BIO STATISTICS COLLABORATION OF AUSTRALIA

Study Guide

Causal Statistical Inference (CSI)

Semester 2, 2021

Prepared by:

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

Jessica Kasza and Andrew Forbes
Department of Epidemiology and Preventive Medicine,
Monash University

MONASH University

THE UNIVERSITY OF MELBOURNE

Contents

Causal Statistical Inference (CSI)
Semester 2, 2021

Instructor contact details

Koen Simons

School of Population and Global Health, University of Melbourne,

(03) 9035 7927

koen.simons@unimelb.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.

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.

Prerequisites

Mathematical Background for Biostatistics (MBB); Epidemiology (EPI); Probability and Distribution Theory (PDT)

Co-requisites

nil

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. Four modules involve 2 weeks of study, one with 3 weeks and the final module with 1 week. 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 published articles or textbooks.
3. One or more extended examples illustrating the concepts/methods introduced in the notes
4. Online videos or screencasts summarising the module’s content-related
5. Self-assessed quizzes

Study materials for all Modules are contained in your mail-out package or downloadable from the eLearning unit site. Assignments and supplementary material, such as datasets will be posted to the unit site. Please note that we are not able to post copies of copyright material (journal articles and book extracts)—for these you will have to rely on the hard copy mail-out or 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 eLearning, 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 on your own computer.

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

Method of communication with coordinator(s)

Questions about administrative aspects or course content can be emailed to the coordinator, 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 Discussions tool in the CSI area of BCA's eLearning site. In 2021 we are using the Learning Management system hosted by the University of Sydney. You may be familiar with the system from previous BCA units, and will receive any specific instructions on using the eLearning site this semester from the BCA Coordinating Office. There is also a "Getting Started" document available on the Student Resources page of the BCA website.

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. We will also use Canvas for posting all course materials although some of the core material (particularly selected readings, whose reproduction is subject to copyright considerations) is also sent out in paper form.

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. **The due date for submission of the required exercises from each module is 11:59pm on the day immediately following the completion of the module, as indicated below.**

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 2, 2021 starts on Monday 26 July

Week	Week commencing	Module	Topic	Assessment
1	July 26	module 1	Introduction to causal concepts	
2	August 2	module 1		
3	August 9	module 2	Causal diagrams and directed acyclic graphs	
4	August 16	module 2		Mod 1 & 2 Exercises due August 23
5	August 23	module 3	Time-invariant exposures and propensity scores	
6	August 30	module 3		
7	September 6	module 3		Mod 3 Exercises due September 13
8	September 13	module 4	Marginal structural models for time-varying treatment	
9	September 20	module 4		
	September 27	-	**mid semester break**	Major Assignment 1 Due October 4
10	October 4	-		Mod 4 Exercises due October 11
11	October 11	module 5	Causal mediation analysis	
12	October 18	module 5		Mod 5 Exercises due September October 25
13	October 25	module 6	Complier average causal effects	
14	November 1	-	Major assignment 2	**Major Assignment 2 Due November 14**

Note that the mid-semester break is one week, from September 27 to October 3 and that the first major assignment is due at the end of the semester break. As module 4 exercises are due the week after, no new material will be presented in the week of October 4.

Assessment

Assessment will include two written assignments worth 30% each, to be made available in the middle and at the end of the semester, and to be completed within approximately two weeks. These assignments will be posted on the eLearning site together with an online Announcement broadcasting their availability. In addition, students will be required to submit solutions to selected practical exercises (one from each module except Module 6), worth a total of 40%, by deadlines specified throughout the semester (see table below).

Assessment name	**Assessment type**	**Coverage**	**Learning objectives**	**Weight**
Module 1 & 2 exercises	Assignment	Module 1-2	1,2,3	10%
Module 3 exercises	Assignment	Module 3	1,2,3,4,5,6	10%
Major Assignment 1	Assignment	Modules 1-3	1,2,3,4,5,6	30%
Module 4 exercises	Assignment	Module 4	1,2,3,4,5,6	10%
Module 5 exercises	Assignment	Module 5	1,2,3,4,5,6	10%
Major Assignment 2	Assignment	Modules 1-6	1,2,3,4,5,6	30%
Online quizzes	Non-assessed	Various modules	various	-

You should submit material for assessment using the Assignments tool in eLearning. Where the work involves algebraic derivations that you find easier to complete by hand then you should scan your work to electronic form for submission. This handwritten work should be scanned and collated into a single pdf file and submitted via the eLearning site. In general, we prefer that your work be typed in Word, LaTeX or similar. We recommend you use software you are familiar with. The use of Microsoft's Equation Editor for algebraic work which is now much easier to use than previous versions.. See the BCA Assessment Guide document for specific guidelines on acceptable standards for assessable work.

Please note that the instructors will not answer all questions relating directly to the assessable material until after it has been submitted, particularly where answering the question provides a substantial hint, partial or full solution. We acknowledge the boundary between seeking clarification to help understand the question and seeking help to answer the question is not perfectly clear. Questions such as '*is … a typo and should it be … instead*' can posted on the eLearning

and/or emailed to the coordinator. Question such as 'i*s ... the correct value*' are clearly off-limits, however, issues with software are more ambiguous '*code ... fails with error ...*' may reveal your attempted solution even you didn't intend for that. If you are unsure which category your question belongs to, email is recommended. In addition, we encourage students to discuss the relevant parts of the notes among themselves, via eLearning, as long as **explicit solutions to assessable exercises are not posted for others to use**, and 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

You should submit all your assessment material via eLearning unless otherwise advised. The use of Turnitin for submitting assessment items has been instigated within unit sites. For more detail please see pages 3-5 the BCA Student Assessment Guide.
This guide will also be included in hardcopy in your package of notes.

The BCA pays great attention to academic honesty procedures. Please be sure to familiarise yourself with these procedures and policies at your university of enrolment. Links to these are available in the BCA Student Assessment Guide. When submitting assessments using Turnitin you will need to indicate your compliance with the plagiarism guidelines and policy at your university of enrolment before making the submission.

Late submission of assessments and extension procedure

The standard BCA policy for late penalties for submitted work is a 5% deduction from the earned mark for each day the assessment is late, up to a maximum of 10 days (including weekends and public holidays). 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 currently in draft version and free to download,and therefore recommended.

Hernán MA, Robins JM (2018). *Causal Inference*. Boca Raton: Chapman & Hall/CRC, forthcoming

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 16, 15 or 14. We are not aware of any major differences between Stata versions that affect the material, but minor issues will be pointed out in eLearnings postings. Importantly, whichever version you are using, please ensure that you have performed the online update to the latest updateof 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
- Feedback from non-assessed online quizzes
- Responses to questions posted on Blackboard

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 2020, by Jessica Kasza, Andrew Forbes and Lyle Gurrin. No changes were made.

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.