

# How Long Should We Keep Data? An Evidence-Based Recommendation for Data Retention Using Institutional Meta-Analyses

Drew N. Wright, Michelle R. Demetres, Keith C. Mages, Antonio P. DeRosa, Caroline Jedlicka, Judy Stribling, Becky Baltich Nelson, Diana Delgado

## Objective

The National Institutes of Health (NIH) is committed to ensuring results and data are made available to the public for research they fund, in whole or in part. Therefore, institutions and researchers must maintain accessible, sharable data for published studies. Using published meta-analyses (MAs) we sought to establish an evidence-based recommendation on how long data should be kept and shared.

## Methods

Using Web of Science (Core Collection – Clarivate Analytics) we identified all MAs with an affiliation of Weill Cornell Medicine (WCM) published through November 2019. The following search was used and identified 211 WCM affiliated meta-analyses: OO=(weill) AND TI=("meta-analysis") AND DT=(Article OR Review). The full text of each MA was reviewed. We then looked at all studies that met inclusion criteria for each MA and documented the publication date of the oldest included paper. For all 211 included MAs in this study, journal category (discipline) and impact factor was identified per Thomson Reuter's Journal Citation reports via Web of Science. All data and trends were analyzed using Power BI (Microsoft).

Contact: Michelle R. Demetres  
mrd2006@med.cornell.edu

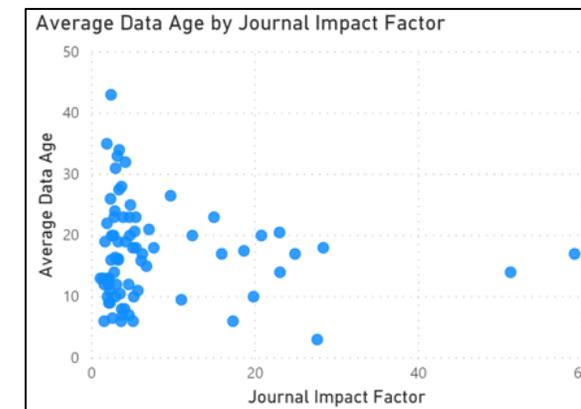
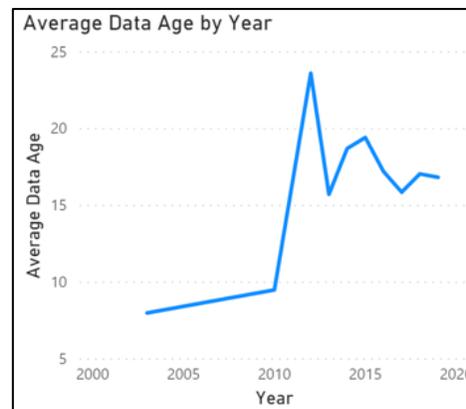


**Weill Cornell Medicine**

Samuel J. Wood Library & C.V. Starr  
Biomedical Information Center

## Results

**The average age of included studies is 17 years. The oldest data set utilized in a MA was 51 years old at the time of publishing, and the youngest was 2.** WCM meta-analyses ranged in discipline, covering 37 out of 76 Web of Science medical-related journal categories. There appeared to be no strong association between data age and impact factor, but there is a possible correlation with data age and year of MA publication suggesting that newer publications include older studies relative to themselves.



## Conclusion

Based on our findings we recommend data be stored for sharing for a minimum of 17 years. This study lays groundwork for larger future studies.