

CRITICAL REVIEW

Reporting of Demographic Variables in Neuropsychological Research: An Update of O’Bryant et al.’s Trends in the Current Literature

Luis D. Medina* , Stephanie Torres, Anthony Gioia, Andrea Ochoa Lopez, James Wang and Paul T. Cirino
Department of Psychology, University of Houston, Houston, TX, USA

(RECEIVED March 6, 2020; FINAL REVISION July 28, 2020; ACCEPTED September 15, 2020; FIRST PUBLISHED ONLINE November 12, 2020)

Abstract

Objective: Demographic trends and the globalization of neuropsychology have led to a push toward inclusivity and diversity in neuropsychological research in order to maintain relevance in the healthcare marketplace. However, in a review of neuropsychological journals, O’Bryant et al. found systematic under-reporting of sample characteristics vital for understanding the generalizability of research findings. We sought to update and expand the findings reported by O’Bryant et al. **Method:** We evaluated 1648 journal articles published between 2016 and 2019 from 7 neuropsychological journals. Of these, 1277 were original research or secondary analyses and were examined further. Articles were coded for reporting of age, sex/gender, years of education, ethnicity/race, socioeconomic status (SES), language, and acculturation. Additionally, we recorded information related to sample size, country, and whether the article focused on a pediatric or adult sample. **Results:** Key variables such as age and sex/gender (both over 95%) as well as education (71%) were frequently reported. Language (20%) and race/ethnicity (36%) were modestly reported, and SES (13%), and acculturation (<1%) were more rarely reported. SES was more commonly reported in pediatric than adult samples, and the opposite was true for education. There were differences between the present results and those of O’Bryant et al., though the same general trends remained. **Conclusions:** Reporting of demographic data in neuropsychological research appears to be slowly changing toward greater comprehensiveness, though clearly more work is needed. Greater systematic reporting of such data is likely to be beneficial for the generalizability and contextualization of neurocognitive function.

Keywords: Sample size, Ethnic groups, Language, Acculturation, Social class

INTRODUCTION

Demographic trends in the United States paint a vivid picture of a changing cultural landscape (O’Bryant et al., 2004). Specifically, there is a downward trend in the percentage of people living below the poverty line as well as in those without a high school education (US Census Bureau, 2019). In regards to ethnoracial diversity, Black/African American individuals continue to comprise close to 13% of the population while the percentages of individuals who identify as Hispanic/Latino, Asian, or of more than one race continue to climb; together, these four groups account for nearly 40% of the US population. Census projections estimate that individuals who identify as White/Caucasian alone will be in

the minority (<50%) by the year 2045. Moreover, the United States is becoming globally representative with notable growth in the number of foreign-born individuals immigrating into the country and recent estimates suggest about 22% of the nation’s population speaks a language other than English at home (Figure 1).

In 2017, reflecting the shifting demographics and the notable advances in research and theory related to the study of multiculturalism, the American Psychological Association adopted its updated multicultural guidelines (American Psychological Association, 2017). In the document, the writers highlight an empirically supported need for consideration of multiple factors that help define an individual’s identity; they also emphasize the intersectionality of identities that enhance our understanding of both between-group and within-group differences. However, while the changing cultural landscape in the United States is evident, our neuropsychological assessment strategies have greatly lagged in

*Correspondence and reprint requests to: Luis D. Medina, PhD, Department of Psychology, University of Houston, 3695 Cullen Blvd., Room 126 Heyne, Houston, TX 77204-5022, USA. Tel.: +1 713 743 9318. Email: LDMedina2@uh.edu

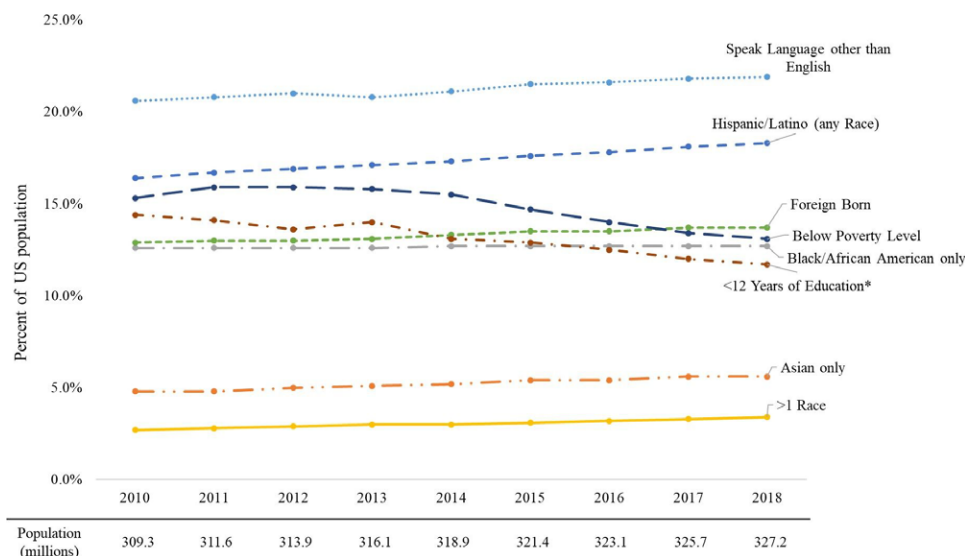


Fig. 1. US demographic trends 2010–2018 (US Census Bureau).

NOTE: *Education estimates are for individual's age 25 and higher. All data are from the American Community Survey for the US Census and reported to have a 0.1% margin of error.

their applicability to under-represented communities. For example, in its Relevance 2050 Initiative, the American Academy of Clinical Neuropsychology provides a grim prediction that our current neuropsychological toolkit will have limited validity for use in the majority of the US population – a full 60% – by the year 2050, rendering the field virtually irrelevant in the healthcare marketplace. The consequences of a lack of cultural contextualization are also likely to be felt internationally, where research and the professionalization of neuropsychology vary considerably. Nearly 15 years ago, a review of the contemporary neuropsychological literature in the field's top journals revealed significant limitations in the interpretability and generalizability of research findings; while some demographic variables, like age, were nearly ubiquitously reported in the literature, other important descriptors, like race/ethnicity, of the samples studied were systematically omitted by manuscript authors (O'Bryant et al., 2004). In addition to limiting the generalizability of study results, these omissions threaten the replicability of the science reported.

The need to understand neuropsychological test performance in the context of age, education, and sex is well-documented (Grant & Adams, 2009; Heaton et al., 1986; Lezak et al., 2004; Seidenberg et al., 1984). Indeed, numerous normative references are now available for neuropsychologists, allowing for the comparison of scores on cognitive measures across multiple domains. These typically include either means and standard deviations or comprehensive tables that allow for the calculation of standardized scores that may control for any combination of age, education, and sex. Relationships between cognitive performance and these demographic variables vary in size from small to very large and depend on the cognitive domain being tested (Table 1). Various reviews and meta-analyses have reported effect sizes in the range of $r = 1.05-.811$ for age and cognition (Salthouse,

2011) and $r = 1.01-.741$ for education. Effect sizes for the relationship between biological sex and cognition have been reported to be in the range of $r = 1.02-.271$ (Zell et al., 2015). Notably, less is known about gender identity and its relationship with cognitive performance.

However, as noted by O'Bryant and colleagues (2004), other demographic variables that should be accounted for in the interpretation of neuropsychological test scores have been identified and include such factors as race/ethnicity, acculturation, language, and socioeconomic status (SES). A large body of literature has demonstrated both direct and indirect (e.g., by pathways related to education quality) statistical effects related to race/ethnicity and cognition. In a study of cognitive performance in older adults, race was significantly associated with test performance – even after controlling for reading level achievement – and ranged in effect sizes from $r = 1.10-.311$ (Morgan et al., 2008). Moreover, evidence of diminishing returns exists suggesting factors like education may not portend the same cognitive outcomes across ethnically and racially diverse groups (Díaz-Venegas et al., 2016), further highlighting the critical need to better characterize the demographic makeup of samples. In a systematic review of the literature examining the relationship between migrant status and cognitive abilities, greater acculturation (i.e., participation in the practices and activities from the dominant culture) was associated with better cognitive performance (Xu et al., 2017). A recent meta-analysis of these relationships suggests effect sizes in the range of $r = 1.25-.661$ (Tan et al., 2020). Not surprisingly, greater proficiency in the language of test administration is associated with better cognitive test performance as well (Black et al., 1999; Yano et al., 2000). Research on post-stroke aphasia and the relationship between basic language abilities and non-language cognitive measures suggests these measures are strongly associated with simple auditory comprehension

Table 1. Ranges of documented effect sizes for the seven demographic variables

Variable	Effect range	Citation
Age	$r = 1.05-.811$	Salthouse (2011)
Education	$r = 1.01-.741$	Opdebeek et al. (2016)
Gender	$r = 1.02-.271$	Zell et al. (2015)
Race/ethnicity	$r = 1.10-.311$	Aiken Morgan et al. (2008)
Acculturation	$r = 1.25-.661$	Tan et al. (2020)
Language	$r = 1.37-.881$	Wall et al. (2017)
Socioeconomic status	$r = 1.02-.471$	Lawson et al. (2018)

Note: Benchmarks for effect sizes provided by Cohen (1988) and Rosenthal (1996) suggest the following thresholds: small, $r = .10$; medium, $r = .30$; large, $r = .50$; very large, $r = .70$.

($r = 1.37-.821$) and naming abilities ($r = 1.42-.881$) (Wall et al., 2017). However, there is also compelling evidence that speaking more than one language is associated with several cognitive benefits, such as in attention and working memory (Adesope et al., 2010). While the potential benefits of bilingualism may be moderated by various factors, such as immigration and the context of acquiring the multiple languages, these benefits may also be greater with the increasing number of languages spoken (Quinteros Baumgart & Billick, 2018). In relation to SES, indicators ranging from the individual level to the community and societal levels have also been identified as moderators of cognitive performance across the lifespan suggesting higher SES is related to better cognitive test performance (Wong & Edwards, 2013; Wu et al., 2015). In a meta-analysis on SES and executive functioning in a pediatric population, SES was associated with cognitive outcomes with effect sizes in the range of $r = 1.02-.471$ (Lawson et al., 2018). Moreover, how these various demographic variables intersect and relate to structural and functional brain variables as well as behavioral and cognitive outcomes can be used to inform research and clinical practice (Dotson & Duarte, 2020).

Any observed test score can be attributed to actual ability, systematic error variance, and random measurement error; effects related to demographic variables may contribute to systematic error. However, a lack of consistent reporting limits our understanding of these effects and has implications on downstream interpretation and treatment planning. Therefore, when these variables are not systematically collected, reported, and included in analyses, the results may be confounded and potentially miss an opportunity to identify significant differences in certain demographic groups. Once systematic reporting of these demographic variables occurs in research samples, downstream clinical decisions such as test interpretation, differential diagnosis, prognosis, and treatment planning can be considered in the context of these demographic modifiers.

Numerous guidelines for reporting research have been recommended for various types of studies. Indeed, over 400 reporting guidelines are currently registered by the

Enhancing the QUALity and Transparency Of health Research (EQUATOR) Network (<http://www.equator-network.org>) and cover diverse scientific approaches such as clinical trials, observational study, and meta-analyses. Such guidelines, like Strengthening the Reporting of Observational Studies in Epidemiology (STROBE; <http://www.strobe-statement.org/>) and Consolidated Standards of Reporting Trials (CONSORT; <http://www.consort-statement.org/>) typically provide structured checklists for authors, editors, and reviewers to follow in order to gauge the inclusion of critical elements. However, while guidance is given as to the description of research participants, this is often vague or limited in scope (e.g., eligibility criteria, source of recruitment, settings where data are collected). Moreover, these guidelines are not systematically followed by authors, requested by journals, or enforced by editors and reviewers. Indeed, a critical review of the most commonly used reporting guidelines concluded that the inconsistent implementation or enforcement of these guidelines may explain why these guidelines may be related to only a moderate improvement in the quality of research reporting (Johansen & Thomsen, 2016).

The Present Study

Although more work is necessary in studying a wide range of individuals from diverse backgrounds, progress in the field has been made to resolve the discrepancy of reporting in these understudied populations. This has included making recommendations for neuropsychological score interpretation as well as the development of normative data for ethnically and linguistically diverse groups (e.g., Mungas et al., 2004; O'Driscoll & Shaikh, 2017; Rossetti et al., 2019). Given the growing awareness that sociodemographic and cultural factors are important for the interpretation of neuropsychological test performance, and emerging demographic trends toward greater diversity, a broad goal is to update and expand the review of O'Bryant et al. (2004) by evaluating the extent to which a suite of sociodemographic variables is reported in the neuropsychological literature. We include the same journals and variables as O'Bryant et al. (2004), but extend their findings by also including more pediatric-focused journals, examining a more recent time span, and coding for additional characteristics (sample size, the proportion of non-Hispanic White in the sample, whether the study included pediatric or adult populations, and country). Within this structure, we propose several hypotheses. (1) Given their nearly ubiquitous reporting in O'Bryant et al., 2004, we expect to see a high proportion of studies that report age and sex. Similarly, given that acculturation was nearly universally not reported, we expect similar proportions of studies that report acculturation. (2) Relative to O'Bryant et al. (2004), we expect to see increases in the proportion of studies that report race/ethnicity, given low rates in that study, coupled with the more recent sampling and pushes in the field for inclusiveness and diversity (e.g., Falk et al., 2013; Rad, Martingano, & Ginges,

2018). We also expect an increase in reporting of education. (3) We expect that non-Hispanic White participants will be over-represented, but we do not have specific hypotheses regarding expectations for sample size, country, or whether the studies included pediatric and/or adult populations. (4) In contrasting pediatric versus adult-focused studies, we expect pediatric studies to be more likely than adult studies to code for factors such as SES and language, given that these variables are tied to many pediatric-focused outcomes.

METHODS

Descriptive Information

Journals/timeframe

We selected common neuropsychological journals. First, we included the five journals that were coded in the original review (*Archives of Clinical Neuropsychology*, *Journal of Clinical and Experimental Neuropsychology*, *Journal of the International Neuropsychological Society*, *Neuropsychology*, and *The Clinical Neuropsychologist*). Second, we added journals with a developmental focus (*Child Neuropsychology* and *Developmental Neuropsychology*) in order to track trends across both pediatric and adult literature since adult-focused neuropsychology journals were sampled well by O'Bryant et al. (2004). In an effort to remain as current as possible while capturing a similar number of articles, we coded journals from April 1, 2016 to March 31, 2019. Although this duration is shorter than that of the original report, we covered more journals, and ended up with a sample of articles similar to the original report; O'Bryant et al. (2004) coded 1440 articles from 5 journals, and the present study evaluated 1648 articles from 7 journals, of which 1277 were coded and reported below (Figure 2).

Types of articles considered

Broadly, we included manuscripts that reported demographic data for a sample. The only types of manuscripts that were excluded categorically were Position Papers ($k = 41$), Provider Surveys ($k = 22$), Review Papers ($k = 174$), Meta-Analyses ($k = 29$), and Case Studies ($k = 27$). In addition, 78 additional published works were excluded that did not include data/participants (e.g., editorials), or secondary analyses (see further description below) that did not report any demographic data (e.g., referring only to a prior work; methods papers using a prior sample only as a demonstration; simulation studies). Our approach to inclusion/exclusion was consistent with the methods described by O'Bryant et al. (2004). Given the above, this meant that 1277 articles were fully coded for analyses.

Coded articles were designated as either Original Research or Secondary Analysis. Original Research was designated when it was apparent or reasonably inferred that the article was the first and/or only report. For individual or stand-alone studies, this was always the case. For papers known to be or reasonably inferred to be derived from "larger projects," these

were typically coded as secondary analyses. This was the case if there was (a) some direct reference to prior reports (e.g., "further details on this sample can be found in Medina et al.") and/or there was some specific indication that the data had been previously used (e.g., "the data here are from the second time point of the ____ project"; "data were taken from the ____ project"). If data came from records (e.g., "records were reviewed"; "records were pulled from X registry"), these were typically coded as Original Research, unless the above conditions were met. As above, the only papers specifically excluded were those that did not present demographic data at all (e.g., either because they were review/meta-analyses or position papers). The research team coded specific demographic variables based on the criteria below. Additionally, the study team communicated through a monthly call to rectify any specific concerns that team members came across during the coding process.

Sample N

We recorded the N of the sample for which demographic information was reported. This definition resolved the issue where demographics were sometimes provided only on the analyzed sample or sometimes on the larger recruited sample. In the case of multiple groups, we reported the sum total of those groups. If there were multiple experiments, using different samples, we reported the sum total across experiments. Thus if 200 were recruited, but 15 were excluded, and 185 analyzed, then the reported sample could be either 200, 185, or some other number; the "Sample N " was determined by the number of individuals for whom the study provided demographic data.

Developmental population

We coded whether studies included pediatric (age 0–17 years) and/or adult (age 18+ years) samples. Some journals publish works predominantly on pediatric *versus* adult populations, but this often does not adequately represent the reported sample. In fact, from the two developmental journals, 23% of coded articles included adult samples (exclusively or in part), and of the five other journals, 20% of articles included pediatric samples (exclusively or in part). In cases where longitudinal data that spanned the two populations were reported, or where the age range included both pediatric and adult populations (e.g., aged 16–24 years), we coded articles as including both of these population categories. For pure follow-up samples (e.g., a report that follows up on a sample reported 5 years earlier), we coded the age for which data were currently being reported (e.g., if the initial sample was recruited at age 15, but the data for the current paper were collected at age 20, the study would be coded as adult only).

Country

The country was coded by name, according to where the sample was recruited (as opposed to, e.g., the country of the lead/

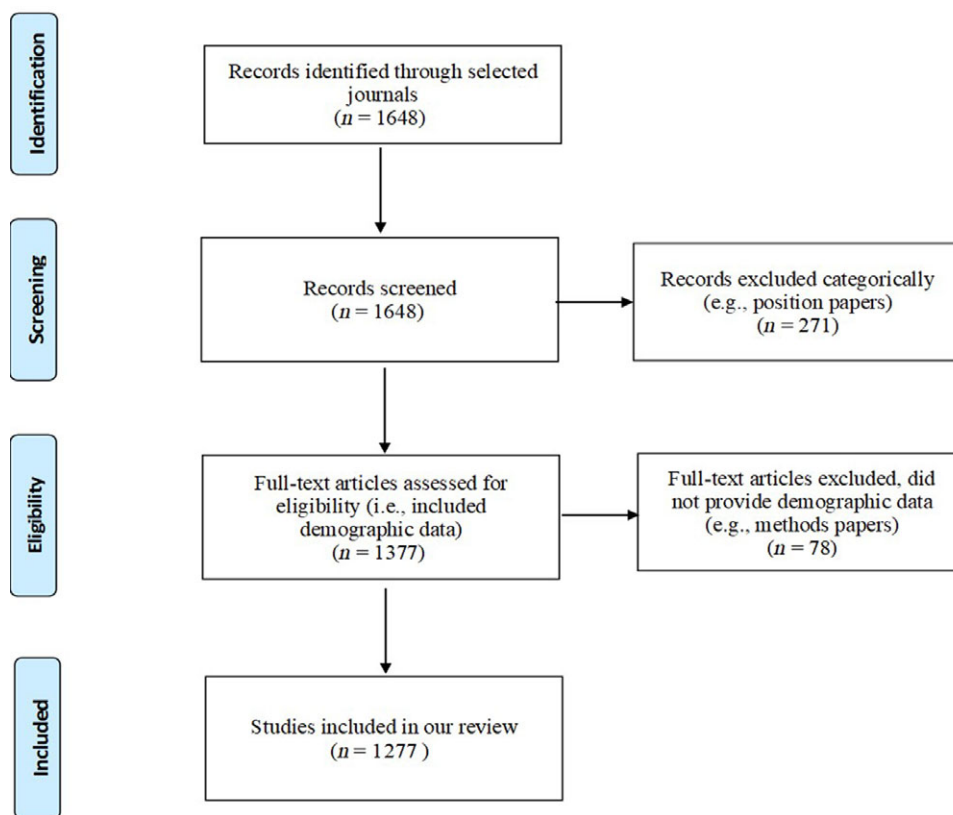


Fig. 2. PRISMA flow diagram.

senior author or where the IRB approval was from). In cases where participants were from multiple countries, we chose to code these as “International.” We also recoded country according to the continent in a *post hoc* manner, for parsimony in reporting.

Coding of Actual Demographic Variables

Age

We coded as 1 when studies reported actual ages (means, SD, etc.). We coded as 0 if only a broad range (“adolescents”; “toddlers”; ages 12–18 years) was provided. We coded age somewhat strictly because of its ubiquity in O’Bryant et al. (2004).

Sex/gender

We coded as 1 when studies reported actual sex/gender or proportions. We coded as 0 if only vague descriptions were given (e.g., “mostly female”). While we acknowledge biological sex differs from gender identity or non-binary gender, we use the term “sex/gender” here to encompass the reporting of both. Notably, despite the distinction between these terms, there remains a prevalent use of both terms in the literature without explicit description justifying the choice in the term. So as to not assume one or the other in the articles reviewed (e.g., the authors reported “gender” but meant “sex”), we refer to sex or gender as reported.

Race/ethnicity

We coded as 1 when studies documented racial/ethnic composition in the majority of cases. For example, if a study reported dichotomies (e.g., % Caucasian), this was still coded as 1. We coded as 0 if race or ethnicity or their proportions were ambiguous (“% were Eastern European,” or “majority were Caucasian”).

%NHW. The percentage of non-Hispanic White, if reported or calculable, was recorded as %NHW. Beyond recording the percentage of White individuals in a sample, this variable distinguished studies that reported the race of each individual within the sample from studies that reported race more generally. For example, in a study that reported a dichotomy such as the number of Hispanic *versus* non-Hispanic individuals, race/ethnicity would still be coded as 1, but %NHW would not be calculable and therefore be left blank.

Education quantity

Education was coded more broadly than some other categories, partly because education means different things for children (tied to how old they are) versus adults (where education typically indicates maximal academic attainment). Education was coded as 1 if the number of years of education or grade level was actually provided (e.g., mean, SD). If a range was given (e.g., grades 4–9, or “all participants had a HS

diploma”), education was also coded 1. However, education was coded 0 even if related data were provided (e.g., % in Special Education), because it was the quantity/level of education that was the variable of interest. For pediatric samples, we coded as 1 when *either* a parental or child education level was given. For example, in infant studies, education *per se* is not relevant, but parental education is often reported. For toddlers, they may or may not be in preschool or parental education may be reported. Similarly, for most age ranges, education could be inferred (“participants were in middle school” or “young adults in college”) and was also coded 1 in such instances.

Socioeconomic status

SES was coded as 1 only if (a) it was specifically identified as such or included the major SES elements (e.g., income, occupation), and (b) went beyond education. For example, if authors documented that they used maternal education level as a proxy for SES without any additional corroborating information, then SES was coded as 0 and education was coded as 1. SES, like education, was coded broadly; if general categories were provided (“sample was predominantly upper-middle class”), then SES was coded 1. Also, as with education, in pediatric samples, if parental SES was given, then SES was coded as 1. Additionally, we coded as 0 if SES was used in analyses but means, SDs, and/or frequencies/percentages were not reported. If occupational categories were provided and presented a clear hierarchy (manual labor *vs.* professional), then we also coded SES as 1.

Language

The language was coded as 1 if the manuscript reported the participants’ language (e.g., some measure of language dominance, the proportion that is bilingual, or the proportion with a second language). The language was also coded as 1 if inclusion/exclusion criteria were clear (e.g., “must be native Swedish speakers,” “must have English as dominant language”). The language was coded as 0 if language ability was ambiguous (“participants were fluent in English”). The language was also coded as 0 if the samples’ language was not explicitly stated, even if the study’s context may suggest a specific language (e.g., a German study might be assumed that its sample spoke German; this, however, was not sufficient reporting according to our definition).

Acculturation

Acculturation was coded as 1 if some specific measure of acculturation was used and data provided, and/or there was some strong indication that all the participants were clearly and uniformly acculturated to the setting (e.g., all participants being of the same ethnicity and language background and having lived in the same small rural town). Otherwise, acculturation was coded 0.

If authors mentioned that groups were “matched” on one or more of the above demographic factors, but did not report evidencing data, we coded as 0. We also coded as 0 in cases where, in secondary analyses, authors referred to a previous report or larger study for any of these demographic variables (though as noted, we did not include secondary analyses if no demographic data were reported). In these cases, although data may be available, specific demographic breakdowns or values were not made immediately evident to the reader, and therefore constituted a lack of reporting.

Analyses

Preliminary analyses included an evaluation of inter-rater reliability coding. While it may seem self-evident whether or not a given datum was reported, because journals and authors vary considerably, there were not always clear-cut statements on which to evaluate whether or not a given piece of information was recorded. In fact, all team members participated in finalizing the definitions relayed above to promote consistency. Once all definitions had been agreed and finalized, we double-coded 10 manuscripts from each journal (70 altogether, or 5.5% of total) for agreement purposes. We computed both percent agreements and Kappa coefficients for each of the 13 variables considered (see Table 2 for values and confidence intervals). These were all acceptable, and indicate that agreement, particularly on key variables, was excellent.

Primary analyses were largely descriptive. However, for comparisons of equivalence or increases/decreases relative to benchmarks, we used a single sample proportion *z*-test from a binomial distribution (Pan, 2002) to statistically compare these frequencies. A specification of a margin is required for equivalence tests; for these, we used 5%, or else the maximum allowed utilizing a 5% would cause the margins to not be bound by 0 or 1. Two-sided *p*-values were used, at .05. Further, in comparing pediatric versus adult, we used chi-square analyses to compare proportions among those reporting either or both of these populations.

RESULTS

The primary descriptive results appear in Table 3. As with O’Bryant et al. (2004), the most frequently reported categories were age, sex/gender, and education (see Table 4 for direct comparisons). We first expected that age, sex/gender, and acculturation would be similar to previously reported rates. We found that for age (reported nearly ubiquitously), the obtained value was greater than the lower margin and less than the upper margin ($z_{\text{lower}} = 33.62$, $p < .001$; $z_{\text{upper}} = -1.66$, $p < .048$), indicating that our obtained value was equivalent to that of O’Bryant et al. (2004). However, for sex/gender, the present reporting was not equivalent ($z_{\text{lower}} = 23.74$, $p < .001$; $z_{\text{upper}} = +5.49$, $p < .001$), indicating that our obtained value was higher than that found in O’Bryant et al. (2004), $z = 8.80$, $p < .001$. Similarly, for

Table 2. Inter-rater reliability

	Demographic variable	% Agreement	Kappa coefficients	95% confidence interval
Primary demographics: seven variables included in O'Bryant et al. (2004)	Education	92.9	0.83	0.68–0.97
	Race/ethnicity	91.4	0.83	0.70–0.96
	Age	100	N/A	N/A
	Sex/gender	95.7	N/A	N/A
	Acculturation	100	N/A	N/A
	Language	95.7	0.84	0.67–1.02
	SES	98.6	0.90	0.71–1.09
	Primary demographics	96.3	0.93	0.89–0.96
Secondary demographics: three additional demographic variables	Article type	80	0.52	0.30–0.75
	Adult	98.6	0.97	0.91–1.03
	Child	100	1.0	1.0–1.0
	Secondary demographics	92.9	0.86	0.79–0.93
Primary and secondary demographics	Total 10 demographics	95.3	0.91	0.87–0.94
Other publication characteristics	Country	82.9	0.65	0.52–0.78
	Sample <i>N</i>	88.6	0.93*	0.89–0.96
	% NHW	95.7	0.95*	0.89–0.97

Note. Kappa coefficients and CIs could not be computed for age and acculturation due to % agreement being 100%. Kappa coefficients and CIs could not be computed for sex/gender due to one of the required cells being 0 (other raters never coded “not reported” for sex/gender). CIs = confidence intervals.

*Sample *N* and % NHW do not have computed Kappa coefficients because these are not categorical variables; values provided are intraclass correlation coefficients.

acculturation, reporting was not equivalent ($z_{\text{lower}} = 2.40$, $p < .009$; $z_{\text{upper}} = 1.46$, $p = .928$), indicating that our obtained value was higher than that found in O'Bryant et al. (2004), $z = 4.18$, $p < .001$.

We expected that both race/ethnicity and education would be more frequently reported. This was the case for race/ethnicity, $z = 12.63$, $p < .001$, but not for education, $z = -5.46$, $p < .001$, which was in fact *less* frequently reported compared to O'Bryant et al. (2004).

For SES and language, these were more frequently reported, $z = 5.48$, $p < .001$ and $z = 8.25$, $p < .001$, respectively, but we expected these to be more reported in pediatric than adult samples. For this comparison, statistics were evaluated excluding the studies that contained both sample groups ($k = 1162$). SES was in fact reported more commonly in pediatric (30.7%) than in adult (5.7%) samples, $\chi^2(1) = 131.2$, $p < .001$ (the value was 18.3% for studies that included both pediatric and adult samples). However, the language was reported similarly across these two populations, $p > .05$, though as noted the overall value was higher than in O'Bryant et al. (2004).

Exploratory evaluation of rates of reporting on other variables did not reveal differences between adult and pediatric samples, on age, sex/gender, acculturation, or race/ethnicity, but education was in fact much more commonly reported in adult (86.8%) than pediatric (39.1%) samples, $\chi^2(1) = 273.4$, $p < .001$.

Considering variables not included in O'Bryant et al., it is interesting to note that the proportion of non-Hispanic White participants across studies was 68% (median, 75%). It should also be noted that this value is only calculable on studies that reported this proportion, which was only approximately one-third of the sample ($n = 414$).

Another interesting value examined was study origin, where the overwhelming proportion of studies included participants from either North America (52%) or Europe (27%).

As a follow-up analysis, we examined the extent to which the rates of reporting were different in original versus secondary analyses (since there is likely variability in the extent to which researchers recapitulate descriptive data that exist elsewhere). However, for most all variables, there was no difference in such rates, $p > .05$, with the exception of race/ethnicity, though in this case, this variable was *more* frequently reported in secondary (50%) relative to original (31%) reports, $\chi^2(1) = 35.5$, $p < .001$.

DISCUSSION

We sought to review articles from neuropsychological journals published over a 3-year period in order to compare the reporting of demographic variables (e.g., age, education, sex/gender, race/ethnicity, and language) to the previously reported rates by O'Bryant et al. (2004). The data were encouraging in some respects, demonstrating increases in reporting of language, race/ethnicity, and even sex/gender (and SES, at least for pediatric populations). However, as intimated 15 years ago by O'Bryant et al. (2004), there remains much work to do. For example, although reporting rates of race/ethnicity and SES and language have increased, they are still quite low. Although the fact that SES as reported in adult studies was lower than even in O'Bryant et al. (2004), this is likely a function of the lenient criteria and the explicit separation of SES from education in the current study. Lower rates of education reporting in pediatric samples appear counterintuitive, although this may in part be due to the fact that young children may not have years of education (though as noted, education was coded even if parent information was provided). It is important to note that the adult journals did not differ in percent reported for education from

Table 3. Descriptive results

Journal	k	Age	EDU	Sex/gender	R/E	ACC	LANG	SES	Adult	Pediatric	Original	% NHW	Sample N
ACN	195	98.7	83.1	93.9	44.1	0.0	15.4	3.6	91.3	16.4	97.4	67.9	137.0
CNP	167	100	38.9	98.2	35.9	0.6	15.0	32.9	17.4	98.8	68.9	62.9	93.0
DVN	95	100	39.0	84.2	28.4	0.0	21.1	22.1	32.6	83.2	74.7	56.2	69.0
JCEN	208	100	77.9	98.1	31.7	0.0	20.7	9.1	89.4	20.7	72.6	70.4	89.5
JINS	240	100	76.7	96.7	37.1	0.0	17.1	12.5	85.8	21.3	50.0	71.1	90.0
NEU	233	100	74.7	97.0	23.2	0.9	22.3	12.9	82.4	22.8	97.0	69.2	69.0
TCN	139	99.3	91.4	98.6	56.1	2.2	27.3	6.5	88.5	17.3	59.7	72.1	169.0
Total	1277	99.8	71.3*	96.0*	36.0*	0.5*	19.5*	13.4*	74.0	35.0	74.9	68.2	99.0

Note: *k* is the number of reviewed articles and included in the comparisons for each journal. The value for Sample *N* is a median value (as original variable was quite skewed and kurtotic) of the samples described in each article. All other values are percentages.

*Significant at $p < .01$ compared to 2004 data; all significant values indicate increases in reporting except for EDU (i.e., education), which demonstrated a decline in reporting. As noted, adult and pediatric do not add up to 100% given that some studies include both populations. ACN = *Archives of Clinical Neuropsychology*; CNP = *Child Neuropsychology*; DVN = *Developmental Neuropsychology*; JCEN = *Journal of Clinical and Experimental Neuropsychology*; JINS = *Journal of the International Neuropsychological Society*; NEU = *Neuropsychology*; TCN = *The Clinical Neuropsychologist*; EDU = education; R/E = race/ethnicity; ACC = acculturation; LANG = language; SES = socioeconomic status; NHW = non-Hispanic White. Original is the proportion of studies with original data, its converse is secondary data reporting.

the original reporting by O'Bryant et al. (2004). In sum, an important takeaway is that not enough information is routinely provided in neuropsychological research regarding race/ethnicity, language status, acculturation, or SES.

It should be noted that the present study focused on the presence of these characteristics being reported. This is a fairly "low bar," and of course is far different than considering these variables analytically (either as covariates, as matching variables, or as an area of scientific focus). Considering such variables routinely, particularly beyond age and sex/gender, may allow for broader applicability of the principles and findings reported. While we mostly limited our review to the demographic variables reviewed by O'Bryant et al. (2004), other variables [e.g., rural vs. urban living environment (Saenz et al., 2018)] may provide an additional valuable sample characterization. Exploration of reporting patterns by subfields of research and practice (e.g., epilepsy vs. neurodegenerative disease) may also provide insights into specific and systematic oversights. In the end, we believe that more routine reporting of such data (both rote reporting as well as consideration in analyses) will serve to make neuropsychological research more inclusive and generalizable, as well as positively affect test interpretation and downstream diagnosis, recommendations, and treatment planning. Particularly as many of these factors are easily obtained *via* self-report, there is little reason not to include them. Acculturation seems the closest to a construct here, and might require specific assessment, though SES typically does as well.

These results highlight two related, albeit distinct, critical issues with neuropsychological literature: representation/generalizability and transparency. This is particularly the case since neuropsychological research occurs across many countries, and more consistent delineation of sample characteristics could improve international communicability.

Regarding representation and generalizability, to take the example of ethnicity/race, non-Hispanic White individuals comprise 60% of the US population (US Census Bureau,

2019). Of the studies for which this information was available (414/1277), the median proportion of non-Hispanic White participants was 75%. In a review of the behavioral science literature, Henrich, Heine, and Norenzayan (2010) noted that broad claims about human behavior are often derived from convenience samples drawn entirely from what they termed WEIRD (Western, Educated, Industrialized, Rich, and Democratic) societies. In their findings, the authors reported how these samples are often, and paradoxically, not representative of the larger population and identified significant variations in brain-behavior relationships highly relevant to the field of neuropsychology – including visuospatial abilities, memory, attention, categorization, and reasoning – threatening notions of the universality of "basic" psychological processes. In a push for characterizing a more representative brain in a cross-disciplinary manner, the population neuroscience literature has explicitly outlined various goals and frameworks for integrating brain-behavior research with representative samples (Falk et al., 2013). The lack of representative samples prevents our profession from developing and providing culturally and linguistically appropriate assessment tools and, consequently, maintains health disparities. Despite numerous documented barriers to recruitment and inclusion (Robinson & Trochim, 2007), under-represented communities do not appear to be less willing to participate in health research (Wendler et al., 2006). While a discussion of the science of recruitment and inclusion for research is beyond the scope of this review, various critical issues and approaches to the effective inclusion of culturally and linguistically diverse communities in research protocols have been reviewed elsewhere (e.g., Hughson et al., 2016; Nápoles et al., 2011; Paskett et al., 2008; Torres-Ruiz et al., 2018; Yancey et al., 2006).

The second critical issue highlighted by the current findings relates to transparency. Adequate description and characterization of study samples are crucial for the replicability of study findings and can help elucidate the reasons for variability when results are not replicated. The overwhelming majority of neuropsychological tests are developed primarily

Table 4. Comparison of percentages with original O’Bryant et al. (2004) results

Journal	Age		Education		Sex/gender		Race/ethnicity		Acculturation		Language		Socioeconomic status	
	Past	Now	Past	Now	Past	Now	Past	Now	Past	Now	Past	Now	Past	Now
ACN	97.3	98.7	82.3	83.1	82.3	93.9	35.0	44.1	0.0	0.0	10.2	15.4	11.1	3.6
JCEN	99.0	100	77.1	77.9	92.3	98.1	20.0	31.7	0.0	0.0	12.0	20.7	8.5	9.1
JINS	97.5	100	72.9	76.7	87.9	96.7	22.9	37.1	0.4	0.0	15.7	17.1	10.4	12.5
NEU	95.7	100	74.5	74.7	84.9	97.0	10.8	23.2	0.0	0.9	13.8	22.3	8.0	12.9
TCN	98.1	99.3	85.6	91.4	90.9	98.6	24.5	56.1	0.0	2.2	6.3	27.3	7.7	6.5
Total	97.6	99.6	77.7	80.8	88.0	89.6	21.5	38.4	0.1	0.6	12.0	20.6	9.0	8.9

Note: All values are percentages. Total values are averages across the journals reviewed and included in the table (does not include percentages of all journals reviewed and included in Table 3). ACN = Archives of Clinical Neuropsychology; JCEN = Journal of Clinical and Experimental Neuropsychology; JINS = Journal of the International Neuropsychological Society; NEU = Neuropsychology; TCN = The Clinical Neuropsychologist.

in the context of WEIRD samples and what is known about performance on these tests is likely restricted to our understanding of these specific samples (Dotson & Duarte, 2020). Honest disclosure of the demographic makeup of samples will allow for the evaluation of the generalizability of reported findings. Moreover, the selection of appropriate tests for a given group will be facilitated. The under-reporting of sample characteristics is not unique to neuropsychology and has not been helped by establishing reporting guidelines (Johansen & Thomsen, 2016). Rad and colleagues (2018) similarly found that important factors, like ethnicity/race and SES, were largely unreported in broader literature related to psychological research. The authors similarly noted that, in addition to omitting this information, researchers in over 80% of the articles reviewed did not discuss their results in the context of the population sampled (Rad et al., 2018). Rad et al. (2018) provide several recommendations for both authors and journals, summarized here. For authors, they recommend (1) reporting of more comprehensive demographic characteristics; (2) justification of the sample studied; (3) contextualization of findings to the population sampled; (4) discussion of the generalizability of the findings (including with regard to culture and context); and (5) analytical investigation of existing diversity (e.g., moderator effects of sex/gender). For journals, Rad and colleagues (2018) recommend (1) consideration of sample diversity as a paper characteristic that can increase its overall contribution; (2) incentivizing and/or rewarding authors who sample diverse populations (e.g., through diversity badges); and (3) explicitly establishing diversity targets (e.g., setting a goal that at least 50% of papers for a given year will sample populations that deviate from typically sampled populations on at least one dimension). We agree with these recommendations, and suspect that many neuropsychological researchers implicitly consider the above in their work; doing so more explicitly is likely to be of benefit.

Taken together, these recommendations would arguably target both the lack of representativeness and the lack of transparency. The American Psychological Association, in their Journal Article Reporting Standards (JARS), lists sampling, recruitment, and participant selection as basic methodological factors to be described when characterizing participant samples in manuscripts (Appelbaum et al., 2018). As our cultural landscape becomes increasingly heterogeneous and neuropsychology becomes more international, expanding our reporting standards to incorporate recommendations for greater inclusion and transparency is not only warranted – it may be vital for our profession to stay relevant both in the healthcare marketplace and as a scientific discipline.

CONFLICT OF INTEREST

The authors have nothing to disclose.

ACKNOWLEDGEMENTS

No funding is associated with the work reported.

REFERENCES

- Adesope, O.O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research, 80*(2), 207–245. <https://doi.org/10.3102/0034654310368803>
- American Psychological Association. (2017). *Multicultural Guidelines: An Ecological Approach to Context, Identity, and Intersectionality*. Retrieved from <http://www.apa.org/about/policy/multicultural-guidelines.pdf>
- Appelbaum, M., Cooper, H., Kline, R.B., Mayo-Wilson, E., Nezu, A.M., & Rao, S.M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist, 73*(1), 3–25. <https://doi.org/10.1037/amp0000191>
- Black, S.A., Espino, D.V., Mahurin, R., Lichtenstein, M.J., Hazuda, H.P., Fabrizio, D., Ray, L.A., & Markides, K.S. (1999). The influence of noncognitive factors on the mini-mental state examination in older Mexican-Americans: Findings from the Hispanic EPESE. *Journal of Clinical Epidemiology, 52*(11), 1095–1102. [https://doi.org/10.1016/S0895-4356\(99\)00100-6](https://doi.org/10.1016/S0895-4356(99)00100-6)
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum. <https://agris.fao.org/agris-search/search.do?recordID=US201300292307>
- Díaz-Venegas, C., Downer, B., Langa, K.M., & Wong, R. (2016). Racial and ethnic differences in cognitive function among older adults in the USA: Cognition of US older adults by race/ethnicity. *International Journal of Geriatric Psychiatry, 31*(9), 1004–1012. <https://doi.org/10.1002/gps.4410>
- Dotson, V.M. & Duarte, A. (2020). The importance of diversity in cognitive neuroscience. *Annals of the New York Academy of Sciences, 1464*(1), 181–191. <https://doi.org/10.1111/nyas.14268>
- Falk, E.B., Hyde, L.W., Mitchell, C., Faul, J., Gonzalez, R., Heitzeg, M.M., . . . Schulenberg, J. (2013). What is a representative brain? Neuroscience meets population science. *Proceedings of the National Academy of Sciences, 110*(44), 17615–17622. <https://doi.org/10.1073/pnas.1310134110>
- Grant, I. & Adams, K.M. (2009). *Neuropsychological Assessment of Neuropsychiatric and Neuromedical Disorders*. New York: Oxford University Press.
- Heaton, R.K., Grant, I., & Matthews, C.G. (1986). Differences in neuropsychological test performance associated with age, education, and sex. In I. Grant & K.M. Adams (Eds.), *Neuropsychological Assessment of Neuropsychiatric Disorders* (pp. 100–120). New York: Oxford University Press.
- Henrich, J., Heine, S.J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*(2–3), 61–83. <https://doi.org/10.1017/S0140525X0999152X>
- Hughson, J., Woodward-Kron, R., Parker, A., Hajek, J., Bresin, A., Knoch, U., Phan, T., & Story, D. (2016). A review of approaches to improve participation of culturally and linguistically diverse populations in clinical trials. *Trials, 17*(1), 263. <https://doi.org/10.1186/s13063-016-1384-3>
- Johansen, M. & Thomsen, S.F. (2016). Guidelines for reporting medical research: A critical appraisal. *International Scholarly Research Notices, 2016*, 1346026. <https://doi.org/10.1155/2016/1346026>
- Lawson, G.M., Hook, C.J., & Farah, M.J. (2018). A meta-analysis of the relationship between socioeconomic status and executive function performance among children. *Developmental Science, 21*(2), e12529. <https://doi.org/10.1111/desc.12529>
- Lezak, M.D., Howieson, D.B., Loring, D.W., Hannay, H.J., & Fischer, J.S. (2004). *Neuropsychological Assessment* (4th ed.). New York: Oxford University Press.
- Morgan, A.A., Marsiske, M., & Whitfield, K.E. (2008). Characterizing and explaining differences in cognitive test performance between African American and European American older adults. *Experimental Aging Research, 34*(1), 80–100. <https://doi.org/10.1080/03610730701776427>
- Mungas, D., Reed, B.R., Crane, P.K., Haan, M.N., & González, H. (2004). Spanish and English Neuropsychological Assessment Scales (SENAS): Further development and psychometric characteristics. *Psychological Assessment, 16*(4), 347–359. <https://doi.org/10.1037/1040-3590.16.4.347>
- Nápoles, A.M., Chadiha, L.A., & Resource Centers for Minority Aging Research. (2011). Advancing the science of recruitment and retention of ethnically diverse populations. *The Gerontologist, 51*(Suppl 1), S142–S146. <https://doi.org/10.1093/geront/gnr019>
- O’Bryant, S.E., O’Jile, J.R., & McCaffrey, R.J. (2004). Reporting of demographic variables in neuropsychological research: Trends in the current literature. *The Clinical Neuropsychologist, 18*(2), 229–233. <https://doi.org/10.1080/13854040490501439>
- O’Driscoll, C., & Shaikh, M. (2017). Cross-cultural applicability of the Montreal Cognitive Assessment (MoCA): A systematic review. *Journal of Alzheimer’s Disease, 58*(3), 789–801. <https://doi.org/10.3233/JAD-161042>
- Opdebeeck, C., Martyr, A., & Clare, L. (2016). Cognitive reserve and cognitive function in healthy older people: A meta-analysis. *Aging, Neuropsychology, and Cognition, 23*(1), 40–60. <https://doi.org/10.1080/13825585.2015.1041450>
- Pan, W. (2002). Approximate confidence intervals for one proportion and difference of two proportions. *Computational Statistics & Data Analysis, 40*(1), 143–157. [https://doi.org/10.1016/S0167-9473\(01\)00107-4](https://doi.org/10.1016/S0167-9473(01)00107-4)
- Paskett, E.D., Reeves, K.W., McLaughlin, J.M., Katz, M.L., McAlearney, A.S., Ruffin, M.T., . . . Gehlert, S. (2008). Recruitment of minority and underserved populations in the United States: The centers for population health and health disparities experience. *Contemporary Clinical Trials, 29*(6), 847–861. <https://doi.org/10.1016/j.cct.2008.07.006>
- Quinteros Baumgart, C., & Billick, S.B. (2018). Positive cognitive effects of bilingualism and multilingualism on cerebral function: A review. *Psychiatric Quarterly, 89*(2), 273–283. <https://doi.org/10.1007/s1126-017-9532-9>
- Rad, M.S., Martingano, A.J., & Ginges, J. (2018). Toward a psychology of *Homo sapiens*: Making psychological science more representative of the human population. *Proceedings of the National Academy of Sciences, 115*(45), 11401–11405. <https://doi.org/10.1073/pnas.1721165115>
- Robinson, J.M. & Trochim, W.M.K. (2007). An examination of community members’, researchers’ and health professionals’ perceptions of barriers to minority participation in medical research: An application of concept mapping. *Ethnicity & Health, 12*(5), 521–539. <https://doi.org/10.1080/13557850701616987>
- Rosenthal, J.A. (1996). Qualitative descriptors of strength of association and effect size. *Journal of Social Service Research, 21*(4), 37–59. https://doi.org/10.1300/J079v21n04_02
- Rossetti, H.C., Smith, E.E., Hynan, L.S., Lacritz, L.H., Cullum, C.M., Van Wright, A., & Weiner, M.F. (2019). Detection of mild cognitive impairment among community-dwelling African Americans using the Montreal cognitive assessment. *Archives*

- of *Clinical Neuropsychology*, 34(6), 809–813. <https://doi.org/10.1093/arclin/acy091>
- Saenz, J.L., Downer, B., Garcia, M.A., & Wong, R. (2018). Cognition and context: Rural–Urban differences in cognitive aging among older Mexican adults. *Journal of Aging and Health*, 30(6), 965–986. <https://doi.org/10.1177/0898264317703560>
- Salthouse, T.A. (2011). Neuroanatomical substrates of age-related cognitive decline. *Psychological Bulletin*, 137(5), 753–784. <https://doi.org/10.1037/a0023262>
- Seidenberg, M., Gamache, M.P., Beck, N.C., Smith, M., Giordani, B., Berent, S., . . . Boll, T.J. (1984). Subject variables and performance on the Halstead Neuropsychological Test Battery: A multivariate analysis. *Journal of Consulting and Clinical Psychology*, 52(4), 658–662. <https://doi.org/10.1037/0022-006X.52.4.658>
- Tan, Y.W., Burgess, G.H., & Green, R.J. (2020). The effects of acculturation on neuropsychological test performance: A systematic literature review. *The Clinical Neuropsychologist*, 1–31. <https://doi.org/10.1080/13854046.2020.1714740>
- Torres-Ruiz, M., Robinson-Ector, K., Atkinson, D., Trotter, J., Anise, A., & Clauser, S. (2018). A portfolio analysis of culturally tailored trials to address health and healthcare disparities. *International Journal of Environmental Research and Public Health*, 15(9). <https://doi.org/10.3390/ijerph15091859>
- US Census Bureau. (2019). *American Community Survey*. Retrieved from <https://data.census.gov/cedsci/>
- Wall, K.J., Cumming, T.B., & Copland, D.A. (2017). Determining the association between language and cognitive tests in poststroke aphasia. *Frontiers in Neurology*, 8. <https://doi.org/10.3389/fneur.2017.00149>
- Wendler, D., Kington, R., Madans, J., Van Wye, G., Christ-Schmidt, H., Pratt, L.A., . . . Emanuel, E. (2006). Are racial and ethnic minorities less willing to participate in health research? *PLoS Medicine*, 3(2), e19. <https://doi.org/10.1371/journal.pmed.0030019>
- Wong, H.S. & Edwards, P. (2013). Nature or nurture: A systematic review of the effect of socio-economic status on the developmental and cognitive outcomes of children born preterm. *Maternal and Child Health Journal*, 17(9), 1689–1700. <https://doi.org/10.1007/s10995-012-1183-8>
- Wu, Y.-T., Prina, A.M., & Brayne, C. (2015). The association between community environment and cognitive function: A systematic review. *Social Psychiatry and Psychiatric Epidemiology*, 50(3), 351–362. <https://doi.org/10.1007/s00127-014-0945-6>
- Xu, H., Zhang, Y., & Wu, B. (2017). Association between migration and cognitive status among middle-aged and older adults: A systematic review. *BMC Geriatrics*, 17(1), 184. <https://doi.org/10.1186/s12877-017-0585-2>
- Yancey, A.K., Ortega, A.N., & Kumanyika, S.K. (2006). Effective recruitment and retention of minority research participants. *Annual Review of Public Health*, 27(1), 1–28. <https://doi.org/10.1146/annurev.publhealth.27.021405.102113>
- Yano, K., Grove, J.S., Masaki, K.H., White, L.R., Petrovitch, H., Chen, R., . . . Curb, J.D. (2000). The effects of childhood residence in Japan and testing language on cognitive performance in late life among Japanese American men in Hawaii. *Journal of the American Geriatrics Society*, 48(2), 199–204. <https://doi.org/10.1111/j.1532-5415.2000.tb03912.x>
- Zell, E., Krizan, Z., & Teeter, S.R. (2015). Evaluating gender similarities and differences using metasynthesis. *American Psychologist*, 70(1), 10–20. <https://doi.org/10.1037/a0038208>