RELIABILITY AND VALIDITY OF PHYSICAL FITNESS TESTS IN PEOPLE WITH MENTAL DISORDERS: A SYSTEMATIC REVIEW AND META-ANALYSIS.
Abstract

Background: Several tests are available to assess the different components of physical fitness, including cardiorespiratory fitness, muscular strength, and flexibility. However, the reliability and validity of physical fitness tests in people with mental disorders has not been meta-analysed. Aims: To examine the reliability, concurrent and convergent validity of physical fitness tests in people with mental disorders. Methods: Studies evaluating the reliability, concurrent and convergent validity of physical fitness tests in people with mental disorders were searched from major databases until January 20th, 2020. Random-effects meta-analyses were performed pooling (1) reliability: test-retest correlations at two-time points; (2) convergent validity between submaximal tests and maximal protocols; or (3) concurrent validity between two submaximal tests. Associations are presented using $r$-values and 95% confidence intervals. Methodological quality was assessed using the Quality Appraisal of Reliability Studies and the Critical Appraisal Tool. Results: A total of 11 studies (N=504; 34% females) were included. Reliability of the fitness tests, produced $r$-values ranging from moderate [balance test-EUROFIT; $r=0.75 (0.60 – 0.85), p=0.0001$] to very strong [explosive leg power EUROFIT; $r=0.96 (0.93 – 0.97, p=0.0001$]. Convergent validity between the six-minute walk test (6MWT) and submaximal cardiorespiratory tests was moderate [0.57 (0.26–0.77); $p=.0001$]. Concurrent validity between the two-minute walk test (2MWT) and 6MWT [$r= 0.86 (0.39–0.97); p=.0004$] was strong. Conclusion: The present study demonstrates that physical fitness tests are reliable and valid in people with mental disorders.

Keywords: Exercise capacity; aerobic; mental disorders;
Introduction

The life expectancy of people with mental illness is shortened by more than ten years compared to the general population (Walker et al., 2015). A significant proportion of this gap in life expectancy is due to the increased prevalence of cardiometabolic diseases (Correll et al., 2017; Firth et al., 2019a; Vancampfort et al., 2016a, 2015d). Although genetic factors and shared pathophysiological mechanisms contribute, treatment-related factors and unhealthy lifestyle habits (i.e. low physical activity) also play a prominent role in this increased risk (Firth et al., 2019a).

Physical activity is a key determinant of physical fitness. Physical fitness comprises multiple components, including cardiorespiratory fitness, muscle endurance, muscular strength, body composition, balance and flexibility (Caspersen et al., 1985; Vanhees et al., 2005). In the general population, some components of physical fitness, such as cardiorespiratory capacity and handgrip strength, are independent predictors of all-cause mortality (Firth et al., 2019b; García-Hermoso et al., 2018; Kodama, 2009).

Previous research demonstrated that people with mental disorders have lower physical fitness levels, although it can be improved with physical activity (Vancampfort et al., 2017). Also, there is compelling evidence that, in this vulnerable population, higher levels of physical fitness are associated with a better quality of life and less symptom severity (Vancampfort et al., 2013a, 2013b).

Testing physical fitness is necessary to: (1) determine the exercise workload needed to provide health benefits and minimize the risk of injury, (2) evaluate physical functioning, and (3) monitor the progress of physical activity interventions (Meyer and Broocks, 2000; Stubbs et al., 2018). Previous evidence demonstrates the validity of many physical fitness tests in the general population (Bennett et al., 2016; Scalco et al., 2018). Since people with mental disorders often experience fluctuations in their energy and motivation levels (Taylor et al., 2003), or may avoid or abandon vigorous exercise fearing an exacerbation of somatic and anxiety symptoms (Vliet et al., 1999), it is important to examine the validity and reliability of physical fitness tests in this population.

To the best of our knowledge, no systematic review or meta-analysis has summarized the validity and reliability of physical fitness tests in people with mental disorders. Findings from this meta-analysis could assist exercise professionals in selecting the most reliable and valid tests to apply in this population. This study aimed to review the literature that evaluated the concurrent and convergent validity and reliability of physical fitness tests. 
of fitness tests, such as cardiorespiratory fitness, exercise functional capacity, muscular strength, balance, flexibility, speed and agility in people with mental disorders.

Methods

Protocol and registration

The present systematic review followed the PRISMA guidelines (Moher et al., 2009). The study protocol is available upon request.

Eligibility criteria

Studies were included if they: 1) included clinically diagnosed patients with mental disorders, such as major depressive disorder, bipolar disorder, anxiety or stress-related disorders, psychotic disorders or substance abuse/dependence, according to the Diagnostic and Statistical Manual for mental Disorders (DSM-V or 5) (Hasin et al., 2013) or the international classification of diseases (ICD-10) (World Health Organization, 2016); (2) assessed the test-retest correlation (reliability), or the correlation between a submaximal cardiorespiratory test against a maximal cardiorespiratory test (convergent validity), or other submaximal validated test (concurrent validity).

Searches

Relevant articles were searched in the following databases: PubMed, PsycINFO, SportDiscus and Web of Science, from inception to January 20th 2020, using the following search terms: (depression OR bipolar and related disorders OR schizophrenia OR psychosis) AND (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 searches were slightly adapted depending on the specific requirements of the database. The string search for each database can be found in the supplementary material 1.

The screening for relevant articles was performed in two steps: First, the titles and abstracts of relevant articles were read. Second, potentially relevant articles, on the basis of their title and abstract, were read in full according to the inclusion and exclusion
criteria. The two steps were performed by two independent reviewers (FS and DT) until consensus. No third reviewer was required.

**Data extraction**

All data were extracted by two (FS and DT) independent reviewers. Information about, volunteers; diagnoses, gender and age, measures; test utilized, retest period and author were extracted.

**Study quality assessment**

The within-study risk of bias was assessed using the Quality Appraisal of Reliability Studies (QAREL) for the convergent validity (Lucas et al., 2010). The QAREL is a checklist comprised of 11 items and has a maximum score of 110%. These questions can be answered as “yes” (good quality), “no” (poor quality) or “unclear” (insufficient information). For reliability risk of bias, we used the Critical Appraisal Tool (CAT) (Brink and Louw, 2011). The CAT scale has 13 items and we eliminated items 4, 5, 6 and 8, as these items do not apply to reliability studies (Muñoz-Bermejo et al., 2019). 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 and have a maximum score of 90%.

In each table, there is a column with the final percentage (%) representing the methodological quality of each study. Studies were considered high quality if they scored above 45% (Muñoz-Bermejo et al., 2019).

**Statistical Analysis**

Meta-analyses were performed for assessing the reliability, convergent validity and concurrent validity of the physical fitness tests. Correlations (r) and sample size data from each study were pooled together to calculate reliability and validity parameters. To determine test-retest reliability, meta-analyses were performed for: 1) each fitness tests, separately; 2) functional exercise capacity tests (walking tests), and 3) submaximal cardiovascular capacity (bike tests). A convergent validity meta-analysis was performed pooling data of studies associating submaximal tests and maximal protocols for functional exercise capacity tests. To determine concurrent validity, a meta-analysis was performed
on the correlation between two submaximal tests (2-min and 6-min walk tests). For all analysis, the correlational values \(r\), \(\rho\), \(ICC\) were converted into Fischer z scores and then, Fischer z scores were reconverted into Spearman correlation \(r\). Outcomes were provided in \(r\) values, together with the 95% confidence interval (95%CI). Correlations were classified as: 0 to 0.19, very weak; 0.2 to 0.39, weak; 0.40 to 0.59, moderate; 0.6 to 0.79, strong; 0.8 to 0.9, very strong; and 1.0, perfect correlation (Walker & Almond, 2010). The \(Q\) and \(I^2\) statistic were used to assess and quantify the heterogeneity, respectively. Scores of \(<25\%\), 25-50% and \(>50\%\) indicated low, moderate and high heterogeneity, respectively (Higgins et al., 2003). Analyses were run with random-effect (DerSimonian and Laird, 1986). All analyses were performed using Comprehensive Meta-Analysis software (version 3) (Borenstein et al., 2013).

### Results

#### Studies selection

The initial search yielded 17,295 results. After removal of duplicates and exclusion at the title plus abstract level, 17,223 abstracts were considered. At the full-text review stage, 72 studies were considered. Finally, a total of 11 studies were included. The flowchart of study selection can be seen in figure 1.

#### Insert figure 1 here

#### Studies and participants characteristics

Across eleven studies, a total of 504 (34% women) participants, with a median age of 37.4 (95% CI: 34.1 – 42.2 years) were included. Studies were conducted in several countries, including Belgium (Knapen et al., 2003; Vancampfort et al., 2010, 2011, 2014, 2015a, 2015c, 2016c, 2019b), Uganda (Vancampfort et al., 2019c, 2019a) and Portugal (Gomes et al., 2016). A total of 5 studies assessed the test-retest reliability of functional walk tests, which are a measure-of-proxy for cardiorespiratory fitness (Gomes et al., 2016; Vancampfort et al., 2011, 2016c, 2019c, 2019a). Three studies investigated the test-retest reliability of fitness tests, measured by the EUROFIT test, including flamingo balance (FBA) (balance), plate tapping (PLT) (speed), sit-and-reach (SAR) (flexibility), standing broad jump (SBJ) (explosive strength), handgrip strength (HGR), sit-ups (SUP)
(muscular endurance), the 5m shuttle run (SHR) (speed and agility) (Vancampfort et al., 2019b, 2015c, 2011), 3 studies measured the 6MWT (Gomes et al., 2016; Vancampfort et al., 2015a, 2010) and 2 studies measured the 2MWT (Vancampfort et al., 2019a, 2019c). Two studies assessed the test-retest reliability of a submaximal cardiorespiratory fitness tests, one study tested the Franz test (Physical Work Capacity, PWC) (Knapen et al., 2003) and one study the Åstrand-Rhyming test (Vancampfort et al., 2014). Two studies assessed the convergent validity of the 6MWT (Gomes et al., 2016; Vancampfort et al., 2015a) and 2 studies for concurrent validity of the 2MWT against the 6MWT (Vancampfort et al., 2019c, 2019a). The details of all studies are shown in Table 1.

**Meta-analysis**

**Test-Retest Reliability of Fitness Tests**

The test-retest reliability of each individual test ranged from moderate [balance test-EUROFIT, \( r=0.75 \) (95%CI: 0.60 – 0.85), \( p>0.0001 \)], to very strong [explosive leg power test-EUROFIT, \( r=0.96 \) (95%CI: 0.93 – 0.97), \( p>0.0001 \)]. Results of individual tests are shown in Table 2.

**Test-Retest Reliability of Functional Exercise Tests**

We found a very strong test-retest reliability for functional exercise capacity tests \( r=0.95 \) (95%CI: 0.94 – 0.97), \( p<0.0001 \) with a high heterogeneity \( (I^2= 57.16\%) \). Data are presented in a forest plot, see Figure 2.

**Test-Retest Reliability of Submaximal Cardiorespiratory Fitness Tests**

We found a very strong test-retest reliability of submaximal cardiorespiratory tests \( r=0.85 \) (95%CI: 0.77 – 0.91), \( p<0.0001 \). The analysis presented a high heterogeneity \( (I^2= 58.35\%) \). Data are presented in a forest plot, see Figure 3.
Convergent Validity of the 6MWT test

We found a moderate convergent validity of the 6MWT \( r=0.57 \) (95%CI: 0.26 – 0.77), \( p=0.0001 \). There was no heterogeneity (I\(^2\) = 0). Data are presented in a forest plot, see Figure 4.

Insert figure 4 here

Concurrent Validity of the 2MWT against the 6MWT

We found a strong correlation between the 6MWT and 2MWT \( r=0.86 \) (95%CI: 0.39 – 0.97) \( p=0.0004 \). The analysis presented high heterogeneity (I\(^2\) = 94.60%). Data are presented in a forest plot, see Figure 5.

Insert figure 5 here

Discussion

To our knowledge, the present meta-analysis is the first to evaluate the reliability and concurrent and convergent validity of physical fitness tests in people with mental disorders. Our main findings are: (1) physical fitness tests can be reliably used in people with a mental disorder; (2) the 6MWT has a moderate convergent validity to estimate cardiorespiratory capacity, and (3) the 2MWT has a strong concurrent validity with the 6MWT.

Mental health disorders are associated with a high prevalence of somatic comorbidity, particularly cardiovascular disease (Batelaan et al., 2016). It is now recognized that a complete patient assessment should involve screening physical activity habits (Correll et al., 2017), and with this, implementing appropriate tests of physical fitness to develop optimal exercise prescriptions (Hallgren et al., 2017).

Of the included studies, the EUROFIT was the only battery used to assess a range of physical fitness components, such as cardiorespiratory fitness, muscular strength, balance, flexibility, speed and agility in this population. While people with mental
disorders present fluctuations in their energy and motivation levels (Firth et al., 2016), fear of somatic symptom exacerbation when performing exercise (Ho et al., 2018), muscular weakness and other abnormal myopathic and electrophysiological features (Reznik et al., 2000), as well as, cognitive symptoms that could potentially affect their performance in such tests (Firth et al., 2018a), the EUROFIT battery has demonstrated value to provide reliable assessments in people with mental disorders. The EUROFIT battery is a safe, and inexpensive protocol that can be implemented in low-resource settings.

We have found evidence that the 6MWT is a reliable test, with a moderate correlation with submaximal or maximal cardiorespiratory tests in people with mental disorders. The 6MWT walk test does not require equipment, such as an ergometer, it is relatively low-cost and easier to administer than cycle or treadmill tests (Solway et al., 2001). Therefore, these tests can be easily be implemented in clinical practice to assess functional exercise capacity levels and to estimate cardiorespiratory fitness. Moreover, the 2MWT showed high convergent validity when compared to the 2MWT. Therefore, in settings with limited staff or time constraints, the 2MWT can be used reliably.

Physical fitness components are associated with better cardiometabolic functioning (Gorczynski and Faulkner, 2010), lesser severity of psychiatric symptoms (Firth et al., 2017), increased quality of life (Vancampfort et al., 2015b), preservation of brain health (Firth et al., 2018b). Given these important health benefits, and the relative ease with which fitness can reliably be assessed by exercise professionals, we argue that physical fitness tests should be incorporated in routine mental health care. Their inclusion may also help to provide a more holistic and comprehensive assessment of the patient and their level of physical functioning.

Some limitations are acknowledged. The number of studies identified was small, therefore, our analyses of test validity may suffer from low statistical power. Also, due the small number of studies, the heterogeneity could not be explored. The EUROFIT battery was the only test examined for most components, and it is not possible to extend these results to other physical fitness tests. Also, worth mentioning that medications may impaired the performance of physical tests regardless of the time of day it is ingested. However, antipsychotic medication can provide unpleasant side-effects, such as feeling of fatigue, and may favour poor performance in the physical test by providing body balance, and lower muscular strength (Perez-Cruzado et al., 2017; Vancampfort et al.,
Lastly, we were not able to explore whether age, sex, body mass index, diagnosis and symptom severity impact the use of such tests.

**Conclusion**

The present review indicates that physical fitness tests are reliable and valid to assess physical fitness components in people with mental disorders. Maximal aerobic exercise testing is the gold standard to assess a person’s cardiorespiratory fitness level, the 6MWT and the 2MWT can be used as a measure-of-proxy for assessing cardiorespiratory fitness in people with mental disorders, especially in low resource settings.

**References**


https://doi.org/10.1016/j.neuroimage.2017.11.007


https://doi.org/10.1136/bmj.327.7414.557


Association (EPA), supported by the International Organization of Physical Therapists in Mental . Eur. Psychiatry 54, 124–144.
https://doi.org/10.1016/j.eurpsy.2018.07.004

https://doi.org/10.1097/01.PSY.0000088580.28749.7F

https://doi.org/10.1002/pri.1821

https://doi.org/10.1016/j.psychres.2015.09.039

https://doi.org/10.1002/wps.20309

https://doi.org/10.1016/j.psychres.2014.08.049

https://doi.org/10.1111/inm.12145

https://doi.org/10.1016/j.psychres.2018.11.052

Vancampfort, D., Kimbowa, S., Basangwa, D., Smith, L., Stubbs, B., Damme, T. Van,


components in people with schizophrenia and related psychotic disorders, bipolar disorder and major depressive disorder: A systematic review and meta-analysis. World Psychiatry 14, 339–347. https://doi.org/10.1002/wps.20252


Table 1. General characteristics of studies with validity and reliability of fitness variables measures in patients’ with mental disorders.
Note: PWCF: Physical work capacity female; PCWm: Physical work capacity men; 6MWT: six minute walk test; 2MWT: two minutes walk test; VO2max: maximal oxygen uptake.

Table 2. Subgroup analysis of tests used in Test-Retest Reliability of Fitness Tests.
<table>
<thead>
<tr>
<th>Analysis</th>
<th>Number of RCTs</th>
<th>Meta-analysis</th>
<th>Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>SMD</td>
<td>95% CI</td>
</tr>
<tr>
<td>Test protocol</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2MWT</td>
<td>2</td>
<td>0.95</td>
<td>0.907</td>
</tr>
<tr>
<td>6MWT</td>
<td>3</td>
<td>0.96</td>
<td>0.938</td>
</tr>
<tr>
<td>Astrand-Rhyming</td>
<td>1</td>
<td>0.92</td>
<td>0.805</td>
</tr>
<tr>
<td>PWC 130</td>
<td>2</td>
<td>0.81</td>
<td>0.609</td>
</tr>
<tr>
<td>PWC 150</td>
<td>2</td>
<td>0.84</td>
<td>0.934</td>
</tr>
<tr>
<td>Flamingo Balance - EFT</td>
<td>3</td>
<td>0.75</td>
<td>0.812</td>
</tr>
<tr>
<td>Handgrip Strength - EFT</td>
<td>3</td>
<td>0.96</td>
<td>0.931</td>
</tr>
<tr>
<td>Plate Tapping - EFT</td>
<td>3</td>
<td>0.88</td>
<td>0.938</td>
</tr>
<tr>
<td>Sit-And-Reach - EFT</td>
<td>3</td>
<td>0.96</td>
<td>0.887</td>
</tr>
<tr>
<td>Standing Broad Jump - EFT</td>
<td>3</td>
<td>0.96</td>
<td>0.903</td>
</tr>
<tr>
<td>Shuttle Run - EFT</td>
<td>3</td>
<td>0.93</td>
<td>0.654</td>
</tr>
<tr>
<td>Sit-Ups - EFT</td>
<td>3</td>
<td>0.94</td>
<td>0.679</td>
</tr>
</tbody>
</table>

Note: RCT: Randomized Clinical Trials. SMD: Standardized mean difference. 2MWT: Two minutes walk test; 6MWT: Two minutes walk test; PWC: Physical Work Capacity; EFT: Eurofit Test