

**THE AMERICAN PSYCHOLOGICAL ASSOCIATION *PUBLICATION MANUAL*
SIXTH EDITION: IMPLICATIONS FOR STATISTICS EDUCATION**

Fiona Fidler

School of Psychological Science, La Trobe University, Australia
f.fidler@latrobe.edu.au

The American Psychological Association (APA) Publication Manual sets the editorial standards for over 1000 journals in the behavioural, life and social sciences. Well known for its referencing standards, the Manual is also an authoritative source of statistical advice for many journals. It is therefore crucial that statistics education in these disciplines address its requirements and recommendations. The sixth edition of the Manual (published 2010) includes new guidelines about reporting effect sizes, confidence intervals and meta-analysis. These changes are a response to calls for the statistical reform of psychology—in particular, calls to decrease reliance on Null Hypothesis Significance Testing—which have been made with increasing vigor over the last 60 years. This paper critically reviews the new guidelines and discusses the implications for statistics teaching within psychology and other disciplines.

In addition to the 59 journals published by the American Psychological Association (APA), there are “at least a thousand other journals in psychology, the behavioural sciences, nursing and personnel administration [that] use the *Manual* as their style guide” (APA, 2001, p. xxi). It is “the single text which virtually every psychologist, of whatever sub-speciality, has contact with at some point in their career” (Budge & Katz, 1995, p. 218). Best known for its advice on formatting and referencing, the *Manual* also—increasingly—offers advice on statistical reporting. It is important for statistics educators who teach students in these disciplines to stay up to date with changes in the *Manual*’s statistical reporting advice and, in particular, to appreciate the shift in thinking that the sixth edition represents.

STATISTICAL REFORM AND THE APA *PUBLICATION MANUAL*

For over six decades psychology and many other life and social sciences have been dominated by Null Hypothesis Significance Testing (NHST). In response to mounting criticisms of the practice, the APA established a Task Force on Statistical Inference in 1999 to investigate a proposal to ban NHST from its journals. The Task Force did not enforce a ban but it did subsequently recommend de-emphasising NHST in favour of estimation. The sixth edition of the *Publication Manual* (released in 2009, copyright 2010) offers by far the strongest support for this recommendation.

Because it so widely known and influential, the *Manual* has often been identified as an important vehicle for statistical reform and re-education within psychology: “the *APA Publication Manual* and similar manuals are the ultimate change agents” (Kirk, 2001, p. 217). Whilst some statistical reform recommendations have had a mention in previous editions of the *Manual* they have been unaccompanied by practical advice and examples, and therefore have been relatively ineffective at motivating changes in practice or teaching (Fidler, 2002; Finch, Thomason & Cumming, 2002).

The 4th edition (1994) was the first to mention statistical power and effect sizes, but their introduction was brief and little practical advice was given: “take seriously the statistical power considerations associated with your tests of hypotheses” (p. 16) and “[you are] encouraged to provide effect-size information” (p. 18). The omission of prescriptive detail is unusual in a text like the *Manual*, which for other topics is remarkably specific. As Kirk (2001) put it:

If the 1994 edition of the APA manual can tell authors what to capitalize, how to reduce bias in language, when to use a semicolon, how to abbreviate states and territories, and principles for arranging entries in a reference list, surely the next edition can provide detailed guidance about good statistical practices (p. 217).

The 5th edition (2001) was the first to recommend CIs. Yet whilst the text claimed “they are, in general, the best reporting strategy” (p. 22), all examples of statistical reporting failed to include any CIs. For example, they were strikingly absent from tables and manuscript templates. Researchers were once again left without practical advice of the kind Kirk called for, and the uneven coverage for statistical reporting continued. Immense detail was provided about how to report p values for varying situations, but no advice was given on how to construct error bars for figures or add effect sizes and CIs to tables or text.

The sixth edition (2010) therefore represents an important step for statistical reform in psychology, and other disciplines bound to the *Manual*. Firstly, it specifies a format for reporting CIs and gives many examples. Secondly, it encourages meta-analysis in many places, and gives detailed standards for reporting meta-analyses. These developments are important because they encourage psychology not only to shift emphasis away from NHST but also, more fundamentally, to think quantitatively and cumulatively.

THE STATISTICS CLASSROOM AND THE SIXTH EDITION

As mentioned, the single most important change in the sixth edition is the move away from sole reliance on NHST to a more complete and quantitative reporting of results: “APA stresses that NHST is but a starting point and that additional reporting elements such as effect sizes, confidence intervals, and extensive description are needed” (p. 33). Coordinators of service courses in relevant disciplines should be aware of this shift. The remainder of this paper highlights key topics in a way that I hope is easily transported to the classroom.

NHST. When reporting NHST the *Manual*'s primary recommendation is to “report exact p values (e.g., $p = .031$)” (p. 114). Relative p values (e.g., $p < .05$) or asterisks to signal relative p values are only permitted if necessary to achieve clarity in tables or figures. The term ‘significant’ on its own can be easily mistaken to mean clinical or practical importance, rather than a ‘statistically significant’ difference or correlation. In most cases the *Manual* uses the expression “statistically significant” to avoid this ambiguity. There is even one example where the *Manual* follows Kline's (2004, pp. 86-88) advice to omit the word significant altogether: “...were statistically different, $F(4, 132) = 13.62, p < .001$ ” (APA, 2009, p. 94). As in the two previous editions, researchers are urged to take power seriously (p.30). For the first time, CIs are presented as an alternative to power in the planning stages of an experiment (p.31).

Unfortunately the *Manual* stops short of issuing warnings about common misinterpretations of NHST, e.g., the inverse probability fallacy and the replication fallacy. Warnings of this kind are especially important in student training, since misconceptions about NHST are widespread and robust (e.g., Haller & Krauss, 2002). Fortunately, Kline (2004, Chapter 3) provides excellent coverage on this topic, and is recommended reading on the APA style website.

Effect Sizes

There are three things to note about the *Manual*'s recommendation to report Effect Sizes. Firstly, effect sizes can be reported in original units, such as mean differences, and/or they can be reported as a standardised or units-free measure. Effect sizes “are often most easily understood when reported in original units”, such as “the mean number of questions answered correctly; kg/month for a regression slope”, but it can also be useful to report an effect size “in some standardized or units-free unit (e.g., as a Cohen's d value) or a standardized regression weight” (all from p.34). The latter helps with comparisons across studies and with meta-analyses, but may not be as intuitively meaningful as reports in original units for individual studies. One solution is to report both!

Secondly, the *Manual* stresses the importance of reporting effect sizes for statistically non-significant effects as well as statistically significant ones: “Mention all relevant results...; be sure to include small effect sizes (or statistically nonsignificant findings)...” (p. 32). This is important because of the common flawed practice of reporting statistically non-significant results as merely *ns*. Reporting only *ns* makes statistically non-significant results impossible to interpret and impedes future meta-analysis.

Finally, and perhaps most importantly, the *Manual* encourages not only the reporting of effect sizes, but also their interpretation: “Wherever possible, base discussion and interpretation of results on point and interval estimates” (p. 34). This means conclusions should be based on more than simple accept-reject statements based on $p < .05$; they should be meaningful quantitative statements about the size and importance of the effects.

Confidence Intervals

CIs are strongly recommended: “It is almost always necessary to include some measure of effect size.... Whenever possible, provide a confidence interval for each effect size” (p. 34). Whereas previous editions failed to provide details of how to report a CI, the sixth edition specifies a reporting format. The format for reporting a CI in text is: “ $d = 0.65$, 95% CI [0.35, 0.95]” (p. 117). If further intervals are reported in the same paragraph, the “95% CI” may be omitted if the meaning is clear. The units of measurement should not be repeated in the CI: “ $M = 30.5$ cm, 99% CI [18.0, 43.0]” (p. 117). So far as we know, there is no other discipline that has a single format for reporting CIs in text that is well-established and widely-used. The sixth edition also introduces additional statistical abbreviations, including CI (confidence interval), ES (effect size), and *LL* (lower limit of a CI) and *UL* (upper limit) (see Table 4.5, pp. 119-123). The abbreviation should be italicized if used as a symbol to refer to a numerical value, but not if it stands for words.

As mentioned above, the sixth edition highlights the use of CI in *a priori* sample size planning, as an alternative to power calculations, e.g., “calculations based on a chosen target precision (confidence interval width) to determine sample sizes” (p. 31). This is an important step. ‘Precision’, i.e., CI width, may be a more intuitive way for students and researchers to think about effect size and statistical power issues.

The sixth edition also demonstrates how to include range of common types of ESs and CIs in results tables. CIs should also be included in tables by using either the [..., ...] format, or separate columns for the lower (*LL*) and upper (*UL*) limits of the intervals. “When a table includes point estimates, for example, means, correlations, or regression slopes, it should also, where possible, include confidence intervals” (p. 138). There are four sample tables that include CIs, and the intervals are for odds ratios (p. 139), means (p. 140), linear contrasts (p. 143, but note the column of contrast values has been omitted), and raw regression coefficients (p. 144). Both styles (e.g., [..., ...] and the separate *LL* and *UL* columns are illustrated.

The sixth edition states that “Figures can be used to illustrate the results... with error bars representing precision of the... estimates” (p. 151). This follows the previous advice of the APA Task Force which recommended that “all figures, include graphical representations of interval estimates whenever possible” (Wilkinson et al., 1999, p. 601). The *Manual* is short on advice about constructing figures with appropriate error bars, or how to interpret figures with error bars. However, the APA style website refers researchers to the excellent paper by Cumming and Finch (2005).

This statement about figures is definitely worth pointing out to students: “If your graph includes error bars, explain whether they represent standard deviations, standard errors, confidence limits, or ranges” (p. 160). Every figure with error bars must state what the bars represent.

Meta-analysis

The sixth edition offers strong motivation for thinking meta-analytically: “Your work will more easily become a part of the cumulative knowledge of the field if you include enough statistical information to allow its inclusion in future meta-analyses” (p. 34). To this end, an entirely new section has been added, the Meta Analysis Reporting Standards (MARS). MARS is offered in an appendix to the sixth edition and provides a detailed checklist of what should be reported in a meta-analysis. In addition, three pages are devoted to a sample meta-analysis manuscript (pp. 57-59), and two sub-sections (pp. 36-37, 183) give meta-analysis guidelines.

CONCLUSION

The sixth edition of the *APA Publication Manual* represents an important step in the statistical sophistication of not only psychology, but the many other disciplines that rely on the *Manual* for their journal reporting standards—sociology, education, nursing and management

amongst them. The changes in the *Manual* that have been outlined in this paper are promising for statistical reform. If widely adopted they provide the opportunity for these disciplines to become cumulative sciences of estimation that deal with quantifying uncertainty, rather than simply making dichotomous accept-reject decisions on the basis of individual studies. Statistical educators have a duty to ensure that students are adequately trained in 'the new stats'.

REFERENCES

- American Psychological Association. (1994). *Publication manual of the American Psychological Association* (4th ed.). Washington, DC: Author.
- American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.
- American Psychological Association. (2010a). *Publication manual of the APA* (6th ed.). Washington, DC: Author.
- Budge, G., & Katz, B. (1995). Constructing psychological knowledge: Reflections on science, scientists and epistemology in the *APA Publication Manual*. *Theory & Psychology*, 5, 217-231.
- Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals, and how to read pictures of data. *American Psychologist*, 60, 170-180. Online: www.apastyle.org/manual/related/cumming-and-finch.pdf.
- Fidler, F. (2002). The fifth edition of the *APA Publication Manual*: Why its statistics recommendations are so controversial. *Educational and Psychological Measurement*, 62, 749-770.
- Finch, S., Thomason, N., & Cumming, G. (2002). Past and future American Psychological Association guidelines for statistical practice. *Theory & Psychology*, 12, 825-853.
- Haller, H., & Krauss, S. (2002). Misinterpretations of significance: A problem students share with their teachers? *Methods of Psychological Research*, 7, 1-20.
- Kirk, R.E. (2001). Promoting good statistical practices: Some suggestions. *Educational and Psychological Measurement*, 61, 213-218.
- Kline, R. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research*. Washington, DC: APA. Online: www.apastyle.org/manual/related/kline-2004.pdf.
- Wilkinson, L., & Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594-604. Online: www.apastyle.org/manual/related/wilkinson-1999.pdf.