Proposta TFM MESIO

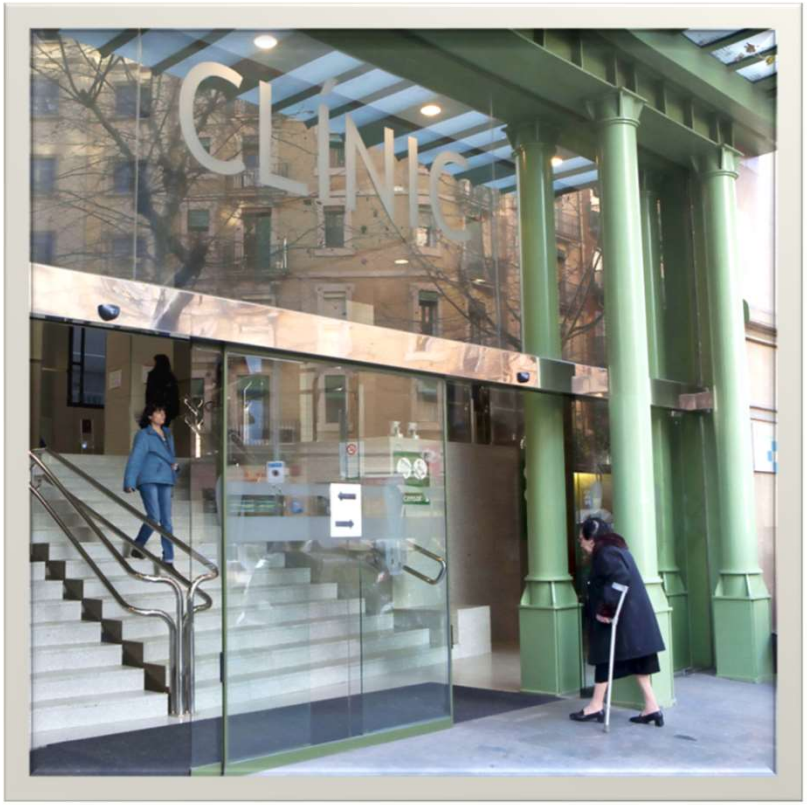
Roger Borràs
Maig 2023

Introduction

Qui sóm?

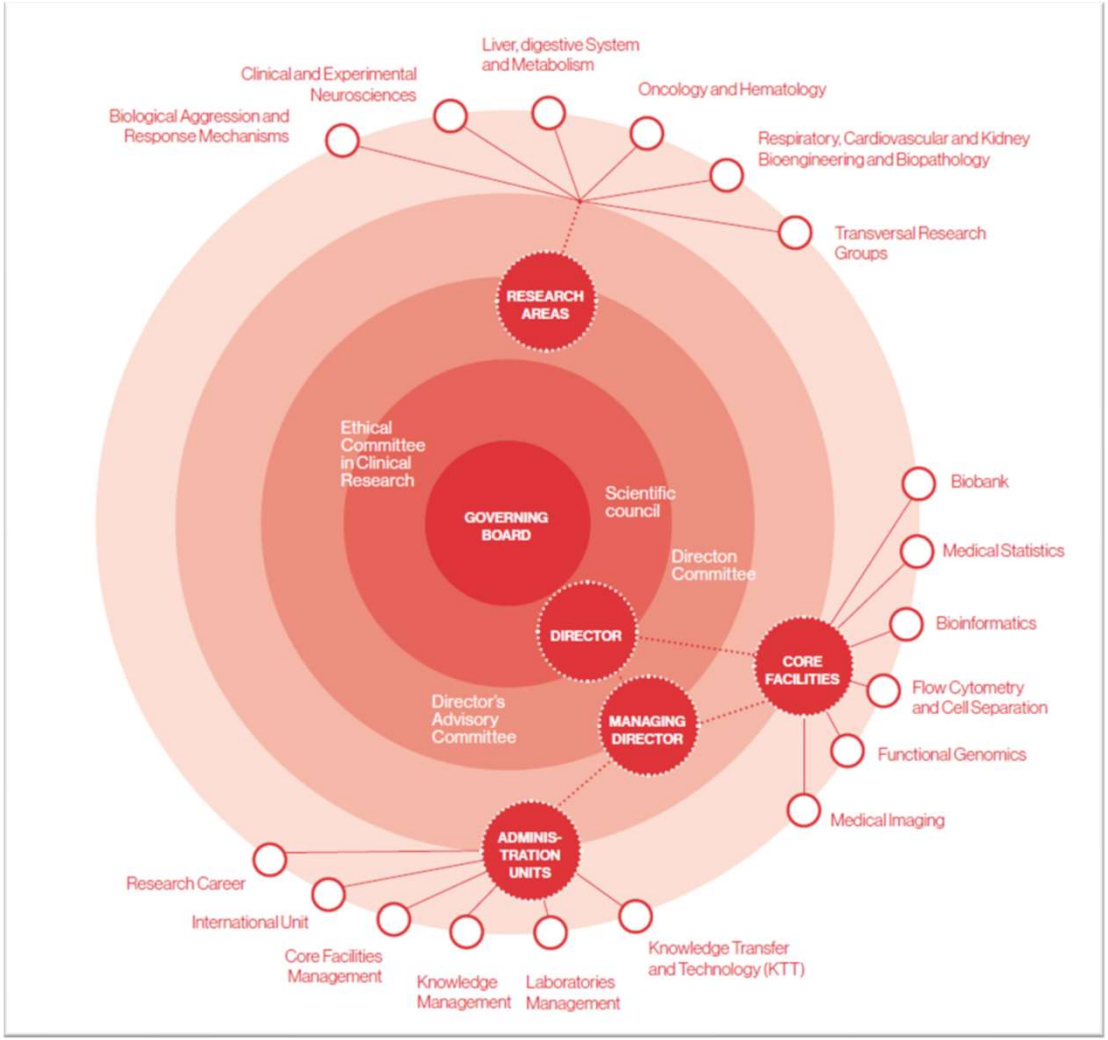

IDIBAPS
FUNDACIÓ CLÍNIC BARCELONA
Assistència
Recerca
Docència i Formació
Coneix-nos
Inici › Recerca
On comença la medicina del demà
Fem recerca biomèdica d'excel·lència per millorar la salut de les persones


SALUT I INNOVACIÓ BASADA EN 4 PILARS
Assistència
Recerca
Docència
Organització


CLINIC

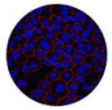
Clínic Barcelona | Universitat de Barcelona

IDIBAPS



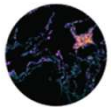
Clínic Barcelona | Universitat de Barcelona

IDIBAPS



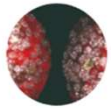
Biological aggression and response mechanisms

More information



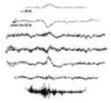
Respiratory, cardiovascular and renal pathobiology and bioengineering

More information



Liver, digestive system and metabolism

More information



Clinical and experimental neuroscience

More information

Oncology and haematology

More information

Transversal research groups

More information

Publication output and indicators.

	OUTPUT	MEAN IF	%Q1	%D1	%OA
Original articles	1,815	7.7	65%	29%	70%
Reviews	362	8.2	64%	23%	72%
Editorials	72	8.2	60%	33%	61%
Clinical guidelines	72	8.3	46%	26%	79%
Others: Letters & Case reports	262	8.7	55%	36%	60%
Consortium publications	125	14.5	77%	49%	85%
TOTAL	2,708	8.2	64%	30%	69%

Recerca

Àrea 1
Agressió biològica i mecanismes de resposta

Àrea 2
Biopatologia i bioenginyeria respiratòria, càrdio-vascular i renal

- Aterosclerosi i malaltia coronària
- Aritmies, Resincronització i Imatge Cardíaques
- Malalties nefro-urològiques i Trasplant Renal
- Risc cardiovascular, nutrició i envelliment
- Biofísica respiratòria i bioenginyeria
- Investigació aplicada en malalties respiratòries infeccioses, malalt crític i càncer de pulmó
- Mecanismes fisiopatològics de les malalties respiratòries
- Immunoal·lèrgia Respiratòria Clínica i Experimental (IRCE)

Àrea 2

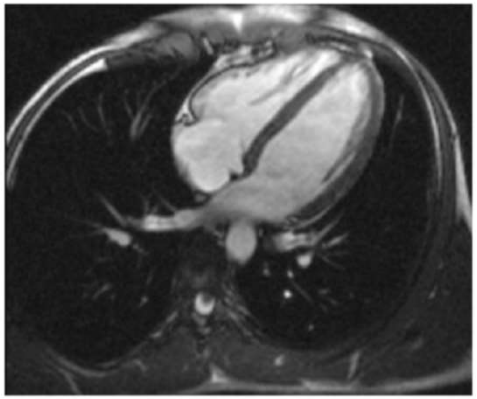
Aritmies, Resincronització i Imatge Cardíaques

Informació de l'equip | **Membres de l'equip** | Producció científica | Projectes de recerca

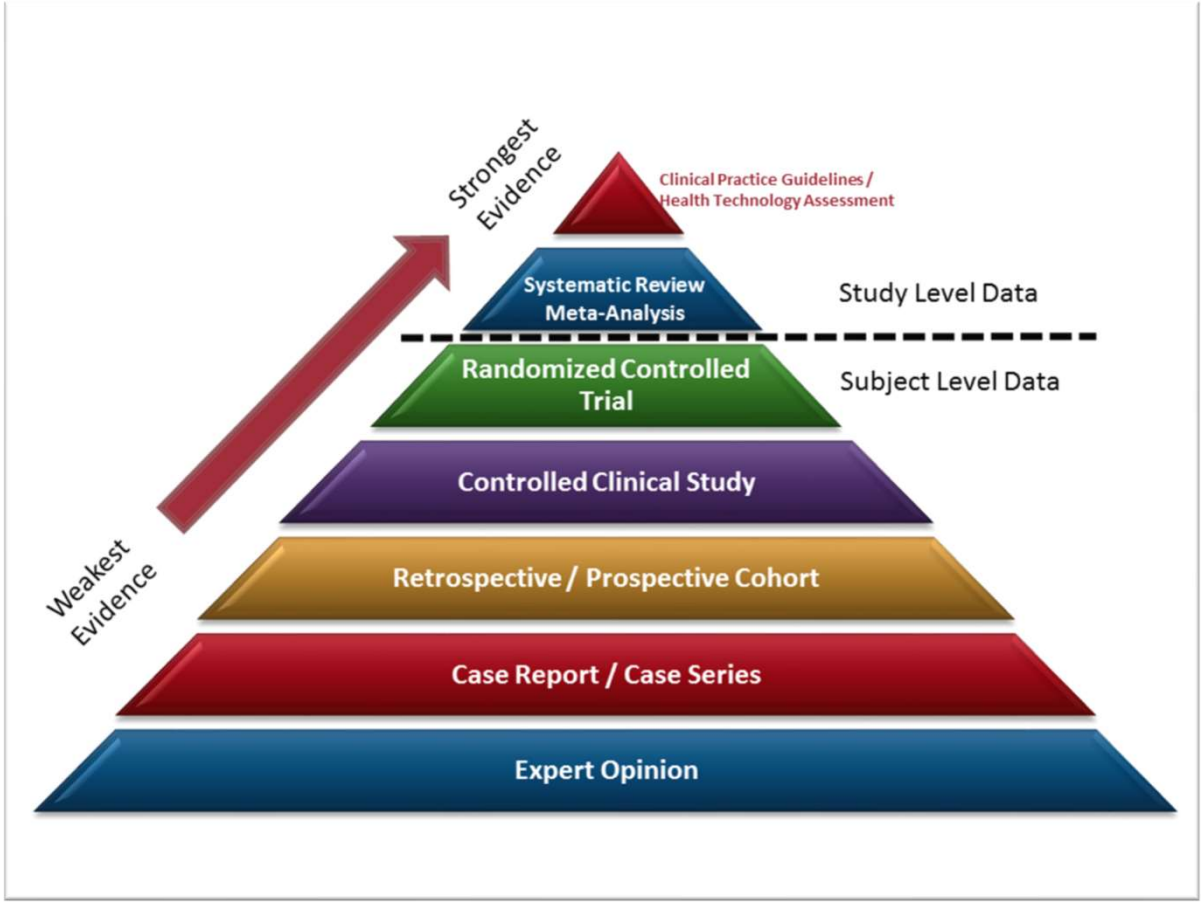
Cap d'equip

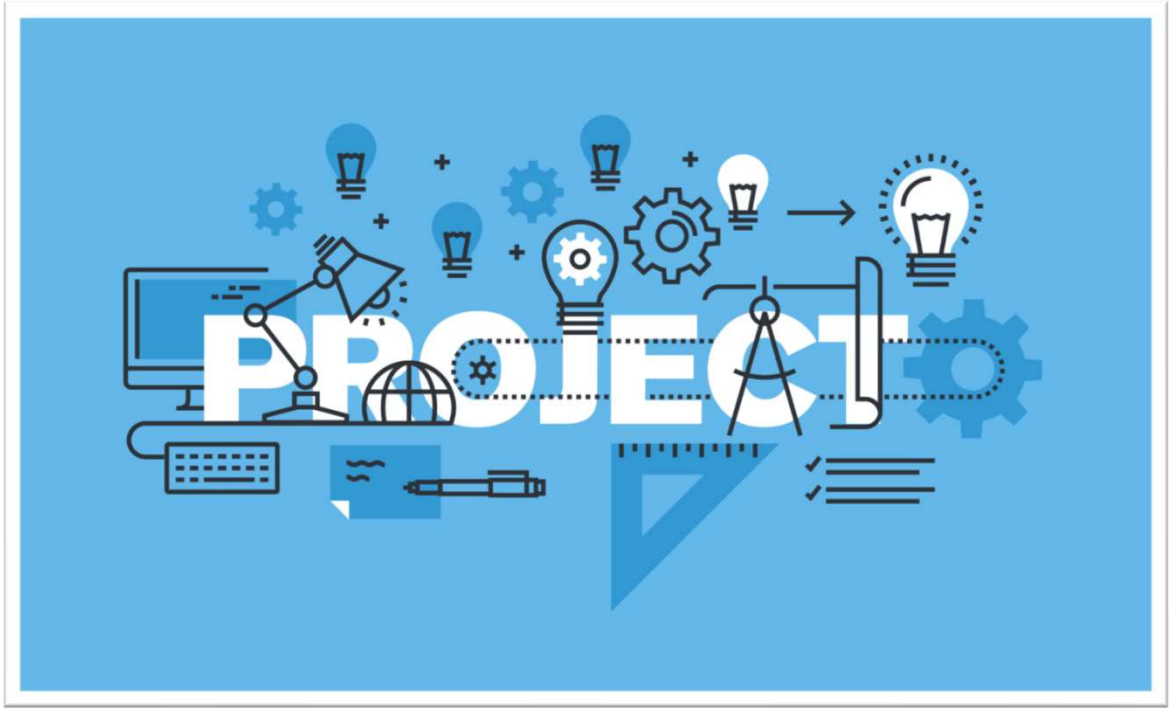
Josep Brugada
(Hospital Clínic)
JBRUGADA@clinic.cat


Examination
Inpatient health monitoring
Patient-generated health data
Medical imaging
Laboratory results



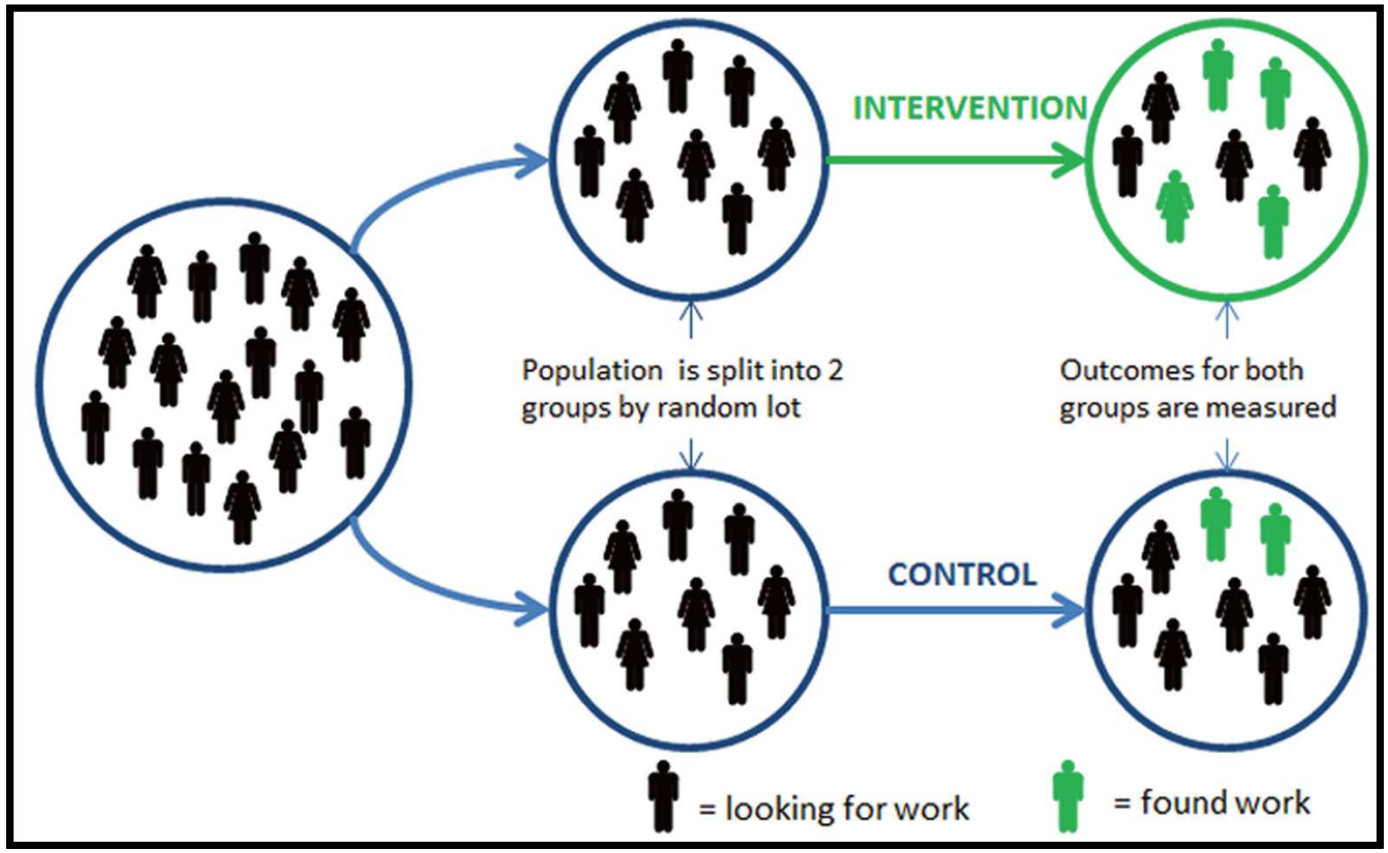
Evidence-Based Medicine



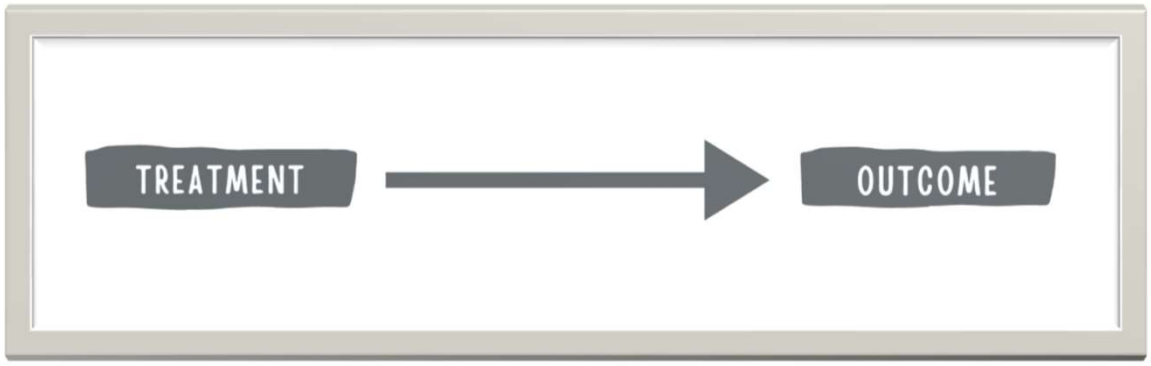


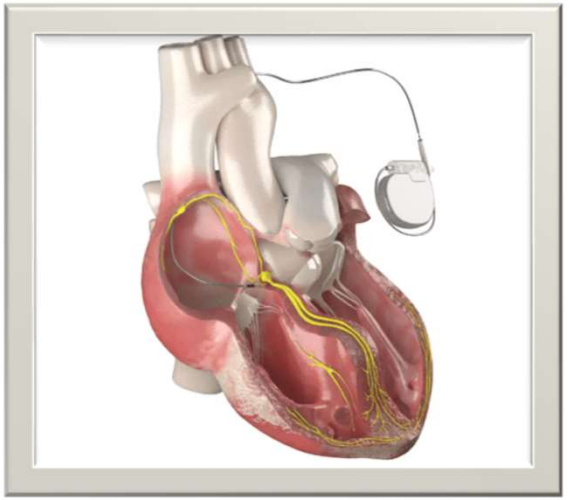
Project proposal

In RCTs...


INTERVENTION
Population is split into 2 groups by random lot
Outcomes for both groups are measured
CONTROL
= looking for work
= found work

Clínic Barcelona | Universitat de Barcelona



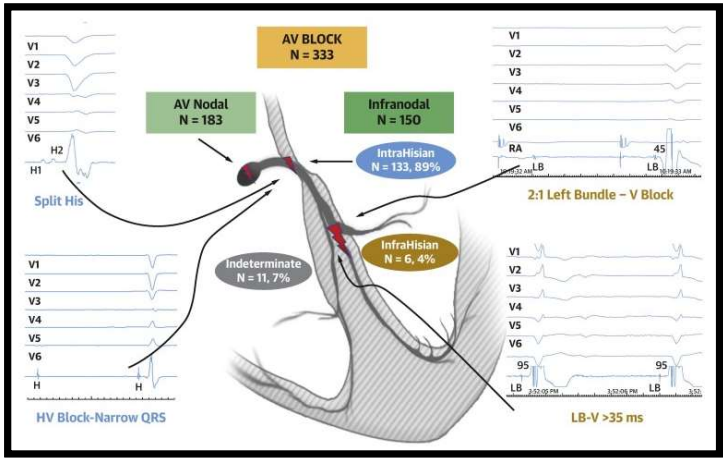


CIED - PHYSIOLOGICAL PACING

Conduction System Pacing vs Biventricular Pacing in Heart Failure and Wide QRS Patients

LEVEL-AT Trial

Margarida Pujol-Lopez, MD,[a,b] Rafael Jiménez-Arjona, MD,[a] Paz Garre, BEng,[a] Eduard Guasch, MD, PhD,[a,b,c] Roger Borràs, MSc,[a,b,d] Adelina Doltra, MD, PhD,[a,b] Elisenda Ferró, MEng,[e,f] Cora García-Ribas, MD, PhD,[a,f] Mireia Niebla, RN,[a] Esther Carro, RN,[a] Jose L. Puente, MD,[a] Sara Vázquez-Calvo, MD,[a] Eric Invers-Rubio, BEng,[b] Ivo Roca-Luque, MD, PhD,[a,b,c] M. Ángeles Castel, MD, PhD,[a,b,c] Elena Arbelo, MD, PhD,[a,b,c] Marta Sitges, MD, PhD,[a,b,c] Josep Brugada, MD, PhD,[a,b,c] José M. Tolosana, MD, PhD,[a,b,c,*] Lluís Mont, MD, PhD[a,b,c,*]



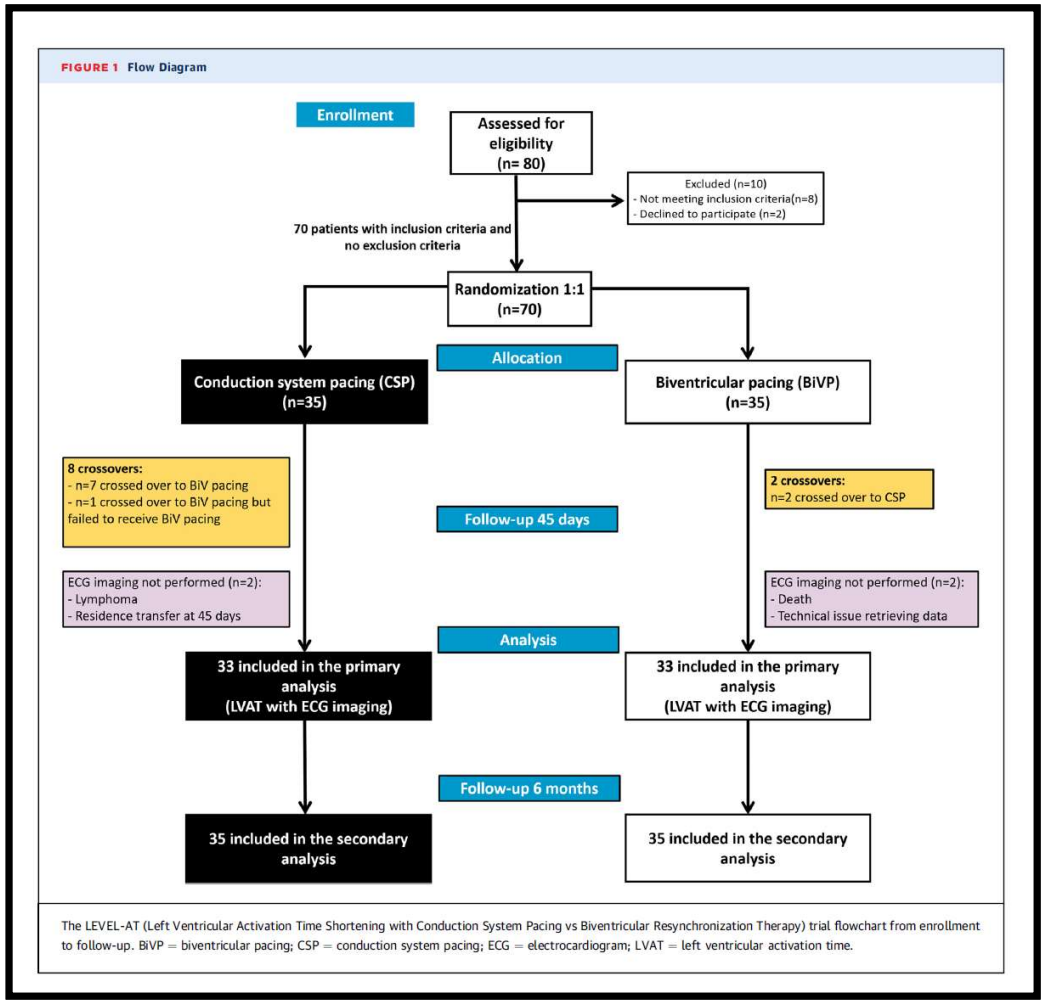
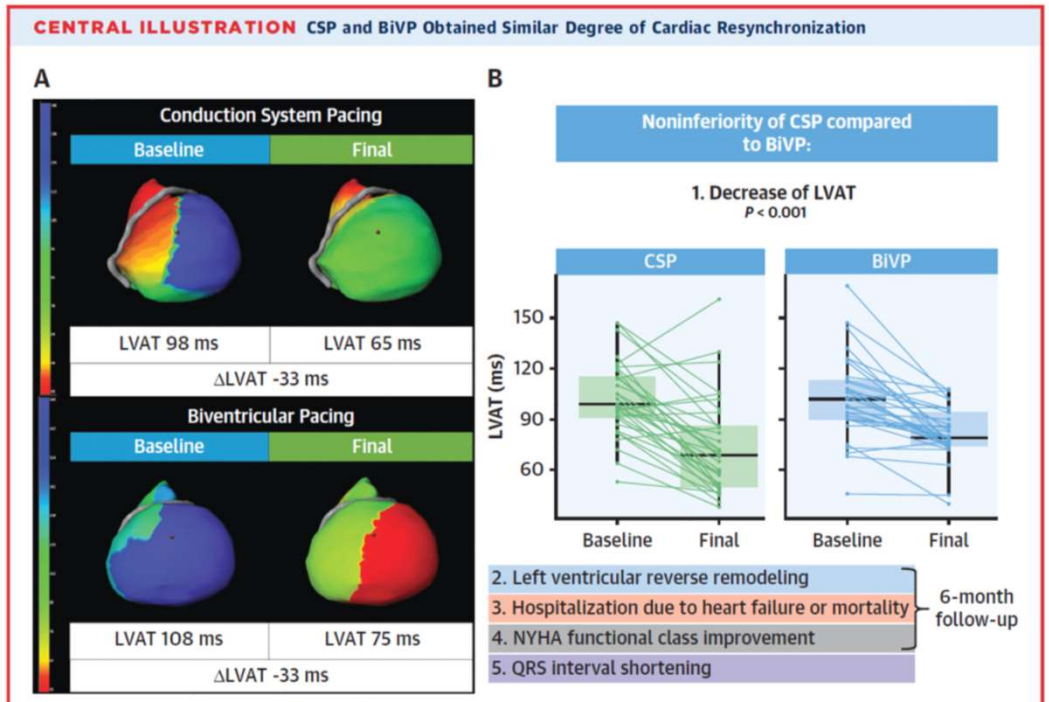


FIGURE 1 Flow Diagram
Enrollment
Assessed for eligibility (n= 80)
Excluded (n=10)
- Not meeting inclusion criteria(n=8)
- Declined to participate (n=2)
70 patients with inclusion criteria and no exclusion criteria
Randomization 1:1 (n=70)
Allocation
Conduction system pacing (CSP) (n=35)
Biventricular pacing (BiVP) (n=35)
8 crossovers:
- n=7 crossed over to BiV pacing
- n=1 crossed over to BiV pacing but failed to receive BiV pacing
2 crossovers:
n=2 crossed over to CSP
Follow-up 45 days
ECG imaging not performed (n=2):
- Lymphoma
- Residence transfer at 45 days
ECG imaging not performed (n=2):
- Death
- Technical issue retrieving data
Analysis
33 included in the primary analysis (LVAT with ECG imaging)
33 included in the primary analysis (LVAT with ECG imaging)
Follow-up 6 months
35 included in the secondary analysis
35 included in the secondary analysis
The LEVEL-AT (Left Ventricular Activation Time Shortening with Conduction System Pacing vs Biventricular Resynchronization Therapy) trial flowchart from enrollment to follow-up. BiVP = biventricular pacing; CSP = conduction system pacing; ECG = electrocardiogram; LVAT = left ventricular activation time.



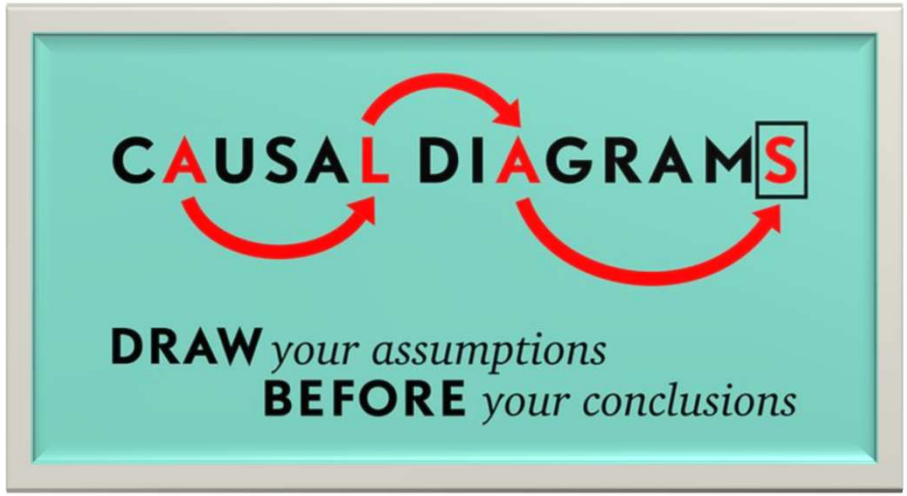
(A) Example of left ventricular (LV) activation time (LVAT) shortening with conduction system pacing (CSP) **(top)** and with biventricular pacing (BiVP) **(bottom)**. Both cases show long LVAT baseline with delayed activation of the left ventricle **(blue)**. Both CSP and BiVP showed a similar decrease in LVAT measured with electrocardiographic imaging and faster activation of the left ventricle **(green and red)**. **(B)** Both CSP and BiVP resulted in similar (noninferior) decrease of LVAT, LV reverse remodeling, heart failure hospitalizations or mortality, improvement in New York Heart Association (NYHA) functional class at 6 months, and QRS shortening. ΔLVAT = (final LVAT − baseline LVAT).

TABLE 2 Endpoints for Intention to Treat

	CSP	BiVP	Mean Difference (95% CI)	*P* Value for Noninferiority	*P* Value for Superiority
Primary					
n	33	33			
Delta LVAT, ms	−28 ± 26	−21 ± 20	−6.8 (−18.3 to 4.6)	<0.001	0.24
Secondary					
n	35	35			
Final QRS duration, ms	125 ± 18	129 ± 13	−4.7 (−12.2 to 2.9)	<0.001	0.23
Delta QRS, ms	−53 ± 20	−48 ± 20	−4.4 (−13.9 to 5.0)	<0.001	0.35
Correction of SF, mm	−1.9 ± 1.8	−2.0 ± 1.7	0.1 (−0.8 to 0.9)	0.15	0.89
Delta LVEF at 6 mo, %					
6 mo - baseline	12.2 ± 9	13.1 ± 9	−0.9 (−5.3 to 3.6)	0.07	0.69
Delta LVESV at 6 mo, mL					
6 mo - baseline	−37 ± 59	−30 ± 41	−8 (−33 to 17)	0.04	0.55
Delta NYHA functional class at 6 mo	−0.8 ± 0.8	−0.4 ± 0.8	−0.3 (−0.7 to 0.0)	<0.001	0.08
Heart failure hospitalization or mortality at 6 mo	2.9 (1)	11.4 (4)	−9% (−21% to 4%)	0.002	0.16
Heart failure hospitalization	2.9 (1)	8.6 (3)			
Mortality	0 (0)	5.7 (2*)			

Values are n, mean ± SD, or n (%), unless otherwise indicated. *1 hospitalized patient later died.
LVESV = left ventricular end systolic volume; SF = septal flash; other abbreviations as in Table 1.

Clínic Barcelona | Universitat de Barcelona



Bone Marrow Transplantation (2020) 55:4–8
https://doi.org/10.1038/s41409-018-0424-x

EBMT

EDITORIAL

Check for updates

Causal inference in randomized clinical trials

Cheng Zheng[1] · Ran Dai[2] · Robert Peter Gale[3] · Mei-Jie Zhang[4]

Received: 7 November 2018 / Accepted: 12 November 2018 / Published online: 26 March 2019

Received: 3 March 2021 | Revised: 8 October 2021 | Accepted: 11 October 2021
DOI: 10.1002/sim.9234

TUTORIAL IN BIOSTATISTICS

Statistics in Medicine WILEY

Introduction to computational causal inference using reproducible Stata, R, and Python code: A tutorial

Matthew J. Smith[1] | Mohammad A. Mansournia[2] | Camille Maringe[1] | Paul N. Zivich[3,4] | Stephen R. Cole[3] | Clémence Leyrat[1] | Aurélien Belot[1] | Bernard Rachet[1] | Miguel A. Luque-Fernandez[1,5,6]

[1]Inequalities in Cancer Outcomes Network, Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK
[2]Department of Epidemiology and Biostatistics, Tehran University of Medical Sciences, Tehran, Iran
[3]Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA
[4]Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA
[5]Non-communicable Disease and Cancer Epidemiology Group, Instituto de Investigacion Biosanitaria de Granada (ibs.GRANADA), Andalusian School of Public Health, University of Granada, Granada, Spain
[6]Biomedical Network Research Centers of Epidemiology and Public Health (CIBERESP), Madrid, Spain

Correspondence
Miguel A. Luque-Fernandez, Inequalities in Cancer Outcomes Network, Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, Keppel Street, Bloomsbury, London WC1E 7HT, UK.
Email: miguel-angel.luque@lshtm.ac.uk

Funding information
Cancer Research UK; Instituto de Salud Carlos III

Abstract
The main purpose of many medical studies is to estimate the effects of a treatment or exposure on an outcome. However, it is not always possible to randomize the study participants to a particular treatment, therefore observational study designs may be used. There are major challenges with observational studies; one of which is confounding. Controlling for confounding is commonly performed by direct adjustment of measured confounders; although, sometimes this approach is suboptimal due to modeling assumptions and misspecification. Recent advances in the field of causal inference have dealt with confounding by building on classical standardization methods. However, these recent advances have progressed quickly with a relative paucity of computational-oriented applied tutorials contributing to some confusion in the use of these methods among applied researchers. In this tutorial, we show the computational implementation of different causal inference estimators from a historical perspective where new estimators were developed to overcome the limitations of the previous estimators (ie, nonparametric and parametric g-formula, inverse probability weighting, double-robust, and data-adaptive estimators). We illustrate the implementation of different methods using an empirical example from the Connors study based on intensive care medicine, and most importantly, we provide reproducible and commented code in Stata, R, and Python for researchers to adapt in their own observational study. The code can be accessed at https://github.com/migariane/Tutorial_Computational_Causal_Inference_Estimators.

KEYWORDS
causal inference, double-robust methods, g-formula, G-methods, inverse probability weighting, machine learning, propensity score, regression adjustment, targeted maximum likelihood estimation

Clínic Barcelona | Universitat de Barcelona

Definition 1: Individual treatment effect

The individual treatment effect, δ_i, equals $Y_i^1 - Y_i^0$

Definition 2: Average treatment effect (ATE)

The average treatment effect is the population average of all i individual treatment effects

$$\begin{aligned} E[\delta_i] &= E[Y_i^1 - Y_i^0] \\ &= E[Y_i^1] - E[Y_i^0] \end{aligned}$$

Definition 3: Switching equation

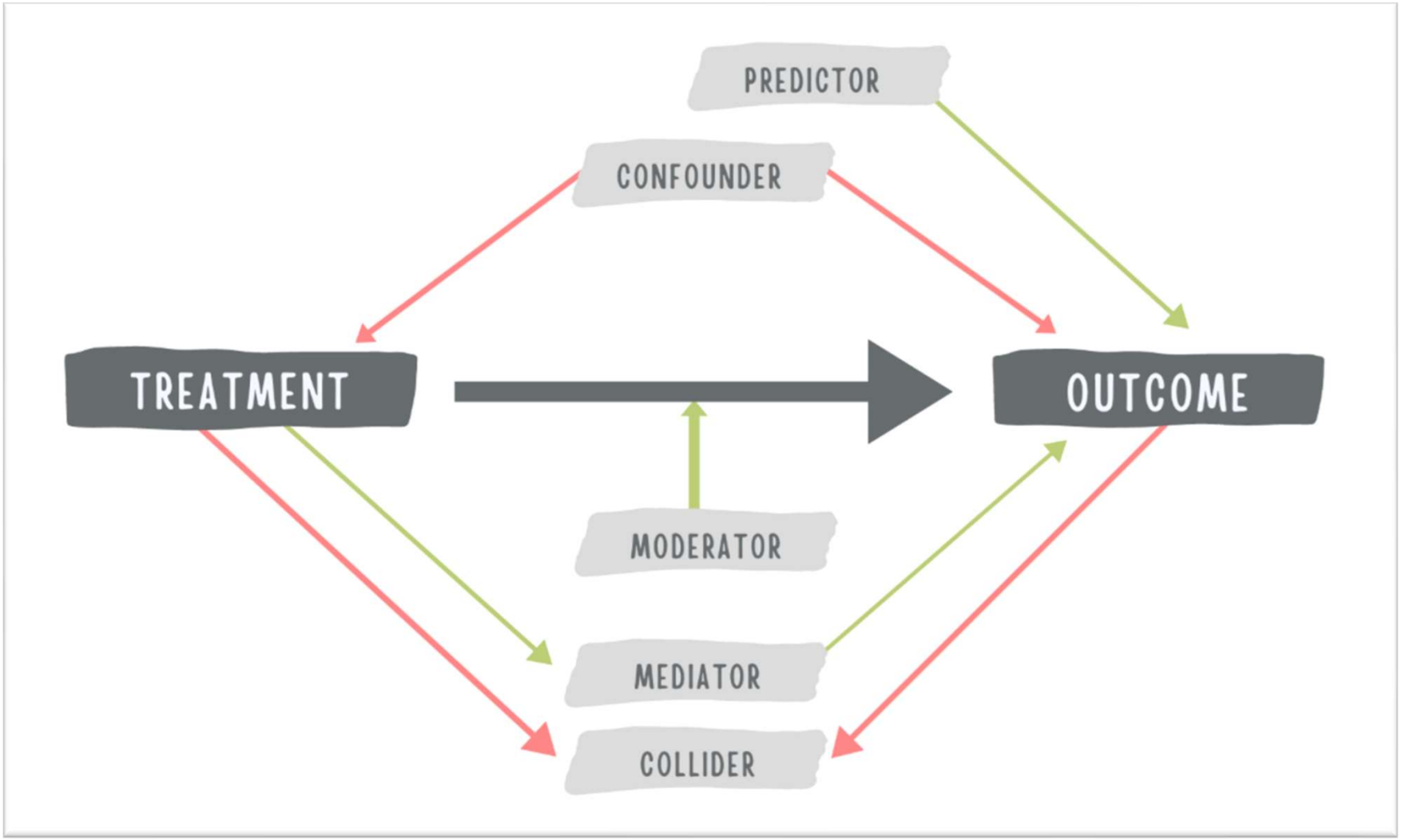
An individual's observed health outcomes, Y, is determined by treatment assignment, D_i, and corresponding potential outcomes:

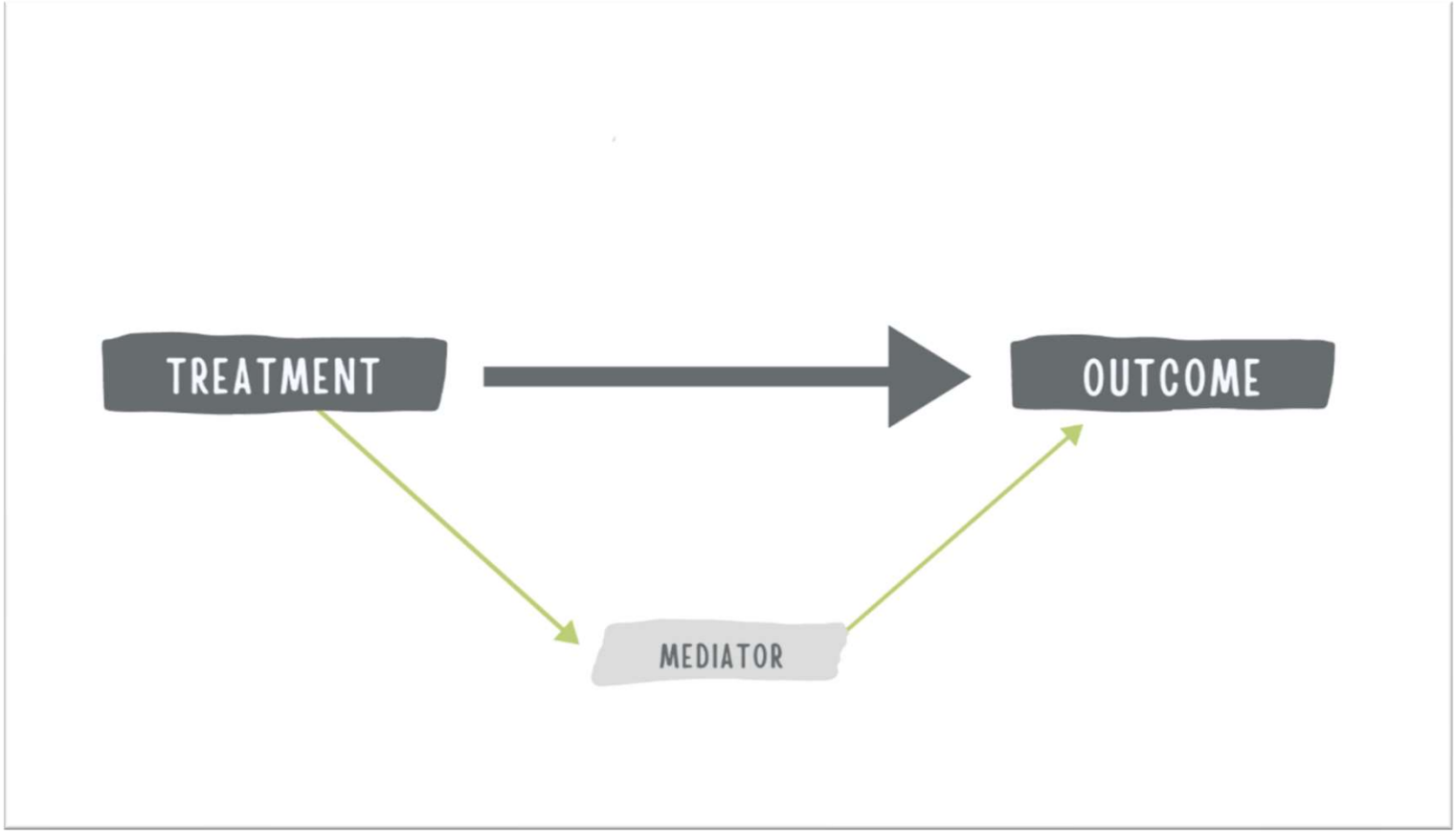
$$Y_i = D_i Y_i^1 + (1 - D_i) Y_i^0$$

$$Y_i = \begin{cases} Y_i^1 \text{ if } D_i = 1 \\ Y_i^0 \text{ if } D_i = 0 \end{cases}$$

Definition 4: Fundamental problem of causal inference

It is impossible to observe both Y_i^1 and Y_i^0 for the same individual and so individual causal effects, δ_i, are unknowable.


PREDICTOR
CONFOUNDER
TREATMENT
OUTCOME
MODERATOR
MEDIATOR
COLLIDER


TREATMENT
OUTCOME
MEDIATOR

Mediation Analysis

A mediation model seeks to identify and explain the mechanism or process that underlies an observed relationship between an independent variable and a dependent variable via the inclusion of a third hypothetical variable, known as a mediator variable (also a mediating variable, intermediary variable, or intervening variable). Rather than a direct causal relationship between the independent variable and the dependent variable, a mediation model proposes that the independent variable influences the mediator variable, which in turn influences the dependent variable. Thus, the mediator variable serves to clarify the nature of the relationship between the independent and dependent variables.

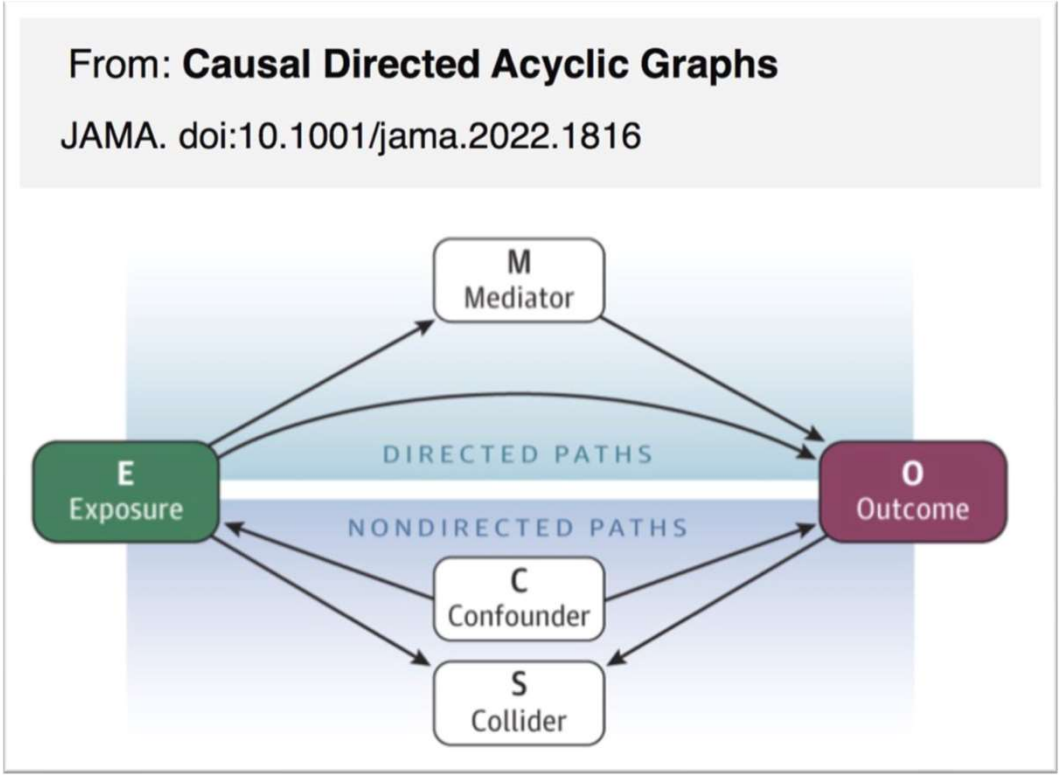


Table 1. Unit-Specific Quantities Defined in Potential Outcomes Unlock Many Causal Estimands for Inquiry

Estimand name	Mathematical statement	DAG	Reference	Colloquial terms
Average treatment effect	$\frac{1}{n}\sum_i \left(Y_i(d') - Y_i(d)\right)$	$D \longrightarrow Y$	Morgan and Winship (2015)	Effect
Conditional average treatment effect	$\frac{1}{n_x}\sum_{i:X_i=x} \left(Y_i(d') - Y_i(d)\right)$	$X \longrightarrow D \longrightarrow Y$	Athey and Imbens (2016)	Effect heterogeneity or moderation
Causal interaction	$\frac{1}{n}\sum_i \Big(\left(Y_i(a',d') - Y_i(a',d)\right) - \left(Y_i(a,d') - Y_i(a,d)\right)\Big)$	$A \to Y \leftarrow D$	Vanderweele (2015)	Joint treatment effect
Controlled direct effect	$\frac{1}{n}\sum_i \left(Y_i(d',m) - Y_i(d,m)\right)$	$D \to Y$, M	Acharya et al. (2016)	Mediation (Illustrations: Example 2)
Natural direct effect	$\frac{1}{n}\sum_i \Big(Y_i(d', M_i(d)) - Y_i(d, M_i(d))\Big)$	$D \to Y$, M	Imai et al. (2011)	Mediation (Part B of the Online Supplement)
Effect of time-varying treatment	$\frac{1}{n}\sum_i \left(Y_i(d_1', d_2') - Y_i(d_1, d_2)\right)$	$D_1 \to D_2 \to Y$	Wodtke et al. (2011)	Cumulative effect

Clínic Barcelona | Universitat de Barcelona

METHODOLOGY Open Access

Investigating causal mechanisms in randomised controlled trials

Hopin Lee[1,2*], Robert D. Herbert[3,4], Sarah E. Lamb[1], Anne M. Moseley[5] and James H. McAuley[3,6]

Abstract

Introduction: In some randomised trials, the primary interest is in the mechanisms by which an intervention exerts its effects on health outcomes. That is, clinicians and policy-makers may be interested in how the intervention works (or why it does not work) through hypothesised causal mechanisms. In this article, we highlight the value of understanding causal mechanisms in randomised trials by applying causal mediation analysis to two randomised trials of complex interventions.

Main body: In the first example, we examine a potential mechanism by which an exercise programme for rheumatoid arthritis of the hand could improve hand function. In the second example, we explore why a rehabilitation programme for ankle fractures failed to improve lower-limb function through hypothesised mechanisms. We outline critical assumptions that are required for making valid causal inferences from these analyses, and provide results of sensitivity analyses that are used to assess the degree to which the estimated causal mediation effects could have been biased by residual confounding.

Conclusion: This paper demonstrates how the application of causal mediation analyses to randomised trials can identify the mechanisms by which complex interventions exert their effects. We discuss methodological issues and assumptions that should be considered when mediation analyses of randomised trials are used to inform clinical practice and policy decisions.

Keywords: Mechanism, Mediation analysis, Complex interventions, Causal inference, Musculoskeletal system, Rheumatoid arthritis, Ankle fractures, Exercise therapy

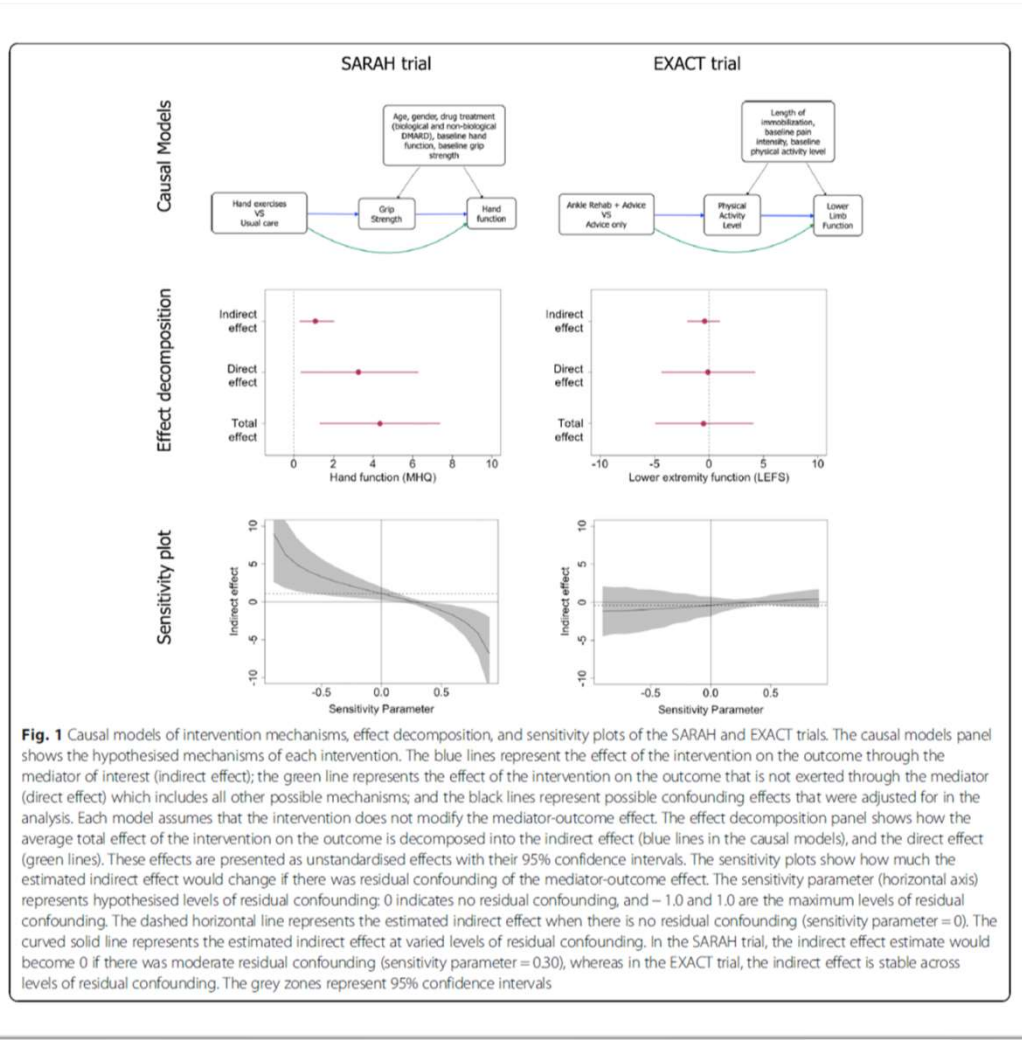
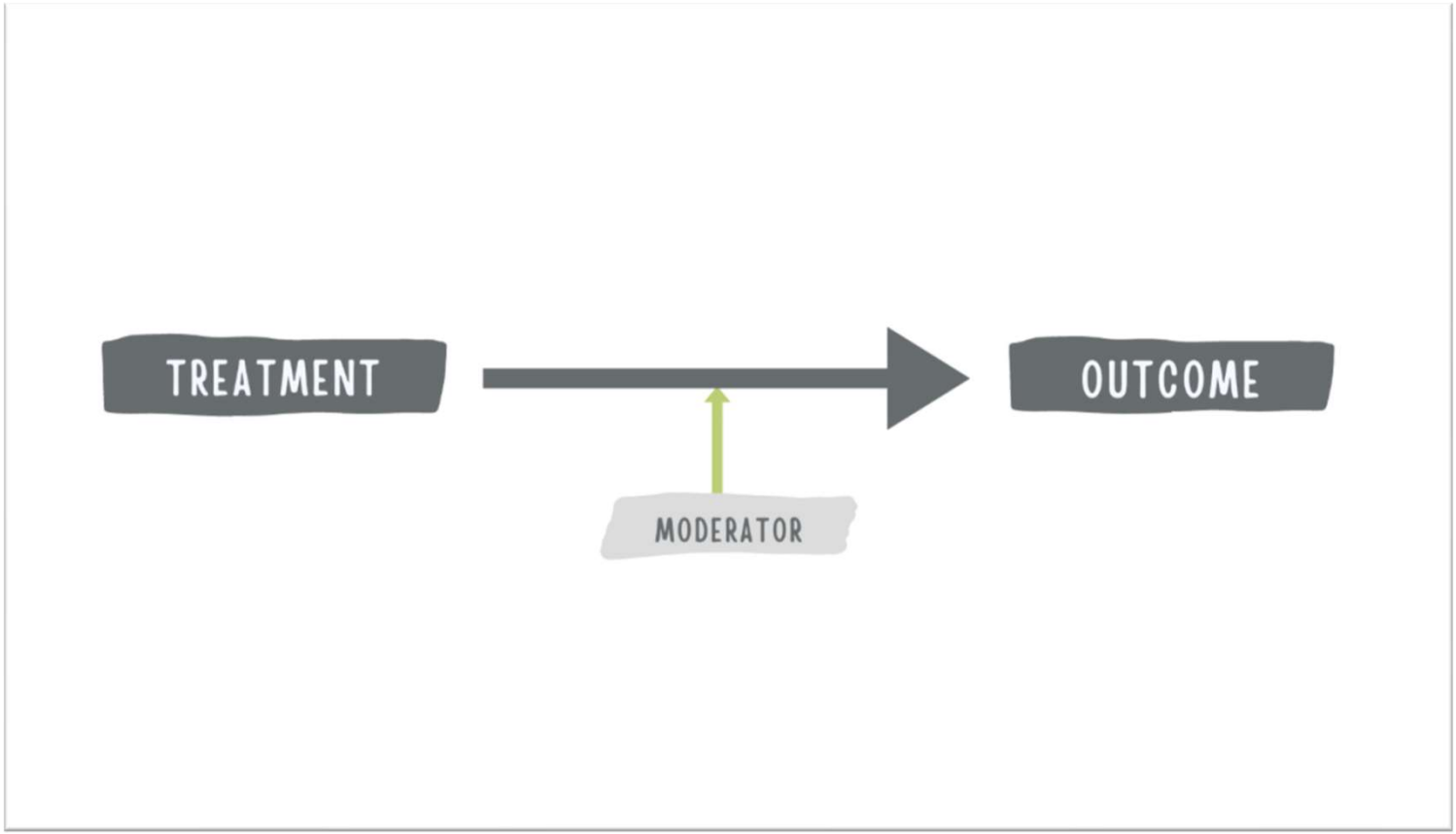


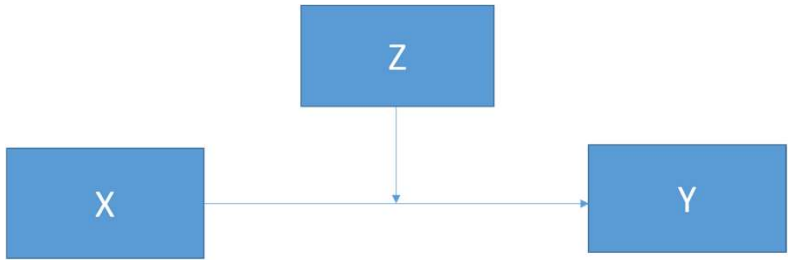
Fig. 1 Causal models of intervention mechanisms, effect decomposition, and sensitivity plots of the SARAH and EXACT trials. The causal models panel shows the hypothesised mechanisms of each intervention. The blue lines represent the effect of the intervention on the outcome through the mediator of interest (indirect effect); the green line represents the effect of the intervention on the outcome that is not exerted through the mediator (direct effect) which includes all other possible mechanisms; and the black lines represent possible confounding effects that were adjusted for in the analysis. Each model assumes that the intervention does not modify the mediator-outcome effect. The effect decomposition panel shows how the average total effect of the intervention on the outcome is decomposed into the indirect effect (blue lines in the causal models), and the direct effect (green lines). These effects are presented as unstandardised effects with their 95% confidence intervals. The sensitivity plots show how much the estimated indirect effect would change if there was residual confounding of the mediator-outcome effect. The sensitivity parameter (horizontal axis) represents hypothesised levels of residual confounding: 0 indicates no residual confounding, and − 1.0 and 1.0 are the maximum levels of residual confounding. The dashed horizontal line represents the estimated indirect effect when there is no residual confounding (sensitivity parameter = 0). The curved solid line represents the estimated indirect effect at varied levels of residual confounding. In the SARAH trial, the indirect effect estimate would become 0 if there was moderate residual confounding (sensitivity parameter = 0.30), whereas in the EXACT trial, the indirect effect is stable across levels of residual confounding. The grey zones represent 95% confidence intervals

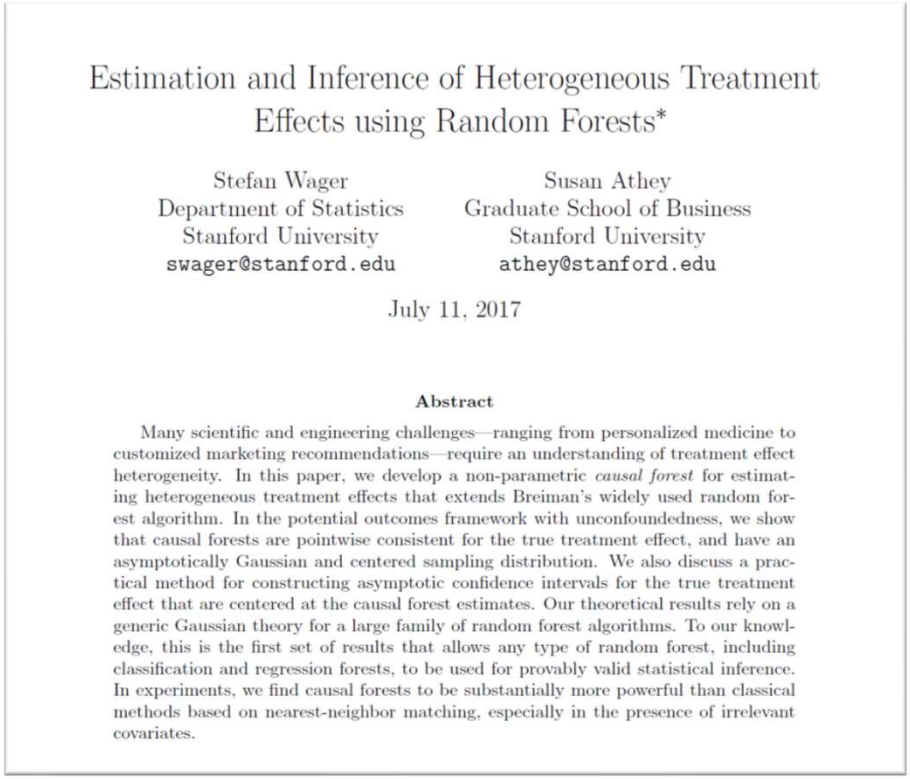
- Detecció de diferents mecanismes segons el grup (buscar asimetries de mediació)
- Diferents outcomes?


TREATMENT
OUTCOME
MODERATOR

Conditional Average Treatment Effects (CATE)

The CATE is the treatment effect in a subgroup of the population, while the ATE is the treatment effect in the population at large. If the composition of variables that modify the treatment effect differs between the subgroup and the population, then CATE will not equal ATE.





Estimation and Inference of Heterogeneous Treatment Effects using Random Forests*

Stefan Wager
Department of Statistics
Stanford University
swager@stanford.edu

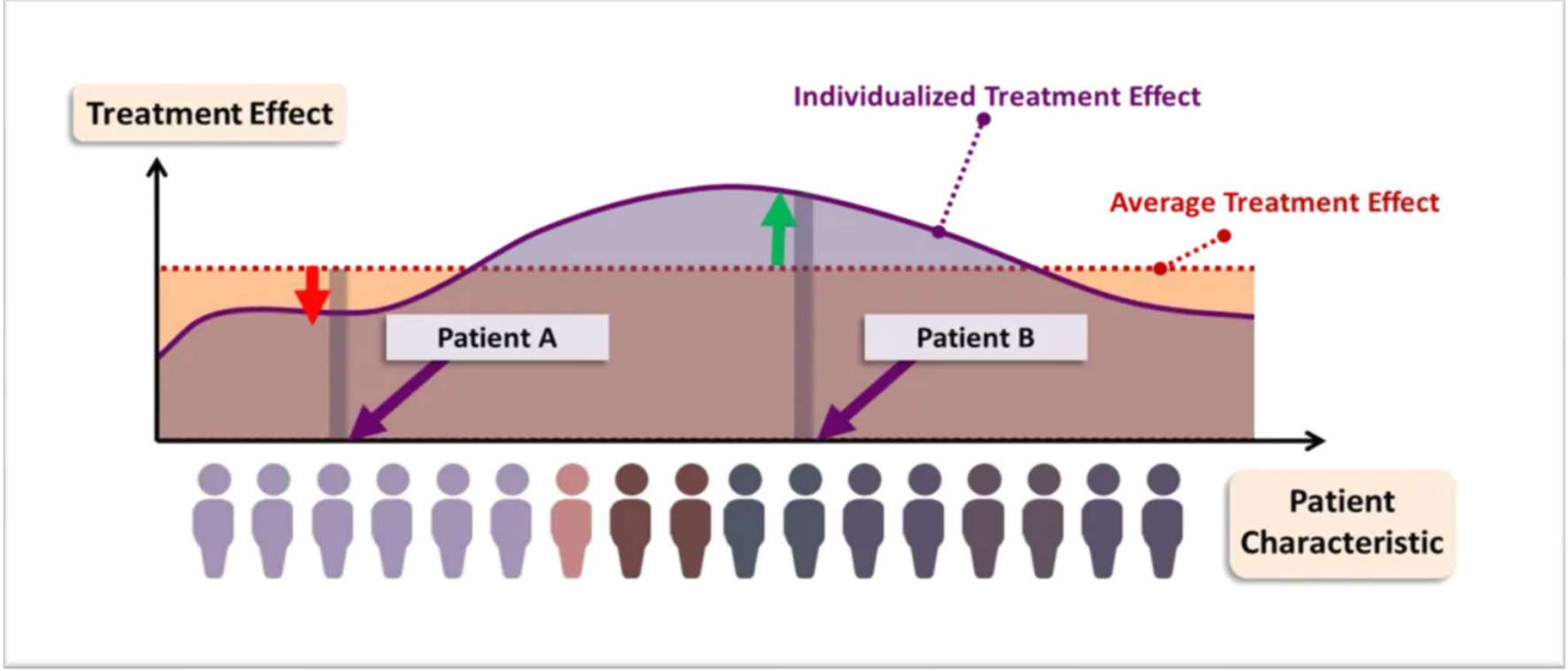
Susan Athey
Graduate School of Business
Stanford University
athey@stanford.edu

July 11, 2017

Abstract

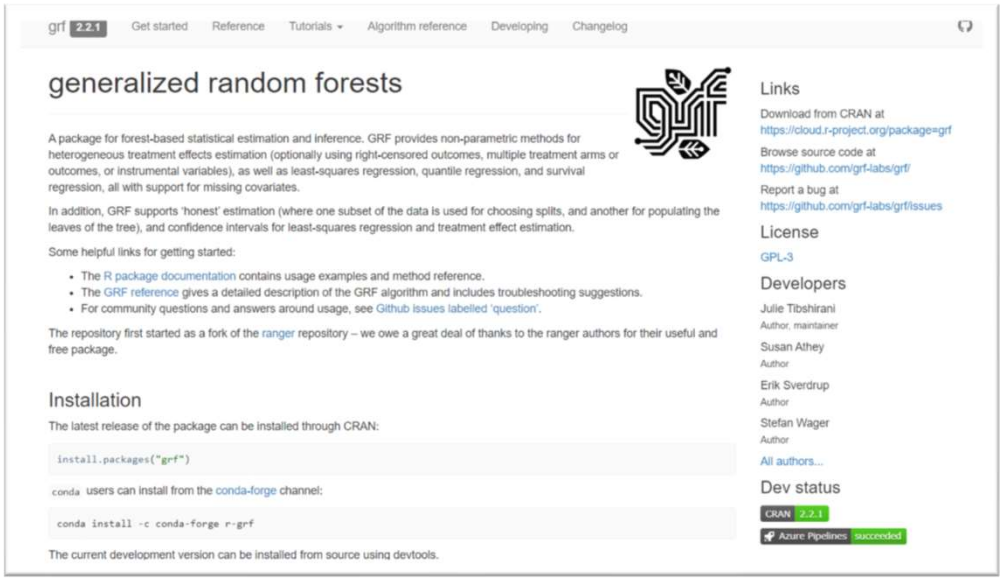
Many scientific and engineering challenges—ranging from personalized medicine to customized marketing recommendations—require an understanding of treatment effect heterogeneity. In this paper, we develop a non-parametric *causal forest* for estimating heterogeneous treatment effects that extends Breiman's widely used random forest algorithm. In the potential outcomes framework with unconfoundedness, we show that causal forests are pointwise consistent for the true treatment effect, and have an asymptotically Gaussian and centered sampling distribution. We also discuss a practical method for constructing asymptotic confidence intervals for the true treatment effect that are centered at the causal forest estimates. Our theoretical results rely on a generic Gaussian theory for a large family of random forest algorithms. To our knowledge, this is the first set of results that allows any type of random forest, including classification and regression forests, to be used for provably valid statistical inference. In experiments, we find causal forests to be substantially more powerful than classical methods based on nearest-neighbor matching, especially in the presence of irrelevant covariates.

"Many scientic and engineering challenges - ranging from personalized medicine to customized marketing recommendations- require an understanding of treatment effect heterogeneity."



- Detecció de variables HTE segons el grup per definer millor el target

Clínic Barcelona | Universitat de Barcelona



grf 2.2.1 Get started Reference Tutorials Algorithm reference Developing Changelog

generalized random forests

A package for forest-based statistical estimation and inference. GRF provides non-parametric methods for heterogeneous treatment effects estimation (optionally using right-censored outcomes, multiple treatment arms or outcomes, or instrumental variables), as well as least-squares regression, quantile regression, and survival regression, all with support for missing covariates.

In addition, GRF supports 'honest' estimation (where one subset of the data is used for choosing splits, and another for populating the leaves of the tree), and confidence intervals for least-squares regression and treatment effect estimation.

Some helpful links for getting started:

- The R package documentation contains usage examples and method reference.
- The GRF reference gives a detailed description of the GRF algorithm and includes troubleshooting suggestions.
- For community questions and answers around usage, see Github issues labelled 'question'.

The repository first started as a fork of the ranger repository – we owe a great deal of thanks to the ranger authors for their useful and free package.

Installation

The latest release of the package can be installed through CRAN:

```
install.packages("grf")
```

`conda` users can install from the conda-forge channel:

```
conda install -c conda-forge r-grf
```

The current development version can be installed from source using devtools.

Links

Download from CRAN at https://cloud.r-project.org/package=grf

Browse source code at https://github.com/grf-labs/grf/

Report a bug at https://github.com/grf-labs/grf/issues

License

GPL-3

Developers

Julie Tibshirani
Author, maintainer

Susan Athey
Author

Erik Sverdrup
Author

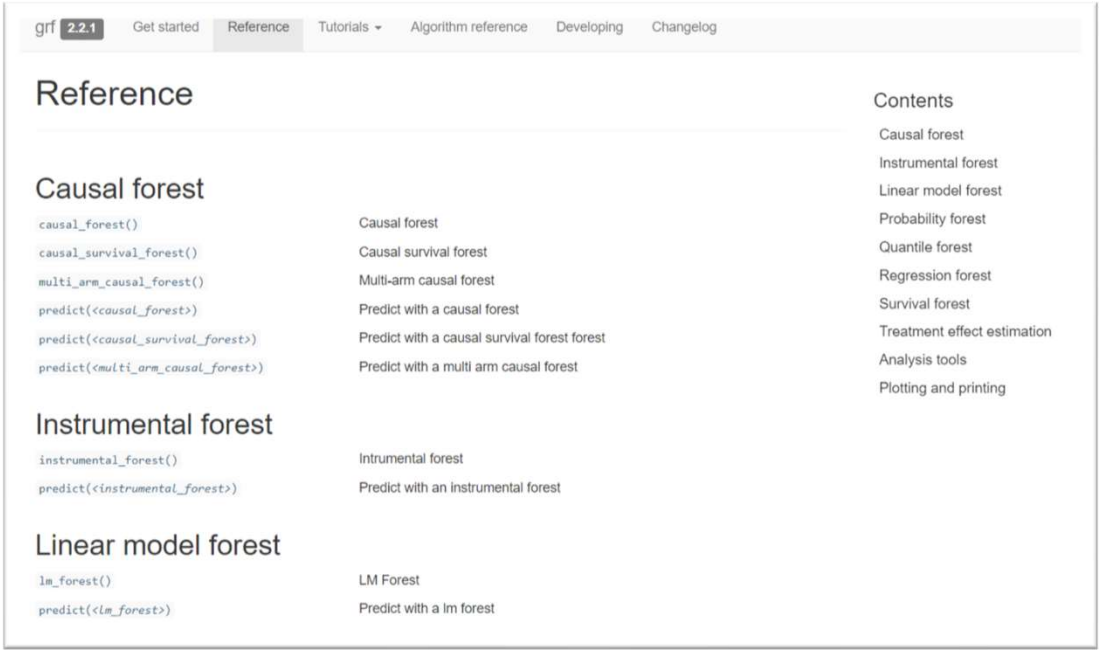
Stefan Wager
Author

All authors...

Dev status

CRAN 2.2.1

Azure Pipelines succeeded



grf 2.2.1 Get started Reference Tutorials Algorithm reference Developing Changelog

Reference

Causal forest

`causal_forest()`	Causal forest
`causal_survival_forest()`	Causal survival forest
`multi_arm_causal_forest()`	Multi-arm causal forest
`predict(<causal_forest>)`	Predict with a causal forest
`predict(<causal_survival_forest>)`	Predict with a causal survival forest forest
`predict(<multi_arm_causal_forest>)`	Predict with a multi arm causal forest

Instrumental forest

`instrumental_forest()`	Intrumental forest
`predict(<instrumental_forest>)`	Predict with an instrumental forest

Linear model forest

`lm_forest()`	LM Forest
`predict(<lm_forest>)`	Predict with a lm forest

Contents

Clínic Barcelona | Universitat de Barcelona

Opinion

EDITORIAL

Toward Personalizing Care
Assessing Heterogeneity of Treatment Effects in Randomized Trials

Issa J. Dahabreh, MD, ScD; Dhruv S. Kazi, MD, MS

Clinicians know that individual patients may respond differently to a given treatment and that the overall treatment effect reported in a randomized trial of the treatment may not be directly applicable to all patients in clinical practice.[1] Determining the treatment effect for an individual patient involves a comparison of the outcome when that patient is exposed to the treatment vs the outcome of the same patient exposed to a control treatment at the same time, a comparison impossible to make in conventional parallel-group trial designs. A practical alternative is to examine heterogeneity of (variation in) treatment effects across groups of patients, categorized by baseline demographic or clinical characteristics, such as age or risk factors for the outcome.[2]

Related article

In this issue of *JAMA*, Goligher and colleagues[3] explore the ability of contemporary statistical techniques to detect heterogeneity of treatment effects using pooled data from 3 randomized platform trials assessing the effect of therapeutic-dose heparin on organ support-free days and all-cause mortality in patients hospitalized for COVID-19 in the early pandemic. They compare 3 approaches for identifying heterogeneity of treatment effects: (1) traditional one-variable-at-a-time subgroup analyses; (2) risk score analyses, in which patients are grouped by predicted risk of trial outcomes; and (3) effect score analyses, in which patients are grouped by predicted treatment effect. The 3 approaches yielded congruent results, suggesting that patients with a body mass index (BMI) of less than 30 and those with moderate severity COVID-19 at presentation appeared to benefit, whereas those with a BMI of 30 or greater and severe COVID-19 at presentation did not benefit and may have been harmed. These findings highlight the need to evaluate heterogeneity of treatment effects in randomized trials: had the trials evaluated the effect of therapeutic heparin in patients hospitalized with COVID-19 without stratifying by disease severity, the overall treatment effect may have been close to null, obscuring signals of differential benefit and harm across clinically meaningful patient subgroups.

This Editorial will attempt to explain the rationale for Goligher and colleagues' efforts, place them in broader methodological context, and offer suggestions for future assessments of heterogeneity in randomized trials.

Traditional subgroup analyses to examine heterogeneity of treatment effects are ubiquitous in the medical literature. Investigators group trial participants by clinical variables (eg, disease severity or BMI categories) and assess whether effects are heterogeneous across subgroups, one variable at a time (Goligher and colleagues' first approach). As common as the approach is, it presents several challenges.[4,5] Trials are typically statistically underpowered to detect differences between subgroups, so there is high risk of false-negative findings. At the same time, performing multiple subgroup analyses increases the risk of false-positive findings. Although these challenges can be addressed by approaches such as rigorous prespecification of comparisons, multiplicity adjustments for hypothesis testing, and hierarchical modeling, a key practical limitation remains: one-variable-at-a-time subgroup analyses are difficult to use for clinical decision-making because each patient belongs to multiple subgroups and each subgroup may have a different magnitude and direction of treatment effect (eg, a patient can have a BMI of less than 30 and severe COVID-19).[2] Thus, traditional one-variable-at-a-time subgroup analyses may be useful as exploratory or descriptive analyses, and may produce population-level insights, but multiple variables have to be jointly considered to generate clinically relevant assessment of heterogeneity of treatment effects.

One way to integrate information from multiple variables is to examine heterogeneity of treatment effects over the predicted risk of a trial outcome (Goligher and colleagues' second approach).[6] A well-calibrated risk model is used to integrate multiple variables into a single "risk score" variable that captures risk of the outcome without treatment, followed by an examination of whether treatment effects vary over the risk score.[7] In practice, risk score analyses typically have 3 steps: first, a risk score—internally developed using the trial data or externally developed using independent data—is used to group trial participants by level of predicted risk; next, risk group-specific treatment effects are estimated; and, finally, the treatment effects are examined for heterogeneity. There are several advantages to this approach. Clinicians intuitively incorporate risk into clinical decision-making, validated risk scores are widely used in clinical practice, and risk is correlated with treatment benefit. To the extent that the risk score captures variation in risk, it should be able to identify groups of patients who are unlikely to benefit from treatment as well as groups that have the potential to benefit. Furthermore, by reducing multiple variables into a single score, risk score approaches avoid the multiplicity issues of one-variable-at-a-time subgroup analyses. These attractive features may explain the increasing popularity of risk score analyses in randomized trials and the emphasis on such approaches in recent methodological recommendations.[8]

But risk score analyses may not fully capture heterogeneity of treatment effects because risk of an outcome in the absence of treatment may not strongly correlate with benefit or harm from treatment. For example, among patients hospitalized with COVID-19, a patient with a BMI of less than 30 and

Research

JAMA | Original Investigation | CARING FOR THE CRITICALLY ILL PATIENT

Heterogeneous Treatment Effects of Therapeutic-Dose Heparin in Patients Hospitalized for COVID-19

Ewan C. Goligher, MD, PhD; Patrick R. Lawler, MD, MPH; Thomas P. Jensen, MS; Victor Talisa, PhD; Lindsay R. Berry, PhD; Elizabeth Lorenzi, PhD; Bryan J. McVerry, MD; Chung-Chou Ho Chang, PhD; Eric Leifer, PhD; Charlotte Bradbury, MD, PhD; Jeffrey Berger, MD; Beverly J. Hunt, MD, PhD; Lana A. Castellucci, MD; Lucy Z. Kornblith, MD; Anthony C. Gordon, MD; Colin McArthur, MD; Steven Webb, MD; Judith Hochman, MD; Matthew D. Neal, MD; Ryan Zarychanski, MD, MSc; Scott Berry, PhD; Derek C. Angus, MD, MPH; for the REMAP-CAP, ATTACC, and ACTIV-4a Investigators

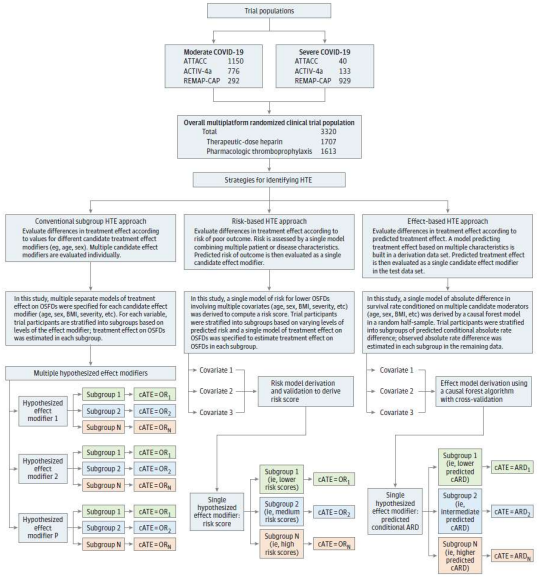
IMPORTANCE Randomized clinical trials (RCTs) of therapeutic-dose heparin in patients hospitalized with COVID-19 produced conflicting results, possibly due to heterogeneity of treatment effect (HTE) across individuals. Better understanding of HTE could facilitate individualized clinical decision-making.

OBJECTIVE To evaluate HTE of therapeutic-dose heparin for patients hospitalized for COVID-19 and to compare approaches to assessing HTE.

DESIGN, SETTING, AND PARTICIPANTS Exploratory analysis of a multiplatform adaptive RCT of therapeutic-dose heparin vs usual care pharmacologic thromboprophylaxis in 3320 patients hospitalized for COVID-19 enrolled in North America, South America, Europe, Asia, and Australia between April 2020 and January 2021. Heterogeneity of treatment effect was assessed 3 ways: using (1) conventional subgroup analyses of baseline characteristics, (2) a multivariable outcome prediction model (risk-based approach), and (3) a multivariable causal forest model (effect-based approach). Analyses primarily used bayesian statistics, consistent with the original trial.

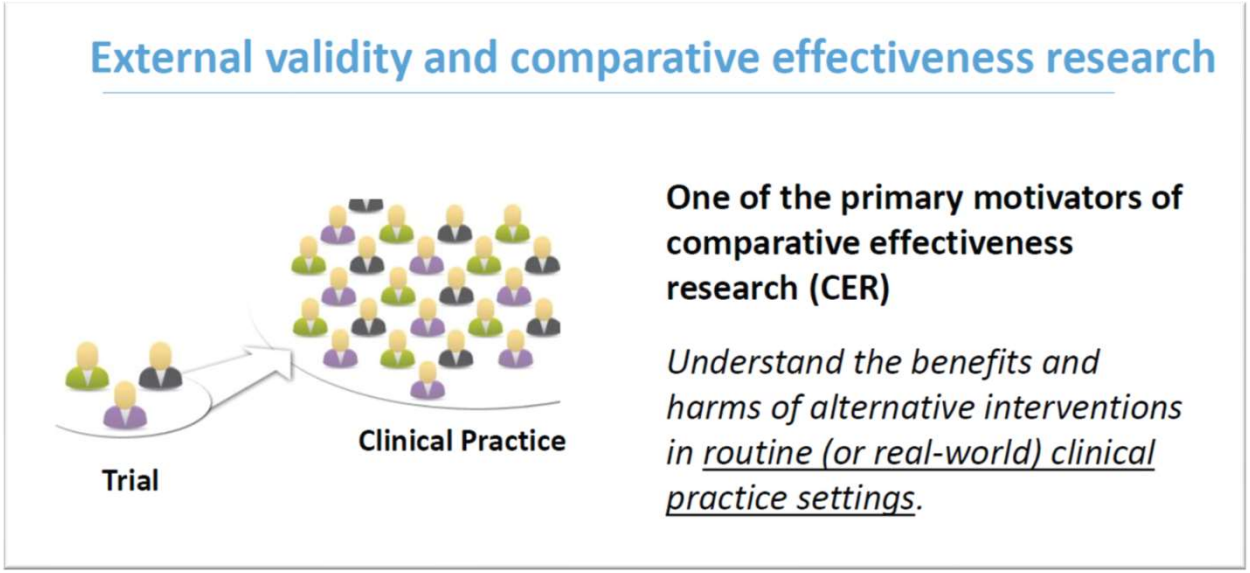
Editorial
Supplemental content

Figure 1. Three Strategies to Evaluate HTE of Therapeutic-Dose Heparin for COVID-19



Transportability problem

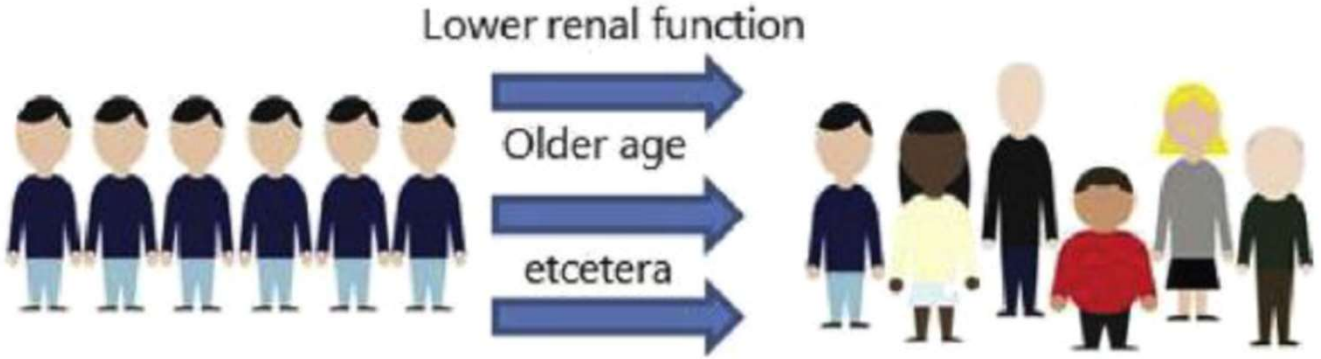
The study of transportability aims to identify conditions under which causal information learned from experiments can be reused in a different environment where only passive observations can be collected.



Why would treatment effects differ in trials and target populations?

Modifiers of drug effects

(i.e., treatment effect heterogeneity)



Clínic Barcelona | Universitat de Barcelona

A FRAMEWORK FOR GENERALIZATION AND TRANSPORTATION OF CAUSAL ESTIMATES UNDER COVARIATE SHIFT

APOORVA LAL, WENJING ZHENG, AND SIMON EJDEMYR

Randomized experiments are an excellent tool for estimating *internally valid* causal effects with the sample at hand, but their *external* validity is frequently questioned. While classical results on the estimation of Population Average Treatment Effects (PATE) implicitly assume random selection into experiments, this is typically far from true in many medical, social-scientific, and industry experiments. When the experimental sample is different from the target sample along observable or unobservable dimensions (termed *covariate shift* in the causal learning literature), experimental estimates may be of limited use for policy decisions. We cast this as a sample selection problem and propose methods to re-weight the doubly-robust scores from experimental subjects to estimate treatment effects in the overall sample (=: *generalization*) or in an alternate target sample (=: *transportation*). We implement these estimators in the open-source package `causalTransportR`[1] and illustrate its performance in a simulation study and discuss diagnostics to evaluate its performance.

Methods

We observe n iid copies of $(\mathbf{X}_i, S_i, S_i A_i, S_i Y_i)_{i=1}^n$, where covariates $\mathbf{X}_i \in \mathbb{R}^p$, treatment $A_i \in \mathcal{A} := \{0, \ldots, K\}$, outcome $Y_i \in \mathbb{R}$, and selection indicator $S_i \in \{0, 1\}$ is a function of pre-treatment variables and is not affected by treatment. In other words, we observe $(\mathbf{X}_i, A_i, Y_i)_{i=1}^{N_1}$ for observations with $S_i = 1$ (henceforth the *study* sample $\mathcal{S}_1$), and only $(\mathbf{X}_i)_{i=N_1+1}^{N}$ for observations with $S_i = 0$ (henceforth the *external* sample $\mathcal{S}_0$). The *overall* sample is $\mathcal{S} := \mathcal{S}_1 \cup \mathcal{S}_0$.

Estimands. We write counterfactual means as $\phi = \mathbb{E}\left[Y^{a,S=1}\right]$ for generalizability and $\mathbb{E}\left[Y^a | S = 0\right]$ for transportability, and contrasts between such counterfactual means under any two treatment levels a, a' represent the average treatment effects (ATE). 'Standard' estimation of effects in the study sample under unconfoundedness is a well-studied and largely resolved problem (see [10] for a review). We study the generalization and transportation problems in the present paper. To this end, we make the following assumptions:

(1) Consistency / SUTVA : $Y_i = 1_{A_i=a} Y_i^a$
(2) Ignorability of Treatment: $Y^0, \ldots, Y^a \perp\!\!\!\perp A | X = x, S = 1$
(3) Overlap
    (a) Treatment overlap: $0 < \Pr(A = a | \mathbf{X} = \mathbf{x}, S = 1) < 1$

E-mail addresses: apoorval@stanford.edu, wzheng@netflix.com, sejdemyr@netflix.com.
Date: October 22, 2022.
Key words and phrases. experimentation, generalization, transportation, bridging.
[1]Available at https://github.com/Netflix-Skunkworks/causalTransportR

A Review of Generalizability and Transportability

Irina Degtiar and Sherri Rose
Harvard T.H. Chan School of Public Health and Stanford University

Abstract. When assessing causal effects, determining the target population to which the results are intended to generalize is a critical decision. Randomized and observational studies each have strengths and limitations for estimating causal effects in a target population. Estimates from randomized data may have internal validity but are often not representative of the target population. Observational data may better reflect the target population, and hence be more likely to have external validity, but are subject to potential bias due to unmeasured confounding. While much of the causal inference literature has focused on addressing internal validity bias, both internal and external validity are necessary for unbiased estimates in a target population. This paper presents a framework for addressing external validity bias, including a synthesis of approaches for generalizability and transportability, the assumptions they require, as well as tests for the heterogeneity of treatment effects and differences between study and target populations.

MSC 2010 subject classifications: Primary 62-2, Statistics Research exposition; secondary 62G05, Statistics Nonparametric inference Estimation.
Key words and phrases: generalizability, transportability, external validity, treatment effect heterogeneity, causal inference.

1. BACKGROUND

The goal of causal inference is often to gain understanding of a particular target population based on study findings. The true underlying causal effect will typically vary with the definition of the chosen target population. However, samples unrepresentative of the target population arise frequently in studies ranging from randomized controlled trials (RCTs) in clinical medicine to policy research (Bell et al., 2016; Kennedy-Martin et al., 2015; Allcott, 2015). In a clinical trial setting, physicians may be left interpreting evidence from RCTs with patients who have demographics and comorbidities that are quite different from

Irina Degtiar is a PhD candidate at the Department of Biostatistics, Harvard T.H. Chan School of Public Health, 655 Huntington Ave, Boston, MA 02115, USA (email: idegtiar@g.harvard.edu). Sherri Rose is an Associate Professor at the Center for Health Policy and Center for Primary Care and Outcomes Research, Stanford University, 615 Crothers Way, CA 94305, USA (email: sherrirose@stanford.edu).

- Que hagi fet l'itinerari de ciències de la salut del MESIO
- Especial motivació per la recerca biomèdica

rborras@recerca.clinic.cat

Gràcies per l'atenció!