

Supplementary information

**Revisiting the theoretical and
methodological foundations of depression
measurement**

In the format provided by the
authors and unedited

Revisiting the theoretical and methodological foundations of depression measurement

Supplementary Information

Eiko I. Fried (1)*

Jessica K. Flake (2)

Donald J. Robinaugh (3,4)

(1) Department of Clinical Psychology, Leiden University, The Netherlands

(2) Department of Psychology, McGill University, Canada

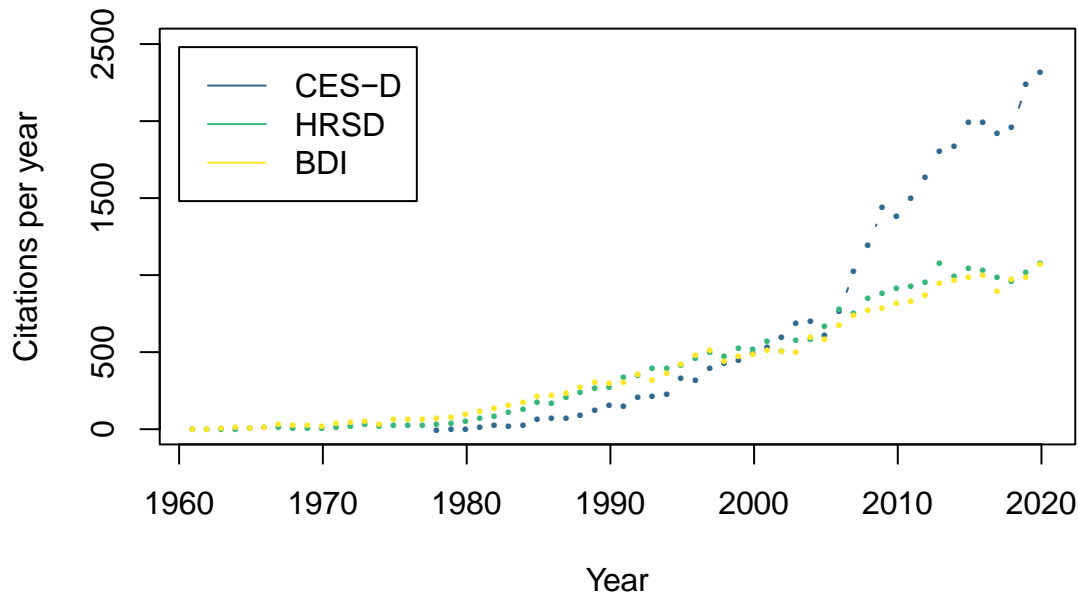
(3) Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts, US

(4) Department of Applied Psychology, Northeastern University, US

*Corresponding author: eikofried@gmail.com

Supplementary Note 1

Supplementary Figure 1 shows citations per year for the CES-D, HRSD, and BDI.



The figure is fully reproducible using supplementary code and data available at <https://osf.io/7dp5s/>. Data are based on a web of science export from April 29 2021.

Supplementary Note 2

Figure 2 is fully reproducible using supplementary code and data available at <https://osf.io/7dp5s/>.

For the categorical figure (Figure 2a), we simulated data with the same N as for the observed DASS-42 data around 2 distributions of means 9 and 34, with a standard deviation of 3 in both cases.

For the figure showing the 14-item depression subscale of the DASS-42 (Figure 2b), we used an open dataset downloaded from openpsychometrics.org on July 19 2021; the corresponding row on the website to access the dataset is entitled "8/14/2018 // Answers to the Depression Anxiety Stress Scales // 42 DASS items, 30+ personality and demographic items // 39,775 // DASS_data_21.02.19.zip". These data include response time (in milliseconds). To ensure that we were relying on reliable data, we deleted entries in which answering one of the 14 DASS depression items took less than 2,500ms or more than 120,000ms, resulting in 12,614 valid responses (of 39,775 total recorded responses). We rescored items on a 0-3 (rather than 1-4) range, as per the official DASS-42 scoring form, and added the 14 items to a sum score. We

dichotomized participants at the cutoff of >21 , which corresponds to severe depression in the DASS-42 manual. We note that this dataset is likely an example of survey topic participation bias, which led to higher than usual depression scores (<https://openpsychometrics.org/info/survey-topic-participation-bias/>) .

For the 9-item PHQ-9 figure c), we used summary statistics provided in supplementary table 3 of the paper by Levis et al. 2018 (<https://doi.org/10.1192/bjp.2018.54>).

Supplementary Note 3

We performed one simulation. We set a seed to make this simulation result reproducible, with the goal of obtaining an MDD prevalence of 0.30 and a kappa of 0.28, as reported in the DSM-5 field trials for Major Depressive Disorder (MDD). To do so, we simulated data for $N=100$ from a correlation matrix that specifies relations between the

- 1. true state (MDD or not)
- 2. the first clinician (MDD diagnosis or not)
- 3. the second clinician (MDD diagnosis or not).

The relation among clinicians was set to reach a kappa of 0.28; the relation between clinicians and the ground truth was set to the maximally possible value for the simulation to produce results (it is mathematically impossible to have a relation between clinicians and ground truth of 1, and a relation between clinicians of 0.28).

The simulation resulted in a depression prevalence of .30, and a kappa coefficient of .29, very close to our desired outcome (meaning that our resulting graph, if anything, errs on the side of overestimating similarity among two clinicians). R code for the analyses that resulted in Figure 3 can be found at <https://osf.io/7dp5s/>.