ReBAC
Revista Brasileira de Análise do Comportamento
Brazilian Journal of Behavior Analysis

ISSN 1807-8338 - Versão Impressa | ISSN 2526-6551 - Versão Eletrônica
periodicos.ufpa.br/index.php/rebac

Sample Size in Behavioral Research: A Systematic Review of JEAB and JABA from 2009 to 2018

Tamanho de amostra em pesquisa do comportamento: uma revisão sistemática de JEAB e JABA de 2009 a 2018

Albert J. Schrimp [1]
James D. Griffith [1]
Kathryn Potoczak [1]
Thomas C. Hatvany [1]
Amber E.Q. Norwood [1]
Ashley A. Conley [1]

[1]Shippensburg University

Abstract

The current research conducted a systematic review on sample size and the use of inferential statistics in basic and applied behavioral research by assessing all experimental studies from 2009 to 2018 in the Journal of the Experimental Analysis of Behavior (JEAB) and the Journal of Applied Behavior Analysis (JABA) which was 1,155 articles. The use or non-use of inferential statistics in behavioral research remains controversial as visual inspection has deep historical roots. JEAB had a median number of eight subjects and JABA had a median number of four subjects, which was statistically different using Mood's median test. In addition, articles in JEAB were more likely to use inferential statistics compared to JABA. In general, inferential statistics were used in the presence of larger sample sizes, however, the use of animal subjects was associated with smaller sample sizes. Although patterns of the use of inferential statistics varied across journal, sample size, and species, this does not preclude the use of statistical methods by applied behavioral researchers, which should be used to support and confirm visual inspections of data.

Keywords: sample size, inferential statistics, behavioral research, systematic review.

Resumo

Esta pesquisa realizou uma revisão sistemática sobre o tamanho da amostra e o uso de estatísticas inferenciais na pesquisa comportamental básica e aplicada, avaliando todos os estudos experimentais de 2009 a 2018 no Journal of the Experimental Analysis of Behavior (JEAB) e no Journal of Applied Behavior Analysis (JABA) que somaram 1.155 artigos. O uso ou não de estatísticas inferenciais na pesquisa comportamental permanece controverso, pois a inspeção visual tem profundas raízes históricas. O JEAB teve um número mediano de oito sujeitos e o JABA teve um número mediano de quatro sujeitos, o que foi estatisticamente diferente usando o teste de mediana de Mood. Além disso, os artigos do JEAB eram mais propensos a usar estatísticas inferenciais em comparação com o JABA. Em geral, a estatística inferencial foi usada na presença de tamanhos amostrais maiores, no entanto, o uso de sujeitos animais foi associado a tamanhos amostrais menores. Embora os padrões de uso de estatísticas inferenciais variem entre periódicos, tamanho da amostra e espécie, isso não impede o uso de métodos estatísticos por pesquisadores comportamentais aplicados, métodos estes que devem ser usados para apoiar e confirmar inspeções visuais de dados.

Palavras-chave: tamanho de amostra, estatística, pesquisa comportamental, revisão sistemática.

✉ TCHatvany@ship.edu

DOI: http://dx.doi.org/10.18542/rebac.v18i2.13634

Systematic analyses of studies comparing trends in basic and applied behavioral research journals have assessed a variety of variables. The variables that have been examined include stimulus control (Starin, 1987), authorship and citation practices (Dymond, 1997; Elliott et al., 2005; Poling et al., 1994; Virues-Ortega, Hurtado-Parrado, Cox, & Pear, 2014), and the use of the term frequency (Carr, Nosik, & Luke, 2018). However, the authors are unaware of any studies that have performed a systematic analysis on a comparison of sample size across journals. Zimmerman, Watkins, and Poling (2015) did a content analysis on articles published in JEAB and reported that pigeons and humans were the most frequently used subjects. After the 1970s, pigeons and rats were used less often, and humans were used more often. The total and average number of subjects used across species were also provided, although no comparisons were made as it was not the focus of the study.

The issue of sample size is particularly relevant as psychology has seen significant efforts to encourage larger sample sizes in research, both recently (LeBel, Campbell, & Loving, 2017) and in past years (Marszalek, Barber, Kohlhart, & Holmes, 2011). Behavioral psychology has not been immune to this push, even though the field has historically utilized small-n research, focusing on continuous measurement, within-subject comparisons, and the impact of environmental variables on individual behavior (Sidman, 1960; Skinner, 1938). Shifting from the foundation of basic non-human research to applied research and treatment of humans reinforced the focus on behavior at the individual level. Thus, the need for inventions with “slam-bang” effects (Kazdin, 2011), or those that can easily be discerned through visual inspection and have a major impact on behavior, became the norm within the field of applied behavior analysis.

However, this certainly does not mean small-n research is immune from issues of current concern in the field of psychology, such as replication of effects (Hantula, 2019), and there have long been concerns regarding the subjectivity of visual inspection as the sole means of determination for treatment effects (Young, 2018). In addition, Ator (1999) suggested that those conducting behavioral research should be familiar with inferential statistics. The justification for doing so is to advocate for the appropriate use or non-use of inferential statistics given the research design and research questions so that the larger research community outside of behavioral research better understands the rationale for selecting a particular data analytic approach. Similarly, it has been suggested that the use of inferential statistics in behavioral research should be justified by authors, rather than simply used by default (Hopkins, Cole, & Mason, 1998).

Some within the field have lobbied for the increased use of between-subject designs, including the use of inferential statistics, in behavioral research, as this is considered the gold standard for establishing an intervention as evidence-based (Kazdin, 2011), and may be key to greater mainstream acceptance and usage of behavioral treatment techniques (Smith, 2012). Additionally, in light of findings that demonstrate that visual inspection is a subjective practice, moving towards the use of inferential statistics may be necessary (Fisch, 1998; Lane & Gast, 2013). However, inferential statistics require sufficient statistical power, which often requires an increased sample size (Cohen, 2013). A systematic examination of this topic within behavioral research comparing journals with a basic and applied emphasis would fill a current information gap in the field.

Number of subjects

Although a comparison of the number of subjects used in basic and applied behavioral research journals has not been directly investigated, the usage of small-n research designs has been examined (Beavers, & Iwata, 2013; Cooper, Heron, & Heward, 2007; Hanley et al., 2003; Kyonka et al., 2019; Zimmermann et al., 2015). Zimmerman and colleagues (2015) indicated an increase in the usage of inferential statistics in JEAB over time. It is of note that small-n research has shown to be vital for behavioral research because of the field’s individualized approach. Behaviors and their functions can vary immensely between individuals, and consequently, unique behavior change programs need to be developed and assessed by those in the field and often comparison between subjects is not possible (Kazdin, 2011). Additionally, in applied behavior analysis, research and treatment often occur simultaneously, thus necessitating an individualized approach.

Inferential statistics

The use of inferential statistics in behavioral research has been debated for decades (Barron, 1999; Branch, 1999; Sidman, 1960). However, recently there has been a call for increased use of statistics in behavioral research (Young, 2018). Despite the debate around inferential statistics and their uses in behavioral research, one thing that is clear is that relying on visual inspection alone is insufficient or potentially biased in some situations (Fisch, 1998).

While the use of inferential statistics has been examined in the behavioral literature (Kyonka, Mitchell, & Bizo, 2019; Zimmerman, Watkins, & Poling, 2015), no research to date has examined the use of inferential statistics in JABA, and no analyses have evaluated inferential statistics in relationship to sample size. Sample size is one of several determinates of statistical power, which is necessary for inferential statistics to be a valuable tool (Cohen, 2013). Statistical power itself is a concern in psychological research, where the estimated power across disciplines and in psychological research as a whole is underpowered (Maxwell, 2004; Maxwell, Lau, & Howard, 2015). In addition, past research has demonstrated that many researchers underestimate what is required for sufficient statistical power (Bakker, Hartgerink, Wicherts, & van der Maas, 2016). The choice to use or not use inferential statistics is based upon a variety of factors in a given study in addition to study design, sample size selection, and statistical power. Also related to those study characteristics is the species under investigation.

Type of species

Previous research has explored the different species of subjects used in behavioral research (Schmorrow, 1993; Zimmerman et al., 2015). The present research does not intend to explore which species are primarily present in behavioral research, but rather to understand how the use of different species might affect sample size. Access to research subjects of certain species is often limited by several practical factors, including lab space, size of the animal, expenses related to care, and animal life cycle (Bacchetti, Deeks & McCune, 2011). To date, however, the authors are unaware of any research that has examined how species and sample size are related to each other. Because the use of non-human species is common in JEAB (Schmorrow, 1993; Zimmerman et al., 2015), the present research aims to provide information regarding species use in behavioral research to better contextualize findings regarding sample size and the use of inferential statistics.

The Present Study

The present study examined sample size in experimental research studies published in JEAB and JABA, comparing articles in these basic and applied behavioral flagship research journals, respectively. Previous researchers have noted a high volume of small-n research designs that are best suited for sample sizes that are too small to be analyzed using traditional statistical methods (Cooper et al., 2007; Zimmermann et al., 2015); thus, the current study sought to answer how many subjects are used in small-n research, and if that differs across behavioral journals that emphasize basic and applied research. Additionally, the present study examined this behavioral research within the context of aspects of the research process that affect sample size (e.g., species under study) and are affected by sample size (e.g., statistical power and the use inferential statistics). These results will provide context for basic and applied behavioral research regarding sample size, species, and the use of inferential statistics.

Method

Sample and screening

The study analyzed articles published in JEAB and JABA from 2009-2018 resulting in 551 JEAB articles and 895 JABA articles. All 1,446 articles were examined, and articles were eliminated if they were review articles, technical reports, content analyses, or other non-experimental articles. This resulted in 412 (74.77%) JEAB articles and 743 (83.02%) JABA articles, for a total of 1,155 experimental articles.

Coding Procedures

After coding for experimental studies, all 1,155 remaining articles were coded for the number of subjects, number of studies, use of inferential statistics, and the use of human or non-human subjects. The coding procedure required raters to read the study abstract, method, and results (often a combined results and discussion section) sections of each study presented in each article.

Number of subjects was operationally defined as the number of different subjects used across all studies within each article and corresponded to the number of subjects used in the data analysis. Subjects whose data was not analyzed due to attrition during the study or other related factors that made their data unusable and were not analyzed in the article were not recorded. Coding of sample size was further broken down for both the article total and by study within the article; thus, the number of studies per article was also coded. If an article was composed of multiple studies with independent subjects, the number of different or independent subjects used across all the studies was recorded for the article total sample size. For each individual study, the sample size was recorded separately.

The use of inferential statistics was defined as the authors reporting a Frequentist or Bayesian statistical analysis conducted on data that was collected in the studies presented in the articles.

Human and non-human were coded based on how the subjects in these studies were described by the authors in the articles. Participants referred to as human, students, children or by name were coded as human and all others as non-humans.

Interrater Agreement

Interrater agreement was calculated for all steps of coding. Three research assistants were provided with operational definitions of the variables and discussions and clarifications were provided during a training session. Two research assistants independently coded all variables. When a discrepancy existed, the third research assistant would code the discrepant case and that coding was used in the analyses. The interrater reliability was as follows: Whether an article was an experimental study had an interrater agreement of 99.24%; the coding of total article sample size had an interrater agreement of 97.32%; the coding of individual study sample size had an interrater agreement of 95.24%; the coding of the use of inferential statistics had an interrater agreement of 95.76%; the coding of human or non-human subjects had an interrater agreement of 100%.

Results

Results indicated that the number of subjects used across the 10 years assessed in the present study ranged from 1 to 102,368 with a median value of 5 across all articles. The sample sizes in JEAB ranged from 1 to 486 with a median of 8, and the sample sizes in JABA ranged from 1 to 102,368 with a median of 4. The means from each journal have not been presented due to outliers and high level of skewness which would misrepresent the data if means were provided. Overall, the analyses demonstrated that a majority of the article sample sizes fell at the lower end of the data range. The medians of the two journals were compared using Mood's Median test. This test was selected because the data was not normally distributed for either JEAB or JABA with a skewness of 4.89 (S.E. = .12) and 27.23 (S.E. = .09), respectively. The results of the median test suggest that the median of JEAB (Mdn = 8) and JABA (Mdn = 4) were significantly different from one another χ^2 (1, N = 1,154) = 190.11, p < .001). Neither journal demonstrated trends in sample size differences over the 10-year period.

The data were further examined by creating categories of sample sizes. The categorizations of sample sizes included: 1, 2-4, 5-9, 10-24, 25-99, and >99. A chi-square analysis was conducted on the categories of sample sizes across journal, yielding a significant finding, χ^2 (1, N = 1,154) = 201.00, p < .001. In order to compare JEAB and JABA for each category, post hoc procedures detailed by Beasley and Schumacker (1995) were followed where Z-scores were derived using the standardized adjusted residuals and a Bonferroni adjustment (i.e., .05/6 = .008) was used as the adjusted p-value criteria to control for Type I errors. Next, p-values were calculated for each Z-score and assessed accordingly. Table 1 illustrates the frequency, percentage, Z-scores, and p-values for each category of sample size across journal. Consistent with the differences in medians, the findings indicate studies with a single subject, or 2-4 subjects are more prevalent in JABA, whereas studies with sample sizes of 5-9, 10-24, and 25-99 occur more often in JEAB.

Table 1
Frequency and Percentages of Samples Sizes in JEAB and JABA Articles

Sample Size	JEAB	JABA	Z
1	3 (0.73%)	89 (11.98%)	6.76 (p<.001) *
2 - 4	76 (18.45%)	369 (49.66%)	10.44 (p<.001) *
5 - 9	140 (33.98%)	137 (18.44%)	5.93 (p<.001) *
10 - 24	102 (24.76%)	77 (10.36%)	6.48 (p<.001) *
25 - 99	68 (16.50%)	52 (7.00%)	5.07 (p<.001) *
>99	23 (5.58%)	19 (2.56%)	2.63 (p = .009)

Note: N for JEAB = 412, N for JABA = 743. *indicates statistical significance after Bonferroni adjustment

The results were then broken down by individual study. First, JEAB had a higher average number of studies per paper of 1.55, as opposed to JABA's 1.17; this was analyzed using a Mann-Whitney test, $U(N_{JABA} = 743, N_{JEAB} = 410)$ = 115,230, z = -67.10, p < .001. Second, sample size was examined at the study level which revealed the same general trend of JEAB having more participants per study than JABA χ^2 (1, N = 1,504) = 133.23, p < .001). With the average study in JEAB having a median of 6 (range 1 -426), and the average study in JABA having a median of 4 (range 1 - 102,368).

The use of inferential statistics was also different between the two journals as significantly more articles used inferential statistics (χ^2 with Yate's continuity correction, (1, N = 1,154) = 341.23, p < .001) in JEAB (63.26%) as compared to JABA (11.17%). Sample size was also a significant predictor of the use of inferential statistics χ^2 (1, N = 1,154) = 226.13, p < .001). Specifically, studies that used inferential statistics had a median sample size of 13 (range 1 – 1822) whereas the median sample size for studies that did not use inferential statistics was 4 (range 1 - 102,368). This prompted a more detailed examination of studies that would be categorized as small-n. Thus, all studies with sample sizes less than 5 were extracted for a follow-up analysis in order to compare journals relative to the use of inferential statistics in an effort to control for the effect of sample size and make comparisons across journals. Even with small-n studies, a similar pattern emerged such that 42.3% of articles in JEAB used inferential statistics compared to only 3.3% of articles in JABA χ^2 (1, N = 539) = 125.43, p < .001).

To analyze the effect of human and non-human subjects across both journals, 11 articles were initially excluded as they had both human and non-human participants; however, including them on either side (with the humans or with the non-humans) had no impact on the outcome of the results. When all articles across JEAB and JABA were combined, the use of species other than humans was a significant predictor of sample size χ^2 (1, N = 1,143) = 58.56, p < .001). Within this combined analysis, the median sample size for human subject research was 4 (range 1 - 102,368) and research using non-human subjects had a median sample of 6 (range 1 - 94). However, when articles within JEAB were examined separately, the opposite relationship was found as the use of species other than humans was associated with a smaller sample size χ^2 (1, N = 403) = 62.34, p < .001). Studies in JEAB using non-human subjects had a median sample of 6 (range 1 - 80) whereas studies using human subjects had a median of 20 (ranger 1 - 486). The same analysis was not possible in JABA as it lacked sufficient studies using non-human participants, leading us to infer that the overall effect is driven by, on average, smaller samples in JABA, which are substantially more likely to consist of humans (χ^2 with Yate's continuity correction, (1, N = 1,143) = 556.21, p < .001), with 59.29% of articles in JEAB using non-human subjects and 1.23% of articles in JABA using non-human subjects. Thus, it appears that the use of non-human subjects was associated with a smaller median sample size in experimental behavioral research and that there is insufficient data to draw a conclusion for applied research.

Discussion

Research findings

The results indicate that behavioral studies published in JABA have smaller sample sizes than studies published in JEAB. This pattern was observed in the comparison of medians and the categories of sample sizes. However, a closer examination of the data yielded an unusual pattern such that the nine studies with the largest sample sizes were all published in JABA (i.e., sample sizes ranging from 508 to 102,368). Studies in both journals ranged from single participant studies to some extremely large studies. The largest study in JEAB during this time period had 486 participants and focused on applying behavioral principles in the prisoner's dilemma game (Locey, Safin, & Rachlin, 2013). The largest study in JABA during this time period was of 102,368 individuals examining how food purchasing behavior could be influenced through environmental modification (Sigurdsson, Larsen, & Gunnarsson, 2013). Additionally, the results demonstrated larger sample sizes were associated with the use of inferential statistics and the use of human subjects. These results were expected as the use of inferential statistics frequently relies on sufficient sample sizes to achieve statistical power (Cohen, 2013) and practical considerations related to the access and maintenance of non-human subjects typically limits the ability to obtain a large sample (Bacchetti, Deeks, & McCune, 2011). JABA also demonstrated substantially less use of inferential statistics. This was also expected given the reduced sample size, but is in line with the preference for small-n studies for reasons of feasibility in applied research (Kazdin, 2011)

Sample size in behavioral research

One of the initial decisions when designing a research study is the number of subjects to include. To understand why the trends related to sample size may have occurred, previously assessed variables need to be considered. The most important of these related variables are the research designs used in behavioral research. As noted in previous literature, behavioral researchers have typically used small-n research designs that can assess samples as small as one subject (Cooper et al., 2007; Kazdin, 2011). Because many subjects exhibit behaviors that are unique in their topography and overall function, assessing subjects using larger, between-subject designs is often not feasible (Kazdin, 2011). It appears as though articles in JABA conduct studies with smaller sample sizes where inferential statistics may be used less often, and Branch (1999) suggested there may be a general reluctance among behavior analysts to use null hypothesis significance testing (NHST). However, articles in JEAB were more prone to use statistical analyses. In fact, other researchers (Kyonka, et al., 2019; Zimmerman, et al., 2015) have recently shown that there has been an increase in the use of inferential statistics in JEAB over time, perhaps due to concerns related to the resulting statistical power of a study, as other journals and grant applications may require or prefer power analyses and/or inferential statistics.

Also of note was that when specifically examining studies with small sample sizes (i.e., <5), over 40% of articles in JEAB used inferential statistics compared to 3% in JABA. There is a growing literature addressing the issue of using data analytic techniques in addition to visual analysis. Some approaches that have recently been suggested include the use of linear mixed-effects modeling (Wiley & Rapp, 2019), interrupted time series (Harrington & Velicer, 2015), simulation modeling (Borckardt & Nash, 2014), generalized logistic model (Verboon & Peters, 2020), mediation analysis(Geuke, Maric, Miočević, Wolters, & Haan, 2019), calculating effect sizes (Hedges, Pustejofsky, & Shadish, 2012), and applying Bayesian models (Natesun, 2019). Thus, there are ample data analytic options available for small-n research designs.

The statistical power of a study is assessed through a power analysis and may be becoming a more important issue in behavioral research. A power analysis indicates the probability a study will lead to the rejection of the null hypothesis and that the result of the study is not due to chance, resulting in either Type I or Type II error (Cohen, 2013). Among the multiple factors which influence the power of a study, one of the most important, as indicated by Cohen (2013), is the sample size of the study. In a series of simulation studies, statistical tests were found to likely run into either Type I or Type II error when sample sizes were equal to or less than five, unless extremely large effect sizes were present. However, sample sizes above five, which made up most of the articles assessed, could allow for the effective use of statistical tests, despite the presence of low power that could hinder their chances of publication (de Winter, 2013). With the median sample size in JABA in the present study being four, most of the empirical articles would fall within the category assessed in the simulation study, supporting the use of small-n research designs.

This may also support recent increases in JEAB in the use of inferential statistics (Kyonka et al., 2019; Zimmermann et al., 2015) which does contrast with the views of some (e.g., Branch, 1999) who posited that inferential statistics are not particularly useful in studies of behavior. With the power of a study still an important factor, some researchers may favor alternative data analytic methods (e.g., visual inspection) when using smaller sample sizes. Additionally, without the burden of having to meet a desired statistical power to effectively assess their results while conducting small-n research, behavioral researchers can focus their attention on developing an intervention that best serves the subjects they are assessing in their studies.

Limitations

Although the findings of this systematic investigation are substantial, there are some limitations that should be noted. First, the type of research (basic or applied) was not coded in the current study. Although both journals generally publish research in their respective areas, articles featuring applied research do appear in JEAB, and more basic research does appear in JABA. Future research exploring sample size and these content areas may benefit from examining this additional variable.

Second, the lack of reported sample sizes in some empirical studies was an issue. Thirty-one studies had empirical data that was reported but lacked reported sample sizes. Many of these studies were field studies where collecting sample size data was not feasible. For example, O'Connor, Lerman, Fritz, and Hodde (2010) conducted a recycling intervention. Their measurement involved recycled bottle counts, which was not representative of the number of individuals involved in the intervention. Although the number of studies without reported sample sizes is small in comparison to the overall data set, the vast majority of these studies appeared in JABA (n =29) and not JEAB (n = 2). This limitation might be addressed in future research by comparing other variables, such as relative foot-traffic or site location counts, among those empirical studies without reported sample sizes.

Third, the type of study design used by the researchers was not coded. Several factors related to sample size selection were not coded for in this investigation such as number of independent variables and their levels, whether those independent variables were between or within subject variables, or the number of time measurements. Additionally, actual statistical power of each study was not coded for either. All these factors may be important in relation to the usage of inferential statistics and might represent an interesting study. All of these are connected to sample size selection and statistical power and are relevant factors that researchers should consider when deciding on a sample size for their research (Cohen, 2013). Finally, a variable related to this topic, measurement accuracy, was not coded. Measurement accuracy, or rather inaccuracy, does affect statistical power. However, we would like to note that measurement accuracy should be ensured for reasons beyond that of ensuring statistical power, and while it is related to the decision to use inferential statistics, it is first and foremost related to the decision to use that measurement tool.

Future directions

The aim of the current study was to explore and compare the relative sample sizes within these areas of behavioral research. The authors made no attempt to compare relative conclusions of the results or findings of these studies, or to assess their ability to be reproduced. However, it should be noted that the sample sizes explored in many of these studies are comparatively small when compared to other fields of psychology. For example, Marszalek and colleagues (2011) examined sample sizes across four top journals in psychology in different content areas over several different years, 2006 being the most relevant to this present research. They found that the smallest median from 2006 was in experimental psychology with a median number of subjects of 12, with only 25% of studies having 10 or fewer participants (Marszalek et al., 2011). Given the nature and history of most behavioral research studies in basic and applied settings, it was not surprising that the median was five. However, this discrepancy may be a source of contention between the behavioral discipline and other areas of psychology.

Those conducting future behavioral small-n research may consider combining traditional visual inspection methodologies and other inferential statistics appropriate for small-n or even single case designs (e.g., Borckardt & Nash, 2014; Geuke, Maric, Miocevic, Wolters, & Hann, 2019; Harrington & Velicer, 2015; Hedges, Pustejofsky, & Shadish, 2012; Natesun, 2019; Verboon & Peters, 2020; Wiley & Rapp, 2019) or by increasing statistical power by increasing one's sample size or change in their study design (Cohen, 2013). Either way it is our recommendation that future behavioral research using inferential statistics carefully consider statistical power.

It should be noted that the use of Null hypothesis testing (NHST) itself is often controversial with many detractors (Szucs, & Ioannidis, 2017; Wagenmakers, 2007; Wasserstein, Schirm, & Lazar, 2019), however, many alternatives exist within inferential statistics such as Bayesian inference (Wagenmakers, 2007), which is currently on the rise within psychology (van de Schoot, Winter, Ryan, Zondervan-Zwijnenburg, & Depaoli, 2017). Further, particularly for behavioral research, there are other lesser well-known approaches such as OOM (Grice, Yepez, Wilson, & Shoda, 2016) and using participants as effect sizes (Grice et al., 2020). While behaviorists might be reluctant to adopt NHST inference, other inferential options might serve behaviorism well. Using inferential statistics in conjunction with more traditional visual inspection methods would allow both basic and applied behavioral researchers, but particularly applied researchers, to further confirm their analyses; this might represent a new form of mixed methodology that might be acceptable to behavioral and non-behavioral researchers so that some data analytic commonality exists.

Conclusion

The present study analyzed the differences between sample size and use of inferential statistics in basic and applied behavioral research over a ten-year period in the two flagship behavioral research journals, JEAB and JABA. The findings suggest that, in general, basic research tends to have larger sample sizes, despite some significantly large studies in JABA and the practical limitations of using non-human subjects. Generally, basic research is more likely to use inferential statistics and this may be related to the larger sample sizes used in the sub-discipline. Although the nature of applied behavioral research generally lends itself to having a smaller sample size, this does not preclude the use of inferential tools by researchers to support and confirm the findings of visual analyses.

Declaration of Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Contribution of each author

All the authors have equally contributed to complete the manuscript.

Copyright

References

Ator, N.A. (1999). Statistical inference in behavior analysis: Environmental determinants? *The Behavior Analyst, 22*(3), 93-97. https://doi.org/10.1007/BF03391985

Bacchetti, P., Deeks, S. G., & McCune, J. M. (2011). Breaking free of sample size dogma to perform innovative translational research. *Science of Translational Medicine, 3*(87), 1-4. https://doi.org/10.1126/scitranslmed.3001628

Bakker, M., Hartgerink, C. H. J., Wicherts, J. M., & van der Maas, H. L. J. (2016). Researchers' intuitions about power in psychological research. *Psychological Science, 27*(8), 1069–1077. https://doi.org/10.1177/0956797616647519

Baron, A. (1999). Statistical inference in behavior analysis: Friend or foe? *The Behavior Analyst, 22*(2), 83–85. doi:10.1007/bf03391983

Beasley, T. M., & Schumacker, R. E. (1995). Multiple regression approach to analyzing contingency tables: Post hoc and planned comparison procedures. *The Journal of Experimental Education, 64*(1), 79-93. https://doi.org/10.1080/00220973.1995.9943797

Beavers, G. A., & Iwata, B. A. (2011). Prevalence of multiply controlled problem behavior. *Journal of Applied Behavior Analysis, 44*(3), 593–597. https://doi.org/10.1901/jaba.2011.44-593

Borckardt, J. J., & Nash, M. R. (2014). Simulation modelling analysis for small sets of single-subject data collected over time. *Neuropsychological Rehabilitation, 24*(34), 492–506. doi:10.1080/ 09602011.2014.895390

Branch, M. N. (1999). Statistical inference in behavior analysis: Some things significance testing does and does not do. *The Behavior Analyst, 22*(2), 87-92. https://doi.org/10.1007/BF03391984

Carr, J. E., Nosik, M. R., & Luke, M. M. (2018). On the use of the term 'frequency' in applied behavior analysis. *Journal of Applied Behavior Analysis, 51*(2), 436–439. https://doi.org/10.1002/jaba.449

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences.* Routledge. https://doi.org/10.4324/9780203771587

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nded.). Upper Saddle River, NJ: Pearson Education, Inc.

de Winter, J.C.F. (2013). Using the Student's t-test with extremely small sample sizes. *Practical Assessment, Research, and Evaluation, 18*(10), https://doi.org/10.7275/e4r6-dj05

Dymond, S. (1997). International publication trends in the experimental analysis of behavior. *The Behavior Analyst, 20*(2), 109–119. https://doi.org/10.1007/BF03392768

Elliott, A. J., Morgan, K., Fuqua, R. W., Ehrhardt, K., & Poling, A. (2005). Self- and cross-citations in the Journal of Applied Behavior Analysis and the Journal of the Experimental Analysis of Behavior: 1993–2003. *Journal of Applied Behavior Analysis, 38*(4), 559–563. https://doi.org/10.1901/jaba.2005.133-04

Fisch, G. S. (1998). Visual inspection of data revisited: Do the eyes still have it? *The Behavior Analyst*, *21*(1), 111–123. https://doi.org/10.1007/bf03392786

Geuke, G.G.M., Maric, M., Miocevic, M., Wolters, L.H., & Haan, E. (2019). Testing mediators of youth intervention outcomes using single-case experimental designs. *New Directions for Child and Adolescent Development, 2019*(167), 39-64. https://doi.org/10.1002/cad.20310

Grice, J. W., Medellin, E., Jones, I., Horvath, S., McDaniel, H., O'lansen, C., & Baker, M. (2020). Persons as effect sizes. *Advances in Methods and Practices in Psychological Science, 3*(4), 443–455. https://doi.org/10.1177/2515245920922982

Grice, J. W., Yepez, M., Wilson, N. L., & Shoda, Y. (2016). Observation-oriented modeling: Going beyond "Is it all a matter of chance"? *Educational and Psychological Measurement, 77*(5), 855–867. https://doi.org/10.1177/0013164416667985

Hanley, G. P., Iwata, B. A., & McCord, B. E. (2003). Functional analysis of problem behavior: A review. *Journal of Applied Behavior Analysis, 36*(2), 147–185. https://doi.org/10.1901/jaba.2003.36-147

Hantula, D. A. (2019). Replication and reliability in behavior science and behavior analysis: A call for a conversation. *Perspectives on Behavior Science, 42*(2), 1-11. https://doi.org/10.1007/s40614-019-00194-2

Harrington, M., & Velicer, W. F. (2015). Comparing visual and statistical analysis in single-case studies using published studies. *Multivariate Behavioral Research, 50*(2), 162-183. https://doi.org/10.1080/00273171.2014.973989

Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2012). A standardized mean difference effect size for single case designs. *Research Synthesis Methods, 3*, 224 – 239. https://doi.org/10.1002/jrsm.1052

Hopkins, B.L., Cole, B.L., & Mason, T.L. (1998). A critique of the usefulness of Inferential statistics in applied behavior analysis. *The Behavior Analyst, 21*(1), 125-37. https://doi.org/10.1007/BF03392787

Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings* (2nd ed.). New York, NY, US: Oxford University Press.

Kyonka, E. G. E., Mitchell, S. H., & Bizo, L. A. (2019). Beyond inference by eye: Statistical and graphing practices in JEAB, 1992-2017. *Journal of the Experimental Analysis of Behavior, 111*(2), 155-165. https://doi.org/10.1002/jeab.509

Lane, J. D. & Gast., D. L. (2014). Visual analysis in single case experimental design studies: Brief review and guidelines. *Neuropsychological Rehabilitation, 24*(3-4), 445-463. https://doi.org/10.1080/09602011.2013.815636

LeBel, E. P., Campbell, L., & Loving, T. J. (2017). Benefits of open and high-powered research outweigh costs. *Journal of Personality and Social Psychology, 113*(2), 230–243. https://doi.org/10.1037/pspi0000049

Locey, M. L., Safin, V., & Rachlin, H. (2013). Social discounting and the prisoner's dilemma game. *Journal of the Experimental Analysis of Behavior, 99*(1), 85–97. https://doi.org/10.1002/jeab.3

Marszalek, J. M., Barber, C., Kohlhart, J., & Holmes, C. B. (2011). Sample size in psychological research over the past 30 years. *Perceptual and Motor Skills, 112*(2), 331–348. https://doi.org/10.2466/03.11.PMS.112.2.331-348

Maxwell, S. E. (2004). The persistence of underpowered studies in psychological research: Causes, consequences, and remedies. *Psychological Methods, 9*(2), 147–163. https://doi.org/10.1037/1082-989X.9.2.147

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does "failure to replicate" really mean? *American Psychologist, 70*(6), 487–498. https://doi.org/10.1037/a0039400

Natesan, P. (2019). Fitting Bayesian models for single-case experimental designs: A tutorial. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 15*(4), 147-156. https://doi.org/10.1027/1614-2241/a000180

O'Connor, R. T., Lerman, D. C., Fritz, J. N., & Hodde, H. B. (2010). Effects of number and location of bins on plastic recycling at a university. *Journal of Applied Behavior Analysis, 43*(4), 711-715. https://doi.org/10.1901/jaba.2010.43-711

Poling, A., Alling, K., & Fuqua, R. W. (1994). Self- and cross-citations in the Journal of Applied Behavior Analysis and the Journal of the Experimental Analysis of Behavior: 1983–1992. *Journal of Applied Behavior Analysis, 27*(4), 729–731. https://doi.org/10.1901/jaba.1994.27-729

Schmorrow, D. D. (1993). *The use of nonhuman subjects in behavior analysis: A review of JEAB studies.* [Unpublished doctoral dissertation]. Western Michigan University. https://scholarworks.wmich.edu/dissertations/1881

Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology.* Basic Books, Inc.

Sigurdsson, V., Larsen, N. M., & Gunnarsson, D. (2013). Healthy food products at the point of purchase: an in-store experimental analysis. *Journal of Applied Behavior Analysis, 47*(1), 151–154. https://doi.org/10.1002/jaba.91

Skinner, B.F. (1938) *The behavior of organisms: An experimental analysis.* D. Appleton-Century.

Smith, T. (2012). Evolution of research on interventions for individuals with autism spectrum disorder: Implications for behavior analysts. *The Behavior Analyst, 35*, 101-113. https://doi.org/10.1007/BF03392269

Starin, S. P. (1987). The trend of stimulus control publications. *The Behavior Analyst, 10*(1), 133–134. https://doi.org/10.1007/bf03392423

Szucs, D., & Ioannidis, J. P. A. (2017). When null hypothesis significance testing is unsuitable for research: A reassessment. *Frontiers in Human Neuroscience, 11*. https://doi.org/10.3389/fnhum.2017.00390

van de Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijnenburg, M., & Depaoli, S. (2017). A systematic review of Bayesian articles in psychology: The last 25 years. *Psychological Methods, 22*(2), 217–239. https://doi.org/10.1037/met0000100

Verboon, P., & Peters, G.J.Y. (2020). Applying the generalized logistic model in single case designs: Modeling treatment-induced shifts. *Behavior Modification, 44*(1), 27-48. https://doi.org/10.1177/0145445518791255

Virues-Ortega, J., Hurtado-Parrado, C., Cox, A. D., & Pear, J. J. (2014). Analysis of the interaction between experimental and applied behavior analysis. *Journal of Applied Behavior Analysis, 47*(2), 380–403. https://doi.org/10.1002/jaba.124

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review, 14*, 779–804.

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond “p < 0.05.” *The American Statistician, 73*(sup1), 1–19. https://doi.org/10.1080/00031305.2019.1583913

Wiley, R. W, & Rapp, B. (2019). Statistical analysis in small-N designs: Using linear mixed-effects modeling for evaluating intervention effectiveness, *Aphasiology*, *33*, 1- 30. https://doi.org/10.1080/02687038.2018.1454884

Young, M. E. (2018). A place for statistics in behavior analysis. *Behavior Analysis: Research and Practice, 18*(2), 193-202. http://dx.doi.org/10.1037/bar0000099

Zimmermann, Z. J., Watkins, E. E., & Poling, A. (2015). JEAB research over time: Species used, experimental designs, statistical analyses, and sex of subjects. *The Behavior Analyst, 38*, 203- 218. https://doi.org/10.1007/s40614-015-0034-5

Submitted on: 20/04/2021
Accept on: 16/04/2022