RELIABILITY AND CONVERGENT VALIDITY OF SELF-REPORTED PHYSICAL ACTIVITY QUESTIONNAIRES FOR PEOPLE WITH MENTAL DISORDERS: A SYSTEMATIC REVIEW AND META-ANALYSIS.

Vagner Deuel de Oliveira Tavares [1], Nicole Leite Galvão-Coelho [1,2], Joseph Firth [3], Simon Rosenbaum [4], Brendon Stubbs [5], Lee Smith [6], Davy Vancampfort [7], Felipe Barreto Schuch [8]

[1] Laboratory of Hormone Measurement, Department of Physiology and Behavior, Federal University of Rio Grande do Norte, Natal, Brazil.

[2] National Institute of Science and Technology in Translational Medicine, Brazil.

[3] Division of Psychology and Mental Health, University of Manchester, Manchester, UK

[4] School of Psychiatry, Faculty of Medicine, UNSW Sydney, Australia

[5] 1- Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom; 2- South London and Maudsley NHS Foundation Trust, Denmark Hill, London SE5 8AZ, United Kingdom.

[6] The Cambridge Centre for Sport and Exercise Sciences, Anglia Ruskin University, Cambridge, UK

[7] 1- KU Leuven Department of Rehabilitation Sciences, Leuven, Belgiam; 2- UP KU Leuven, Leuven-Kortenberg, Belgium

[8] Department of Sports Methods and Techniques, Federal University of Santa Maria, Santa Maria, Brazil.

Address: Campus Universitário UFRN - Lagoa Nova, CEP 59.072-970 | Natal - RN, Brasil

Corresponding author: Email: deueltavares@gmail.com

Conflicts of interest: None declared.

Abstract

To examine the reliability and convergent validity of self-reported questionnaires (SRQs) to measure physical activity (PA) **Methods:** Systematic review with meta-analysis. Studies evaluating the validity and reliability of SRQs to assess PA in people with mental disorders (January 20th, 2020). Random-effects meta-analyses were performed pooling 1) test-retest correlations, or 2) the convergent validity between the SRQs and objective measures (e.g.: accelerometry). Associations in *r* values, with the 95% confidence interval (95%CI). Methodological quality was assessed. **Results:** A total of 9 unique studies (N=1,344; 40.5% females) were included. We found a moderate correlation test-retest reliability for PA SRQs in the assessment of vigorous PA [r=0.69 (0.38 – 0.85); p = 0.001], moderate to vigorous PA [r=0.63 (0.25 – 0.84); p = 0.003], moderate PA [r=0.63 (0.39 – 0.79); p = 0.001], and good correlation total PA [r=0.75 (0.37 – 0.92); p = 0.001]. SRQs have moderate correlations with objective measures for moderate to vigorous PA [r=0.25 (0.18 – 0.32); p = 0.0001], total PA [r=0.47 (0.28 – 0.62); p = 0.005], a poor correlation for moderate PA [r=0.18 (0.03 – 0.36); p = 0.047], and no correlation with vigorous PA [r=0.06 (-0.10 – 0.22); p = 0.440]. **Conclusion:** Current evidence indicates that SRQs are reliable over time to assess moderate, vigorous and total PA levels and valid when assessing moderate PA.

Keywords: Self-reported questionnaires; Mental disorders; Reliability; Validity

Background

Mental disorders include a group of heterogeneous disorders, such as major depressive disorder (MDD), bipolar disorder (BD), schizophrenia, and anxiety disorders (AD) that causes significant distress and impairment of personal functioning [1]. People with mental disorders have a 10 to 20 years shortened life expectancy compared to the general population [2].This premature mortality is partially explained by the increased exposure to lifestyle risk factors, such as sedentary behavior and low levels of physical activity (PA) that in turn, are associated with an increased risk of physical morbidities, such as obesity, diabetes and cardiovascular diseases [2].

A robust body of evidence demonstrates that people with mental disorders spend more time in sedentary behavior and less time engaged in PA than people without mental disorders [3,4]. For example, people with severe mental illness are 50% more likely than general population to not adhere to the global recommendation of 150 minutes of moderate to vigorous physical activity per week [3,5], possibly due to lack of motivation and/or pleasure (avolia and anhedonia), seen in different psychiatric symptoms. PA can reduce psychiatric symptoms and is a protective factor against incident depression, anxiety, and schizophrenia [6–12], while reducing time spent sedentary during leisure is associated with lower levels of depression and anxiety [13]. Moreover, PA tends to improve neurobiological and cognitive changes observed in mental disorders [14–16].

Although previous evidence has demonstrated clear cross-sectional and longitudinal associations between PA and mental disorders, most studies rely upon the use of self-reported questionnaires (SRQ) to assess PA [5,17]. SRQs are widely used, cheap and easily administered tools to assess PA. However, it is not clear whether these tools provide accurate and reliable assessments of PA in people with mental disorders. A previous systematic review discussed the psychometric proprieties of SRQs and its use in people with mental disorders [18], but, to the best of our knowledge, no meta-analyses has pooled the reliability (test-retest parameters) or the validity (against an objective measure, such as pedometers and accelerometers) of these measures. Therefore, our objective was to summarize and meta-analyze the reliability and convergent validity of SRQs for assessing PA in people with mental disorders.

Methods

Protocol and registration

The present systematic review followed the PRISMA guidelines [19] and was registered in the International Prospective Register of Systematic Reviews (PROSPERO) on number CRD42020161309.

Eligibility criteria

Studies were included if they: 1) Included clinically diagnosed patients with mental disorders, according to the Diagnostic and Statistical Manual for mental Disorders (DSM-IV or 5) (American Psychiatric Association et al., 2013) or the international classification of diseases (ICD-10) [21] criteria; 2) Assessed the validity of SRQs measuring PA, against objective measures (e.g.: pedometers and accelerometers), or assessed the test-retest reliability, (e.g.: readministered of the questionnaire over a period ranging from 2 to 15 days); 3) Utilized SRQs with known reliability or validity in the general population.

Search

Searchers were performed on PubMed, PsycINFO, SPORTDiscus, and Web of Science until January 20th, 2020. The search strategy used in PubMed was: (depression OR Bipolar and Related Disorders OR schizophrenia OR psychosis) (oxygen consumption OR Vo2 OR oxygen uptake OR aerobic OR cardiovascular OR cardiopulmonary exercise test OR cardiopulmonary fitness OR physical fitness OR fitness OR physical functional performance OR exercise OR leisure activit* OR Physical activity OR muscle strength OR muscle power OR stretching OR walking OR ability) AND (Valid* OR test-retest OR Reliabilit* OR reproducibility). The search strategy for other databases were slightly adapted. The string search for each database can be found in supplementary materials 1.

Data extraction

All data was extracted by two independent reviewers (DT and FS). Data on diagnoses, sample size, sex (% women), age, measure, SRQs measure, objective PA measure, retest period, and correlation values (Pearson's correlation, Intraclass Correlation, or others) were retrieved from individual studies.

Study quality assessment

The risk of bias within-study was assessed using the Quality Appraisal of Reliability Studies (QAREL) for the convergent validity, [22] and the Critical Appraisal Tool (CAT) for reliability [23]. The QAREL is a checklist composed of 11 items. Items 1 and 2 are related to bias and representativeness of subjects and evaluators. Items 3-8 relate to the order that subjects and raters were examined. Item 9 is related to the interval of time between measurements. Item 10 checks whether the test has been applied and interpreted correctly. Item 11 considers the reliability of the statistical analysis. Each item is answered as "yes" (good quality), "no" (poor quality) or "unclear" (insufficient information).

The CAT scale has 13 items. Item 1 is related to the characterization of the subjects. Item 2 is related to the characteristics of the raters. Items 3 and 4 verify the adequacy of the design. Item 5 examines intra-rater reliability. Items 6 and 7 measure aspects of the order of examination and period between the reference standard and index test. Item 8 is related to the adequacy of the interval between measurements. Items from 9 - 11 are examine procedure details. Item 12 is related to the sample composition. Item 13 assesses the statistical details. The CAT scales can be scored as "yes", "no", or "not applicable" for research methods to be able to distinguish between high- and low-quality studies. However, we eliminated items 4,5, 6 and 8, as these items are not applicable to reliability studies.

The QAREL and CAT scales have a maximum score of 110% and 90%, respectively. In each table, there is a column with the final percentage (%) accounting for the methodological quality of each study. Studies were considered high quality if they scored above 45% [24].

Table 1. Description of characteristics of studies in meta-analyses of reliability and convergent validity questionnaires assessing physical activity.

Insert table 1 here…

Statistical Analysis

The meta-analysis was conducted using the transformation of correlational effects (e.g.: *r, rho, ICC)* into Fischer z scores and then reconverting to Spearman correlation (*r)* of each individual study. A random-effects meta-analysis was performed pooling the

converted Spearman correlation for test-retest reliability and for validity, for each PA intensity the amount of energy expended when a person is at rest (light 1.1 to 2.9 times; moderate 3.0 to 5.9 times; vigorous 6.0 or more times and total is any duration may be included in the accumulated total volume of PA) [25], were calculated, whenever sufficient data. Associations were provided in *r* values, together with the 95% confidence interval (95%CI). Correlations for reliability test-retest were classified as excellent ($r = > 0.90$), good ($r = 0.75$ - 0.90), moderate ($r = 0.50$ - 0.75) poor ($r = < 0.50$) [26] and correlations for convergent validity classified as excellent ($r = > 0.75$), good ($r = 0.50$ - 0.75), moderate ($r = 0.25 – 0.50$) poor ($r = < 0.25$) [27]. The I^2 statistic was used to quantify the proportion of the variance in observed effects variation in true effects. An I^2 higher than >50% was considered as an indicative of substantial variation in observed effects [28]. We evaluated the presence of publication bias using the Begg and Mazumdar [29] and Egger tests [30]. In case we detect significant publication bias, the Duval and Tweedie trim and fill technique was applied to adjust and recalculate the new effect [31]. Statistical significance was set at $p<0.05$ and all analyses were performed using Comprehensive Meta-Analysis software, version 3.0 (Biostat. Inc).

Results

Studies selections

The initial search yielded 17,295 results. After the removal of duplicates and exclusion at the title plus abstract level, 17,240 abstracts were considered. After the full-text review stage, 66 studies were considered and a total of 8 unique studies were included in the review (fig. 1). One additional study was found in hand-searches Rosenbaum et al., (2020)[32].

Studies and participants characteristics

Across the nine unique studies [32–40], a total of 1,344 participants were included with median age 43.7 and confidence interval (CI) (34.9 – 51.5) (40.5% women). Studies have been conducted in several countries such as the USA (Lindamer et al., 2008), Belgium [40], the UK [39,42], China [38] and Canada [34,35]. One study was a multicenter study (Rosenbaum et al., 2020). A total of 7 studies assessed the convergent validity of SRQs against objective measures such as accelerometers or Sensewear armbands [32–37,39], and 6 studies assessed the test-retest reliability of the SRQS [32,34,35,37,38]. Among the 9 included

studies, 4 tested the validity or reliability of the International Physical Activity Questionnaire IPAQ-SF [34–36,40].

Figure 1 **–** Flow chart of study selection

Insert figure 1 here…

Systematic Review

Differences between self-reported questionnaires and objective measures of physical activity

Four studies reported on differences between SRQ to objective measures with regards to over/underestimation of PA in SMI. First, Firth et al., (2017) [36] found that whereas SRQ measures placed individuals with schizophrenia as roughly equal to the general population (falling within the 47th percentile rank for total activity), the accelerometry measures in contrast showed that, on average, people with schizophrenia engaged in less PA than 80% of the general population. Second, Duncan et al., (2016) [34] found that SRQ measures in individuals with schizophrenia for moderate to vigorous PA is underestimated at 17 minutes with limits agreement from 145 to 111 minutes. Third, Vancampfort et al., (2017) [40] also found SRQs overestimated activity, with outpatients with first episode psychosis overestimated their physical activity levels by between 35% and 50%. Finally, Soundy et al., (2007) [39] found that SRQ overestimated moderate physical activity at 16.9 minutes with limits agreement -87.5 to 121.3 minutes per day and underestimated vigorous physical activity -10.4 with limits agreement -58.9 to 38.1 minutes per day.

Meta-analyses

Test-retest reliability for physical activity self-reported questionnaires

A moderate correlation for the test-retest reliability of self-reported vigorous PA [r=0.69 (0.38 – 0.85); p = 0.001; I^2= 94.29, fail safe N = 444], no significant *Kendall's rank* correlation coefficient was recorded (τ = 0.00, p = 1.00) indicating significant funnel plot symmetry. A moderate correlation for moderate to vigorous PA [r=0.63 (0.25 – 0.84); p = 0.003; I^2= 95.45], due to the low number of studies (<2) no fail safe N and *Kendall's rank* correlation coefficient was conducted. A moderate correlation for moderate PA [r=0.63 (0.39 – 0.79); p = 0.0001; I^2= 92.56, fail safe N = 32], no significant *Kendall's rank* correlation coefficient was recorded (τ = 0.66, p = 0.29) indicating significant funnel

plot symmetry. A good correlation for total PA [r=0.75 (0.37 – 0.92); p = 0.001; I^2= 92.56, fail safe N = 78], and all results was found (see supplementary figures). The visual inspection of the funnel plot suggests the existence of publication bias in the total PA analysis. However, the Egger (intercept = 8.48, p = 0.09) and the Begg and Mazumdar tests (tau = 0.50, p = 0.30) did not. The visual inspection of the funnel plot, the Egger and the Begg and Mazumdar tests did not suggest the presence of publication bias in the vigorous, moderate to vigorous or moderate PA meta-analyses.

Convergent validity of self-reported questionnaire against objective measures

We did not find a significant correlation between self-reported and objective measures of vigorous PA [r=0.06 (-0.10 – 0.22); p = 0.440; I^2= 0, fail safe N = 00], no significant *Kendall's rank* correlation coefficient was recorded (τ = 0.00, p = 1.00) indicating significant funnel plot symmetry. A moderate correlation between self-reported and objective measures of for moderate to vigorous PA [r=0.25 (0.18 – 0.32); p = 0.0001; I^2= 0], due to the low number of studies (<2) no fail safe N and *Kendall's rank* correlation coefficient was conducted. A poor correlation between self-reported and objective measures of moderate PA [r=0.18 (0.03 – 0.36); p = 0.047; I^2= 12.02, fail safe N = 13], no significant *Kendall's rank* correlation coefficient was recorded (τ = 0.00, p = 1.00) indicating significant funnel plot symmetry. A moderate correlation between self-reported and objective measures of total PA [r=0.47 (0.28 – 0.62); p = 0.0001; I^2= 0, fail safe N = 3], no significant *Kendall's rank* correlation coefficient was recorded (τ = 0.00, p = 1.00) indicating significant funnel plot symmetry, and all results was found (see supplementary figures). The visual inspection of the funnel plot, and the Egger and Begg and Mazumdar tests did not indicate the presence of publication bias in any of the convergent validity analyses.

Study quality assessment

The studies had on average a low risk of bias (mean = 50.6%; range = 40% to 60%) in the QAREL scale for reliability assessment. Only one study was considered of a high risk of bias [37]. All included studies had a low risk of bias in the CAT scale for convergent validity analysis (mean=76.1%; range = 40% to 100%).

Discussion

To the best of our knowledge, this is the first study performing a meta-analysis to evaluate the reliability and convergent validity SRQs to assess PA in people with mental disorders. The findings demonstrate a moderate test-retest reliability for PA SRQs to assess moderate, moderate to vigorous, vigorous and good test-retest reliability for total PA. However, we did find a moderate correlation between SRQs and objective measures when assessing moderate to vigorous PA, for total PA, and a poor correlation to moderate PA. No association was found between subjective and objective vigorous PA assessments.

In a previous narrative synthesis, Soundy et al. (2014), highlighted some limitations, such as insufficient reporting of the ICC values in validation studies. However, Soundy et al. (2014) did not perform a meta-analysis quantifying the reliability of SRQs. In the present study, we found that SRQs are consistent measures for assessing moderate, vigorous and total PA across different time points. However, convergent validity was poor when compared to objective measures.

In the present review, we considered objective measures based on accelerometry as the most accurate and reliable PA assessment that is traditionally used as the gold standard in the literature to test the convergent validity of SRQs [43,44]. Accelerometry, has increased accuracy and reliability, distinguishing between intensity levels when compared to the SRQ's [45,46]. These are validated with equations and are objective measures that have been recommended to serve as a reference criterion in the validation of subjective measures [43,44]. Soundy et al (2014) concluded that there is no precision capability at individual levels of PA. These findings are supported by the present study that did not find convergent validity of SRQ to assess vigorous PA and only a poor validity to assess moderate PA. This limited concordance can be explained by three potential reasons: 1) people with mental illness might have a different perception of PA intensity. For example, previous studies have found that higher negative feelings were linked to a greater physical exhaustion following exercise, and thus leading to an overestimated perception of exercise intensity [47]; 2) people with mental disorders experience cognitive deficits, which might be increasing the risk of recall bias [48]; and 3) the use of SRQs to assess PA might suffer from the social desirability bias, [36,49]. Although of occurs associations of SRQs for poor to moderate related between levels PA and accelerometers in different populations [50,51]. These findings are in accordance with

previous systematic reviews examining the direction and magnitude of differences in PA between SRQ versus objective measures shows that SRQ's generally overestimate PA in people with mental disorders.

Practical implications

Promoting habitual PA in people with mental disorders is a priority, due to its benefits in the physical and mental health of people with mental disorders. Accordingly, precise assessments of PA are necessary for people with mental disorders. Our findings suggest that SRQs may be a reliable option for surveillance and to monitor PA levels across time. According to the results of the present review, it can be suggested that SRQ's with recall periods of seven days have moderate to good test-retest reliability for PA. Therefore, this should be utilized for PA measures, besides reliability, is more one option to cost-effective when the longer-term health benefits and costs are considered [52]. However, they might not reflect the actual time spent in vigorous and moderate PA. For an accurate assessment of vigorous and moderate PA, objective measures seem to be preferable. On the other hand, objective measures may not be a feasible tool for most contexts and for wide use in clinical practice once objective measures are more expensive [53,54]. Moreover, objective measures do not measure the type and the contexts of PA [46,53]. Nonetheless, given the increasingly high rates of smartphone usage across the globe [55], the built-in accelerometers within these devices may present a novel option for widescale objective physical activity monitoring in mental illness. While the accuracy of these devices may not currently meet the 'gold standard' levels of validity, the greater availability and accessibility of smartphones indicates that these may ultimately present a favorable option, particularly for long-term monitoring, compared to self-report measures or existing monitoring devices. Furthermore, as the capabilities of smartphones are constantly improving, advances in these technologies could also present new solutions for objectively measuring type and context of PA, which is missing from current tools. Thus, increased research using these devices, and staying up-to-date with ongoing improvements in technology, will lead to great progress in the field of PA assessments.

Limitations

There are some limitations that should be considered in the present study. First, all studies were conducted in developed countries, and generalization of these findings to low-and middle-income countries is limited, only one study was developed at 40% of countries with low and middle-income [32]. Second, yet we considered accelerometry as

the gold standard to access PA. Accelerometry has its limitations. For example, swimming and strength training are not captured [56], for reasons of being utilized on the hip, no detect upper body movement and do not capture energy expenditure associated with weight support [57,58]. Also, accelerometry do not capture information on PA domain (leisure, transportation, work, household). Despite some limitations, the accelerometry may be a tool to better identify optimal PA levels [59]. Third, the number of studies identified is small, therefore, we may suffer from low statistical power. Fourth, there are not enough studies for pooling the proprieties of SRQs to assess light PA. Lastly, although we have found a substantial variance on the observed effects between studies. However, the small number of studies precludes exploring potential moderators. Further studies should explore whether age, gender, diagnosis, or the PA instrument used can explain the between study variance of effects.

Conclusion

The results indicated that SRQs are reliable to assess moderate, vigorous and total PA levels. However, SRQs have limited validity when compared to objective measures for assessing moderate and vigorous PA in people with mental disorders.

Declarations

- **Competing interests:** The authors declare that they have no competing interests
- **Funding:** None declared.
- **Acknowledgements:** None declared.

References

1. Lawrence D, Kisely S, Pais J. The epidemiology of excess mortality in people with mental illness. *Can J Psychiatry*. 2010;55(12):752-760. doi:10.1177/070674371005501202

2. Firth J, Siddiqi N, Koyanagi A, et al. The Lancet Psychiatry Commission: a blueprint for protecting physical health in people with mental illness. *The Lancet Psychiatry*. 2019;6(8):675-712. doi:10.1016/s2215-0366(19)30132-4

3. Vancampfort D, Firth J, Schuch FB, et al. Sedentary behavior and physical activity levels in people with schizophrenia, bipolar disorder and major depressive disorder: a global systematic review and meta-analysis. *World*

Psychiatry. 2017;16(3):308-315. doi:10.1002/wps.20458

4. Stubbs B, Williams J, Gaughran F, Craig T. How sedentary are people with psychosis? A systematic review and meta-analysis. *Schizophr Res*. 2016;171(1-3):103-109. doi:10.1016/j.schres.2016.01.034

5. Schuch F, Vancampfort D, Firth J, et al. Physical activity and sedentary behavior in people with major depressive disorder: A systematic review and meta-analysis. *J Affect Disord*. 2017;210:139-150. doi:10.1016/j.jad.2016.10.050

6. Firth J, Rosenbaum S, Stubbs B, Gorczynski P, Yung AR, Vancampfort D. Motivating factors and barriers towards exercise in severe mental illness: A systematic review and meta-analysis. *Psychol Med*. 2016;46(14):2869-2881. doi:10.1017/S0033291716001732

7. Rosenbaum S, Tiedemann A, Sherrington C, Curtis J, Ward PB. Physical activity interventions for people with mental illness: A systematic review and meta-analysis. *J Clin Psychiatry*. 2014;75(9):964-974. doi:10.4088/JCP.13r08765

8. Stubbs B, Vancampfort D, Rosenbaum S, et al. An examination of the anxiolytic effects of exercise for people with anxiety and stress-related disorders: A meta-analysis. *Psychiatry Res*. 2017;249:102-108. doi:10.1016/j.psychres.2016.12.020

9. Ashdown-Franks G, Firth J, Carney R, et al. Exercise as Medicine for Mental and Substance Use Disorders: A Meta-review of the Benefits for Neuropsychiatric and Cognitive Outcomes. *Sport Med*. 2019;(0123456789). doi:10.1007/s40279-019-01187-6

10. Brokmeier L, Firth J, Vancampfort D, et al. Does physical activity reduce the risk of psychosis? A systematic review and meta-analysis of prospective studies. *Submitted*. Published online 2019:112675. doi:10.1016/j.psychres.2019.112675

11. Schuch FB, Vancampfort D, Firth J, et al. Physical activity and incident depression: A meta-analysis of prospective cohort studies. *Am J Psychiatry*. 2018;175(7):631-648. doi:10.1176/appi.ajp.2018.17111194

12. Schuch FB, Stubbs B. The Role of Exercise in Preventing and Treating Depression. *Curr Sports Med Rep*. 2019;18(8):299-304. doi:10.1249/jsr.0000000000000620

13. Hallgren M, Nguyen TTD, Owen N, et al. Associations of sedentary behavior in leisure and occupational contexts with symptoms of depression and anxiety. *Prev Med (Baltim)*. 2020;133(February):106021. doi:10.1016/j.ypmed.2020.106021

14. Griffin ÉW, Mullally S, Foley C, Warmington SA, O'Mara SM, Kelly ÁM. Aerobic exercise improves hippocampal function and increases BDNF in the serum of young adult males. *Physiol Behav*. 2011;104(5):934-941. doi:10.1016/j.physbeh.2011.06.005

15. Loprinzi PD, Herod SM, Cardinal BJ, Noakes TD. Physical activity and the brain: A review of this dynamic, bi-directional relationship. *Brain Res*. 2013;1539:95-104. doi:10.1016/j.brainres.2013.10.004

16. Schuch FB, Deslandes AC, Stubbs B, Gosmann NP, Silva CTB da, Fleck MP de A. Neurobiological effects of exercise on major depressive disorder: A systematic review. *Neurosci Biobehav Rev*. 2016;61:1-11. doi:10.1016/j.neubiorev.2015.11.012

17. Vancampfort D, Firth J, Schuch F, et al. Physical activity and sedentary behavior in people with bipolar disorder: A systematic review and meta-analysis. *J Affect Disord*. 2016;201:145-152. doi:10.1016/j.jad.2016.05.020

18. Soundy A, Roskell C, Stubbs B, Vancampfort D. Selection, Use and Psychometric Properties of Physical Activity Measures to Assess Individuals with Severe Mental Illness: A Narrative Synthesis. *Arch Psychiatr Nurs*. 2014;28(2):135-151. doi:10.1016/j.apnu.2013.12.002

19. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ*. 2009;339(7716):332-336. doi:10.1136/bmj.b2535

20. America Psychiatric Association et al. Diagnostic and statistical manual of mental disorders. *Am Psychiatr Pub*. Published online 2013. doi:10.1176/appi.books.9780890425596

21. World Health Organization. International Statistical Classification of Diseases and Related Health Problems. *World Heal Organ*. 2016;10:xii, 362 p. doi:10.1002/1520-6505(2000)9:5<201::AID-EVAN2>3.3.CO;2-P

22. Lucas NP, Macaskill P, Irwig L, Bogduk N. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). *J Clin Epidemiol*. 2010;63(8):854-861. doi:10.1016/j.jclinepi.2009.10.002

23. Brink Y, Louw QA. Clinical instruments: Reliability and validity critical appraisal. *J Eval Clin Pract*. 2011;18(6):1126-1132. doi:10.1111/j.1365-2753.2011.01707.x

24. Muñoz-Bermejo L, Pérez-Gómez J, Manzano F, Collado-Mateo D, Villafaina S, Adsuar JC. Reliability of isokinetic knee strength measurements in children: A systematic review and meta-analysis. *PLoS One*. 2019;14(12):1-15. doi:10.1371/journal.pone.0226274

25. Piercy KL, Troiano RP, Ballard RM, et al. The physical activity guidelines for Americans. *JAMA - J Am Med Assoc*. 2018;320(19):2020-2028. doi:10.1001/jama.2018.14854

26. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. 2016;15(2):155-163. doi:10.1016/j.jcm.2016.02.012

27. Portney L, Watkins M. Foundations of clinical research: Applications to practice. *New Jersey Prentice Hall*. Published online 2009. http://www.r2library.com/public/ResourceDetail.aspx?authCheck=true&resid=765.

28. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions. *Cochrane Collab*. Published online 2011. Available from www.handbook.cochrane.org

29. Begg CB, Mazumdar M. Operating Characteristics of a Rank Correlation Test for Publication Bias. *Biometrics*. 1994;50(4):1088. doi:10.2307/2533446

30. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *Br Med J*. 1997;315(7109):629-634. doi:10.1136/bmj.316.7129.469

31. Duval S, Tweedie R. Trim and Fill: A Simple Funnel-Plot-Based Method. *Biometrics*. 2000;56(June):455-463.

32. Rosenbaum, Simon and Morell, Rachel and Abdel-Baki, Amal and Ahamadpanah, Mohammad and Anilkumar, Thekkethayyil V and Baie, Lara and Bauman, Adrian and Bender, Stefan and Han, Justin Boyan and Brand, Serge and Bratland-Sanda, Solfrid and Bueno-Antequera, PB. Assessing physical activity in people with mental illness: 23- country reliability and validity of the Simple Physical Activity Questionnaire (SIMPAQ). *BMC Psychiatry*. Published online 2020:1-37. doi:10.21203/rs.2.12937/v2

33. Dubbert PM, White JD, Grothe KB, Jile JO, Kirchner KA. Physical Activity in Patients who are Severely Mentally Ill : Feasibility of Assessment for Clinical and Research Applications. 2006;20(5):205-209. doi:10.1016/j.apnu.2006.04.002

34. Duncan MJ, Arbour-nicitopoulos K, Subramanieapillai M, Remington G, Faulkner G. Revisiting the International Physical Activity Questionnaire (IPAQ): Assessing physical activity among individuals with schizophrenia. *Schizophr Res*. Published online 2016:6-11. doi:10.1016/j.schres.2016.09.010

35. Faulkner G, Cohn T, Remington G. Validation of a physical activity assessment tool for individuals with schizophrenia. 2006;82:225-231. doi:10.1016/j.schres.2005.10.020

36. Firth J, Stubbs B, Vancampfort D, et al. The validity and value of self-reported physical activity and accelerometry in people with schizophrenia: A population-scale study of the UK biobank. *Schizophr Bull*. 2017;44(6):1293-1300. doi:10.1093/schbul/sbx149

37. Lindamer LA, Mckibbin C, Norman GJ, et al. Assessment of physical activity in middle-aged and older adults with schizophrenia. 2008;104:294-301. doi:10.1016/j.schres.2008.04.040

38. Ma W, Chiang L, Yen W, Huang L, Tsai T. International Journal of Nursing Studies 3-Month Physical Activity Checklist : Development and validation with Taiwanese adults with mental illness. *Int J Nurs Stud*. 2011;48(12):1517-1521. doi:10.1016/j.ijnurstu.2011.06.005

39. Soundy A, Taylor A, Faulkner G, Rowlands A. Psychometric Properties of the 7-Day Physical Activity Recall Questionnaire in Individuals with Severe Mental Illness. 2007;21(6):309-316. doi:10.1016/j.apnu.2007.03.001

40. Vancampfort D, Hert M De, Simon IM, Stubbs B, Damme T Van, Probst M. Validity and correlates of the International Physical Activity Questionnaire in first-episode psychosis. 2017;(February):1-6. doi:10.1111/eip.12521

41. Andorko ND, Rakhshan-rouhakhtar P, Hinkle C, et al. Assessing validity of retrospective recall of physical activity in individuals with psychosis-like experiences. 2019;273(December 2018):211-217. doi:10.1016/j.psychres.2019.01.029

42. Firth J, Stubbs B, Vancampfort D, Schuch FB, Rosenbaum S. The Validity and Value of Self-reported Physical Activity and Accelerometry in People With Schizophrenia : A Population-Scale Study of the UK Biobank. Published online 2017:1-8. doi:10.1093/schbul/sbx149

43. Sirard JR, Pate RR. Physical activity assessment in children and adolescents. *Sport Med*. 2001;31(6):439-454. doi:10.2165/00007256-200131060-00004

44. Dollman J, Okely AD, Hardy L, Timperio A, Salmon J, Hills AP. A hitchhiker's guide to assessing young people's physical activity: Deciding what method to use. *J Sci Med Sport*. 2009;12(5):518-525. doi:10.1016/j.jsams.2008.09.007

45. Tudor-Locke CE, Myers AM. Challenges and opportunities for measuring physical activity in sedentary adults. *Sport Med*. 2001;31(2):91-100. doi:10.2165/00007256-200131020-00002

46. Harris TJ, Owen CG, Victor CR, Adams R, Ekelund U, Cook DG. A comparison of questionnaire, accelerometer, and pedometer: Measures in older people. *Med Sci Sports Exerc*. 2009;41(7):1392-1402. doi:10.1249/MSS.0b013e31819b3533

47. Blanchard CM, Rodgers WM, Courneya KS, Spence JC. Moderators of the exercise/feeling-state relationship: The influence of self-efficacy, baseline, and in-task feeling states at moderate- and high-intensity exercise. *J Appl Soc Psychol*. 2002;32(7):1379-1395. doi:10.1111/j.1559-1816.2002.tb01442.x

48. East-Richard, C., R. -Mercier, A., Nadeau, D., & Cellard C. Transdiagnostic neurocognitive deficits in psychiatry: A review of meta-analyses. *Can Psychol Can*. Published online 2019. doi:10.1037/cap0000196

49. Sallis JF, Saelens BE. Assessment of physical activity by self-report: Status,

limitations, and future directions. *Res Q Exerc Sport*. 2000;71(July):1-14. doi:10.1080/02701367.2000.11082780

50. Craig CL, Marshall AL, Sjöström M, et al. International physical activity questionnaire: 12-Country reliability and validity. *Med Sci Sports Exerc*. 2003;35(8):1381-1395. doi:10.1249/01.MSS.0000078924.61453.FB

51. Kim Y, Park I, Kang M. Convergent validity of the International Physical Activity Questionnaire (IPAQ): Meta-analysis. *Public Health Nutr*. 2013;16(3):440-452. doi:10.1017/S1368980012002996

52. GC V, Wilson ECF, Suhrcke M, Hardeman W, Sutton S. Are brief interventions to increase physical activity cost-effective? A systematic review. *Br J Sports Med*. 2016;50(7):408-417. doi:10.1136/bjsports-2015-094655

53. Ferrari P, Friedenreich C, Matthews CE. The role of measurement error in estimating levels of physical activity. *Am J Epidemiol*. 2007;166(7):832-840. doi:10.1093/aje/kwm148

54. Skender S, Schrotz-King P, Böhm J, et al. Repeat physical activity measurement by accelerometry among colorectal cancer patients - Feasibility and minimal number of days of monitoring. *BMC Res Notes*. 2015;8(1):1-8. doi:10.1186/s13104-015-1168-y

55. Joseph Firth, Torous J, Stubbs B, et al. The “online brain”: how the Internet may be changing our cognition Joseph. *World Psychiatry*. 2019;18(2):119-129. doi:10.1002/wps.20617

56. Shepard LA. Evaluating Test Validity. *Rev Res Educ*. 1993;19(1):405-450. doi:10.3102/0091732X019001405

57. Jakicic JM, Winters C, Lagally K, Ho J, Robertson RJ, Wing RR. The accuracy of the TriTrac-R3D accelerometer to estimate energy expenditure. *Med Sci Sport Exerc*. 1999;31(5):747-754. doi:10.1097/00005768-199905000-00020

58. D H, K M, C B, E D, P F. Validity of accelerometry for the assessment of moderate intensity physical activity in the field. *Med Sci Sports Exerc*. 2000;32(9):S442-S449.

59. Atienza AA, Moser RP, Perna F, et al. Self-Reported and Objectively Measured

Activity Related to Biomarkers Using NHANES. *Med Sci Sport Exerc*. 2011;43(5):815-821. doi:IO.1249/MSS.ObO13e3181fdfc32

60. Taylor AH, Ussher MH, Guy Faulkner. The acute effects of exercise on cigarette cravings, withdrawal symptoms, affect and smoking behaviour: a systematic review. *Addict J (Abingdon, England)*. Published online 2007. doi:10.1111/j.1360-0443.2006.01739.x

Table 1.

Author	Diagnoses	Participants (n)	Women%	Age (yrs)	Questionnaire	Time Period Measure	Objective Measure	Time Objective Measure	Retest Period
Duncan et al, 2016	Schizophrenia	113	40	41	IPAQ-SF	Last 7 days	Accelerometer (wGT3X+)	7 days	3 weeks
Lindamer et al., 2008	Schizophrenia	54	40.7	50	YPAS	Typical week	Accelerometer (Actigraph)	7 days	1 week
Ma et al, 2011	Multiple mental disorders	60	-	36.3	3MPAC	Last 7 days	-	-	2 weeks
Dubbert et al, 2006	Multiple disorders	20	-	46.4	PWA	Last 7 days	Accelerometer (RT3)	3 days	-
Soundy, 2007	Multiple mental disorders	14	28.7	52.9	7DR	Last 7 days	Accelerometer (RT3)	7 days	2 weeks
Faulkner et al, 2006	Schizophrenia	35	27	39.6	IPAQ-SF	Last 7 days	Accelerometer (RT3)	7 days	-
Firth et al, 2017	Schizophrenia	39	45	54.1	IPAQ-SF	Last 7 days	Accelerometer (Activity AX3)	7 days	-
Vancampfort et al., 2017	Psychosis-Like	19	21	24.4	IPAQ-SF	Last 7 days	SWA	5 days	-
(Rosenbaum et al., 2020)	Multiple mental disorders	1,010	44	?	SIMPAQ	Last 7 days	Accelerometer (Actigraph GT3x)	7 days	1 week

Description of characteristics of studies in meta-analyses of reliability and convergent validity questionnaires assessing physical activity.

Note: Ypas: Yale Physical Activity Scale; 3MPAC: 3-Month Physical Activity Checklist; PWA: Past week activity; IPAQ-SF: International physical activity questionnaire short form; SWA: Sensewear Armband; 7DR: 7-Day Physical Activity Recall.

Figure 1 – Flow chart of study selection