“工欲效其事，必先信其器”——个体差异的科学靠谱吗？

邢秀侠¹, 左西年^{2*}

1. 北京工业大学应用数理学院, 北京 100124;

2. 中国科学院心理研究所磁共振成像研究中心, 北京 100101

* 联系人, E-mail: zuoxn@psych.ac.cn

“To do a valid job, must make tools reliable first” —A decent science of individual differences?

Xiuxia Xing¹ & Xinian Zuo^{2*}

¹ College of Applied Sciences, Beijing University of Technology, Beijing 100124, China;

² Magnetic Resonance Imaging Research Center, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China

* Corresponding author, E-mail: zuoxn@psych.ac.cn

doi: 10.1360/TB-2019-0270

作为人类特有的行为, 科学研究是社会文明的重要推动力量之一. 近年来, 研究的可重复性问题成为科学关注的焦点, 从心理科学到临床医学等领域, 研究的可重复性成为巨大挑战. 生命科学研究的共同特点之一是对测量工具的需求, 先进的技术会促进更为精准的测量, 提升研究可信度. 测量理论中的信(可信)效(有效)度概念在不同学科都有涉及, 特别是在心理科学和医学中有明确的统计学界定, 但在其他学科未被充分认识, 尤其是交叉学科.

人与人之间为何表现出如此巨大的差别? 这是个体差异的科学问题, 一直以来是生命科学领域备受关注的焦点. *Nat Hum Behav* 于 2019 年 6 月 28 日在线发表题为“Harnessing reliability for neuroscience research”的评论文章^[1], 以神经科学为例, 聚焦神经影像技术, 提出了个体差异测量信效度统计学框架. 在此框架下, 个体差异的总测量由三部分组成(图 1(a)): 研究对象特异的变化(疾病或特质测量)、研究对象非特异的变化(干扰和污染测量)、随机错误(随机噪声测量). 个体差异的测量采用重复测量设计(图 1(b)), 图中相同颜色表示同一个人, 不同颜色表示不同的人, 个体间差异测量包含“疾病或特质测量”与“干扰与污染测量”的总和, 而随机噪声测量则是个体内差异测量. 个体差异测量的信度是个体间差异测量在总测量中所占比例, 而疾病或特质特异变化所占比例则是个体差异测量的效度. 由此, 测量的信度就像一个瓶子的盖子一样, 牢牢地限制住了测量的效度, 不可信的测量永远不可能有效; 与此同时, 测量的个体间差异越大, 其信度越高, 测量个体内差异越小, 其信度越高; 最后, 测量信度越高,

其检测统计效应所需样本量越小.

那么, 到底什么原因使得个体测量的信度如此重要呢? 从理论上讲, 结合上述 3 条测量信效度的统计定律, 加上在实际应用和实践中, 对于测量的效度是无法直接进行测量的(否则就没必要进行研究了, 因为疾病或特质有效性已经解决的话, 就意味着相应科学问题已经回答). 从实际应用上讲, 教育实践中的“因材施教”和临床实践中的“精准医疗”都体现出个体差异测量研究的价值, 高信度测量意味更易于区别不同的学生或病人, 而在不同场合的测量稳定性也更好. 综合来看, 高信度测量对个体差异研究和应用转化至关重要.

近 10 年来, 神经影像因为其安全性和高时空精度的优势, 已经积累了大量数据, 成千上万的人脑影像已经上线并公开, 涵盖了人类在不同发展阶段和各类脑疾病障碍上的影像^[2]. 由此而催生了开放式神经科学的出现, 推动了大型化脑科学(比如人工智能和脑疾病生物标记物)研究. 个体差异研究的基础是统计力度, 其决定了检测实验效应的能力. 大样本量是提高统计力度的因素之一, 然而如果测量信度不够, 就会产生对大样本量的不必要需求. 在此评论文章中, 研究团队采用蒙特卡洛方法对信度、样本量和效应量之间的关系进行了数值模拟. 结果揭示: 在神经影像领域, 潜在效应量较小, 测量的信度局限将会极大地增加研究对样本量的需求. 神经影像测量的信度研究表明: 现有数据中极少有足够的个体数据能获得高度可信的脑连接测量. 各国推出的各类大型脑计划中, 个体差异的基础和转化研究(教育和临床)是中国脑计划的核心和特色,

图1 (网络版彩色)个体差异测量的组成(a)与信效度(b)

Figure 1 (Color online) Individual differences-reliability (a) and validity as well as composition (b)

首要解决的基础科学问题是测量信效度, 据此优化和标准化大型数据测量规范, 防止产生大量的低质量数据样本。

神经影像领域的各类实验所采用测量的信度水平不一。以磁共振成像为例, 脑形态测量最为可信, 信度可达到临床转化研究对信度的要求。功能磁共振的测量则要低很多, 并且依赖于扫描时间长短, 其中认知任务功能磁共振测量实验设计直接来源于实验心理学领域的行为实验, 个体间差异受到局限, 导致其测量信度较低。从统计学的角度, 低可信度的测量加上小样本将增加科学研究的假阳性率, 降低了研究的可重复性, 基因组学已经将其视为领域挑战并计划长期研究来试图克服。反过来, 基于高可信度测量的研究结果在科学期刊上出现得更为频繁, 比如: 人脑各种脑网络中, 默认网络和额定网络的测量可信度更高^[3,4]。因此各类研究更易于检测到其统计效应, 但可能这只是其测量信度高的一种表现, 而并不是这些脑网络对实验效应特异性的体现。

整个生命科学领域应当关注测量信度的评测与优化。特别是神经科学领域, 通常对这一问题未给予足够重视,

无论测量手段是影像、电生理、神经炎症标记物、微生物组学, 还是认知神经科学范式、私人穿戴设备等, 都应将测量信度及其决定因素作为基本问题来研究。近 10 年来, 中国科学院心理研究所与国际合作团队一起, 针对个体差异测量理论进行了长期而系统的部署, 重点培育“心理行为的个体差异及其毕生发展规律与应用”研究方向, 领衔建立了“国际信度与可重复性联盟^[5]”, 提出了大型脑智毕生发展项目“彩巢计划——成长在中国^[6,7]”, 参与北京市脑计划, 创建主办双年度“国际人脑发展会议^[8]”, 为参与和推动国家脑计划做出了具有国际影响力的贡献。

《论语·卫灵公》写到“工欲善其事, 必先利其器”, 比喻要做好一件事, 准备工作非常重要。神经科学家致力于探索大脑真相(效度), 并将相关基础研究成果转化为临床工具, 检测和优化测量的信度必须成为这些转化研究的前提和准备^[9-11], 需要科研人员改进当前研究实践、科研基金管理部门的支持, 共同产生开放社区资源以用于这些基本特性的定量化, 正所谓“工欲效其事, 必先信其器”。不久的将来, 期待科学揭开脑与行为关系的真相, 造福人类。

推荐阅读文献

- 1 Zuo X N, Xu T, Milham M P. Harnessing reliability for neuroscience research. *Nat Hum Behav*, 2019, doi: 10.1038/s41562-019-0655-x
- 2 Milham M P, Craddock R C, Son J J, et al. Assessment of the impact of shared brain imaging data on the scientific literature. *Nat Commun*, 2018, 9: 2818
- 3 Zuo X N, Xing X X. Test-retest reliabilities of resting-state FMRI measurements in human brain functional connectomics: A systems neuroscience perspective. *Neurosci Biobehav Rev*, 2014, 45: 100-118
- 4 Chen B, Xu T, Zhou C, et al. Individual variability and test-retest reliability revealed by ten repeated resting-state brain scans over one

- month. PLoS One, 2015, 10: e0144963
- 5 Zuo X N, Anderson J S, Bellec P, et al. An open science resource for establishing reliability and reproducibility in functional connectomics. *Sci Data*, 2014, 1: 140049
 - 6 Zuo X N, He Y, Betzel R F, et al. Human connectomics across the life span. *Trends Cogn Sci*, 2017, 21: 32–45
 - 7 Yang N, He Y, Zhang Z, et al. Chinese Color Nest Project: Growing up in China (in Chinese). *Chin Sci Bull*, 2017, 62: 3008–3022 [杨宁, 何叶, 张喆, 等. 彩巢计划——“成长在中国”. *科学通报*, 2017, 62: 3008–3022]
 - 8 Zuo X N, He Y, Su X Q, et al. Developmental population neuroscience: Emerging from ICHBD. *Sci Bull*, 2018, 63: 331–332
 - 9 Xing X X, Zuo X N. The anatomy of reliability: A must read for future human brain mapping. *Sci Bull*, 2018, 63: 1606–1607
 - 10 Zuo X N, Biswal B B, Poldrack R A. Editorial: Reliability and reproducibility in functional connectomics. *Front Neurosci*, 2019, 13: 117
 - 11 Xu T, Yang Z, Jiang L L, et al. A connectome computation system (CCS) for discovery science of brain. *Sci Bull*, 2015, 60: 86–95