NIHR Statistics Group – Imaging Studies section

27th April 2015

Scarman House Conference Centre, University of Warwick

Statistical Issues in Clinical Trials of Inflammatory Bowel Disease using Endoscopy

Introduction

Tom Fanshawe and Sue Mallett opened the meeting by giving an overview of common statistical issues in imaging studies and of the aim of the group and the meetings which are planned to occur twice a year focusing on specific methodological topics and different imaging modalities.

This meeting focused on four important statistical considerations:

- Sample size considerations in imaging studies
- Inter-rater agreement and the effect of subjective image assessment
- Reproducibility of endpoint assessment
- Agreement and variability between repeated imaging time-points

Clinical Background

The first session was about Centralised Reading of Endoscopy for Clinical Trials in Inflammatory Bowel Disease and was presented by Dr Vip Jairath, a NIHR Clinical Trials Fellow at the Oxford Clinical Trials Research Unit and Clinician in Gastroenterology and Inflammatory Bowel Disease at the John Radcliffe Hospital in Oxford.

He first introduced the different types of images used: still images (X-rays), dynamic images (MRI, PET-CT) and video images (angiography, endoscopy); and the idea of centralised reading of imaging, which occurs when experts located off-site independently assess images, blinded to information about the treatment that was given. He also presented some guidance and guidelines for evaluating imaging endpoints, such as the FDA's Guidance for Clinical Trial Imaging Endpoint Process and the European Medicines Agency's guidelines on clinical evaluation of diagnostic events.

He then focused on inflammatory bowel disease (IBD), how it is diagnosed (endoscopic assessment followed by confirmatory biopsy) and the aims and nature of treatment. As the typical goals for this disease are improving quality of life and control of symptoms, the endpoints of interest include patient-reported outcome measures (PROMs), biochemical indices, endoscopy findings, imaging and histology.

Measures used in this area include:

- Crohn's Disease Activity Index (http://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0048967) which in some studies has shown poor correlation of symptoms with inflammation.

- Mayo Disease Activity Index for Ulcerative Colitis (http://www.nature.com/ajg/journal/v104/n6/fig_tab/ajg200983t1.html). This has grades from 0 to 3 (from less to more severe) in 4 domains (stool frequency, rectal bleeding, endoscopy findings and physician's global assessment), some of which are subjective in nature.

Using a phase 3 study comparing the anti-inflammatory drug 5-ASA versus placebo as an example, Dr Jairath highlighted the problem of inter-observer variability and reduction of bias in trial inclusion criteria by using endoscopy to identify patients with a certain degree of disease.

Statistical considerations

A key issue in ulcerative colitis drug intervention trials is the variation of outcome response in the placebo arm. Several studies have shown that factors contributing to this heterogeneity include the duration of patient follow up, the number of follow up visits, baseline disease severity, the lack of standardised outcome measurements and bias in disease ascertainment. This variation in response can impact the power of a trial.

Another issue is disagreement between clinicians and how this is reflected in the data and the level of bias that may be introduced. The level of expertise of clinicians adds further complexity together with the often subjective nature of the assessment of the images. In this clinical area, there is no gold standard for comparison beyond consensus-based methods, which are not well described in the literature in this area.

Discussion

After this talk, Susan Dutton led a structured discussion where attendees were given four scenarios, representing the four statistical considerations listed above:

1. **Sample size considerations** – methods such as the Bland-Altman method for agreement between the gold standard and a new modality (Bland JM & Altman DG (1986), Statistical methods for assessing agreement between two methods of clinical measurement. Lancet: 327; 307-310) recommend a sample size of at least 100 participants as a rule of thumb (https://www-users.york.ac.uk/~mb55/meas/sizemeth.htm). For imaging, the gold standard may be histopathology or another imaging modality. However, there are imaging types which can be very expensive so the number of participants who can be treated is very limited (e.g. the F-MISO PET scan). Expense therefore could affect statistical power for these studies, thus affecting the development and use of the new techniques.
2. **Independent central review of endoscopy imaging as an eligibility criterion** – As a means to validate imaging outcomes, an independent reviewer can be used to quantify the level of agreement between clinicians. Some clinicians may be over-optimistic to enrol participants and may recruit participants that are later found to be ineligible. If

reviewers have different opinions on a participant's eligibility, the decision on whether the participant should be recruited is not clear, and may require consensus panel assessment. Implications on power and analysis were discussed.

3. **Independent central review of endoscopy imaging as an outcome criterion** - Unresolved issues discussed included the number of central reviewers needed, the associated cost implications, and methods to account for disagreements between the different reviewers in the analysis. A drawback of central review is the difficulty in making sure that information is presented to the central reader in the same way that would be assessed by an on-site observer.
4. **Repeatability** – Often, images will be taken at multiple time points. It is difficult to ascertain whether the images are taken at the same location in the body in each scan. Unresolved issues discussed included how to increase the accuracy of true changes over time and how to account for this in the analysis, and how to identify measurement error.

Sponsorship

This event was sponsored by the University of Warwick Department of Statistics, a University of Warwick Mathematics and Statistics WAMP Platform Grant, and the West Midlands Health Informatics Network.

Acknowledgements

Thanks to Tom Nichols for arranging the venue and sponsorship, and to Susan Dutton, Victoria Strauss, Dr Vip Jairath and the Imaging Studies section steering group for developing and advising on the scientific content.

Future meetings

The next meeting is planned for October/November 2015 – details to be announced.