**对*Lau Lilleholt(2023)*研究结果的计算可复现性检验**

小组成员分工

|  |  |  |  |
| --- | --- | --- | --- |
| 组长 | 阎佳怡 | | |
| 组员 | 蒋承欢、朱妍懿 | | |
| 分工 | | | |
| 数据分析 | 阎佳怡（80%）、蒋承欢（10%）、朱妍懿（10%） | PPT 制作 | 蒋承欢（40%）、朱妍懿（30%）、阎佳怡（30%） |
| 文字报告制作 | 蒋承欢（20%）、阎佳怡（80%） | PPT 展示 | 阎佳怡 |

\* 同一名同学可负责多个部分；如同一内容由多位同学负责，可按百分比注明贡献占比

**摘要**：流行疲劳被定义为一种逐渐出现的主观疲倦和疲惫状态，以及遵循推荐的健康保护行为（包括在大流行期间让自己了解情况）的普遍消退，学界一直在争论其存在和性质。本文介绍了流行疲劳量表，并使用来自丹麦的面板调查和丹麦和德国的重复横断面调查的数据 (总体 *N* = 34,582) 展示了流行疲劳在 COVID-19 大流行期间是如何演变的。流行疲劳的相关矩阵表明流行疲劳与人们自我报告的对推荐的健康保护行为的依从性呈负相关。预先注册的在线实验 (*N* = 1584) 中流行疲劳的 (去) 激励方面进一步表明流行疲劳会对人们坚持推荐的健康保护行为的意愿产生负面影响。综上所述，这些发现不仅为流行疲劳的存在提供了证据，还为其心理和行为关联提供了证据。

**关键词**：计算可复现性; 流行疲劳; 量表; 拟合指数

**Keyword**: Computational reproducibility; Pandemic fatigue; Scale; Fit Index

**1 引言**

**1.1 所选文献信息**

**表 1 文献信息表**

|  |  |  |  |
| --- | --- | --- | --- |
| **1 文献基本信息** | | | |
| 所选文献 | Lilleholt, L., Zettler, I., Betsch, C., & Böhm, R. (2023). Development and validation of the pandemic fatigue scale. *Nature Communications*, *14*(1). <https://doi.org/10.1038/s41467-023-42063->2 | | |
| 数据来源 | 在线实验和丹麦和德国重复横断面数据以及结果、代码和图表来自<https://doi.org/10.17605/OSF.IO/XD463> | | |
| **2 文献选取** | | | |
| 文献主题是否包含不止一篇研究？ | * 是，且包含元分析研究 * 是，但不包含元分析研究 * 否 | 文献此前被其他研究者重复过？ | * 是(https://webofscience.clarivate.cn/wos/alldb/full-record/WOS:001325870700004) * 否 |
| **3 研究假设选取** | | | |
| 重复的研究假设 | (指出检验了原文献哪部分的哪些假设结果，如  **研究一：流行疲劳量表(PFS)的开发和验证**  假设1a探索性因素分析：流行疲劳量表(PFS)的项目确定以及模型假设  假设1b验证性因素分析：流行疲劳量表(PFS)的三种模型在丹麦和德国横断面数据中的验证  假设1c内部一致性：量表的内部一致性可接受  **研究二：流行疲劳量表(PFS)的测量不变性测试**  假设2量表在丹麦和德国对流行疲劳的测量相似  **研究三：疫情疲劳随时间的发展**  假设3和丹麦和德国重复横断面数据中，流行疲劳随着时间的推移而增加  **研究四：流行疲劳的相关矩阵**  假设4丹麦和德国重复横断面数据中，流行疲劳与社会人口统计学（年龄、性别、是否有工作、受教育年限）和性格维度（情绪、外向性、诚实谦逊、宜人性、愤怒、尽责性、对经验的开放性）部分相关  相关  **研究五：流行疲劳和坚持健康保护行为的意图**  假设5丹麦和德国重复横断面数据中，流行疲劳和推荐的健康保护行为，包括遵守物理距离措施的倾向、佩戴口罩、坚持卫生习惯的倾向、让自己了解疫情情况（信息寻求）相关  **研究六：在线实验**  假设6证实上述所有结果  假设6a在线实验高疲劳组比低疲劳组在疫情疲劳程度上有显著差异  假设6b在线实验三组在行为意向上有差异 | | |
|  | * 是，但是没有具体结果 * 否 | 文献共几个实验，重复的研究假设是第几个实验中的？ | 共6个部分，重复了全部部分，除了丹麦面板数据相关的分析（缺少数据，研究3/4/5使用该数据）  有假设可供验证的是第6部分在线实验 |
| **4 数据集选取** | | | |
| 是否采用原始数据？ | * 是（使用清洗后数据） * 否 | 是否对样本量进行修改？ | * 是(说明原因) * 否 |
| 若修改，报告原文样本量大小和修改后的样本量大小 |  | 若修改，报告使用 G-power 计算的修改后的样本量对应的效应量 |  |

**1.2 文献介绍**

一些国家的公众对健康保护行为的依从性逐渐下降的一个可能解释是流行疲劳增加但是却无法直接观察到。然而，观察到的公众对健康保护行为的依从性下降是否可以归因于流行疲劳，一直存在争议。概念化流行疲劳时，首先要更普遍地考虑疲劳的性质，然后阐明什么使流行疲劳独一无二并且在概念上与相关结构不同。要测试新结构的存在和概念化，则拥有一个可靠的测量工具并为所提出的测量工具的结构效度提供证据，包括内容、收敛性和有效性。所以文章制订了流行疲劳量表，并且检验其信效度，在此基础上拟合了理论模型，并且揭示了流行疲劳及其子因素的影响因素与被影响的变量。

**2 方法**

**2.1 样本**

来自the COVID-19 Snapshot MOnitoring (COSMO) 项目的数据，使用了来自丹麦重复横断面调查的 25 波 (2020.10.19–2021.09.20)、丹麦面板调查的9 波 (2020.10.19–2021.06.21) 和德国重复横断面调查的 18 波 (2020.10.27–2021.09.07) 的数据。为控制时间依赖性背景因素（即每百万人中的 COVID-19 新病例和死亡人数、COVID-19 繁殖率和政策严格性）的影响，进一步使用来自(https://ourworldindata.org的 COVID-19 数据) 的数据在此次复现中没有找到,丹麦和德国的重复横断面调查和丹麦面板调查均未使用统计方法预先确定样本量。

**2.2 原研究方法简介**

陈述中省略基础包如dplyr、tidyverse

**2.2.1 量表开发：通过EFA和CFA构建并验证疫情疲劳量表的因子结构。**

1）探索性因子分析（EFA）

本部分使用stats包提取疫情疲劳相关条目：计算峰度偏度（describe、mardia）、计算KMO（KMO，括号内为函数名，下同）和Bartlett球形检验（Bartlett）、碎石图（Scree）、评价碎石图（nFactors）、计算Very Simple Structure and Velicer's MAP（vss）、计算相关矩阵（cor.test）、使用pearson相关系数进行项目缩减（pairs.panels）、基于多水平相关性的探索性因子分析（polychoric）。

paran包评估主成分分析得到的主成分分析和因子分析（fa）得到的因子：平行分析（fa.parallel）、因子提取与旋转，使用了最小残差法（minres）和斜交旋转（oblimin）。

2）验证性因子分析（CFA）

除了上述包，还使用了OpenMx、semPlot包，用于绘制SEM（结构方程模型）路径图。本部分进行了稳健最大似然估计（ML），计算模型拟合指标：比较拟合指数（CFI）、Tucker-Lewis指数（TLI）、近似误差均方根（RMSEA）、标准化均方根误差（SRMR）、绘制路径图。

3）信度与效度分析

本部分使用psych计算了Cronbach's Alpha信度系数（alpha）、计算MacDonald's Omega（reliability函数，针对二阶因子信度）。

**2.2.2 跨组等价性：测量不变性**（Measurement Invariance）**分析确保量表在不同群体间可比。**

分别使用稳健最大似然估计（Robust Maximum Likelihood Estimation）和稳健对角加权最小二乘估计（Robust Diagonally Weighted Least Squares Estimation）来计算构型不变性（模型结构跨组一致）、度量不变性（因子载荷跨组相等）、标量不变性（截距跨组相等），用于检验模型在不同国家（丹麦vs德国）间的不变性，确保测量等价性，并且当部分条目（如PANDEMIC\_FATIGUE\_2\_MB）不满足不变性时，释放该条目的因子载荷进行部分不变性检验。使用cfa和summary函数。

**2.2.3 回归模型探讨相关因素**

计算pearson相关系数（corr.test），然后进行了回归分析（lm）计算不同疲劳条件下的行为意向差异并分析疫情疲劳的影响因素（如年龄、认知风险等）。呈现回归结果（export\_summs、styler），使用普通最小二乘回归（OLS），还使用了Hmisc包

混合效应模型（丹麦面板数据，由于数据缺失，此次未验证）文中使用lme4包

**2.2.4 实验数据验证因果效应**

使用美国实验数据样本进行CFA，验证疲劳对行为意向的因果影响（如高疲劳组行为意向更低）。进行了回归分析（lm），对所有量表变量（情感风险、信任、忧虑、物理距离、卫生、全面流行病疲劳量表、信息疲劳因素、行为疲劳因素）计算信度（alpha），并且对比不同估计方法（ML vs. DWLS），目的是在新样本（如美国）中验证模型的普适性。

**2.2.5 绘图与结果可视化**

使用corrplot绘制相关矩阵热图，使用semplot绘制模型示意图，forestplot包绘制森林图，使用ggplot2绘制雨云图（实验数据分布）。

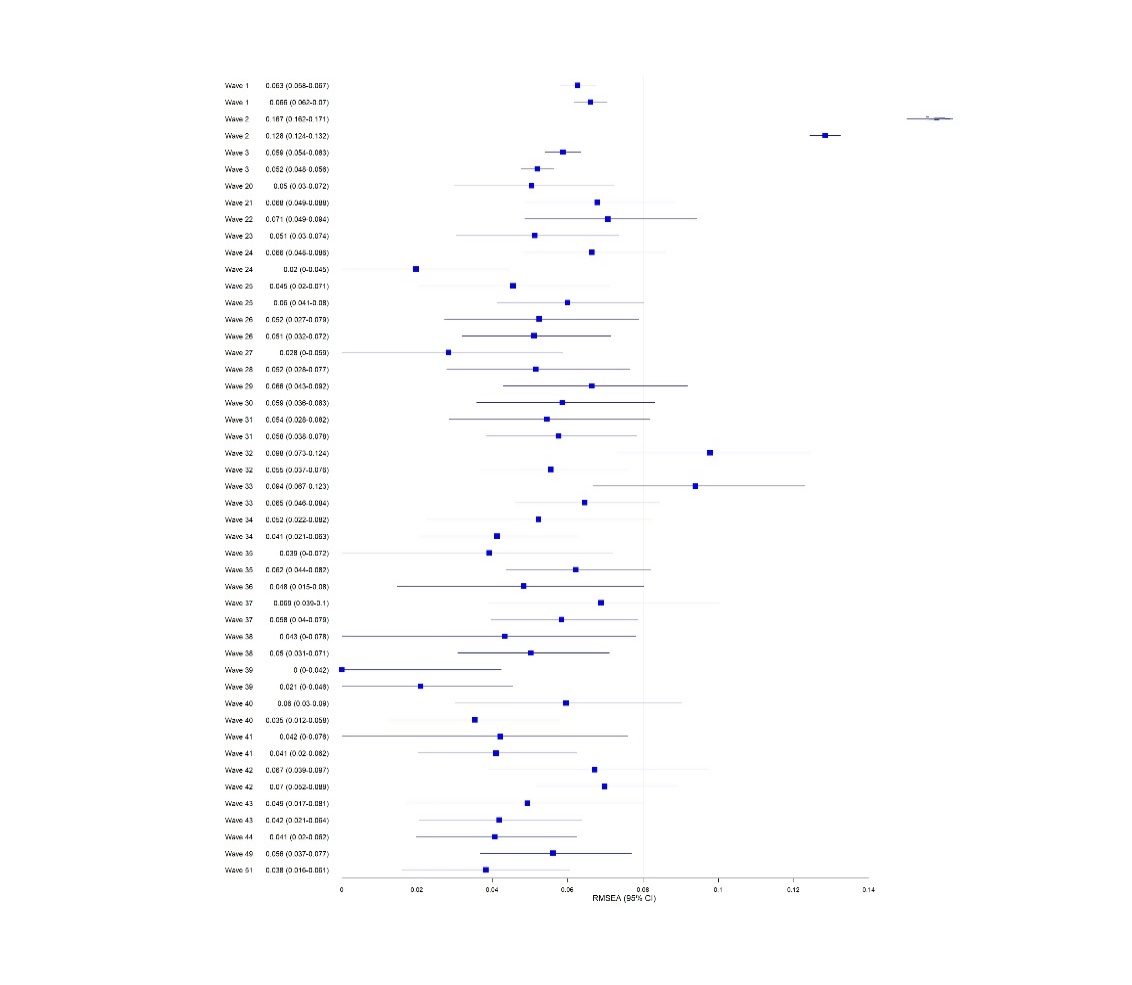
**2.3 重复思路说明**

除了没有丹麦panel面板数据相关的分析之外，所有结果均进行了复现。

因为paran包在分析初期不可用，使用额外的psych包，其中的principal函数用于执行PCA（主成分分析）、fa函数用于执行EFA（因子分析），fa.parallel函数、用于执行平行分析。主成分分析时由于各个包之间函数的区别将estimator换成"ML"，即稳健最大似然估计（处理非正态数据），若数据有缺失，使用"ml"全息极大似然处理缺失值。后来解决了paran包不可用的问题，重新计算后发现用两个包的结果是一样的。

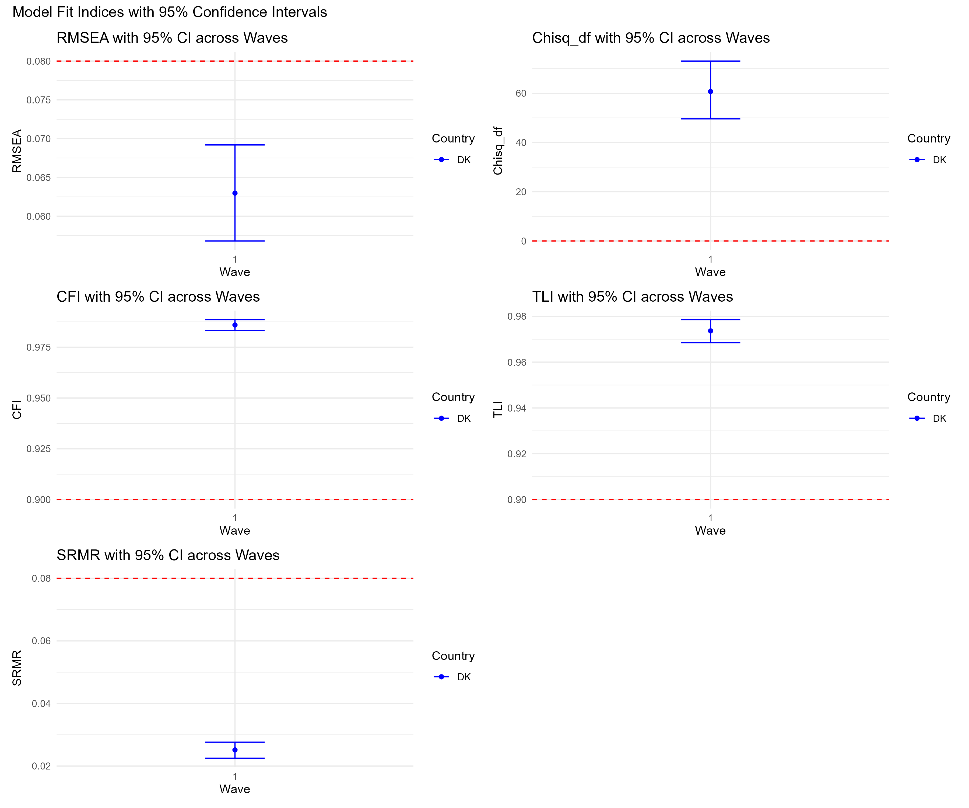
因为OpenMx与semPlot包在分析初期不可用，使用PhantomJS和lavaan包。lavaan 是一个基于 R 语言的结构方程建模（SEM）软件包，具有模型构建与分析、参数估计与模型拟合、图形化展示等功能。后因lavaan包无法满足所需功能，解决OpenMx与semPlot包不可用的问题，重新做了图表。

在核对数据时在RMSEA的数值分布和所报告的范围略有不符，对于每一波的数据应用了forestplot包绘制森林图，内容是每一波数据的RMSEA范围，如图1。对于丹麦的数据，将所有波的比较拟合指数（CFI）、Tucker-Lewis指数（TLI）、近似误差均方根（RMSEA）、标准化均方根误差（SRMR）做了NHST推断性统计，如图2，发现与文章中报告的数据范围一致。



**图1丹麦和德国的RMSEA森林图**

每个wave中两个数据分别来自丹麦和德国



**图2 丹麦模型拟合指数的置信区间**

**3 结果**

**3.1 使用与原文献相同方法的推断性统计**

原文献推断性统计重复的结果详见PPT以及文件夹figure & results，这里只有较为简单的在线实验的方差分析结果呈现，由此汇总成表格：

**表 2 在线实验三组在疫情疲劳程度上的差异**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | 均值差 | *t* | | *p* | Cohen's d | 调整后p值 |
| 原文献  报告结果 | 高疲劳(N = 502) vs 低疲劳(N = 530) | 0.47 | | 5.43 | <0.001 | 0.34 | <0.001 |
| 低疲劳 vs 控制组(N = 552) | -0.21 | | -2.49 | 0.013 | -0.15 | 0.038 |
| 高疲劳 vs 控制组 | 0.26 | | 2.92 | 0.004 | 0.18 | 0.011 |
| 本研究 | 高疲劳(N = 502) vs 低疲劳(N = 530) | 0.47 | | 5.43 | <0.001 | 0.34 | 2.11e-07 |
| 低疲劳 vs 控制组(N = 552) | -0.21 | | -2.49 | 0.013 | -0.15 | 0.03876 |
| 高疲劳 vs 控制组 | 0.26 | | 2.92 | 0.004 | 0.18 | 0.010569 |
| *δ* | | 均为0 | | 均为0 | 均为0 | 均为0 | 均为0 |
| 评级（满分为5） | | 5 | | 5 | 5 | 5 | 5 |

**表 3 在线实验三组在行为意向上的差异**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | 均值差 | *t* | *p* | Cohen's d | 调整后p |
| 原文献  报告结果 | 高疲劳(N = 502) vs 低疲劳(N = 530) | 0.30 | 4.13 | <0.001 | 0.26 | <0.001 |
| 低疲劳 vs 控制组(N = 552) | 0.21 | 2.98 | 0.003 | 0.19 | 0.009 |
| 高疲劳 vs 控制组 | 0.09 | 1.24 | 0.213 | 0.08 | 0.640 |
| 本研究 | 高疲劳(N = 502) vs 低疲劳(N = 530) | 0.30 | 4.13 | <0.001 | 0.26 | 0.000116 |
| 低疲劳 vs 控制组(N = 552) | 0.21 | 2.98 | 0.003 | 0.19 | 0.008949 |
| 高疲劳 vs 控制组 | 0.09 | 1.24 | 0.213 | 0.08 | 0.640200 |
| *δ* | | 均为0 | 均为0 | 均为0 | 均为0 | 均为0 |
| 评级（满分为5） | | 5 | 5 | 5 | 5 | 5 |

**3.2 对原文计算可复现性进行评估**

**3.2.1 使用与原文献相同方法**

报告原文献的值的评级分布、推论的一致情况，整理成表格，如下表所示：

**表 4 结果可复现性的评估表**

|  |  |  |
| --- | --- | --- |
| 可复现性情况 | 数量及占比 | |
| *N* | *%* |
| 完全一致(*δ* = 0%) | 未知 | 100 |
| 偏差较小(0% < *δ* < 10%) |  |  |
| 偏差较小(*δ* ＞ 10%) |  |  |
| 因舍入导致的偏差 |  |  |

**表 5 推论的一致性的评估表(原分析方法)**

|  |  |  |
| --- | --- | --- |
| 推论的一致性 | 数量及占比 | |
| *N \** | *%* |
| 一致 | 未知 | 100 |
| 不一致 |  |  |

**4 讨论**

**4.1 计算可复现性检验结果分析**

结合下表，对原文献进行分析，推测可能导致可复现性检验结果差异的原因。对于重要的原因，逐段进行展开说明。

**表 6计算上（不）可重复的原因分析表**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **可能原因** | | **研究一** | **研究二** | **研究三** | **研究四** | **研究五** | **研究六** | | |
| **数据开放获取特定问题** | 没有提供原始数据 |  |  | √ | √ | √ | |  |
| 没有提供处理后的数据 |  |  | √ | √ | √ | |  |
| 没有提供数据处理过程的描述或代码 |  |  |  |  |  | |  |

**4.2 其他思考**

编制一个量表真难啊。

**参考文献(APA格式)**

Lilleholt, L., Zettler, I., Betsch, C., & Böhm, R. (2023). Development and validation of the pandemic fatigue scale. *Nature Communications*, *14*(1). <https://doi.org/10.1038/s41467-023-42063-2>

Muller, K., & Cohen, J. (1989). Statistical Power Analysis for the Behavioral Sciences. *Technometrics*, *31*(4), 499. https://doi.org/10.2307/1270020

Nakagawa, S., & Schielzeth, H. (2012). A general and simple method for obtaining R2 from generalized linear mixed‐effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Wiley, M., & Wiley, J. F. (2019). Advanced R statistical programming and data models. In *Apress eBooks*. https://doi.org/10.1007/978-1-4842-2872-2

Krekel, C., Swanke, S., De Neve, J., & Fancourt, D. (2020). Are Happier People More Compliant? Global Evidence from Three Large-Scale Surveys During Covid-19 Lockdowns. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3691403

Hoffman, L. (2014). *Longitudinal Analysis: Modeling Within-Person Fluctuation and Change*. http://ci.nii.ac.jp/ncid/BB18162885

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of Within-Person and Between-Person effects in longitudinal models of change. *Carolina Digital Repository (University of North Carolina at Chapel Hill)*. https://doi.org/10.17615/79gv-qb47