes treatment or exposure variation irrelevance, i.e., well-defined exposures that can map onto conceivable interventions with unique potential outcomes.

B2: Conditional Exchangeability Over Exposure in the Observational Study Population: *$E(Y_x|X=x,C=c,V=v,S=1) = E(Y_x|C=c,V=v,S=1)$ for every combination of c and v with $f_{C,V}(c,v,S=1) > 0$.*

This holds if the union of C and V is sufficient for confounding control in the presence of selection.

B3: Positivity of Exposure Probability in the Observational Study Population: *For an exposure X=x,*

$$P(X = x|C=c,V=v,S=1) > 0$$

for every combination of c and v with $f_{C,V}(c,v,S=1) > 0$.

There is a non-zero probability that X is in either the exposed or unexposed group for each combination of c and v in S=1.

B4: Conditional Exchangeability over Selection in Measure Between the Observational Study Population and the Target Population: *For the exposed group X = 1 and the unexposed group X = 0,* $E(Y_{x=1} - Y_{x=0}|V=v,S=0) = E(Y_{x=1} - Y_{x=0}|V=v,S=1)$ *for each v with* $f_V(v,S=0) > 0$.

Equivalently, V is S-admissible on the risk difference scale. V can be reduced to a set of EMM.

B5: Positivity of the Probability of Participation in the Observational Study Population: $P(S=1|V=v) > 0$ *for each v with* $f_V(v,S=0) > 0$.

Each covariate pattern within the target population has a non-zero probability of occurring in the study population.

Assumptions B1 through B5 are necessary for the identification of the TACE when the study population consists of observational data.

Observational Transportability Formula

Under B1 through B5,

$$TACE_{OBS} = \sum_{C,V} g(C \cup V)P(C|V,S=1)P(V|S=0)$$

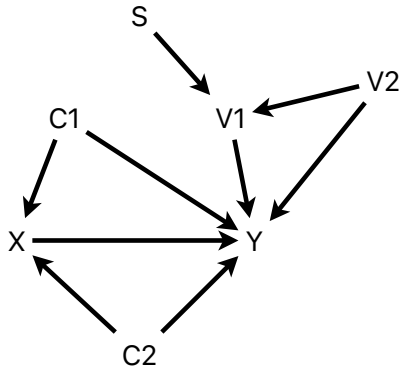
The proof of the formula is in the Appendix. The result demonstrates that the following information is sufficient for the TACE to be calculable:

- A list of variables that is S-admissible on the risk difference scale, denoted by V.
- A list of variables that is conditionally exchangeable over exposure in union with V and in the presence of selection, denoted by C.
- Observational data on C and V within the study population.
- Observational data on V within the target population.
- Exposure and outcome data in the study population.

Illustration

To ingrain the concepts of observational transportability, the variable selection procedure and TACE calculation is presented under a hypothetical scenario. Consider the data-generating process encapsulated by the S-DAG in Figure 2. Suppose one is interested in utilizing the observational transportability formula to calculate $TACE_{OBS}$. They can do so by combining information leveraged from the S-DAG and appropriate data.

Figure 2:



S-DAG. X is a cause of Y. The union of V1 and V2 differs in distribution. C1 and C2 are confounders.

To generate $TACE_{OBS}$, they first need to guarantee S-admissibility on the risk difference scale. The union of V1 and V2 is S-admissible in distribution. They are limited to non-graphical means for assessing effect-measure modification. Suppose that they went through a non-graphical approach to determine that V2 is an EMM but V1 is not. With this information, V2 is S-admissible on the risk difference scale. For the next step, V2 needs to be extended by a set of variables to ensure conditionally exchangeability over exposure in the presence of selection. The union of C1 and C2 is sufficient. Given exposure and outcome data in the study population, the final check for identification is observational data on C1, C2, and V2 in the study population and V2 in the target population. If fulfilled, the TACE is obtainable and represented by the following expression:

$$TACE_{OBS} = \sum_{C1,C2,V2} g(C1 \cup C2 \cup V2)P(C1,C2|V2,S=1)P(V2|S=0).$$

This assumes no positivity or counterfactual consistency violations (assumptions B1, B3, and B5).

Application

This section aims to provide an illustrative example of observational transportability using real-world data. Considering the case of blood lead levels and hypertension, cross-sectional data from NHANES is used to estimate an SACE. Then, the formula for observational transportability is applied to obtain the TACE in BRFSS where data is only needed for some EMM. Lastly, a comparison is made of the SACE and the TACE in addition to their confidence intervals. Since the study population consists of cross-sectional data, results may be influenced by reverse causality.

Humans are exposed to lead through a variety of sources, including lead-based paint, contaminated soil, and drinking water [9]. Lead has also been shown to have a detrimental impact on human health. According to the Mayo Clinic, lead poisoning has been linked to high blood pressure, joint and muscle pain, difficulties with memory or concentration, and other adverse symptoms [10]. As of 2022, researchers estimate that more than 170 million people in the United States have been exposed to harmful levels of lead in their early childhood [11]. The connection of lead exposure to hypertension has been a subject of debate for years.

Study Population

Study population data was collected from the 2011-2012 National Health and Nutrition Examination Survey (NHANES) [12] conducted by the Centers for Disease Control and Prevention (CDC). NHANES is a nationally representative cohort that utilizes questionnaires, laboratory samples, and other physical examinations to assess the health of adults and children in the United States. The original data for this study consisted of 8 NHANES cross-sectional datasets [13, 14, 15, 16, 17, 18, 19, 20] collected from 2011 to 2012 that were merged by a unique subject identifier, yielding 8,956 subjects in total. Patients who were younger than 20 years old or had missing values for blood lead level, diastolic blood pressure, systolic blood pressure, age, body mass index (BMI), sex, race and ethnicity, education, smoking status, alcohol consumption, sleep duration, or reported days of poor mental health were excluded, leaving 2,823 subjects in the final analytic sample. Participants with missing data included those whose lead content was below the detection level (<0.25 μg/dL) and those who did not know or refused to provide information on smoking status, alcohol consumption, average sleep duration, or days of poor mental health.

Exposure, Outcome, and Confounder Assessment (NHANES)

Blood lead measurements were taken by NHANES from blood samples procured in a laboratory. Diastolic and systolic blood pressure measurements were taken from mobile examination centers. Following the American Heart Association's definition of hypertension as of 2023 [21], participants with a systolic blood pressure reading of 130mmHg or above or a diastolic blood pressure of 80mmHg or above were defined as having hypertension.

Age (continuous), sex (binary), race and ethnicity (categorical), educational attainment (categorical), BMI (continuous), smoking behavior (categorical), and alcohol use (continuous) were determined through multiple NHANES questionnaires as confounders in accordance with an existing cross-sectional study on blood lead and blood pressure in NHANES [22]. We assume that these confounders constitute a set that is sufficient for conditional exchangeability over exposure in the presence of selection. The list may not be exhaustive. The categories under race and ethnicity were "Mexican American", "Other Hispanic", "Non-Hispanic White", "Non-Hispanic Black", and "Other Race". Education was classified by highest level of attainment, with three categories: "< High School" for those that did not graduate high school, "High School" for those that graduated but did not obtain a postsecondary degree, and "> High School" for those that obtained a postsecondary degree.

For smoking behavior, participants were categorized as current smokers if they reported smoking at least 100 cigarettes in their life and additionally reported that they still smoke "every day" or "some days", former smokers if they reported smoking at least 100 cigarettes in their lifetime and additionally reported currently smoking "not at all", and never smokers if they reported smoking fewer than 100 cigarettes in their lifetime. For alcohol consumption,

participants were asked how many drinks they had on an average drinking day during the last 12 months.

Target Population

Target population data was acquired from the 2022 Behavioral Risk Factor Surveillance System (BRFSS) conducted by the CDC [23]. The data consisted of a single file with 445,132 participants. Participants younger than 20 years old and those with missing or unknown values for the final choice of EMM – sleep duration and stress – were filtered out. After applying this exclusion criteria, 164,801 participants were available for analysis.

Effect-Measure Modification Assessment (NHANES and BRFSS)

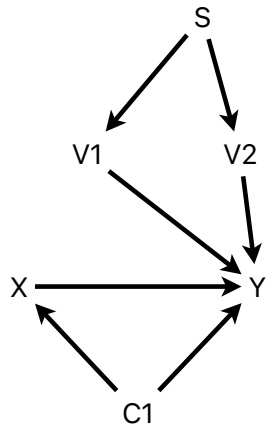
The search for EMM available for analysis not already identified as confounders was informed by a study on the modifying effect of stress and a few other articles on the general risk factors for hypertension [24, 25]. Although there were several candidates like physical activity and salt consumption [26], many potential EMM were not measured in both datasets under a universal standard. Among those that were equivalently measured in both populations and available for analysis were average sleep duration per night, measured in hours, and a question with identical wording that asked for the number of days that a participant's mental health was "not good" in the past 30 days. In NHANES, sleep duration was preemptively capped at 12 hours.

Effect measure modification was assessed over the risk difference scale by fitting separate linear regression models for days of poor mental health (or stress) and sleep duration in the study population, adjusting for the exposure and confounders, and obtaining a *p*-value of the Wald statistic corresponding to the interaction term. Utilizing an alpha value of 0.05, there was not significant evidence against the null hypothesis of no effect measure modification for days of

poor mental health ($p = 0.4371$) and sleep duration ($p = 0.4398$) in the study population. After applying the respective exclusion criteria for the target population, a Kolmogorov-Smirnov test was performed for each potential EMM to determine if there is evidence for heterogeneity in distribution between the study and target populations. There was significant evidence of heterogeneity in distribution for days of poor mental health ($p < 2.2e\text{-}16$) but not sleep duration ($p = 0.1546$).

To summarize, there was not quality evidence of effect-measure modification due to stress or sleep duration, and only one of the covariates varied in distribution. However, for the purpose of illustration, both variables were treated as EMM that form an S-admissible set on the risk difference scale. Figure 3 depicts the hypothetical data-generating process of blood lead and hypertension based on the variable definitions.

Figure 3:



S-DAG of the hypothesized data-generating process for blood lead (X) on hypertension risk (Y). C1 designates a covariate in a set C of all proposed confounders: age, sex, race and ethnicity, educational attainment, BMI, smoking behavior, and alcohol use (only one is shown to prevent cluttering; they all share similar causal relations). V1 (stress) and V2 (sleep duration) are treated as EMM that vary in distribution. Accordingly, they are dependent on selection (S) and are risk factors for Y. Assuming an accurate data-generating process, V = {V1, V2} is sufficient

for S-admissibility on the risk difference scale. Moreover, the union of C and V is sufficient for confounding control in the presence of selection.

Statistical Methods

The true SACE and TACE and their 95% confidence intervals (CI) were estimated using a repeated multi-step process involving bootstrapping. In total, 1,000 bootstrapped datasets were created from the study population data. Each dataset yielded a sample SACE and a sample TACE. The following procedure was repeated for each bootstrapped dataset:

- Calculating the SACE: A logistic regression model was learned on the exposure, confounders, differing EMM (sleep and days of poor mental health), and outcome within the dataset, including variables representing interaction with the exposure for each EMM in the regression. Counterfactual data was simulated by duplicating the data frame, increasing the value of the exposure by 1, updating the interaction variables, and creating predictions with this new dataset. The SACE was calculated by taking the mean difference of the predicted and observed values of the outcome. The estimate was then recorded in a list of all previously computed SACEs.
- Augmenting the target population: Dummy data on the exposure and covariates were generated in the target population by repeatedly copying the study population data into a new data frame until the number of rows surpassed that of the original target population. Subsequently, the resulting data frame was trimmed to align with the row count of the target population. Dummy data for sleep and days of poor mental health was replaced by EMM data in the target population. Variables measuring interaction were then updated as a product of the true EMM data and the fake exposure data.
- Calculating the TACE: Predictions were made from the model on the resulting data frame. Counterfactual data was generated under a similar process as before in the target

data. The TACE was obtained by taking the mean difference of the predicted values before and after intervening on X in the target data. This TACE was recorded in a list of all previously computed TACEs.

After the last repetition, the average of all SACEs and a 95% CI were obtained using the 2.5th and 97.5th quantiles of the 1,000 iterations. An average and 95% CI was additionally obtained for the TACEs. These averages are treated as the final values of the SACE and TACE. Analyses were conducted using R 4.3.0 and RStudio Version 2023.09.1+494.

Results

Table 1 illustrates the means and proportions for the exposure, confounders, and potential EMM, and outcome in the study and target populations after applying the respective exclusion criteria. A mean lead level of 1.55 μg/dL was obtained in NHANES and 41.4% of the participants in the study population exhibited hypertension. There was no information on exposure or outcome data in BRFSS. NHANES participants leaned heavily towards being male, while BRFSS participants were mostly female. BRFSS participants were older and demonstrated higher levels of educational attainment. BMI distributions did not noticeably vary. BRFSS participants were less likely to be current smokers and more likely to be never smokers in comparison to NHANES. Participants in NHANES had significantly fewer days of poor mental health than those in BRFSS. Both populations exhibited a similar pattern in sleep duration. Summary statistics on alcohol consumption and race distribution could not be assessed in identical fashion for the target population and are therefore missing from the table.

Table 1: Study and target population demographics[a,b]

Variable	Study Population (NHANES) (n = 2,823)	Target Population (BRFSS) (n = 164,801)
Exposure		

Lead Level (μg/dL)	1.55 ± 2.04	-
Confounders		
Age (yrs)	46.04 ± 17.63	49.73 ± 17.25*
BMI (kg/m²)	28.64 ± 6.73	29.05 ± 7.23
Alcoholic Drinks per Day	2.81 ± 2.99	-**
Sex		
Male	55.6	39.8
Female	44.4	60.2
Race and Ethnicity		
Mexican American	8.6	-
Other Hispanic	9.5	-
Non-Hispanic White	41.4	-
Non-Hispanic Black	24.8	-
Other Race	14.4	-
Highest Education		
< High School	18.1	5.3
High School	19.4	22.5
> High School	62.5	72.2
Smoking Status		
Current	23.3	15.5
Former	23.7	27.1
Never	53.0	57.5
Potential EMM (Non-confounders)		
Days of Poor Mental Health	3.97 ± 7.78	11.10 ± 10.15
Sleep Duration (hours)	6.79 ± 1.34	6.82 ± 1.64
Outcome		
Has Hypertension		
Yes	41.4	-
No	58.6	-

[a]Continuous data is presented as a mean ± standard deviation, and categorical variables are measured in terms of percentages.
[b]- indicates that a measurement could not be taken for the target population.
*An imputation procedure was used for age in BRFSS
**There is a similar question in BRFSS for alcohol consumption, but it includes days where a person did not drink in the measured average.

Means and proportions of the study demographics in NHANES and BRFSS. In the analysis, days of poor mental health and sleep duration are treated as a set of EMM that vary in distribution between NHANES and BRFSS.

Table 2: SACE and TACE

Effect Measure	**NHANES (95% CI)**	**BRFSS (95% CI)**
Average Causal Effect (ACE)	0.0073 (-0.0131, 0.0289)	0.0096 (-0.0057, 0.0274)

Final estimates of the SACE and TACE obtained by averaging samples generated by 1,000 bootstrapped datasets.

As illustrated in **Table 2**, an SACE of 0.0073 was obtained, meaning that a one μg/dL increase in blood lead exposure was associated with a 0.73% increase in hypertension risk in the study population. The 95% confidence interval for this estimate ranged from -0.0131 to 0.0289. A TACE of 0.0096 was obtained, indicating a 0.96% increase in hypertension risk per one μg/dL increase in blood lead exposure in the target population. The 95% CI for this estimate ranged from -0.0057 to 0.0274. The 95% CIs contained 0, suggesting that blood lead content was not associated with a statistically significant increase in risk of hypertension in the study and target populations.

Discussion

This study set out to derive and illustrate a methodology for utilizing information from an observational dataset to obtain an ACE for a target population of interest. The identifiability assumptions were stated, and a formula was obtained for the TACE when the study population corresponds to an RCT. From there, a formula was derived for the TACE when the study population is an observational dataset. Unlike the RCT case, the derivation for observational transportability relied on identifying a set sufficient for confounding control as conditional exchangeability over exposure could not be assumed by design.

The methodology was applied in a real-world setting, utilizing cross-sectional NHANES data on blood lead and hypertension to estimate the TACE in BRFSS, a telephone survey. The literature on blood lead and hypertension was consulted to create a list of measurable confounders in the study population and a set of measurable potential EMM in both populations. Weak evidence of effect measure modification was found for days of poor mental health and sleep duration, and strong evidence of heterogeneity in distribution was found for days of poor mental health. Regardless, both variables were treated as EMM that differ in distribution, necessitating their measurement in both the study and target populations. Finally, the SACE and TACE were assessed from 1,000 bootstrapped datasets and averages were taken. A final SACE of 0.0073 and a final TACE of 0.0096 were obtained. Blood lead was not associated with a statistically significant increased risk of hypertension in either dataset.

The ability to generate causal insights from observational transportability may be limited by poor-quality study data. In an observational study, high-quality data is relied upon, meaning that the exposure, confounders, and outcome are adequately measured and with minimal

selection bias or uncontrolled confounding. The same standard of high-quality data is required when transporting causal estimates from an observational setting. In addition, causal estimates obtained from cross-sectional datasets are generally unreliable as a temporal relationship between the exposure and outcome cannot be established. More work needs to be done to demonstrate how observational transportability can be utilized in a longitudinal setting where the TACE can more reliably be interpreted causally.

Estimating an ACE through observational transportability poses some additional challenges that are distinct from those found in a typical observational study. Unlike in a typical observational study, one is met with the task of identifying a set that is S-admissible on the risk difference scale. As such, knowledge of the data-generating process must include differences in the distribution of the covariates between the study and target population and how these variables relate to the outcome. Variables in both populations must be measured by a universal standard such that heterogeneity in distribution is not a product of variations in measurement but actual differences between the two populations.

Moreover, the nature of the data-generating process itself may pose problems for estimating the TACE under observational transportability. A presupposition this study makes is the absence of downstream factors of the exposure besides Y that depend on selection. Therefore, the application of observation transportability under this study's framework is potentially limited to circumstances where other effects of X, such as mediators, do not drive selection. In addition, not all settings are amenable to transportability, regardless of the study design of the source population. One example of this is when there are no measured intermediate variables between the outcome and selection. Estimates are not transportable in this scenario [2].

Despite its limitations, observational transportability can be a powerful tool for causal inference. Observational transportability can be useful in certain situations where traditional observational analysis fails to generate an ACE that can be interpreted causally. Standard methods for observational analysis are prone to bias when data on exposure, outcome, or confounder data in the population of interest is incomplete or mismeasured and when only cross-sectional data is collected. By incorporating information on effect modifiers that differ between the study and the target population, observational transportability grants the freedom to generate causal insights from data that is less than ideal. The concept allows new populations to be studied in detail, including those that contribute significantly to the exposed population or carry a major portion of the disease burden. The importance of high-quality study population data when transporting is emphasized; causal estimates generated from poor-quality study data are meaningless.

This study opens an alternative approach for estimating the TACE of any population that does not require performing additional observational studies. Suppose one is interested in measuring the TACE of an exposure-outcome relationship of interest. An unbiased estimate for the TACE is feasible with i) a well-defined model that controls for confounding and incorporates all EMM in a set that is S-admissible in distribution and ii) target population data of all EMM specified in the model that differ in distribution. While this is a promising new approach for epidemiological research, the standard format poses some barriers to applying observational transportability in practice. For one, most epidemiological studies do not provide a record of the formulas they used in their analyses. In addition, non-confounding EMM are justifiably ignored in models when the primary goal is to estimate a risk difference. Researchers in the field should

move towards publishing their models as a general practice with an eye towards reducing the time and cost of population-level research through observational transportability.

Appendix

RCT Transportability Formula Proof

$TACE_{RCT} = E(Y_{x=1} - Y_{x=0}|S=0)$

$= \sum_{Vx} E(Y_{x=1} - Y_{x=0}|V_x,S=0)P(V_x|S=0)$ (PL)

$= \sum_{V} E(Y_{x=1} - Y_{x=0}|V,S=0)P(V|S=0)$ (Counterfactual equivalence)

$= \sum_{V} E(Y_{x=1} - Y_{x=0}|V,S=1)P(V|S=0)$ (A4 and A5)

$= \sum_{V} E(Y_{x=1} - Y_{x=0}|X,V,S=1)P(V|S=0)$ (A2 and A3)

$= \sum_{V}[E(Y_{x=1}|X=1,V,S=1) - E(Y_{x=0}|X=0,V,S=1)]P(V|S=0)$

$= \sum_{V}[E(Y|X=1,V,S=1) - E(Y|X=0,V,S=1)]P(V|S=0)$ (A1)

$= \sum_{V} g(V)P(V|S=0)$

Observational Transportability Formula Proof

$TACE_{OBS} = E(Y_{x=1} - Y_{x=0}|S=0)$

$= \sum_{Vx} E(Y_{x=1} - Y_{x=0}|V_X,S=0)P(V_X|S=0)$ (PL)

$= \sum_{V} E(Y_{x=1} - Y_{x=0}|V,S=0)P(V|S=0)$ (Counterfactual equivalence)

$= \sum_{V} E(Y_{x=1} - Y_{x=0}|V,S=1)P(V|S=0)$ (B4 and B5)

$= \sum_{Cx,V} E(Y_{x=1} - Y_{x=0}|C_X,V,S=1)P(C_X|V,S=1)P(V|S=0)$

$= \sum_{C,V} E(Y_{x=1} - Y_{x=0}|C,V,S=1)P(C|V,S=1)P(V|S=0)$

$= \sum_{C,V}[E(Y_{x=1}|C,V,S=1) - E(Y_{x=0}|C,V,S=1)]P(C|V,S=1)P(V|S=0)$

$= \sum_{C,V}[E(Y_{x=1}|X=1,C,V,S=1) - E(Y_{x=0}|X=0,C,V,S=1)]P(C|V,S=1)P(V|S=0)$ (B2 and B3)

$= \sum_{C,V}[E(Y|X=1,C,V,S=1) - E(Y|X=0,C,V,S=1)]P(C|V,S=1)P(V|S=0)$ (B1)

$= \sum_{C,V} g(C \cup V)P(C|V,S=1)P(V|S=0)$

Discussion

[1] Pearl, J., & Bareinboim, E. (2011). Transportability of causal and statistical relations: A formal approach. *Proceedings of the AAAI Conference on Artificial Intelligence*, *25*(1), 247–254. https://doi.org/10.1609/aaai.v25i1.7861

[2] Bareinboim, E., & Pearl, J. (2013). A general algorithm for deciding transportability of experimental results. *Journal of Causal Inference*, *1*(1), 107–134. https://doi.org/10.1515/jci-2012-0004

[3] Pearl, J., & Bareinboim, E. (2014). External validity: From do-calculus to transportability across populations. *Statistical Science*, *29*(4). https://doi.org/10.1214/14-sts486

[4] Degtiar, I., & Rose, S. (2023). A review of Generalizability and transportability. *Annual Review of Statistics and Its Application*, *10*(1), 501–524. https://doi.org/10.1146/annurev-statistics-042522-103837

[5] Dahabreh, I. J., Petito, L. C., Robertson, S. E., Hernán, M. A., & Steingrimsson, J. A. (2020). Toward causally interpretable meta-analysis. *Epidemiology*, *31*(3), 334–344. https://doi.org/10.1097/ede.0000000000001177

[6] Mehrotra, M. L., Westreich, D., Glymour, M. M., Geng, E., & Glidden, D. V. (2021). Transporting subgroup analyses of randomized controlled trials for planning implementation of new interventions. *American Journal of Epidemiology*, *190*(8), 1671–1680. https://doi.org/10.1093/aje/kwab045

[7] Inoue, K., Hsu, W., Arah, O. A., Prosper, A. E., Aberle, D. R., & Bui, A. A. T. (2021). Generalizability and transportability of the National Lung Screening Trial Data: Extending trial results to different populations. *Cancer Epidemiology, Biomarkers & Prevention*, *30*(12), 2227–2234. https://doi.org/10.1158/1055-9965.epi-21-0585

[8] Pearl, J. (2015). Generalizing experimental findings. *Journal of Causal Inference*, *3*(2), 259–266. https://doi.org/10.1515/jci-2015-0025

[9] Sciences--EPH-PHS--4300, E. P. H. (n.d.). *Common sources of lead poisoning*. Washington State Department of Health. https://doh.wa.gov/community-and-environment/contaminants/lead/common-sources-lead-poisoning

[10] Mayo Foundation for Medical Education and Research. (2022, January 21). *Lead poisoning*. Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/lead-poisoning/symptoms-causes/syc-20354717

[11] McFarland, M. J., Hauer, M. E., & Reuben, A. (2022). Half of US population exposed to adverse lead levels in early childhood. *Proceedings of the National Academy of Sciences*, *119*(11). https://doi.org/10.1073/pnas.2118631119

[12] Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Data. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2012, https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?BeginYear=2011

[13] Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Questionnaire (or Examination Protocol, or Laboratory Protocol). Hyattsville, MD: U.S. Department of Health and Human Services, Centers

for Disease Control and Prevention, 2012. https://wwwn.cdc.gov/Nchs/Nhanes/2011-2012/PBCD_G.htm#LBXBPB
[14] Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Examination Protocol. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2012. https://wwwn.cdc.gov/Nchs/Nhanes/2011-2012/BPX_G.htm
[15] Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Questionnaire. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2012. https://wwwn.cdc.gov/Nchs/Nhanes/2011-2012/DEMO_G.htm
[16] Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Examination Protocol. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2012. https://wwwn.cdc.gov/Nchs/Nhanes/2011-2012/BMX_G.htm#BMXBMI
[17] Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Questionnaire. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2012. https://wwwn.cdc.gov/Nchs/Nhanes/2011-2012/SMQ_G.htm
[18] Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Questionnaire. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2012. https://wwwn.cdc.gov/Nchs/Nhanes/2011-2012/ALQ_G.htm#ALQ130
[19] Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Questionnaire. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2012. https://wwwn.cdc.gov/Nchs/Nhanes/2011-2012/HSQ_G.htm#HSQ480
[20] Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Questionnaire. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2012. https://wwwn.cdc.gov/Nchs/Nhanes/2011-2012/SLQ_G.htm#SLD010H
[21] *The facts about high blood pressure*. www.heart.org. (2023, May 25). https://www.heart.org/en/health-topics/high-blood-pressure/the-facts-about-high-blood-pressure
[22] Huang, Z. (2022). Association between blood lead level with high blood pressure in US (Nhanes 1999–2018). *Frontiers in Public Health*, *10*. https://doi.org/10.3389/fpubh.2022.836357
[23] United States Department of Health and Human Services. Centers for Disease Control and Prevention. Behavioral Risk Factor Surveillance System (BRFSS), 2022. https://www.cdc.gov/brfss/annual_data/annual_2022.html
[24] Peters, J. L., Kubzansky, L., McNeely, E., Schwartz, J., Spiro, A., Sparrow, D., Wright, R. O., Nie, H., & Hu, H. (2007). Stress as a potential modifier of the impact of lead levels on blood pressure: The Normative Aging Study. *Environmental Health Perspectives*, *115*(8), 1154–1159. https://doi.org/10.1289/ehp.10002
[25] Li, C., & Shang, S. (2021). Relationship between sleep and hypertension: Findings from the NHANES (2007–2014). *International Journal of Environmental Research and Public Health*, *18*(15), 7867. https://doi.org/10.3390/ijerph18157867

[26] Meher, M., Pradhan, S., & Pradhan, S. R. (2023). Risk factors associated with hypertension in young adults: A systematic review. *Cureus*. https://doi.org/10.7759/cureus.37467