

## Radix

radix is a new R package, that is based on the [Distill web framework](#) to bring scientific or technical writing using R Markdown that is native to the web. The [Distill web framework](#) is used in the [Distill Research Journal](#) which publishes research on machine learning.

Getting started on radix is easy, especially if you know how to use R Markdown and use the RStudio IDE. The best way to get started is to follow the instruction in this website article [here](#).

## Blogdown

Blogdown is an R package that uses [Hugo](#) in the backend to generate the website. blogdown also can use [Jekyll](#) or [Hexo](#) as the generator in place of Hugo however some features are only supported in Hugo. While the package name signifies that the primary motivation is for

blogging, blogdown is not limited to blogs and is flexible to create any website.

Getting started on blogdown is slightly involved in the beginning but it becomes relatively easy once you have become accustomed to the workflow. It is easier if someone initially demonstrates you how to do it however if you do not have anyone around you to demonstrate it, you may try following the set of slides referenced here. More comprehensive details about the blogdown package is best explained in the [book](#) by Xie et al. (2017).

## Conclusion

So, radix or blogdown? This is ultimately up to you. One thing I'm sure is that writing articles using R Markdown documents has aided greatly in my scientific writing.

# STRENGTHENING ANALYTICAL THINKING FOR OBSERVATIONAL STUDIES (STRATOS): INTRODUCING THE CAUSAL INFERENCE TOPIC GROUP (TG7)

Ingeborg Waernbaum<sup>1</sup>, Bianca De Stavola<sup>2</sup>, Erica Moodie<sup>3</sup>, Saskia le Cessie<sup>4</sup>, Els Goetghebeur<sup>5</sup> on behalf of STRATOS TG7

<sup>1</sup> Department of Statistics, USBE, Umeå University, SW.

<sup>2</sup> GOS Institute of Child Health, University College London, U.K..

<sup>3</sup> Department of Epidemiology, Biostatistics and Occupational Health, McGill University, CA.

<sup>4</sup> Department of Clinical Epidemiology and Department of Biomedical Data Sciences, Leiden University, NL.

<sup>5</sup> Department of Applied Mathematics, Computer Science and Statistics, Ghent University, BE.

When the STRATOS initiative engaged in its mission to support practicing statisticians in their access to state of the art methods, 'Causal Inference' became one of the initial topic groups. It is not only concerned with important scientific questions at the heart of most disciplines and directed towards interventions that change the world [1, 2], but there was also a surge of new methodology over the past decades. Methods vary widely in their approach and involve challenges at the conceptual as well as the technical level. Today, adapted software facilitates calculation of causal estimators. For their causal interpretation however, they rely on more than the usual modeling constraints. Underlying assumptions are

often cast in terms of *potential outcomes* and involve parallel worlds where the observation units are subject to different exposure levels than those that were observed, with a shift in expected potential outcomes as a result [3]. Causal parameters (estimands) are then defined as contrasts between potential outcome distributions, e.g., the expected difference between potential outcomes under two different levels of exposure: the observed exposure and the complementary exposure that was not observed. Unfortunately, plausibility of the assumptions needed for consistent estimation can not be evaluated in terms of observed data alone.

An important challenge presented by methods for causal inference lies in formulating practical questions in terms of targeted causal estimands which embody the scientific interest. Another challenge lies in translating assumptions into plausible constraints for the subject matter [4]. Hence, causal analyses are characterized by the special precautions required for the process of finding balance between the causal questions and observed data and assumptions.

There are very helpful papers with explanation on specific types of methods [5,6], but little guidance on how to choose the most appropriate method for a specific question. This is where TG7 aims to contribute. We set out to give an overview of principles of causal inference and causal effect estimation, emphasizing logic and intuition. We

develop the approach from carefully specifying the causal question in context, over making assumptions explicit in the subject matter causal framework, to constructing the various estimators. We describe classes of estimators under both the *no unmeasured confounding* and *instrumental variable* [7,8] assumptions.

Given a sufficient set of confounders [9], a balanced comparison of treatment groups can be achieved by adjusting for them directly through stratification, matching, regression modeling or inverse probability weighting, or indirectly by conditioning on the treatment propensity score. One then averages these conditional outcomes per treatment over well selected covariate distributions, representing the target population which may or may not coincide with the study population, or (un)treated subpopulations. This leads to estimators of corresponding estimands, most relevant to inform specific policy questions. We plead for explicit reporting in published reports of which estimand is envisaged and estimated for what population mix or subset.

To facilitate comprehension and enable comparison of advantages and disadvantages of various estimators, we developed a 'simulation learner'. This simulation tool is built on an existing data structure, but generates, in addition to the 'observed' data, potential outcomes under a range of possible exposures or treatments for the same observation units. The simulated dataset as well as

the R-code that generated it can be found on the TG7 website [www.ofcaus.org](http://www.ofcaus.org). The website also includes slides, practicals and solutions with code (R, Stata, SAS) from courses we taught on the topic. The material from the courses is currently being written as a tutorial paper.

TG7 recently welcomed two new members: Vanessa Didelez (Leibniz Institute for Prevention Research and Epidemiology – BIPS and University of Bremen) and Martin Wolkewitz (Institute of Medical Biometry and Statistics, Division Methods in Clinical Epidemiology, University of Freiburg). The TG also have two affiliated members: Niels Keiding (Section of Biostatistics, University of Copenhagen) and Michael Wallace (Statistics and Actuarial Science, University of Waterloo).

We aim to continue our work on the initial guidance of point exposures to extensions for longitudinal data and dynamic treatments [10, 11]. Projects related to more specialized areas are also in a planning phase. Practical application of the principled causal inference methods will be confronted with all the usual complications of variable selection, missing data, survival type outcomes, measurement error, high dimensional data and more. We are therefore looking forward to interacting with our colleagues from these topic groups to further guidance and methods development in this area.

#### References

[1] Pearl, J. Causality. (2009). Cambridge university press.

[2] Hernán M.A., Robins JM (2018). Causal Inference. Boca Raton: Chapman & Hall/CRC, forthcoming.

[3] Rubin D.B.(1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*. 66:688 – 701

[4] Goetghebeur, E., De Stavola, B., Moodie, E. Waernbaum, I. and le Cessie, S. (2016). "The Statistics of Tragedy" or "the Tragedy of Statistics"? . Letter to *Significance*, p.46, February 2016.

[5] Bang, H. and Robins, J.M.(2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* 61: 962-973.

[6] Austin, P.C., and Stuart, E.A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in Medicine* 34: 3661-3679.

[7] Vansteelandt, S. and Goetghebeur E. (2003). Causal inference with generalized structural mean models. *Journal of the Royal Statistical Society, Series B*, 65: 817 – 835.

[8] Boef, A., le Cessie, S., Dekkers, O. et al. (2016). Physician's prescribing preference as an instrumental Variable. *Epidemiology*, 27: 276--283.

[9] de Luna, X., Waernbaum, I. and Richardson, T.S. (2011). Covariate selection for the non-parametric estimation of an average treatment effect. *Biometrika* 98:861-875.

[10] Chakraborty, B. and Moodie, E.E. (2013). Statistical meth-

ods for dynamic treatment regimes. *New York: Springer*.

[11] Daniel, R., Cousens, S., De Stavola, B., Kenward, M. and Sterne, J. (2013). Methods for dealing with time-dependent confounding. *Statistics in Medicine* 32: 1584-1618.

## Mathematical Riddle

The Solution (in integers) to the last issue's Mathematical Riddle is:

The graphic displays three mathematical equations using symbols:

$$\frac{\text{Star}(15)}{5} + \frac{\text{Triangle}(27)}{3} = 12$$

$$\text{Square}(18) - \text{Star}(15) = \frac{\text{Triangle}(27)}{\text{Square}(9)}$$

$$\frac{\text{Square}(18)}{36} = \frac{\text{Heart}(1)}{2}$$

In the top right corner, there is a logo for "אלברטו" (Alberto) with the text "אתגרי למידה בראש סוב" and a colorful star icon.

The seven individuals who answered correctly were:

1. David Baird (VSN (NZ) Limited New Zealand)
2. Cole Tim (Population Policy and Practice Program London, UK)
3. Ron Mowers (Retired from Syngenta (Seed/AG company))
4. Van Burgel (Western Australian Government Department of Primary Industries and Regional Development, Albany)
5. Naoki Ishizuka (Cancer Institute Hospital Japanese Foundation Cancer research, Japan)
6. Chen Michael (Marvel Company, Israel)
7. Josef Levi (Clalit Research Unit, Israel)