FRIDAY, 14 JUNE 2019

15:30:00 – 17:00:00

Missing data in clinical research: to impute or not to impute, that is the question

SP0156 MISSING DATA: IS IT ALL THE SAME?

Stian Lydersen. Norwegian University of Science and Technology, Department of mental health, Trondheim, Norway

Introduction: In most studies, there are “holes” in the data set, such that data are missing partly or completely for some of the subjects. This results in reduced statistical power. More seriously, missing data may cause biased results, since data are usually not missing completely at random.

Missing data mechanisms: In which way, and to what extent, does the probability that data are missing, depend on observed and/or unobserved data values? This is called the missing data mechanism, and is important for an appropriate choice of analysis method. Three types of missing data mechanisms have been defined, see for example (Sterne et al. 2009).

MCAR: *Missing completely at random.* There are no systematic differences between the missing values and the observed values. The probability that values are missing does not depend on any of the data, observed or unobserved.

MAR: *Missing at random.* The probability that values are missing may depend on observed data values, but not on unobserved data values. For example, some variables are missing more frequently for patients with higher age. Data are missing at random *conditionally* on observed data.

MNAR: *Missing not at random.*

Even after the observed data are taken into account, systematic differences remain between the missing values and the observed values: The probability that values are missing depends on unobserved data values as well.

It is possible to distinguish between MCAR and MAR by inspecting the data at hand. But it is never possible to determine whether data are MNAR from your data.

Methods for handling missing data

Some commonly used methods, from the less to the more complex ones, are listed below, with an indication of when they give unbiased estimates:

- Complete case analysis (disregarding cases with partially missing data) (MCAR)
- Single imputation methods (sometimes under MAR but underestimates uncertainty)
- Multiple imputation (MAR)
- Full information maximum likelihood (MAR)
- Linear mixed models in longitudinal studies (MAR)

In longitudinal studies, Last observation carried forward (LOCF) is a simple method for imputing missing values. But it is not unbiased under any sensible assumptions, and should not be used (Lydersen 2019).

Reporting: In general, report the amount of missing data in the different variables, and how this was handled in the analysis. This is recommended in the Consort Guidelines (Moher et al. 2010) as well as the STROBE Statement (STROBE 2014), see also (Lydersen 2015)

REFERENCES:

- [1] Lydersen, S. 2015. Statistical review: frequently given comments. *Ann. Rheum. Dis.*, 74, (2) 323-325. Lydersen, S. 2019. Last observation carried forward - simple but not recommended. In press, *Tidsskrift for Den norske legeforening*
- [2] Moher, D., Hopewell, S., Schulz, K.F., Montori, V., Gotzsche, P.C., Devereaux, P.J., Elbourne, D., Egger, M., & Altman, D.G. 2010. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ*, 340, c869
- [3] Sterne, J.A., White, I.R., Carlin, J.B., Spratt, M., Royston, P., Kenward, M. G., Wood, A.M., & Carpenter, J.R. 2009. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*, 338, b2393
- [4] STROBE Statement 2014. Strengthening the reporting of observational studies in epidemiology.

Disclosure of Interests: None declared

DOI: 10.1136/annrheumdis-2019-eular.8530

SP0157

PRO SESSION: IMPUTATION IS A POWERFUL ANALYTICAL TOOL, WHEN APPROPRIATELY USED

Loreto Carmona. Instituto de Salud Musculoesquelética, Research, Madrid, Spain

Analysing a study with missing data very likely will yield biased results. Of course, the first principle is to avoid missing data by designing the study and the data collection very thoughtfully. If the study is already done and data collected, then one needs to study how important, and what type of missingness is affecting our data. Simple *ad hoc* methods, such as LOCF, or imputing the mean or other value are generally invalid. Multiple imputations need additional work and expertise, but generally yield better results, closer to reality.

Many statistical programmes have macros available for imputing missing values and analysing imputed data. We will review how these work. Understanding what multiple imputations do will make the decision clearer.

Disclosure of Interests: Loreto Carmona Grant/research support from: Abbvie, Actelion, Astellas, BMS, Eisay, Gebro Pharma, Grünenthal, Leo Pharma, Lilly, MSD, Novartis, Pfizer, Roche, Sanofi-Aventis and UCB Pharma, Paid instructor for: Novartis

DOI: 10.1136/annrheumdis-2019-eular.8440

SP0158

CON SESSION: PREPARE YOUR DATA COLLECTION WELL: IMPUTATION DOES NOT REPLACE ‘REAL DATA’

Maxime Dougados.

The recently proposed sophisticated statistical methods to impute missing data are very attractive.

For the clinical researchers in charge of data collection, it might be tempting to “neglect” the quality of data collection in favor of the time spent by one database manager/statistician to impute missing data.

This presentation will be a plea in favor to make all the possible efforts to collect appropriately the data in clinical research studies/trials. In particular, emphasis will be placed on the importance of writing a specific section in any research protocol describing the techniques used to avoid these missing data and also the a priori decided technic that will be used to manage the missing data. One point of importance in this section: to specify in advance the maximum percentage of missing data over which no imputation can be made.

Two different situations:

- Cross-sectional studies. Any item (for example HLA B27, MRI SIJ inflammation positivity) should be considered as mandatory or not before running the study.

- Longitudinal studies: any longitudinal protocol should include a section detailing the technic to appropriately follow the patients in order to avoid the “lost of follow-up patients”. This is particularly important in longitudinal follow-up of patients suffering from a recent onset disease (e.g. early synovitis cohorts, early inflammatory back pain cohorts) since it is well known that the “lost of follow-up” have specific baseline characteristics suggesting a less severe disease. It is also important to decrease at a minimum level the “lost of follow-up” patients participating at a therapeutic trial/registry since the reason of “lost of follow-up” might be related to a specific drug effect.

Disclosure of Interests: maxime dougados Grant/research support from: Eli Lilly and Company, Pfizer, AbbVie, and UCB Pharma, Consultant for: Eli Lilly and Company, Pfizer, AbbVie, and UCB Pharma

DOI: 10.1136/annrheumdis-2019-eular.8463

FRIDAY, 14 JUNE 2019

15:30:00 – 17:00:00

Don't panic – round table discussion on risk perception

SP0159

WHAT DO WE KNOW – WHAT SHALL WE DO – WHAT DO WE TELL

Timothy R. Radstake. University Medical Center Utrecht, Netherlands

There is an increasing need for patients to be involved in research as well as matters that can influence their own clinical journey. One aspect of science that links optimal clinical care is the increased use of biomarkers for so-called personalized medicine approaches.