BIO STATISTICS COLLABORATION OF AUSTRALIA

Study Guide

CAUSAL INFERENCE (CSI)

Semester 1, 2024

Prepared by:

Jessica Kasza and Andrew Forbes
School of Public Health and Preventive Medicine,
Monash University

Lyle Gurrin
School of Population and Global Health, University of Melbourne,

MONASH University

THE UNIVERSITY OF MELBOURNE

Contents

Instructor contact details 2
Background 2
Context within the program 2
Prerequisites 3
Co-requisites 3
Unit summary 3
Workload requirements 3
Learning Outcomes 3
Unit content 4
Recommended approaches to study 4
Method of communication with coordinator(s) 5
Module descriptions 5
Unit schedule 7
Assessment 8
Submission of assessments and academic honesty policy 8
Late submission of assessments and extension procedure 9
Learning resources 9
Software 9
Feedback 9
Changes to CSI since last delivery, including changes in response to student evaluation 10
Acknowledgments 10

Causal Statistical Inference (CSI)
Semester 1, 2024

Instructor contact details

Jessica Kasza	Andrew Forbes
School of Public Health and Preventive Medicine, Monash University Tel: (03) 9903 0597	School of Public Health and Preventive Medicine, Monash University Tel: (03) 9903 0580
Email: jessica.kasza@monash.edu	Email: andrew.forbes@monash.edu

If you have any general BCA queries, please contact: Jacqueline Vaughan or Emily Higginson at the BCA Coordinating Office on 02 9562 5076/54 or email bca@sydney.edu.au

Background

Randomised clinical trials are well established as the gold standard for attributing causation to differences in outcomes between treatment groups. In contrast, results from observational studies are typically framed using associations between exposure and outcome, with arguments for causation largely based on qualitative criteria, such as those of Bradford-Hill. In 1974 a language for defining and a framework for assessing causation in observational studies based on counterfactual (or potential) outcomes became available, and in the mid-1990s graphical criteria for causation were developed in the form of causal diagrams. These general frameworks subsumed existing approaches, resolving long-standing paradoxes, and have been extended to cover many situations in the study design and analysis of both observational studies and randomised trials where there is a desire to make inferences that have a causal interpretation. This framework allows investigators to readily state the assumptions required for such inferences, a limited number of which may be directly assessable from the data. There are, however, key assumptions common to all designs and analysis approaches that necessitate external, expert knowledge to enable causal inference.

Context within the program

Regression modelling and "effect measures" are ubiquitous in the study of biostatistics and epidemiology, including within the BCA program. Interpretation of results are typically cautiously worded in terms of "association" rather than "causation". This unit involves careful study of assumptions required to make causal interpretations

from statistical associations, clarifies the objectives from regression modelling in terms of description, prediction or causal explanation, and presents modern methods designed to facilitate causal interpretations under appropriate assumptions.

Prerequisites

Epidemiology (EPI) plus either (a) Mathematical Background for Biostatistics (MBB) and Probability and Distribution Theory (PDT); or (b) Mathematical Foundations for Biostatistics (MFB); plus a multivariable regression modelling unit (e.g. LMR, RM1 or a regression unit in a health course such as Master of Public Health).

Co-requisites

Nil

Unit summary

This unit covers the core concepts and methods to enable the assessment of causation from randomised and observational studies. The foundational concepts and notation of counterfactual (or potential) outcomes are first presented to define precisely the meaning of the causal effect of a treatment or exposure, and the concept of the target trial is discussed. Causal diagrams, in the form of directed acyclic graphs (DAGs), are presented as a visual representation of a postulated underlying causal structure, together with the key properties and rules used to determine the likely existence of confounding, selection and other biases that prevent unbiased estimation of causal effects. The unit then proceeds to methods of estimating average causal effects. Propensity score methods for assessment of single time point treatments/exposures are studied in detail, followed by longitudinal designs, in which the treatments/exposure and their effects can vary over time. The study of mediation effects, which are explanatory methods for determination of mechanisms and pathways by which an exposure affects and outcome, are covered using language and methods of the potential outcomes framework. The unit concludes with a discussion of what to do when participants do not fully adhere to treatments, with an introduction to principal stratification and instrumental variables.
Throughout the unit, comparisons will be made with 'conventional' statistical methods, outlining the conditions under which a causal interpretation of inferences from the application of these methods can and cannot be made. Methods of analysis will use Stata and R software. Emphasis throughout will be placed on careful statements of assumptions required for valid causal inference and interpretation of results of data analyses in that light.

Workload requirements

The expected workload for this unit is 10-12 hours per week on average, consisting of guided readings, discussion posts, independent study and completion of assessment tasks.

Learning Outcomes

At the completion of this unit students should be able to:

1. Use counterfactuals (potential outcomes) to precisely define causal effects
2. Describe the differences between association and causation, and the fundamental assumptions required for causation
3. Construct causal diagrams and use them to identify potential sources of bias
4. Implement causal inference methods, using software, for single time point and longitudinal exposures, and for mediation analyses
5. Interpret results of analyses in light of the causal assumptions required
6. Effectively communicate results of causal analyses in language suitable for a clinical or epidemiological journal

Unit content

The unit is divided into 6 modules, summarised in more detail below. Modules 1, 2, 4 and 5 involve 2 weeks of study, Module 3 is 3 weeks long and the final module is 1 week long. Each module generally includes the following material:

1. Module notes describing concepts and methods, and including some exercises of a more “theoretical” nature.
2. Selected readings from the recommended textbook or published articles
3. One or more extended examples illustrating the concepts/methods introduced in the notes
4. Online videos or screencasts summarising the module’s content
5. An online ‘live’ tutorial, with date and time chosen by a Doodle poll of student preferences. The recording of the tutorial will be made available online for students unable to make it to the live session, or for later viewing.
6. Self-assessed quizzes

Study materials for all Modules are downloadable from the Canvas eLearning unit site. Assignments and supplementary material, such as datasets will also be posted to the unit site. Please note that we are not able to post copies of copyright material online (journal articles and book extracts)—for these you will have to rely on resources from your home university’s library.

Recommended approaches to study

Students should work through each module systematically, following the module notes and any readings referred to, and working through the accompanying exercises. You will learn a lot more efficiently if you tackle the exercises systematically as you work through the notes. You are encouraged to post any content-related questions to the Canvas discussion board, whether they relate directly to a given exercise, or are a request for clarification or further explanation of an area in the notes. You should also work through all of the computational examples in the notes for yourself.

Outline solutions to the exercises in each module (except those to be submitted for assessment, as described below) will be posted online at the midway point of the

allocated time for the module. This is intended to encourage you to attack the exercises independently (or via the eLearning site), and yet not make you wait too long to see the sketch solutions.

Make the most of this unit by engaging with coordinators and fellow students on the Discussion Board and in Tutorials. These are safe spaces to discuss the course material and related ideas and students are encouraged to make the most of them by engaging in respectful discussion.

Questions about Assignments should be directed to the coordinator in the first instance to avoid any Academic Honestly issues.

Method of communication with coordinator(s)

Questions about administrative aspects or course content can be emailed to the coordinator(s), and when doing so please use "CSI:" in the Subject line of your email to assist in keeping track of our email messages. Coordinator/s will be available to answer questions related to the module notes and practical exercises, and to address any other issues that require clarification. However, please note that instructors are not necessarily available every day of the week and you should expect that it may take a day or so to respond to questions (possibly longer over weekends and during breaks!).

We strongly recommend that you post content-related questions to the Discussion board on the unit site.

Relying on Canvas for content-related communication and problem-solving will enable other students to benefit from responses and indeed to respond themselves, and we try to encourage as much interaction as possible within the class through this medium.

Module descriptions

Below is an outline of the study modules, followed by a timetable and assessment description table

Each module is scheduled to begin on a Monday and conclude on the Sunday of the following week.

Module 1: Introduction to causal concepts (2 weeks)

Potential outcomes and counterfactuals
Definition of causal effects
Assumptions required for measures of association to have a causal interpretation
General concepts of estimation of causal effects in randomised controlled trials and observational studies

Module 2: Causal diagrams and directed acyclic graphs (2 weeks)

Drawing a causal diagram
Association versus causation in causal diagrams
Graph terminology, properties, and interpretation

Recognising biases in diagrams and using diagrams to guide design and analysis

Module 3: Time-invariant exposures and propensity scores (3 weeks)

Introduction to propensity scores – the 'why and what'
General operation of propensity scores
Propensity score adjustment methods: matching, stratification, regression, and inverse probability weighting - the 'how'
Guidelines for the estimation of causal effects using propensity scores

Module 4: Marginal structural models (MSM) for time-varying treatments (2 weeks)

Key characteristics of the source data
The problem with standard methods
Specifying marginal structural models and their estimation
Required assumptions
Issues specific to estimation of MSMs for survival outcomes

Module 5: Causal mediation analysis (2 weeks)

Critique of the standard "Baron and Kenny" approach used in the social sciences
Modern approaches to mediation
    Definition of controlled direct, natural direct, and natural indirect effects
    Estimation methods
Interventional effects
Multiple mediators

Module 6: Complier average causal effects (1 week)

Intention to treat versus per protocol effects
Principal stratification and causal effects when there is imperfect compliance
Instrumental variables and two-stage least squares regression

Unit schedule

Semester 1, 2024 starts on Monday 26 February

Week	Week commencing	Module	Topic	Assessment
1	February 26	Module 1	Introduction to causal concepts	
2	March 4	Module 1		
3	March 11	Module 2	Causal diagrams and directed acyclic graphs	
4	March 18	Module 2		Recorded Presentation due March 25 (20%)
5	March 25	Module 3	Time-invariant exposures and propensity scores	
6	April 1	Mid-semester break		
7	April 8	Module 3		
8	April 15	Module 3		Major Assignment 1 due April 22 (30%)
9	April 22	Module 4	Marginal structural models (MSM) for time-varying treatments	
10	April 29	Module 4		Short Assessment due May 6 (20%)
11	May 6	Module 5	Causal mediation analysis	
12	May 13	Module 5		
13	May 20	Module 6	Complier average causal effects	
			Major Assignment 2	Major Assignment 2 due Monday 3 June (30%)

Note that the week starting April 1 has no new material presented.

Assessment

Assessment will include one pre-recorded presentation task (worth 20%), one short written assessment worth 20% and two major assignments worth 30% each. The assessment tasks will be made available approximately two weeks prior to their due date. The **due date for submission of these assessment tasks is 11:59pm on the day indicated in the unit schedule table above.**

Assessment name	Assessment type	Coverage	Learning objectives	Weight
Recorded Presentation	Recorded presentation	Modules 1-2	1,2,3	20%
Major Assignment 1	Written assignment	Modules 1-3	1,2,3,4,5	30%
Short Assessment	Written assignment	Modules 1-4	1,2,3,4,5	20%
Major Assignment 2	Written assignment	Modules 1-6	1,2,3,4,5,6	30%
Online quizzes	Not assessed	Various modules	Various	Not assessed

You should submit material for assessments using the Assignments tool on the unit site. We strongly recommend you become familiar with equation typesetting software such as Microsoft's Equation Editor for algebraic work. You may submit neatly handwritten work, however please note that marks will potentially be lost if the solution cannot be understood by the markers due to unclear or illegible writing. Handwritten work should be scanned and collated into a single pdf file and submitted via the unit site. See the BCA Assessment Guide available on the Student Resources page on Canvas for specific guidelines on acceptable standards for assessable work.

Students are encouraged to discuss relevant topics in the Discussion Board. However, please avoid posting questions relating directly to assessable material. These should be emailed to the Unit Coordinator in the first instance.

Explicit solutions to assessable exercises should not be posted for others to use. Each student's submitted work must be clearly their own, with anything derived from other students' discussion contributions clearly attributed to the source.

Submission of assessments and academic honesty policy

All assessment material should be submitted via the relevant Assessment module in Canvas unless otherwise advised. Turnitin plagiarism detection is applied to all submissions. For detailed information, please see the BCA Assessment Guide, which includes links to the Academic Honesty policies at member universities. Please familiarise yourself with the procedures and policies at your home university. You will

need to indicate your compliance with the plagiarism guidelines and policy at your home university.

A special note regarding "contract cheating" sites: Unfortunately, there have been instances in the past of students using such websites to post assignment questions and receive solutions (usually for a fee). We have arrangements with these sites to identify the student posting questions or accessing the solutions, and such students will be referred to and face disciplinary processes at their home university.

Use of ChatGPT and other generative AI tools in assessment tasks: The assessment tasks in this Unit have been designed to be challenging, authentic and complex. Although individual assessment components may provide specific guidance regarding the use of generative AI tools (e.g., ChatGPT), successful completion of these components will require students to critically engage in specific contexts and tasks for which artificial intelligence will provide only limited support and guidance. In all cases, a failure to reference the use of generative AI may constitute student misconduct under the Student Code of Conduct of your University of enrolment. To successfully complete assessment tasks, students will be required to demonstrate detailed comprehension of their submission independent of AI tools.

Late submission of assessments and extension procedure

We adhere to standard BCA policy for late penalties for submitted work, i.e. a 5% deduction from the earned mark for each day the assessment is late, up to a maximum of 50%. Extensions are possible, but these need to be applied for (by email) as early as possible. The Unit Coordinator is not able to approve extensions beyond three days; for extensions beyond three days you need to apply to your home university, using their standard procedures.

Learning resources

There is no single prescribed text for the subject, but a number of reference books are suggested as background material (list below). The first book in the list is the one that we find closest to our materials. It is free to download, and therefore recommended.

Hernán MA, Robins JM (2020). *Causal Inference: What if.* Boca Raton: Chapman & Hall/CRC.

> Please note that section and page numbers for *Causal Inference: What if* in the module notes will refer to the version of this book dated January 2 2024.

Other suggested reading materials

Morgan S, Winship C. *Counterfactuals and Causal Inference. Methods and Principles for the Social Sciences*. 2nd edition, Cambridge University Press, 2015

Imbens G, Rubin D. *Causal Inference for Statistics, Social and Biomedical Sciences*. Cambridge University Press, 2015.

Pearl J, Mackenzie D. *The Book of Why: The New Science of Cause and Effect.* Basic Books, 2018. *[This is a conversational style book written for a general science audience rather than being a textbook on its own.]*

Software

For this subject you will need to have access to, and a working familiarity with, either Stata or R.

We expect most of you would be using Stata 18, 17, 16, or 15. We are not aware of any major differences between Stata versions that affect the material, but minor issues will be pointed out in discussion board postings. Importantly, whichever version you are using, please ensure that you have performed the online update to the latest update of that version. (Use the command **update query**)

Feedback

Our feedback to you:

The types of feedback you can expect to receive in this unit are:

- Formal individual feedback on submitted exercises assignments
- Responses to questions posted on Canvas

Your feedback to us:

One of the formal ways students have to provide feedback on teaching and their learning experience is through the BCA student evaluations at the end of each unit. The feedback is anonymous and provides the BCA with evidence of aspects that students are satisfied with and areas for improvement.

Changes to CSI since last delivery, including changes in response to student evaluation

CSI was last delivered in Semester 2, 2023, by Jessica Kasza; feedback from students on that delivery of the unit was very positive. Since then, we have reduced the number of assessment tasks from 6 to 4; some typos in the notes and exercise solutions have been corrected; some aspects of the notes have been clarified; and references to the Hernán and Robins text have been updated to align with the latest version.

Acknowledgments

We thank John Carlin, Margarita Moreno-Betancur, and Tibor Schuster for several helpful discussions and help with preparation of short course materials on which some of the notes are based.