Accounting for Missing Data and Dropout

Course developed by
Deborah H. Glueck and Keith E. Muller

Slides developed by Jessica R. Shaw, Keith E. Muller, Albert D. Ritzhaupt and Deborah H. Glueck

 1

Learning Objectives

Use an example trial to illustrate data missing at random.

Demonstrate how to increase sample size to account for data missing at random.

2

An example will help illustrate how to deal with missing data

With missing data, investigators need to check why the data is missing.

If the data is missing at random, a simple calculation can help adjust sample size.

Power and sample size sections in grant proposals should account for missing data.

3

Recall the longitudinal study of pain perceived after a root canal

Vignette

Researchers conducted a study to determine if patients who are instructed to use a sensory focus have a different pattern of long-term memory of pain than patients who did not. Patients were selected and randomly assigned to either intervention or no intervention.

Logan, Baron, and Kohout, 1995

4

Recall the longitudinal study of pain perceived after a root canal

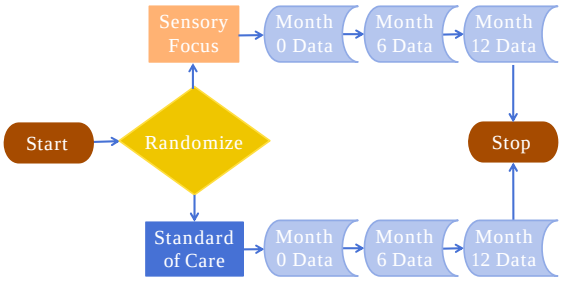
Vignette, continued

Patients in the intervention group listened to automated audio instructions to pay close attention only to the physical sensations in their mouth. Patients in the no intervention group listened to automated audio instruction on a neutral topic to control for media and attention effects.

Logan, Baron, and Kohout, 1995

5

Observed pain after root canal was measured at 0, 6, and 12 months



Logan, Baron, and Kohout, 1995

6

The scientists wished to compare the experience of the two groups

Measurements:

Patients were asked to rate their memories of pain at baseline, and at 6 and 12 months

Logan, Baron, and Kohout, 1995

7

The scientists wished to compare the experience of the two groups

Scientific goal:

The goal of the study was to compare the intervention group to the standard of care group in terms of their pattern of memory over time.

Logan, Baron, and Kohout, 1995

8

The scientists wished to compare the experience of the two groups

Null hypothesis:

There is no difference in the pattern of pain over time between groups. In other words, there exists no group by time interaction in patients' recall of pain.

Logan, Baron, and Kohout, 1995

9

The scientists planned a repeated measures analysis of variance

Planned analysis:

A repeated measures analysis of variance allows testing all trends across time and their interaction with treatment.

Logan, Baron, and Kohout, 1995

10

The scientists planned a repeated measures analysis of variance

Model:

The scientists planned to fit a general linear multivariate model, with the repeated measurements of memory of pain as the outcome, and indicator variables for group as the predictors.

Logan, Baron, and Kohout, 1995

11

The scientists planned a repeated measures analysis of variance

Statistical test:

The scientists planned to use a Hotelling-Lawley test to test interaction between group and time with an 0.01 Type I error rate.

12

The scientists understood the missing data pattern

Over 12 months, researchers expected a

25% loss to follow up.

Logan, Baron, and Kohout, 1995

13

The scientists understood the missing data pattern

They recognized that patients left the clinic because their insurance status changed or because their employers changed insurance.

Logan, Baron, and Kohout, 1995

14

The scientists understood the missing data pattern

Further, they discerned that leaving was unassociated with the memory of pain, the treatment, income, or any other factor they could measure.

Logan, Baron, and Kohout, 1995

15

FIRST DRAFT OF SAMPLE SIZE SUMMARY

16

Preliminary calculations found a required sample size of 44 participants

We plan a repeated measures ANOVA using the Hotelling-Lawley Trace to test for a time by treatment interaction. Based on previous studies, we predict that repeated measures of pain recall will have a standard deviation of 0.98.

Logan, Baron, and Kohout, 1995

17

Preliminary calculations found a required sample size of 44 participants

The correlation in pain recall between baseline and 6 months will be 0.5. Informed by clinical experience, we predict that the correlation will decrease slowly over time. Thus, we anticipate a correlation of 0.4 between pain recall measures at baseline and 12 months.

Logan, Baron, and Kohout, 1995

18

Preliminary calculations found a required sample size of 44 participants

For a desired power of 0.90 and a Type I error rate of 0.01, we need to enroll 44 participants to detect a mean difference of 1.2.

Logan, Baron, and Kohout, 1995

19

ACCOUNTING FOR MISSING DATA

20

The required sample size must be adjusted in anticipation of missing data

Inflate sample size by the 25% anticipated loss to follow up.

$$\frac{44}{1 - 0.25} = 58.6$$

21

Round up to 60 for division into two treatment groups of equal size

$$\frac{44}{1-0.25} = 58.6$$

$$\approx 60$$

22

Revise the sample size summary to include the adjusted sample size

Over 12 months, we expect 25% loss to follow up. We will inflate the sample size target of 44 by 25% to account for the attrition, for a total enrollment goal of 60 participants, or 30 participants enrolled per treatment arm.

23

Recently published results give better adjustments

Ringham et al. (2015, *Statistics in Medicine*) gave new and more accurate approximations to adjust for missing data.

Code available at www.SampleSizeShop.org

24

Review Summary

- We start with our sample size that is based on preliminary calculations, and understand the missing data pattern we can expect
- $\frac{\textit{Original sample size}}{1 - (\%\ \textit{loss to follow up})} = \textit{Adjusted sample size}$
- Report what you did that lead to your adjusted sample size in the sample size summary

25