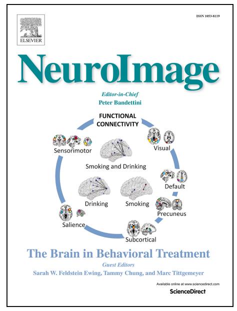
Accepted Manuscript



Data quality over data quantity in computational cognitive neuroscience

Antonio Kolossa, Bruno Kopp

PII: S1053-8119(18)30005-3

DOI: 10.1016/j.neuroimage.2018.01.005

Reference: YNIMG 14618

To appear in: *NeuroImage*

Received Date: 26 August 2017

Revised Date: 28 November 2017

Accepted Date: 3 January 2018

Please cite this article as: Kolossa, A., Kopp, B., Data quality over data quantity in computational cognitive neuroscience, *NeuroImage* (2018), doi: 10.1016/j.neuroimage.2018.01.005.

Data Quality over Data Quantity in Computational Cognitive Neuroscience

Antonio Kolossa[1], Bruno Kopp[1*]

Abstract—**We analyzed factors that may hamper the advancement of computational cognitive neuroscience (CCN). These factors include a particular statistical mindset, which paves the way for the dominance of statistical power theory and a preoccupation with statistical replicability in the behavioral and neural sciences. Exclusive statistical concerns about sampling error occur at the cost of an inadequate representation of the problem of measurement error. We contrasted the manipulation of data quantity (sampling error, by varying the number of subjects) against the manipulation of data quality (measurement error, by varying the number of data per subject) in a simulated Bayesian model identifiability study. The results were clear-cut in showing that - across all levels of signal-to-noise ratios - varying the number of subjects was completely inconsequential, whereas the number of data per subject exerted massive effects on model identifiability. These results emphasize data quality over data quantity, and they call for the integration of statistics and measurement theory.**

Index Terms—**computational modeling, functional brain imaging, signal-to-noise ratio, reliability, replicability**

I. Introduction

Many empirical studies across a broad range of behavioral sciences may be associated with insufficient statistical power (Cohen, 1988). Insufficient statistical power typically occurs in small sample-size studies, in which small-to-medium effect sizes are under scrutiny. Under circumstances of insufficient statistical power, the inflation of false positive discoveries erodes our confidence in the factual significance of 'statistically significant' findings, and it endangers their replication (Chase and Chase, 1976; Sedlmeier and Gigerenzer, 1989; Button et al., 2013; Krzywinski and Altman, 2013; Miller and Ulrich, 2016; Munafò et al., 2017; Szucs and Ioannidis, 2017). The recognition of the problems that are associated with insufficient statistical power, with increased false positive discovery rates, and hence with inadequate replicability, led to many methodological refinements of - and meta-statistical discussions about - null hypothesis testing (Cumming, 2014; Gelman and Carlin, 2014; Halsey et al., 2015; Harlow et al., 2016; Kruschke and Liddell, 2017). The development of new computer-based technologies allowed hitherto inconceivable developments, most notably the open science framework (OSF; `https://osf.io` (Nosek et al., 2015)), and it led to several multi-lab collaboration initiatives (e.g., Hagger et al., 2016; Wagenmakers et al., 2016). We collectively refer to these consensual reactions to the replication crisis in behavioral sciences as the big/open/pre-registered data remedy.

[1]Hannover Medical School, Hannover, Germany. Department of Neurology, e-mail: antonio.kolossa@gmail.com
[1*]Corresponding author: B. Kopp, Department of Neurology, Hannover Medical School, Hannover, Germany; Carl-Neuberg-Str. 1, 30625 Hannover, Germany. Tel.: +49 511 532 2439, e-mail: kopp.bruno@mh-hannover.de.

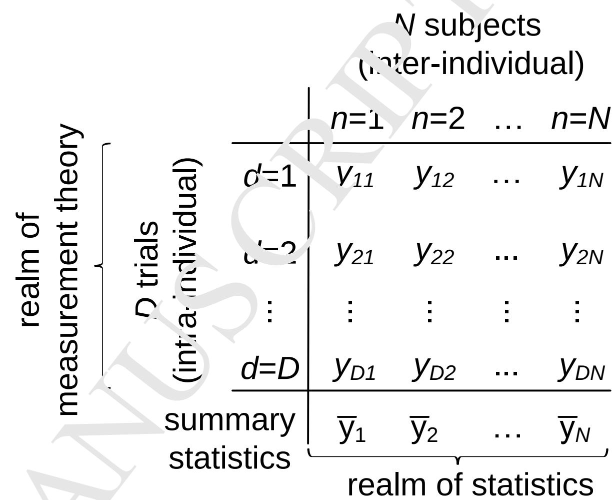


Fig. 1. A typical data matrix. Summary statistics represent usually a measure of central tendency (mean or median), and they enter statistical analysis.

While there is nothing wrong with the big/open/pre-registered data remedy, another dimension of the typical scientific data matrix receives comparatively little attention. This dimension can be thought of as the intra-individual (in contrast to the inter-individual) dimension of measurement. Figure 1 outlines our framework. Let us define the number of subjects (i.e., sample size N) as the (inter-individual) *quantity* dimension of such a data matrix (we have just rehearsed that we should ensure sufficiently large N to optimize statistical power, given typical small-to-medium effect sizes), which is subject to sampling error as described above. The second dimension constitutes the number of data per subject (here referred to as D), and we term this dimension of a data matrix its (intra-individual) *quality* dimension, which is subject to measurement error. In psychometric theory (Nunnally and Bernstein, 1994; Raykov and Marcoulides, 2011; Revelle, 2014), reliability quantifies measurement error, whereas the calculation of signal-to-noise ratios (SNR) occurs more often as the quantification of measurement error in neuroscience (Schimmel, 1967; Coppola et al., 1978; Başar, 1980; Möcks et al., 1988; Raz et al., 1988; Puce et al., 1994; beim Graben, 2001; Paukkunen et al., 2010; Kolossa and Kopp, 2016).

We present a short summary of some essential background material for a deeper understanding of our D (a variable