

Brodeur, Abel; Jakobsson, Niklas; Moore, Kristjan; Barbarioli, Bruno

Working Paper

A comment on "Validation of the CRAVE-C scale in Chinese adults: a four-study examination of competing motivations for physical activity versus rest"

I4R Discussion Paper Series, No. 324

Provided in Cooperation with:

The Institute for Replication (I4R)

Suggested Citation: Brodeur, Abel; Jakobsson, Niklas; Moore, Kristjan; Barbarioli, Bruno (2026) : A comment on "Validation of the CRAVE-C scale in Chinese adults: a four-study examination of competing motivations for physical activity versus rest", I4R Discussion Paper Series, No. 324, Institute for Replication (I4R), s.l.

This Version is available at:

<https://hdl.handle.net/10419/344063>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

No. 324

DISCUSSION PAPER SERIES

A comment on “Validation of the CRAVE-C scale in Chinese adults: a four-study examination of competing motivations for physical activity versus rest”

Abel Brodeur

Kristjan Moore

Niklas Jakobsson

Bruno Barbarioli

This paper received a response:

Cheng, Z., A. Taylor, M.A. Stults-Kolehmainen, M. Gerber, F. Herold, M. Ross, G. Ash, A.F. Kramer, M. Zhao. 2026. Crave-C post-publication evidence package. *I4R Discussion Paper Series* No. 325. Institute for Replication.

September 2026

I4R DISCUSSION PAPER SERIES

I4R DP No. 324

A comment on “Validation of the CRAVE-C scale in Chinese adults: a four-study examination of competing motivations for physical activity versus rest”

Abel Brodeur^{1,2}, Niklas Jakobsson³, Kristjan Moore, Bruno Barbarioli¹

¹Institute for Replication

²University of Ottawa/Canada

³Karlstad University, Karlstad/Sweden

SEPTEMBER 2026

Any opinions in this paper are those of the author(s) and not those of the Institute for Replication (I4R). Research published in this series may include views on policy, but I4R takes no institutional policy positions.

I4R Discussion Papers are research papers of the Institute for Replication which are widely circulated to promote replications and meta-scientific work in the social sciences. Provided in cooperation with EconStor, a service of the [ZBW – Leibniz Information Centre for Economics](#), and [RWI – Leibniz Institute for Economic Research](#), I4R Discussion Papers are among others listed in RePEc (see IDEAS, EconPapers). Complete list of all I4R DPs - downloadable for free at the I4R website.

I4R Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

Editors

Abel Brodeur
University of Ottawa

Jörg Ankel-Peters
RWI – Leibniz Institute for Economic Research

E-Mail: joerg.peters@rwi-essen.de
RWI – Leibniz Institute for Economic Research

Hohenzollernstraße 1-3
45128 Essen/Germany

www.i4replication.org

ISSN: 2752-1931

A comment on “Validation of the CRAVE-C scale in Chinese adults: a four-study examination of competing motivations for physical activity versus rest”

Abel Brodeur^{1*}, Niklas Jakobsson², Kristjan Moore³, Bruno Barbarioli¹

¹ Institute for Replication, Department of Economics, University of Ottawa

² Karlstad Business School, Karlstad University, Karlstad, Sweden

³ Independent researcher, Reykjavik, Iceland

* Correspondence: Abel Brodeur, abrodeur@uottawa.ca

Original study: Cheng, Taylor, Stults-Kolehmainen et al. (2024)

Frontiers in Psychology, 15:1467949. DOI: 10.3389/fpsyg.2024.1467949

Replication package: OSF DOI 10.17605/OSF.IO/CFPNJ

Abstract

Inspection of the publicly archived replication package for Cheng et al. (2024) reveals several data integrity concerns across three of the four reported studies. In Study 1, 18 survey sessions contain rows sharing identical participant numbers, timestamps, IP addresses, and contact details, indicating repeated submissions by the same respondent or duplicates that were not screened out before analysis. In Study 2, the healthy control group contains 500 observations despite the paper stating 481, and the $N = 500$ mean matches the reported statistic exactly. The clinical group subscale scores contain 17 values exceeding the scale maximum of 50, and all Study 4 item responses fall in the range 1–11 rather than the stated 0–10, suggesting a possible scale mismatch between groups; if unharmonised, the paper's primary finding — that chronic illness patients show higher Rest desire than controls — disappears entirely. The abstract additionally states the wrong direction for the psychiatric Move comparison. Study 3 data are absent from the repository, and Study 4 statistics cannot be reproduced from the released files.

Study 1: Confirmatory Factor Analysis (N = 481)

1.1 Duplicate Survey Submissions

Study 1 collected data via the "Questionnaire Star" (问卷星) platform in April 2023. The dataset contains 481 rows. Matching rows by the combination of submission timestamp and IP address identifies 18 sessions in which more than one row shares both values. These 18 sessions contribute 38 rows in total and 20 extra (redundant) observations, all of which are retained in the analytical sample.

Evidence of repeated submission. For all 18 sessions, the following are consistent across copies: participant number, submission timestamp, completion duration, IP address, email or phone number, and all 13 CRAVE item responses, although a small number of demographic fields differ across copies. Two sessions produced triplicate entries. The affected participant numbers are: 13, 29, 70, 113, 119, 161, 212, 240, 288, 299, 309, 329, 334, 359, 406, 416, 437, 449. Representative examples:

Table 1. Examples of repeated or near-repeated rows in the Study 1 data file.

Rows	Timestamp	Count
9, 10, 11	2023/4/2 14:16:09	3
3, 4	2023/4/2 13:12:42	2
47, 48	2023/4/3 13:18:19	2
120, 121, 122	2023/4/7 19:19:18	3
148, 149	2023/4/8 16:26:58	2

Scale of coincidence. The 13-item CRAVE scale, with items scored 0–10, yields approximately 3.45×10^{13} possible response patterns. Entries sharing the same participant number, IP address, timestamp, and email or phone number are overwhelmingly consistent with repeated submissions by the same respondent or duplicates rather than independent participants who happened to give identical responses.

Study 2: Psychiatric Disorders, Chronic Illness, and Healthy Controls

2.1 Contradictory summary statistics, sample counts, and primary data between sources

The Study 2 data file contains 500 healthy-control Move/Rest score pairs. The paper states that the 481 Study 1 participants were reused as healthy controls. The discrepancy is not attributable to blank or administrative rows: all 500 rows contain numerical scores. Furthermore, the overall means computed from all 500 rows match the values reported in the text to three decimal places.

Table 2. Contradictory summary statistics between sources for Study 2.

Statistic	From Study 2 file (control N=500)	From Study 1 file (control N=481)	Reported in text	Shown in Figure 2
Healthy control Rest mean	25.946	22.474	25.95	23.94
Healthy control Move mean	33.312	33.825	25.39	33.83
Chronic illness Rest mean	31.18	NA	31.18	35.64
Psychiatric illness Rest mean	35.64	NA	35.64	31.18

The analyses therefore, appear to have used N = 500 rather than the stated 481. The source of the additional 19 observations is not described anywhere in the paper or supplementary materials.

Additionally, the values shown in Figure 2 are inconsistent with values quoted in the text and those calculated directly from the data (Table 2). For the controls, the mean values in Figure 2 for Rest and Move contradict both the values reported in the text and those calculated directly from the Study 1 and Study 2 data files. In the Rest panel, it is also apparent that the axis labels for chronic illness and psychiatric illness have been swapped. Additionally, the Move summary statistics for controls quoted in the text mistakenly repeat the statistics given for psychiatric patients – “(25.39 ± 13.47)”.

A more serious issue is observed when comparing data points for the controls provided in the Study 1 and Study 2 data files. Although the two sources show similar summary statistics (Table 2) and Pearson correlations between Rest and Move (Study 1 $r = -0.50$, Study 2 $r = -0.45$), the observations are substantially non-overlapping. Of the 363 unique Rest-Move value pairs in the Study 2 data file, 233 do not appear in the Study 1 data file. Non-overlapping values are also seen for the Rest and Move variables individually: for example, Move scores of 13 and 5 are only recorded in the Study 1 data file, while Move scores of 4, 9, and 10 are only recorded in the Study 2 data file.

2.2 Scale Mismatch Between Healthy Controls and Clinical Groups

Observation. The CRAVE-C items are stated to range from 0 ("not at all") to 10 ("more than ever"). The Study 1 spreadsheet is consistent with this: all 13 CRAVE items have a minimum value of 0 and a maximum of 10. In Study 4, however, the same 13 items have a minimum of 1 (never 0) and a maximum of 11 in the baseline file (see section 4.2). The post-test file provides per-item responses, and these, too, fall within the 1–11 range; the presence of a Move subscale score of 52 (above the 50-point maximum of a five-item, 0–10 scale) confirms that the inappropriate scale was used here as well. Every participant in Study 4 who chose the highest response option received a value of 11 rather than 10, and no one scored 0 on any item. Study 2's healthy control scores derive from the Study 1 data (scale 0–10); the source of the Study 2 clinical group scores is not specified.

Indicator in Study 2 data. Seventeen subscale values in the clinical groups exceed 50 — the maximum possible on a five-item, 0–10 scale — with values reaching 55. Additionally, the minimum score across the 190 observations is 5. These are consistent with the items having been scored on a 1–11 scale (maximum per item = 11; five items $\times$ 11 = 55). None of the 500 healthy control scores in the Study 2 file nor the 481 healthy control scores in the Study 1 file exceeds 50, and 14 values lower than 5 are observed.

Implications if unharmonised. If the clinical group scores in Study 2 were computed from 1–11 responses without subtracting 1 from each item, every clinical subscale score would be inflated by 5 points relative to the healthy controls (5 items $\times$ 1-point offset). Applying that correction changes the between-group comparisons substantially:

Table 3. Statistical consequences of incorrect scaling of clinical Rest and Move scores.

Comparison	Control mean	Clinical (archived)	Clinical (–5 correction)	p (in text)	p (from archived)	p (corrected)
Normal vs Chronic — REST	25.95	31.18	26.18	< 0.001	1.5×10^{-4}	0.866
Normal vs Psych — REST	25.95	35.64	30.64	< 0.001	7.0×10^{-8}	0.008
Normal vs Chronic — MOVE	33.31	35.81	30.81	0.070	0.070	0.069
Normal vs Psych — MOVE	33.31	25.39	20.39	< 0.001	2.3×10^{-5}	9.0×10^{-12}

It is clear that the mean values and *p*-values reported in the text and in Figure 2 correspond to the uncorrected archived values.

Chronic illness Rest finding. If the archived clinical scores were analysed without harmonisation to the 0–10 scale, the finding that "patients with chronic illnesses had a significantly higher 'Rest' desire than healthy controls" would be entirely attributable to the scale offset: the 5.23-point raw difference is virtually identical to the 5-point mechanical inflation, and correcting for it yields $t(560) = -0.17$, $p = 0.866$.

2.3 Abstract Direction Error

The abstract states: *"Patients with psychiatric disorders had a significantly higher 'Move' desire than healthy controls."* The archived data show the reverse: psychiatric Move = 25.39 (SD = 13.47) versus healthy control Move = 33.31 (SD = 10.06), $t(531) = 4.28$, $p < 0.001$, with psychiatric patients having the lower value. The same is true for the corrected values.

The study's other significant comparative claims — that psychiatric patients show higher Rest desire and that chronic illness patients show higher Rest desire — are directionally consistent with the archived data (before any scale correction). This reversal therefore appears to be a transcription error in the abstract rather than a systematic data problem. Regardless of cause, the stated direction of a key finding is incorrect and requires a formal correction.

Study 3: Cardiorespiratory Fitness (N = 83)

The OSF replication package does not contain participant-level data for Study 3. Study 3 therefore cannot be independently reproduced from the released materials, despite the article stating that the datasets are available through the OSF repository.

Study 4: Athletes vs Non-Athletes, HIIT Intervention (N = 42)

4.1 Published Statistics Are Not Reproduced by the Released Data

The article reports baseline and post-HIIT Move and Rest statistics and paired t-tests for athletes (n = 22) and non-athletes (n = 20) in Table 5. Recomputing these from the released spreadsheets yields means that broadly agree with the published values but standard deviations and t-statistics that differ substantially across all four within-group comparisons. It should be noted that the data files are poorly labeled: the coding of nearly all categorical variables is unexplained, and columns for pre- and post-test scores have identical names. This ambiguity can be resolved by calculating summary statistics on the data files and comparing them to values reported in the paper.

Table 4. Inconsistencies in means, standard deviations, and t-statistics reported in the text and those calculated from the data file for Study 4. Values in pre-test and post-test cells represent mean $\pm$ standard deviation. Inconsistencies between values reported in the main text and those calculated from the data files are underlined.

Variable	Group	Pre-test		Post-test		Pre vs post t-statistic	
		Main text	Data files	Main text	Data files	Main text	Data files
Move	Athlete	35.23 $\pm$ 7.58	35.27 $\pm$ 6.35	42.56 $\pm$ 8.56	42.45 $\pm$ 4.09	<u>5.20</u>	<u>3.88</u>
Move	Non-athlete	33.75 $\pm$ 6.05	33.75 $\pm$ 4.02	25.45 $\pm$ 5.01	26.30 $\pm$ 5.72	<u>-3.11</u>	<u>-5.12</u>
Rest	Athlete	32.41 $\pm$ 9.19	32.41 $\pm$ 5.59	24.55 $\pm$ 7.20	24.55 $\pm$ 5.02	<u>-3.55</u>	<u>-5.46</u>
Rest	Non-athlete	25.25 $\pm$ 6.13	25.25 $\pm$ 6.16	34.65 $\pm$ 8.40	34.65 $\pm$ 7.53	<u>6.88</u>	<u>4.06</u>
Age	Athlete	20.36 $\pm$ 0.95	20.41 $\pm$ 0.85	NA	NA	NA	NA

Age	Non-athlete	19. <u>35</u> ± 1.04	19. <u>60</u> ± 0.88	NA	NA	NA	NA
-----	-------------	----------------------	----------------------	----	----	----	----

The consistent underestimation of variance shown in Table 4, as well as minor discrepancies in mean values for sample ages and athlete Move scores, suggests the archived files cover the correct number of participants but that analyses were affected by calculation errors. No alternative scoring formula recovers the article's SDs from the available item-level responses.

Inspection of the scoring formulas embedded in the two spreadsheets reveals that the Move subscale is not computed identically across studies. The Study 1 spreadsheet sums items C1, C2, C6, C9, and C13 ("I want to go out and about"), whereas the Study 4 spreadsheets sum items C1, C2, C5, C6, and C9 ("I want to burn calories"). The Rest formula is consistent across both files (C4, C8, C10, C11, C12). This cross-study inconsistency in subscale composition is not described in the methods.

4.2 Out-of-Range Item Responses

As mentioned in section 2.2, in both the baseline and post-test files, CRAVE item values of exactly 11 appear systematically. At baseline, 16 of 42 participants (38%) have at least one item = 11; at post-test, 22 of 42 (52%). Every out-of-range value is exactly 11 — never 12 or higher — consistent with a response platform presenting a range of "1-11" while the range stated in the text and observed in the Study 1 control data is "0-10." Supporting this, a value of zero is never observed in the 1,092 item measurements in the Study 4 data file, while the Study 1 data file has values of 0 for 343 of 6,253 measurements (13 items × 481 participants, with no missing values).

Replication Package

All code for the above analyses can be found here: <https://osf.io/zydn3/>.