



# Loss to follow-up and missing data: important issues that can affect your study results

Juliana Carvalho Ferreira<sup>1,2,a</sup>, Cecília Maria Patino<sup>1,3,b</sup>

Investigators used data from the São Paulo State Tuberculosis Control Program database to evaluate the association between the place of diagnosis and treatment outcomes. They reported that 25% of new tuberculosis cases were diagnosed in emergency facilities and that, in comparison with patients diagnosed in outpatient settings, they were more likely to have unsuccessful treatment outcomes, including loss to follow-up (in 12%) and death (in 10%).<sup>(1)</sup>

## LOSS TO FOLLOW-UP AND TYPES OF MISSING DATA

To study associations of an exposure (in cohort studies) or of an intervention (in randomized controlled trials) with clinical outcomes prospectively, investigators collect exposure/intervention information at study entry and then follow study participants over time and collect outcome data. If follow-up is incomplete or interrupted, leading to missing data at the end of the study, this could impact the internal validity of the study. Participants with missing data, compared with those with complete data, may differ systematically, for example when loss to follow-up is related to the death of participants.

Missing data occur for multiple reasons: 1. the variable of interest is not measured by the research team (e.g., forgetting to measure weight at baseline); 2. the study participant misses a scheduled study visit or test; or 3. the variable is measured but the team fails to register the variable value on the data collection form. However, loss to follow-up, where patient data is not available until the end of the follow-up period, is the most critical mechanism of missing data, because it might include missing outcome data, which is crucial to answer the research question.

Missing data are classified as missing completely at random (MCAR), when missingness is not related to the exposure, covariates, or the outcome; missing at random (MAR), when missingness is related to the exposure or confounders, but not the outcome; and missing not at random (MNAR), when missingness might be related to the outcome (Chart 1).<sup>(2)</sup> Loss to follow-up and missing data can threaten the internal validity of a study even if the mechanism is MCAR, in which we consider that the remaining participants are a random sample of the initial study population, because the study power will be decreased. If the mechanism is MAR, adjustments and imputation methods can be used, but that might introduce biases to the study. If the mechanism is MNAR, there is a serious risk of biased results.<sup>(3)</sup>

## HOW TO DEAL WITH LOSS TO FOLLOW-UP AND MISSING DATA

The best strategy to avert missing data is to prevent loss to follow-up. Designing the study carefully, training staff, implementing data quality procedures, and developing mechanisms to retain and contact participants are key. Additionally, there are statistical methods available to deal with missing data, but these procedures should be planned a priori and with consultation of a biostatistician.<sup>(3)</sup> However, there are situations when investigators might not overcome problems related to missing data because the mechanism of missingness is MNAR. In that case, losses to follow-up of 20%, for example, can result in serious biases and, therefore, should not be considered "acceptable". Remember, missing data are common and best practices include thinking about it early on when defining the research question and writing the protocol.

**Chart 1.** Types of missing data and strategies to minimize them.

Type of missing data	Example	Strategies to minimize missing data
MCAR	Participant moves to another state and abandons the study; a test result is lost at the lab	Develop standardized collection forms; monitor data quality; keep participant contact information up to date
MAR	In a cohort of COPD patients, participants with mild disease are more likely to abandon the study because they are asymptomatic	Offer benefits and incentives to retain participants; regularly contact participants; conduct a pilot study to identify risk factors for loss to follow-up; and develop strategies to overcome them
MNAR	Loss to follow-up is higher among tuberculosis patients who have serious adverse events due to tuberculosis drugs than among patients who tolerate treatment, and treatment nonadherence is related to death	Offer adequate support for study participants; develop strategies to retain participants with a high risk of loss to follow-up; and develop alternative methods to measure the outcome even for participants lost to follow up

MCAR: missing completely at random; MAR: missing at random; and MNAR: missing not at random.

## REFERENCES

1. Ranzani OT, Rodrigues LC, Waldman EA, Prina E, Carvalho CRR. Who are the patients with tuberculosis who are diagnosed in emergency facilities? An analysis of treatment outcomes in the state of São Paulo, Brazil. *J Bras Pneumol.* 2018;44(2):125-133. <https://doi.org/10.1590/1806-37562017000000384>
2. Pedersen AB, Mikkelsen EM, Cronin-Fenton D, Kristensen NR, Pham TM, Pedersen L, et al. Missing data and multiple imputation in clinical epidemiological research. *Clin Epidemiol.* 2017;9:157-166. <https://doi.org/10.2147/CLEP.S129785>
3. Kristman V, Manno M, Côté P. Loss to follow-up in cohort studies: how much is too much? *Eur J Epidemiol.* 2004;19(8):751-60.

1. Methods in Epidemiologic, Clinical, and Operations Research—MECOR-program, American Thoracic Society/Asociación Latinoamericana del Tórax, Montevideo, Uruguay.  
2. Divisão de Pneumologia, Instituto do Coração, Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo, São Paulo (SP) Brasil.  
3. Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA.  
a. <http://orcid.org/0000-0001-5742-2157>; b. <http://orcid.org/0000-0001-6548-1384>