The Gender Congruence and Life Satisfaction Scale (GCLS): Development and Validation of a Scale to Measure Outcomes from Transgender Health Services

Abstract

Background: It is vital that the treatment offered at transgender health services can be evaluated to ensure a high quality of care. However, the tools currently used to evaluate treatment at transgender health services are limited by mainly focusing on mental health or because they have been developed for binary transgender people only. This study therefore aimed to develop and validate a tool that addresses these limitations. The Gender Congruence and Life Satisfaction Scale (GCLS) was developed through reviewing the literature, conducting interviews with transgender people, and holding discussions with experts working in transgender healthcare. An initial pool of items was developed and feedback on these was obtained. The tool was then validated.

Method: For the validation of the tool a total of 789 participants (451 transgender [171 transgender females, 147 transgender males, 133 non-binary identifying people], and 338 cisgender [254 females, 84 males]) were recruited from the United Kingdom to test the factor structure and validity of the GCLS.

Results: Exploratory factor analysis retained 38 items which formed seven subscales (psychological functioning; genitalia; social gender role recognition; physical and emotional intimacy; chest; other secondary sex characteristics; and life satisfaction). These seven subscales were found to have good internal consistency and convergent validity. The GCLS was also found to be capable of discriminating between groups (e.g., people who have and have not undergone gender affirming medical interventions). Transgender and cisgender subscale norms are provided for the GCLS.

Conclusion: The GCLS is a suitable tool to use with the transgender population to measure health-related outcomes for both clinical and research purposes.

Key words transgender; non-binary identities; gender congruence; gender distress; mental health; life satisfaction; scale; treatment outcome

Introduction

Transgender people are those who experience a discrepancy between the gender they were assigned at birth and their gender identity. Transgender males are people who were assigned female at birth, on the basis of their sexual characteristics, but identify as male. Transgender females are people who were assigned male at birth, on the basis of their sexual characteristics, but identity as female. Some transgender people may also identify between or outside the binary gender spectrum. These people may choose to identify as gender neutral (feeling that one is neither male nor female), non-gender (having no gender in relation to presentation), or gender queer (identifying and presenting in a way that is outside the gender dichotomy of male and female) (Arcelus & Bouman, 2017; Richards, Bouman, & Barker, 2017, Richards et al., 2016). Some people may also be more fluid with their gender identity whereby they do not have a fixed gender and it can therefore vary over time. Cisgender people do not experience such discrepancy (Arcelus & Bouman, 2017; Bouman et al., 2017). Many transgender people experience high levels of distress due to the discrepancy between their birth-assigned gender and gender identity (e.g., Beek, Kreukels, Cohen-Kettenis, & Steensma, 2015). To alleviate this distress, many transgender people will approach transgender health services in order to access gender affirming medical interventions (GAMI) to help them transition to the gender they identify with. Health professionals working at transgender health services may start their patients on cross-sex hormone treatment to induce either masculinization (with testosterone) or feminization (with estrogen and often with testosterone-blocking medication), depending on the patient's gender identity (Coleman et al., 2012; Wylie et al., 2014). After living as their experienced gender, transgender people, if they wish to do so, can be referred for gender affirming surgery (Coleman et al., 2012; Wylie et al., 2014).

Transgender health services throughout Europe and North America have seen a substantial increase in the number of referrals in recent years. This has put a strain on these services, especially in relation to waiting times for assessment and treatment (Aitken et al., 2015; de Vries, Kreukels, T'Sjoen, Ålgars, & Mattila, 2015). In the United Kingdom (UK) patients can expect to wait more than one year before their first appointment at a transgender health service (Bouman & Richards, 2013; UK Trans Info, 2016). In light of this, evaluating the care and treatment received at these services is important. This will allow for the quality of care to be improved and will also facilitate the identification of service and personal factors associated with positive and negative outcomes (e.g., Dawson, Doll, Fitzpatrick, Jenkinson, & Carr, 2010). If factors associated with positive and negative outcomes of transgender health services can be identified, patient-centered services can be created which provide extra support for patients who are vulnerable to poorer outcomes. Identification of these factors may also help to make the treatment process timely and more efficient for patients who are deemed not to be vulnerable. Patient-centered services are crucial in improving patient outcomes (e.g., Lauver et al., 2002). In addition, within nationalized social healthcare systems which are present in most European countries, access to care, including transgender healthcare is free at the point of access for all citizens. However, these healthcare systems have limited resources and, at some point, rationing decisions may have to be made. Tools that predict which interventions are associated with good outcomes are vital to assist with making decisions regarding the allocation of healthcare resources and to explore how to improve overall patient outcomes.

Research has explored patient treatment outcomes post-gender affirming interventions, such as cross-sex hormones and gender affirming genital surgery. On the whole, these studies have demonstrated that mental health and quality of life improve following gender affirming

medical interventions (e.g., Dhejne, Van Vlerken, Heylens, & Arcelus, 2016; Jones, Haycraft, Murjan, & Arcelus, 2016; Marshall, Claes, Bouman, Witcomb, & Arcelus, 2015; Murad et al., 2010; Van de Grift et al., 2018; Witcomb et al., 2018). The main focus of outcome evaluations at transgender health services has often been based around mental health symptoms (e.g., anxiety, depression) (e.g., Bouman et al., 2016, 2017; Heylens et al., 2014; Murad et al., 2010; Witcomb et al., 2018). While mental health is evidently an important aspect to include within any outcome assessment of gender affirming medical interventions, given the high prevalence of mental health problems transgender people often experience pre-gender affirming medical interventions (Dhejne et al., 2016), it should not be the sole focus. Researchers have argued that improvements in mental health should not be the only focus of transgender health service evaluations, as mental health interventions are generally not provided by transgender health services (Arcelus & Bouman, 2015; Dhejne et al., 2016). Patient outcomes will also be biased by levels of mental health problems experienced pre-treatment. Moreover, the treatment pathways for transgender people consist of multidisciplinary treatment options, which further complicates the evaluation of these interventions.

One of the most important outcomes that should be measured after gender affirming medical interventions is a change in the distress and unhappiness a person experiences with their experienced gender and body as a result of their gender identity being at odds with their assigned gender (i.e., gender incongruence). Measurement of this outcome is often neglected due to the lack of available measures that have been developed with the transgender population and are capable of assessing a change in distress and unhappiness a person experiences with their gender and body. The few measures that are available to assess gender distress (such as the Utrecht Gender Dysphoria Scale; Cohen-Kettenis & van Goozen, 1997)

and body dissatisfaction (such as the Hamburg Body Drawing Scale; Becker et al., 2016), although useful, are limited. These measures have been developed with the binary gender system in mind (i.e., male or female) which is problematic in light of an increasing number of people identifying as non-binary or outside the gender binary (Beek et al., 2015; Clarke, Veale, Townsend, Frohard-Dourlent, & Saewyc, 2018; Richards et al., 2016, 2017); also, participants are asked to complete different versions depending on the gender they were assigned at birth. Measures developed to assess treatment outcomes within the transgender population therefore need to ensure they encapsulate people with non-binary gender identities and be gender neutral (i.e., applicable to all genders). A gender neutral measure also allows clinicians to assess outcomes using the same tool throughout the medical transition.
In addition, for outcome evaluations to be meaningful, the measures employed must have been developed for, and validated with, the population in question (Dawson et al., 2010).

Currently, mental health and quality of life measures that are used to assess patient outcomes at transgender health services have often been developed for use with other specific (and dissimilar) populations (e.g., Eating Disorder Inventory-2; Garner, 1991) or the general population (e.g., Hospital Anxiety and Depression Scale; Zigmond & Snaith, 1983). Therefore, these measures are unlikely to be specific enough (i.e., unlikely to ask about mental health problems in relation to gender distress) to be used with the transgender population for meaningful evaluations. Due to the lack of a suitable, validated mental health and quality of life measure, transgender health services usually invite patients to complete a series of measures that assess different constructs relating to mental health and quality of life to ensure evaluations are comprehensive. Research has found respondent burden to be great when patients are asked to complete multiple questionnaires, especially when these questionnaires assess similar constructs (Rolstad, Adler, & Ryden, 2011; Turner et al., 2007).

Respondent burden can affect the quality of data gathered and may reduce response rates (Diehr, Chen, Patrick, Feng, & Yasui, 2005; Snyder et al., 2007). In light of this, treatment evaluations collected at transgender health services may be of a poor quality as they might have been affected by respondent burden.

Aims

The objective of this study was to develop a self-report tool that was capable of assessing the outcomes of transgender health services. Patients are referred to transgender health services as a consequence of experiencing gender incongruence and therefore in the short-term the main aim of the tool was to assess levels of gender distress, gender congruence, and associated mental well-being. It was also expected that the tool would be able to assess long-term outcomes by measuring levels of life satisfaction and psychological well-being. Such a tool is imperative to ensure evaluations are mapped onto the most important aims of gender affirming medical treatment and to allow meaningful and efficient evaluations to take place at transgender health services. The tool that was developed in the current study has been named the Gender Congruence and Life Satisfaction Scale (GCLS). Details of the development of the GCLS will be provided in the methods section. Once the tool had been developed, the first aim of this study was to explore the factor structure of the GCLS. The second aim was to explore the convergent and discriminant validity of the GCLS (types of construct validity). The final aim was to determine whether the GCLS can distinguish between subgroups (i.e., transgender and cisgender people) and be sensitive to changes in gender distress, gender incongruence, associated mental well-being, and life satisfaction throughout the treatment process (known-groups validity; a further type of construct validity). The final aim therefore also provides an opportunity to pilot the GCLS with transgender people and provide subscale norms.

Method

Participants and recruitment

Two different groups of participants were involved in this study: transgender people and cisgender (non-trans) people. Part of the study sample comprised transgender people who were invited to take part from a national transgender health service within the UK. This is a National Health Service (NHS) funded center, which offers assessment and treatment to transgender people who are pursuing, or are considering, medical transition. The center is one of the larger transgender healthcare services in Europe and receives around 1000 referrals a year from England and Wales. This service accepts referrals from people aged 17 and over. Clinicians at the service informed participants of the existence of the study and provided them with the information sheet and details about how to participate.

Some transgender people and all of the cisgender participants were recruited through the community via snowball sampling. All transgender and cisgender people recruited through the community were required to be aged 18 or over. This was achieved by sharing an online link to the study with transgender support organizations (for transgender people only), via social media websites, and by email. The content of this advertisement was the same for transgender and cisgender people recruited through the community. Participants who took part in the online survey were then asked to pass the link on to others in their network.

Recruitment took place over four months in 2016. A total of 458 transgender participants were invited. However, from this sample, seven people were removed as they either had provided no information about their gender assigned at birth and gender identity (n = 3) or had not indicated whether or not their gender assigned at birth and gender identity were the same (n = 4). Thus, the final sample consisted of 451 transgender people. In addition, a total

of 375 cisgender participants were invited. Of these participants, 37 were removed as they reported a gender identity that was different from the gender they were assigned at birth. The final cisgender sample consisted of 338 participants. The final sample size therefore included 789 participants (338 identified as cisgender [42.8 %] and 451 as transgender [57.2 %]).

Procedure

After participants had read through the information sheet and decided whether or not they would like to take part, they were invited to complete the first iteration of the newly developed tool (GCLS). Participants recruited from the transgender health service were either asked to complete a paper questionnaire pack in their own time and return this to the service in a pre-paid envelope or complete the online survey. Community participants were asked to complete the survey online. No paper alternative was offered. Informed consent was obtained from all participants prior to taking part in the study.

The study was approved by an NHS research ethics committee and by the Research and Development Department of Nottinghamshire Healthcare NHS Foundation Trust (16/EM/0183) in line with Health Research Authority guidance (HRA, 2013), as well as Loughborough University Research Ethics Committee.

Development of the Gender Congruence and Life Satisfaction Scale

Several processes were undertaken to complete the development of the GCLS. First, systematic reviews of the available literature on treatment outcomes, including body dissatisfaction (see Jones et al., 2016) and mental health were conducted (see Dhejne et al., 2016). This was followed by a review of existing body dissatisfaction measures that are currently used to assess patient outcomes at transgender health services (Jones et al., 2016).

Next, in-depth interviews with 14 transgender people attending a national transgender health service in the UK were undertaken as part of a larger study (Jones, Arcelus, Bouman, & Haycraft, 2017). These interviews highlighted how distress and dissatisfaction with gender, associated mental well-being, and life satisfaction improved over the treatment process. The study provided valuable information regarding the most important objectives of transgender health services from the patients' point of view. Following this, several discussions and focus groups took place with clinicians from different transgender health services in the UK, Sweden and Belgium. All of the outlined processes and studies mentioned above informed the first draft of the GCLS. The items within this first draft were developed as a result of the authors reviewing other self-report measures relating to gender incongruence, gender distress, mental well-being (including body dissatisfaction), and life satisfaction used with the transgender and cisgender population. This promoted a discussion among the research team about some of the limitations of these measures (e.g., not transgender specific, developed within the binary gender system). Some of the limitations were identified through earlier phases of development (i.e., the review of existing measure used to assess body dissatisfaction). The team then worked together to develop a list of items while taking into consideration some of the limitations identified in the previous phase. To be a valid item it had to: 1) be related to gender incongruence, gender distress, mental well-being or life satisfaction; 2) assess this construct in relation to gender incongruence and/or gender distress (e.g., I have felt extremely distressed when looking at my genitals); 3) be gender neutral. This process resulted in 85 items being developed.

The first draft of the GCLS was then discussed with the Nottingham Centre for Transgender Health Service User Research Advisory Group (SURAG). This group consisted of 21 people, who were attending, or had attended the service and agreed to provide feedback for research

projects taking place at the service. This consultation resulted in a revised draft of the GCLS. The authors also worked in consultation with transgender people in the community, who were recruited through charities and support organizations for transgender and LGBT people within the UK. This included asking people to comment on the tool and the items that comprised it. Feedback was collected which allowed the development of the next draft of the scale, which was shared with several clinicians and academics in the field, including clinical academics from outside the UK (Sweden and Belgium). Following feedback, several drafts were developed which were discussed with the above groups until a tool that satisfied everyone was created. This process resulted in a tool with 42 items initially being created.

Measures

Socio-demographic questions: Information about participants' age, their gender assigned at birth and gender identity was collected. Participants were invited to provide information (if applicable) about their gender transition, including the amount of time they spend living in their experienced gender in their daily lives (less than 50 %, more than 50 %, or 100 % of the time). They were also asked whether they were using cross-sex hormones and/or blockers (if relevant) and whether they had undergone any gender affirming surgery ("Yes" or "No" response style). On the online survey, these questions were only visible to individuals who indicated incongruence between their birth and experienced gender (i.e., these questions were not visible to cisgender participants).

Gender Congruence and Life Satisfaction Scale (GCLS; Jones, Bouman, Haycraft, & Arcelus, 2018; see Appendix): This 38-item scale aims to assess change and measure improvements in gender (in)congruence, related mental well-being, and life satisfaction throughout the process of undergoing gender affirming medical interventions. The scale is

independent of gender assigned at birth and was developed to be relevant to people who identify as male, female, as well as those who identify outside the binary gender system (e.g., non-binary identities). The GCLS asked respondents to think about how they have felt over the last 6 months and to rate their responses on a 5-point Likert scale (always = 1; never = 5). The 6 month time frame was selected to reflect the time transgender people have to of experienced gender incongruence and distress for to receive an official diagnosis of Gender Dysphoria as per the Diagnostic and Statistical Manual of Mental Disorders (DSM), fifth edition (American Psychiatric Association (APA), 2013). Although the GCLS is not a diagnostic tool, this time frame is felt to be important within the field of transgender health. Mean scores were calculated for each subscale. In the final (38 item) version of the scale, 11 items were reverse coded (i.e., never = 1; always = 5). A higher score indicates a greater gender congruence, greater gender-related well-being, and greater life satisfaction. In contrast, a lower score indicates lower gender congruence, poorer gender-related well-being, and poorer general life satisfaction.

Hamburg Body Drawing Scale (HBDS; Becker et al., 2016): This scale was originally developed by Appelt and Strauss (1988) for use with individuals with different forms of psychoendocrinological disorders and assesses satisfaction with 33 different body parts. Becker et al. (2016) recently validated the scale with the transgender population and adapted it to include male and female specific subscales. There is also an item that assesses overall satisfaction with appearance. This was the only item used in the current study and the reason for including it was to assess the convergent validity of the GCLS (i.e., to determine whether two measures that theoretically should be related, are related). The male and female subscales were not used within the current study as they were thought not appropriate to use with non-binary identifying people. Participants were asked to rate their responses on a 5-point Likert

scale (1 = very dissatisfied; 5 = very satisfied) and therefore a higher score indicates a higher level of body satisfaction. Becker et al. (2016) found the HBDS subscales to have good reliability in a transgender population (α = .62-.91). As it was not possible to conduct reliability analysis with just one item, this was not calculated in the current study.

Transgender Congruence Scale (TCS; Kozee, Tylka, & Bauerband, 2012): This measure has 12 questions, of which nine correspond with "appearance congruence" *(e.g., I do not feel that my appearance reflects my gender identity; I am generally comfortable with how others perceive my gender identity when they look at me)* and three relating to "gender identity acceptance" *(e.g., I am not proud of my gender identity).* Mean scores were calculated for each subscale. A global score can also be calculated by finding the mean of all responses. In the current study, only the 9-item appearance congruence subscale was used to assess convergent validity. Responses were rated on a 5-point scale from 1= strongly disagree to 5= strongly agree. A higher score indicates a higher level of transgender congruence. This measure has been found to have good reliability (α = .93; Kozee et al., 2012) and in the current sample, the appearance congruence subscale had excellent reliability (α = .96).

World Health Organization Quality of Life-BREF (WHOQOL-BREF; Harper, 1998): The WHOQOL-BREF has 26 items which measures four domains: psychological, physical, relationships, and environment. There are also two items which are assessed separately, that ask about overall quality of life and overall health. In the current study the psychological and relationships subscales were used as well as the item that assessed overall quality of life, so as to assess convergent validity. The questions were rated on a 5-point Likert scale (1 = very dissatisfied; 5 = very satisfied) and a higher score indicates a higher quality of life. The subscale scores were calculated using the mean and then multiplied by four. This is to make

the scores from the WHOQOL-BREF comparable to the WHOQOL-100 (Harper, 1998), which is the longer original questionnaire. The WHOQOL-BREF has been found to have good to excellent reliability and validity in the general population and in clinical populations (e.g., rehabilitation, primary care, mental health) across 23 countries (Skevington, Lofty, & O'Connell, 2004). In this study, the psychological (α = .89) and relationships (α = .76) subscales had very good reliability.

Internet Gaming Disorder Scale-Short Form (IGDS; Pontes & Griffiths, 2015): This measure is a brief 9-item questionnaire that assesses internet gaming behavior *(e.g., Do you systematically fail when trying to control or cease your gaming activity?).* Participants were asked to rate their responses on a 5-point Likert scale (1 = never; 5 = very often). Scores range from 9-45 and were calculated by summing all responses. A higher score indicates more problematic gaming behavior. The measure has good reliability (α = .87) (Pontes & Griffiths, 2015) and in this study had excellent reliability (α = .92). This measure was included within the current study to assess discriminant validity (i.e., to determine whether two measures were not related, as would theoretically be expected).

Statistical Analysis

For the first aim of the study, principal component analysis was conducted with data from the transgender participants only to determine the factor structure of GCLS. For factor analysis, Comrey and Lee (1992) suggested that 300 participants is a good sample size and therefore the number of transgender participants (*n* = 451) is considered adequate. For the second aim, convergent and discriminant validity testing was conducted. To assess convergent and discriminant validity, one-tailed Spearman's Rho correlations (as the data were not normally distributed) were conducted between the GCLS and the WHOQOL, HBDS, TCS and IGDS

with the transgender participants only (n = 451). As multiple comparisons were conducted, Bonferroni corrections were used (.05 ÷ 13 = .004). For the final aim, known-groups validity testing (another form of construct validity) was conducted. This analysis determines whether known-groups within the dataset, in this circumstance transgender people at different stages of their transition (i.e., cross-sex hormones versus no cross-sex hormones) and cisgender people score in a theoretically expected way on the GCLS (e.g., transgender people score lower than cisgender people indicating a poorer outcome). Norm values were also generated for the GCLS subscales, global scale and clusters explored among the different groups. One-tailed Mann-Whitney U tests were conducted for this aim as the data were not normally distributed. The significance level was set at $p < .05$.

Results

Descriptive statistics relating to the gender identities of the transgender people are displayed in Table 1. The mean age of the transgender participants was 36.94 (*SD* = 15.46) and ranged from 17 to 77 years. Of the 451 transgender people, 189 (41.9 %) had not undergone any gender affirming medical intervention, 145 (32.2 %) had taken cross-sex hormone treatment only, 22 (4.9%) had taken cross-sex hormones and undergone mastectomy only, 92 (20.4 %) had taken cross-sex hormone treatment and undergone genital surgery (+ / - mastectomy) and 3 (.7) had undergone genital surgery but were not taking cross-sex hormone treatment. In the cisgender sample, 84 (24.9 %) were male and 254 female (75.1 %). The mean age of the cisgender sample was 36.52 (*SD* = 12.23) and ranged from 19 to 70 years.

INSERT TABLE 1 ABOUT HERE

Aim 1: Factor structure of GCLS

Preliminary analyses

To explore the factor structure of the GCLS, principal component analysis was conducted. It was expected that the items on the GCLS would be related and so therefore oblique (direct oblimin criterion) rotation was employed. As preliminary analysis, multicollinearity and singularity were assessed. The determinant was above .00001 and therefore multicollinearity and singularity were not a cause for concern. Bartlett's Test of Sphericity was also significant ($p < .001$) suggesting that correlations between variables were significantly different from zero and therefore it was concluded that variables were adequately related to find clusters within the dataset.

Exploratory factor analysis and item elimination

Analysis to determine the number of factors to retain was conducted. Kaiser (1960) suggested that factors with an eigenvalue greater than one should be retained. Kaiser's criteria is seen to be accurate when the sample size exceeds 250 (current sample $n = 451$) and the average communality is equal or greater than .6 (.67 in the current study; Field, 2013). Seven factors with an eigenvalue greater than one were identified and explained 67.30 % of the total variance. In contrast, the scree plot suggested five factors should be extracted (Cattell, 1966). Courtney and Gordon (2013) recommended that other statistical techniques (other than the eigenvalue and scree plot) should also be used to establish the number of factors to extract. Additional analysis was therefore conducted and Ruscio and Roche's (2012) comparative data technique suggested seven factors should be retained, as did Velicer's (1976) minimum average partial (MAP) test. Courtney and Gordon (2013) found Ruscio and Roche's (2012) comparative data technique to have 87.14 % accuracy and Velicer's (1976) MAP test to have 59.6 % accuracy in determining the number of factors to retain. Based on these analyses, seven factors were explored.

Item retention and elimination were subsequently considered in accordance with several criteria. Stevens (2002) suggested that for a sample size greater than 300 (current sample *n* = 451), item loadings should be greater than .30 to be retained within a factor. All item loadings in the current study were above .30 and therefore all 42 items were retained at this stage. Next, the interpretability of items was explored by an expert panel working in the area of transgender health (*n* = 8) and any items which were not found to theoretically or conceptually fit within the factor they had been placed were excluded. Based on this review, items 21, 27, 28 and 30 were excluded. When items cross-loaded simultaneously onto two (or more) factors with a difference of less than .10, the face validity of the item was considered. This applied to items 10, 19, 26 and 41. Items 10 (-.39), 26 (.42) and 41 (-.39) and were felt to conceptually fit better within the factors that they loaded highest with and therefore were retained within these factors. It was also felt that item 19 did not conceptually fit within factor 1 (the factor which the item loaded the highest, .41), however it also loaded highly onto factor 3 (.39) and was felt to conceptually fit better within this factor (in comparison to factor 1). Therefore, item 19 was retained within factor 3. In total, four items were removed which resulted in the revised version of the GCLS comprising 38 items.

Analysis of the remaining 38 items

The remaining 38 items were then subjected to a second principle components analysis with oblique (direct oblimin criterion) rotation. Seven factors had an eigenvalue greater than one, which explained a large proportion of the overall variance (68.01 %; see Table 2). All item loadings were greater than .30 and all items conceptually and theoretically fitted within the factor that they had been placed. The only exception to this rule was item 19, which loaded highly onto factor 1 and factor 3 and was moved from factor 1 to 3 as it was felt to better

conceptually fit within factor 3. Therefore 38 items were retained to comprise the final version of the GCLS.

The first factor (10 items) included items that related to psychological functioning associated with gender (in)congruence and was labelled “psychological functioning”. The second factor (6 items) included questions that pertained to distress and incongruence relating to the genitals and therefore was named “genitalia”. The third factor (4 items) included questions that asked about the degree to which participants were satisfied with how others perceived their gender role and therefore these items were categorized as “social gender role recognition”. The fourth factor (4 items) asked participants about satisfaction with their physical and emotional relationships and therefore was named “physical and emotional intimacy”. The fifth factor (4 items) was labelled “chest” as it included questions that asked about distress and incongruence with the chest. The sixth factor (3 items) included questions that were related to distress and incongruence experienced in relation to non-genital secondary sex characteristics (e.g., hair, voice) and therefore these questions were categorized as “other secondary sex characteristics”. The seventh factor (7 items) included questions that assessed general life satisfaction (not related to gender incongruence) and therefore was labelled “life satisfaction”. From reviewing these subscales, they appeared to conceptually cluster into two overarching themes; one that directly assessed the degree of gender congruence (factors 2, 3, 5 and 6) and another that assessed gender-related mental well-being and general life satisfaction (factors 1, 4 and 7). The clusters can be used by calculating the mean scores for the involved items.

Internal consistency

Internal consistency for each of the GCLS subscales and the global score (comprising all 38 items) were calculated. All seven subscales were found to have good (> .7) internal consistency (Nunnally, 1978; see Table 2) and the internal consistency for the global score was excellent (α = .95).

INSERT TABLE 2 ABOUT HERE

Aim 2: Convergent and discriminant validity

In order to test the construct validity of the GCLS (i.e., the degree to which the GCLS measures what it claims to), Spearman's Rho correlations were conducted between the GCLS and the WHOQOL, HBDS, TCS and IGDS with the transgender participants only (n = 451). Construct validity comprises convergent and discriminant validity. For the GCLS to show *convergent* validity (i.e., to determine whether two measures that theoretically should be related, are related), it would be expected to have a moderate to high correlation with the WHOQOL, HBDS and TCS (Evans, 1996). As might be expected, the GCLS physical and emotional intimacy subscale had a strong significant association with the WHOQOL relationships subscale, as did the GCLS psychological subscale and the WHOQOL psychological subscale. Additionally, the GCLS chest, GCLS genitalia and GCLS other secondary sex characteristics subscales had a moderate to strong association with the HBDS overall satisfaction scale. The GCLS social gender role recognition, GCLS other secondary sex characteristics, GCLS chest and GCLS genitalia subscales also all had a moderate to strong significant association with the TCS appearance congruence subscale. The strongest significant association was found between the GCLS life satisfaction and the WHOQOL overall quality of life subscale. These results confirm the convergent validity of the GCLS (see Table 3).

Next, correlations were run to determine *discriminant* validity; that is, to test whether the GCLS and IGDS are unrelated (i.e., a weak correlation) as would theoretically be expected. Table 4 shows that the GCLS subscales and the IGDS were unrelated as the two measures are weakly associated with one another (e.g., below $r = .39$; Evans, 1996). This suggests that the GCLS has discriminant validity. It is likely that some of these correlations will have reached statistical significance due to the large sample size of the current study (Field, 2013).

INSERT TABLE 3 ABOUT HERE

INSERT TABLE 4 ABOUT HERE

Aim 3: Known-groups' validity

To determine whether the GCLS is capable of distinguishing between subgroups (e.g., cisgender people and transgender people who have had no gender affirming medical intervention) and to determine whether groups score in a theoretically expected way (e.g., people who have undergone gender affirming medical treatments will be expected to score higher than people who have not undergone gender affirming medical treatments on all subscales of the GCLS) known groups validity (a further type of construct validity) was conducted using Mann-Whitney U tests (Hattie & Cooksey, 1984). Known-groups testing has previously been used to assess the construct validity of self-report measures (e.g., Alvarenga, Scagliusi, & Philippi, 2010). In the current study, some of these analyses were conducted between known-groups within the transgender sample, and some between transgender and cisgender participants. These analyses are not exhaustive but provide subscale and cluster norms for the different groups explored.

Initially, responses on the GCLS for transgender participants who had not undergone any gender affirming medical treatment (n = 189) and cisgender people (n = 338) were compared. It was found that the cisgender participants scored significantly higher (indicating a more positive outcome) on all GCLS subscales and the global scale than transgender people who had not undergone any gender affirming medical interventions (see Table 5). Responses between the two groups were also compared on the GCLS clusters previously identified. It was found that cisgender people scored significantly higher (indicative of a more positive outcome) than transgender people who were yet to undergo gender affirming medical interventions on the gender congruence cluster and the gender-related mental well-being and life satisfaction cluster (see Table 5). All effect sizes for the comparative analysis were large (Field, 2009).

INSERT TABLE 5 ABOUT HERE

Next, a Mann-Whitney U test was conducted to compare responses on the GCLS between transgender males who had not undergone any gender affirming medical intervention (n = 46) and transgender males who had taken cross-sex hormone treatment and undergone chest reconstructive (but not genital) surgery (n = 17). Within the UK (which is where these data were collected) chest reconstructive surgery is funded by the NHS. However, breast augmentation surgery is not funded, hence these analyses were only conducted in relation to transgender males. The analyses demonstrated that transgender males who had taken cross-sex hormones and undergone chest reconstructive scored significantly higher (more positive outcome) on the psychological functioning, social gender role recognition, chest, other secondary sex characteristics, and life satisfaction subscales of the GCLS as well as the global scale, than transgender males who had not undergone any gender affirming medial

intervention (see Table 6). No significant differences were found between the two groups on the genitalia and physical and emotional intimacy subscales of the GCLS. The two groups were also compared on the GCLS clusters and it was found that transgender males who had taken cross-sex hormones and undergone chest reconstructive surgery scored significantly higher (more positive outcome) on the gender congruence cluster and the gender-related mental well-being and life satisfaction cluster than transgender males who had not undergone any gender affirming medical interventions (see Table 6). Effect sizes for the comparative analysis were large (social gender role recognition; chest; other secondary sex characteristics; global scale; and cluster one), medium (psychological functioning; cluster two) and small (genitalia; physical and emotional intimacy; life satisfaction) (Field, 2009).

INSERT TABLE 6 ABOUT HERE

Finally, a Mann-Whitney U test was conducted between responses on the GCLS for transgender participants who were yet to undergo any gender affirming medical intervention (n = 189) and people who had taken cross-sex hormone treatment and undergone genital surgery (+ / - chest reconstructive surgery; n = 92). People who had taken cross-sex hormone treatment and undergone genital surgery (+ / - chest reconstructive surgery) scored significantly higher (more positive outcome) on all GCLS subscales and the global scale compared to people who had not undergone any gender affirming medical interventions (see Table 7). The groups were also compared on the two clusters and it was found that transgender people who had taken cross-sex hormone treatment and undergone genital surgery (+ / - chest reconstructive surgery) scored significantly higher (more positive outcome) on the gender congruence cluster and the gender-related mental well-being and life satisfaction cluster than people who had not undergone any gender affirming medical

intervention (see Table 7). Effect sizes for the comparative analysis were large (social gender role recognition; chest; cluster one), medium (psychological functioning; other secondary sex characteristics; life satisfaction; global scale; cluster two) and small (genitalia; physical and emotional intimacy) (Field, 2009).

INSERT TABLE 7 ABOUT HERE

Summary of analysis

Factor analysis on items of the GCLS supported a 7-factor solution with 38 items retained in total. The seven subscales are: psychological functioning; genitalia; social gender role recognition; physical and emotional intimacy; chest; other secondary sex characteristics; and life satisfaction. The subscales can be categorized into two clusters; 1) gender congruence (genitalia, chest, other secondary sex characteristics and social gender role recognition), and 2) gender-related mental well-being and general life satisfaction (physical and emotional intimacy, psychological functioning and life satisfaction). A higher score indicates a greater gender congruence, greater gender-related well-being, and greater life satisfaction. The GCLS subscales and the global scale have undergone vigorous testing and have demonstrated good reliability and validity. Subscale, global scale and cluster norms for transgender people at different stages of medical transition, as well as cisgender people have been generated. The known groups' analysis also demonstrated that the measure is capable of discriminating between groups of interest (e.g., people who have and have not undergone gender affirming medical interventions) and highlights how gender incongruence, gender-related mental well-being and general life satisfaction improve over the course of the medical transition. It can therefore be concluded that the GCLS is a suitable tool to use with the transgender population

to measure gender congruence and life satisfaction, and - in particular - improvements in these that are likely to occur during the treatment process.

Discussion

The main aim of this study was to develop and evaluate a new transgender health outcome measure: the Gender Congruence and Life Satisfaction Scale (GCLS) which is capable of measuring changes in gender congruence and life satisfaction over the course of gender affirming medical interventions. The GCLS is the first transgender health outcome measure that can be used independently of gender assigned at birth and gender identity. This study confirmed that the GCLS is a valid and reliable measure to use with the transgender population and is capable of producing high quality outcome data for clinical and research purposes.

Clinical utility of the GCLS

Having a brief, self-report questionnaire that is capable of assessing transgender health outcomes is important in order to measure the quality of care that is offered at transgender health services (Dawson et al., 2010), to enable transgender patients to be followed-up over time (both before and after gender affirming medical interventions), and to determine what transgender healthcare needs people require. The latter may help with financial planning and future workforce and service development in the field of transgender healthcare.

The GCLS was developed in collaboration with the transgender population and healthcare professionals working in the field of transgender health. This has ensured that the GCLS is meaningful and relevant for use as an outcome measure for people attending transgender health services (T'Sjoen, Motmans, Arcelus & Bouman, 2017). Factor analysis suggested the retention of 38 items, which comprised seven subscales (psychological functioning, genitalia,

social gender role recognition, physical and emotional intimacy, chest, other secondary sex characteristics, and life satisfaction) that were clustered into two themes: 1) gender congruence and 2) gender-related mental well-being and general life satisfaction. The two clusters have clinical relevance and applicability as some transgender people may or may not experience mental health problems in association with the distress and unhappiness they experience with their gender (e.g., Dhejne et al., 2016). The existence of two separate clusters therefore allows these factors to be considered separately. In addition, the outcome of transgender health services can be divided into outcomes related to the symptomatology that the individual presents to the service with (i.e., gender incongruence and gender dysphoria) and outcomes that relate to psychological well-being (life satisfaction, and mental health). Being able to review the scores on these two clusters may allow clinicians and healthcare professionals to deliver a patient-centered service. The delivery of patient-centered interventions is crucial to improve patient outcomes (e.g., Lauver et al., 2002). The mean scores given for each subscale, the global score and cluster of the GCLS for transgender people at different stages of medical transition within the current study can be used as norms. Subscale and cluster norms have also been provided for cisgender people; a sample that may provide a useful reference point for both future clinical and research purposes.

Research utility of the GCLS

It is expected that the GCLS may not only be a useful clinical tool, but may also aid research in the field of transgender health. As part of the validation process, the GCLS has already demonstrated that it can distinguish between subgroups of transgender individuals who are at different stages of medical transition, which makes it a suitable research tool. The current study found that transgender people who had not undergone any gender affirming medical interventions reported worse outcomes on all GCLS subscales, the global scale and two

clusters than both cisgender people and transgender people who had taken cross-sex hormone treatment and undergone genital surgery (+ / - chest reconstructive surgery). These findings support previous research that has shown patients to report less distress with their gender, lower levels of body dissatisfaction, better mental well-being, and greater life satisfaction after they had undergone gender affirming surgery (e.g., Dhejne et al., 2016; Jones et al., 2016; Marinkovic and Newfield, 2017; Marshall et al., 2015; Murad et al., 2010; van de Grift et al., 2018; Witcomb et al., 2018).

This study also found that transgender males who had not undergone any gender affirming medical intervention reported worse outcomes on the chest, other secondary sex characteristics, social gender role recognition, psychological functioning, and life satisfaction subscales of the GCLS, as well as the global scale and the two clusters, than transgender males who had taken cross-sex hormone treatment and undergone chest reconstructive surgery. There was no significant difference between the two groups on the genitalia and physical and emotional intimacy subscales of the GCLS. This finding supports previous research, which has found dissatisfaction and distress with the genitalia to be increasingly prevalent post-chest reconstructive surgery in treatment-seeking transgender men (van de Grift et al., 2016). Research with cisgender populations has found body dissatisfaction (Woertman & van den Brink, 2012), as well as specific genital dissatisfaction (Schick, Calabrese, Rima, & Zucker, 2010), to affect sexual satisfaction and therefore the high levels of dissatisfaction and distress experienced in relation to the genitalia in transgender men in the current study (pre-genital surgery) is likely to impact on their physical relationships. The analysis we conducted with known-groups demonstrates the sensitivity of the GCLS and its ability to distinguish between outcomes at different stages of medical transition (e.g., chest distress dissipates post-chest reconstructive surgery). Having a validated tool that is capable

of having this level of sensitivity may be extremely useful in advancing research in transgender health. For instance, when the length of time a person has been on cross-sex hormone treatment is known, longitudinal research may be able to identify a time frame in which long-term outcomes, such as psychological functioning and life satisfaction, improve. Establishing this in different subgroups (e.g., non-binary identifying people) is also important. The GCLS has been developed independent of gender assigned at birth and gender identity and therefore may also be an appropriate measure to use with non-binary identifying people. Research with this population is in its infancy, but suggests poorer mental health than binary transgender people (Rimes, Goodship, Ussher, Baker, & West, 2017; Scottish Trans Alliance, 2015; Thorne, Witcomb, Nieder, Nixon, & Arcelus, 2018; Warren, Smalley, & Barefoot, 2016). Some of the subscales from the GCLS may also be used to advance knowledge concerning factors that predict good outcomes in transgender people, for example, factors that predict life satisfaction. Establishing factors that predict a good outcome among non-binary identifying people is also essential as the factors are likely to differ to what is seen among transgender people who identify within the binary gender system.

The GCLS is the first transgender health outcome measure to be developed and validated with the transgender population and is capable of assessing important treatment outcomes for treatment seeking binary and non-binary identifying people. However, there are some limitations to be considered. First, the GCLS was only validated in an English speaking adult population. Future research should focus on validating the GCLS in other languages and other age groups (e.g., children and adolescents). Second, although gender incongruence, related mental well-being, and life satisfaction were all found to improve over the course of gender affirming medical interventions in the current study, it must be considered that these conclusions were based on cross-sectional data and test re-test reliability was not established.

In addition to this, data were not collected in relation to whether or not participants were accessing transgender health services (i.e., were treatment seeking) or the length of time on cross-sex hormone treatment. To further explore the short and long-term outcomes of gender affirming medical interventions, longitudinal research that uses the GCLS in treatment seeking transgender people needs to be conducted. The gender of the cisgender participants within this study was also skewed (i.e., the majority were cisgender women). Future research should look to employ the GCLS with a larger sample of cisgender males to compare scores with transgender males. Contrary to expectations, some subscales of the GCLS significantly correlated with the IGDS. This is likely to be due to the large sample size in the current study (Field, 2013). This finding may warrant further research into gaming behavior of transgender people.

In conclusion, the findings from this study suggest that the GCLS is a suitable and robust measure to assess treatment outcomes in relation to gender congruence, related mental well-being, and life satisfaction within the transgender population in both a clinical and research capacity. Having a measure that is capable of assessing these outcomes may allow for the quality of gender affirming interventions to be improved within transgender health services, and thus may improve the quality of life of transgender people and their families.

References

Aitken, M., Steensma, T. D., Blanchard, R., VanderLaan, D. P., Wood, H. F. A., Spegg, C., … Zucker, K. J. (2015). Evidence for an altered sex ratio in clinic-referred adolescents with gender dysphoria. *Journal of Sexual Medicine, 12*, 756-763. doi: 10.1111/jsm.12817

Alvarenga, M. S., Scagliusi, F. B., & Philippi, S. T. (2010). Development and validity of the disordered eating attitude scale (DEAS). *Perceptual and Motor Skills, 110*, 379-395. doi: 10.2466/PMS.110.2.379-395

American Psychological Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5 (5th ed.).* Washington DC: American Psychiatric Association.

Appelt, H., & Strauß, B. (1988). *Psychoendokrinologische Gynäkologie: Ergebnisse und Perspektiven.* [Psycho-endocrinological Gynaecology: results and prospects]. Stuttgart: Enke.

Arcelus, J., & Bouman, W.P. (2015). Current and future direction of gender dysphoria and gender incongruence research. *Journal of Sexual Medicine, 12*, 2226-2228. doi: http://dx.doi.org/10.1111/jsm.13065

Arcelus, J., & Bouman, W.P. (2017). Language and terminology. In: W.P. Bouman & J. Arcelus (Eds.), *The Transgender Handbook – A Guide for Transgender People, their Families and Professionals* (pp.1-12). New York: Nova Publishers.

Becker, I., Nieder, T. O., Cerwenka, S., Briken, P., Kreukels, B. P., Cohen-Kettenis, P. T., ... Richter-Appelt, H. (2016). Body image in young gender dysphoric adults: A European multi-center study. *Archives of Sexual Behaviour, 45*, 559-574. doi: 10.1007/s10508-015-0527-z

Beek, T. F., Kreukels, B. P. C., Cohen-Kettenis, P. T., & Steensma, T. D. (2015). Partial gender request and underlying motives of applicants for gender affirming interventions. *Journal of Sexual Medicine, 12*, 2201-2205. doi: 10.1111/jsm.13033

Bouman, W.P., Claes, L., Brewin, N., Crawford, J.R., Millet, N., Fernandez-Aranda, F., & Arcelus, J. (2017). Transgender and anxiety: A comparative study between transgender people and the general population. *International Journal of Transgenderism*, 18(1), 16-26.

Bouman, W.P., Davey, A., Meyer, C., Witcomb, G.L., & Arcelus, J. (2016). Predictors of psychological well-being among trans individuals. *Sexual and Relationship Therapy*, 31(3), 357-373.

Bouman, W.P., & Richards, C. (2013). Diagnostic and treatment issues for people with gender dysphoria in the United Kingdom. *Sexual and Relationship Therapy*, *28*, 165-171. doi: http://dx.doi.org/10.1080/14681994.2013.819222

Bouman, W. P., Suess Schwend, A., Motmans, J., Smiley, A., Safer, J. D., Deutch, M. B., …Winter, S. (2017). Language and transgender health. *International Journal of Transgenderism*, *18*(1), 1-6. doi: http://dx.doi.org/10.1080/15532739.2016.1262127

Cattell, R. B. (1966). The scree-test for the number of factors. *Multivariate Behavioral Research, 1*, 245-276. doi:10.1207/s15327906mbr0102_10

Clarke, B.A., Veale, J.F., Townsend, M., Frohard-Dourlent, H., & Saewyc, E. (2018). Non-binary youth: Access to gender-affirming primary health care. *International Journal of Transgenderism*. Advance online publication. doi: 10.1080/15532739.2017.1394954

Cohen-Kettenis, P. T., & van Goozen, S. H. M. (1997). Sex reassignment of adolescent transsexuals: A follow-up study. *Journal of the American Academy of Child and Adolescent Psychiatry*, *36*, 263–271. doi: https://doi.org/10.1097/00004583-199702000-00017

Coleman, E., Bockting, W., Botzer, M., Cohen-Kettenis, P., DeCuypere, G., Feldman, J., … Zucker, K. (2012). Standards of care for the health of transsexual, transgender, and gender-nonconforming people, Version 7. *International Journal of Transgenderism*, *13*, 165-232. doi: http://dx.doi.org/10.1080/15532739.2011.700873

Comrey, A. L., & Lee, H. B. (1992). *A First Course in Factor Analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Courtney, M. G. R., & Gordon, M. (2013). Determining the number of factors to retain in EFA: Using the SPSS R-Menu v2.0 to make more judicious estimations. *Practical Assessment, Research & Evaluation, 18*(8), 1-14.

Dawson, J., Doll, H., Fitzpatrick, R., Jenkinson, C., & Carr, A. J. (2010). The routine use of patient reported outcome measures in healthcare settings. *British Medical Journal, 340*, c186. doi: https://doi.org/10.1136/bmj.c186

Dhejne, C., Van Vlerken, R., Heylens, G., & Arcelus, J. (2016). Mental health and gender dysphoria: A review of the literature. *International Review of Psychiatry, 28*, 44-57. doi: http://dx.doi.org/10.3109/09540261.2015.1115753

de Vries, A. L. C., Kreukels, B. P. C., T'Sjoen, G., Ålgars, M. & Mattila, A. (2015). Increase of referrals to gender identity clinics: A European trend? In: Transgender Healthcare in Europe. Book of Abstracts. pp10. Ghent, Belgium: European Professional Association of Transgender Health (EPATH). Retrieved from: http://epath.eu/wp-content/uploads/2014/07/EPATH-2015-Book-of-Abstracts.pdf

Diehr, P., Chen, L., Patrick, D., Feng, Z., & Yasui, Y. (2005). Reliability, effect size, and responsiveness of health status measures in the design of randomized and cluster-randomized trials. *Contemporary Clinical Trials*, *26*, 45-58. doi: https://doi.org/10.1016/j.cct.2004.11.014

Evans, J. D. (1996). *Straightforward statistics for the behavioral Sciences*. Pacific Grove, CA: Brooks/Cole Publishing.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics (4th ed.*). Thousand Oaks, CA: Sage.

Garner, D.M. (1991). *Eating Disorder Inventory-2. Professional manual*. Odessa, FL: Psychological Assessment Research, Inc.

Harper, A. (1998). Development of the World Health Organization WHOQOL-BREF quality of life assessment. *Psychological Medicine*, *28*, 551-558.

Hattie, J., & Cooksey, R. W. (1984). Procedures for assessing the validities of tests using the "known-groups" method. *Applied Psychological Measurement, 8*, 295-305.

Health Research Authority (HRA) (2013). *Guidance for NHS Research Studies*. London: Health Research Authority.

Heylens, G., Verroken, C., De Cock, S., T'Sjoen, G., & De Cuypere, G. (2014). Effects of different steps in gender reassignment therapy on psychopathology: a prospective study of persons with a Gender Identity Disorder. *Journal of Sexual Medicine, 11*, 119-126. doi: 10.1111/jsm.12363

Jones, B. A., Arcelus, J., Bouman, W. P., & Haycraft, E. (2017). Barriers and facilitators of physical activity and sport participation among young transgender adults who are medically transitioning. *International Journal of Transgenderism*, *18*, 227-238. doi: http://dx.doi.org/10.1080/15532739.2017.1293581

Jones, B. A., Haycraft, E., Murjan, S., & Arcelus, J. (2016). Body dissatisfaction and disordered eating in trans people: A systematic review of the literature. *International Review of Psychiatry*, *28*, 81-94. doi: http://dx.doi.org/10.3109/09540261.2015.1089217

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*, 141-151.

Kozee, H. B., Tylka, T. L., & Bauerband, L. A. (2012). Measuring transgender individuals' comfort with gender identity and appearance: Development and validation of the Transgender Congruence Scale. *Psychology of Women Quarterly*, *36*, 179–196. doi: 10.1177/0361684312442161

Lauver, D. R., Ward, S. E., Heidrich, S. M., Keller, M. L., Bowers, B. J., Brennan, P. F., ... Wells, T. J. (2002). Patient-centered interventions. *Research in Nursing & Health*, *25*, 246-255. doi: 10.1002/nur.10044

Marinkovic, M., & Newfield, R.S. (2017). Chest reconstructive surgeries in transmasculine youth: Experience from one pediatric center. *International Journal of Transgenderism*, *18*(4), 376-381.

Marshall, E., Claes, L., Bouman, W. P., Witcomb, G. L., & Arcelus, J. (2016). Non-suicidal self-injury and suicidality in trans people: a systematic review of the literature. *International Review of Psychiatry*, *28*(1), 58-69.

Murad, M. H., Elamin, M. B., Garcia, M. Z., Mullan, R. J., Murad, A., Erwin, P. J., & Montori, V. M. (2010). Hormonal therapy and sex reassignment: A systematic review and meta-analysis of quality of life and psychosocial outcomes. *Clinical Endocrinology*, *72*, 214-231. doi: 10.1111/j.1365-2265.2009.03625.x

Nunnally, J. C. (1978). *Psychometric Theory*. Second edition. New York: McGraw-Hill.

Pontes, H. M., & Griffiths, M. D. (2015). Measuring DSM-5 Internet Gaming Disorder: Development and validation of a short psychometric scale. *Computers in Human Behavior*, *45*, 137-143. doi: https://doi.org/10.1016/j.chb.2014.12.006

Richards, C., Bouman, W.P., & Barker, M-J. (2017). *Genderqueer and Non-binary Genders*. London: Palgrave Macmillan.

Richards, C., Bouman, W. P., Seal, L., Barker, M. J., Nieder, T. O., & T'Sjoen, G. (2016). Non-binary or genderqueer genders. *International Review of Psychiatry*, *28*, 95–102. doi: http://dx.doi.org/10.3109/09540261.2015.1106446

Rimes, K.A., Goodship, N., Ussher, G, Baker, D., & West, E. (2017). Non-binary and binary transgender youth: Comparison of mental health, self-harm, suicidality, substance use, and

victimization experiences. *International Journal of Transgenderism*. Advance online publication. doi: 10.1080/15532739.2017.1370627

Rolstad, S., Adler, J., & Rydén, A. (2011). Response burden and questionnaire length: is shorter better? a review and meta-analysis. *Value in Health, 14*, 1101-1108. doi: https://doi.org/10.1016/j.jval.2011.06.003

Ruscio, J., & Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychological Assessment, 24*, 282-292. doi: http://psycnet.apa.org/doi/10.1037/a0025697

Schick, V. R., Calabrese, S. K., Rima, B. N., & Zucker, A. N. (2010). Genital appearance dissatisfaction: Implications for women's genital image self-consciousness, sexual esteem, sexual satisfaction, and sexual risk. *Psychology of Women Quarterly, 34*, 394-404.

Scottish Trans Alliance (2015). *Non-binary people's experiences in the UK*. Edinburgh: Scottish Trans.

Skevington, S. M., Lotfy, M., & O'Connell, K. A. (2004). The World Health Organization's WHOQOL-BREF quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL group. *Quality of Life Research, 13*, 299-310. doi: 10.1023/0000018486.91360.00

Snyder, C. F., Watson, M. E., Jackson, J. D., Cella, D., Halyard, M. Y., & Mayo/FDA Patient-Reported Outcomes Consensus Meeting Group (2007). Patient-reported outcome instrument selection: Designing a measurement strategy. *Value in Health, 10*, S76-S85. doi: https://doi.org/10.1111/j.1524-4733.2007.00270.x

Stevens, J. P. (2002). *Applied Multivariate Statistics for the Social Sciences (4th ed.)*. Hillsdale, NJ: Erlbaum.

Thorne, N., Witcomb, G.L., Nieder, T. O., Nixon, E., & Arcelus, J. (2018). A comparison of mental health symptomatology and levels of social support in young treatment seeking

transgender individuals who identify as binary and non-binary. *International Journal of Transgenderism.* Advance online publication.

Turner, R. R., Quittner, A. L., Parasuraman, B. M., Kallich, J. D., Cleeland, C. S., & Mayo/FDA Patient-Reported Outcomes Consensus Meeting Group (2007). Patient-reported outcomes: Instrument development and selection issues. *Value in Health*, *10*, S86-S93. doi: https://doi.org/10.1111/j.1524-4733.2007.00271.x

UK Trans Info. (2016). *Current waiting times and patient population for gender identity services in the UK*. Retrieved from: https://uktrans.info/waitingtimes

T'Sjoen, G., Motmans, J., Arcelus, J., & Bouman, W.P. (2017). The need of patient involvement in transgender healthcare research. *Journal of Sexual Medicine*, 14(12), 1494-1495.

van de Grift, T. C., Kreukels, B. P. C., Elfering, L., Özer, M., Bouman, M-B., Buncamper, M. E., ... Mullender, M. G. (2016). Body image in transmen: Multidimensional measurement and the effects of mastectomy. *Journal of Sexual Medicine*, *13*, 1778-1786. doi: https://doi.org/10.1016/j.jsxm.2016.09.003

van de Grift, T.C., Elaut, E., Cerwenka, S.C., Cohen-Kettenis, P.T., & Kreukels, B.P.C. (2018). Surgical satisfaction, quality of life, and their association after gender-affirming surgery: A follow-up study. *Journal of Sexual and Marital Therapy*, *44*(2), 138-148.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*, 321-327.

Warren, J.C., Bryant Smalley, K., & Nikki Barefoot, K. (2016). Psychological well-being among transgender and genderqueer individuals. *International Journal of Transgenderism*, 17 (3-4), 114-123. doi: http://dx.doi.org/10.1080/15532739.2016.1216344

Witcomb, G.L., Bouman, W.P., Claes, L., Brewin, N., Crawford, J., & Arcelus, J. (2018). Levels of depression in transgender people and its predictors: Results of a large

matched control study with transgender people accessing clinical services. *Journal of Affective Disorders*. Advanced online publication. doi: 10.1016/j.jad.2018.02.051

Woertman, L., & van den Brink, F. (2012). Body image and female sexual functioning and behavior: A review. *Journal of Sex Research, 49*, 184-211. doi: http://dx.doi.org/10.1080/00224499.2012.658586

Wylie, K., Barrett, J., Besser, M., Bouman, W. P., Bridgman, M., Clayton, A., …Ward, D. (2014). Good practice guidelines for the assessment and treatment of adults with Gender Dysphoria. *Sexual and Relationship Therapy, 29*, 154-214. doi: http://dx.doi.org/10.1080/14681994.2014.883353

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica, 67*, 361-370. doi: 10.1111/j.1600-0447.1983.tb09716.x

Appendices

Gender Congruence and Life Satisfaction Scale (GCLS)

Jones, Bouman, Haycraft and Arcelus

Citation and tool use: The GCLS is freely available for use. Please contact the authors if you have any questions about the scale:

Beth Jones: B.Jones@lboro.ac.uk
Prof Jon Arcelus: Jon.Arcelus@nottingham.ac.uk

If you would like to use the tool in a different language (i.e., translate it) then please contact the authors as they may already have it available in your language. This will avoid several versions of the tool being created in the same language. We hope you find this tool useful.

Scoring instructions: Items are scored on a 5-point scale (always=1; never=5). Items with an asterisk (indicated below) need to be reverse scored (always=5, never=1). A **higher score** therefore indicates a **more positive outcome** (higher gender congruence, better gender-related mental well-being, and better general life satisfaction).

To score item 26, if the participant indicates (by ticking the box) that they have had genital surgery, score their response as missing data. If they have not ticked the box, score their response as for the other GCLS items (always=1; never=5).

To obtain subscale scores and the global score, calculate the mean score. The global score comprises all 38 items. The items that comprise each subscale are listed below:

1. **Genitalia**: 14, 21, 25*, 26, 27 and 29 (6 items)
2. **Chest**: 15, 18, 28 and 30* (4 items)
3. **Other secondary sex characteristics**: 17, 23 and 24 (3 items)
4. **Social gender role recognition**: 16*, 19, 20* and 22* (4 items)
5. **Physical and emotional intimacy**: 3, 5, 32* and 33* (4 items)
6. **Psychological functioning**: 1, 2, 4, 6, 7, 8, 9, 11, 12 and 13 (10 items)
7. **Life satisfaction**: 10, 31*, 34*, 35, 36*, 37 and 38* (7 items)

The subscales can also be grouped into two clusters. Cluster 1, **gender congruence**, comprises subscales 1-4, and cluster 2, **gender related mental well-being and life satisfaction**, comprises subscale 5-7. To obtain the cluster scores, calculate the mean of all the items in that cluster:

Cluster 1 (gender congruence): 14, 15, 16*, 17, 18, 19, 20*, 21, 22*, 23, 24, 25*, 26, 27, 28, 29, 30*
Cluster 2 (gender-related mental well-being and general life satisfaction): 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 31*, 32*, 33*, 34*, 35, 36*, 37, 38*

Gender Congruence and Life Satisfaction Scale (GCLS)

Below is a range of statements about how you might feel in relation to your gender, mental well-being, and life satisfaction. Please respond to each statement, thinking about how frequently you have felt like this in the past **6 months**.

Please rate each statement as: **NEVER (N), RARELY (R), SOMETIMES (S), OFTEN (O) or ALWAYS (A).**

Please note that when talking about ***'gender identity'*** we mean one's internal sense of one's self as a man, a woman, or some other gender.

In the past **6 months,** due to the distress about my gender (i.e., the distress caused as the gender I was assigned at birth does not match with my gender identity):					
1. I have avoided social situations and/or social interactions	N	R	S	O	A
2. I have not gone to school/college/work	N	R	S	O	A
3. I have not been able to have emotional relationships with other people	N	R	S	O	A
4. I have suffered from anxiety	N	R	S	O	A
5. I have not been able to be physically intimate with other people	N	R	S	O	A
6. I have been unable to leave the house	N	R	S	O	A
7. I have found it difficult to make friends	N	R	S	O	A
8. I have thought about cutting or hurting my chest, genitals and/or surrounding areas	N	R	S	O	A
9. I have felt that life is meaningless	N	R	S	O	A
10. I have not enjoyed life	N	R	S	O	A
11. I have not engaged in leisure activities	N	R	S	O	A
12. I have suffered from low mood	N	R	S	O	A
13. I have thought about hurting myself or taking my own life	N	R	S	O	A

GCLS continues over…

Please rate each statement as: **NEVER (N), RARELY (R), SOMETIMES (S), OFTEN (O) or ALWAYS (A).**

In the past **6 months:**					
14. I have felt distressed when touching my genitals as they do not match my gender identity	N	R	S	O	A
15. I have felt so distressed about my chest that I have not been able to have a fulfilling life	N	R	S	O	A
16. I have felt comfortable with how others have perceived my gender	N	R	S	O	A
17. I have felt that my body hair conflicts with my gender identity, either because I have it and do not like it or because I would like to have it	N	R	S	O	A
18. I have felt like my chest does not match my gender identity	N	R	S	O	A
19. I have found it distressing that others do not address me according to my gender identity	N	R	S	O	A
20. I have felt satisfied with the pronouns that others use when talking about me	N	R	S	O	A
21. I have felt unhappy about my genitalia since they do not match my gender identity	N	R	S	O	A
22. I have felt comfortable with how other people perceive my gender based on my physical appearance	N	R	S	O	A
23. I have felt that my voice has affected the way other people have perceived my gender identity which has been distressing for me	N	R	S	O	A
24. I have felt that my facial hair conflicts with my gender identity, either because I have it and do not like it or because I would like to have it	N	R	S	O	A
25. I have felt that my genitals do match with my gender identity	N	R	S	O	A
26. I have felt that genital surgery will address the unhappiness I experience in relation to my gender ☐ Tick here if you have already had genital surgery (unless you feel you need more)	N	R	S	O	A
27. I have been unable to have a fulfilling life because of the distress relating to my genitalia	N	R	S	O	A
28. I have felt extremely distressed when looking at my chest	N	R	S	O	A
29. I have felt extremely distressed when looking at my genitals	N	R	S	O	A
30. I have felt satisfied with my chest	N	R	S	O	A

GCLS continues over…

Please rate each statement as: **NEVER (N), RARELY (R), SOMETIMES (S), OFTEN (O) or ALWAYS (A).**

Next, we would like to know how satisfied you have been with your life for the last **6 months:**					
31. I have felt satisfied at school/college/work	N	R	S	O	A
32. I have felt satisfied with my emotional relationship(s)	N	R	S	O	A
33. I have felt satisfied with my sex life	N	R	S	O	A
34. I have felt satisfied in my leisure activities and hobbies	N	R	S	O	A
35. I have not felt satisfied with my friends	N	R	S	O	A
36. I have felt satisfied with the support I have received from other significant people	N	R	S	O	A
37. I have not felt satisfied with my health	N	R	S	O	A
38. I have felt satisfied with life in general	N	R	S	O	A