Journal of Affective Disorders 408 (2026) 121898

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research paper

Clinical utility of a quick and easy-to-use international tool for assessing and identifying impaired physical fitness in people with severe mental illness – The PsychiActive project

Alvaro Lopez-Moral [a,b], Diego Munguia-Izquierdo [a,c,*], Javier Bueno-Antequera [a,c]

[a] Physical Performance & Sports Research Center (CIRFD), Department of Sports and Computer Science, Section of Physical Education and Sports, Faculty of Sports Sciences, Universidad Pablo de Olavide, Sevilla, Spain
[b] Department of Physical Education and Sport, Universidad de Sevilla, Sevilla, Spain
[c] Research Group in Development Movimiento Humano, Universidad de Zaragoza, Zaragoza, Spain

ARTICLE INFO

Keywords:
Fitness assessment
Early detection
Physical health screening
Functional impairment
Psychiatric rehabilitation
Validity
Reliability

ABSTRACT

Background: Severe mental illness (SMI) is a leading cause of mortality, disability, and frailty, largely driven by physical multimorbidity. Physical fitness is a strong predictor of health outcomes, yet routine assessment is rarely implemented in SMI due to limited resources. Self-reported tools may offer a feasible alternative, although their validity and clinical utility remain unclear. This study evaluated the validity, reliability, and clinical utility of the International Fitness Scale (IFIS) in SMI.

Methods: In a multicenter cross-sectional study, 234 adults with SMI (18–65 years, 62 females) completed the IFIS and a battery of objective physical fitness tests. Convergent validity was assessed using ANOVA and ANCOVA. Test–retest reliability over two weeks was assessed using weighted kappa and percentage agreement. Clinical utility was evaluated using area under the curve, sensitivity, specificity, and positive predictive value.

Results: The IFIS effectively discriminated objective physical fitness across all domains and response categories in adults with severe mental illness ($p < 0.005$), with lower self-reported fitness consistently associated with poorer objective performance. Test–retest reliability was fair for the five-response scale ($\kappa = 0.22$–0.32) and improved after category reduction ($\kappa = 0.26$–0.50). The highest positive predictive value was observed for muscular strength (75%) and cardiorespiratory fitness (46%). The IFIS showed moderate discriminatory accuracy for cardiorespiratory fitness and flexibility (AUC = 0.64 for both domains). Mean completion time was 1.8 ± 1.2 min.

Conclusions: The IFIS is a valid, reliable, and time-efficient tool for identifying impaired physical fitness in adults with SMI.

1. Introduction

People with severe mental illness (SMI) have a life expectancy that is reduced by up to 20 years, primarily due to physical conditions such as cardiovascular disease, respiratory disorders, and sarcopenic obesity, which translate into impaired physical fitness (Correll et al., 2017; Goldfarb et al., 2022). People with SMI are more than twice as likely to develop physical multimorbidity than the general population (odds ratio = 2.4), being almost four times higher for those younger than 40 years (odds ratio = 3.99) (Halstead et al., 2024). This increased physical disease burden contributes to higher rates of premature mortality (Walker et al., 2015), disability (Weye et al., 2021), and frailty (Pearson et al., 2022). As a leading cause of global mortality and disability, SMI imposes a global economic burden comparable to that of cardiovascular diseases, surpassing the impact of cancer, respiratory diseases, and diabetes (Trautmann et al., 2016). These disparities underscore the urgent need for integrated physical healthcare in mental health strategies to address this critical situation in SMI (Wykes et al., 2015).

A critical approach to mitigating the increased physical disease burden in people with SMI is early identification of individuals at risk, enabling timely therapeutic interventions. Meta-analyses and overviews of meta-analyses representing over 20.9 million observations from 199 unique cohort studies consistently demonstrate that physical fitness is a key determinant of mortality and morbidity across populations, with a

* Corresponding author at: Departamento de Deporte e Informática, Universidad Pablo de Olavide, Carretera Utrera Km. 1, s/n, 41013, Sevilla, Spain.
E-mail address: dmunizq@upo.es (D. Munguia-Izquierdo).

https://doi.org/10.1016/j.jad.2026.121898
Received 16 February 2026; Received in revised form 26 April 2026; Accepted 27 April 2026
Available online 29 April 2026

stronger predictive value than traditional risk factors such as smoking, hypertension, hypercholesterolemia, and type 2 diabetes (García-Hermoso et al., 2018; Lang et al., 2024). Meta-analyses, systematic reviews, and cohort studies have shown that individuals with SMI have poorer fitness levels compared to the general population, equivalent to those of individuals 30 years older, increasing the risk of cardiovascular disease, sarcopenic obesity, and other adverse health outcomes (Scheewe et al., 2019; Vancampfort et al., 2016; Vancampfort et al., 2017a; Zechner et al., 2022). Although physical fitness and functional impairment are conceptually distinct constructs, evidence suggests that physical fitness is associated with physical functioning in a modest but clinically meaningful way, particularly in domains such as aerobic capacity, muscle strength, and balance, without implying a one-to-one relationship (Brobakken et al., 2022; Perez-Cruzado et al., 2018). Given its prognostic relevance, scientific and health organizations (such as the American Heart Association) strongly recommend a routine fitness assessment as a clinical vital sign (Ross et al., 2016). Therefore, integrating routine fitness assessments into mental health settings could enhance treatment strategies and reduce the burden of SMI.

Routine fitness assessments are rarely conducted in individuals with SMI due to limited resources, time constraints, and insufficient training in mental health settings. The scarcity of validated fitness tools further limits clinical implementation, as many methods are complex, physically demanding, and poorly tolerated (de Oliveira Tavares et al., 2021; Lopez-Moral et al., 2025b). Self-reported fitness tools may help address this gap by providing a rapid, low-burden, and cost-effective approach for identifying individuals with impaired physical fitness and have shown meaningful associations with objective fitness indicators and relevant health outcomes (Sanchez-Lastra et al., 2021). For healthcare providers, such tools offer a ‘first-step’ screening that requires no specialized equipment or extensive training, allowing efficient triage, while for patients, early detection enables targeted interventions to mitigate functional decline and reduce the 20-year mortality gap associated with SMI. The International Fitness Scale (IFIS) has demonstrated its validity across various populations, including children (Ortega et al., 2011; Sánchez-López et al., 2015, 2023), adolescents (Ramírez-Vélez et al., 2017), older adults (Merellano-Navarro et al., 2017), pregnant women (Romero-Gallardo et al., 2020), and patients with chronic diseases such as fibromyalgia (Álvarez-Gallardo et al., 2016), demonstrating moderate to substantial reliability (Pereira et al., 2020). Completion takes approximately 3 min (Español-Moya and Ramírez-Vélez, 2014), making it a brief and efficient tool for use in clinical settings. Considering that cognitive deficits due to symptom severity, along with physical and psychological conditions, may influence self-assessments (Sheffield et al., 2018), the validity and reliability of this tool in SMI remain unexplored, underscoring the need for validation to enhance its clinical utility.

This study aimed to 1) examine the convergent validity of the IFIS in discriminating between different levels of objectively measured physical fitness in individuals with SMI, 2) assess its test-retest reliability in this population, and 3) determine its clinical utility in identifying individuals with impaired physical fitness. Based on previous research, we hypothesized that the IFIS would effectively discriminate between physical fitness levels and demonstrate moderate to substantial test-retest reliability in individuals with SMI (Álvarez-Gallardo et al., 2016; Pereira et al., 2020).

2. Methodology

2.1. Study design and participants

This multicenter cross-sectional study was conducted between 2018 and 2022 as part of the PsychiActive project, which aims to assess the impact of exercise as a therapeutic intervention for individuals with SMI (Lopez-Moral et al., 2025a, 2025b). A convenience sample of adults diagnosed with SMI, including schizophrenia, schizoaffective disorder, bipolar disorder, major depressive disorder, personality disorders, and severe stress-related disorders, based on the International Classification of Diseases 10th Revision (ICD-10) criteria and stabilized on antipsychotic medication, was recruited from seven outpatient mental health settings in southern Spain. The exclusion criteria included clinical instability (presence of acute psychiatric symptoms, recent changes in clinical status within the last three months, or treatment changes in the previous month), comorbid substance abuse, and evidence of uncontrolled cardiovascular, neuromuscular, and endocrine disorders. Two evaluation sessions were scheduled, with an average interval of two weeks between them. For both evaluations, the participants first completed the IFIS questionnaire and then the anthropometric and fitness tests in a 90-min session. Maintaining a consistent order minimized potential biases. An exercise physiologist with over six years of experience assessing fitness in adults with SMI conducted the sessions, ensuring standardization, reliability, and consistency. Age, diagnosis, and duration of illness were obtained from each participant's medical records. Psychiatric symptoms over the previous week were assessed using the Brief Symptom Inventory-18 (Derogatis, 2001), with scores ranging from 0 to 72; higher scores reflect greater severity.

The authors assert that all procedures contributing to this work complied with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration (1975, revised in 2013). The procedure was approved by the ethics committees of the Virgen Macarena and Virgen del Rocío University Hospitals (1674-N-17) and followed the “Strengthening the Reporting of Observational Studies in Epidemiology” (STROBE) statement, as well as the recommendations for reporting the results of studies of instrument and scale development and testing (Streiner and Kottner, 2014). The participants provided informed consent before enrolment and did not receive compensation for their participation.

2.2. Measures

2.2.1. Anthropometric measurement

Height was measured with a stadiometer (Seca 711, Hamburg, Germany), and body weight via an InBody 770 device (Biospace, Seoul, South Korea). Each measurement was conducted twice, and the average of the two readings was calculated for the analyses. The formula for calculating BMI was body mass (kg) divided by height squared (m^2). The measurements were conducted under resting conditions after a 3-h fasting period and without fluid intake. The participants were instructed to remove all metal objects, wear minimal clothing, and urinate before the measurement. The measurements were performed in a ventilated examination room under controlled temperature and humidity conditions.

2.2.2. Objectively measured physical fitness

Physical fitness was assessed using the *“Senior Fitness Test battery”* (Rikli and Jones, 1999), which offers a standardized evaluation of cardiorespiratory endurance, muscular strength, speed and agility, and flexibility. These methods are reliable, cost-effective, and easy to administer. In addition, they require minimal complexity or physical exertion, making them well-suited for individuals with SMI.

- Cardiorespiratory fitness was assessed with the “*6-minute walk test*,” which consists of walking around a 46-m rectangular space. The objective is to cover the greatest distance possible in 6 min.
- Lower body muscular strength was assessed with the “*30-second chair stand test*,” which consists of standing up and sitting down on a chair as many times as possible, with arms crossed on the chest. A valid repetition was considered when a participant achieved a full body extension.
- Upper body muscular strength was assessed with the “*arm curl test*,” which consists of performing the greatest number of elbow flexion–extensions for 30 s with a dumbbell (3.6 kg for men and 2.3 kg

for women) while sitting on a chair. A valid repetition was considered when a complete elbow flexion was performed. Additionally, "*handgrip strength*" (Tomkinson et al., 2025) was measured using a dynamometer (TKK 5401 Grip-D, Takey, Tokyo, Japan). The participants stood with their feet shoulder-width apart and arms fully extended, applying maximum grip force as quickly and strongly as possible for five seconds.

- Lower body flexibility was assessed with the "*chair sit-and-reach test*," which consists of moving the hands together towards the tips of the toes, trying to reach and/or surpass them if possible. The elbows had to be extended, and the knee of the foot on which the test was performed had to be flexed.
- Upper body flexibility was assessed with the *"back scratch test,"* which consists of trying to touch and/or extend the tips of the fingers on the back. One hand is placed behind the shoulder on the same side with the palm down and the other behind the back with the palm up.
- Agility/dynamic balance was assessed with the "*8-foot up-and-go test*," which consists of getting up from a chair, walking 2.44 m, turning 180°, and walking back to sit on the same chair in the shortest time possible.

The tests were performed in the following order, with at least 1 min of rest between attempts and/or 3 min between tests to reduce the influence of fatigue: the handgrip strength test, back scratch test, chair sit-and-reach test, arm curl test, 8-ft up-and-go test, 30-s chair stand test, and 6-min walk test. All participants performed several submaximal familiarization trials prior to the recorded attempts to ensure correct understanding and standardized execution. The best of two attempts for each test was used in the analysis, except for the 30-s chair stand test and the 6-min walk test, which were performed only once.

More details on test execution and scoring are available in the original text (Rikli and Jones, 1999) and the free-access video tutorials prepared by the researchers themselves (available at https://upotv.upo.es/series/58da216a238583e0478b48f0). Additionally, a tutorial on the handgrip test is available at https://youtu.be/r8wSXq0NWzY?si=pcb3kJCTRCHjRx7A.

2.2.3. Self-reported physical fitness

Self-reported physical fitness was assessed using the Spanish version of the IFIS. The IFIS consists of five questions that ask participants to rate their physical fitness compared to the average person of the same age and condition (in this study, referring to peers of similar age and sex within their mental health settings) using a five-point Likert scale (very poor, poor, acceptable, good, very good). The first question evaluates overall physical fitness, while the remaining four assess specific components: cardiorespiratory fitness, muscular strength, speed and agility, and flexibility. Originally developed in English, the IFIS has been validated in nine European countries and their respective languages (German, Austrian German, Greek, Flemish, French, Hungarian, Italian, Spanish, and Swedish) as part of the HELENA project (Ortega et al., 2011). Further details, including access to the instrument, are available on the PROFITH research group's official website (https://profith.ugr.es/recursos/como-evaluar-la-condicion-fisica-en-de-forma-utoreportada-intenational-fitness-scale/).

2.3. Statistical analysis

In accordance with previous methodologies (Merellano-Navarro et al., 2017; Romero-Gallardo et al., 2020), and due to the low frequency of extreme responses, "very good" and "good" were combined, as were "poor" and "very poor," for analysis. All analyses were conducted considering age and sex, either through statistical adjustment or the use of age- and sex-specific reference standards.

The convergent validity of the IFIS in distinguishing objectively measured fitness levels based on self-reported responses in individuals with SMI was assessed using ANOVA (unadjusted) and ANCOVA (adjusted for sex and age), using data from the first evaluation session only. The objective fitness measures were dependent variables, and the self-reported responses were fixed factors. Pairwise comparisons ("very poor/poor" vs. "acceptable," "acceptable" vs. "good/very good," and "very poor/poor" vs. "good/very good") were conducted with Bonferroni adjustments. Effect size interpretation for ηp^2 (partial eta squared) followed the benchmarks: small ($\eta p^2 = 0.01$), medium ($\eta p^2 = 0.06$), and large ($\eta p^2 = 0.14$) effects (Lakens, 2013).

The mean and standard error was calculated for each response. Standardized effect sizes were calculated using Hedges' g (Cohen's d adjusted for small sample bias) with 95% confidence intervals. An online calculator was used (Uanhoro, 2020a). The formulas for Hedges' g and its CI are available on GitHub (Uanhoro, 2020b), with Hedges' $g =$ Cohen's $d \times (1 - 3/(4(n1 + n2) - 9))$. A Hedges' $g \geq 0.50$ was considered a minimally important difference (Norman et al., 2003). Standardized scores [(*value* – *mean*) / *SD*] were computed to compare performance across self-reported fitness responses. Composite z-scores were then calculated for each fitness component by averaging the standardized scores of the relevant tests: muscle strength (handgrip, arm curl, 30-s chair stand), flexibility (back scratch, chair sit-and-reach), and overall fitness (mean of all components).

The test–retest reliability of the IFIS was evaluated using weighted kappa coefficients, recommended for ordered categorical data (Cohen, 1968), along with percentage agreement. Weighted kappa accounts for exact matches and assigns partial credit to adjacent responses. Standard errors and 95% confidence intervals were calculated for each kappa coefficient to assess estimate precision. Agreement strength was classified as follows (Landis and Koch, 1977): ≤0 = poor, 0.01–0.20 = slight, 0.21–0.40 = fair, 0.41–0.60 = moderate, 0.61–0.80 = substantial, and 0.81–1 = almost perfect. Agreement levels were determined by calculating the difference between the initial test and the retest, with a difference of 0 indicating "perfect" agreement (identical responses) and ± 1 classified as "perfect-acceptable" agreement.

To assess the clinical utility of the IFIS as a screening tool for identifying individuals at risk of impaired physical fitness requiring prioritized intervention, the Area Under the Curve (ROC-AUC), sensitivity, specificity, and positive predictive values (PPV) were calculated. Impaired physical fitness was defined as a fitness level below the 25th percentile relative to age- and sex-specific reference standards (Bueno-Antequera et al., 2025). Analyses examined the ability of IFIS response categories, particularly the "very poor" and "very poor/poor" classifications, to discriminate between individuals above and below these normative thresholds. These analyses provide a quantitative evaluation of the IFIS's screening performance in identifying individuals with potentially clinically relevant low fitness levels in mental health settings, rather than its association with broader clinical outcomes.

Statistical analyses were performed using SPSS Statistics for Windows, Version 27.0 (IBM Corp., Armonk, NY, USA), with significance set at $p < 0.05$.

3. Results

A total of 304 individuals were screened for eligibility. Of these, 234 were included in the analysis, and 70 were excluded (33 due to age > 65 years and 37 due to incomplete data). Table 1 presents data from 234 participants (171 men, 63 women) with a mean age of 40.6 ± 9.2 years, ranging from 18 to 65 years. The most common diagnoses were schizophrenia, schizotypal, and delusional disorders (67%). Significant sex differences were found in height ($p < 0.001$) and weight ($p < 0.001$), with women exhibiting higher psychiatric symptom severity ($p < 0.05$). Women performed worse than men in all physical fitness tests except for the composite scores. The older participants had lower cardiorespiratory fitness, muscular strength, and speed and agility than the younger group ($p < 0.05$). All participants demonstrated full comprehension of the tool, with a time to complete of 1.8 ± 1.2 min.

Table 1
Sociodemographic characteristics.

		All	SEX				AGE			
		$n = 234$	Men $n = 171$	Women $n = 63$	*p*-value	ES (g) [95% IC]	Age ≤ 41 $n = 126$	Age > 41 $n = 108$	*p*-value	ES (g) [95% IC]
Age (years)		40.6 ± 9.2	40.7 ± 9.4	40.7 ± 8.8	0.913	0.00 [−0.29, 0.29]	33.8 ± 5.7	48.6 ± 5.6	< 0.001	−2.61 [−2.97, −2.27]
Height (cm)		171.8 ± 8.4	175.4 ± 6.1	162.1 ± 5.7	< 0.001	2.21 [1.86, 2.57]	172.9 ± 8.4	170.7 ± 8.4	< 0.05	0.26 [0.00, 0.52]
Weight (kg)		89.4 ± 19.2	92.4 ± 17.6	81.5 ± 21.1	< 0.001	0.58 [0.29, 0.88]	90.6 ± 19.6	88.2 ± 18.7	0.354	0.12 [−0.13, 0.38]
BMI (kg/m^2)		30.2 ± 5.9	29.9 ± 5.3	30.8 ± 7.2	0.286	−0.15 [−0.44, 0.14]	30.2 ± 6.1	30.2 ± 5.6	0.935	0.00 [−0.26, 0.26]
Illness duration (years)		15.2 ± 9.3	15.9 ± 9.0	13.4 ± 9.9	0.116	0.26 [−0.03, 0.55]	10.7 ± 6.8	21.5 ± 8.7	< 0.001	−1.39 [−1.68, −1.11]
Severity of psychiatric symptoms (0–72)[a]		16.1 ± 13.2	14.4 ± 11.8	20.8 ± 16.0	< 0.05	−0.49 [−0.78, −0.20]	16.9 ± 12.6	15.3 ± 14.1	0.350	0.12 [−0.14, 0.38]
Objective physical fitness										
OVERALL	*Z*-score overall fitness	−0.002 ± 0.4	−0.003 ± 0.4	0.000 ± 0.8	0.955	0.00 [−0.29, 0.29]	0.037 ± 0.4	−0.050 ± 0.4	0.087	0.25 [0.00, 0.51]
CRF	6-min walk test (m)	574.1 ± 99.6	592.8 ± 95.8	523.2 ± 92.2	< 0.001	0.73 [0.44, 1.03]	586.1 ± 99.1	560.0 ± 98.7	< 0.05	0.26 [0.01, 0.52]
MF	Handgrip (kg)	42.2 ± 11.0	47.0 ± 8.3	29.1 ± 5.4	< 0.001	2.33 [1.98, 2.70]	43.4 ± 10.9	40.8 ± 11.0	0.070	0.24 [−0.02, 0.50]
	Arm curl test (rep)	27.4 ± 6.5	28.0 ± 6.7	25.9 ± 5.5	< 0.05	0.33 [0.04, 0.62]	28.7 ± 6.8	25.8 ± 5.7	< 0.05	0.46 [0.20, 0.72]
	30-Second chair test (rep)	23.0 ± 6.4	23.6 ± 6.3	21.5 ± 6.3	<0.05	0.33 [0.04, 0.62]	24.3 ± 6.0	21.6 ± 6.6	<0.05	0.43 [0.17, 0.69]
	Z-score muscular fitness	0.000 ± 0.8	0.001 ± 0.8	0.000 ± 0.8	0.991	0.00 [−0.29, 0.29]	0.184 ± 0.7	−0.217 ± 0.7	< 0.001	0.57 [0.31, 0.83]
SP-AG	8-Foot up-and-go test (s)[b]	4.7 ± 1.2	4.6 ± 1.3	5.2 ± 0.9	< 0.05	−0.50 [−0.79, −0.21]	4.5 ± 1.3	5.0 ± 1.1	< 0.05	−0.41 [−0.67, −0.15]
	Back scratch test (cm)	−3.8 ± 11.7	−4.7 ± 12.4	−1.5 ± 9.4	< 0.05	−0.27 [−0.56, 0.02]	−2.6 ± 10.6	−5.2 ± 12.9	0.095	0.22 [−0.04, 0.48]
FLEX	Chair sit-and-reach test (cm)	−1.8 ± 12.3	−3.8 ± 12.5	3.5 ± 10.0	< 0.001	−0.61 [−0.91, −0.32]	−1.3 ± 12.2	−2.5 ± 12.4	0.449	0.10 [−0.16, 0.35]
	Z-score flexibility	0.000 ± 0.8	−0.001 ± 0.9	0.000 ± 0.8	0.993	0.00 [−0.29, 0.29]	0.079 ± 0.8	−0.093 ± 0.9	0.117	0.24 [−0.02, 0.49]
Diagnoses[c]										
Schizophrenia, schizotypal and delusional disorders		142 (67%)	116 (75%)	26 (47%)			77 (69%)	65 (65%)		
Mood [affective] disorders		41 (19%)	24 (15%)	17 (30%)			20 (18%)	21 (21%)		
Neurotic, stress-related and somatoform disorders		11 (5%)	3 (2%)	8 (14%)			7 (6%)	4 (4%)		
Disorders of adult personality and behavior		18 (9%)	13 (8%)	8 (9%)			8 (7%)	10 (10%)		

The distribution of the IFIS responses across the fitness components, with the extreme responses grouped, is presented in Table 2. The majority of the participants rated their fitness levels as "acceptable" (41% in overall fitness), a pattern consistent across all components, indicating a central tendency in self-perceived fitness. Women self-reported lower physical fitness responses than men in the components of cardiorespiratory fitness, muscular fitness, and speed and agility ($p < 0.05$). No differences in the distribution of responses were observed between the age groups. The distribution of the IFIS with the extreme responses ungrouped is presented in Supplementary Table 1.

Table 3 presents the between-responses differences in objectively measured physical fitness, showing significant differences ($p < 0.005$) across all IFIS responses and fitness components, even after adjusting for age and sex. The ANOVA results indicated medium effect sizes ($\eta p^2 \geq 0.06$) for all tests. Pairwise comparisons revealed that individuals in the "very poor/poor" response group had significantly lower performance than those in the "good/very good" response group ($p < 0.05$) across overall fitness and components. These differences exceeded the threshold for minimally important differences in all fitness tests (i.e., Hedges' $g \geq 0.50$) except for the back scratch test ($g = -0.32$). After adjusting for age and sex, the IFIS remained effective in distinguishing between these responses, particularly in flexibility and muscular fitness, where it was able to differentiate among all three responses ($p < 0.05$).

The test-retest reliability of the IFIS is shown in Table 4. Weighted kappa (κ) coefficients ranged from 0.22 to 0.32 for the five-response scale and from 0.26 to 0.50 for the three-response scale. The mean perfect agreement was observed in 46% (range 41–50) for the five-response scale and 59% (range 55–63) for the three-response scale, with perfect-acceptable agreement increasing to 89%.

Table 5 shows the clinical utility of the IFIS in identifying impaired physical fitness. In the grouped responses, the highest PPV was observed in muscular strength (75%; Handgrip) and cardiorespiratory fitness (46%; 6-min walk test), followed by flexibility (37%; Back Scratch test). Regarding the ROC-AUC, the scale demonstrated its highest discriminatory capacity in cardiorespiratory fitness (0.64, 95% CI: 0.55–0.73; 6-min walk test) and flexibility (0.64, 95% CI: 0.55–0.73; Back Scratch test). While specificity remained robust across all dimensions (72–85%), the sensitivity was higher for flexibility (52%; Back Scratch test) and cardiorespiratory fitness (44%; 6-min walk test).

Table 2
Responses were obtained from the International Fitness Scale, with extreme responses grouped.

Components		Responses: Very Poor/ Poor (1)	Acceptable (3)	Good/Very good (5)
Overall	All	51 (22%)	96 (41%)	87 (37%)
Sex	Men	34 (20%)	69 (40%)	68 (40%)
	Women	17 (27%)	27 (43%)	19 (30%)
	p-value		0.320	
Age	≤ 41 years	29 (23%)	49 (39%)	48 (38%)
	> 41 years	22 (20%)	47 (44%)	39 (36%)
	p-value		0.759	
Cardiorespiratory fitness	All	55 (23%)	90 (39%)	89 (38%)
Sex	Men	36 (21%)	62 (36%)	73 (43%)
	Women	19 (30%)	28 (45%)	16 (25%)
	p-value		<0.05	
Age	≤ 41 years	27 (21%)	48 (38%)	51 (41%)
	> 41 years	28 (26%)	42 (39%)	38 (35%)
	p-value		0.626	
Muscular fitness	All	56 (24%)	91 (39%)	87 (37%)
Sex	Men	30 (18%)	69 (40%)	72 (42%)
	Women	26 (41%)	22 (35%)	15 (24%)
	p-value		<0.001	
Age	≤ 41 years	31 (25%)	51 (40%)	44 (35%)
	> 41 years	25 (23%)	40 (37%)	43 (40%)
	p-value		0.740	
Speed and Agility	All	57 (24%)	92 (39%)	85 (36%)
Sex	Men	33 (19%)	74 (43%)	64 (37%)
	Women	24 (38%)	18 (29%)	21 (33%)
	p-value		<0.05	
Age	≤ 41 years	29 (23%)	52 (41%)	45 (36%)
	> 41 years	28 (26%)	40 (37%)	40 (37%)
	p-value		0.781	
Flexibility	All	70 (30%)	91 (39%)	73 (31%)
Sex	Men	52 (30%)	69 (40%)	50 (30%)
	Women	18 (29%)	22 (35%)	23 (36%)
	p-value		0.554	
Age	≤ 41 years	40 (32%)	50 (40%)	36 (28%)
	> 41 years	30 (28%)	41 (38%)	37 (34%)
	p-value		0.621	

Note. Values expressed as frequency (%).

4. Discussion

This study has demonstrated the utility of the IFIS as a valid, reliable, and easy-to-use tool for quickly identifying impaired physical fitness in adults with SMI. The IFIS effectively distinguished differences across all responses and fitness components, even after adjusting for age and sex. Specifically, the participants with the "very poor/poor" responses exhibited significantly lower performance than those with the "good/very good" responses across all fitness tests. The test-retest reliability showed fair agreement, improving when reduced from five to three responses. Regarding its clinical utility, the IFIS identified impaired physical fitness in muscular strength (75% PPV) and cardiorespiratory fitness (46% PPV), followed by flexibility (37% PPV) within the "very poor/poor" category. These values, alongside an AUC of 0.64 for both the cardiorespiratory and flexibility domains, indicate the scale's potential for identifying individuals with SMI at high risk of physical impairment. Consequently, its application may assist in the clinical screening of patients who would benefit from targeted exercise interventions.

Our study revealed a central tendency in IFIS responses, with low frequencies in extremes, grouped for analysis. This pattern is consistent with findings in older adults (Merellano-Navarro et al., 2017), young adults (Ortega et al., 2013), and pregnant women (Romero-Gallardo et al., 2020). However, it differs in younger populations, such as children (Sánchez-López et al., 2023) and adolescents (Ramírez-Vélez et al., 2017), where only one extreme response is grouped. The central tendency in self-reported fitness among individuals with SMI can be attributed to poor physical health (Halstead et al., 2024; Vancampfort et al., 2016) and sedentary behavior (Bueno-Antequera et al., 2018; Vancampfort et al., 2017b), limiting physical activity and objective fitness benchmarks. These factors make it difficult for individuals to differentiate their fitness levels from their peers, leading to moderate self-assessments. A similar trend is observed in other chronic diseases, such as fibromyalgia (Álvarez-Gallardo et al., 2016), where individuals also report more neutral responses.

The IFIS effectively discriminated between different fitness levels in adults with SMI, demonstrating convergent validity through significant differences between self-reported and objectively measured fitness levels across all components. Pairwise comparisons revealed its ability to distinguish between fitness levels, particularly in extreme responses, even after adjusting for age and sex. However, a greater discriminative capacity was observed at the lower end of the scale, with more significant differences between the "very poor/poor" and "acceptable" groups than between the "acceptable" and "good/very good" groups. This suggests that lower fitness levels may be more easily perceived and

Table 3
Means and SE of measured physical fitness by self-reported physical fitness responses in SMI.

		Very Poor/ Poor (1)		Acceptable (3)		Good/ Very Good (5)			Pairwise comparisons between groups					
									1–3		3–5		1–5	
		n (%)	Mean (SE)	*n* (%)	Mean (SE)	*n* (%)	Mean (SE)	*p*-value	*p*-value	ES (g) [95% IC]	*p*-value	ES (g) [95% IC]	*p*-value	ES (g) [95% IC]
Unadjusted models														
Overall	Z-score fitness	48 (21%)	−0.2 (0.1)	95 (41%)	0.0 (0.0)	87 (38%)	0.1 (0.0)	< 0.001	0.001	−0.62 [−0.98,-0.27]	0.143	−0.26 [−0.55, 0.04]	< 0.001	−0.75 [−1.11,-0.39]
CRF	6-min walk test (m)	55 (24%)	510.7 (12.5)	90 (38%)	581.9 (9.8)	89 (38%)	605.4 (9.9)	< 0.001	< 0.001	−0.76 [−1.16,-0.36]	0.279	−0.25 [−0.64, 0.15]	< 0.001	−1.04 [−1.45,-0.64]
MF	Handgrip (kg)	56 (24%)	36.2 (1.4)	91 (39%)	42.3 (1.1)	87 (37%)	45.9 (1.1)	< 0.001	0.002	−0.59 [−0.93,-0.25]	0.070	−0.34 [−0.64,-0.05]	< 0.001	−0.93 [−1.28,-0.58]
	Arm curl test (rep)	55 (24%)	24.6 (0.8)	91 (39%)	26.9 (0.6)	87 (37%)	29.7 (0.7)	< 0.001	0.087	−0.39 [−0.73, 0.06]	0.011	−0.39 [−0.73, 0.06]	< 0.001	−0.74 [−1.10,-0.40]
	30-Second chair test (rep)	54 (24%)	20.5 (0.8)	91 (39%)	23.0 (0.7)	85 (37%)	24.7 (0.7)	0.001	0.064	−0.42 [−0.76,-0.08]	0.205	−0.28 [−0.57, 0.02]	< 0.001	−0.64 [−0.99,-0.29]
	Z-score MF	54 (24%)	−0.3 (0.1)	91 (39%)	0.0 (0.1)	85 (37%)	0.3 (0.1)	< 0.001	0.063	−0.42 [−0.77,-0.09]	0.015	−0.43 [−0.73,-0.13]	< 0.001	−0.77 [−1.12,-0.42]
SP-AG	8-Foot up-and-go test (s) [a]	56 (24%)	5.4 (0.2)	92 (40%)	4.6 (0.1)	85 (36%)	4.5 (0.1)	< 0.001	<0.001	0.56 [0.22, 0.90]	1.000	0.14 [−0.15, 0.44]	< 0.001	0.71 [0.36, 1.06]
FLEX	Back scratch test (cm)	70 (30%)	−8.4 (1.4)	91 (39%)	−3.5 (1.2)	73 (31%)	0.1 (1.3)	< 0.001	0.020	−0.44 [−0.76,-0.12]	0.135	−0.32 [−0.63,-0.01]	< 0.001	−0.32 [−0.63,-0.01]
	Chair sit-and-reach test (cm)	69 (30%)	−5.6 (1.4)	91 (39%)	−2.0 (1.3)	73 (31%)	2.0 (1.4)	0.001	0.182	−0.29 [−0.60, 0.03]	0.092	−0.34 [−0.65,-0.03]	0.001	−0.69 [−1.03,−0.35]
	Z-score FLEX	69 (30%)	-0.3 (0.1)	91 (39%)	0.0 (0.1)	73 (31%)	0.3 (0.1)	< 0.001	0.017	−0.45 [−0.77,-0.14]	0.050	−0.36 [−0.68,-0.05]	< 0.001	−0.36 [−0.68,-0.05]
Adjusted models[b]														
Overall	Z-score fitness	48 (21%)	−0.2 (0.1)	95 (41%)	0.0 (0.0)	87 (38%)	0.1 (0.0)	< 0.001	0.003	−0.62 [−0.98,-0.27]	0.309	−0.26 [−0.55, 0.04]	< 0.001	−0.75 [−1.12,-0.39]
CRF	6-Minute walk test (m)	55 (24%)	517.0 (11.9)	90 (38%)	583.4 (9.3)	89 (38%)	599.9 (9.4)	< 0.001	< 0.001	−0.59 [−0.95,−0.24]	0.651	-0.24 [−0.53, 0.05]	< 0.001	−0.83 [−1.20,-0.47]
MF	Handgrip (kg)	56 (24%)	39.5 (1.0)	91 (39%)	41.9 (0.8)	87 (37%)	44.2 (0.8)	0.002	0.187	−0.38 [−0.72,-0.04]	0.114	−0.26 [−0.56, 0.03]	0.001	−0.59 [−0.94,-0.24]
	Arm curl test (rep)	55 (24%)	24.9 (0.8)	91 (39%)	27.0 (0.6)	87 (37%)	29.5 (0.6)	< 0.001	0.127	−0.27 [−0.61, 0.07]	0.878	−0.44 [−0.74, −0.14]	< 0.001	−0.70 [−1.06, −0.36]
	30-Second chair test (rep)	54 (24%)	20.9 (0.8)	91 (39%)	23.0 (0.6)	85 (37%)	24.5 (0.6)	0.003	0.118	−0.36 [−0.70, −0.02]	0.301	−0.25 [−0.54, 0.05]	0.002	−0.60 [−0.95, −0.25]
	Z-score MF	54 (24%)	−0.4 (0.1)	91 (39%)	0.0 (0.1)	85 (37%)	0.3 (0.1)	< 0.001	0.022	−0.47 [−0.81, −0.13]	0.012	−0.43 [−0.73, −0.13]	< 0.001	−0.89 [−1.25, −0.53]
SP-AG	8-Foot up-and-go test (s)	56 (24%)	5.2 (0.2)	92 (40%)	4.6 (0.1)	85 (36%)	4.5 (0.1)	< 0.001	0.003	0.57 [0.23, 0.91]	1.000	0.10 [−0.19, 0.40]	< 0.001	0.67 [0.33, 1.02]
FLEX	Back scratch test (cm)	70 (30%)	−8.4 (1.3)	91 (39%)	−3.4 (1.2)	73 (31%)	0.0 (1.3)	< 0.001	0.016	−0.46 [−0.77, −0.14]	0.168	−0.30 [−0.61, 0.01]	< 0.001	−0.76 [−1.10, −0.42]
	Chair sit-and-reach test (cm)	69 (30%)	−5.6 (1.4)	91 (39%)	−1.8 (1.2)	73 (31%)	1.8 (1.4)	0.001	0.125	−0.33 [−0.64, −0.01]	0.149	−0.31 [−0.62, 0.00]	0.001	−0.63 [−0.97, −0.30]
	Z-score FLEX	69 (30%)	−0.4 (0.1)	91 (39%)	0.0 (0.1)	73 (31%)	0.3 (0.1)	< 0.001	0.015	−0.45 [−0.77, −0.13]	0.047	−0.38 [−0.69, −0.07]	< 0.001	−0.83 [−1.18, −0.49]

Note. SE: Standard error. CRF: cardiorespiratory fitness; MF: muscular fitness; FLEX: flexibility; SP-AG: speed and agility. 8 ft up-and-go test (s)[a]: Lower time indicates better performance. ES (g) [95% CI]: effect size using Hedges's g with a 95% confidence interval. Adjusted model [b]: analysis of covariance adjusted for sex and age. Pairwise comparisons were performed using Bonferroni-adjusted post hoc tests. Significant differences ($p < 0.05$).

Table 4
Test-retest reliability and agreement analysis of the International Fitness Scale (IFIS) in SMI ($n = 230$), considering the number of responses.

	5 Responses (ungrouped)					3 Responses (grouped)			
IFIS items	**Weighted kappa coefficients**	**Standard error**	**95% Confidence interval**	**Perfect agreement**	**Perfect-acceptable agreement**	**Weighted kappa coefficients**	**Standard error**	**95% Confidence interval**	**Perfect agreement**
Overall	0.28	0.05	(0.19, 0.65)	47	88	0.33	0.07	(0.18, 0.48)	59
Cardiorespiratory fitness	0.22	0.05	(0.13, 0.47)	41	86	0.26	0.07	(0.11, 0.40)	55
Muscular fitness	0.32	0.05	(0.23, 0.77)	50	89	0.50	0.07	(0.36, 0.63)	63
Speed and Agility	0.26	0.04	(0.17, 0.59)	46	86	0.36	0.07	(0.19, 0.48)	59
Flexibility	0.28	0.05	(0.19, 0.66)	47	90	0.38	0.07	(0.24, 0.52)	59

Note. Perfect agreement (%): percentage of responses with exact agreement. Perfect-acceptable agreement (%): percentage of responses with exact agreement or ± 1 value.

Table 5
Clinical utility of the IFIS in identifying impaired physical fitness levels compared to reference values.

Components		3 Responses (grouped)				5 Responses (ungrouped)			
	Fitness tests	PPV (%)	ROC-AUC (IC 95%)	Sensibility (%)	Specificity (%)	PPV (%)	ROC-AUC (IC 95%)	Sensibility (%)	Specificity (%)
CRF	6-Minute walk test	45	0.64 (0.55, 0.73)	45	83	50	0.55 (0.46, 0.64)	14	96
MF	Handgrip	75	0.57 (0.50, 0.65)	30	85	71	0.52 (0.44, 0.59)	8	95
	Arm curl test	20	0.56 (0.45, 0.67)	34	78	18	0.51 (0.40, 0.62)	9	93
	30-Second chair test	36	0.62 (0.52, 0.71)	43	81	41	0.55 (0.45, 0.64)	15	95
SP-AG	8-Foot up-and-go test	12	0.42 (0.34, 0.50)	12	72	20	0.50 (0.40, 0.58)	5	93
FLEX	Back scratch test	37	0.64 (0.55, 0.73)	52	76	44	0.55 (0.45, 0.64)	14	95
	Chair sit-and-reach test	31	0.58 (0.49, 0.67)	42	74	13	0.48 (0.40, 0.57)	4	92

Note. PPV: positive predictive value; ROC–AUC: area under the receiver operating characteristic curve; 95% CI: 95% confidence interval.

reported, whereas differentiating between moderate and high fitness may be more challenging, likely reflecting less precise self-perception at higher levels. Clinically, this is particularly relevant, as it supports the IFIS as a useful tool for accurately identifying individuals at higher risk of impaired physical fitness.

Similar patterns of discrimination and validity were observed in other populations, including children (Sánchez-López et al., 2023) and young adults (Ortega et al., 2013), as well as pregnant women (Romero-Gallardo et al., 2020) and patients with chronic diseases such as fibromyalgia (Álvarez-Gallardo et al., 2016). In women with fibromyalgia, the IFIS showed reduced ability to discriminate between extreme fitness levels, particularly in cardiorespiratory fitness, muscular strength, and speed and agility, likely due to the under-representation of extreme categories (Álvarez-Gallardo et al., 2016). Similarly, in preschoolers, the IFIS demonstrated low agreement with objectively measured fitness when completed by an external proxy (Sánchez-López et al., 2023). The reliance on subjective perceptions and the limited representation of extreme responses may constrain the tool's discriminative capacity and overall validity. Simplifying response options and increasing sample sizes may help to enhance its performance. It is also important to note that the IFIS was originally developed for youth as part of the HELENA project (Ortega et al., 2011) and has been extensively validated in children and adolescents, whereas validation data in general adult populations remain comparatively scarce.

The two-week test-retest reliability of the IFIS in adults with SMI demonstrated fair agreement for the five-response scale ($\kappa = 0.22$–0.32) and the three-response scale across all components ($\kappa = 0.26$–0.38), with the exception of muscular fitness, which showed moderate agreement ($\kappa = 0.50$). These values are lower than those reported in a recent systematic review (Pereira et al., 2020). In that review, weighted kappa values ranged from 0.40 to 0.99, with over 50% of the components achieving $\kappa \geq 0.60$. The agreement was particularly strong in younger populations (children: $\kappa = 0.90$–0.98; adolescents: $\kappa = 0.73$–0.80). Adults exhibited lower κ values (0.46–0.62). Despite this fair agreement, perfect agreement was observed in 47% of the participants, increasing to 89% when including perfect-acceptable agreement. This aligns with the findings of Ortega et al. (2011) in adolescents (65% and 97%, respectively). The systematic review by Pereira et al. (2020) suggests that factors such as the participant's health status, test-retest interval, questionnaire administration, and comprehension may affect reliability. In our study, we standardized the evaluation process with a two-week test-retest interval, a fixed test order, and expert administration, ensuring the participants fully understood the questionnaire and reducing the risk of bias. However, health-related factors in SMI, such as age and cognitive impairments associated with antipsychotic medication or disease symptoms, may explain the lower reliability scores observed (Sheffield et al., 2018). In addition, individuals with SMI often present low levels of physical activity and high sedentary behavior, which may limit prior experience with structured fitness tasks and contribute to less precise self-perceptions (Vancampfort et al., 2017a). These factors may impair accurate self-assessment, especially when distinguishing between closely related response options (e.g., "very poor" and "poor"). This may partly explain the observed central tendency, as well as the higher agreement with the three-response scale compared to the five-response format. Simplifying the IFIS to three responses could reduce cognitive demands, and, thus, enhance the reliability in SMI populations.

To the authors' knowledge, this is the first study evaluating the clinical utility of the IFIS for identifying individuals with impaired physical fitness. The analysis revealed that within the "very poor/poor" category, the IFIS effectively identified approximately one in every two individuals at risk, particularly in muscular strength and cardiorespiratory fitness, a component the American Heart Association considers a vital sign and advocates for its routine assessment (Ross et al., 2016). Epidemiological evidence consistently demonstrates an inverse relationship between cardiorespiratory fitness and all-cause mortality, independent of age, sex, race/ethnicity, and traditional cardiovascular risk factors, such as cholesterol levels, blood pressure, obesity, smoking, family history, glycemia, and type 2 diabetes (Lang et al., 2024). This is particularly useful for screening individuals with SMI, as low cardiorespiratory fitness is a primary factor contributing to the high prevalence of mortality and disability, especially from cardiovascular disease (Correll et al., 2017). Furthermore, muscular strength serves as an independent predictor of mortality and demonstrates a protective effect against metabolic syndrome, hypertension, and age-related adiposity (García-Hermoso et al., 2018). Given the high prevalence of sarcopenic obesity and frailty in SMI, the capacity of the IFIS to detect low strength levels is essential for prioritizing interventions that mitigate disability and functional dependence. From a public health perspective, the IFIS could enhance routine screening in SMI, facilitating targeted interventions to reduce physical disparities and the associated health burden.

4.1. Limitations and strengths

One of the main limitations of this study is the predominance of male participants with SMI, which may limit the generalizability of validity results, particularly to females. The study also did not account for SMI-specific factors, such as psychotic symptomatology, medication side effects, and cognitive impairments common in SMI, which can affect memory, self-assessment, and, consequently, the validity and reliability of the IFIS. These factors also hinder the accurate differentiation between similar responses, which may explain the low percentage of extreme responses on the IFIS and the reliability of the results. In addition, exposure to the physical fitness tests may have influenced participants' self-perceptions between assessments, potentially affecting retest responses. The relatively low mean severity of psychiatric symptoms observed in the sample may also limit generalizability, as participants were predominantly clinically stable outpatients, and the use of a convenience sample requiring attendance at the assessment site may have underrepresented individuals with poorer health status or greater functional limitations. Future research should focus on recruiting participants with a broader range of fitness levels, enhancing female representation, and accounting for SMI-specific factors to improve the validity and reliability of the IFIS.

Despite its limitations, this is the first study to evaluate the convergent validity, reliability, and clinical utility of the IFIS in individuals with SMI. We standardized the evaluation process with a two-week test-retest interval, fixed test order, and expert administration, ensuring full participant understanding and reducing the risk of bias. Additionally, our larger sample size (234 participants) compared to previous studies (19–82 participants) has strengthened the robustness of the findings (de Oliveira Tavares et al., 2021; Lopez-Moral et al., 2025a, 2025b). Collapsing the response categories into a three-response format did not compromise internal consistency, which remained stable across formats, while improving sensitivity and overall discrimination (Supplementary Table 2). This makes the three-response version particularly useful as a screening tool in contexts where early detection of individuals with impaired physical fitness is a priority. The inclusion of a diverse range of SMI types enhances the generalizability of our results, supporting their applicability in real-world clinical settings.

The clinical utility of the IFIS was assessed for detecting impaired physical fitness, improving routine screening, and identifying physical limitations across different fitness components, which can guide more specific interventions. The inclusion of a diverse range of SMI types enhances the generalizability of our results, supporting their applicability in real-world clinical settings.

5. Conclusions

This study demonstrates the utility of the IFIS as a valid and reliable tool for identifying impaired physical fitness in individuals with SMI. Its ability to distinguish fitness levels and detect those at risk highlights its clinical utility. Given its ease of use and cost-effectiveness, the IFIS could facilitate routine fitness assessments in mental health settings, supporting early interventions to mitigate health risks.

CRediT authorship contribution statement

Alvaro Lopez-Moral: Writing – original draft, Formal analysis, Conceptualization. **Diego Munguia-Izquierdo:** Writing – review & editing, Supervision, Funding acquisition. **Javier Bueno-Antequera:** Writing – review & editing, Supervision, Investigation.

Patient consent statement

Consent for publication was obtained from each participant included in this study. The participants were informed about the nature, purpose, and potential implications of the research, and they provided written consent, acknowledging their understanding and agreement to the publication of the findings.

Ethics approval statement

The authors assert that all procedures associated with this work complied with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2013. All procedures involving human subjects/patients were approved by the ethics committees of the Virgen Macarena and Virgen del Rocío University Hospitals (1674-N-17).

Funding

This work was funded by Proyectos I + D + i 2020 (Ref. PID2020-118262RB-I00) from the Spanish Ministry of Science and Innovation. It was also supported by the CTS-948 Research Group, Universidad Pablo de Olavide, Junta de Andalucía, and FEDER Programme 2014–2020 (Ref. UPO-1262802). ALM was supported by Universidad Pablo de Olavide (Ref. PPI1803). JBA was supported by the Operational Programme FEDER Andalucía 2014–2020 (Ref. PAC2042) and by the Junta de Andalucía through the Consejería de Transformación Económica, Industria, Conocimiento y Universidades (Ref. POSTDOC_21_00059). Funding for open access publishing: Universidad Pablo de Olavide/CBUA. The funders were not involved in the design of the study, the collection, analysis, and interpretation of the data, the writing of the report, or the decision to publish.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank all the participants, FAISEM, and the Hospital Virgen de Valme for their support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jad.2026.121898.

Data availability

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

Álvarez-Gallardo, I.C., Soriano-Maldonado, A., Segura-Jiménez, V., Carbonell-Baeza, A., Estévez-López, F., McVeigh, J.G., Delgado-Fernández, M., Ortega, F.B., 2016. International FItness scale (IFIS): construct validity and reliability in women with fibromyalgia: the al-Ándalus project. Arch. Phys. Med. Rehabil. 97 (3), 395–404. https://doi.org/10.1016/j.apmr.2015.08.416.

Brobakken, M.F., Nygård, M., Wang, E., 2022. Physical health impairment and exercise as medicine in severe mental disorders: A narrative review. Sports Medicine - Open 8 (1), 115. https://doi.org/10.1186/s40798-022-00490-3.

Bueno-Antequera, J., Munguía-Izquierdo, D., Oviedo-Caro, M. Á., & López-Moral, Á. (2025). Fitness reference values to identify health risks in severe mental disorders: A multicenter cross-sectional study from the PsychiActive project. PREPRINT (version 1) available at Research Square. Doi:10.21203/rs.3.rs-6375080/v1.

Bueno-Antequera, J., Oviedo-Caro, M.Á., Munguía-Izquierdo, D., 2018. Sedentary behaviour, physical activity, cardiorespiratory fitness and cardiometabolic risk in psychosis: the PsychiActive project. Schizophr. Res. 195, 142–148. https://doi.org/10.1016/j.schres.2017.10.012.

Cohen, J., 1968. Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. Psychol. Bull. 70 (4), 213–220. https://doi.org/10.1037/h0026256.

Correll, C.U., Solmi, M., Veronese, N., Bortolato, B., Rosson, S., Santonastaso, P., Thapa-Chhetri, N., Fornaro, M., Gallicchio, D., Collantoni, E., Pigato, G., Favaro, A., Monaco, F., Kohler, C., Vancampfort, D., Ward, P.B., Gaughran, F., Carvalho, A.F., Stubbs, B., 2017. Prevalence, incidence and mortality from cardiovascular disease in patients with pooled and specific severe mental illness: a large-scale meta-analysis of 3,211,768 patients and 113,383,368 controls. World Psychiatry 16 (2), 163–180. https://doi.org/10.1002/wps.20420.

de Oliveira Tavares, V.D., Vancampfort, D., Hallgren, M., Heissel, A., Chaparro, C.G.A.P., Solmi, M., Tempest, G.D., de Oliveira Neto, L., Galvão-Coelho, N.L., Firth, J., Schuch, F.B., 2021. Reliability and validity of physical fitness tests in people with mental disorders: A systematic review and meta-analysis. Physiother. Res. Int. 26 (3). https://doi.org/10.1002/pri.1904.

Derogatis, L.R., 2001. Brief Symptom Inventory 18. Johns Hopkins University, Baltimore.

Español-Moya, M.N., Ramírez-Vélez, R., 2014. Validación del cuestionario International Fitness Scale (IFIS) en sujetos colombianos de entre 18 y 30 años de edad. Rev. Esp. Salud Publica 88 (2), 271–278. https://doi.org/10.4321/S1135-57272014000200009.

García-Hermoso, A., Cavero-Redondo, I., Ramírez-Vélez, R., Ruiz, J.R., Ortega, F.B., Lee, D.-C., Martínez-Vizcaíno, V., 2018. Muscular strength as a predictor of all-cause mortality in an apparently healthy population: A systematic review and Meta-analysis of data from approximately 2 million men and women. Arch. Phys. Med. Rehabil. 99 (10), 2100–2113.e5. https://doi.org/10.1016/j.apmr.2018.01.008.

Goldfarb, M., De Hert, M., Detraux, J., Di Palo, K., Munir, H., Music, S., Piña, I., Ringen, P.A., 2022. Severe mental illness and cardiovascular disease: JACC state-of-the-art review. J. Am. Coll. Cardiol. 80 (9), 918–933. https://doi.org/10.1016/j.jacc.2022.06.017.

Halstead, S., Cao, C., Høgnason Mohr, G., Ebdrup, B.H., Pillinger, T., McCutcheon, R.A., Firth, J., Siskind, D., Warren, N., 2024. Prevalence of multimorbidity in people with and without severe mental illness: a systematic review and meta-analysis. Lancet Psychiatry 11 (6), 431–442. https://doi.org/10.1016/S2215-0366(24)00091-9.

Lakens, D., 2013. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front. Psychol. 4. https://doi.org/10.3389/fpsyg.2013.00863.

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33 (1), 159. https://doi.org/10.2307/2529310.

Lang, J.J., Prince, S.A., Merucci, K., Cadenas-Sanchez, C., Chaput, J.-P., Fraser, B.J., Manyanga, T., McGrath, R., Ortega, F.B., Singh, B., Tomkinson, G.R., 2024. Cardiorespiratory fitness is a strong and consistent predictor of morbidity and mortality among adults: an overview of meta-analyses representing over 20.9 million observations from 199 unique cohort studies. Br. J. Sports Med. 58 (10), 556–566. https://doi.org/10.1136/bjsports-2023-107849.

Lopez-Moral, A., Munguia-Izquierdo, D., Borrueco-Sanchez, J., Lopez-Sanchez, C., Bueno-Antequera, J., 2025a. PowerONyou: enhancing well-being through exercise for individuals with severe mental disorders in ambulatory mental health clinics in southern Spain. British Journal of sports medicine, bjsports-2024-109533. https://doi.org/10.1136/bjsports-2024-109533.

Lopez-Moral, A., Munguia-Izquierdo, D., Bueno-Antequera, J., 2025b. Assessing a clinical vital sign in severe mental illness: validation study of the 5 sit-to-stand test for monitoring muscle strength – the PsychiActive project. BJPsych Open 11 (1), e24. https://doi.org/10.1192/bjo.2024.842.

Merellano-Navarro, E., Collado-Mateo, D., García-Rubio, J., Gusi, N., Olivares, P.R., 2017. Validity of the international fitness scale "IFIS" in older adults. Exp. Gerontol. 95, 77–81. https://doi.org/10.1016/j.exger.2017.05.001.

Norman, G., Sloan, J., Wyrwich, K., 2003. Interpretation of changes in health-related quality of life the remarkable universality of half a standard deviation. Med. Care 41 (5), 582–592. https://doi.org/10.1097/00005650-200305000-00004.

Ortega, F.B., Ruiz, J.R., España-Romero, V., Vicente-Rodriguez, G., Martínez-Gómez, D., Manios, Y., Béghin, L., Molnar, D., Widhalm, K., Moreno, L.A., Sjöström, M., Castillo, M.J., 2011. The international fitness scale (IFIS): usefulness of self-reported fitness in youth. Int. J. Epidemiol. 40 (3), 701–711. https://doi.org/10.1093/ije/dyr039.

Ortega, F.B., Sánchez-López, M., Solera-Martínez, M., Fernández-Sánchez, A., Sjöström, M., Martínez-Vizcaino, V., 2013. Self-reported and measured cardiorespiratory fitness similarly predict cardiovascular disease risk in young adults. Scand. J. Med. Sci. Sports 23 (6), 749–757. https://doi.org/10.1111/j.1600-0838.2012.01454.x.

Pearson, E., Siskind, D., Hubbard, R.E., Gordon, E.H., Coulson, E.J., Warren, N., 2022. Frailty and severe mental illness: A systematic review and narrative synthesis. J. Psychiatr. Res. 147, 166–175. https://doi.org/10.1016/j.jpsychires.2022.01.014.

Pereira, D.A., Correia Júnior, J.L., Carvas Junior, N., Freitas-Dias, R., 2020. Reliability of questionnaire The International Fitness Scale: a systematic review and meta-analysis. Einstein *(Sao Paulo, Brazil)* 18, eRW5232. https://doi.org/10.31744/einstein_journal/2020rw5232.

Perez-Cruzado, D., Cuesta-Vargas, A.I., Vera-Garcia, E., Mayoral-Cleries, F., 2018. The relationship between quality of life and physical fitness in people with severe mental illness. Health Qual. Life Outcomes 16 (1), 82. https://doi.org/10.1186/s12955-018-0909-8.

Ramírez-Vélez, R., Cruz-Salazar, S.M., Martínez, M., Cadore, E.L., Alonso-Martinez, A.M., Correa-Bautista, J.E., Izquierdo, M., Ortega, F.B., García-Hermoso, A., 2017. Construct validity and test–retest reliability of the international fitness scale (IFIS) in Colombian children and adolescents aged 9–17.9 years: the FUPRECOL study. PeerJ 5, e3351. https://doi.org/10.7717/peerj.3351.

Rikli, R.E., Jones, C.J., 1999. Development and validation of a functional fitness test for community-residing older adults. J. Aging Phys. Act. 7 (2), 129–161. https://doi.org/10.1123/japa.7.2.129.

Romero-Gallardo, L., Soriano-Maldonado, A., Ocón-Hernández, O., Acosta-Manzano, P., Coll-Risco, I., Borges-Cosic, M., Ortega, F.B., Aparicio, V.A., 2020. International fitness scale—IFIS: validity and association with health-related quality of life in pregnant women. Scand. J. Med. Sci. Sports 30 (3), 505–514. https://doi.org/10.1111/sms.13584.

Ross, R., Blair, S.N., Arena, R., Church, T.S., Després, J.-P., Franklin, B.A., Haskell, W.L., Kaminsky, L.A., Levine, B.D., Lavie, C.J., Myers, J., Niebauer, J., Sallis, R., Sawada, S.S., Sui, X., Wisløff, U., 2016. Importance of assessing cardiorespiratory fitness in clinical practice: A case for fitness as a clinical vital sign: A scientific statement from the American Heart Association. Circulation 134 (24). https://doi.org/10.1161/CIR.0000000000000461.

Sanchez-Lastra, M.A., Varela, S., Martínez-Aldao, D., Ayán, C., 2021. Questionnaires for assessing self-perceived physical fitness: A systematic review. Exp. Gerontol. 152, 111463. https://doi.org/10.1016/j.exger.2021.111463.

Sánchez-López, M., García-Hermoso, A., Ortega, F.B., Moliner-Urdiales, D., Labayen, I., Castro-Piñero, J., Benito, P.J., Vicente-Rodríguez, G., Sanchis-Moysi, J., Cantallops, J., Artero, E.G., Martínez-Vizcaíno, V., 2023. Validity and reliability of the international fitness scale (IFIS) in preschool children. Eur. J. Sport Sci. 23 (5), 818–828. https://doi.org/10.1080/17461391.2022.2049884.

Sánchez-López, M., Martínez-Vizcaíno, V., García-Hermoso, A., Jiménez-Pavón, D., Ortega, F.B., 2015. Construct validity and test-retest reliability of the international fitness scale (IFIS) in Spanish children aged 9-12 years. Scand. J. Med. Sci. Sports 25 (4), 543–551. https://doi.org/10.1111/sms.12267.

Scheewe, T.W., Jörg, F., Takken, T., Deenik, J., Vancampfort, D., Backx, F.J.G., Cahn, W., 2019. Low physical activity and cardiorespiratory fitness in people with schizophrenia: A comparison with matched healthy controls and associations with mental and physical health. Front. Psychol. 10. https://doi.org/10.3389/fpsyt.2019.00087.

Sheffield, J.M., Karcher, N.R., Barch, D.M., 2018. Cognitive deficits in psychotic disorders: A lifespan perspective. Neuropsychol. Rev. 28 (4), 509–533. https://doi.org/10.1007/s11065-018-9388-2.

Streiner, D.L., Kottner, J., 2014. Recommendations for reporting the results of studies of instrument and scale development and testing. J. Adv. Nurs. 70 (9), 1970–1979. https://doi.org/10.1111/jan.12402.

Tomkinson, G.R., Lang, J.J., Rubín, L., McGrath, R., Gower, B., Boyle, T., Klug, M.G., Mayhew, A.J., Blake, H.T., Ortega, F.B., Cadenas-Sanchez, C., Magnussen, C.G., Fraser, B.J., Kidokoro, T., Liu, Y., Christensen, K., Leong, D.P., Aadahl, M., Abdin, E., Yu, R., 2025. International norms for adult handgrip strength: a systematic review of data on 2.4 million adults aged 20 to 100+ years from 69 countries and regions. J. Sport Health Sci. 14, 101014. https://doi.org/10.1016/j.jshs.2024.101014.

Trautmann, S., Rehm, J., Wittchen, H., 2016. The economic costs of mental disorders. EMBO Rep. 17 (9), 1245–1249. https://doi.org/10.15252/embr.201642951.

Uanhoro, J.O., 2020, September 1. Effect Size Calculators. Https://Effect-Sizecalculator.Herokuapp.Com.

Uanhoro, J.O., 2020, September 10. Cohens d calculators. Https://Github.Com/Stonegold546/Cohens_d_calculators/Blob/Master/README.Md.

Vancampfort, D., Firth, J., Schuch, F.B., Rosenbaum, S., Mugisha, J., Hallgren, M., Probst, M., Ward, P.B., Gaughran, F., De Hert, M., Carvalho, A.F., Stubbs, B., 2017a. 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 16 (3), 308–315. https://doi.org/10.1002/wps.20458.

Vancampfort, D., Rosenbaum, S., Schuch, F., Ward, P.B., Richards, J., Mugisha, J., Probst, M., Stubbs, B., 2017b. Cardiorespiratory fitness in severe mental illness: A systematic review and Meta-analysis. Sports Med. 47 (2), 343–352. https://doi.org/10.1007/s40279-016-0574-1.

Vancampfort, D., Stubbs, B., Sienaert, P., Wyckaert, S., De Hert, M., Soundy, A., Probst, M., 2016. A comparison of physical fitness in patients with bipolar disorder, schizophrenia and healthy controls. Disabil. Rehabil. 38 (20), 2047–2051. https://doi.org/10.3109/09638288.2015.1114037.

Walker, E.R., McGee, R.E., Druss, B.G., 2015. Mortality in mental disorders and global disease burden implications. JAMA Psychiatry 72 (4), 334. https://doi.org/10.1001/jamapsychiatry.2014.2502.

Weye, N., Santomauro, D.F., Agerbo, E., Christensen, M.K., Iburg, K.M., Momen, N.C., Mortensen, P.B., Pedersen, C.B., Whiteford, H.A., McGrath, J.J., Plana-Ripoll, O., 2021. Register-based metrics of years lived with disability associated with mental

and substance use disorders: a register-based cohort study in Denmark. Lancet Psychiatry 8 (4), 310–319. https://doi.org/10.1016/S2215-0366(21)00029-8.
Wykes, T., Haro, J.M., Belli, S.R., Obradors-Tarragó, C., Arango, C., Ayuso-Mateos, J.L., Bitter, I., Brunn, M., Chevreul, K., Demotes-Mainard, J., Elfeddali, I., Evans-Lacko, S., Fiorillo, A., Forsman, A.K., Hazo, J.-B., Kuepper, R., Knappe, S., Leboyer, M., Lewis, S.W., Wittchen, H.-U., 2015. Mental health research priorities for Europe. Lancet Psychiatry 2 (11), 1036–1042. https://doi.org/10.1016/S2215-0366(15)00332-6.
Zechner, M.R., Anderson, E.Z., Gill, K.J., 2022. Old before their time: comparisons of people with SMI and healthy older adults. Psychiatr. Rehabil. J. 45 (1), 89–94. https://doi.org/10.1037/prj0000496.