

# Statistics Corner: Making Tables

Kamal Kishore<sup>1</sup>, Vidushi Jaswal<sup>2</sup>

## REALITY CHECK

The data presentation is an art based on science. In general, raw data are difficult to understand and present. Therefore, information from data is appropriately summarized and presented in texts and tables. The tables emphasize individual and precise values. Further, these are important and integral components of scientific research. A young trainee has been asked to prepare dummy tables by the mentor. The trainee has a general idea about the tables, but did not hear about the dummy tables. Therefore, the trainee wants to know the answers to a few questions before proceeding ahead with the assigned task.

- What are the components of the table?
- What are the characteristics of a good table?
- What are dummy tables?
- What are the advantages of dummy tables?

**Keywords:** Dummy tables, Good table, Tables, Tabulating data

*Journal of Postgraduate Medicine, Education and Research* (2020); 10.5005/jp-journals-10028-1393

## INTRODUCTION

The data collection is a crucial and integral component of academic writing. Researchers collect a large chunk of data to fulfill the objectives of the study. However, it is not feasible to present and make sense from large raw datasets. Therefore, data are summarized and presented in a lucid text, tables, and graphs for a better understanding. The text, tables, and charts primarily come under the results section, which perhaps can be called the soul of the study. The tables are visual elements that aid in speeding up the comprehension and interpretation for the readers. Therefore, the inclusion of specific text, tables, and figures need careful attention of the researchers.

The initial tables and graphs in the analysis are included to describe the characteristics of the study participants. Subsequently, the tables and figures with  $p$  values are used to support or refute the study hypothesis. Therefore, researchers need to carefully think and decide on the content and sequence of data to be included in tables. It is good to come up with dummy tables at the study conceptualization stage. The preparation of dummy tables will facilitate the systematic and structured evaluation of data in contrast to unscientific data torturing<sup>1</sup> approach.

The analysis of data before finalization of tables and figures often forces investigators for  $p$ -hacking.<sup>2</sup> In other words, it motivates the researchers to report statistically significant results ( $p < 0.05$ ) rather than the actual findings of interest at the study conceptualization stage. It further leads to loads of output which become challenging to compile. Data torturing is unstructured and unscientific; therefore, researchers should refrain from taking this approach. Data torturing is touted as one of the reasons for the replicability crisis in science.<sup>3</sup>

## TABLES

Tables are crucial for the presentation of data. Generally, academic institutions and publication houses have guidelines on including numbers and the formats of the tables. Still, when it comes to finalization, many researchers miss the specifics of preparing informative tabular display. Researchers may also consult reputed

<sup>1</sup>Department of Biostatistics, Postgraduate Institute of Medical Education and Research, Chandigarh, India

<sup>2</sup>Department of Psychology, Mehr Chand Mahajan DAV College for Women, Chandigarh, India

**Corresponding Author:** Kamal Kishore, Department of Biostatistics, Postgraduate Institute of Medical Education and Research, Chandigarh, India, Phone: +91 9591349768, e-mail: kkishore.pgi@gmail.com

**How to cite this article:** Kishore K, Jaswal V. Statistics Corner: Making Tables. *J Postgrad Med Edu Res* 2020;54(3):169–171.

**Source of support:** Nil

**Conflict of interest:** None

journals for refining and updating the tables. To a casual reader, it appears tables have only rows and columns as two components. However, a data table consists of five critical components named as title, row stubs, column headers and subheadings, footnotes, and body.<sup>4</sup> A complete data table is self-explanatory, where a reader does not require consulting the text for understanding it.

It is better to prepare a table keeping in mind “what” (outcomes of interest, such as, demographics or clinical characteristics), “where” (place), and “when” (time, such as, a year) with the help of “why” (is this table supplementing the study objective/s) in the study. The column headers (domains of importance, such as, area, time, or groups) come at the top of the row with the appropriate unit of measurements. The row stubs are usually the extreme left column with independent variables as its component. However, depending on the number of levels in both rows and columns, these are interchangeable. The footnotes are used to define the usage of unconventional symbols (such as, †, ‡, §, ¶, \*, S), signs (such as, □, ♦, ♪, ♫, ☼, ☺), and abbreviations (such as, LR—likelihood ratio), and the acknowledgement (for the replication of a partial or complete table) for making a self-explanatory table. The font size for footnotes is usually smaller, and they come at the bottom of the table. Finally, the body of the table reflects the results of the study. The adherence to a subset of requirements leads to an example of a poorly designed table. Table 1A demonstrates multiple issues in a single table for display purposes.

**Table 1A:** Demographics and clinical details

|              | Baseline    | First follow-up | Second follow-up | Third follow-up |
|--------------|-------------|-----------------|------------------|-----------------|
| Participants | 100         | na              | 88               | 77.5            |
| Male         | 60%         | —               | 71.59            | 70.3%           |
| SBP          | 130.5       | 141             | 142.53           | 137.333         |
| HR           | 63.73 (10)  | 65.11 (11)      | 70.22 (12)       | 74.44 (14)      |
| Weight       | 64.5 (15.2) | 65 (14)         | 63 (13.3)        | 59 (14.2)       |

However, a lousy table may have single or multiple problems as summarized below.

- The title is not appropriate
  - No information about the study place.
  - No information about the study group.
- The column headings and body (data) are not aligned properly
  - Either column headings or subheadings or both do not follow a systematic pattern.
  - Data are not aligned with headings, nor it is right-, left-, or center-justified.
- Lack of uniformity principle
  - No values in the brackets for SBP.
  - Poor and illogical selection of decimal places.
- Absence of footnote
  - The symbol (—) and abbreviation (na) are not explained.
  - The units of the variables are not included either with headings or footnotes.
- The shading and lines
  - Too many lines increase the cost (more ink), cognitive load, and time.
  - Difficult to read and interpret.
- Width and indicators
  - There are no indicators to segregate continuous or categorical data.
  - The table width is spread across the page for no reason.

**Good Table**

The general guidelines and important components of the table are discussed in the preceding section. However, there are subtle issues that demand the attention of the researchers to construct a good table. Many times, the body of the table is devoid of fundamental characteristics, such as, units, numerical precision, zeroes, and uniform justification. It needs appropriate and consistent usage of units to convey information in a row or a column. Similarly, it is good practice to arrange data uniformly on the right or left in each cell of the table. Many studies report numerical values with high precision with the help of decimals. A researcher needs to think carefully about retaining the number of decimals for presenting the data. For example, does it make a difference to report systolic blood pressure as 131.2 in comparison with 131 mm Hg? Clinical judgment rather than software-based reporting should take precedence for finalizing the table values. Further, the values with multiple zeros, such as, RBC count should be presented as  $5.2 \times 10^6 \mu\text{L}$  as compared to 5,200,000  $\mu\text{L}$ . Readers can consult both Tables 1A and 1B for making quick comparisons. Table 1A looks more concise, organized, and scientific as compared to Table 1A for

**Table 1B:** The baseline and follow-up clinical characteristics of the hospitalized patients recruited in the surgery unit of Post Graduate Institute of Medical Education & Research, Chandigarh, India, in 2017\*

| Characteristics      | Baseline    | First follow-up | Second follow-up | Third follow-up |
|----------------------|-------------|-----------------|------------------|-----------------|
| Participants—no. (%) | 200 (100.0) | na              | 176 (88.0)       | 155 (77.5)      |
| Male—no. (%)         | 120 (60.0)  | —               | 100 (51.8)       | 115 (74.1)      |
| SBP (mm Hg)          | 130 ± 10.6  | 141 ± 11.6      | 142 ± 12.0       | 137 ± 12.5      |
| HR (beats/minute)    | 63 ± 10.0   | 65 ± 11.1       | 70 ± 12.2        | 74 ± 14.3       |
| Weight (kg)          | 64.5 ± 15.4 | 65 ± 14         | 63 ± 13.3        | 59 ± 14.2       |

\*The values with a plus-minus sign are mean ± SD. na, the value was not available

—no. (%) cannot be calculated as the gender status for few patients was missing

Note: A hypothetical dataset is used for the demonstration purpose

**Table 1C:** A heat map of the baseline and follow-up clinical characteristics of the hospitalized patients recruited in the surgery unit of Post Graduate Institute of Medical Education & Research, Chandigarh, India, in 2017\*

| Characteristics      | Baseline    | First follow-up | Second follow-up | Third follow-up |
|----------------------|-------------|-----------------|------------------|-----------------|
| Participants—no. (%) | 200 (100.0) | na              | 176 (88.0)       | 155 (77.5)      |
| Male—no. (%)         | 120 (60.0)  | —               | 100 (51.8)       | 115 (74.1)      |
| SBP (mm Hg)          | 130 ± 10.6  | 141 ± 11.6      | 142 ± 12.0       | 137 ± 12.5      |
| HR (beats/minute)    | 63 ± 10.0   | 65 ± 11.1       | 70 ± 12.2        | 74 ± 14.3       |
| Weight (kg)          | 64.5 ± 15.4 | 65 ± 14         | 63 ± 13.3        | 59 ± 14.2       |

\*The values with a plus-minus sign are mean ± SD. na, the value was not available

—no. (%) cannot be calculated as the gender status for few patients was missing

Note: A hypothetical dataset is used for the demonstration purpose

the same data. A scientifically correct table follows the principle of parsimony and gives complete information without going into the details of the study text.

**Heat Maps**

The hallmark of the table is the presentation of variables in different units with accuracy. However, it becomes challenging to interpret and process the tables with an increase in the number of variables and comparison groups. Further, it is hard to visualize trends and associations with the help of tables. Readers will take different messages from the same table, which leads to confusion and shifts the flow of the discussion from the intended objectives. The heat maps can be used to update regular tables for a quick visualization of the information of importance. The numerical data with varying hue of colors produce heat maps. It is easy to generate heat maps in Microsoft Excel® (Microsoft, WA, USA) under conditional formatting option available in the “Home” tab. The default green, yellow, and red colors can be changed with other color combinations. The highest and lowest values in the table represent the dark shade of red and green, respectively (Table 1C).

**Dummy Tables**

Dummy tables are the planned empty tables with rows and columns titles. The dummy tables drive the sequential flow of data analysis



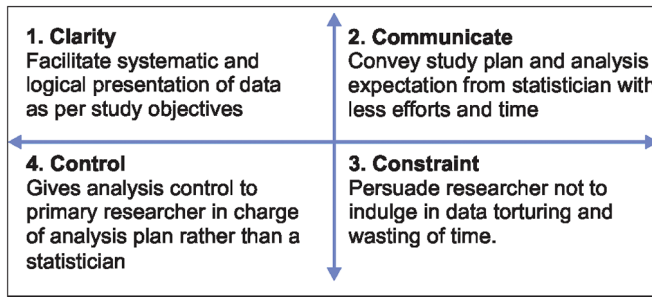


Fig. 1: Four Cs summarizing the importance of dummy tables

and discussion with various stakeholders, including statisticians. The intersection of rows and columns is known as cells. The values in the cells (body) of the dummy tables are filled after data collection and analysis. Dummy tables serve as the link between research questions, hypotheses, and data analysis plan. Preparing dummy tables at the protocol stage saves time and facilitates reporting of results as per objectives and guards against data torturing. The four Cs in Figure 1 summarize the importance of dummy tables.

**CONCLUSION**

The tables are an incredible tool to display a complex aggregate of numbers and figures in a clear, concise, and comprehensible

manner. The researchers need to carefully think about the essential components of the table before finalization. The decision should be derived from the study objectives and the research hypotheses. It is recommended to prepare dummy tables before initiating or meeting statisticians for data analysis. The preparation of dummy tables will improve clarity in communication and reduce clutter while self-analyzing or discussing an analysis plan with a statistician.

**ACKNOWLEDGMENTS**

The authors acknowledge Dr Nitasha and Dr Minakshi for their valuable time and inputs to improve the quality of the article.

**REFERENCES**

1. Mills JL. Data torturing. *N Engl J Med* 1993;329(16):1196–1199. DOI: 10.1056/NEJM199310143291613.
2. Wicherts JM, Veldkamp CLS, Augusteijn HEM, et al. Degrees of freedom in planning, running, analysing, and reporting psychological studies: a checklist to avoid p-hacking. *Front Psychol* 2016;7:1832.
3. Peng R. The reproducibility crisis in science: a statistical counterattack. *Significance* 2015;12:30–32. DOI: 10.1111/j.1740-9713.2015.00827.x.
4. United Nations Economic Commission for Europe. *Making Data Meaningful Part 2: A guide to presenting statistics*. Geneva 2009.