Acta Psychologica 249 (2024) 104457

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

# The bias and precision of reporting the average age of human participants

Jeroen B.J. Smeets

*Department of Human Movement Sciences, Vrije Universiteit Amsterdam, van der Boechorststraat 9, NL-1081 BT Amsterdam, the Netherlands.*

ARTICLE INFO

*Keywords:*
Age factors
Development
Uncertainty
Age distribution
PsycINFO **2240** Statistics & Mathematics

ABSTRACT

Most research papers in psychology study the behaviour of a sample of participants. To characterise this sample, authors report various characteristics, frequently including the mean age and the associated standard deviation. However, based on reports from authors who publish in Acta Psychologica and from respondents on X/Twitter, the present paper shows that some authors use rounded-down ages whereas others don't, which lead to an uncertainty of 0.5 year in the average age. The results furthermore show that the authors tend to report the average age with two decimals precision, irrespective of the uncertainty of this average. I recommend that publications should explicitly mention how the average age is determined and report its value using a number of decimals that reflects its uncertainty.

## 1. Introduction

Most studies involving human participants provide some information on the age of the participants. It is in general relevant to have an idea of the age of the participants because the behaviour under study is likely to depend in some way on age. Indeed, the publication manual of the American Psychological Association requires you to report age and do so preferably by providing extensive summary statistics (e.g. not only the range but also the mean, median and standard deviation; American Psychological Association, 2020). However, age reporting has an important caveat: it is very uncommon to report one's age with fractions or decimals (with some exceptions, e.g. Townsend, 1982). Age is generally reported as rounded down, i.e. a new-born baby has the age of zero and will keep this age until its first birthday (although some cultures used to round the age up; Coale & Li, 1991). How does this rounding down to whole numbers when reporting one's age affect the summary statistics of the age distribution of participants?

A simple example shows that rounding down can have considerable effects. As an example, I will consider the age distribution of a youth soccer team, which (in the Netherlands) consists of children who are born in the same year (Koninklijke Nederlandse Voetbalbond, 2024). If the team would consist of children born in 2010, all children will report to be 13 years old on New Year's Eve 2023, so their rounded down age will be 13. The distribution of actual ages will be uniform 13.000 to 13.999, which corresponds to an average of 13.5. Basing the average on a rounded down age thus introduces a bias of 0.5 year. The use of rounded down age will also affect the standard deviation. As all children will all report the same (rounded down) age on New Year's Eve, the standard deviation will be zero. The standard deviation of the rounded down age will increase during the first half of the year (up to 0.5 on July 1st) and will then decrease again to zero (Fig. 1). The actual ages have a constant standard deviation, which equals the standard deviation of uniform distribution of one year, which is $1/\sqrt{12} \approx 0.289$. Rounding down thus can affect both the mean and standard deviation: on New Year's Eve 2023, the rounded down age is 13.000 ± 0.000 (μ ± σ), whereas the actual ages of the children will be 13.500 ± 0.289. As the bias in the reported mean age that is introduced by rounding down the individual age is exactly 0.5, one can correct for this when reading a paper. However, to do so, one should know that authors have based their mean on rounded down ages. The first question I will address in this paper is: do authors consistently base the average age they report on rounded down ages?

A separate issue is how many decimals one should use so report the mean age. In the above reasoning, I based my calculations on a perfectly uniform distribution of ages and reported the values in three decimals. In real situations, we have samples of a limited number of participants, so the distribution won't be perfectly uniform. If the sample-size is 9 (not unreasonable for the team of the example), the uncertainty of the estimate of the mean (the standard error of the mean, SEM) is $\sigma/\sqrt{N} \approx 0.1$ year. In this case, it would make more sense to report the mean age (and standard deviation) with no more decimals than the SEM, so only one decimal: 13.5 ± 0.3. I here follow an uncertainty-based rule that has been for instance proposed by Cole (2015): report the mean with a number of decimals that matches the one-significant digit SEM. In other words: report only meaningful digits. This uncertainty-based rule can be refined by combining the SEM with other error-sources (Cousineau,

*E-mail address:* J.B.J.Smeets@vu.nl.

https://doi.org/10.1016/j.actpsy.2024.104457
Received 30 April 2024; Received in revised form 21 July 2024; Accepted 7 August 2024
Available online 10 August 2024
0001-6918/© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

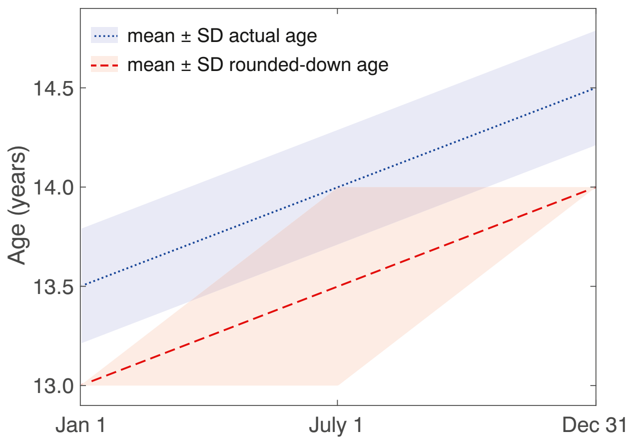


**Fig. 1.** The development of the average age (lines) in 2024 for all children born in 2010. The average actual age (expressed in fraction of years, dotted) is always half a year above the age rounded down to an integer (dashed). Whereas the standard deviation (shaded area) of the actual age remains constant at 0.28, the standard deviation of the rounded-down age varies from zero at the turn of the year to 0.5 halfway the year. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2020).

At several places, one can find rules for reporting means that differ from the precion-based rule discussed in the previous paragraph. For instance, the APA rules "When reporting data measured on integer scales …, report means and standard deviations to one decimal place" (American Psychological Association, 2020, p180) in line with the advice to report means with "no more than one extra decimal place over the raw data" (Altman & Bland, 1996, p572). This rule is independent of the SEM, in contrast with the uncertainty-based rule. In cases in which you have thousands of participants (e.g. Yang, Craig, Anderson, Ross, & Muntaner, 2024), the SEM is often lower than 0.09 year, so one can report the mean using two decimals according to the uncertainty-based rule, whereas in the cases in which you have only 10 participants and a SEM of more than 1 year (e.g. Bretschneider, Meyer, & Asbrock, 2023), one should report the mean without any decimals. Do authors consistently follow the APA rule to determine the number of decimals, or do they consider the uncertainty of the mean?

## 2. Methods

The aim of this paper is to find whether authors use a consistent way of reporting age. As the way ages are reported might be affected by editorial instructions of the various, journals, I decided to restrict myself to a single journal. If inconsistency in age-reporting is a serious issue, it would be visible in checking a few dozens of papers. Therefore, I decided to use all papers in the three most recent volumes of Acta Psychologica at the time of writing (240–242).

This strategy resulted in 81 papers which I checked for reports on age. Only a few papers indicated how they obtained the values for the age. To obtain information on whether authors used the actual age or the rounded-down age, I contacted the corresponding author of each study and asked how the age was determined. I provided four possible answers:

a) by asking the participants their age (in years)
b) by asking the participants their date of birth and calculating their age (in years)
c) by asking the participants their date of birth and calculating their age (in years and days)
d) other (please specify)

After one email and one reminder, I obtained answers for 38 of the 65 papers volumes 240–242.

I used X (Twitter) to obtain additional information from people in my network. Given my own research field, they might be within the field of perception and action control. Here I formulated the question as "A question for those who have human participants in their research. How do you determine the age of your participants (with which resolution)?" with shortened versions of the same four options as I used for the authors, without the "please specify" with option 'd'. This poll received 72 answers.

Thirdly, I determined for the 81 papers whether mean ages and the associated standard deviations were mentioned. If so, I noted these values as well as the number of participants and the standard deviation. I noted the number of decimals of the mean and standard deviation. In some cases, the mean and standard deviation were provided with a different number of decimals. In these cases, I used the highest of the two for the analysis, assuming that the reason for the difference was that one value had a "0" as the last decimal (which was omitted). If a paper reported the results of more than one group of participants, I used the data for two groups; if there were more than two groups, I used the ones with the largest and smallest number of participants and/or standard error of the mean age. In this way, I obtained data from 65 groups of participants. The groups were variable in size (ranging from 10 to 45,018) and in the homogeneity of ages (σ ranging from 0.34 to 32 years).

## 3. Results

For both groups (authors of Acta Psychologica papers and people on X/Twitter), about 60 % reported obtaining ages by asking the participants; 40 % used other methods, including ones that produced ages at a better resolution than a year (Fig. 2). The authors specified what 'other' method they used and reported most frequently to obtain the age in month resolution or to use age-ranges. The authors who used age-ranges did not report the average age in their paper.

The precision with which the mean age was reported ranged from zero to three decimals, with two decimals occurring most frequently (in 45 of the 65 studies), and three decimals occurring only once (Fig. 3). The number of decimals appears to be independent of the uncertainty and the number of participants, and in general more decimals are reported than I would expect based on either the APA rule or on the uncertainty (SEM).

## 4. Discussion

The aim of this paper is to find whether authors use a consistent way of reporting age. My results (Fig. 2) show that researchers use various methods to determine the age of their participants. Most common is the rounded-down age (in years), but also methods that have a better resolution are used by a substantial number of authors. This finding holds not only for the authors who publish in Acta Psychologica, but also for respondents on X/Twitter. The results furthermore show that the authors of papers in Acta Psychologica tend to report the average age with two decimals precision, irrespective of the uncertainty of their average (Fig. 3).

When using the rounded-down age as most authors do, one underestimates the age by half a year (Fig. 1). As authors don't disclose how they obtain the values for the age, it is impossible for the reader to decide whether the value should be corrected for this underestimation or not. When reporting the age of groups with a narrow age-range, such as a school class, also the standard deviation might be biased (shaded areas in Fig. 1). Are these biases problematic? I think that for most studies, the bias of half a year is not very problematic, as the average age is more than 50 times larger than the potential bias). Of course, it might be problematic in developmental studies in which half a year is considerable. Moreover, when studying development, one may tend to use groups with a small age-range, which might introduce a bias in the

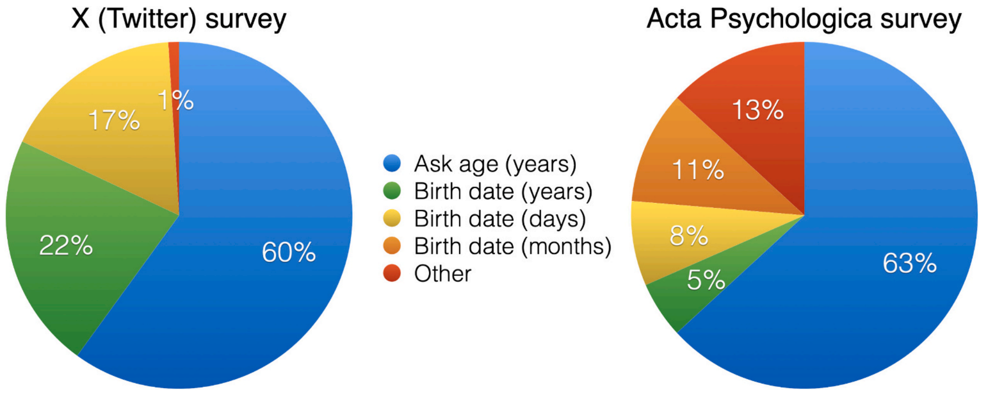


**Fig. 2.** How behavioural scientists obtain the ages of their participants. As the authors could specify the 'other' category, I created for part of the 'other' responses an additional category 'birth date (months)'.

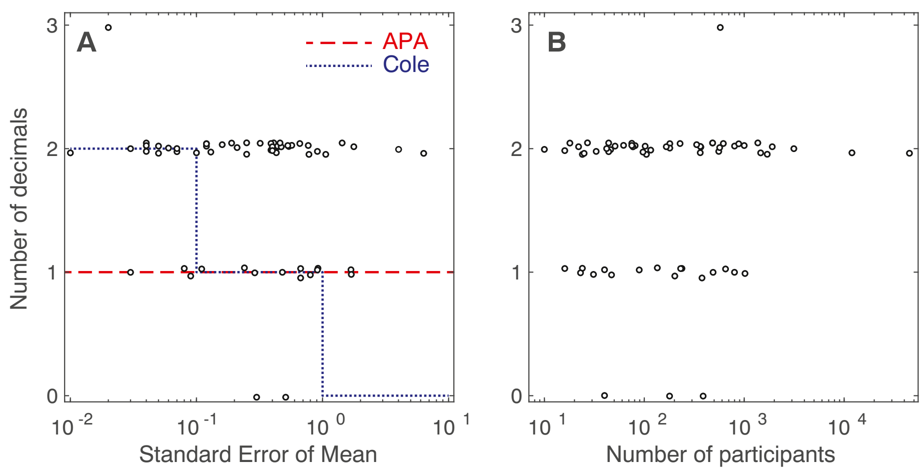


**Fig. 3.** The relation between the number of decimals used to report the mean age and parameters of the data. The number of decimals with which the age is reported ranges from zero the three, independent of the standard error of mean (A) and independent of the number of participants (B). Circles are values from a group of participants. The dashed line is the APA recommendation (American Psychological Association, 2020) and the dotted line indicates the number of meaningful decimals given the uncertainty (SEM; Cole, 2015). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

estimation of variability.

The authors quite consistently reported the mean using two decimals. This is neither in line with the APA-rules nor with the uncertainty rule (Fig. 3); it might be caused by SPSS's default setting to display two decimals, or by earlier versions of the APA publication manual (American Psychological Association, 2010). Indeed, Acta's 'Guide for authors' does not specify that APA style should be used (except for references) nor does it give other instructions on formatting descriptive statistics. Measuring age at a resolution of a year does not prohibit obtaining an average with a low uncertainty if the number of participants is high, as the deviations introduced by the low-resolution will be random and thus average out. The rule of APA therefore does not make much sense. Better make your estimate of uncertainty on the basis of the data (Cousineau, 2020) to determine a sensible number of decimals.

## 5. Conclusion

Authors tend to suggest a low uncertainty (in the order of weeks) despite some of them introducing a bias of half a year. I have two recommendations to solve this issue:

1) To be able to judge the presence of a bias, publications should explicitly mention whether the average age is based on the actual age or the rounded-down age.
2) The number of decimals used to display an average should correspond to the uncertainty with which that average is known.

## CRediT authorship contribution statement

**Jeroen B.J. Smeets:** Writing – original draft, Visualization, Methodology, Conceptualization.

## Declaration of competing interest

None.

## Data availability

Data will be made available on request.

## References

Altman, D. G., & Bland, J. M. (1996). Presentation of numerical data. *BMJ, 312*(7030), 572. https://doi.org/10.1136/bmj.312.7030.572
American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). American Psychological Association.
American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th ed.). American Psychological Association. https://doi.org/10.1037/0000165-000

Bretschneider, M., Meyer, B., & Asbrock, F. (2023). The impact of bionic prostheses on users' self-perceptions: A qualitative study. *Acta Psychologica, 241*, Article 104085. https://doi.org/10.1016/j.actpsy.2023.104085

Coale, A. J., & Li, S. (1991). The effect of age misreporting in China on the calculation of mortality rates at very high ages. *Demography, 28*(2), 293–301. https://doi.org/10.2307/2061281

Cole, T. J. (2015). Too many digits: The presentation of numerical data. *Archives of Disease in Childhood, 100*(7), 608. https://doi.org/10.1136/archdischild-2014-307149

Cousineau, D. (2020). How many decimals? Rounding descriptive and inferential statistics based on measurement precision. *Journal of Mathematical Psychology, 97*, Article 102362. https://doi.org/10.1016/j.jmp.2020.102362

Koninklijke Nederlandse Voetbalbond. (2024). Handboek Competitie Veldvoetbal. https://www.knvb.nl/downloads/bestand/28298/knvbhandboek-veldvoetbal-2024-25.

Townsend, S. (1982). *The secret diary of Adrian mole, aged 13 3/4*. Methuen.

Yang, W., Craig, S. L., Anderson, J. A. E., Ross, L. E., & Muntaner, C. (2024). Sexual orientation and cognition in aging populations: Results from the Canadian longitudinal study on aging. *Acta Psychologica, 242*, Article 104117. https://doi.org/10.1016/j.actpsy.2023.104117