Request for Permission to Include Previously Published or Co-Authored Material

Date: *17 May 2021*

From: *Professor Mark J. van der Laan*

To: Associate Dean of the Graduate Division for Academic Affairs
Graduate Services: Degrees, 318 Sproul Hall

Re: *Nima Hejazi, SID 22669337, Biostatistics PhD, nhejazi@berkeley.edu*

I am writing to ask for permission for the dissertation of the above named student to use co-authored and/or previously published material, as follows:

- N.S. Hejazi, M.J. van der Laan, H.E. Janes, P.B. Gilbert, and D.C. Benkeser. Efficient nonparametric inference on the effects of stochastic interventions under two-phase sampling, with applications to vaccine efficacy trials. *Biometrics*, 2020. Online at <https://doi.org/10.1111/biom.13375>.

- N.S. Hejazi and D.C. Benkeser. txshift: Efficient estimation of the causal effects of stochastic interventions in R. *Journal of Open Source Software*, 5(54), 2020. Online at <https://doi.org/10.21105/joss.02447>.

The student named was involved in the published/co-authored research in a role that makes it appropriate to claim a role as author of the work as an original contribution to research. The previously published work forms part of a larger coherent argument appropriate for the graduate degree for which the student is a candidate.

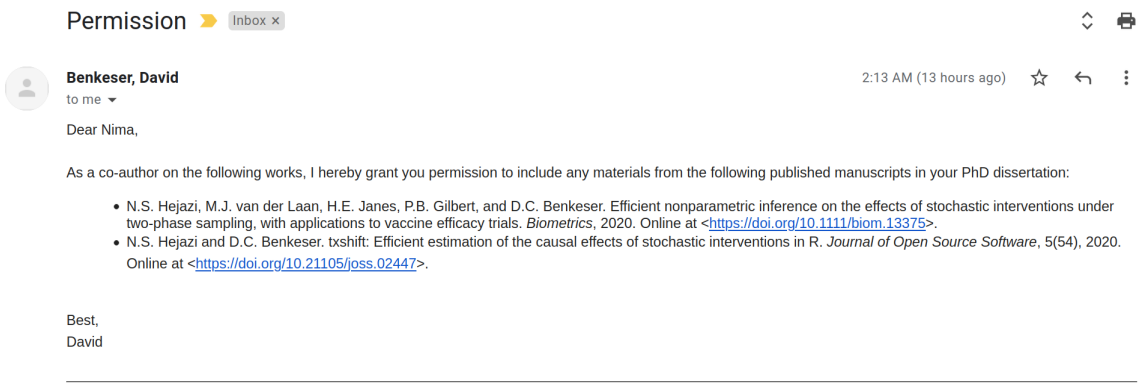
Statements agreeing to the use of co-authored work are included with this letter.

For the first listed publication, all secondary authors made contributions to the research ideas expressed in the manuscript as well as to its final written form; however, Nima Hejazi should be considered as the primary contributor to this line of research. Statements from all co-authors (myself, Holly Janes, Peter Gilbert, and David Benkeser), provided via email, are included below in this document.

For the second listed publication, the main research effort is a contribution of a software tool that provides access to techniques described in the first publication. This work was led by Nima Hejazi, with advising support offered by David Benkeser, whose permission, provided by email, is copied below in this document.

Permission statements to include the publications in the student's dissertation:

Permission granted (for both publications) by David Benkeser:



Permission

Inbox ×

Benkeser, David 2:13 AM (13 hours ago)

to me

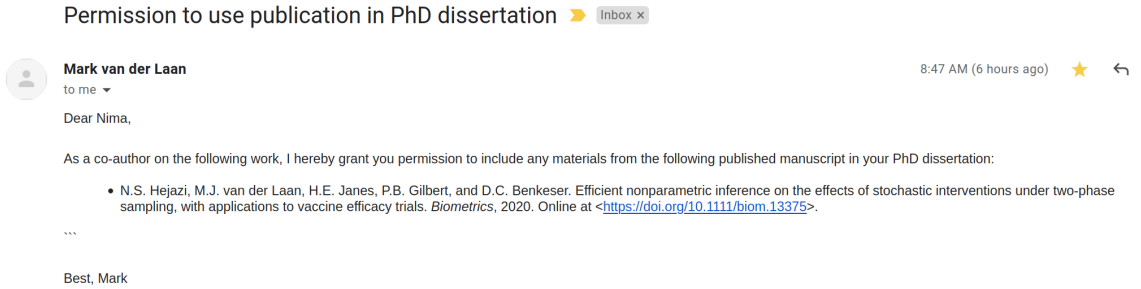
Dear Nima,

As a co-author on the following works, I hereby grant you permission to include any materials from the following published manuscripts in your PhD dissertation:

- N.S. Hejazi, M.J. van der Laan, H.E. Janes, P.B. Gilbert, and D.C. Benkeser. Efficient nonparametric inference on the effects of stochastic interventions under two-phase sampling, with applications to vaccine efficacy trials. *Biometrics*, 2020. Online at <https://doi.org/10.1111/biom.13375>.
- N.S. Hejazi and D.C. Benkeser. txshift: Efficient estimation of the causal effects of stochastic interventions in R. *Journal of Open Source Software*, 5(54), 2020. Online at <https://doi.org/10.21105/joss.02447>.

Best,
David

Permission granted (for first publication) by Mark van der Laan:



Permission to use publication in PhD dissertation

Inbox ×

Mark van der Laan 8:47 AM (6 hours ago)

to me

Dear Nima,

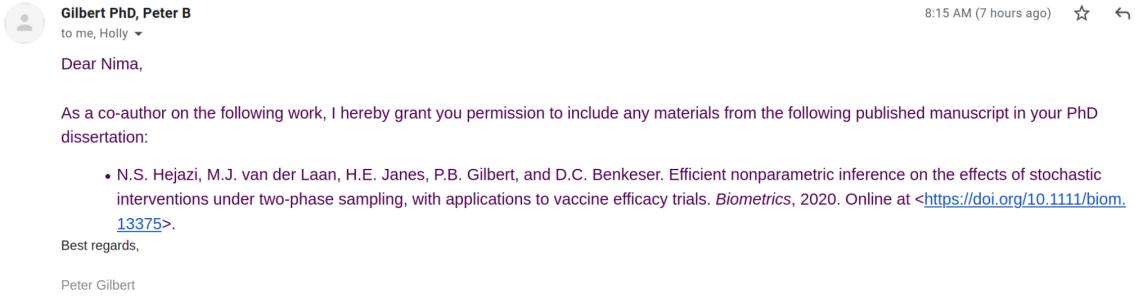
As a co-author on the following work, I hereby grant you permission to include any materials from the following published manuscript in your PhD dissertation:

- N.S. Hejazi, M.J. van der Laan, H.E. Janes, P.B. Gilbert, and D.C. Benkeser. Efficient nonparametric inference on the effects of stochastic interventions under two-phase sampling, with applications to vaccine efficacy trials. *Biometrics*, 2020. Online at <https://doi.org/10.1111/biom.13375>.

```

Best, Mark

Permission granted (for first publication) by Peter Gilbert:



**Gilbert PhD, Peter B** 8:15 AM (7 hours ago)

to me, Holly

Dear Nima,

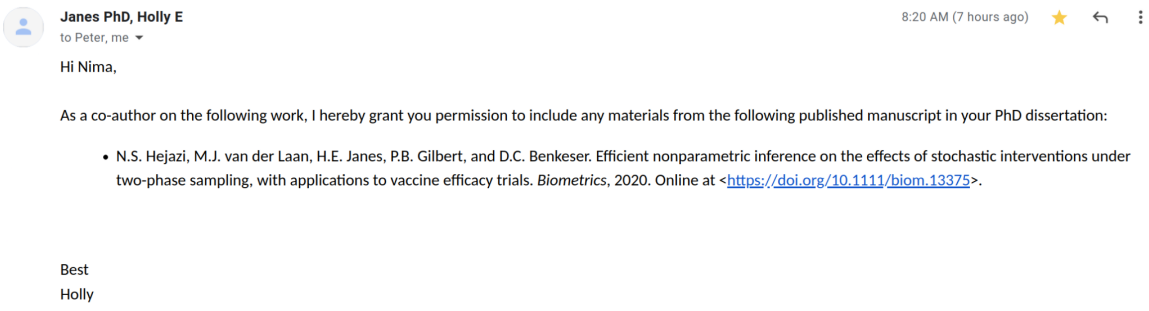
As a co-author on the following work, I hereby grant you permission to include any materials from the following published manuscript in your PhD dissertation:

- N.S. Hejazi, M.J. van der Laan, H.E. Janes, P.B. Gilbert, and D.C. Benkeser. Efficient nonparametric inference on the effects of stochastic interventions under two-phase sampling, with applications to vaccine efficacy trials. *Biometrics*, 2020. Online at <https://doi.org/10.1111/biom.13375>.

Best regards,

Peter Gilbert

Permission granted (for first publication) by Holly Janes:



**Janes PhD, Holly E** 8:20 AM (7 hours ago)

to Peter, me

Hi Nima,

As a co-author on the following work, I hereby grant you permission to include any materials from the following published manuscript in your PhD dissertation:

- N.S. Hejazi, M.J. van der Laan, H.E. Janes, P.B. Gilbert, and D.C. Benkeser. Efficient nonparametric inference on the effects of stochastic interventions under two-phase sampling, with applications to vaccine efficacy trials. *Biometrics*, 2020. Online at <https://doi.org/10.1111/biom.13375>.

Best
Holly

**Attached: First page of publication #1 and of publication #2.**
```

Received: 2 April 2020 | Revised: 2 August 2020 | Accepted: 1 September 2020

DOI: 10.1111/biom.13375

BIOMETRIC METHODOLOGY

Biometrics A JOURNAL OF THE INTERNATIONAL BIOMETRIC SOCIETY WILEY

Efficient nonparametric inference on the effects of stochastic interventions under two-phase sampling, with applications to vaccine efficacy trials

Nima S. Hejazi[1,2] | **Mark J. van der Laan**[3,4] | **Holly E. Janes**[5,6,7] | **Peter B. Gilbert**[5,6,7] | **David C. Benkeser**[8]

[1] Graduate Group in Biostatistics, University of California, Berkeley, California

[2] Center for Computational Biology, University of California, Berkeley, California

[3] Division of Epidemiology & Biostatistics, School of Public Health, University of California, Berkeley, California

[4] Department of Statistics, University of California, Berkeley, California

[5] Vaccine & Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, Washington

[6] Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington

[7] Department of Biostatistics, University of Washington, Seattle, Washington

[8] Department of Biostatistics & Computational Biology, Rollins School of Public Health, Emory University, Atlanta, Georgia

Correspondence
Nima S. Hejazi, Graduate Group in Biostatistics, Center for Computational Biology, University of California, Berkeley, CA 94720.
Email: nhejazi@berkeley.edu

Funding information
U.S. National Library of Medicine, Grant/Award Number: T32LM012417

Abstract

The advent and subsequent widespread availability of preventive vaccines has altered the course of public health over the past century. Despite this success, effective vaccines to prevent many high-burden diseases, including human immunodeficiency virus (HIV), have been slow to develop. Vaccine development can be aided by the identification of immune response markers that serve as effective surrogates for clinically significant infection or disease endpoints. However, measuring immune response marker activity is often costly, which has motivated the usage of two-phase sampling for immune response evaluation in clinical trials of preventive vaccines. In such trials, the measurement of immunological markers is performed on a subset of trial participants, where enrollment in this second phase is potentially contingent on the observed study outcome and other participant-level information. We propose nonparametric methodology for efficiently estimating a counterfactual parameter that quantifies the impact of a given immune response marker on the subsequent probability of infection. Along the way, we fill in theoretical gaps pertaining to the asymptotic behavior of nonparametric efficient estimators in the context of two-phase sampling, including a multiple robustness property enjoyed by our estimators. Techniques for constructing confidence intervals and hypothesis tests are presented, and an open source software implementation of the methodology, the `txshift` `R` package, is introduced. We illustrate the proposed techniques using data from a recent preventive HIV vaccine efficacy trial.

KEYWORDS
causal inference, stochastic interventions, targeted minimum loss estimation, two-phase sampling, vaccine efficacy

1 | INTRODUCTION

Ascertaining the population-level causal effects of exposures is a common goal in scientific research. Such effects can be formulated via summaries of the distribution of *counterfactual random variables*, which describe the values a measurement would have taken if a particular level of exposure were assigned to the unit. Often, the exposure

The Journal of Open Source Software

txshift: Efficient estimation of the causal effects of stochastic interventions in R

Nima S. Hejazi[1, 2] **and David Benkeser**[3]

1 Graduate Group in Biostatistics, University of California, Berkeley **2** Center for Computational Biology, University of California, Berkeley **3** Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University

DOI: 10.21105/joss.02447

Software

- Review
- Repository
- Archive

Editor: Marcos Vital

Reviewers:

- @klmedeiros
- @joethorley

Submitted: 20 May 2020
Published: 07 October 2020

Summary

Statistical causal inference has traditionally focused on effects defined by inflexible static interventions, applicable only to binary or categorical exposures. The evaluation of such interventions is often plagued by many problems, both theoretical (e.g., non-identification) and practical (e.g., positivity violations); however, stochastic interventions provide a promising solution to these fundamental issues (Díaz & van der Laan, 2018). The `txshift` R package provides researchers in (bio)statistics, epidemiology, health policy, economics, and related disciplines with access to state-of-the-art statistical methodology for evaluating the causal effects of stochastic shift interventions on *continuous-valued* exposures. `txshift` estimates the causal effects of modified treatment policies (or "feasible interventions"), which take into account the natural value of an exposure in assigning an intervention level. To accommodate use in study designs incorporating outcome-dependent two-phase sampling (e.g., case-control), the package provides two types of modern corrections, both rooted in semiparametric theory, for constructing unbiased and efficient estimates, despite the significant limitations induced by such designs. Thus, `txshift` makes possible the estimation of the causal effects of stochastic interventions in experimental and observational study settings subject to real-world design limitations that commonly arise in modern scientific practice.

Statement of Need

Researchers seeking to build upon or apply cutting-edge statistical approaches for causal inference often face significant obstacles: such methods are usually not accompanied by robust, well-tested, and well-documented software packages. Yet coding such methods from scratch is often impractical for the applied researcher, as understanding the theoretical underpinnings of these methods requires advanced training, severely complicating the assessment and testing of bespoke causal inference software. What's more, even when such software tools exist, they are usually minimal implementations, providing support only for deploying the statistical method in problem settings untouched by the complexities of real-world data. The `txshift` R package solves this problem by providing an open source tool for evaluating the causal effects of flexible, stochastic interventions, applicable to categorical or continuous-valued exposures, while providing corrections for appropriately handling data generated by commonly used but complex two-phase sampling designs.

Background

Causal inference has traditionally focused on the effects of static interventions, under which the magnitude of the exposure is set to a fixed, prespecified value for each unit. The evaluation