



# International Journal of Clinical and Health Psychology

www.elsevier.es/ijchp



ORIGINAL ARTICLE

## Validity of Categories Related to Gender Identity in ICD-11 and DSM-5 Among Transgender Individuals who Seek Gender-Affirming Medical Procedures



Rebeca Robles<sup>a,\*</sup>, Jared W. Keeley<sup>b</sup>, Hamid Vega-Ramírez<sup>a</sup>, Jeremy Cruz-Islas<sup>c</sup>, Victor Rodríguez-Pérez<sup>d</sup>, Pratap Sharan<sup>e</sup>, Shivani Purnima<sup>e</sup>, Ravindra Rao<sup>e</sup>, María Inés Rodríguez-Lobato<sup>f</sup>, Bianca Soll<sup>f</sup>, Françoise Askevis-Leherpeux<sup>g</sup>, Jean-Luc Roelandt<sup>g</sup>, Megan Campbell<sup>h</sup>, Gerhard Grobler<sup>i</sup>, Dan J. Stein<sup>h</sup>, Brigitte Khoury<sup>j</sup>, Joseph El Khoury<sup>j</sup>, Ana Fresán<sup>a</sup>, María-Elena Medina-Mora<sup>a</sup>, Geoffrey M. Reed<sup>k</sup>

<sup>a</sup> Ramón de la Fuente Muñiz National Institute of Psychiatry, Mexico

<sup>b</sup> Virginia Commonwealth University, USA

<sup>c</sup> Centro de Atención Transgénero Integral, Mexico

<sup>d</sup> Condesa Specialized Clinic, Ministry of Health, Mexico

<sup>e</sup> All India Institute of Medical Sciences, India

<sup>f</sup> Hospital de Clínicas de Porto Alegre (HCPA), Universidade Federal do Rio Grande do Sul (UFRGS), Brazil

<sup>g</sup> EPSM Lille Métropole, INSERM – Eceve, France

<sup>h</sup> University of Cape Town, South Africa

<sup>i</sup> University of Pretoria, South Africa

<sup>j</sup> American University of Beirut, Lebanon

<sup>k</sup> Columbia University Vagelos College of Physicians and Surgeons, USA

Received 20 August 2021; accepted 2 November 2021

Available online 30 November 2021

### KEY WORDS

ICD-11;  
DSM-5;  
Gender incongruence;  
Gender dysphoria;  
Ex post facto study;

### Abstract

*Background/Objective:* The most recent versions of the two main mental disorders classifications—the World Health Organization's ICD-11 and the American Psychiatric Association's DSM-5—differ substantially in their diagnostic categories related to transgender identity. ICD-11 gender incongruence (GI), in contrast to DSM-5 gender dysphoria (GD), is explicitly not a mental disorder; neither distress nor dysfunction is a required feature. The objective was compared

At the time of the study R. Robles, J.W. Kelley, P. Sharan, D. Stein, B. Khoury, and M.E. Medina-Mora were Consultants to the ICD-11 Field Studies Coordination Group, reporting to the WHO International Advisory Group for the Revision of ICD-10 Mental and Behavioural Disorders; and G.M. Reed was a member of the WHO Secretariat, Department of Mental Health and Substance Abuse, WHO. The WHO Department of Mental Health and Substance Abuse provided consultation regarding the focus and methodology of the study, none of these entities dictated the final form of the study. Unless specifically stated, the views expressed in this article are those of the authors and do not represent the official policies or positions of WHO or of the ICD-11 Field Studies Coordination Group.

\* Corresponding author: Calzada México Xochimilco 101, Col. San Lorenzo Huipulco. Tlalpan, Ciudad de México, 14370, Mexico.  
E-mail address: reberobles@imp.edu.mx (R. Robles).

<https://doi.org/10.1016/j.ijchp.2021.100281>

1697-2600/© 2021 Asociación Española de Psicología Conductual. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

ICD-11 and DSM-5 diagnostic requirements in terms of their sensitivity, specificity, discriminability and ability to predict the use of gender-affirming medical procedures. *Method:* A total of 649 of transgender adults in six countries completed a retrospective structured interview. *Results:* Using ROC analysis, sensitivity of the diagnostic requirements was equivalent for both systems, but ICD-11 showed greater specificity than DSM-5. Regression analyses indicated that history of hormones and/or surgery was predicted by variables that are an intrinsic aspect of GI/GD more than by distress and dysfunction. IRT analyses showed that the ICD-11 diagnostic formulation was more parsimonious and contained more information about caseness than the DSM-5 model. *Conclusions:* This study supports the ICD-11 position that GI/GD is not a mental disorder; additional diagnostic requirements of distress and/or dysfunction in DSM-5 reduce the predictive power of the diagnostic model.

© 2021 Asociación Española de Psicología Conductual. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## PALABRAS CLAVE

CIE-11;  
DSM-5;  
Discordancia de género;  
Disforia de género;  
Estudio ex post facto

## Validez de las categorías relacionadas con la identidad de género en la CIE-11 y el DSM-5 entre personas transgénero que buscan procedimientos médicos que afirmen su género

### Resumen

*Antecedentes/Objetivo:* Las versiones más recientes de las clasificaciones de trastornos mentales –CIE-11 de la Organización Mundial de la Salud y DSM-5 de la Asociación Psiquiátrica Americana– difieren en sus categorías diagnósticas relacionadas con la identidad transgénero. La discordancia de género (DiscG) de la CIE-11, en contraste con la disforia de género (DisfG) del DSM-5, no es considerada un trastorno mental; el estrés y la disfunción no son características requeridas para el diagnóstico. El objetivo fue comparar los requisitos diagnósticos de la CIE-11 y el DSM-5 en términos de sensibilidad, especificidad y capacidad para discriminar casos y predecir el uso de procedimientos médicos de afirmación de género. *Método:* 649 adultos transgénero de seis países completaron una entrevista estructurada retrospectiva. *Resultados:* De acuerdo con el análisis ROC, la sensibilidad de ambos sistemas fue equivalente, aunque la CIE-11 mostró mayor especificidad que el DSM-5. Los análisis de regresión indicaron que la historia de uso de hormonas o cirugía se predijo por variables intrínsecas a la DiscG/DisfG y no por el estrés o disfunción. Según los análisis de respuesta al ítem (TRI) la formación CIE-11 resulta más parsimoniosa y contiene mayor información sobre los casos. *Conclusiones:* Se aporta evidencia a favor de que la DiscG/DisfG no es un trastorno mental; los criterios diagnósticos adicionales de estrés y/o disfunción del DSM-5 reducen su poder predictivo.

© 2021 Asociación Española de Psicología Conductual. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

The classification of transgender identity as a mental disorder has imposed social, legal and health burdens on this population. At the same time, not having a diagnostic category in the classification of health conditions could be “a significant impediment for transgender people seeking access to medical treatment, and is therefore not recommended” (Drescher et al., 2012, pp. 575). The development processes of the most recent versions of the major global classifications –the Eleventh Revision of the International Classification of Diseases (ICD-11) and the Fifth Edition of the DSM (DSM-5)– therefore attempted to find a balance between the concerns related to the stigmatization of mental disorders and the need for diagnostic categories that support the provision of high-quality, affordable healthcare services this population might require (Drescher et al., 2012). However, their solutions were substantially different.

The World Health Organization (WHO) reconceptualized ICD-10’s Gender Identity Disorders as Gender Incongruence (GI), including and moved its two component categories Gender Incongruence in Childhood and Gender Incongruence

in Adolescence and Adulthood (without the label “disorder”) out of the chapter on Mental, Behavioral or Neurodevelopmental Disorders to a new ICD-11 chapter on Conditions Related to Sexual Health. One rationale for this change was that these conditions fail to satisfy one of the main definitional requirements of mental disorders: the presence of distress and/or dysfunction caused by the condition itself (Askevis-Leherpeux et al., 2019; Campbell et al., 2018; Khoury et al., 2020; Lobato et al., 2019; Robles et al., 2016).

In DSM-5, the names of the related diagnostic categories have been changed to “Gender dysphoria in children” and “Gender dysphoria in adolescents or adults”, and diagnostic formulations requires distress or dysfunction related to gender dysphoria (GD), which since DSM-IV has been a requirement for virtually all mental disorders (Stein et al., 2020). Both the ICD-11 GI label and the DSM-5 GD label are less stigmatizing and focus more on the subjective experience of transgender people, but in DSM-5 the term “dysphoria” and the category’s presence in a classification of mental

disorders connote that psychological distress is a primary feature (Stein et al., 2020), while it may or may not be present in the ICD-11 GI formulation.

Another difference between the ICD-11 and DSM-5 diagnostic formulations is the required duration for a diagnosis of GI/GD. DSM-5 requires a minimum duration of six months for a diagnosis of GD in adolescence and adulthood. In contrast, ICD-11 provides more flexibility by indicating that marked GI/GD should be present for “several months” in order to assign a diagnosis and reduce risks by increasing access to gender-affirming medical procedures (GAMP) under medical supervision.

Despite these central differences, both diagnostic systems coincide in the four main manifestations of the central clinical feature of GI/GD: (1) a strong dislike or discomfort with the one’s primary and/or secondary sex characteristics, (2) a strong desire to be rid of some of all of one’s primary and/or secondary sex characteristics, (3) a strong desire to have the primary and/or secondary sex characteristics (or appearance) of the experienced gender, and (4) a strong desire to be treated (to live and be accepted) as a person of the experienced gender.

The principal aim of the present study was to explore whether the ICD-11 or DSM-5 diagnostic requirements were better able to establish the presence of GI/GD using Item Response Theory (IRT) analysis (treating each set of diagnostic requirements as a measurement scale and each diagnostic feature as an item of the corresponding scale). Diagnostic requirements were evaluated retrospectively through detailed interviews focusing on self-identified transgender adults’ experiences at the time of their adolescence, when secondary sex characteristics began to appear and is frequently described as a very stressful time for transgender people (Mirabella et al., 2020). This strategy intended to avoid the artificial advantage for ICD-11 (which does not require the presence of distress/dysfunction) that would have resulted from application of the diagnostic requirements at the moment of the study (i.e., during adulthood), when it is more probable that transgender people and others around them would have adapted to their transgender status and as a result distress/dysfunction would be much less likely. IRT analysis describes the relationship between a latent trait or construct to be assessed (in this case, the diagnosis of GI/GD); the properties of the components or requirements or items on a scale (in this case, each of the four main manifestations of GI/GD plus distress or dysfunction in DSM-5); and respondents’ answers to the individual items (diagnostic features/criteria) (Yang & Kao, 2014). This allows the estimation of the ability of the items to discriminate between cases and non-cases (which are referred to as discrimination and difficulty). Additionally, we evaluate the sensitivity and specificity of both sets of diagnostic requirements, as well as whether history of GAMP (hormones and/or surgery) was predicted more by variables that are an intrinsic aspect of GI/GD than by experienced distress/dysfunction. We hypothesized that the essential features for GI in ICD-11 (without distress/dysfunction) would provide more information on caseness and have greater specificity than the DSM-5 criteria for GD, and that variables intrinsically part of GI/GD would predict a history of GAMP better than distress/dysfunction. Although the construct of GAMP is an evolving one (Deutsch & Feldman, 2013), use of hormone

treatment and surgery can be seen as an intermediate outcome validator of GI/GD in our sample.

## Method

This multi-site study used a cross-sectional descriptive design drawing on a purposive sample (Etikan et al., 2016) of transgender adults from six countries (Brazil, France, India, Lebanon, Mexico, and South Africa). All procedures were approved by Ethics Committees (IRBs) of all participating institutions.

## Participants

Participants were adult individuals who identified themselves as transgender who were receiving health services at one of the study sites (or other institutions in the case of France). These included the Hospital de Clínicas de Porto Alegre, and the Hospital das Clínicas of the University of São Paulo, in Brazil; two sites providing HIV prevention services for the transgender population in Delhi, India accessed by the All India Institute of Medical Sciences; the Arab Foundation for Freedoms and Equality and the MARSА sexual health clinic, in Beirut, Lebanon; the Condesa Specialized Clinic of the Ministry of Health, in Mexico City; and the Steve Biko Academic Hospital Gender Reassignment Clinic in Pretoria, the Groote Schuur Transgender Clinic at the University of Cape Town, and the Triangle Project in Cape Town, South Africa. In France, participants were recruited at the Maison Dispersée de Santé, which supports transgender people in their affirmation process.

## Measures, procedures and data analysis

The materials for the study were developed in English and Spanish and then translated from English into Arabic, French, Hindi and Portuguese by bilingual mental health professionals using a thorough forward and back-translation process. Written informed consent was obtained from all participants and then an appointment was scheduled with a trained research assistant who administered a structured interview. At the beginning of the interview, the research assistant asked about participants’ sociodemographic characteristics and history of GAMP. Then, the research assistant identified the participant’s specific period during which she/he became aware of the development of own secondary sex characteristics (subsequently referred to as index period). This was possible using an interview strategy that has proved suitable for evaluating the presence of the diagnostic requirements during a specified period (First et al., 1996). First, the research assistant asked participants to focus on the “age when they first became consciously aware that they might be transgender and would need to do something about it” (which in all cases was during childhood), and then on the “age they first became aware of the development of secondary sex characteristics associated with a non-preferred gender” (which in all cases was at some point during adolescence). Participants were then instructed to answer the following questions based on their feelings, thoughts, and experiences at that particular time and age. In addition, all subsequent questions began with the following phase:

“During the time we are talking about (index period), at age...”. This retrospective interview covered the presence and duration of a variety of aspects of the individuals’ experience related to their GI/GD during the interview index period, which corresponds to the diagnostic requirements for GI/GD ICD-11 or DSM-5. Answer options to the questions covering the diagnostic features of GI/GD other than distress or dysfunction were binary (Yes/No). Experience of distress related to their transgender identity was assessed by asking: “During the time we are talking about, (index period), at age... did you experience psychological distress related to your gender identity?”; while related dysfunction was assessed using an adaptation of the Sheehan Disability Scale (Sheehan & Sheehan-Harnett, 1996) referring to the index period, and asking about the extent to which they felt that any reported disruption in functioning was related to their gender identity. Distress and dysfunction were collapsed into a single variable for this analysis, corresponding to this element of the DSM-5 diagnostic criteria.

Data analyses employed as hypothesis probes included regressions to predict history of GAMP, receiver operating characteristic (ROC) curves to calculate the sensitivity and specificity of both ICD-11 and DSM-5, and an IRT analysis to evaluate the ability of each diagnostic requirement to predict who qualifies for a diagnosis under either ICD-11 or DSM-5. The variables used to represent GI/GD diagnostic features were minimally related (all phi coefficients  $< .23$ ) and thus treated as independent for the purpose of these analyses.

## Results

A total of 679 transgender adults were invited to participate in the study; only twenty potential participants from India and ten from Mexico were excluded from the final sample of the study. In India, ten potential participants did not consent to participate, while ten who consented and participated were excluded from the final database given discrepancies/errors in the information provided. In Mexico, five did not consent to participate, and five did not provide sufficient information to allow for analysis. Thus, the final sample included 649 transgender adults. Table 1 presents the number of participants by country, their demographic characteristics and history of GAMP.

The presence of GI/GD features was consistent across countries except for the strong desire to be treated as a person of the experienced gender (Table 2). Participants from Brazil, Lebanon, and South Africa were most likely to report this, while participants from France were markedly less likely to do so.

### Predicting GAMP

When examined separately, the *strong desire to be rid of some of all of one’s primary and/or secondary sex characteristics* ( $\chi^2(1) = 12.31, p < .001$ ) and the *strong dislike or discomfort with the one’s primary and/or secondary sex characteristics* ( $\chi^2(1) = 7.26, p < .01$ ) were the only GI/GD features related to use of GAMP. When all variables were entered into a simultaneous prediction of history of GAMP in a logistic regression, the pattern was the same with only these two aspects of GI/GD being significant predictors

( $B = -1.05, p < .01$ ;  $B = -0.84, p < .05$ , respectively). Distress and dysfunction were not significant predictors of GAMP in any of these analyses.

### ROC curve analysis

The ROC curve analysis assessed the extent to which the different diagnostic guidelines in ICD-11 and criteria in DSM-5 contributed to distinguishing individuals who met the overall diagnostic requirements for that system from those who did not. Table 3 shows the area under the curve (AUC) for each diagnostic guideline. In keeping with the respective diagnostic requirements, the presence of distress and/or dysfunction was not included in determining an ICD-11 diagnosis, although it was for DSM-5. The four main diagnostic requirements (manifestations of GI/GD shared in both systems) provided more information (i.e., higher AUCs) for an ICD-11 diagnosis than for a DSM-5 diagnosis. As expected, distress and/or dysfunction was the least informative characteristic for ICD-11, as it is not a diagnostic requirement. In contrast, distress and/or dysfunction was the major determinant of a DSM-5 diagnosis, with the other criteria being barely above chance levels (i.e., .50). The sensitivity of each diagnostic requirement was equivalent across the two systems. However, the ICD-11 diagnostic requirements showed higher specificity, meaning that they were less likely to be associated with non-cases in the ICD-11 than in the DSM-5.

### IRT models

A more refined examination of the value of each diagnostic requirement comes from IRT analysis. For ICD-11, the IRT model included the four main manifestations of GI/GD shared in both systems. For DSM-5, the model included these four variables plus a duration requirement and the presence of distress/dysfunction related to GI/GD. The ICD-11 model did not include a duration variable since everyone met its durational requirement (several months with marked GI/GD). In this context, item difficulty refers to the threshold at which a person met the requirement; lower (including more negative) values indicate that the threshold was “easier” to meet. Item discriminability refers to the diagnostic requirement’s ability to predict who qualifies for the condition; higher values indicate a sharper distinction between people with and without GI/GD.

A reduced model, where all diagnostic features were constrained to have the same difficulty, provided a significantly worse fit to the data for both ICD-11 ( $F(1, 4) = 627.39, p < .001$ ) and DSM-5 ( $F(1,6) = 824.07, p < .001$ ). This finding indicates that the diagnostic features should have different difficulty values, i.e., the diagnostic requirements correspond to different thresholds of intensity in GI/GD. Similarly, a reduced model where all diagnostic features were constrained to have the same discriminability provided a significantly worse fit to the data for ICD-11 ( $F(1, 4) = 11.79, p < .05$ ), but not for DSM-5 ( $F(1, 6) = 10.29, p = .11$ ). Thus, the ICD-11 model justifies having different discriminability parameters for each diagnostic requirement. For DSM-5, there was not enough evidence to justify differing discriminabilities (i.e., they are all statistically equal to 1). The ICD-11 model having different discriminability values means that the diagnostic requirements varied in the degree to

**Table 1** Demographic information and history of MGAP.

	Brazil	France	India	Lebanon	Mexico	South Africa	Total
<i>n</i>	103	72	139	28	250	57	649
Age <i>M</i> ( <i>SD</i> )	30 (8.8)	27.71 (9.67)	26.9 (8.81)	29.64 (8.77)	30.82 (10.25)	28.68 (7.51)	29.27 (9.48)
Years of education <i>M</i> ( <i>SD</i> )	11.99 (3.46)	13.83 (2.52)	10.56 (6.01)	11.36 (3.42)	12.20 (3.54)	13.28 (2.39)	12.06 (4.13)
Marital status <i>n</i> (%)							
Partnered	22 (21.4)	18 (25)	44 (31.7)	6 (21.4)	43 (17.2)	18 (31.6)	151 (23.3)
Single	81 (78.6)	54 (75)	94 (68.3)	22 (78.6)	207 (82.8)	39 (68.4)	497 (76.7)
Primary occupation <i>n</i> (%)							
Employed	82 (79.6)	61 (83.9)	107 (77)	14 (50)	180 (72)	35 (61.4)	446 (68.7)
Student	2 (1.9)	23 (31.9)	16 (11.5)	3 (10.7)	36 (14.4)	9 (15.8)	89 (13.7)
Homemaker	5 (4.9)	-	-	-	9 (3.6)	-	14 (2.2)
Retired/unemployed	14 (13.6)	21 (29.2)	16 (11.5)	11 (39.3)	25 (10)	13 (22.8)	100 (15.4)
Gender identity <i>n</i> (%)							
Women	79 (76.7)	39 (54.2)	109 (78.4)	14 (50)	125 (50)	28 (49.1)	394 (60.7)
Man	24 (23.3)	23 (31.9)	2 (1.4)	2 (7.1)	29 (11.6)	9 (15.8)	89 (13.7)
Genderqueer	-	3 (4.2)	19 (13.)	3 (10.7)	4 (1.6)	-	29 (4.5)
Agender	-	-	9 (6.5)	-	1 (0.4)	-	10 (1.6)
Transwoman	-	3 (4.2)	-	6 (21.4)	74 (29.6)	15 (26.3)	98 (15.1)
Transman	-	4 (5.6)	-	-	17 (6.8)	5 (8.8)	26 (4)
Other	-	-	-	3 (10.)	-	-	3 (0.5)
Sex assigned at birth <i>n</i> (%)							
Female	24 (23.3)	28 (38.9)	2 (1.4)	3 (10.7)	46 (18.4)	14 (24.6)	117 (18)
Male	79 (76.7)	44 (61.1)	137 (98.6)	23 (82.1)	202 (80.8)	43 (75.4)	528 (81.4)
Intersex	-	-	-	2 (7.1)	2 (0.8)	-	4 (0.6)
Gender-affirming medical procedures (GAMP) <i>n</i> (%)							
Hormones – Yes*	85 (97.7)	60 (100)	16 (88.9)	16 (80)	182 (98.9)	20 (90.9)	379 (96.9)
Surgery – Yes*	34 (39.1)	23 (38.3)	7 (38.9)	11 (55)	36 (19.6)	12 (54.5)	123 (31.5)
Other* -Yes	87 (84.5)	60 (83.3)	18 (12.9)	20 (71.4)	184 (73.6)	22 (38.6)	391 (60.2)

Note: *M* = mean; *SD* = standard deviation; *n* = frequency; % = percent.

\* Other includes mold release, collagen, permanent hair removal, etc.

**Table 2** Frequency (%) of gender incongruence/dysphoria features by country.

	Brazil	France	India	Lebanon	Mexico	South Africa	Total
Rid desire	103 (100)	72 (100)	117 (84.2)	28 (100)	227 (90.8)	57 (100)	604 (93.1)
Body dislike	103 (100)	69 (95.8)	119 (85.6)	27 (96.4)	240 (96)	57 (100)	615 (94.8)
Appearance desire	98 (95.1)	63 (87.5)	133 (95.7)	28 (100)	243 (97.2)	56 (98.2)	621 (95.7)
Acceptance desire	91 (88.3)	26 (36.1)	94 (67.6)	26 (92.9)	156 (62.4)	49 (86)	442 (68.1)
Distress and/or dysfunction	96 (93.2)	63 (87.5)	121 (87.1)	26 (92.9)	228 (91.2)	51 (89.5)	585 (90.1)

Note: Rid desire = strong desire to be rid of some of all of one's primary and/or secondary sex characteristics; Body dislike = strong dislike or discomfort with the one's primary and/or secondary sex characteristics; Appearance desire = strong desire to have the primary and/or secondary sex characteristics of the experienced gender; Acceptance desire = strong desire to be treated (to live and be accepted) as a person of the experienced gender.  
 $\chi^2(5) = 14.72, p < .05$ .

which they differentiated individuals who do or do not have GI/GD.

It is not possible to directly compare the ICD-11 and DSM-5 IRT models because they are based on a different number of variables. However, the ICD-11 model appears more parsimonious (i.e. lower log likelihood, AIC, and BIC values; see Table 4). Further, the ICD-11 model contains slightly more information about the "caseness" of the individual than the constrained DSM-5 model. Additionally, this information is more applicable to the central portion of the latent attribute (i.e., between  $\pm 4$  standard deviations from the mean) for ICD-11 (information = 5.46, 88.43%) than DSM-5 (information = 3.80, 63.81%). In other words, a large portion of the information drawn from the DSM-5 diagnostic requirements is only useful at very low levels of GI/GD, whereas the ICD-11 requirements are applicable across the full range of GI/GD.

Regarding the performance of individual diagnostic requirements (see Table 4 and Figures 1 and 2), the strong desire to be treated (to live and be accepted) as a person of the experienced gender had the highest difficulty as well as the highest discriminability relative to the other features; while the strong desire to be rid of some of all of one's primary and/or secondary sex characteristics had the lowest difficulty and discriminability.

From the perspective of considering the diagnostic requirements to be a "test" of a latent construct of GI/GD, it is desirable to have items that vary in their difficulty and discriminability. The ICD-11 model does an acceptable job of meeting this goal (Figure 1), meaning that its diagnostic features can detect a spectrum of GI/GD. The DSM-5 diagnostic features follow a similar pattern (Figure 2). However, its features are more tightly bound, with the exception of the desire to be treated as a person of the experienced gender (which had high difficulty), and the duration requirement (which had very low difficulty, meaning that nearly all people in the sample met this requirement). The other symptoms have less spread than ICD-11, indicating that they are somewhat redundant in determining a DSM-5 diagnosis.

## Discussion

The principal aim of the present international multisite field study was to compare the ICD-11 and the DSM-5 diagnostic criteria for GI/GD in transgender adults in terms of their sensitivity, specificity, discriminability and ability to predict history of GAMP. Our analyses found that the use of GAMP was predicted only by variables related to marked GI/GD rather than by the presence of distress/dysfunction, which are diagnostic criteria in the DSM-5 but not in the ICD-11. These findings support the external validity of the ICD-11's diagnostic guidelines for GI in adolescence and adulthood. Although the possibility that distress/dysfunction might be one of the reasons to seek GAMP cannot be ruled out, there are transgender people that do not exhibit important levels of distress/dysfunction and does really want and use GAMP; the characteristics related in a more unequivocal way to the use of GAMP are the marked gender incongruence manifestations, which should be then the only criteria for the diagnosis of the condition, for example.

**Table 3** ROC summaries for ICD-11 and DSM-5.

	ICD-11				DSM-5			
	AUC	95% CI	Sensitivity	Specificity	AUC	95% CI	Sensitivity	Specificity
Rid desire	.83	.69, .97	.95	.71	.56	.49, .64	.94	.18
Body dislike	.77	.60, .93	.96	.57	.60	.53, .68	.97	.24
Appearance desire	.77	.61, .93	.97	.57	.56	.49, .64	.97	.15
Acceptance desire	.85	.80, .90	.70	1	.61	.54, .68	.71	.51
Distress and/or dysfunction*	.68	.52, .84	.91	.43	.95	.91, .99	1	.89

Note: AUC = Area Under the Curve; 95% CI = 95% Confidence Interval for AUC; Rid desire = strong desire to be rid of some of all of one's primary and/or secondary sex characteristics; Body dislike = strong dislike or discomfort with the one's primary and/or secondary sex characteristics; Appearance desire = strong desire to have the primary and/or secondary sex characteristics of the experienced gender; Acceptance desire = strong desire to be treated (to live and be accepted) as a person of the experienced gender.

\* Distress and dysfunction were not included as determinants of a diagnosis for ICD-11, but they were for DSM-5.

Additionally, although ROC statistics reflect an equivalent sensitivity of each diagnostic requirement across the two systems, ICD-11 showed higher specificity than DSM-5. The higher likelihood of ICD-11 correctly identifying those who have the condition or true positives and, at the same time, not categorizing certain people as having the condition when in fact they do not have it (i.e., avoiding false positives) (Trevethan, 2017) suggests that boundaries with gender-diverse individuals who are less likely to be interested in GAMP could be better established using the ICD-11 (given that, in contrast to transgender people, those gender non-conforming may feel comfortable with their biological sex without conforming to perceived societal roles, for example) (Institute of Medicine, 2011). This finding stands out because one could argue that the duration and distress/dysfunction criteria in DSM-5 are designed to

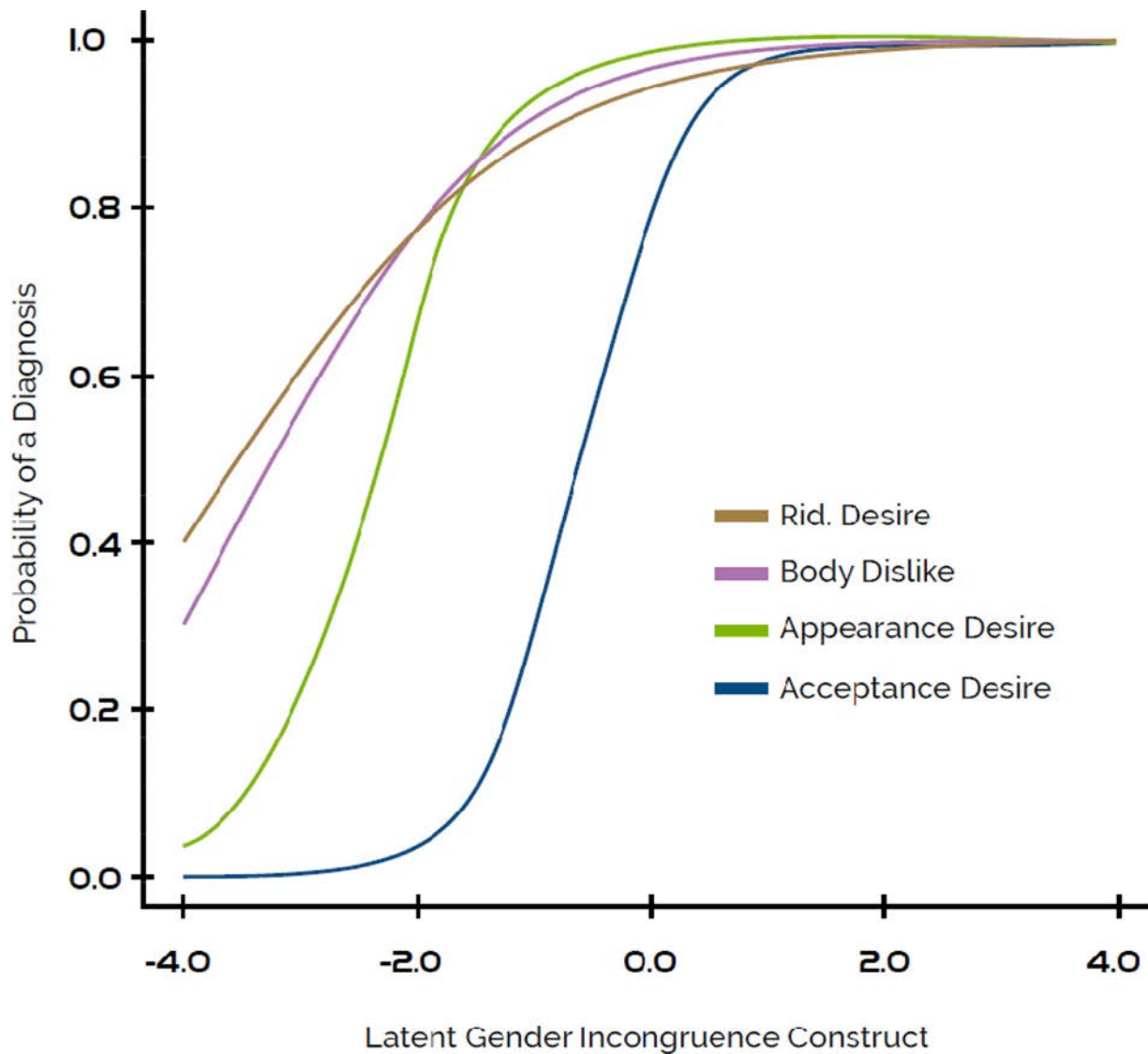
exclude false positives, but this study suggests that they may fail to achieve that goal.

Consistent with these findings, IRT analyses showed that the ICD-11 model appears more parsimonious and contains slightly more information about the "caseness" than the constrained DSM-5 model. Statistically, our analytic approach advantaged the DSM-5 by including more variables in the analysis (i.e., a duration requirement and distress/dysfunction), yet this model did not capture as much information. One possible interpretation of this finding is that the duration of one's GI/GD symptoms and the resultant distress/dysfunction are not meaningful indicators of GI/GD. For duration, this is almost certainly because participants in all countries described their experience of gender identity as very long-term, if not lifelong. For distress/dysfunction, this may be related to our previous finding that, in this

**Table 4** IRT model statistics.

	ICD-11	DSM-5	DSM-5 (constrained)
Log Likelihood	-790.28	-999.21	-1004.35
AIC	1596.56	2022.42	2032.71
BIC	1632.36	2076.13	2086.41
Item Difficulty			
Rid desire	-3.51	-3.2	-3.01
Body dislike	-3.2	-2.7	-3.33
Appearance desire	-2.33	-2.33	-3.55
Acceptance desire	-0.59	-0.68	-0.91
Distress/Dysfunction	-	-2.97	-2.59
Duration	-	-4.54	-6.99
Item Discriminability			
Rid desire	0.82	0.92	1
Body dislike	1.05	1.34	1
Appearance desire	1.96	1.96	1
Acceptance desire	2.35	1.64	1
Distress/Dysfunction	-	0.83	1
Duration	-	1.76	1
Total Information	6.17	8.44	5.95

Note: AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; Rid desire = strong desire to be rid of some of all of one's primary and/or secondary sex characteristics; Body dislike = strong dislike or discomfort with the one's primary and/or secondary sex characteristics; Appearance desire = strong desire to have the primary and/or secondary sex characteristics of the experienced gender; Acceptance desire = strong desire to be treated (to live and be accepted) as a person of the experienced gender.



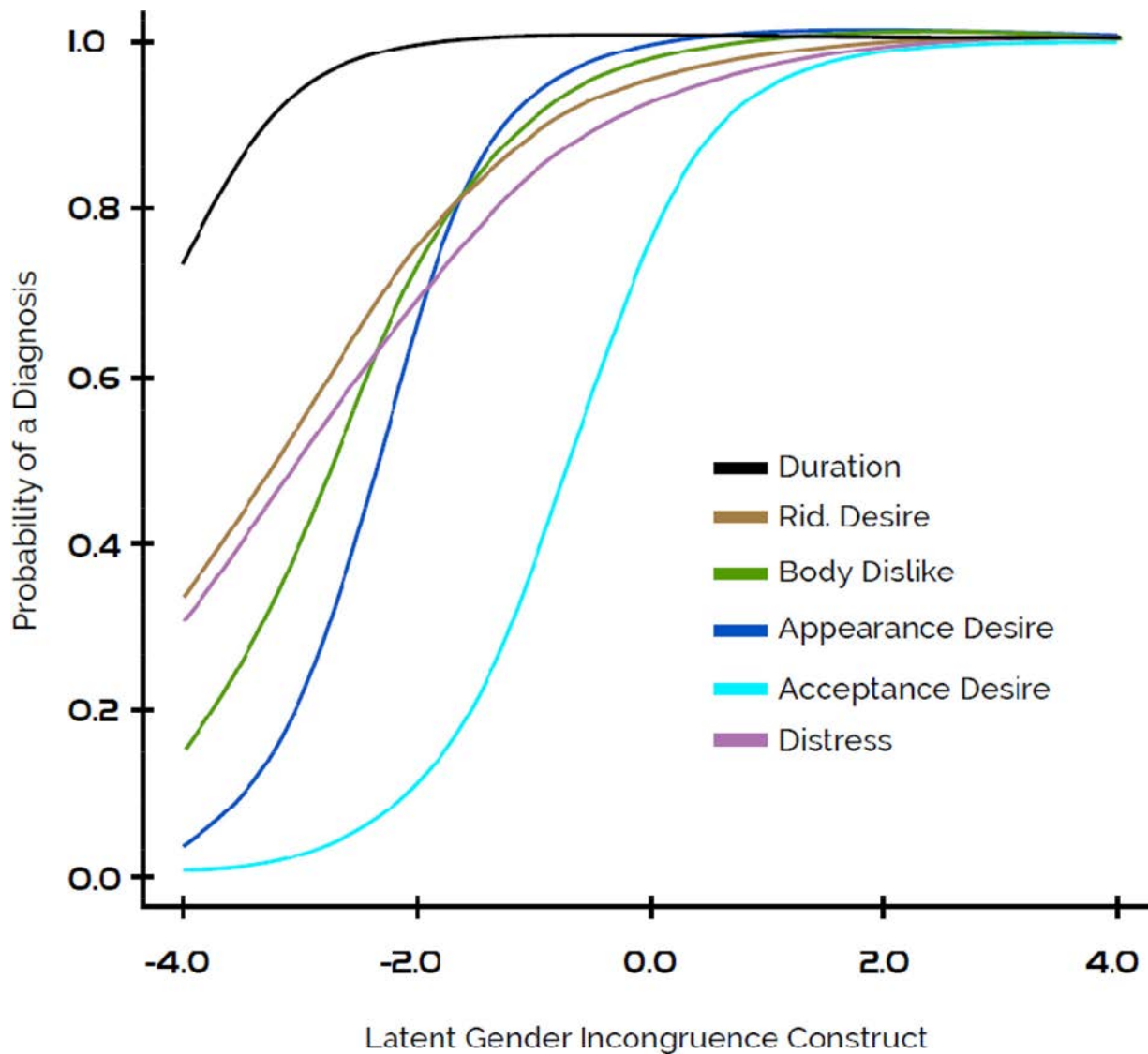
**Figure 1** ICD-11: Item characteristic curves. *Note:* Rid. Desire = strong desire to be rid of some of all of one's primary and/or secondary sex characteristics; Body Dislike = strong dislike or discomfort with the one's primary and/or secondary sex characteristics; Appearance Desire = strong desire to have the primary and/or secondary sex characteristics of the experienced gender; Acceptance Desire = strong desire to be treated (to live and be accepted) as a person of the experienced gender.

sample, distress and dysfunction were more strongly associated with experiences of stigmatization and violence than with the experience of GI *per se* (Askevis-Leherpeux et al., 2019; Campbell et al., 2018; Khoury et al., 2020; Lobato et al., 2019; Robles et al., 2016). Moreover, IRT analyses testing a latent construct of GI/GD showed that the ICD-11 guidelines are better at detecting the spectrum of GI/GD that exists, characterizing cases at all levels and intensities of GI/GD, while at the same time avoiding the redundancy of additional criteria included in DSM-5.

Interestingly, these analyses also revealed that the strong desire to be treated as a person of the experienced gender was reported by the lowest proportion of individuals (i.e., highest difficulty), possibly reflecting the fact that this characteristic of marked GI/GD may be one of the last features expressed by transgender people, given the social stigma in many countries. Our South African sample, which reported

the lowest level of violence related to their gender identity (42.1%) exhibited a higher frequency of participants asking to be treated as a person of the experienced gender (60%), in contrast with the Mexican sample, which reported one of the highest levels of social violence related to their gender identity (63%) and only 43% asking to be treated as a person of the experienced gender (Askevis-Leherpeux et al., 2019; Campbell et al., 2018; Khoury et al., 2020; Lobato et al., 2019; Robles et al., 2016).

This marked GI/GD feature is also a strong indicator of being transgender (i.e., highest discriminability), which is in line with previous evidence indicating that among transgender people, those with more GI tended to disclose their transgender identity at younger ages (Maguen et al., 2007). In contrast, the strong desire to be rid of some of all of one's primary and/or secondary sex characteristics was reported by the highest proportion of the sample (lowest difficulty)



**Figure 2** DSM-5: Item characteristic curves. *Note:* Rid. Desire = strong desire to be rid of some of all of one's primary and/or secondary sex characteristics; Body Dislike = strong dislike or discomfort with the one's primary and/or secondary sex characteristics; Appearance Desire = strong desire to have the primary and/or secondary sex characteristics of the experienced gender; Acceptance Desire = strong desire to be treated (to live and be accepted) as a person of the experienced gender; Distress = Distress or dysfunction.

although its ability to predict overall group membership was spread over a wider range of individuals (lowest discriminability).

The main limitation of the study is related to its non-representative and volunteer sample of transgender individuals receiving health services. Thus, results should not be taken as estimates of the prevalence of particular indicators or other epidemiological parameters. Transgender individuals who do not seek GAMP offered in the participating would not have been included in the sample. It may also underrepresent individuals who experience greater barriers to accessing care, among whom the prevalence of different components of GI/GD and the predictors of GAMP may be different. However, both ICD and DSM are intended to be applied in health care systems with people who are similar to those included in our sample, often in order to make decisions about the appropriateness of GAMP.

Another possible limitation is that the data examined were based on participants' recollection and reconstruction of their experiences of GI/GD, distress/dysfunction at a younger age, which might have been influenced by subsequent experiences. A prospective study enrolling adolescents prior to receiving any GAMP could provide valuable information, although because prospective participants would be difficult to identify unless they presented for treatment, other methods would be required to capture those who do not use GAMP. However, our interview was designed to generate information that was as specific and accurate as possible regarding the most relevant time period, and although a recall bias could be present as in all studies using retrospective diagnostic or epidemiological interviews, this issue applied equally to all participants and should not have systematically affected the results of the study.

Finally, a third evident limitation is the categorical (binary) measurement level of distress/dysfunction, which might have conditioned the results. Confirmatory studies using continuous measurements of these variables are warranted.

Results of this large international, multilingual study indicate that the distress/dysfunction diagnostic criteria in DSM-5 do not provide additional useful information in identifying transgender people seeking GAMP. Rather, this study provides support for the placement of GI categories outside the ICD-11 chapter on Mental, Behavioral or Neurodevelopmental Disorders. This change has important clinical and policy implications. Clinicians' use of ICD-11 can help them to identify and provide health services for transgender people in timely manner in a wider variety of settings, which could decrease the harmful use of GAMP without medical supervision, a practice that is very common among transgender people in some countries (e.g., Robles et al., 2016).

Stakeholders around the world could employ the ICD-11 to support policies and programs and actions to reduce experiences of social rejection and violence in this population through the destigmatization and promotion of human and health rights of transgender people. The change also has important policy implications because it removes the justification for past insidious practice in some countries of denying transgender people certain rights (e.g., legal recognition of gender identity, reproductive rights, child custody) because they have a mental disorder -although ICD stills a disease classification.

## Funding

This study was partially funded by unrestricted grants by the World Health Organization Department of Mental Health and Substance Abuse to participating institutions in Brazil, India, Lebanon, Mexico, and South Africa for the development and implementation of ICD-11 field studies. Financial support for the French study was provided by the Direction Générale de la Santé (DGS), Ministry of Health, France.

## References

- Askevis-Leherpeux, F., de la Chenelière, M., Baleige, A., Chouchane, S., Martin, M. J., Robles-García, R., Fresán, A., Quach, A., Stona, A.-C., Reed, G., & Roelandt, J. L. (2019). Why and how to support depsychiatrisation of adult transidentity in ICD-11: A French study. *European Psychiatry, 59*, 8-14. <https://doi.org/10.1016/j.eurpsy.2019.03.005>.
- Campbell, M. M., Fresán, A., Addinall, R. M., Böhmke, W., Grobler, G. P., Marais, A., Wilson, D., Stein, D. J., Reed, G. M., & Robles, R. (2018). Experiences of gender incongruence and the relationship between social exclusion, psychological distress, and dysfunction among South African transgender adults: a field-study for ICD-11. *Annals of Clinical Psychiatry, 30*, 168-174.
- Deutsch, M. B., & Feldman, J. L. (2013). Updated recommendations from the world professional association for transgender health standards of care. *American Family Physician, 87*, 89-93.
- Drescher, J., Cohen-Kettenis, P., & Winter, S. (2012). Minding the body: Situating gender identity diagnoses in the ICD-11. *International Review of Psychiatry, 24*, 568-577. <https://doi.org/10.3109/09540261.2012.741575>.
- Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics, 5*, 1-4. <https://doi.org/10.11648/j.ajtas.20160501.11>.
- First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1996). *Structured clinical interview for DSM-IV axis I disorders, clinical version (SCID-CV)*. American Psychiatric Press.
- Institute of Medicine. (2011). *The Health of Lesbian, Gay, Bisexual, and Transgender People: Building a Foundation for Better Understanding*. National Academies Press.
- Khoury, B., El Khoury, J., Fresán, A., Daouk, S., Khauli, N., Ramadan, Z., Reed, G. M., & Robles, R. (2020). The ICD-11 classification of gender incongruence of adolescence and adulthood: Adequacy among transgender people in Lebanon. *Culture, Health & Sexuality, 23*, 131-142. <https://doi.org/10.1080/13691058.2019.1692907>.
- Lobato, M. I., Soll, B. M., Brandelli Costa, A., Saadeh, A., Gagliotti, D. A., Fresán, A., Reed, G., & Robles, R. (2019). Psychological distress among transgender people in Brazil: Frequency, intensity and social causation—an ICD-11 field study. *Brazilian Journal of Psychiatry, 41*, 310-315. <https://doi.org/10.1590/1516-4446-2018-0052>.
- Maguen, S., Shipherd, J. C., Harris, H. N., & Welch, L. P. (2007). Prevalence and predictors of disclosure of transgender identity. *International Journal of Sexual Health, 19*, 3-13. [https://doi.org/10.1300/J514v19n01\\_02](https://doi.org/10.1300/J514v19n01_02).
- Mirabella, M., Giovanardi, G., Fortunato, A., Senofonte, G., Lombardo, F., Lingiardi, V., & Speranza, A. M. (2020). The body I live in. perceptions and meanings of body dissatisfaction in young transgender adults: A qualitative study. *Journal of Clinical Medicine, 9*, Article 3733. <https://doi.org/10.3390/jcm9113733>.
- Robles, R., Fresán, A., Vega-Ramírez, H., Cruz-Islas, J., Rodríguez-Pérez, V., Domínguez-Martínez, T., & Reed, G. M. (2016). Removing transgender identity from the classification of mental disorders: a Mexican field study for ICD-11. *The Lancet Psychiatry, 3*, 850-859. [https://doi.org/10.1016/S2215-0366\(16\)30165-1](https://doi.org/10.1016/S2215-0366(16)30165-1).
- Sheehan, D. V., & Sheehan-Harnett, R. (1996). The measurement of disability. *International Clinical Psychopharmacology, 11*(Suppl. 3), 89-95. <https://doi.org/10.1097/00004850-199606003-00015>.
- Stein, D. J., Szatmari, P., Gaebel, W., Berk, M., Vieta, E., Maj, M., de Vries, Y. A., Roest, A. M., de Jonge, P., Maercker, A., Brewin, Ch. R., Pike, K. M., Grilo, C. M., Fineberg, N. A., Briken, P., Cohen-Kettenis, P. T., & Reed, G. M. (2020). Mental, behavioral and neurodevelopmental disorders in the ICD-11: An international perspective on key changes and controversies. *BMC Medicine, 18*, 1-24. <https://doi.org/10.1186/s12916-020-1495-2>.
- Trevethan, R. (2017). Sensitivity, specificity, and predictive values: Foundations, plabilities, and pitfalls in research and practice. *Frontiers in Public Health, 5*, Article 307. <https://doi.org/10.3389/fpubh.2017.00307>.
- Yang, F. M., & Kao, S. T. (2014). Item response theory for measurement validity. *Shanghai Archives of Psychiatry, 26*, 171-177. <https://doi.org/10.3969/j.issn.1002-0829.2014.03.010>.