

# STRengthening Analytical Thinking for Observational Studies (STRATOS): Update on the Missing Data Topic Group (TG1)

James Carpenter (1) and Katherine J Lee (2)

(1) Department of Medical Statistics, London School of Hygiene & Tropical Medicine, and MRC Clinical Trials Unit at UCL, London. Email: [james.carpenter@lshtm.ac.uk](mailto:james.carpenter@lshtm.ac.uk)

(2) Murdoch Children's Research Institute, University of Melbourne, Australia. Email: [katherine.lee@mcri.edu.au](mailto:katherine.lee@mcri.edu.au)

The aim and scope of the STRATOS missing data topic group were described in the 2017 Biometric Bulletin 34, 11-13. In the four years since then, the group has conducted a number of projects and also developed stimulating plans for future activities. We are pleased to take this opportunity to present an update from our group.

## 1. The TARMOS framework

In response to the issues we raised in our previous Biometric Bulletin article, we developed a framework for the Treatment And Reporting of Missing data in Observational Studies (the TARMOS framework), with the aim of (1) giving experienced analysts the confidence to apply missing data methodology routinely, and (2) giving analysts with less formal statistical training a practical overview of the issues and how to address them. The resulting article [1] was published open-access in the Journal of Clinical Epidemiology earlier this year.

The paper gives a step-by-step illustration of the framework's application to a dataset from the Avon Longitudinal Study of Parents and Children, exploring the effect of teenage smoking on educational attainment, and also presents a flowchart for selecting an appropriate method to handle missing data. In writing, we seek to encourage researchers to think systematically about missing data, and put a strong emphasis on pre-planning the statistical analysis and transparency in reporting results. We hope that using the TARMOS framework will increase the reproducibility of research findings.

Needless to say, getting a group of statisticians to agree on such a framework was quite a task – and we are therefore delighted that readers' have received the paper positively, with one tweet commenting 'This is a crystal clear paper about how to handle missing data in observational studies'.

## 2. A comparison of complete case analysis, inverse probability weighting and multiple imputation

Analysts only have to have a brief interaction with the missing

data literature to become aware of the three broad methodologies for handling missing data: (i) restricting the analysis to those with complete data (complete-case (CC) analysis, the usual default in software); (ii) re-weight the observed data to make it representative of the full data set (inverse probability weighting, IPW) or (iii) use multiple imputation (MI) to provide multiple 'complete' data sets, each of which can be analysed as originally intended with the results being combined for final inference using Rubin's rules.

However, few analysts are aware of the relative suitability of these techniques in specific situations.

Led by Rod Little, the topic group has written a manuscript to clarify this, with social, rather than medical, analysts as the primary audience. We address the following questions (a) when should we use MI rather than IPW or CC, and (b) if MI is not appropriate, when should IPW be used instead of CC? The answers are illustrated using a freely-available cohort of UK educational data. The manuscript has received a favourable first review, and we hope it will be published in the first half of 2022.

## 3. Symposium at the World Congress of Epidemiology, September 2021.

A highlight of our dissemination activity was the Topic Group's symposium at the recent World Congress of Epidemiology. This featured four presentations, three of which were given by members of the topic group, on selection of appropriate methods for analysis of partially observed data, practical methods for exploring the sensitivity of conclusions to the untestable assumptions about the missing data mechanism and our framework for improving the reporting of analyses affected by missing data.

## 4. Is multiple imputation always the answer?

Finally, our TG contributed to the discussions that led to a stimulating paper exploring whether multiple imputation is always the answer for missing data in statistical analysis [2].

**Looking forward**, we are excited about the momentum building behind our key projects, specifically

### 1. Tutorial for analysis of data that are missing not at random

This work is focusing on how to frame and report sensitivity analyses when it is suspected that data may be missing not at random. We seek to address the same audience as for the TARMOS paper in the Journal of Clinical Epidemiology. We will be advocating for researchers to use a pattern mixture approach, implemented via multiple imputation, focusing on the one or two key variables with non-trivial proportions of missing data in turn. We will complement the article with code for the examples in R and Stata.

### 2. Joint project with the Initial Data Analysis topic group (TG3)

This paper, aimed at researchers who have only received limited formal statistical training, will use a case study to present the key elements of initial data analysis when there is a non-trivial proportion of missing data. The aim is to provide a structured approach which will substantially improve the reproducibility of research, clearly differentiating between hypothesis testing and hypothesis generation.

3. Joint project on missing data in causal models with the Causal Inference topic group (TG7)

At the last IBC 'in Seoul', James Carpenter contributed a presentation on recent developments in handling missing data in causal inference settings, such as marginal structural models, highlighting the assumptions the various methods make, and the importance that they are plausible in order for inference to be reliable. We are developing work with the Causal Inference topic group based on this presentation in order to ensure that a wide pool of analysts are aware of these challenges and how to address them.

4. Towards structured reporting of analyses with missing data

At the recent STRATOS symposium, a key theme that emerged was the importance of structured reporting in order to help both researchers and journal editors promote reproducible research. Clearly this has implications when analysing partially observed data, because handling missing values entails a substantial number of extra models, each with its associated modelling decisions. We will explore whether there is scope for the group to make a substantive contribution in this area.

### Website

We are also delighted that we have been able to launch a [website for our topic group](#), which contains details of the activities of the group and its membership.

### Membership

Members of the Missing Data Topic Group are: James Carpenter (co-chair), LSHTM, UK; Katherine Lee (co-chair), Murdoch Children's Research Institute, Melbourne, Australia; Melanie Bell, University of Arizona, USA; Rosie Cornish, University of Bristol, Bristol, UK; Els Goetghebeur, Ghent University, Belgium; Joseph Hogan, Brown University, USA; Rod Little, University of Michigan, USA; Rheanna Mainzer, Murdoch Children's Research Institute, Melbourne, Australia; Andrea Rotnitzky, Harvard University, USA; Kate Tilling, University of Bristol, Bristol, UK;

**In conclusion**, we'd like to emphasise that we are keen to approach issues regarding missing data in an inclusive way – and so are keen to stimulate interactions with other researchers. In particular, we welcome comments and suggestions on our past, current and planned work, and also suggestions for illustrative datasets (which should be publicly available). From time to time we will need additional help with various projects, and James and Kate would be happy to hear from readers who are interested in contributing in this way.

### References

[1] Lee KJ, Tilling K, Cornish RP, Little RJ, Bell ML, Goetghebeur E, Hogan JW, Carpenter JR for the STRATOS initiative (2021): Framework for the Treatment And Reporting of Missing data in Observational Studies: The TARMOS framework. *Journal of Clinical Epidemiology*, 134, 79-88. DOI: <https://doi.org/10.1016/j.jclinepi.2021.01.008>

[2] Rachael A Hughes, Jon Heron, Jonathan A C Sterne, Kate Tilling, Accounting for missing data in statistical analyses: multiple imputation is not always the answer, *International Journal of Epidemiology*, Volume 48, Issue 4, August 2019, Pages 1294–1304, <https://doi.org/10.1093/ije/dyz032>

# Anniversary Corner – An Update on Plans for the Society's 75th Anniversary

With the start of a new year, we welcome a new column to the *Biometric Bulletin* that is focused on the upcoming 75th anniversary of the Society. Look for this column in all 2022 issues of the *Bulletin*, starting with this one.

You may not have even realized that the Society was about to celebrate this major milestone... its 75th anniversary, this coming September. That's because planning for the celebration is still ongoing. But very soon, we'll share more about how leaders, members, Regions and friends can all get involved in the celebration! Of course, every great celebration involves a few good stories. And the Society fits into our own personal stories in many different ways. To start off our celebration, we invite you to share your stories with us, then with the rest of the membership. Look for a separate announcement on this initiative soon.

Some of you reading this column might be inspired to share some of your own reflections about the society and its 75th birthday. We welcome you to reach out by sending a message to [ibs@biometricsociety.org](mailto:ibs@biometricsociety.org). We'd be happy to consider including your thoughts and ideas about the anniversary in our planning or in future *Bulletin* issues or other IBS publications (with your permission of course).

Our Past Presidents are also discussing how the Society has been shaped by its members over the years, and vice versa. We look forward to offering reflections on the Society's first 75 years online and in social media between now and September, but also "live" in Riga, Latvia during a special anniversary session planned for IBC 2022.

And we've already announced in this issue of the *Bulletin* the fantastic news that one of our Past Presidents, Lynne Billard (ENAR), is working with a professional publisher to bring the first 75 years of the Society's existence to readers in book form. As members, you can be sure that more information on this initiative will be forthcoming.

Shortly after our new website was launched, the Society took the opportunity to refresh its "[history](#)" page to better reflect some of the major milestones and moments. As you prepare for the 75th anniversary of the Society, we encourage you to check out this page, if you haven't already.

There's much more to come. But in 75 years, haven't we already proven as much? We look forward to celebrating with you in 2022!

Respectfully,  
Peter Doherty, CAE  
Executive Director"