Chapter 7
Recording Scientific Data

7.1 Using Statistics in Place of the Study Question to Frame Results

We become so enamored with the output of our statistical programs and our statistical understanding that sometimes our narrative reads like the output of our statistical analysis program. You know you are making this mistake when words like "association," "analysis," or "relationship" are the subject of a sentence or when the name of variables used to code the data appears in the manuscript.

The point of analysis of health data is not mathematical output but what these results mean in terms of the lives and health of people. The results should be expressed and communicated with other health professionals in terms of the research question.

Examples of the error		Alternative, better options	
✗	Father's literacy was associated with immunization status (p=.007).	✓	Children whose fathers were educated were more likely to be completely immunized than children of uneducated fathers (84% versus 44%, p =.007).
✗	In simple regression analysis, education and pregnancy status give highly significant relationship, while language and counseled by give significant relationship on screening.	✓	Women who were educated, who spoke Hindi, and who benefited from counseling from a physician were more likely to consent to the screening test.
✗	The analysis of association among the independent variables showed that there is an association between the main exposure variable (Distgrp2) and the costgrp and between costgrp and the duration of disease (Durdgrp2).	✓	People who lived farther from health facilities spent more money per visit to the health-care facility.

S. Luby, D. L. Southern, *The Pathway to Publishing: A Guide to Quantitative Writing in the Health Sciences*, https://doi.org/10.1007/978-3-030-98175-4_7

7.1.1 Framing Narrative Results Around P-Values

A p-value assesses the probability that results as extreme as observed in the analyzed groups could have arisen by chance enrollment of a nonrepresentative study population. Scientific authors should assess how likely chance is a credible explanation for observed differences, but a p-value < 0.05 does not prove an association is causal. It does not provide insight on whether the association is due to bias. It does not assess whether the association is due to confounding.

A low p-value conflates whether an association between exposure and outcome have a large effect (which may have quite important impacts on the scientific or public health implications of the results) or whether there is a small or even trivial effect in a large number of observations. (If the sample size is one million, all of the p-values will be <0.001.)

As the authors of a formal assessment of the use of p-values in biomedical literature noted, "p-values do not provide a direct estimate of how likely a result is true or of how likely the null hypothesis is ('there is no effect') true. Moreover, they do not convey whether a result is clinically or biologically significant. P-values depend not only on the data but also on the statistical method used, the assumptions made, and the appropriateness of these assumptions" [15].

In short, p-values are silent on most important dimensions of assessing valid scientific inference. Presentations of results should not be framed around p-values. Indeed, framing results around p-values communicates to the reader that the author has a naïve approach to data interpretation. Instead, frame results around effect sizes and presenting work in an order so that readers can consider issues of confounding, bias, and dose effect and present p-values like a footnote, not as a central finding. Think of "statistical significance" as only an issue of second-order concern, that is, if there is a difference that is potentially meaningful and interesting, it provides a test of whether this difference is likely due to chance selection of a nonrepresentative study population.

Framing a scientific narrative around p-values also encourages a naïve dichotomous conceptuality, that is, that a factor is either present or absent. Science is characterized less by this sort of absolute binary frames and more about measuring degrees of difference.

The editors of the *International Journal of Epidemiology* explain their perspective on this issue. "We actively discourage use of the term 'statistically significant' or just 'significant' and statements in method sections such as 'findings at p<0.05 were considered significant.' Where used, we ask authors to provide effect estimates with confidence intervals and exact P values, and to refrain from the use of the term 'significant' in either the results or discussion section of their papers. Our justification of this position is given in Sterne J, Davey-Smith G. 'Sifting the evidence – What's wrong with significance tests?' *BMJ* 2001: 322:226-231."

Examples of the error		Alternative, better options	
✗	When we looked at the contamination of each toy ball separately, two toys did not reach statistical significance for fecal coliform contamination.	✓	When we compared fecal coliform contamination between groups for each toy ball separately, toys were consistently less contaminated in the cleaner households compared to the less clean households. However, the comparison between groups of fecal coliform contamination of toys 2 and 4 did not reach statistical significance (Table 2).
✗	Compared with persons who contracted Nipah infection from another person, Nipah cases who drank raw date palm sap were more likely to develop convulsion (log rank p-value <0.001), altered mental status (log rank p-value <0.001), and die (log rank p-value <0.001).	✓	Compared with persons who contracted Nipah infection from another person, Nipah cases who drank raw date palm sap were three times more likely to develop convulsions, 50% more likely to develop altered mental status, and 58% more likely to die (log rank p-values all <0.001).

7.2 Not Presenting the Core Data

Your most engaged readers are not only interested in your conclusions. They want to look at the data and draw their own conclusions. This is the essence of science-reflective consideration of empiric observations. Your manuscript should present the data in a way that allows the reader to form an independent opinion as to whether the data were analyzed properly and interpreted prudently. As a matter of transparency, the reader should be able to redo the key calculations. Thus, basic frequencies, rates, or means comparing groups on your central findings are crucial.

A common variant of this error occurs when comparison between groups is limited to measures of association or percentages without the underlying numbers. In its most extreme form, the measure of association is omitted entirely. Only a p-value is presented (see Error 7.1.1).

Examples of the error		Alternative, better options	
✗	Most subjects (62%) were not aware of ….	✓	Of 113 subjects, 70 (62%) were not aware of…. *[always show numerators and denominators in the calculation of proportions.]*
✗	There was a significant difference in the proportion of case-patients and control-subjects who reported eating the potato salad (p=0.0001).	✓	Of the 42 case-patients, 30 (71%) reported eating the potato salad compared with 19 of the 120 control-subjects (16%, odds ratio=13.3 p<0.01).
✗	Proportions only in the tables	✓	Always provide numerators and denominators.

7.3 Using Too Many Decimal Places

When the results of a study are presented with an excessive number of decimals, communication between the writer and the reader is impaired. The extra digits distract the reader from the message. Presenting too many decimal places also implies a precision that the data generally lack.

This error is most commonly seen with percentages. Data are presented as percentages, for example, 39%, rather than as frequencies, for example, 321/815, so that it is easier to remember and compare one group or scenario to another. Although 10,000 decimal places are a more precise report of the percentage, it is burdensome for the reader. For example, if 13 of 17 enrolled study subjects have a particular characteristic, this can be reported as 76%, 76.5%, 76.47%, 76.461%, 76.46706.... With a powerful enough calculating program, you could report thousands or millions of decimal places.

However, reported percentages with multiple decimal places are no longer easy to remember and compare. Active readers who want to understand the meaning of your scientific writing will often compare reported numbers to each other. It is much easier for readers to compare numbers and to perform mental arithmetic on rounded numbers. Thus, wherever possible, note percentages without decimal places. Only include decimals if the percentage is less than 10, and the figures beyond the decimal point have public health significance.

Similarly, when people report relative risk or confidence intervals, they often report it to two decimal places, for example, the statement that people who ate goat curry were three times more likely to become ill than persons who did not (relative risk of 3.24, 95% confidence interval CI=0.74–12.99, p value=.143). Can your investigation reliably estimate the relative risk and the confidence interval to two decimal places? If the study cannot support such precision, then you should not imply that level of precision by reporting the extra decimal places.

One rule of thumb for confidence intervals for odds ratio is that they should not have more than two meaningful figures. Whether or not these figures are decimals or not depends upon where the odds ratio fit on a log scale. Remember that the odds ratios for "protective exposures" and "risk factors" are symmetrical around the number one on a log scale. Thus, reporting an odds ratio of 243 represents the same amount of precision as an odds ratio of 24.3, an odds ratio of 2.43, and an odds ratio of 0.243. Thus, try to round up (add or subtract digits) so that you always display two meaningful figures, for example, 24, 2.4, or .24.

Examples of the error		Alternative, better options	
✗	The prevalence of active trachoma was 21.01% (95% confidence interval: 6.23–36.77%).	✓	The prevalence of active trachoma was 21% (95% confidence interval: 6.2–37%).
✗	People who ate goat curry were three times more likely to become ill than persons who did not (relative risk of 3.24, 95% confidence interval CI=0.74–12.99, p-value=0.143).	✓	People who ate goat curry were three times more likely to become ill than persons who did not (relative risk of 3.2, 95% confidence interval CI=0.74–13, p-value=0.15).

7.4 Using Too Few Decimal Places

In the enthusiasm to avoid using too many decimal places, occasionally, authors present too few. In most contexts, you want to communicate two digits of numerical information (25% is two digits. $1.2 million is two digits). As noted in 7.3 reporting, a percentage greater than 10, adding a third digit, a decimal place, is generally distracting and uninformative. However, if you are reporting an odds ratio or other relevant small number, then it is important to communicate two digits of information (2.1 or 0.63) even if one or more of these digits are decimal places. Count digits, not decimal places!

Examples of the error		Alternative, better options	
✗	Children whose mother completed primary education were less likely to be hospitalized for diarrhea (odds ratio 0.6, 95% confidence interval 0.4, 0.8).	✓	Children whose mother completed primary education were less likely to be hospitalized for diarrhea (odds ratio 0.57, 95% confidence interval 0.42, 0.77).
✗	Ambulatory case-patients spent a median of US$2 (IQR=$1–4) in the public hospitals.	✓	Ambulatory case-patients spent a median of US$1.8 (IQR=$1.1–3.6) in the public hospitals.

7.5 Using Incomplete Headings for Tables and Figures

In a biomedical manuscript, the figures and tables should stand alone. A reader should be able to look at the table or figure, read the title, and understand it. Readers should not have to refer to the narrative methods or results to understand the table or the figures. Thus, a typical heading reporting on a study population should include person, place, and time. The number of study subjects and statistical methodology should be communicated. Use footnotes to explain apparent discrepancies or other issues in the table/figure that benefit from further clarification.

Tables and figures that are developed for slides to accompany verbal presentations are different than tables developed for manuscripts. Slide visuals are designed to be understood quickly. Brief titles for tables and figures in these slides are preferred.

Examples of the error		Alternative, better options:	
✗	Figure 1: Epicurve of the measles outbreak	✓	Figure 1: Cases of measles by date of onset, Chennai City, Tamil Nadu, November 2004
✗	Table 2: Risk factors associated with illness, univariate analysis	✓	Table 2: Characteristics of meningitis case-patients and control subjects, Kano City, Nigeria, March 1996

7.6 Imbalance Between Table and Narrative Presentation of Results

7.6.1 Too Little Narrative Explaining the Tables

Just as tables, figures, and graphs should stand on their own and not require accompanying text, the narrative section of the results should stand alone. A reader should be able to read only the narrative text, not look at any of the figures or tables, and come away with a clear understanding of the important findings from the analysis. This error most commonly takes the form of several well-constructed tables being presented in the results section with only a sentence or two in the narrative results section pointing to each table. The results section should not repeat all the data that is in a table but rather should focus the reader on the highlights. Look at several quality journal articles related to your research question, and note the balance between what is presented in the narrative text and what is presented in the tables. Strive for a similar balance.

Example of the error		Alternative, better option	
✗	Of all the food items, only the vanilla ice cream was associated with illness (Table X).	✓	The risk of illness was estimated according to consumption of each of the eight menu items that were served at the lunch (Table X). Eating vanilla ice cream was the only exposure that was significantly associated with illness (relative risk: 8.6, p=0.001) and that accounted for the majority of cases (population attributable fraction: 86%).

7.6.2 Too Much Narrative Explaining the Tables

Some manuscripts deploy excessive narrative to comment on nearly every number presented in a table. This includes reiterating minor findings that are not relevant to the core issues engaged by the manuscript. A key responsibility of the analyst is to reduce data so it is more easily understandable to the reader. The narrative results section of a manuscript should summarize the primary findings and highlight findings that contribute importantly to the interpretation of the results. Avoid overrepetition of data that is more easily seen and compared in a well-constructed table.

7.6.3 Presenting Results in Narrative that Would Be Clearer in a Table

Comparison of a few numbers can be clearly understood when presented in a narrative paragraph, but when there are many numbers, subgroups, and comparisons, a table is a more efficient format for communication. Tables allow the readers to quickly compare columns and subgroups and so understand the relationships among all of the observations.

Examples of the error	
✗	A total of 4046 blood cultures were performed from study participants hospitalized in the inpatient department of Hospital A (n=2363) and Hospital B (n=1683). Of these, 694 (17%) were positive for *Salmonella* Typhi or Paratyphi. 421 (18%) of blood cultures from inpatients in Hospital A and 208 (12%) in Hospital B grew *S*. Typhi, while 39 (2%) of blood cultures from inpatients in Hospital A and 26 (2%) in Hospital B grew *S*. paratyphi. A total of 4046 blood cultures were performed from study participants enrolled as outpatients department of Hospital A (n=6225) and Hospital B (n=5094). Of these, 694 (6%) were positive for *Salmonella* Typhi or Paratyphi. 435 (7%) of blood cultures from outpatients in Hospital A and 208 (12%) in Hospital B grew *S*. Typhi, while 10 (0.2%) of blood cultures from outpatients in Hospital A and 14 (0.2%) in Hospital B grew *S*. paratyphi.
Alternative, better option	
✓	

		Tested	***Salmonella* typhi**		***Salmonella* paratyphi**	
			No.	**%**	**No.**	**%**
	Hospital A					
	Inpatient	2363	421	18%	39	2%
	Outpatient	6225	435	7%	10	0.2%
	Hospital B					
	Inpatient	1683	208	12%	26	2%
	Outpatient	5094	289	6%	14	0.3%
	Total	**15365**	**1353**		**89**	

7.7 Pointing Too Explicitly to Tables and Figures

In your results section, if the words “Table 1” or “‘Figure 2” are the subject of a sentence, you have likely committed this error. The paper should be organized around the central ideas you want to communicate and that you want the reader to focus on. Thus, lead with your findings, and compose your language around those findings and related ideas rather than around structures, that is, pages, tables, or figures.

Examples of the error		Alternative, better options	
✗	Table 1 describes the forms in which areca nut was used.	✓	Sweetened varieties of areca nut were the most popular (Table 1).
✗	Figure 2 presents the age, sex, and geographic distribution of our sample across the four study districts.	✓	The age, sex, and geographical distribution of the samples was similar across the four study districts (Figure 2).

7.8 Using Inappropriate Figures

Edward Tufte, in his excellent book, "*The Visual Display of Quantitative Information,*" argues that figures for scientific manuscripts should be evaluated using a data to ink ratio. He urges communicating the most data with the least ink. Excessive ink in figures mean they include unnecessary axes, grid lines, borders, 3-D effects, and other elements that do not add substance and make the figures less understandable.

Space is at a premium for print journal editors, who weigh this issue more from the perspective of data to space ratio. Both pie charts and simple frequencies presented as bar charts are inefficient. It is reasonable to assume that the reader of a scientific manuscript understands the difference between 20% and 40% and so does not need it illustrated by comparing relative widths of a pie or relative heights of a bar. A simple table can efficiently present proportions.

Thus, use figures to achieve key communication objectives. Figures are best used in two situations:

1. When they permit presenting a large amount of data in a format that reveals the underlying characteristics of the distribution, for example, scatter plots that show trends
2. When they communicate in a more effective and efficient visual format than can be achieved with a narrative description or a table, for example, a figure that presents multiple components of a phenomenon, such as different age trends by sex, or presents the data in a way that reveals an important relationship

7.9 Generic Data Tables That Lack a Clear Message

There is no single standard format to present data in tables. Tables are an integral element of the broad scientific argument that you compose through your manuscript. Tables should be organized based on the communication objective of the article. Thus, the first step in drafting a table is to identify the communication objective for the table. Examples might be to describe the baseline characteristics of the population, to compare the outcome of a group who received an intervention with the outcome in a nonintervention group, or to compare the characteristics and exposures of persons who became ill with persons who remained well.

First, specify the communication objective of the table. Then construct the table so that the message comes through clearly. The patterns in the data that you are striving to illustrate should be obvious at a glance or at least should be obvious once they have been pointed out by the narrative description in the results section of the manuscript [16]. Just like narrative scientific writing, expect to develop and revise tables through several drafts.

7.10 Table Layout That Impairs Comparisons

An advantage of presenting data in tables, rather than in a narrative paragraph, is that by clearly aligning numbers, different groups and different characteristics can be readily compared. Numbers are easier to compare reading down columns than across rows especially for larger numbers of items. Such comparisons are often the central communication objective of a table. To facilitate comparison, avoid:

- Columns that are too wide. This makes it difficult to compare data between columns. One common form of this error is to set the width of the table column based on the length of the column heading rather than on optimizing column width to permit comparison of data.
- Ordering data haphazardly. Rather than presenting characteristics in the table in alphabetical order, or in the order they were asked in the questionnaire, consider the easiest way for the reader to understand the information. Ordering characteristics from smallest to largest or largest to smallest is an intuitive approach that helps the reader to quickly and easily understand.
- Poorly aligned data that impedes comparison. Align data and decimals so that a vertical list is readily comparable.

Hard to compare	Easier	Still Easier
23 42 34 109 87 42 27 98 114 75	23	23
	42	27
	34	34
	109	42
	87	42
	42	75
	27	87
	98	98
	114	109
	75	114

(These examples and much of the text was contributed by Robert Fontaine with help from ASC Ehrenberg [16].)

Use the table layout effectively to help the viewer -- place numbers for comparison close together

Year	Both Sexes	Male	Female
1973	600	500	99
1970	670	580	87
1968	550	460	89
1966	330	260	71

Draw columns and rows close together

Year	Both Sexes	Male	Female
1973	600	500	99
1970	670	580	87
1968	550	460	89
1966	330	260	71

Move and minimize intervening numbers

	Rate per 1000 (SE)		
Year	Male	Female	All
1993	83 (2.3)	78 (2.2)	80 (1.9)
1994	62 (2.5)	66 (2.7)	63 (1.8)
1995	58 (2.1)	54 (2.0)	56 (1.7)
1996	55 (2.0)	45 (2.0)	51 (1.7)

Remove intervening numbers entirely if consequence minimal

	Rate per 1000[a]		
Year	M	F	All
1993	83	78	80
1994	62	66	63
1995	58	54	56
1996	55	45	51

a. Standard errors for all rates less than 5% of rate.

Organize data by magnitude

Exposure	1000 Cases	Rate	Rate Ratio	p
A	11	2.9	1.3	> 0.10
B	6	9.9	4.3	< 0.001
C	34	5.4	2.3	> 0.1
None	27	2.3	1.0	Ref*

a = p-value
b = reference exposure category

Organize data by magnitude

Exposure	1000 Cases	Rate	Rate Ratio	p[a]
B	6	9.9	4.3	< 0.01
C	34	5.4	2.3	< 0.05
A	11	2.9	1.3	> 0.001
None	27	2.3	1.0	Ref[b]

a. = p-value
b. = reference exposure category

7.11 Using Less Informative Denominators in a Table

Multicolumn tables allow readers to compare characteristics among different subgroups. Authors commonly include percentages to assist data interpretation. Sometimes, authors erroneously report percentages using less informative row totals as the denominator rather than the more informative column denominator.

Using a row total denominator prevents an intuitive comparison of the columns, thereby undermining a primary advantage of presenting data in tables.

Consider the two tables below that describe study subjects some of whom were enrolled near battery recycling sites and others who were enrolled near turmeric processing sites. Some of the study subjects were caregivers, and some were workers.

In the erroneous table at the top, row totals are used to calculate proportions. Thus, among the 188 illiterate study subjects, 74% of them were caregivers who were enrolled at the battery recycling site. By comparison, only 5% were caregivers who were enrolled at the turmeric processing site. This comparison is not particularly informative. It reflects the peculiarity of study subject enrollment. Specifically, that more study subjects were enrolled from battery recycling rather than from turmeric processing. If the study team had spent a few more days at turmeric processing sites, these numbers would be quite different. The proportions do not reflect underlying characteristics of the different groups. The proportion in the total column is particularly uninformative.

The proportions in the alternative table are more informative. They illustrate, for example, that 31% of caregivers at the battery recycling site were illiterate in contrast to 16% in the turmeric processing site and 100% of workers. It suggests that these groups are different.

Examples of the error:

× **Table.** Demographic and socio-economic information of the respondents.

Characteristics	Caregivers			
	Battery recycling site (n=444)	Turmeric processing site (n=57)	Workers (n=40)	Total (N=541)
Age				
Mean	31.4	32.1	37.6	31.9
Educational status				
Illiterate	139 (74%)	9 (5%)	40 (21%)	188 (100%)
Monthly household income				
Mean ± SD	16739	18139	15756	16808
Occupation				
Dependent	422 (89%)	55 (11%)	0 (0%)	477 (100%)

Alternative, better option:

✓ **Table.** Demographic and socio-economic information of the respondents.

Characteristics	Caregivers			
	Battery recycling site (n=444)	Turmeric processing site (n=57)	Workers (n=40)	Total (N=541)
Age				
Mean	31.4	32.1	37.6	31.9
Educational status				
Illiterate	139 (31%)	9 (16%)	40 (100%)	188 (35%)
Monthly household income				
Mean ± SD	16739	18139	15756	16808
Occupation				
Dependent	422 (95%)	55 (97%)	0 (0%)	477 (88%)

7.12 Comparing to a Varying Baseline

We often analyze data where observations are grouped into multiple levels of exposure. In the example below, we have categorized observed handwashing behavior into mutually exclusive categories:

Handwashing after defecation	Group A		Group B		Odds ratio	
	Number	%	Number	%	varying baseline	reference group
No handwashing	75	12%	150	19%	0.6	--
Washed one hand with water alone	150	23%	150	19%	1.3	2.0
Washed both hands with water alone	125	19%	150	19%	1.0	1.7
Washed one hand with soap	150	23%	100	15%	2.1	3.0
Washed both hands with soap	150	23%	200	25%	0.9	1.5
Total	650		750			

The common error is to compare the prevalence of each level of the variable in group A to the prevalence at the same level of the variable in group B. Thus, if we compare the prevalence of washing both hands with water alone, the prevalence is the same (19%) in group A and group B, so we could say that people in group A and B are equally likely to wash both hands with water alone, which is equivalent to an odds ratio of 1.0. The problem with this comparison is that the people who are not washing both hands with water alone are quite a heterogeneous group. Some of them are practicing less intense handwashing (not washing their hands at all or only washing one hand), and others are practicing more intense handwashing. Indeed, even if we have an elevated odds ratio with such a comparison, it is difficult to interpret because we don't know if this elevation results from a difference in more intense or less intense handwashing.

The standard approach to resolve this dilemma is to arrange the exposure level into a mutually exclusive hierarchy. Set the lowest level of exposure as the reference group, and then consider the 2 × 2 table comparing each level of exposure to this reference group. Using this approach illustrated in the final column, we can conclude that compared with group B, group A is more likely to wash either one or both hands with water rather than not washing at all.

7.13 P-Value in a Baseline Table of a Randomized Controlled Trial

In a randomized controlled trial, the intervention is assigned randomly. Therefore, any difference between groups is due to random assignment. A p-value tests whether or not an observed difference is larger than would be expected by chance. It is an irrelevant test in a randomized controlled trial.

If reviewers asked that such a comparison be added to your baseline table, cite the classic article: Altman D. Comparability of randomised groups. *The Statistician* (1985) 34, pp. 125–136. If you believe your reviewer has a sense of humor, you may want to directly quote from Altman, "... Performing a significance test to compare baseline variables is to assess the probability of something having occurred by chance when we know that it did occur by chance. Such a procedure is clearly absurd."

It remains important to assess whether there are meaningful differences between the intervention and control group. If there is imbalance, this suggests that randomization failed to create balanced group. If the baseline characteristics that differ between groups are also associated with the study outcome, then an adjusted analysis will need to be included. The assessment of baseline differences, however, is not based on evaluating a p-value but rather on a judgment of whether the differences in characteristics between groups is large enough that they could plausibly affect the outcome.

Examples of the error:

× **Table 1:** Student household characteristics at baseline (N=1,560)

Description	Intervention (N=688) n (%)	Control (N=872) n (%)	*p* value
Source of household income			
Formal employment/service	405 (59)	479 (55)	0.12
Self-employment	27 (4)	43 (5)	0.34
Casual/contract job	57 (8)	95 (11)	0.08
Day labor	102 (15)	93 (11)	0.01
Rikshaw/van puller	46 (7)	99 (11)	0.002
Driving motor vehicles	51 (7)	63 (7)	0.89
Household size (member): Mean (Std. Dev.)	5.0 (1.6)	4.9 (1.5)	0.32

Alternative, better option:

✓ **Table 1:** Student household characteristics at baseline (N=1,560)

Description	Intervention (N=688) n (%)	Control (N=872) n (%)
Source of household income		
Formal employment/service	405 (59)	479 (55)
Self-employment	27 (4)	43 (5)
Casual/contract job	57 (8)	95 (11)
Day labor	102 (15)	93 (11)
Rikshaw/van puller	46 (7)	99 (11)
Driving motor vehicles	51 (7)	63 (7)
Household size (member): Mean (Std. Dev.)	5.0 (1.6)	4.9 (1.5)

7.14 Using Nonstandard Footnote Symbols in Tables

Footnotes contribute important explanations to data presented in tables. They are useful to clarify an analytic approach, groups being compared, statistical significance, and other explanatory information. Historically, the sequence of footnotes was:

*, †, ‡, §, ||, ¶, **, ††, ‡‡, §§, ||||, ¶¶, etc.

You can find these symbols using the insert symbol feature of Microsoft Word. Note that these symbols should be in superscript.

More recently, some journals have suggested different symbols (most commonly a, b, c, d...). Check with your target journal's instructions to authors to ensure that your notation is consistent with their preference.

7.15 Using the Wrong Symbol to Designate Degree

Wrong example: 4 0C or 4 oC

To use the degree symbol, select the insert symbol feature of Word, select a circle (i.e., not the letter "o" or the number zero), and then make the circle superscript.

Correct example: 4 °C

Recent versions of MS Word include a degree symbol. Go to Insert and then Symbol to find the figure.

7.16 Numbering Figures or Tables out of Sequence

Readers expect and journals require tables and figures to be numbered in the order that they are referred to in the narrative text of the paper (i.e., Table 1, Table 2, Table 3, Figure 1, Figure 2, Figure 3). In addition, each table and figure should be cited in the narrative text (otherwise, readers and editors will assume it is not important and can be dropped).

The most common form of this error is when authors mention an element of complicated data analysis in the methods section and refer to a later table or figure in the manuscript. Usually, the best approach in this situation is to describe the statistical method without pointing to the results table or figure. The problem with citing the advanced table or figure as Table 1 or Figure 1 is that it will confuse readers to have this more complicated analysis presented before the more basic results that build toward the more complicated analysis.

The other common form of this error is renumbering the tables or figures but not updating these numbers in the narrative text.

7.17 Maps with Irrelevant Details

When a map is included in the manuscript, its role is to communicate specific geographical information, for example, the location of the study, spatial relationships among cases, or the spatial distribution of exposures. Inserting a map constructed by someone else that is filled with details that are irrelevant to the communication role for the map, for example, district divisions, rivers, or railroad lines, distracts readers from the message. Draw your own map or begin with a generic map and add the elements that are essential to the message.

Example of the error	Alternative, better option