BEHAV ANALYST (2015) 38:203–218
DOI 10.1007/s40614-015-0034-5

JEAB Research Over Time: Species Used, Experimental Designs, Statistical Analyses, and Sex of Subjects

Zachary J. Zimmermann[1] · **Erin E. Watkins**[1] · **Alan Poling**[1]

Published online: 25 July 2015

Abstract We examined the species used as subjects in every article published in the *Journal of the Experimental Analysis of Behavior* (*JEAB*) from 1958 through 2013. We also determined the sex of subjects in every article with human subjects (N=524) and in an equal number of randomly selected articles with nonhuman subjects, as well as the general type of experimental designs used. Finally, the percentage of articles reporting an inferential statistic was determined at 5-year intervals. In all, 35,317 subjects were studied in 3,084 articles; pigeons ranked first and humans second in number used. Within-subject experimental designs were more popular than between-subjects designs regardless of whether human or nonhuman subjects were studied but were used in a higher percentage of articles with nonhumans (75.4 %) than in articles with humans (68.2 %). The percentage of articles reporting an inferential statistic has increased over time, and more than half of the articles published in 2005 and 2010 reported one. Researchers who publish in *JEAB* frequently depart from Skinner's preferred research strategy, but it is not clear whether such departures are harmful. Finally, the sex of subjects was not reported in a sizable percentage of articles with both human and nonhuman subjects. This is an unfortunate oversight.

Keywords Experimental subjects · Nonhuman · Experimental design · Statistical analyses · Inferential statistics · Gender

Throughout his career, B. F. Skinner favored the use of within-subject experimental designs to intensively study a small number of subjects, whose data were displayed graphically and analyzed via visual inspection alone (e.g., Skinner 1934a, 1934b, 1936, 1940). Sidman (1960) codified this approach to research, which Skinner (1938) termed

✉ Alan Poling
alan.poling@wmich.edu

[1] Department of Psychology, Western Michigan University, 1903 W. Michigan Ave, Kalamazoo, MI 49008-5439, USA

the experimental analysis of behavior (EAB), in *Tactics of Scientific Research.* EAB is very different from the approach to research favored by many psychologists, which involves between-subjects experimental designs and statistical data analysis, and provides substantially different information about behavior and the variables that control it (e.g., Johnston and Pennypacker 2008; Poling, Methot, and LeSage 1995).

In 1958, Skinner et al. founded a journal uniquely dedicated to EAB, as evident in its title, the *Journal of the Experimental of Behavior* (*JEAB*). Since its inception, *JEAB* has been the flagship journal for EAB (Lattal 2008; Saville, Epting, and Buskist 2002), and researchers have plumbed its pages for information about the discipline and how it has changed over time (e.g., Elliott, Morgan, Fuqua, Ehrhardt, and Poling 2005; Poling et al. l98l; Saville, Epting, and Buskist 2002). For example, Foster, Jarema, and Poling (1999) determined the percentage of *JEAB* articles that reported inferential statistics in 1960, 1965, 1970, 1975, 1980, 1985, 1990, and 1995 and found a generally increasing trend across time, which was well described by a regression line with a slope of 1.09. This trend is interesting given that Skinner and other behavior analysts (e.g., Branch 1999; Hopkins, Cole, and Mason 1998; Perone 1999) have been highly critical of the use of inferential statistics. To our knowledge, however, no one has determined if the upward trend reported by Foster et al. continued after 1995. One purpose of the present research was to do so.

A second purpose was to determine the relative popularity of within-subject and between-subjects experimental designs in articles published in *JEAB* from its inception through 2013. The use of inferential statistics and between-subjects research designs go hand and glove; therefore, the data reported by Foster et al. (1999) suggest that the use of between-subjects designs in *JEAB* increased from 1960 through 1995. But this is not necessarily the case, because it is certainly possible to use inferential statistics to analyze the data from within-subject and from mixed within- and between-subjects research, and some authors have proposed that doing so is advantageous (e.g., Gentile, Roden, and Klein 1972; Jones, Vaught, and Weinrott 1977). Furthermore, between-subjects designs do not necessarily require the use of statistical data analyses. Experimental design and data analysis are separate aspects of an experiment, both of which influence what can be learned from a study (Johnston and Pennypacker 2008; Poling et al. 1995), and examining both aspects of studies published in *JEAB* may be of interest to many behavior analysts.

Michael (1974) contended that interest in the use of inferential statistics in behavior analysis was fostered in large part by the misconceptions that it "is not possible to effect adequate levels of experimental control with much human applied research, and that in such cases a significance test would be quite valuable as a judgemental aid…" (p. 647). Although a few applied studies were published in early volumes of *JEAB*, this no longer occurs. But prominent behavior analysts have long argued for the importance of basic studies of human operant behavior (e.g., Hake 1982; Nevin 1982; Skinner 1953), and *JEAB* has for many years been an outlet for this kind of research (Kangas and Hackenberg 2009; Saville et al. 2002; Sokolowski, Disma, and Abramson 2010). Practical constraints may make it more difficult to use within-subject designs to conduct basic research with humans than with laboratory animals, and our informal reading of *JEAB* prior to the present review suggested that a higher proportion of human than nonhuman studies published therein used between-subjects designs and

statistical data analysis. We analyzed *JEAB* studies with human and nonhuman subjects separately to test this proposition.

In addition to recording data about research design and data analysis, we reported the specific type of subject used in each article and their reported sex. Others have previously reported on the species used in *JEAB* articles (Grossett et al. l982; Kangas and Hackenberg 2009; Saville et al. 2002; Sokolowski et al. 2010), a variable that may influence the generality of findings and is of interest for that reason. Therefore, it was worthwhile to attempt to replicate as well as to slightly extend their findings. It is widely recognized that the sex of subjects sometimes, but not invariably, influences how they respond to a given independent variable (Poling et al. 2009; Porter, Christian, and Poling 2003; Watkins, Zimmermann, and Poling 2014). Therefore, the sex of subjects should be routinely reported and, where appropriate, examined as a determinant of behavior.

Method

Data Collection

The present review comprised research articles published in *JEAB* from 1958, when the journal was founded, through 2013. Review articles, technical reports, and other articles that did not describe behavioral studies were excluded. Across all years, a total of 3,084 articles were rated. The total numbers of articles rated for successive decades from the 1950s through the 2010s were 65, 639, 842, 531, 478, 375, and 154, respectively.

A total of 524 articles involving only human subjects, all the articles with only humans that appeared in *JEAB*, were evaluated for research design, statistical analyses, and subject demographics in the present review. To allow for comparison, an additional 524-article sample of articles involving only nonhuman subjects was analyzed in the same manner. Due to the time and resources required to analyze articles in the aforementioned manner, the remaining articles with nonhuman subjects (N=2,036) were rated only for subject demographics. The nonhuman articles chosen for review were selected at random via a random number generator. In this process, all nonhuman articles from each decade first were assigned sequential numbers (e.g., 1 through 65 for the 1950s). Numbers generated by the random number generator were then used to select specific articles from the database for a particular year, with the provision that the number of articles with nonhuman subjects selected for a given year was the same as the number of articles with human subjects published in that year.

A single rater recorded the species used and the total number of subjects reported in each article. Subjects used in more than one experiment in a single publication were counted only once. In cases where authors indicated that data for the same subjects were reported in more than one publication, those subjects were counted only once. When subjects began a study but their data were not reported (as when a subject died during the course of an experiment), those subjects were not counted.

For all articles involving human subjects (N=524) and nonhuman subjects (N=524), the rater also recorded the reported sex of subjects and the type of research design that was used. Designs were categorized as within-subject, between-subjects, or mixed, as

defined by Poling et al. (1995). In brief, experiments in which the effects of the independent variable(s) were ascertained by comparing data from different subjects exposed to different conditions were considered as between-subjects designs. Experiments in which the effects of the independent variable(s) were ascertained by comparing data from the same subjects when they were exposed to different conditions were considered as within-subject designs. Studies that used any combination of these two strategies were considered as mixed designs.

In keeping with the methods of the trend analysis conducted by Foster et al. (1999), which reviewed all articles (human and nonhuman) in 5-year intervals, every article (human and nonhuman) published in 2000, 2005, 2010, and 2013 was examined by the rater to determine whether any inferential statistic, as defined by Foster et al., was reported. Foster et al. did not record data separately for studies involving human and nonhuman subjects, but we did do so. A second rater independently scored 10 % of the overall articles, selected at random. A percentage measure of interobserver agreement was determined by dividing the total number of ratings where the two observers agreed (agreements) by the total number of ratings where they agreed plus the total number where they disagreed (agreements + disagreements) and multiplying by 100. Interobserver agreement was 96 %. The two raters discussed all items where their ratings disagreed and reached consensus on the value that should be, and was, reported.

Results

Species Used From *JEAB*'s inception through 2013, pigeons, rats, humans, nonhuman primates, and miscellaneous other species were used in 47.9, 24.2, 17.0, 7.5, and 5.2 % of the research articles published therein, respectively. Two or more species were used in 1.6 % of these articles. Overall, 0.3 % of the articles included both human and nonhuman subjects. Figure 1 depicts for each decade the percentage of *JEAB* articles using pigeons, rats, nonhuman primates, humans, and other species as subjects. The percentage of studies using two or more species is also shown. These data and all other data reported are presented by decade, rather than by year, to simplify description. The decade beginning with 1950 comprises 2 years (1958 and 1959) and the decade beginning with 2010 comprises 4 years (2010–2013). All other decades comprise 10 years. Our comparison of figures showing data for each year and for each decade indicated that the two forms of data display do not support substantially different conclusions, although the former display illustrates variability not evident in the latter display. Others (Kangas and Hackenberg 2009; Saville et al. 2002; Sokolowski et al. 2010) have published figures presenting data for the subjects used in *JEAB* research during many of the years we examined.

The percentage of studies using pigeon subjects increased over the first three decades of *JEAB*'s existence, peaked in the 1970s and 1980s, fell a bit in the 1990s, and remained fairly stable thereafter. The percentage of studies with rats, the most commonly used species in the first decade, decreased until the 1980s and has been fairly consistent since that time. Studies with humans have generally increased in relative prevalence over time, whereas studies with nonhuman primates have decreased to a very low level. Studies with other species, always rare, decreased from the 1950s through the 1980s, then increased in prevalence in the 1990s to a low level that has

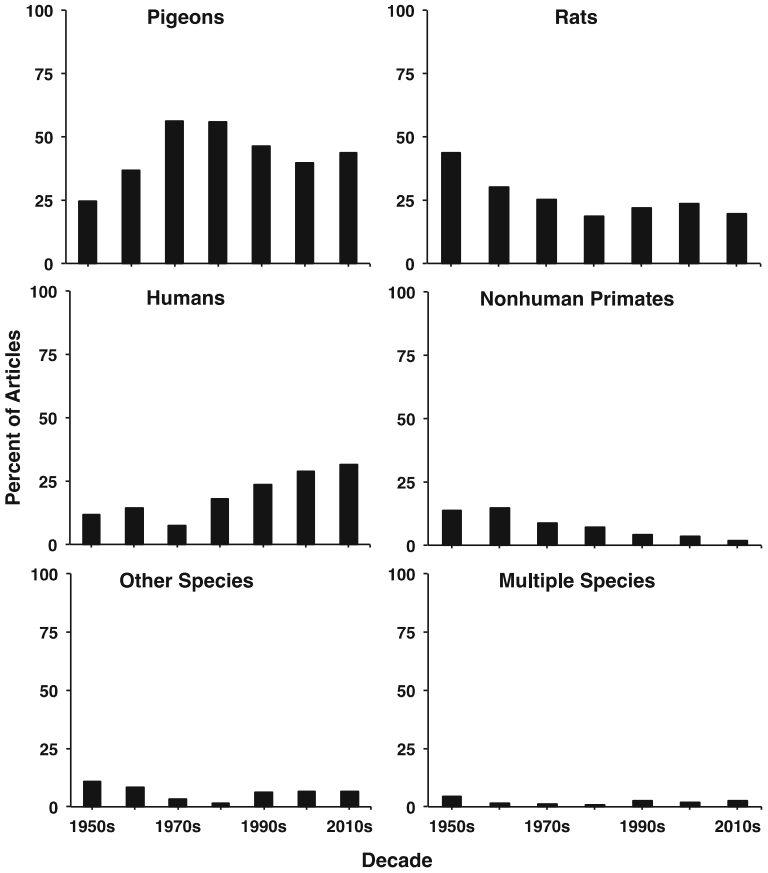

Fig. 1 Percentage of articles in which members of the indicated species served as subjects

remained relatively consistent. Regardless of the decade, very few *JEAB* articles have involved more than one species.

Table 1 shows for each species the total number of subjects used by decade and across all decades combined. Overall, the number of pigeons (11,989) used exceeded the number of subjects of any other species, although nearly as many humans (11,641) were used. For pigeons and rats, the number of subjects used reached a peak in the 1970s and declined thereafter, whereas the number of humans used increased from the 1970s onward. The number of nonhuman primates and other species used was relatively low each decade, with no clear trends evident in these data.

Experimental Design Figure 2 shows for each decade and for all decades combined the percentage of articles with human subjects that used within-subject, between-subjects, and mixed experimental design. Figure 3 provides comparable data for articles with nonhuman subjects. With one exception, perhaps due to the small sample of studies with nonhuman subjects published in the 1950s, a substantial majority of articles involved within-subject designs, regardless of whether the subjects were humans or nonhumans. No clear trends across decades in any of the categories were evident. When data for all individual years were compared via a *t* test, there was a

Table 1 Number of subjects used in *JEAB* studies

Species	Decade							
	1950s	1960s	1970s	1980s	1990s	2000s	2010s	Total
Rats	286	1,779	2,852	1,168	1,445	1,025	449	9,004
Pigeons	82	1,964	4,202	2,341	1,757	1,224	419	11,989
Humans	386	1,436	970	1,599	1,789	3,112	2,349	11,641
Nonhuman primates	28	356	319	148	104	63	8	1,026
Other species	23	362	304	182	146	323	137	1,477
Total	805	5,897	8,647	5,438	5,241	5,747	3,362	35,137

statistically significant difference at the .05 α level between the mean percentage of articles with human subjects that involved within-subject experimental designs (M= 68.2 %; SD=11.4) and the mean percentage of articles with nonhuman subjects (M= 75.4 %; SD= 2.8) that did so (t=−2.22, df=110, p=.028).

For each decade, between-subjects experimental designs were used in a higher percentage of articles with human subjects than in articles with nonhuman subjects. A t test indicated that there was a statistically significant difference at the .05 level between the mean yearly percentage of articles with human subjects that involved between-subjects experimental designs (M=16.4 %; SD=2.8) and the mean percentage of articles with nonhuman subjects (M= 4.3 %; SD= 3.1) that did so (t=5.94, df=110, p<.001).

In general, more articles that involved nonhuman subjects used mixed experimental designs than between-subjects experimental designs, but the reverse was true for articles that involved human subjects. There was, however, no statistically significant difference at the .05 α level in the mean yearly percentage of articles with human

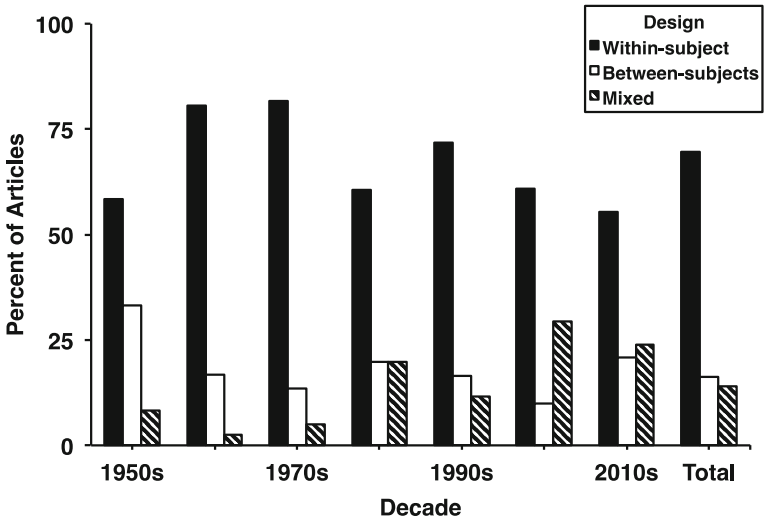

Fig. 2 Percentage of articles with human subjects in which the indicated type of experimental design was used

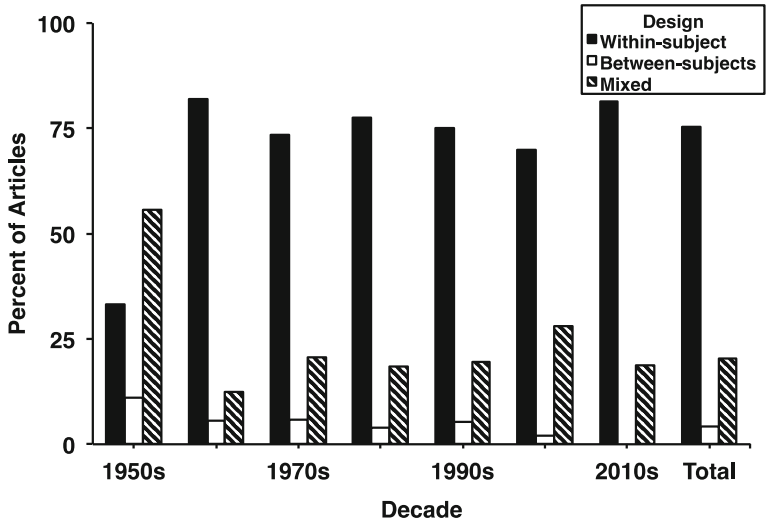

Fig. 3 Percentage of articles with nonhuman subjects in which the indicated type of experimental design was used

subjects that involved mixed experimental designs (M=15.4 %; SD=11.5) and the mean percentage of articles with nonhuman subjects (M= 20.3 %; SD= 2.9) that did so (t=−1.67, df=110, p=.098).

Inferential Statistics Figure 4, which shows data from Foster et al. (1999) as well as our findings, depicts at 5-year intervals the percentage of *JEAB* articles that reported an inferential statistic. Data for articles with both human and nonhuman subjects are included. The regression equation provided in the figure describes the dashed regression line fitted to the data by the method of least squares. That equation accounts for 82 % of the variance in the graphed data and has a slope of 1.08, indicating a progressive increase over time in the percentage of articles reporting an inferential statistic. During 2000, 2005, 2010, and 2013, respectively, 30.0, 90.0, 55.6, and 55.6 % of articles with human subjects reported an inferential statistic, whereas 35.7, 57.6, 69.6, and 44.8 % of articles with nonhuman subjects did so.

Sex of Subjects Table 2 shows for each decade and for all decades combined the percentage of articles with human and nonhuman subjects that did not provide information regarding the sex of subjects. Overall, 25.6 % of the articles with human subjects and 38.5 % of the articles with nonhuman subjects failed to describe the sex of subjects; no clear trends across decades are evident in this measure. Figure 5 shows the percentage of human subjects of reported sex that were males for each decade and for all decades combined. Overall, 50.8 % of the human subjects were male. Male human subjects substantially outnumbered females during the 1950s, but similar numbers of males and females were used in other decades.

Figure 6 shows the percentage of nonhuman subjects of reported sex that were males for each decade and for all decades combined. Data for all nonhuman species are combined in this figure. For all nonhuman species combined across all years, 77.9 % of *JEAB* subjects of reported sex were males. The number of male nonhuman subjects

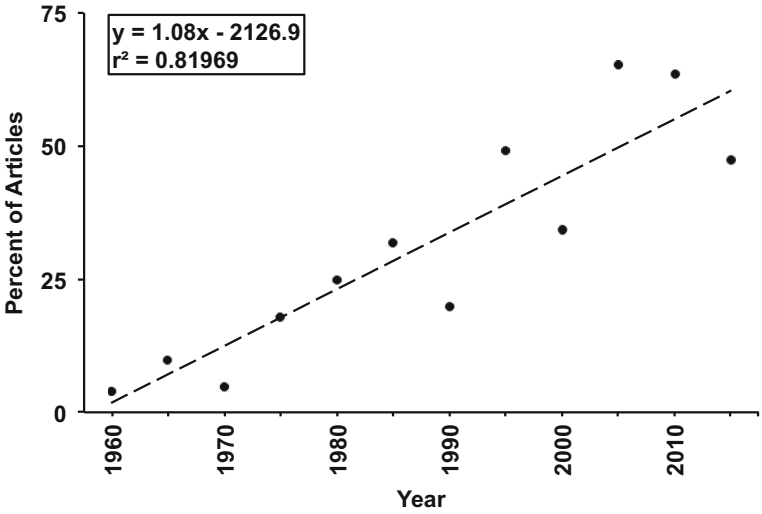

Fig. 4 Percentage of articles reporting inferential statistics

substantially exceeded the number of females during the 1950s and 1970s, but the numbers were similar in all other decades. Figure 7 shows for all species across all years the percentage of subjects of reported sex that were male. Data for humans and nonhumans are included in this figure. The percentage of male subjects substantially exceeded the percentage of female subjects for all nonhuman species except the miscellaneous category ("other species"). Percentages of male subjects were 87, 85.4, 70.2, and 42.5 for nonhuman primates, rats, pigeons, and other species, respectively. As noted previously, a similar number of male and female humans of reported sex were studied, with males accounting for 50.8 % of the total.

Table 3 shows for each decade and for all decades combined the percentage of articles for which the sex of subjects was reported that involved only male subjects, only female subjects, and both male and female subjects. Over all years, 70 % of studies with human subjects used both males and females, 21.3 used only males, and 8.7 % used only females. During the 1960s and 1970s, the category with the most studies was "males only," but that changed to "both males and females" for every decade thereafter. For studies with nonhuman subjects, during each decade, "males only" was the category with the most studies, and "both males and females" was the category with the fewest studies. Overall, 75.2 % of these studies involved only male

Table 2 Percentage of articles not reporting the sex of subjects

Subject type	Decade							
	1950s	1960s	1970s	1980s	1990s	2000s	2010s	Total
Human	50	30.2	31.2	17	22.4	34.9	28	27.3
Nonhuman	12.5	23.3	32.8	42.6	47.9	35.8	54	38.6
Total	31.3	26.7	32	29.8	35.2	35.4	41	32.9

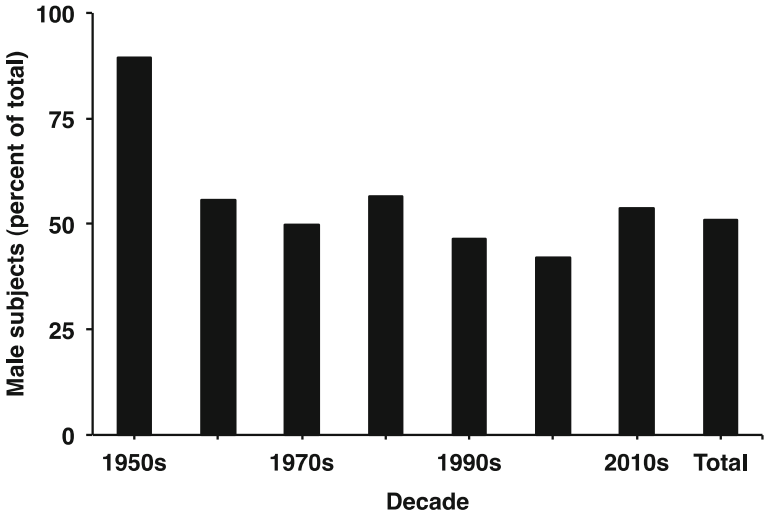

Fig. 5 Percentage of human subjects of reported sex that were males for each decade and for all decades combined

subjects, 15.8 % involved only female subjects, and 9% involved both male and female subjects.

Discussion

Several authors have described and commented on the species used as subjects in *JEAB* articles (e.g., Grossett et al., 1982; Millard 1976; Saville et al. 2002; Sokolowski et al. 2010). Unsurprisingly, the present findings concerning the percentage of articles that

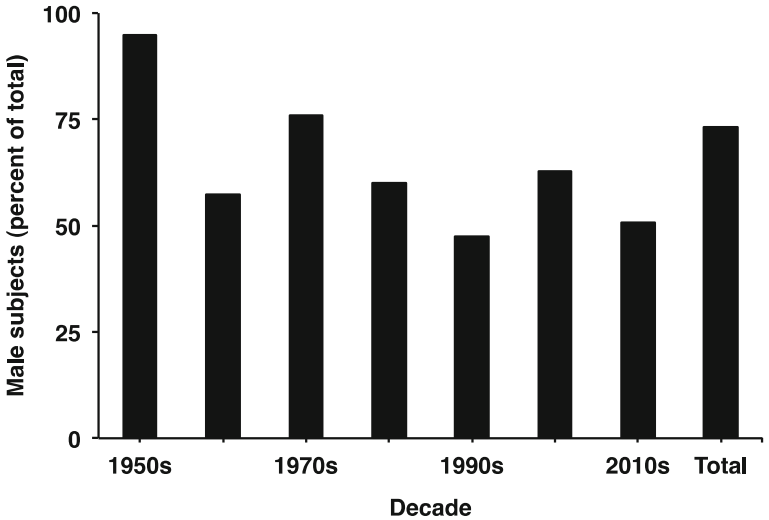

Fig. 6 Percentage of nonhuman subjects of reported sex that were males for each decade and for all decades combined

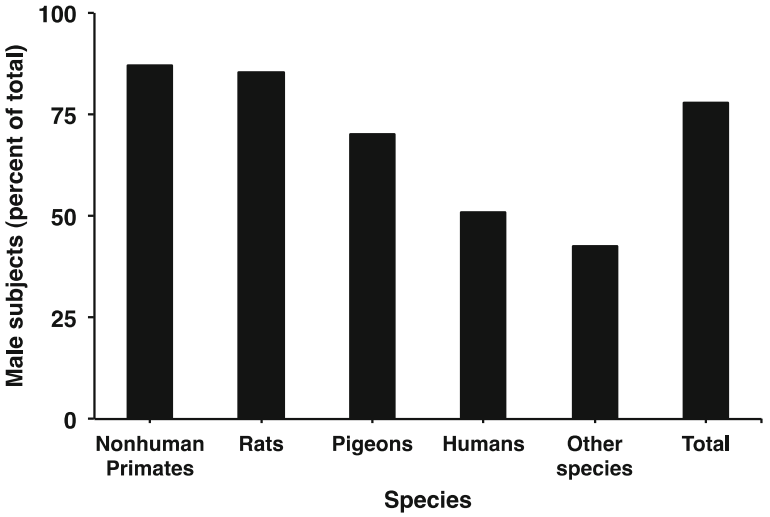

Fig. 7 Percentage of subjects of reported sex that were male for each species across all years

use particular species as subjects are comparable to those of earlier investigations. Collectively, they show that research with humans is growing ever more popular. As Saville et al. (2002) pointed out, many behavior analysts have long argued for the importance of studying human operant behavior (e.g., Hake 1982; Nevin 1982; Skinner 1953), and some predicted in the early 1990s that the number of studies with human participants published in *JEAB* would soon outnumber the number of studies with nonhuman subjects (Dougherty, Nedelmann, and Alfred 1993; Hyten and Reilly 1992). This prediction has not come true, but roughly a quarter of all studies published in the past two decades has involved human subjects. This is a substantial proportion and indicates that human operant research is alive and apparently well. As Critchfield (2011) pointed out, studying humans is essentially the only way to learn about some highly important topics, such as verbal behavior, and is an invaluable component of translational research. But as he also pointed out, practical and ethical considerations pose significant challenges for anyone who studies humans, and research methods may have to be adjusted in view of these challenges. These adjustments may be inconsistent

Table 3 Percentage of articles with only male, only female, and both male and female subjects

Human subjects	Decade							
	1950s	1960s	1970s	1980s	1990s	2000s	2010s	Total
Only males	75	45	31.4	20.5	13.3	4.2	16.7	21.3
Only females	0	20	7.8	9	7.8	5.6	0	8.7
Males/females	25	35	60.8	70.5	78.9	90.2	83.3	70
Nonhuman subjects								
Only males	85.7	74.2	82.9	74.1	80.3	71.4	60.9	75.2
Only females	14.3	18.2	14.6	18.5	11.5	14.3	21.7	15.8
Males/females	0	7.6	2.5	7.4	8.2	14.3	17.4	9

with the tactics and strategies that historically have characterized EAB (e.g., steady-state methodology and within-subject experimental designs).

Prior reports of the species of subjects used in EAB research (Saville, Epting, and Buskist 2002; Sokolowski et al. 2010) did not provide information about the number of individuals of particular species used as subjects. Our findings indicate that authors of *JEAB* articles have studied more than 35,000 subjects altogether, a number that is impressive on the one hand and modest on the other, given that those subjects were the basis of 3,084 articles, many describing multiple experiments. Using as few subjects as possible to answer a research question is an established ethical principle in psychological research, as evident in the American Psychological (2012) *Guidelines for the Care and Use of Animals in Research*, which states: "In proposing a research project, the psychologist should be familiar with the appropriate literature, consider the possibility of non-animal alternatives, and *use procedures that minimize the number of nonhuman animals in research* [italics added]" (p. 4).

On average, articles that used rats, pigeons, nonhuman primates, and other species as subjects studied 12.4, 7.8, 4.0, and 8.2 individuals, respectively. Such low values appear to be possible because more than 75 % of articles involving nonhuman subjects used within-subject designs, an advantage of which is the ability to answer meaningful experimental questions based on detailed study of a few individuals (e.g., Johnston and Pennypacker 2008; Poling et al. 1995). On average, 21.3 subjects were used in each *JEAB* article with human participants, which is substantially more than the mean number for any nonhuman species.

This difference reflects, in part, the use of between-subjects and mixed designs in a higher percentage of articles with human subjects (31.8 %) than in articles with nonhuman studies (24.6 %). Articles with human subjects that used between-subjects or mixed experimental designed involved, on average, 40.6 individuals, whereas the mean value when within-subject designs were used was 14.5. For studies with nonhumans, the respective mean values were 12.6 and 4.3. Of course, means can be misleading, and there was substantial variation across articles in the number of subjects used, regardless of species. For studies with humans, the range was 2 to 486 for studies with between-subjects or mixed designs and 2 to 200 for studies with within-subject designs. Respective values for studies with nonhumans were 2 to 426 and 1 to 88.

Because means alone provide limited information, it is common practice in many disciplines to use inferential statistics whenever means are compared. Foster et al. (1999) previously reported that the percentage of *JEAB* articles reporting an inferential statistic had generally increased across successive 5-year intervals from 1960 through 1995. The present data indicate that this trend has continued, and it is noteworthy that the slope of regression lines fitted to the data reported by Foster et al. and to the present data was almost identical, being 1.09 and 1.08, respectively. Given that Skinner (e.g., 1940) and several other behavior analysts (e.g., Branch 1999; Hopkins et al., 1999; Johnston and Pennypacker 2008; Perone 1999; Poling et al. 1995; Sidman 1960) have argued that inferential statistics are not especially useful in a science of behavior, it is also noteworthy that well over half of the *JEAB* articles published in 2005 and 2010 reported an inferential statistic. Clearly, things have changed with respect to the use of inferential statistics in *JEAB*.

Whether this is a change for the better is open to debate. As Michael (1974) sagely noted, inferential statistics, like figures, are judgment aids that are intended to make it

easier for people to discern the effects of independent variables on dependent variables. Inferential statistics can be useful for analyzing complex data sets and can easily be combined with visual data analysis. As Branch (2014) has noted, however, it is crucial to recognize that inferential statistics do not provide a measure of reliability or replicability, as they are sometimes mistaken to do. Although we did not attempt to quantify the relation, it appeared to us that almost all of the *JEAB* articles that reported an inferential statistic also included a graphic representation and visual analysis of the data. Nearly three decades ago, Huitema (1986) persuasively argued that including inferential statistics in behavior-analytic research articles would substantially increase the probability that those articles would be taken seriously outside the discipline. He wrote:

> Statistical tests have very high credibility to most people in most fields; to act otherwise is to commit political suicide…If an audience is convinced that a researcher has something worthwhile to say on the basis of statistical significance alone, consider how impressed it will be when both statistical significance and a very large visually apparent effect is presented. My suggestion then, is to employ statistical as well as visual methods when you are attempting to communicate with a nonbehavioral audience that is likely to be influenced by the presentation of statistical tests. Infiltration works. Besides that, statistics make people feel good. (p. 229)

Although it is not clear whether *JEAB* articles with inferential statistics were especially likely to delight or impress their readers, whether behavior analysts or not, it is clear that the quality of an article cannot be discerned based solely on the technique used to analyze data or the general design strategy used. Although EAB historically involved the use of within-subject designs to study a small number of subjects whose data was visually analyzed, several variables, including the experimental question and the proposed audience for the research, determine how behavioral research should be conducted (Johnston and Pennypacker 2008; Poling et al. 1995). Our findings regarding the kind and number of subjects used in *JEAB* articles, the research designs employed, and the use of inferential statistics say nothing about the current or future health of the journal or of the discipline that it represents.

The situation is arguably different with respect to the sex of subjects, a variable that has repeatedly been shown to influence behavior and its sensitivity to some, but by no means all, environmental variables (Dalla and Shors 2009). It is beyond our purpose to review the relevant literature, but two examples from behavioral pharmacology should make the point. In one, Campbell, Morgan, and Carroll (2002) showed that the suppressive effects of wheel-running on cocaine self-administration were much greater in female rats than in males. In the second, Zhang et al. (2007) demonstrated that early exposure to high doses of dextromethorphan significantly disrupted male rats' performance in a water maze but not the performance of females. Many other examples of sex differences in the effects of drugs, both quantitative and qualitative, are available (see, e.g., reviews by Franconi, Brunelleschi, Steardo, and Cuomo 2007; Schwartz 2007). The findings of these studies are sufficiently robust that the National Institute on Drug Abuse concluded in 2007 that "studies of drug abuse should routinely address gender issues" (Heading 2008, p. 428) and published a 276-page reporting dealing with the

topic (Volkow 2005–2008). Of course, sex does not always modulate drug effects, but it does so often enough that there is good reason to routinely report the sex of subjects in psychopharmacological research and to analyze data in a way that allows for the detection of sex differences, if any (Poling et al. 2009). The National Institutes of Health (NIH) have announced that new guidelines are in the works that will require inclusion of both male and female subjects in preclinical research, a step in the right direction towards taking into full account the role of sex on biology and behavior (Clayton and Collins 2014). Of course, the NIH guidelines will only directly affect NIH-funded research, although they may have a trickle-down effect on other projects.

To the extent that the sex of subjects modulates the effects of independent variables studied in *JEAB* research, which include but are not primarily drugs, the sex of subjects should also be reported therein. Interestingly, across all years, this was not done in 25.6 % of the articles with human subjects and in 38.6 % of the articles with nonhuman subjects that we examined. Despite the substantial attention paid to sex (or "gender") issues over the past 20 years, the percentages of *JEAB* articles not reporting subjects' sex in the 2000s and 2010s were similar to percentages for earlier decades. This, we feel, is unfortunate. Unless there is good reason not to report the sex of subjects, which it seems would rarely be the case with the kind of research reported in *JEAB*, then it appears that this information should be provided. As Porter et al. (2003) indicated, unless there is clear evidence that sex is irrelevant, if the sex of subjects is not indicated, then readers cannot determine whether reported results should generalize to males only, to females only, or to both. They also cannot ascertain if subjects' sex influenced results. Finally, they cannot determine whether males or females were excluded from participation, which can be an ethical concern with respect to human subjects and topics of clinical relevance (Hohmann and Parron 1996).

Determining whether sex modulates the effects of a given independent variable obviously requires that the effects of that variable be studied in both males and females and that data to be analyzed as a function of sex. Both males and females were studied in a substantial majority of *JEAB* studies involving human subjects, so those studies potentially could have determined whether the sex of subjects influenced results. We did not attempt to determine how many studies made such a determination, but it is heartening that most studies with humans included both sexes and that, overall, a similar number of males and females were studied. Prior studies have shown that in many research areas relevant to behavior analysis, male humans are overrepresented relative to their prevalence in the population of interest. This is the case, for example, in school psychology research (Holverstott, Ehrhardt, Parish, Ervin, and Poling 2002), in organizational behavior management research (Jarema, Snycerski, Bagge, Austin, and Poling 1999), in mental retardation (i.e., cognitive impairment) research (Porter et al. 2003), in autism research (Edwards et al. 2011; Watkins et al. 2014), and in psychopharmacological research (Poling et al. 2009).

When studies with nonhuman subjects are considered, 75.2 % of *JEAB* studies we examined involved only males, only 9.0 % included both males and females, and 77.9 % of all nonhuman subjects were male. Male subjects outnumbered females in studies with pigeons, rats, and nonhuman primates, but not with studies involving miscellaneous other species, which were few in number. Given that relatively few *JEAB* studies with nonhumans included both male and female subjects, it appears that sex was not generally assumed to significantly modulate the effects of the kinds of

independent variables examined in *JEAB* studies with nonhumans. Whether this assumption is warranted cannot be resolved empirically, because few relevant data are available, but it is clear that sex sometimes affects results. Unfortunately, determining whether it does so under the conditions of a given study substantially increases the difficulty of conducting an investigation. At minimum, the number of subjects studied must be increased. In addition, examining possible effects of subjects' sex requires the use of a between-subject or mixed experimental design, neither of which are viewed as traditional components of the Skinnerian approach to research (Poling et al. 1995). Moreover, it is best practice to house female and male laboratory animals in separate colony areas when specifically examining sex as an independent variable (Prendergast, Onishi, and Zucker 2014), which may be difficult or impossible for some researchers. There undoubtedly are good reasons why most *JEAB* studies have used only male nonhuman subjects and why those subjects have been rats or pigeons. We suggest, however, that future researchers consider whether it is appropriate for them to continue these practices and to let logic and evidence, not precedent, guide their decision.

In closing, it is important to emphasize that in determining the sex of subjects and the research designs used we examined every *JEAB* article published from 1958 through 2013 (N=524) that involved human subjects, but we only examined a randomly selected sample (N=524) of articles that involved nonhuman subjects. This tack was taken to make the project feasible but raises the possibility that our findings involving articles with nonhuman subjects are not representative of the population of such studies. Given that our sample size was relatively large and that articles were selected at random, however, it is unlikely that values for the sample and population differ greatly, although examining this possibility might be worthwhile.

References

American Psychological Association (2012). *Guidelines for ethical conduct in the care and use of nonhuman animals in research*. Retrieved January 20, 2015 from http://www.apa.org/science/leadership/care/care-animal-guidelines.pdf.

Branch, M. N. (1999). Statistical inference in behavior analysis: some things statistical testing does and does not do. *The Behavior Analyst, 22*(2), 87–92.

Branch, M. N. (2014). Malignant side effects of null-hypothesis significance testing. *Theory and Psychology, 24*(2), 256–277.

Campbell, U. C., Morgan, A. D., & Carroll, M. E. (2002). Sex differences in the effects of baclofen on the acquisition of intravenous cocaine self-administration in rats. *Drug & Alcohol Dependence, 66*(1), 61–69.

Clayton, J. A., & Collins, F. S. (2014). Policy: NIH to balance sex in cell and animal studies. *Nature, 509*(7500), 282–283.

Critchfield, T. S. (2011). Translational contributions of the experimental analysis of behavior. *The Behavior Analyst, 34*(1), 3–17.

Dalla, C., & Shors, T. J. (2009). Sex differences in learning processes and operant conditioning. *Physiology & Behavior, 97*(2), 229–238.

Dougherty, D. M., Nedelmann, M., & Alfred, M. (1993). An analysis and topical bibliography of the last ten years of human operant behavior: from minority to near majority (1982–1992). *The Psychological Record, 43*, 501–530.

Edwards, T. L., Watkins, E. E., Lotfizadeh, A. D., & Poling, A. (2011). Intervention research to benefit people with autism: how old are the participants? *Research in Autism Spectrum Disorders, 6*(3), 996–999.

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.

Foster, T. M., Jarema, K., & Poling, A. (1999). Inferential statistics: criticised by Sidman (1960), but popular in the *Journal of the Experimental Analysis of Behavior. Behaviour Change, 16*(3), 203–204.

Franconi, F., Brunelleschi, S., Steardo, L., & Cuomo, V. (2007). Gender differences in drug responses. *Pharmacological Research, 55*(2), 81–95.

Gentile, J. R., Roden, A. H., & Klein, R. D. (1972). An analysis-of-variance model for the intrasubject replication design. *Journal of Applied Behavior Analysis, 5*(2), 193–198.

Grossett, D., Roy, S., Sharenow, E., & Poling, A. (1982). Subjects used in *JEAB* research: is the snark a pigeon? *The Behavior Analyst, 5*(2), l89–l90.

Hake, D. F. (1982). The basic-applied continuum and the possible evolution of human operant social and verbal research. *The Behavior Analyst, 5*(1), 21–28.

Heading, C. E. (2008). Gender issues and the pharmacotherapy of substance abuse. *Drugs, 11*(6), 428–432.

Hohmann, A. A., & Parron, D. L. (1996). How the new NIH guidelines on inclusion of women and minorities apply: efficacy trials, effectiveness trials, and validity. *Journal of Consulting and Clinical Psychology, 64*(5), 851–855.

Holverstott, K. M., Ehrhardt, K. E., Parish, T., Ervin, R., & Poling, A. (2002). Females and males as participants in school psychology research: data from four journals over 15 years. *School Psychology International, 23*, 449–457.

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), 125–137.

Huitema, B. E. (1986). Statistical analysis and single-subject designs: some misunderstandings. In A. Poling & R. W. Fuqua (Eds.), *Research methods in applied behavior analysis* (pp. 209–232). New York: Plenum Press.

Hyten, C., & Reilly, M. P. (1992). The renaissance of the experimental analysis of human behavior. *The Behavior Analyst, 15*(2), 109–114.

Jarema, K., Snycerski, S., Bagge, S., Austin, J., & Poling, A. (1999). Participation of women as authors and participants in articles published in the *Journal of Organizational Behavior Management. Journal of Organizational Behavior Management, 19*(1), 85–94.

Johnston, J. M., & Pennypacker, H. S. (2008). *Strategies and tactics of behavioral research* (3rd ed.). New York: Routledge.

Jones, R. R., Vaught, R. S., & Weinrott, M. (1977). Time-series analysis in operant research. *Journal of Applied Behavior Analysis, 10*(1), 151–166.

Kangas, B. D., & Hackenberg, T. D. (2009). On reinforcing human behavior in the laboratory: a brief review and some recommendations. *Experimental Analysis of Human Behavior Bulletin, 27*, 21–26.

Lattal, K. A. (2008). *JEAB* at 50: coevolution of research and technology. *Journal of the Experimental Analysis of Behavior, 89*(1), 129–135.

Michael, J. (1974). Statistical inference for individual organism research: mixed blessing or curse? *Journal of Applied Behavior Analysis, 7*(4), 647–653.

Millard, W. J. (1976). Species preference of experimenters: *Journal of the Experimental Analysis of Behavior. American Psychologist, 31*(12), 894–896.

Nevin, J. A. (1982). Editorial. *Journal of the Experimental Analysis of Behavior, 37*(1), 1–2.

Perone, M. (1999). Statistical inference in behavior analysis: experimental control is better. *The Behavior Analyst, 22*(2), 109–116.

Poling, A., Methot, L. L., & LeSage, M. G. (1995). *Fundamentals of behavior analytic research*. New York: Plenum Press.

Poling, A., Durgin, A., Bradley, K. P., Porter, L. K., Van Wagner, K., Weeden, M., & Panos, J. J. (2009). Females' participation in psychopharmacology research as authors, editors, and subjects. *Experimental and Clinical Psychopharmacology, 17*(2), 105–112.

Poling, A., Picker, M., Grossett, D., Hall-Johnson, E., & Holbrook, M. (1981). The schism between experimental and applied behavior analysis: is it real and who cares? *The Behavior Analyst, 4*(2), 93–l02.

Porter, C. L., Christian, L., & Poling, A. (2003). Research brief: some data concerning the reporting of participants' gender in the mental retardation literature. *Mental Retardation, 41*(2), 75–77.

Prendergast, B. J., Onishi, K. G., & Zucker, I. (2014). Female mice liberated for inclusion in neuroscience and biomedical research. *Neuroscience and Biobehavioral Reviews, 40*, 1–5.

Saville, B. K., Epting, K., & Buskist, W. (2002). Selected publication trends in *JEAB*: implications for the vitality of the experimental analysis of behavior. *The Behavior Analyst, 25*(1), 45–55.

Schwartz, J. B. (2007). The current state of knowledge on age, sex, and their interactions on clinical pharmacology. *Clinical Pharmacology and Therapeutics, 82*(1), 87–96.

Sidman, M. (1960). *Tactics of scientific research: evaluating experimental data in psychology*. New York: Basic Books.

Skinner, B. F. (1934a). The extinction of chained reflexes. *Proceedings of the National Academy of Sciences of the United States of America, 20*(4), 234–237.

Skinner, B. F. (1934b). A discrimination without previous conditioning. *Proceedings of the National Academy of Sciences of the United States of America, 20*(9), 532–536.

Skinner, B. F. (1936). Thirst as an arbitrary drive. *Journal of General Psychology, 15*, 205–210.

Skinner, B. F. (1938). *The behavior of organisms: an experimental analysis*. New York: Appleton.

Skinner, B. F. (1940). A method of obtaining an arbitrary degree of hunger. *Journal of Comparative Psychology, 30*(1), 139–145.

Skinner, B. F. (1953). *Science and human behavior*. New York: Macmillan.

Sokolowski, M. B., Disma, G., & Abramson, C. I. (2010). A paradigm for operant conditioning in blow flies (*Phormia terrae novae* Robineau-Desvoidy, 1830). *Journal of the Experimental Analysis of Behavior, 93*(1), 81–89.

Volkow, N. (2005–2008). *Excerpts from the Director's report to the National Advisory Council on Drug Abuse: women, sex/gender differences and drug abuse*. Washington, DC: National Institute on Drug Abuse.

Watkins, E. E., Zimmermann, Z. J., & Poling, A. (2014). The gender of participants in published research involving people with autism spectrum disorders. *Research in Autism Spectrum Disorders, 8*(2), 143–146.

Zhang, T. Y., Cho, H. J., Lee, S., Lee, J.-H., Choi, S. J., Ryu, V., et al. (2007). Impairments in water maze learning of aged rats that received dextromethorphan repeatedly during the adolescent period. *Psychopharmacology, 191*(1), 171–179.