



Accuracy of psychometric tools in the assessment of personality in adolescents and adults requesting gender-affirming treatments: a systematic review

Lehmann, K., & Leavey, G. (2019). Accuracy of psychometric tools in the assessment of personality in adolescents and adults requesting gender-affirming treatments: a systematic review. *European Psychiatry*, 62, 60-67. Advance online publication. <https://doi.org/10.1016/j.eurpsy.2019.09.004>

[Link to publication record in Ulster University Research Portal](#)

Published in:
European Psychiatry

Publication Status:
Published (in print/issue): 31/10/2019

DOI:
[10.1016/j.eurpsy.2019.09.004](https://doi.org/10.1016/j.eurpsy.2019.09.004)

Document Version
Author Accepted version

General rights
Copyright for the publications made accessible via Ulster University's Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Ulster University's institutional repository that provides access to Ulster's research outputs. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact pure-support@ulster.ac.uk.

Accuracy of psychometric tools in the assessment of personality in adolescents and adults requesting gender-affirming treatments: a systematic review

Katrin Lehmann^{a*} and Professor Gerard Leavey^b

^a*Bamford Centre for Mental Health, Ulster University, Belfast, Northern Ireland;*

^b*Bamford Centre for Mental Health, Ulster University, Belfast, Northern Ireland*

Corresponding author: Katrin Lehmann, Bamford Centre for Mental Health & Wellbeing, Ulster University, Jordanstown Campus, Shore Road, Newtownabbey, County Antrim, BT37 0QB, Northern Ireland, e-mail: lehmann-k@ulster.ac.uk

Background: The assessment and screening for personality disorders in individuals requesting gender affirming treatments may be an important aspect of predicting medical and surgical outcomes for this population, but there is no consensus on how best to do so.

Aims: To review the diagnostic accuracy of psychometric tools used for the assessment of personality disorders in those requesting gender affirming treatments.

Method: A systematic review: Prospero CRD42017078783 [1].

Results: Many studies have focussed on the assessment of personality disorders in this population, but since 1979, only two have used an index and reference test.

Conclusion: There are no agreed reference standards for this population and psychometric tools continue to be scored on reference data from the cisgender (not transgender) population. We need robust evidence on this issue, as individuals may be denied access to gender affirming treatments based on psychometric tools without established reliability in this population.

Introduction

Transgender people (also called trans people) experience an incongruence between their birth assigned sex and their own sense of gender identity [2]. Incongruence can be social, - how others perceive the person based on their birth assigned sex and/ or physical - between an individual's self-identity and their primary or secondary sex characteristics [2]. Trans people living in a largely gender-binary society experience significant levels of social rejection, discrimination, often resulting in poor physical and mental health [3]. Transgender people are a very diverse group, which includes individuals who live with their gender incongruence without transition, others who decide on social transition only without accessing specialist gender services and individuals who purchase their own hormones online [2]. This creates challenges in relation to estimating population sizes and studies on the experiences of individuals who identify as transgender, gender non-conforming or gender-questioning are usually focused on gender affirming treatments from health services, mostly because these individuals are reached relatively easily. Published population studies focused on questioning participants from the general population about their identity report estimates of 0.5% [4] to 1.3% [5] for those who are male assigned at birth, while estimates of 0.4% [4] to 1.2% [5] are reported for those who are female assigned at birth. Using the lower estimates of these studies as an overall mean and extrapolating these figures to a global population of 5.1 billion, [2] calculate a figure of 25 million transgender individuals worldwide.

Gender services across Great Britain have seen a 240% increase in gender-affirming treatment referrals over the past five years [6], similar to trends in most developed countries. Despite this, there is no consensus on optimal assessment of psychological functioning and mental health of individuals requesting gender affirming

treatments [7]. The assessment of personality disorders for those seeking gender-affirming treatment is regarded by some as critical to treatment planning and prognosis on the outcome of medical and surgical interventions [8].

Individuals who experience distress related to their sex assigned at birth can access specialist gender services for psychological, medical and/or surgical interventions. According to the current World Professional Association for Transgender Health (WPATH) guidance [9], anyone seeking gender-affirming treatments (for example hormones or surgical procedures) should complete a comprehensive assessment process which includes a psychological/ psychiatric assessment [10]. Psychiatric assessment in this population has historically consisted of detailed clinical interviews.

Gender Dysphoria

Gender dysphoria is considered a psychiatric diagnosis as part of the current version of the Diagnostic and Statistical Manual of Mental Disorders (DSM) [11] and International Classification of Diseases (ICD-10) [12]. This is the subject of ongoing debates attracting various opinions; from welcoming mental health professionals as part of responsible transgender treatments [13] to removing gender dysphoria from the category of psychiatric diagnoses [14]. Proposed changes in the ICD-11 include a change in terminology from transsexualism categorised under the mental health section to gender incongruence of adolescence and adulthood under the categories of sexual health [15].

Personality Assessment

Personality incorporates traits or constructs that differentiate individuals from others [16], as well as intra-psychic processes which enable individuals to achieve valued and need-fulfilling life-tasks [17]. These life tasks include creating a working model of the self, relating well to other people and reaching occupational goals [16].

The assessment of personality and personality traits using questionnaires for this purpose is undertaken in a variety of occupational, legal and clinical settings. In clinical environments, standardised personality assessments can be a useful guide for selecting treatments for individuals based on their personality traits. For example, cognitive behaviour therapy may be a treatment option more suitable to some individuals with specific personality traits while others might prefer a different modality.

Personality Disorder Assessment

Difficulties in relation to achieving adult life tasks are often associated with personality pathology [18]. Personality disorders are not simply defined as clinically significant extreme personality traits, rather the mechanism that is dysfunctional is stopping the individual from functioning adaptively in society [19]. Personality disorders are pervasive and ingrained [11], however more recent evidence highlights that personality status can change in response to treatment [20]. Since the formulation of DSM-III [11], personality disorders were given a separate axis in the classification which consisted of 11 categories. This categorical system of classifying personality disorders using heterogenous descriptions did not work well in practice and dimensional models based on personality traits rather than behaviours were considered [20]. A dimensional system considers personality on a continuum with normal variations on one end and what is considered a personality disorder at the extreme end of the continuum [20]. The assessment of personality disorders is focused on the functional impact and has evolved over time from a categorical approach in which clinicians attempted to match individuals to multiple categories to a more dimensional approach but agreement between different assessment approaches continues to be an issue. Other concerns related to the stability of current assessment methods, definitions of severity of personality disorders and information sources used for the assessment are unresolved

[20]. Personality is generally assessed using a combination of self-report questionnaires and a structured clinical interview [20]. There are numerous available clinical interview schedules and questionnaires for the assessment of personality disorders, but unfortunately cross-instrument reliability is very poor, which is largely related to the criteria for different personality disorder diagnoses and overlap between categories, often leading to multiple diagnoses [20]. The reported prevalence of personality disorders in those seeking gender-affirming treatments ranges from 4.3% [21] to 81.4% [22], perhaps reflecting the disparate cultural contexts (Italian and Iranian), social norms and inherent difficulties in conducting diagnostic assessments for personality disorders. The diagnosis of personality disorders remains controversial [23], argued by some as carrying more stigma and shame than for other psychiatric diagnoses [24, 25] and consequently impacting on service access [26, 27, 28]. Under the current World Professional Transgender Health (WPATH) guidance [9], access to gender-affirming treatments is the treatment of choice for individuals experiencing gender related distress. While most individuals experience positive changes following hormonal and or surgical interventions, between 1 and 2% of individuals expressed regret and a further 1% attempt suicide [29]. Some studies have delayed or excluded individuals with significant psychopathology from accessing hormonal or surgical interventions [30], while others have suggested a link between post-operative dissatisfaction and pre-operative psychopathology [31].

Objectives

Psychometric assessment tools are commonly used in clinical practice to assess personality in individuals requesting gender-affirming care. However, no formal guidelines for interpreting test data for transgender individuals exist [7]. Other authors

[32, 33] suggest that minority individuals can be overly pathologized in psychometric tests using normative data. Some of the psychometric assessment tools, for example the Minnesota Multiphasic Personality Inventory (MMPI-2) [34], are scored on gender-based norms. It is unclear whether individuals should be evaluated based on their assigned sex at birth, self-identified gender or both and further questions arise for individuals identifying outside the gender binary. To date, the accuracy of psychometric assessment tools for personality disorder assessment in this population has never been examined in a systematic review. In view of concerns about personality disorders as poor prognostic factors and potential reasons for denying access to gender-affirming treatments, we conducted a systematic review.

Aims:

To review:

- What psychometric tests (interventions) and reference tests (comparisons) are used to diagnose personality disorders (outcome) in adolescents and adults who request gender-affirming treatments (population)?
- How accurate are psychometric tests for diagnosing personality disorders compared to reference tests in adolescents and adults requesting gender-affirming treatment.

Method

Protocol and registration

This review is based on the protocol ‘Accuracy of psychometric tools in the assessment of personality in adolescents and adults requesting gender realignment:

protocol for a systematic review [1]. The protocol is accessible via Prospero International prospective register of systematic review: CRD42017078783; available from: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID42017078783.

Eligibility criteria

Diagnostic accuracy assessment studies are commonplace in physical medicine but much less so in mental health practice. Over the past decade the reporting of diagnostic accuracy studies has been under scrutiny and calls for greater transparency in the reporting of studies have led to the Standards for Reporting Diagnostic Accuracy Studies (STARD) [35]. In the case of this review, the STARD standards were not available at the time of reporting of included studies. Accuracy broadly refers to the agreement between the test under study and the reference or standard test [36].

Accuracy refers to high sensitivity, the chances that the test outcome is positive in someone who has the condition, while specificity refers to the probability that the test outcome is negative in someone who does not have the condition [37]. The reference test is the gold standard test available to identify the outcome (personality disorder) in the population (adolescents and adults requesting gender affirming treatments). Tests with 100% sensitivity and specificity are very rare thus the term reference standard rather than gold standard is more appropriate [37]. A gold standard in personality disorder assessments is rare [20]. Challenges in assessment can be linked to the classification of personality disorders, poor inter-rater reliability in clinical assessments [38] and evidence of personality status as unstable [39] rather than fixed over a lifespan. And in the absence of other alternatives, clinical assessments were used as the reference test for the review. Psychometric tools for the assessment of personality disorders were used as index test in this review.

The timing of index and reference tests at various points of contact with specialist gender services (e.g. on assessment, following physical interventions) was considered in the review. Children up to the age of 12 years old were excluded from the review for several reasons. While children, adolescents and adults seek gender affirming treatments based on current WPATH guidance [9], the assessment of personality disorders tends to focus on adolescents and adults [40].

Information sources

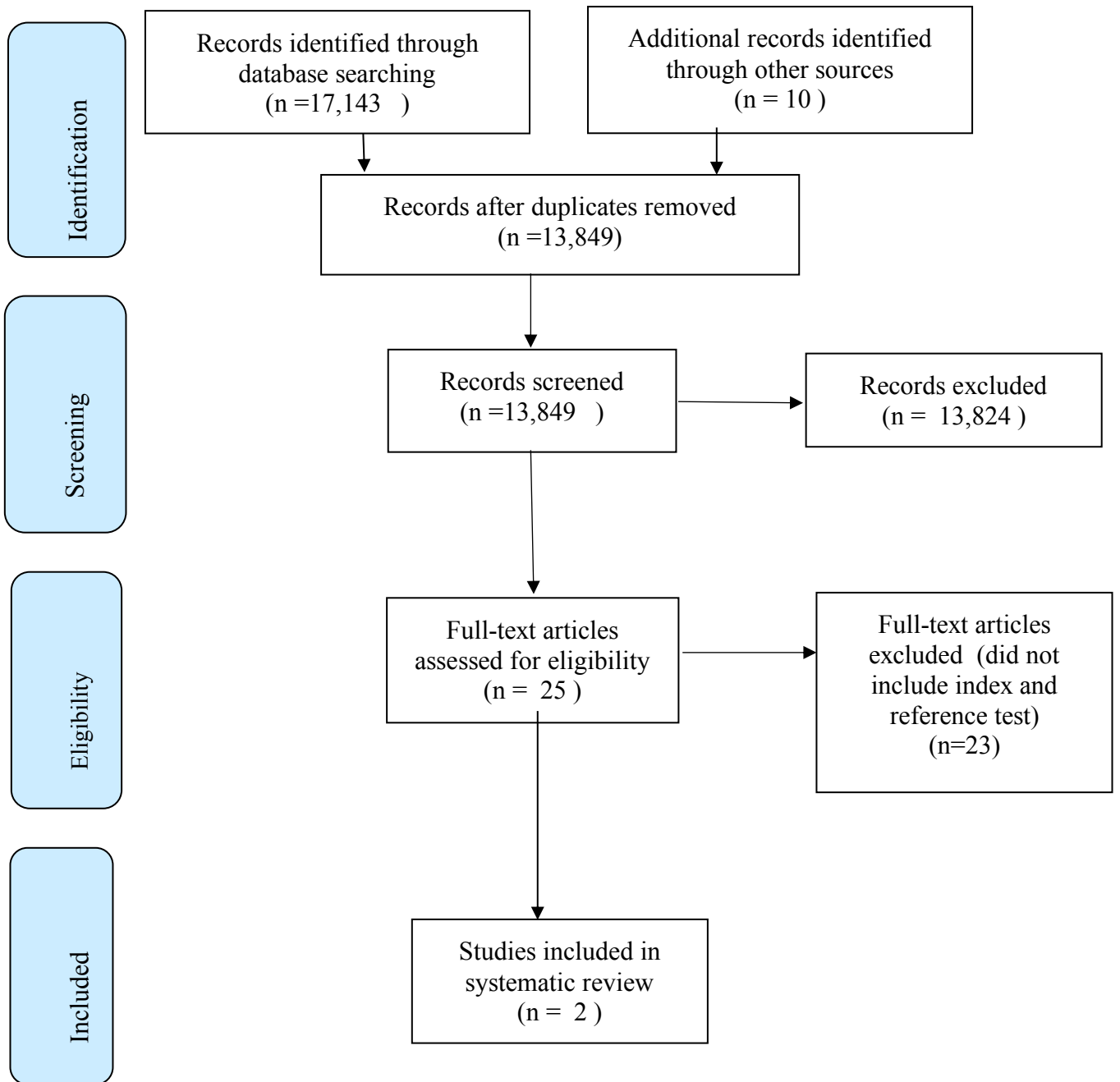
An electronic literature search was conducted using Ovid MEDLINE (1946 to December 2018) • Ovid MEDLINE In-Process & Other Non-Indexed Citations (December 2018), Embase (1980 to December 2018; Ovid), PsycINFO (1887 to December 2018; EBSCOhost), PsycARTICLES (1894 to December 2018; EBSCOhost) and Cochrane Database of Systematic Reviews (CDSR; latest issue, The Cochrane Library). The initial search strategy initially was adapted for individual databases. Searches were limited to the years following the publication of the first WPATH standards in 1979. Language limits were not applied to the searches and translations were sought where possible. As previous searches did not identify any studies clearly identified as diagnostic accuracy studies, a narrow search for study type was not possible.

Example of search strategy

- (1) exp Gender Identity/
- (2) exp Gender Dysphoria/
- (3) exp Transsexualism/
- (4) exp Transgender Persons/
- (5) gender variance mp.
- (6) gender fluid mp.
- (7) sex change mp.
- (8) gender change mp.
- (9) Gender identity mp.
- (10) Gender dysphoria mp.
- (11) Transsexualism mp.
- (12) Transgender mp.
- (13) 1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12
- (14) exp Mental Health/
- (15) exp Psychological Tests/
- (16) psychological needs.mp
- (17) exp Mental Health services/
- (18) Mental Health.mp
- (19) Psychological Test*. Mp
- (20) exp Personality/
- (21) exp personality disorder/
- (22) personality mp.
- (23) personality disorder mp. 24) 14 OR 15 OR 16 OR 17 OR 18 OR 19 OR 20 OR 21 OR 22 OR 23 25)
13 AND 24 26) exp animals/ not humans.sh 27) 25 NOT 26 28) limit (24) to (yr= "1979- Current" and
("all adult (19 plus years)" or "adolescent (13 to 18 years)"))



Table 1: PRISMA Flow Diagram [41]



Data collection process

Both review authors independently screened titles and abstracts returned through the searches against the inclusion criteria. Full text studies were obtained if the titles and abstract met the inclusion criteria or if there was uncertainty. Both reviewers screened full text articles and made decisions about the inclusion of the study. Additional information was sought from the authors if there were uncertainties regarding eligibility of the studies. One of the studies was translated from Dutch into English for further review but did not meet inclusion criteria on closer inspection. Both reviewers independently used a standardised tool to extract the information [42] and completed critical appraisal checklists and then compare findings. Any disagreements were resolved through further discussion.

Data items included inclusion/exclusion criteria for each study, sample size, participant demographics, study methodology, index test description, reference test description, geographical location of data collection, setting of data collection, persons executing and interpreting index tests, persons executing and interpreting reference tests and index/ reference time interval (and treatments carried out in between).

Risk of bias assessment

The QUADAS-2 revised tool [43] was used to assess the methodological quality of each included study in relation to risk of bias in the selection of patients, patient flow, the conduct and interpretation of the index test and the conduct and interpretation of the reference test. The QUADAS-2 tool was piloted prior to use. Both authors independently completed risk of bias assessments on each included study. Disagreements were resolved through further discussion.

Results

Study selection

Overall 13,849 studies were independently screened. By the screening of titles and abstracts 13,824 studies were excluded, while 25 studies were screened for further detailed analysis, from which 23 studies did not meet the eligibility criteria as the studies did not compare a psychometric tool for personality disorder assessment with an index test. Only two studies conducted in 1993 [44] and 2000 [45] met inclusion criteria for the review.

Table 2- Study characteristics of included studies

Author/Date	Bodlund O, Kullgren G, Sundbom E, Höjerback T. Personality traits and disorders among transsexuals. Acta Psychiatr Scand. 1993 Nov 1;88(5):322–7.	Miach PP, Berah EF, Butcher JN, Rouse S. Utility of the MMPI-2 in assessing gender dysphoric patients. Journal of Personality Assessment. 2000 Oct 1;75(2):268-79.
Inclusion/exclusion criteria	Applicants for SRS who met strict criteria related to mental and physical health and social capacity.	Consecutive male-to-female candidates referred for psychological assessment to determine suitability for SRS surgery who attended between May 1993 and October 1996.
Sample size	Index group- 19 applicants for SRS Control group 133 individuals	Index group- 48 individuals diagnosed as transsexual Control group- 34 individuals diagnosed as GIDAANT
Participant demographics	Index group (9 male sex assigned at birth; 10 female sex assigned at birth) Control group (81 women; 52 men)	Index group 48 male sex assigned at birth Control group 34- (male assigned at birth)
Study methodology	Cross sectional survey with comparison group	Cross sectional survey with comparison group
Index test description	SCID used to score number of fulfilled criteria below cut-off level. Total proportion was calculated as global index of personality pathology. Personality disorder diagnosis through combination of cut-off levels for axis II diagnosis (28-30) and global assessment and functioning (GAF) score of <70.	MMPI-2 (excluded if FT score =>90)
Reference test description	Clinical DSM-III-R evaluation (axes I and II) by two psychiatrists. This included several clinical interviews and clinical records.	Psychological assessment by two psychiatrists and a clinical psychologist after several clinical interviews.
Location of data collection	Sweden	Australia
Setting of data collection	University Hospital	Medical Centre GD clinic
Persons executing and interpreting index tests	Two psychiatrists	Two psychiatrists, one clinical psychologist, MMPI interpreted by third author, Minnesota
Persons executing and interpreting reference test	Two psychiatrists	Two psychiatrists, one clinical psychologist
Index/reference time interval (and treatments carried out in between)	Different stages of SRS process	After 6-9 months waiting period following acceptance to clinic.

Risk of bias assessment- study 1[44]

The first study used a total population sample of all individuals requesting gender-affirming care in a Swedish region of 2.5 million inhabitants. All participants met the then criteria for transsexualism to access gender affirming treatments under the then DSM-III-R [46]. Participants were administered index and reference tests at different stages of gender affirming care and included individuals who were established on hormone treatments and individuals who had already completed gender affirming surgical treatments. While it was believed that personality profiles remain stable over a lifetime [47], more recent studies have found positive changes in functioning following testosterone therapy [48]. Comparing individuals at different stages may have impacted on the results of the study. The study [44] used the Swedish version of the SCID screen [49]. The SCID includes 124 questions, with 103 criteria for the assessment of personality disorders. Authors used the SCID to score the number of fulfilled criteria to describe personality traits below cut-off level and to make diagnoses of personality disorder based on DSM-III-R [46] criteria. The authors set additional criteria for the diagnosis of personality disorder by combining cut-off levels for axis II diagnosis (28-30) with a global assessment of functioning (GAF) score [50] of <70. The global assessment of functioning was used in this context as an additional layer for diagnosis and measure for social and occupational functioning of participants. The Global Assessment of Functioning is a numerical scale completed by a clinician, which rates the social, psychological and everyday functioning of an individual [51]. The maximum score which can be attained is 100, indicating the best possible functioning. Two psychiatrists conducted several clinical interviews and reviewed clinical records before making a diagnosis related to personality difficulties based on the DSM-II-R [46]. Based on standards at the time, clinician diagnosis based on all available data [40] was

the accepted diagnostic process. It is unclear whether the two psychiatrists were aware of the results of the index test prior to conducting the reference test. Nineteen participants completed the index and reference standard tests, but it is unclear at what point the tests were completed. Three patients failed to meet other requirements for access to gender-affirming treatments (age, immaturity and alcohol problems) and were excluded. All participants who completed the index and reference test were included in the analysis. The study also included a control group who only completed the index test.

Risk of bias assessment- study 2[45]

The second study enrolled a consecutive sample of males assigned at birth who were referred to a national centre for assessment. It is not stated why only males assigned at birth were included in the study, but the authors differentiate between males assigned at birth meeting criteria for transsexualism under DSM-III-R [46] and those meeting non transsexual gender dysphoria (GIDAANT) under the same criteria. All participants completed index tests on admission to the service and reference tests were completed after 6 months waiting period following admission to the service. As the service accepts referrals for surgical gender affirming procedures it is unclear if any or all participants were already established on hormone treatments prior to admission to the service. The MMPI-2 [34] was administered to all 86 individuals at initial assessment. MMPI-2 sheets were scored and analysed independently by one of the authors in Minnesota. The MMPI-2 [34] was completed on admission to the centre while the reference standard was undertaken after a period of waiting following the admission. The threshold for scoring the MMPI-2 [34] was specified and those with a scale F T score of 90 or above were excluded from the study. The F Score is related to atypical responses on the MMPI-2 and a score of 90 or above creates questions in

relation to the truthfulness of responses [52]. Participants were scored against their sex assigned at birth. Two psychiatrists and one clinical psychologist independently assessed participants. During a subsequent team conference all information was reviewed and diagnoses in relation to a personality disorder were made based on DSM-III-R standards [46]. Disagreements were resolved through further team discussion and through collecting new information to aid the diagnostic process. This process appears to have been conducted independently from the index test, which was scored and processed elsewhere. Four patients were excluded from the data analysis due to MMPI-2 [34] scale F T scores of 90 or above. The MMPI-2 [34] includes F scale items to detect unusual and potentially untruthful ways of responding to test items [52]. The MMPI-2 [34] was conducted at admission to the programme, while psychological testing was conducted after completion of a waiting period of 6 to 9 months after admission to the programme with an appropriate interval between index and reference standard tests.

Table 3- Risk of Bias and Applicability Judgments based QUADAS-2 [43]

Study	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Study 1 [44]	?	😊	😊	😊	😊	😊	😊
Study 2 [45]	?	?	😊	😊	?	😊	😊

😊= low risk; 😞= high risk; ?=unclear risk

Table-4 Prevalence of personality disorders

Study	Sample	Test used	PD prevalence index test	PD prevalence reference test
Study 1 [44]	n=19 transsexuals	SCID & GAF	26.3%	36.8%
Study 2 [45]	n=48 transsexuals	MMPI-2	14.6%	29.2%
Study 2 [45]	n=34 gender identity non-transsexual type (GIDAANT)	MMPI-2	47.1%	64.7%

Table-5 Accuracy, Sensitivity and specificity of index tests compared to reference standards

Index test diagnosis	Reference diagnosis		Total	Accuracy	Sensitivity	Specificity
	(+)	(-)				
SCID & GAF	(+)	(-)		89%	72%	100%
(+)	5	0	5			
(-)	2	12	14			
Total	7	12	19			
MMPI-2 Transsexual group	(+)	(-)		85%	50%	100%
(+)	7	0	7			
(-)	7	34	41			
Total	14	34	48			
MMPI-2 GIDAANT group	(+)	(-)		82%	73%	100%
(+)	16	0	16			
(-)	6	12	18			
Total	22	12	34			

Prevalence of personality disorders

We only included results which were in keeping with the population in this review and therefore results related to a non-representative control group of individuals selected from the general population were not included [44]. In both studies, index tests detected less individuals with personality disorders compared with the reference standards. A significantly higher number of individuals in study 2 [45] belonging to the gender identity non-transsexual type group (GIDAANT) were reported to have personality disorders. Accurate diagnostic assessment can be made in the context of accepted criteria [37], which for both studies relate to a binary measure of absence or presence of personality disorders. Diagnosis of personality disorder was based on an agreed reference standard [46]. We calculated sensitivity and specificity for both studies. Both index tests were over 82% accurate in detecting personality disorders in the three groups. Neither of the index tests created false positives thus a diagnosis of personality disorder in individuals who do not have a personality disorder based on the reference test. While the SCID & GAF [44] and MMPI-2 [45] showed a sensitivity of over 72% in identifying true positives meaning individuals who are diagnosed with a personality disorder on index and reference test, the sensitivity of the MMPI-2 [45] in the transsexual group was only 50%. The interpretation of these results however needs to be completed in the context of other factors.

Discussion

This is the first systematic review on the assessment of personality disorders in gender reassignment. Given the rapid increase in people seeking gender-affirming treatment across Western, developed countries, it is perhaps alarming that we found only two studies and even these have considerable limitations.

In the first study [44], all participants met very stringent criteria for transsexualism [45], functioned well socially and showed no signs of severe mental illness. The authors acknowledge that the sample represents a carefully selected group of individuals as a large proportion requesting gender affirming treatments are excluded because of their mental health, physical conditions or poor social functioning [44]. There is no information on the baseline measures of all participants. It is unclear how many individuals were new to the service; how many had been taking hormonal interventions and how many were waiting or had completed their surgical transition. This is an important issue, as hormonal treatment has shown to improve functioning in other studies [48].

The second study [45] used a representative sample, but only focused on males assigned at birth. This relates to the study differentiating between males assigned at birth meeting transsexual criteria and those meeting gender identity disorder non-transsexual type (GIDAANT) criteria [46]. The difference between criteria for diagnosis of transsexual compared to GIDAANT is largely related to persistent pre-occupation for at least two years with wanting to get rid of sex characteristics assigned at birth [46]. Those meeting GIDAANT criteria, met criteria for discomfort related to sex assigned at birth without the desire to want rid of sex characteristics assigned at birth.

The SCID screen used in the first study [44] is a self-report questionnaire containing 124 questions requiring yes or no responses. Participants were at different stages of their gender affirming care when they completed the SCID screen. It is difficult to know how the timing of the SCID screen impacted on the results. The addition of the Global Assessment of Functioning (GAF) criteria was used to reduce over inclusiveness in the diagnosis of personality disorder. While the GAF has been described as a valid and reliable instrument [53] it has been excluded from the most

recent DSM-5 edition as an inadequate instrument to assess psychiatric functional impairment [54]. The exclusion was based on lack of conceptual clarity of the GAF, questionable psychometric properties of the GAF and its reliance on appropriate training of the clinician to ensure reliability and validity of the instrument [54]. In the GAF any impairments in functioning related to physical or environmental factors are not considered [54]. The GAF's conflation of symptom severity is often not congruent with levels of experienced impairment of functioning [54]. This could have created GAF scores < 70 suggestive of significant impairment when this was not actually the case. Furthermore, physical factors due to gender affirming treatments or stressful environmental factors which may have been experienced by some individuals were not reflected in the scoring system. It is therefore difficult to know how stage of gender affirming care impacted on the GAF score.

The MMPI-2 [34], a 567-item questionnaire administered by a clinician is the most widely used questionnaire for the assessment of personality [47], however it is viewed as a poor assessment tool for the assessment of personality disorders [55]. It is unclear why the MMPI-2 [34] was chosen for this was study other than that it was part of the assessment procedure of the clinic at the time. The MMPI-2 [34] was scored independently based on male normative data. Other research has highlighted that those in the earlier phase of their transition process show higher scores on the MMPI-2 [56], with testosterone treatment reducing MMPI-2 scores in one study [7]. This contrasts with prior studies which suggested that MMPI-2 results remain stable over time even in individuals who complete intensive psychotherapy[47]. The MMPI-2 [34] is based on male or female normative data. Male and female norms for the MMPI-2 [34] were derived from a representative sample of the cisgender (not transgender) population and thus normative data for those requesting gender affirming care does not exist. It has

been suggested that cultural variables can impact MMPI-2 scores [7] with elevation to the psychopathic deviate scale caused by lack of acceptance of transgender people in society [57] or even experienced transphobia.

Reference tests in both studies consisted of detailed psychological and psychiatric assessments by a team of clinicians. It is unclear whether clinicians in the first study [44] were aware of the outcome of the index test prior to conducting the reference test. Given the small total sample in this study clinicians may have already been aware of the absence or presence of personality difficulties as some participants were attending the centre for many years. In the second study [45], the reference test diagnosis was based on a team consensus approach. There is a clear time interval between index and reference test. While the index test was processed and scored elsewhere it is unclear if the clinical team were aware of the results of the index test prior to conducting the reference test.

To establish the sensitivity and specificity of a diagnostic test, the prevalence of the disorder needs to be considered for sample size calculations of cases and controls to be undertaken [36]. Without this knowledge it is impossible to determine whether the study population is representative of the population to which the test will be applied. It is likely that much larger sample sizes including cases and controls would have been required in both studies. While this might be true from a statistical point of view, both studies included the total sample of individuals known to their respective services at a point in time. This clearly creates a range of difficulties for anyone trying to conduct accuracy assessments of psychometric tools in this population. The prevalence of personality disorders in the population was unknown and estimates from other studies range widely [21,22]. Even if it was possible to calculate sample sizes of cases and

controls based on prevalence figures, it may be impossible to recruit enough cases for future studies.

Conclusion

Reference test assessment increased the prevalence of personality disorder in both studies. While personality traits are believed to be stable over many years, it is unclear whether other factors, such as prior exposure to gender affirming treatments, cultural variables or experience of transphobia could have impacted on the index test results of both studies.

Psychometric tools continue to be used to assess personality disorders in this population, despite the absence of normative data for scoring and comparative reference tests. Thus, individuals may be excluded from accessing gender affirming treatments based on clinical practice which does not have any evidence base. There is a clear gap in our current knowledge related to the reliability of psychometric assessment tools in this population. Future studies looking at the accuracy of psychometric assessment tools require larger sample sizes and knowledge of prevalence rates of personality disorders in this population. Tests also need to be developed based on normative data for the transgender not cisgender population. The development of any new normative data in this population will be very complex due to intersectionality. Individuals requesting gender affirming treatments are not a homogenous group and may belong to multiple marginalised groups for example due to their gender identity, ethnicity, sexuality or disability [58]. Without further research and understanding of intersectionality and cultural variables impacting on personality assessments in this population we are at risk of marginalising individuals even further.

Acknowledgments:

Authorship Contribution

All persons (Katrin Lehmann, Professor Gerard Leavey) who meet authorship criteria are listed as authors, and all authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the design of the systematic review, data collection process, analysis and interpretation, drafting and revision of this article.

Funding

Katrin Lehmann is funded through the Public Health Agency Northern Ireland Research and Development Fellowship award to undertake a PhD. No additional funding was obtained for this review.

Conflict of interest

The authors have no conflict of interest to declare.

References:

1. Lehmann K, Leavey G. Accuracy of psychometric tools in the assessment of personality in adolescents and adults requesting gender alignment: protocol for a systematic review. (2017) PROSPERO CRD42017078783
2. Winter S, Diamond M, Green J, Karasic D, Reed T, Whittle S, Wylie K. Transgender people: health at the margins of society. *The Lancet*. 2016 Jul 23;388(10042):390-400.
3. Bockting WO, Miner MH, Swinburne Romine RE, Hamilton A, Coleman E. Stigma, mental health, and resilience in an online sample of the US transgender population. *Am J Public Health*. 2013 May;103(5):943–51.
4. Conron KJ, Scott G, Stowell GS, Landers SJ. Transgender health in Massachusetts: Results from a household probability sample of adults. *Am J Public Health* 2012; 102: 118–22.

5. Clark T, Lucassen M, Bullen M, et al. The health and well-being of transgender high school students: results from the New Zealand adolescent health survey (Youth'12). *J Adolesc Health* 2014; 55: 93–99.
6. Torjesen I. Trans health needs more and better services: increasing capacity, expertise, and integration. *BMJ*. 2018 Aug 8;362:k3371.
7. Keo-Meier CL, Fitzgerald KM. Affirmative Psychological Testing and Neurocognitive Assessment with Transgender Adults. *Psychiatr Clin North Am*. 2017 Mar;40(1):51–64.
8. Duisin D, Batinic B, Barisic J, Djordjevic, ML, Vujovic, S, Bizic, M. Personality Disorders in Persons with Gender Identity Disorder. *Scient World Journal*. 2014; 1-7.
9. Coleman E, Bockting W, Botzer M, Cohen-Kettenis P, DeCuypere G, Feldman J, et al. Standards of Care for the Health of Transsexual, Transgender, and Gender-Nonconforming People, Version 7. *Int J Transgenderism*. 2012 Aug 1;13(4):165–232.
10. Fraser L. Psychotherapy in the World Professional Association for Transgender Health's Standards of Care : Background and Recommendations. *Int J Transgenderism*. 2009 Jul 20;11(2):110–26.
11. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Association; 2013.
12. World Health Organisation. *Manual of the International Statistical Classification of Diseases, Injuries and Causes of Death*. 10th revision. WHO; 1992.
13. Selvaggi G, Giordano S. The Role of Mental Health Professionals in Gender Reassignment Surgeries: Unjust Discrimination or Responsible Care? *Aesthetic Plast Surg*. 2014 Dec 25;38(6):1177–83.
14. Richards C, Arcelus J, Barrett J, Bouman WP, Lenihan P, Lorimer S, et al. Trans is not a disorder – but should still receive funding. *Sex Relatsh Ther*. 2015 Jul 3;30(3):309–13.
15. Thomas R, Pega F, Khosla R, Verster A, Hana T, Say L. Ensuring an inclusive global health agenda for transgender people. *Bull World Health Organ*. 2017;95(2):154–6.
16. Krueger RF, Skodol AE, Livesley WJ, ShROUT PE, Huang Y. Synthesizing dimensional and categorical approaches to personality disorders: refining the research agenda for DSM-V Axis II. *International Journal of Methods in Psychiatric Research*. 2007 Jun;16(S1):S65-73.
17. Westen D, Shedler J, Bradley R. A prototype approach to personality disorder diagnosis. *American Journal of psychiatry*. 2006 May;163(5):846-56.
18. Skodol AE, Gunderson JG, Shea MT, McGlashan TH, Morey LC, Sanislow CA, Bender DS, Grilo CM, Zanarini MC, Yen S, Pagano ME. The collaborative longitudinal personality disorders study (CLPS): Overview and implications. *Journal of personality disorders*. 2005 Oct 1;19(5):487-504.
19. Livesley WJ, Jang KL. Differentiating normal, abnormal, and disordered personality. *European Journal of Personality: Published for the European Association of Personality Psychology*. 2005 Jun;19(4):257-68.
20. Tyrer P, Coombs N, Ibrahim F, Mathilakath A, Bajaj P, Ranger M, Rao B, Din R. Critical developments in the assessment of personality disorder. *The British Journal of Psychiatry*. 2007 May;190(S49):s51-9.
21. Fisher AD, Bandini E, Casale H, Ferruccio N, Meriggiola MC, Gualerzi A, et al. Sociodemographic and Clinical Features of Gender Identity Disorder: An Italian Multicentric Evaluation. *J Sex Med*. 2013 Feb;10(2):408–19.

22. Meybodi AM, Hajebi A, Jolfaei AG. The frequency of personality disorders in patients with gender identity disorder. *Med J Islam Repub Iran*. 2014;28:90.
23. Lewis KL, Grenyer BFS. Borderline Personality or Complex Posttraumatic Stress Disorder? An Update on the Controversy. *Harv Rev Psychiatry*. 2009 Sep;17(5):322–8.
24. Catthoor K, Feenstra DJ, Hutsebaut J, Schrijvers D, Sabbe B. Adolescents with personality disorders suffer from severe psychiatric stigma: evidence from a sample of 131 patients. *Adolesc Health Med Ther*. 2015;6:81–9.
25. Magallón-Neri E, Forns M, Canalda G, De la Fuente JE. 542 – Stigmatization, personality disorders and adolescence. *Eur Psychiatry*. 2013 Jan 1;28:1.
26. Adebawale, Lord. Personality disorder: taking a person-centred approach. *Ment Heal Rev J*. 2010 Dec 14;15(4):6–9.
27. Bodner E, Cohen-Fridel S, Mashiah M, Segal M, Grinshpoon A, Fischel T, et al. The attitudes of psychiatric hospital staff toward hospitalization and treatment of patients with borderline personality disorder. *BMC Psychiatry*. 2015 Dec 22;15(1):2.
28. Latalova K, Ociskova M, Prasko J, Sedlackova Z, Kamaradova D. If You Label Me, Go with Your Therapy Somewhere! Borderline Personality Disorder and Stigma. *Eur Psychiatry*. 2015 Mar;30:1520.
29. Michel A, Anseau M, Legros J., Pitchot W, Mormont C. The transsexual: what about the future? *Eur Psychiatry*. 2002 Oct 1;17(6):353–62.
30. Smith YL, Van Goozen SH, Cohen-Kettenis PT. Adolescents with gender identity disorder who were accepted or rejected for sex reassignment surgery: a prospective follow-up study. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2001 Apr 1;40(4):472-81.
31. De Cuyper G, Elaut E, Heylens G, Van Maele G, Selvaggi G, T’Sjoen G, et al. Long-term follow-up: psychosocial outcome of Belgian transsexuals after sex reassignment surgery. *Sexologies*. 2006 Apr;15(2):126–33.
32. Campbell Jr AL, Ocampo C, Rorie KD, Lewis S, Combs S, Ford-Booker P, Briscoe J, Lewis-Jack O, Brown A, Wood D, Dennis G. Caveats in the neuropsychological assessment of African Americans. *Journal of the National Medical Association*. 2002 Jul;94(7):591.
33. Heaton RK, Taylor MJ, Manly J. Demographic effects and use of demographically corrected norms with the WAIS-III and WMS-III. In *Clinical interpretation of the WAIS-III and WMS-III* 2003 Jan 1 (pp. 181-210). Academic Press.
34. Butcher JN. *Minnesota Multiphasic Personality Inventory-2: Manual for administration, scoring, and interpretation*. University of Minnesota Press; 2001.
35. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, Lijmer JG, Moher D, Rennie D, De Vet HC, Kressel HY. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *Radiology*. 2015 Oct 28;277(3):826-32.
36. Flahault A, Cadilhac M, Thomas G. Sample size calculation should be performed for design accuracy in diagnostic test studies. *Journal of clinical epidemiology*. 2005 Aug 1;58(8):859-62.
37. Knottnerus JA, Muris JW. Assessment of the accuracy of diagnostic tests: the cross-sectional study. *Journal of clinical epidemiology*. 2003 Nov 1;56(11):1118-28.

38. Livesley WJ, Larstone R, editors. Handbook of personality disorders: Theory, research, and treatment. Guilford Publications; 2018 Feb 15.
39. Shea MT, Stout R, Gunderson J, Morey LC, Grilo CM, McGlashan T, Skodol AE, Dolan-Sewell R, Dyck I, Zanarini MC, Keller MB. Short-term diagnostic stability of schizotypal, borderline, avoidant, and obsessive-compulsive personality disorders. *American Journal of Psychiatry*. 2002 Dec 1;159(12):2036-41.
40. Skodol AE, Johnson JG, Cohen P, Sneed JR, Crawford TN. Personality disorder and impaired functioning from adolescence to adulthood. *The British Journal of Psychiatry*. 2007 May;190(5):415-20.
41. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine*. 2009 Aug 18;151(4):264-9.
42. Campbell JM, Klugar M, Ding S, Carmody DP, Hakonsen SJ, Jadotte YT, White S, Munn Z. Diagnostic test accuracy: methods for systematic review and meta-analysis. *International journal of evidence-based healthcare*. 2015 Sep 1;13(3):154-62.
43. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of internal medicine*. 2011 Oct 18;155(8):529-36.
44. Bodlund O, Kullgren G, Sundbom E, Höjerback T. Personality traits and disorders among transsexuals. *Acta Psychiatr Scand*. 1993 Nov 1;88(5):322-7.
45. Miach PP, Berah EF, Butcher JN, Rouse S. Utility of the MMPI-2 in assessing gender dysphoric patients. *Journal of Personality Assessment*. 2000 Oct 1;75(2):268-79.
46. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Association; 1994.
47. Spiro A, Butcher JN, Levenson RM, Aldwin CM, Bosse R. Change and stability in personality: A 5-year study of the MMPI-2 in older men. *Basic sources for the MMPI-2*. 2000:443-63.
48. Keo-Meier CL, Herman LI, Reisner SL, Pardo ST, Sharp C, Babcock JC. Testosterone treatment and MMPI-2 improvement in transgender men: A prospective controlled study. *Journal of consulting and clinical psychology*. 2015 Feb;83(1):143.
49. Ekselius L, Lindström E, von Knorring L, Bodlund O, Kullgren G. SCID II interviews and the SCID Screen questionnaire as diagnostic tools for personality disorders in DSM-III-R. *Acta Psychiatrica Scandinavica*. 1994 Aug;90(2):120-3.
50. Jones SH, Thornicroft G, Coffey M, Dunn G. A brief mental health outcome scale: Reliability and validity of the Global Assessment of Functioning (GAF). *The British Journal of Psychiatry*. 1995 May;166(5):654-9.
51. Hall RC. Global assessment of functioning: a modified scale. *Psychosomatics*. 1995 May 1;36(3):267-75.
52. Wygant DB, Sellbom M, Ben-Porath YS, Stafford KP, Freeman DB, Heilbronner RL. The relation between symptom validity testing and MMPI-2 scores as a function of forensic evaluation context. *Archives of Clinical Neuropsychology*. 2007 May 1;22(4):489-99.
53. Von Korff M, Andrews G, Delves M: Assessing activity limitations and disability among adults, in *The Conceptual Evolution of DSM-5*. Edited by

- Regier DA, Narrow WE, Kuhl EA, et al.
WashingtonDC:AmericanPsychiatricPublishing,Inc.,2011,pp163–88.
54. Gold LH. DSM-5 and the assessment of functioning: The World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0). *Journal of the American Academy of Psychiatry and the Law Online*. 2014 Jun 1;42(2):173-81.
 55. Derksen JJ, Butcher JN. The contribution of the MMPI-2 to the diagnosis of personality disorders. *MMPI-2 A practitioner's Guide*. 2005:99-120
 56. Gómez-Gil E, Vidal-Hagemeyer A, Salamero M. MMPI–2 characteristics of transsexuals requesting sex reassignment: Comparison of patients in prehormonal and presurgical phases. *Journal of personality assessment*. 2008 Jun 26;90(4):368-74.
 57. de Vries AL, Doreleijers TA, Steensma TD, Cohen-Kettenis PT. Psychiatric comorbidity in gender dysphoric adolescents. *Journal of Child Psychology and Psychiatry*. 2011 Nov;52(11):1195-202.
 58. Beattie M, Lenihan P. *Counselling Skills for Working with Gender Diversity and Identity*. Jessica Kingsley Publishers; 2018 Mar 21.