Currently, the local Medicare Administrative Contractors (MACs) determine coverage of gender reassignment surgery on an individual claim basis. The Centers for Medicare & Medicaid Services (CMS) proposes to continue this practice and not issue a National Coverage Determination (NCD) at this time on gender reassignment surgery for Medicare beneficiaries with gender dysphoria. Our review of the clinical evidence for gender reassignment surgery was inconclusive for the Medicare population at large. The low number of clinical studies specifically about Medicare beneficiaries’ health outcomes for gender reassignment surgery and small sample sizes inhibited our ability to create clinical appropriateness criteria for cohorts of Medicare beneficiaries. CMS reaffirms our commitment to eliminate health care disparities and advance the health of all minority populations including the transgender community. This position is in alignment with the Affordable Care Act (ACA) which established the CMS Office of Minority Health and the CMS Minority Health Equity Plan.

Our conclusion that an NCD is not warranted does not dispute the medical necessity of transition-related care on a case-by-case basis in accordance with accepted standards of care. In the absence of a NCD, initial coverage determinations under section 1862(a)(1)(A) of the Social Security Act (the Act) and any other relevant statutory requirements will be made by the local Medicare Administrative Contractors (MACs) on an individual claim basis pursuant to Department of Health and Human Services Departmental Appeals Board Docket No. A-13-87, Decision No. 2576.

While we are not issuing a NCD, CMS supports and encourages the relatively nascent efforts of U.S.-based researchers to provide evidence-based robust clinical guidance on how studies that will fill the evidence gaps and help inform the answer to the question posed in this proposed decision memorandum. Based on the gaps identified in the clinical evidence, these studies should focus on which patients are most likely to best advocate and provide achieve improved health outcomes for the transgender community. This includes our vigorous support with gender reassignment surgery, which types of surgery are most appropriate, and what types of federally-funded quality research for this cohort physician criteria and care setting(s) are needed to ensure that patients achieve improved health outcomes.

We are requesting public comments on this proposed decision memorandum pursuant to section 1862(l)(3)(a) of the Act. We are specifically interested in public comments on the evidence we cited in this decision, comments containing any new evidence that has not been considered, and comments on whether a study could be developed that would support coverage with evidence development (CED), which would only cover gender reassignment surgery for beneficiaries who choose to participate in a clinical study.

---

**Proposed Decision Memo**

To: Administrative File: CAG #00446N
I. Proposed Decision

Currently, the local Medicare Administrative Contractors (MACs) determine coverage of gender reassignment surgery on an individual claim basis. The Centers for Medicare & Medicaid Services (CMS) proposes to continue this practice and not issue a National Coverage Determination (NCD) at this time on gender reassignment surgery for Medicare beneficiaries with gender dysphoria. Our review of the clinical evidence for gender reassignment surgery was inconclusive for the Medicare population at large. The low number of clinical studies specifically about Medicare beneficiaries’ health outcomes for gender reassignment surgery and small sample sizes inhibited our ability to create clinical appropriateness criteria for cohorts of Medicare beneficiaries. The small sizes and limited number of US studies are in part due to the fact that this is a rare disease and that in the United States, public and private health insurance have almost universally precluded payment for gender dysphoria until the last 10-15 years. In addition, because there are adequate studies (mostly from European centers where this care is covered under national health insurance programs) demonstrating that there is some (albeit poorly defined) benefits from treatment and significant harms such as suicidality in untreated gender dysphoria, the gold standard randomized controlled trial of gender reassignment surgery would not be ethically acceptable to perform at this time based on national and international guidelines for ethical performance of human research.

Our conclusion that an NCD is not warranted does not dispute the medical necessity of transition-related care on a case-by-case basis in accordance with accepted standards of care.

---

1. [http://www.fda.gov/Drugs/ResourcesForYou/Consumers/ucm143563.htm](http://www.fda.gov/Drugs/ResourcesForYou/Consumers/ucm143563.htm)
the absence of a NCD, initial coverage determinations under section 1862(a)(1)(A) of the Social Security Act (the Act) and any other relevant statutory requirements will be made by the local Medicare Administrative Contractors (MACs) on an individual claim basis, pursuant to Department of Health and Human Services Departmental Appeals Board Docket No. A-13-87, Decision No. 2576. We also recognize that an NCD on sex reassignment surgery as a whole instead of specific individual types of sex reassignment surgery is a broad scope that conflates the results of specific surgeries that may have more beneficial effects with those for whom the benefits are less or simply less well defined due to the limitations of the existing research.

While we are not issuing a NCD, CMS encourages robust clinical studies that will fill the evidence gaps and help inform the answer to the question posed in this proposed decision memorandum. Based on the gaps identified in the clinical evidence, these studies should focus on which patients are most likely to achieve improved health outcomes with gender reassignment surgery, which types of surgery are most appropriate, and what types of physician criteria and care setting(s) are needed to ensure that patients achieve improved health outcomes.

We are requesting public comments on this proposed decision memorandum pursuant to section 1862(l)(3)(a) of the Act. We are specifically interested in public comments on the evidence we cited in this decision, comments containing any new evidence that has not been considered, and comments on whether a study could be developed that would support coverage with evidence development (CED), which would only cover gender reassignment surgery for beneficiaries who choose to participate in a clinical study.

II. Background

Below is a list of acronyms used throughout this document.

AHRQ - Agency for Healthcare Research and Quality  
AIDS - Acquired Immune Deficiency Syndrome  
ANOVA - Analysis of Variance  
APA - American Psychiatric Association  
APGAR - Adaptability, Partnership Growth, Affection, and Resolve test  
BIQ - Body Image Questionnaire  
BSRI - Bem Sex Role Inventory  
CCEI - Crown Crisp Experimental Index  
CHIS - California Health Interview Survey  
CI - Confidence Interval  
CMS - Centers for Medicare & Medicaid Services  
DAB - Departmental Appeals Board  
DSM - Diagnostic and Statistical Manual of Mental Disorders  
EMBASE - Exerpta Medica database  
FBeK - Fragebogen zur Beurteilung des eigenen Korpers  
FDA - Food and Drug Administration  
FPI-R - Freiburg Personality Inventory  
FSFI - Female Sexual Function Index  
GAF - Global Assessment of Functioning
A. Diagnostic Criteria

The criteria for gender dysphoria or spectrum of related conditions as defined by the American Psychiatric Association (APA) in the Diagnostic and Statistical Manual of Mental Disorders (DSM) has changed over time (See Appendix A).

Gender dysphoria (previously known as gender identity disorder) is a classification used to describe persons who experience significant discontent with their biological sex and/or gender
assigned at birth. Therapeutic options for gender dysphoria include behavioral and psychotherapies, hormonal treatments, and a number of surgeries used for gender reassignment. Speech therapy is also sometimes employed. This proposed decision is only focusing on gender reassignment surgery.

B. Prevalence of Gender Dysphoria

Prevalence of gender dysphoria estimates have been reported by several investigators.

For estimates of transgender individuals in the U.S., we looked at several studies.

The Massachusetts Behavior Risk Factor Surveillance Survey (via telephone) (2007 and 2009) found that 0.5% individuals self-identified as transgender (Conron et al., 2012). This study did not differentiate between those people who had undergone (or desired to undergo) therapies for gender reassignment, the number who would be of interest to this analysis. “Transgender” is an umbrella term that some people self-identify with that can include gender non-conforming people who do not necessarily carry a diagnosis of gender dysphoria but nonetheless identify with the broad definition of the term “transgender” identified 0.5% individuals as transgender (Conron et al., 2012).

Derivative data obtained from the 2004 California Lesbian Gay Bisexual and Transgender (LGBT) Tobacco Survey (via telephone) and the 2009 California Health Interview Survey (CHIS) (via telephone) suggested the LGB population constitutes 3.2% of the California population and that transgender subjects constitute approximately 2% of the California LGBT population and 0.06% of the overall California population (Bye et al., 2005; CHIS 2009; Gates, 2011).

In a recent review of Medicare claims data, CMS estimated that in calendar year 2013 there were at least 4,098 transgender beneficiaries (less than 1% of the Medicare population) who utilized services paid for by Medicare, of which 90% had confirmatory diagnosis, billing codes, or evidence of a hormone therapy prescription. The Medicare transgender population is racially and ethnically diverse (e.g., 74% White, 15% African American) and spans the entire country. The following states have at least 100 transgender beneficiaries: California, Florida, Georgia, Illinois, Massachusetts, Michigan, Minnesota, New York, Pennsylvania, Ohio, Texas, Washington, and Wisconsin. Nearly 80% of transgender beneficiaries are under age 65, including approximately 23% ages 45-54. The most prevalent chronic conditions of note for the transgender population under 65 years of age were depression, major depressive affective disorder, and anxiety. Approximately 75% of transgender Medicare beneficiaries have been affected by depression, which is a disproportionately high amount compared to the Medicare population as a whole with 14% of Medicare fee-for-service beneficiaries suffering from the disease (CMS, Chronic Conditions Among Medicare Beneficiaries, 2012 at https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/chronic-conditions/downloads/2012chartbook.pdf). Based on the claims data, about 48% of transgender beneficiaries use hormone therapy, which are coverable under the Medicare Part D prescription drug benefit program (CMS Office of Minority Health (2015, June). New Directions in CMS Disparities Research: Sexual Orientation & Gender Identity. Paper presented at the Academy
For international comparison purposes, recent estimates of transgender populations in other countries are similar to those in the United States. New Zealand researchers, using passport data, reported a prevalence of 0.0275% for male-to-female adults and 0.0044% female-to-male adults (6:1 ratio) (Veale, 2008). Researchers from a center of transgender treatment and reassignment surgery in Belgium conducted a survey of regional plastic surgeons and reported a prevalence of 0.008% male-to-female and 0.003% female-to-male (ratio 2.7:1) surgically reassigned transsexuals in Belgium (De Cuypere et al., 2007). Swedish researchers, using national mandatory reporting data on those requesting reassignment surgery, reported secular changes over time in that the number of completed reassignment surgeries per application increased markedly in the 1990s; the male-to-female/female-to-male sex ratio changed from 1:1 to 2:1; the age of male-to-female and female-to-male applicants was initially similar, but increased by eight years for male-to-female applicants; and the proportion of foreign born applicants increased (Olsson, Moller 2003).

C. Interventions

Table 1 provides information about some of the types of therapeutic interventions for transgender individuals.

<table>
<thead>
<tr>
<th>Treatment Category</th>
<th>Male to Female</th>
<th>Female to Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>HORMONAL(^1)</td>
<td>Estrogens</td>
<td>Androgens</td>
</tr>
<tr>
<td>Core</td>
<td>Anti-androgens (e.g., spironolactone, 5-α reductase blockers, androgen receptor blockers, GnRH analogues)</td>
<td>Progestins/GnRH analogues for menses suppression as needed after 1 yr of androgens</td>
</tr>
<tr>
<td>SURGICAL(^2,3)</td>
<td>Orchidectomy (testes)</td>
<td>Hysterectomy (uterus) and Salpingo-oophorectomy (fallopian tubes + ovaries)</td>
</tr>
<tr>
<td>Natal Internal Genital Removal</td>
<td>Penectomy</td>
<td>NA</td>
</tr>
<tr>
<td>Natal External Genital Removal</td>
<td>NA</td>
<td>Mastectomy</td>
</tr>
<tr>
<td>Breast Removal</td>
<td>NA</td>
<td>Mastectomy</td>
</tr>
<tr>
<td>Genital Reconstruction(^2)</td>
<td>Vaginoplasty</td>
<td>Metoidioplasty or Phalloplasty</td>
</tr>
<tr>
<td></td>
<td>Clitoroplasty</td>
<td>Inflatable/rigid penile prosthesis</td>
</tr>
<tr>
<td></td>
<td>Labioplasty</td>
<td>insertion</td>
</tr>
<tr>
<td></td>
<td>Urethrostomy</td>
<td>Scrotal reconstruction</td>
</tr>
</tbody>
</table>

RH=gonadotropin releasing hormone  NA=not applicable ?=possible ↑=increased 2o=secondary
\(^1\)—Bowman et al., 2012; Deutch, 2015; Elaut et al., 2010; Gooren et al., 2005, 2013, and 2014; Heresova, 1986; Jacobiet, et al., 2009; Kronawitter et al., 2009; Meuller, 2010; Meyer et al., 1981; Pelusi et al., 2014; Schlatterer et
III. History of Medicare Coverage

CMS does not currently have an NCD on gender reassignment surgery. Previously, NCD 140.3 (“Transsexual Surgery”) barred coverage for gender reassignment surgeries. The HHS Departmental Appeals Board found NCD 140.3 to be invalid in May 2014. The Board's analysis concluded that gender reassignment surgery is a safe and effective treatment for gender dysphoria.

A. Current Request

On December 3, 2015, CMS accepted a formal complete request from a beneficiary to initiate a national coverage analysis (NCA) for gender reassignment surgery.

CMS opened this National Coverage Analysis (NCA) to thoroughly review the evidence to determine whether or not gender reassignment surgery may be covered nationally under the Medicare program.

B. Benefit Category

Medicare is a defined benefit program. For an item or service to be covered by the Medicare program, it must fall within one of the statutorily defined benefit categories as outlined in the Act. For gender reassignment surgery, the following are statutes are applicable to coverage:

Under §1812 (Scope of Part A)
Under §1832 (Scope of Part B)
Under §1861(s) (Definition of Medical and Other Health Services)
Under §1861(s)(1) (Physicians’ Services)

This may not be an exhaustive list of all applicable Medicare benefit categories for this item or service.

IV. Timeline of Recent Activities

Table 2: Timeline of Medicare Coverage Policy Actions for Gender Reassignment Surgery

<table>
<thead>
<tr>
<th>Date</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>August 1, 1989</td>
<td>The Health Care Financing Agency (HCFA; predecessor agency to CMS) published the initial NCD, titled “140.3, Transsexual Surgery” in the Federal Register. (54 Fed. Reg. 34,555, 34,572)</td>
</tr>
</tbody>
</table>
May 30, 2014  The HHS Departmental Appeals Board (DAB) determined that the NCD denying coverage for all transsexual surgery was not valid. As a result, MACs determined coverage on a case-by-case basis.

December 3, 2015  CMS accepts an external request to open an NCD. A tracking sheet was posted on the web site and the initial 30 day public comment period commenced.

January 2, 2016  Initial comment period closed. CMS received 103 comments.

V. FDA Status

Surgical procedures per se are not subject to the Food and Drug Administration’s (FDA) approval.

Inflatable penile prosthetic devices, rigid penile implants, testicular prosthetic implants, and breast implants have been approved/cleared by the FDA.

VI. General Methodological Principles

In general, when making national coverage determinations, CMS evaluates relevant clinical evidence to determine whether or not the evidence is of sufficient quality to support a finding that an item or service is reasonable and necessary for the diagnosis or treatment of illness or injury or to improve the functioning of a malformed body member. (§ 1862 (a)(1)(A)). The evidence may consist of external technology assessments, internal review of published and unpublished studies, recommendations from the Medicare Evidence Development & Coverage Advisory Committee (MEDCAC), evidence-based guidelines, professional society position statements, expert opinion, and public comments.

The overall objective for the critical appraisal of the evidence is to determine to what degree we are confident that: 1) specific clinical question relevant to the coverage request can be answered conclusively; and 2) the extent to which we are confident that the intervention will improve health outcomes for patients.

A detailed account of the methodological principles of study design the agency staff utilizes to assess the relevant literature on a therapeutic or diagnostic item or service for specific conditions can be found in Appendix B. In general, features of clinical studies that improve quality and decrease bias include the selection of a clinically relevant cohort, the consistent use of a single good reference standard, blinding of readers of the index test, and reference test results.

Public commenters sometimes cite the published clinical evidence and provide CMS with useful information. Public comments that provide information based on unpublished evidence, such as the results of individual practitioners or patients, are less rigorous and, therefore, less useful for making a coverage determination. CMS uses the initial comment period to inform the public of its proposed decision. CMS responds in detail to the public comments that were received in response to the proposed decision when it issues the final decision memorandum.
VII. Evidence

A. Introduction

Below is a summary of the evidence we considered during our review, primarily articles about clinical trials published in peer-reviewed medical journals. We considered articles cited by the requestor, in public comments, as well as those found by a CMS literature review. Citations are detailed below.

B. Literature Search Methods

CMS staff extensively searched for primary studies evaluating therapeutic interventions for gender dysphoria. There was particular emphasis on the various surgical interventions, but other treatments including hormone therapy, psychotherapy, psychiatric treatment, ancillary reproductive and gender modifying services, and post-operative surveillance services for natal sex organs were also included because of their serial and sometimes overlapping roles in patient management. The emphasis focused less on specific surgical techniques and more on functional and qualitative outcomes unless specific techniques altered those types of outcomes.

The reviewed evidence included articles obtained by searching literature databases and technology review databases from PubMed (1965 to current date), EMBASE, the Agency for Healthcare Research and Quality (AHRQ), the Blue Cross/Blue Shield Technology Evaluation Center, the Cochrane Collection, the Institute of Medicine, and the National Institute for Health and Care Excellence (NICE) as well as the source material for commentary, guidelines, and formal evidence-based documents published by professional societies. Systematic reviews were used to help locate some of the more obscure publications and abstracts.

Keywords used in the search included: Trans-sexual, transgender, gender identity disorder (syndrome), gender dysphoria and/or hormone therapy, gender surgery, genital surgery, gender reassignment (surgery), sex reassignment (surgery) AND/OR quality of life, satisfaction-regret, psychological function (diagnosis of mood disorders, psychopathology, personality disorders), employment status, relationships, other social function, suicide (attempts), mortality, sexual function, urinary function, and adverse events-reoperations. After the identification of germane publications, CMS also conducted searches on the specific psychometric instruments used by investigators.

Psychometric instruments are scientific tools used to measure individuals' mental capabilities and behavioral style. They are usually in the form of questionnaires that numerically capture responses. These tools are used to create a psychological profile that can address questions about a person’s knowledge, abilities, attitudes and personality traits. In the evaluation of patients with gender dysphoria, it is important that both validity and reliability be assured in the construction of the tool (validity refers to how well the tool actually measures what it was designed to measure, or how well it reflects the reality it claims to represent, while reliability refers to how accurately results of the tool would be replicated in a second identical piece of research). That is because when evaluating patients with this condition most of the variables of interest (e.g.,
satisfaction, anxiety, depression) are latent in nature (not directly observed but are rather inferred) and difficult to quantify objectively.

Studies with robust study designs and larger, defined patient populations assessed with objective endpoints or validated test instruments were given greater weight than small, pilot studies. Reduced consideration was given to studies that were underpowered for the assessment of differences or changes known to be clinically important. Studies with fewer than 30 patients were reviewed and delineated, but excluded from the major analytic framework. Oral presentations, unpublished white papers, and case reports were excluded. Publications in languages other than English were excluded.

Included studies were limited to those with adult subjects. Review and discussion of the management of children and adolescents with the additional considerations of induced pubertal delay are outside the scope of this NCD. In cases where the same population was studied for multiple reasons or where the patient population was expanded over time, the latest and/or most germane sections of the publications were analyzed. The excluded duplicative publications are delineated.

CMS also searched Clinicaltrials.gov to identify relevant clinical trials. CMS looked at trial status including early termination, completed, and ongoing with sponsor update, and ongoing with estimated date of completion. Publications on completed trials were sought. The CMS internal search was limited to articles published prior to March 21, 2016. CMS reviewed results of clinical trials involving adult human subjects; to reports of prospective (e.g., blinded, non-blinded, cross sectional), partially prospective, retrospective longitudinal studies randomized meeting certain criteria.

CMS acknowledges that gender dysphoria is by U.S. definition a rare disease, and a very small subset of the at-large pool of Medicare beneficiaries and therefore, a large prospective, double-blinded, randomized, controlled study of gender reassignment surgery outcomes may be impractical. In addition, given the current evidence base and standards of practice in the U.S. and other developed nations, performing a true randomized controlled trial would be unethical.

C. Discussion of Evidence

The development of an assessment in support of Medicare coverage determinations is based on the same general question for almost all national coverage analyses (NCAs): "Is the evidence sufficient to conclude that the application of the item or service under study will improve health outcomes for Medicare patients?" CMS is interested in answering the following question:

**Is there sufficient evidence to conclude that gender reassignment surgery significantly improves health outcomes for some Medicare beneficiaries with gender dysphoria?**

The evidence reviewed is directed towards answering this question.
1. Internal Technology Assessment

When looking at the studies evaluating gender reassignment surgery for patients with gender dysphoria, we found an array of disparate research designs. Most of the studies were conducted in Europe. Only six studies took place in the U.S. (Ainsworth, Spiegel, 2010; Beatrice, 1985; Meyer, Reter, 1979; Newfield et al., 2006; Lawrence, 2006; Leinung et al., 2013). Most of the studies that evaluated gender dysphoria were descriptive in nature; few made inferences which may be applicable to the Medicare population.

CMS conducted an extensive literature search on gender reassignment related surgical procedures and on facets of gender dysphoria that provide context for this analysis. The latter includes medical and environmental conditions. CMS also explored the relative roles that psychological support, mental health care, cross-sex hormonal therapy, and the various gender reassignment related surgical procedures played in health outcomes.

CMS identified numerous publications related to gender reassignment surgery. A large number of these were case reports, case series with or without descriptive statistics, or studies with population sizes too small to conduct standard parametric statistical analyses. Others addressed issues of surgical technique.

CMS identified and described 36 publications on gender reassignment surgery that included health outcomes. Because the various investigators at a site sometimes conducted serial studies on ever-enlarging cohort populations, studied sub-populations, studied different outcomes, or used different tools to study the same outcomes, not all study populations were unique. To reduce bias from over-lapping populations, only the latest or most germane publication(s) were described. Subsumed publications were delineated.

Of these 36 publications, two publications used different assessment tools on the same population, and, so for the purposes of evaluation, were classified as 1 study (Udeze et al., 2008; Megeri, Khoosal, 2007). For another publication, the complete manuscript could not be located despite an exhaustive search by the Library of Medicine (Barrett, 1998). This precluded adequate review, thus, it was not included. A total of 33 studies were reviewed (See Figure 1). Appendices C, D, and F include more detail of each study.

The publications covered a time span from 1979 to 2015. Over half of the studies were published after 2005.

Figure 1. Studies of Gender Reassignment Surgery (GRS)
ANOVA=Analysis of Variance Normative=Psychometric Tests with known normative for large populations

The studies in Figure 1 are categorized into 3 groups. The first group, depicted by the colored boxes (red, blue, and green), had explicit controls. There was a single randomized study. The remainder in the first group were observational studies. These were subdivided into longitudinal studies and cross-sectional studies. The second group, depicted by black boxes (starting with the surgery only populations box) consisted of surgical series. The third group, depicted by black boxes (starting with mixed population), was composed of mixed populations of patients not stratified by treatment and which included a spectrum of therapeutic interventions.

When looking at the totality of studies, they fell into the following research design groups:

a. **Prospective, non-blinded, observational, cross-sectional studies with no concurrent controls**

Ainsworth and Spiegel conducted a prospective, observational study using a cross-sectional design and a partially self-designed survey tool. The blind status is unknown. Treatment types served as the basis for controls. The investigators, head and neck surgeons who provided facial feminization services, assessed perception of appearance and quality of life in male-to-female subjects with self-reported gender dysphoria. Patients could have received no therapeutic intervention, hormone therapy, reassignment surgery, and/or facial feminization surgery and an unrestricted length of transition. (Transition refers to the time when a transgender person begins to live as the gender with which they identify rather than the gender assigned at birth.) Criteria for the various types of interventions were not available because of the survey design of the study. Patients were recruited via website or at a 2007 health conference. Pre-specified controls to eliminate duplicate responders were not provided. The investigators employed a self-designed Likert-style facial feminization outcomes evaluation questionnaire and a “San Francisco 36” health questionnaire. No citations were provided for the latter. It appears to be the Short-form (SF) 36-version 2. Changes or differences considered to be biologically significant were not pre-specified. Power corrections for multiple comparisons were not provided.

The investigators reported that there were 247 participants. (The numbers of incomplete questionnaires was not reported.) Of the 247 participants, 25 (10.1%) received only primary sex trait reassignment surgery, 28 (11.3%) received facial surgery without primary sex trait reassignment surgery, 47 (19.0%) received both facial and primary sex trait reassignment surgery, and 147 (59.5%) received neither facial nor reassignment surgery. The mean age for each of these cohorts was: 50 (no standard deviation [S.D.]) only reassignment surgery, 51 (no S.D.) only facial surgery, 49 (no S.D.) both types of surgery, and 46 (no S.D.) (neither surgery). Of the surgical cohorts: 100% of those who had undergone primary sex trait reassignment surgery alone used hormone therapy, 86% of those who had undergone facial feminization used hormone therapy, and 98% of those who had undergone both primary sex trait reassignment surgery and facial feminization used hormone therapy. In contrast to the surgical cohorts, 66% of the “no surgery” cohort used hormonal therapy, and a large proportion (27%) had been in transition for less than 1 year.

The investigators reported higher scores on the facial outcomes evaluation in those who had undergone facial feminization. Scores of the surgical cohorts for the presumptive SF-36 comprehensive mental health domain did not differ from the general U.S. female population. Scores of the “no surgery” cohort for the comprehensive mental health domain were statistically lower than those of the general U.S. female population, but within 1 standard deviation of the normative mean. Mean scores of all the gender dysphoric cohorts for the comprehensive physical domain were statistically higher than those of the general female U.S. population, but were well within 1 standard deviation of the normative mean. Analyses of inter-cohort differences for the SF-36 results were not conducted. Although the investigators commented on the potential disproportionate impact of hormone therapy on outcomes and differences in the time in “transition”, they did not conduct any statistical analyses to correct for putative confounding variables.

Motmans et al., conducted a prospective, non-blinded, observational study using a cross-sectional design and a non-specific quality-of-life tool. No concurrent controls were used in this study. Quality of life in this Dutch-speaking population was assessed using the Dutch version of a SF-36 (normative data was used). Participants included subjects who were living in accordance with the preferred gender and who were from a single, unspecified, Belgian university specialty clinic at Ghent. The Dutch version of the SF-36 questionnaire along with its normative data were used. Variables explored included employment, pension status, ability to work, being involved in a relationship. Also explored, was surgical reassignment surgery and the types of surgical interventions. Intragroup comparisons by transgender category were conducted, and the relationships between variables were assessed by analysis of variance (ANOVA) and correlations.

The age of the entire cohort (n=140) was 39.89±10.21 (female-to-male: 37.03±8.51; male-to-female: 42.26±10.39). Results of the analysis revealed that not all female-to-male patients underwent surgical reassignment surgery and, of those who did, not all underwent complete surgical reassignment. The numbers of female-to-male surgical interventions were: mastectomy 55, hysterectomy 55, metadoidoplasty 8 (with 5 of these later having phalloplasty), phalloplasty 40, and implantation of a prosthetic erectile device 20. The frequencies of various male-to-female surgical interventions were: vaginoplasty 48, breast augmentation 39, thyroid cartilage reduction 17, facial feminization 14, and hair transplantation 3.

The final number of subjects with SF-36 scores was 103 (49 [47.6%] female-to-male; 54 [52.4%] male-to-female; ratio 1:1.1). For this measure, the scores for the vitality and mental health domains for the final female-to-male cohort (n= 49 and not limited to those having undergone some element of reassignment surgery) were statistically lower: 60.61±18.16 versus 71.9±18.31 and 71.51±16.40 versus 79.3±16.4 respectively. Scores were not different from the normative data for Dutch women: vitality: 64.3±19.7 or mental health 73.7±18.2. None of the domains of the SF-36 for the final male-to-female cohort (n=54 and not limited to those having undergone some element of reassignment surgery) were statistically different from the normative data for Dutch women.

Analysis of variance indicated that quality-of-life as measured by the SF-36 did not differ by whether female-to-male patients had undergone genital surgery (metadoidoplasty or phalloplasty) or not. Also, ANOVA indicated that quality-of-life as measured by the SF-36 did not differ by whether male-to-female patients had undergone either breast augmentation or genital surgery (vaginoplasty) or not.

Whether there is overlap with the Ghent populations studied by Heylens et al., Weyers et al., or Wierckx et al. is unknown.

Weyers at al. 2009 conducted a prospective, non-blinded, observational study using a cross-sectional design and several measurement instruments including a non-specific quality of life tool and a semi-specific quality of life tool (using normative data) along with 2 self-designed tools.

The investigators assessed general quality of life, sexual function, and body image from the prior 4 weeks in Dutch-speaking male-to-female patients with gender dysphoria who attended a single-center, specialized, comprehensive care university clinic. Investigators used the Dutch version of the SF-36 and results were compared to normative data from Dutch women and U.S. women. The 19 items of the Dutch version of the Female Sexual Function Index (FSFI) were used to measure sexual desire, function, and satisfaction. A self-designed 7 question visual analog scale (VAS) was used to measure satisfaction with gender related body traits and appearance perception by self and others. A self-designed survey measured a broad variety of questions regarding personal medical history, familial medical history, relationships, importance of sex, sexual orientation, gynecologic care, level of regret, and other health concerns. For this study, hormone levels were also obtained.

The study consisted of 50 participants. Analysis of the data revealed that the patient’s average age was 43.1 ±10.4 years, and all of the patients had vaginoplasty. This same population also had undergone additional feminization surgical procedures (breast augmentation 96.0%, facial feminization 36.0%, vocal cord surgery 40.0%, and cricoid cartilage reduction 30.0%). A total of two (4.0%) participants reported “sometimes” regretting reassignment surgery and 23 (46.0%) were not in a relationship. For the cohort, estradiol, testosterone, and sex hormone binding globulin levels were in the expected range for the reassigned gender. The SF-36 survey revealed that the subscale scores of the participants did not differ substantively from those of Dutch and U.S. women. VAS scores of body image were highest for self-image, appearance to others, breasts, and vulva/vagina (approximately 7 to 8 of 10). Scores were lowest for body hair, facial hair, and voice characteristics (approximately 6 to 7 of 10).

The total FSFI score was 16.95±10.04 out of a maximal 36. The FSFI scores averaged 2.8 (6 point maximum): satisfaction 3.46±1.57, desire 3.12±1.47, arousal 2.95±2.17, lubrication 2.39±2.29, orgasm 2.82±2.29, and pain 2.21±2.46. Though these numbers were reported in the study, data on test population controls were not provided. VAS scores of body image were highest for self-image, appearance to others, breasts, and vulva/vagina (approximately 7 to 8 of 10). Scores were lowest for body hair, facial hair, and voice characteristics (approximately 6 to 7 of 10).

A post hoc exploration of the data revealed the following: perceived improvement in general health status was greater in the subset that had undergone reassignment surgery within the last year; sexual orientation impacted the likelihood of being in a relationship; SF-36 scores for vitality, social functioning, and mental health were nominally better for those in relationships, but that overall SF-36 scores did not differ by relationship status; sexual orientation and being in a relationship impacted FSFI scores; and reported sexual function was higher in those with higher satisfaction with regards to their appearance.

Wierckx at al. conducted a prospective, non-blinded, observational study using a cross-sectional design and several measurement instruments (a non-specific quality of life tool with reported normative data along with 3 self-designed tools). The investigators assessed general quality of life, sexual relationships, and surgical complications in Dutch-speaking female-to-male patients with gender dysphoria who attended a single-center, specialized, comprehensive care, university clinic. Investigators used the Dutch version of the SF-36 with 36 questions, 8 subscales, and 2 domains evaluating physical and mental health. Results were compared to normative data from Dutch women and Dutch men. Self-designed questionnaires to evaluate aspects of medical history, sexual functioning (there were separate versions for those with and without partners), and surgical results were also used. The Likert-style format was used for many of the questions.

A total of 79 female-to-male patients with gender dysphoria had undergone reassignment surgery were contacted; however, ultimately, 47 (59.5%) chose to participate. Three additional patients were recruited by other patients. One of the 50 participants was later excluded for undergoing reassignment surgery within the 1 year window. The age of patients was: 30±8.2 years (range 16 to 49) at the time of reassignment surgery and 37.1 ±8.2 years (range 22 to 54) at the time of follow-up. The time since hysterectomy, oopherectomy, and mastectomy was 8 years (range 2 to 22). The patient population had undergone additional surgical procedures: metaidoiplasty (n=9; 18.4%), phalloplasty (n=8 after metaidoiplasty, 38 directly; 93.9% total), and implantation of erectile prosthetic device (n=32; 65.3%). All had started hormonal therapy at least 2 years prior to surgery and continued to use androgens.

The SF-36 survey was completed by 47 (95.9%) participants. The “Vitality” and the “Mental Health” scales were lower than the Dutch male population: 62.1±20.7 versus 71.9±18.3 and 72.6±19.2 versus 79.3±16.4 respectively. These subscale scores were equivalent to the mean scores of the Dutch women.

None of the participants were dissatisfied with their hysterectomy-oopherectomy procedures; 4.1% were dissatisfied with their mastectomies because of extensive scarring; and 2.2% were dissatisfied with their phalloplasties. Of the participants, 17.9% were dissatisfied with the implantation of an erectile prosthetic device; 25 (51.0%) reported at least one post-operative complication associated with phalloplasty (e.g., infection, urethrostasis, or fistula formation); 16 (50.0% of the 32 with an erectile prosthetic device) reported at least one post-operative complication associated with implantation of an erectile prosthetic (e.g., infection, leakage, incorrect positioning, or lack of function).

A total of 18 (36.7%) participants were not in a relationship; 12.2% reported the inability to achieve orgasm with self-stimulation less than half the time; 12.2% did not respond to the question. Of those with partners with partners, 28.5% reported the inability to achieve orgasm with intercourse less than half the time and 9.7% did not respond to this question. Also, 61.3% of those with partners reported (a) no sexual activities (19.4%) or (b) activities once or twice monthly (41.9%), and there were 12.9% who declined to answer.
Post hoc assessments suggested that being in relationship or having undergone phalloplasty did not impact the scores of the SF-36 domains. Also this assessment suggested that for patients in a relationship, sexual satisfaction was related to “Vitality” scores. Finally this assessment suggested a relationship between sexual satisfaction and more frequent orgasm and pleasure with the partner.


Salvador et al. conducted a prospective, non-blinded, observational study using a cross-sectional design (albeit over an extended time interval) and a self-designed quality of life tool. The investigators assessed regret, sexual function, partnerships, and family relationships in patients who had undergone gender reassignment surgery at least 24 months prior.

Out of the 243 enrolled in the clinic over a 10 year interval, 82 underwent sex reassignment surgery. There were 69 participants with a minimum 2-year follow up, of whom 52 agreed to participate in the study for a 75% response rate. The age at follow-up was 36.3±8.9 (range 15-58) years with the time to follow-up being 3.8±1.7 (2-7) years. A total of 46 participants reported pleasurable neo-vaginal sex and post-surgical improvement in the quality of their sexual experience. The quality of sexual intercourse was rated as satisfactory to excellent, average, unsatisfactory, or not applicable in the absence of sexual contact by 84.6%, 9.6%, 1.9%, and 3.8% respectively. Of the participants, 78.8% reported greater ease in initiating and maintaining relationships; 65.4% reported having a partner; 67.3% reported increased frequency of intercourse; 36.8% reported improved familial relationships. No patient reported regret over reassignment surgery. The authors did not provide information about incomplete questionnaires.


Blanchard et al. conducted a prospective, non-blinded, cross-sectional study using a self-designed questionnaire and a non-specific psychological symptom assessment with normative data. The investigators assessed social adjustment and psychopathology in patients with gender dysphoria and who were at least 1 year post gender reassignment surgery. Reassignment surgery was defined as either vaginoplasty or mastectomy/construction of male chest contour with or without nipple transplants, but did not preclude additional procedures. Partner preference was defined as either vaginoplasty or mastectomy/construction of male chest contour with or without nipple transplants, but did not preclude additional procedures. Partner preference was not impact the scores of the SF-36 domains. Also this assessment suggested that for patients in a relationship, sexual satisfaction was related to “Vitality” scores. Finally this assessment suggested a relationship between sexual satisfaction and more frequent orgasm and pleasure with the partner.

Of the 294 patients (111 natal females and 183 natal males, ratio: 1:1.65) initially evaluated, 263 were eligible for the study and 79 patients participated in the study (38 female-to-male; 32 male-to-female with male partner preference; 9 male-to-female with female partner preference). The respective mean ages for these 3 groups were 32.6, 33.2, and 47.7 years with the last group being older statistically (p=0.01). Additional surgeries in female-to-male patients included: oophorectomy/hysterectomy 92.1% and phalloplasty 7.9%. Additional procedures in male-to-
Psychopathology as measured by the Global Severity Index of the SCL-90R was absent in all 3 patient groups. Interpretation did not differ by the sex of the normative cohort.

Of participants, 63.2% of female-to-male patients cohabitated with partners of their natal gender. 46.9% of male-to-female patients with male partner preference cohabitated with partners of their natal gender; 93.7% reported that they would definitely undergo reassignment surgery again. The remaining 6.3% (1 female-to-male; 1 male-to-female with male partner preference; 3 male-to-female with female partner preference) indicated that they probably would undertake the surgery again. Post hoc analysis suggested that the more ambivalent responders had more recently undergone surgery. Of responders, 98.7% indicated that they preferred life in the reassigned gender. The one ambivalent subject was a skilled and well compensated tradesperson who was unable to return to work in her male dominated occupation.


Tsoi conducted a prospective, non-blinded, observational study using a cross-sectional design and a self-designed quality of life tool. The investigators assessed overall life satisfaction, employment, partner status, and sexual function in gender-reassigned persons who had undergone gender reassignment surgery between 1972 and 1988 inclusive and who were approximately 2 to 5 years post-surgery. Acceptance criteria for surgery included good physical health, good mental health, absence of heterosexual tendencies, willingness to undergo hormonal therapy for ≥6 months, and willingness to function in the life of the desired gender for ≥6 months. Tsoi also undertook retrospective identification of variables that could predict outcomes.

The size of the pool of available patients was not identified. Of the 81 participants, 36 were female-to-male (44.4%) and 45 were male-to-female (55.6%) (ratio 1:1.25).

The mean ages at the time of the initial visit and operation were: female-to-male 25.4±4.4 (range 14-36) and 27.4 ±4.0; (range 14-36); male-to-female 22.9±4.6 (range 14-36) and 24.7±4.3 (14-36) years respectively. Of all participants, 14.8% were under age 20 at the time of the initial visit. All were at least 20 at the time of gender reassignment surgery. The reported age of onset was 8.6 years for female-to-male patients and 8.7 years for male-to-female patients.

All participants reported dressing without difficulty in the reassigned gender; 95% of patients reported good or satisfactory adjustment in employment and income status; 72% reported good or satisfactory adjustment in relationships with partners. Although the quality of life tool was self-designed, 81% reported good or satisfactory adjustment to their new gender, and 63% reported good or acceptable satisfaction with sexual activity. Of the female-to-male patients,
39% reported good or acceptable satisfaction with sex organ function in comparison to 91% of male-to-female patients (p<0.001). (The author reported that a fully functioning neo-phallus could not be constructed at the time.) The age of non-intercourse sexual activity was the only predictor of an improved outcome.


Gómez-Gil et al. conducted a prospective, non-blinded observational study using a cross-sectional design and non-specific psychiatric distress tools in Spain. The investigators assessed anxiety and depression in patients with gender dysphoria who attended a single-center specialty clinic with comprehensive endocrine, psychological, psychiatric, and surgical care. The clinic employed World Professional Association for Transgender Health (WPATH) guidelines. Patients were required to have met diagnostic criteria during evaluations by 2 experts. Investigators used the Hospital Anxiety and Depression Scale (HADS) and the Social Anxiety and Distress Scale (SADS) instruments. The SADS total score ranges from 0 to 28, with higher scores indicative of more anxiety. English language normative values are 9.1±8.0. HAD-anxiety and HAD-depression total score ranges from 0 to 21, with higher scores indicative of more pathology. Scores less than 8 are normal. ANOVA was used to explore effects of hormone and surgical treatment.

Of the 200 consecutively selected patients recruited, 187 (93.5% of recruited) were included in the final study population. Of the final study population, 74 (39.6%) were female-to-male patients; 113 (60.4%) were male-to-female patients (ratio 1:1.5); and 120 (64.2%) were using hormones. Of those using hormones, 36 (30.0%) were female-to-male; 84 (70.0%) were male-to-female (ratio 1:2.3). The mean age was 29.87±9.15 (range 15-61). The current age of patients using hormones was 33.6±9.1 (n=120) and older than the age of patients without hormone treatment (25.9±7.5) (p=0.001). The age at hormone initiation, however, was 24.6±8.1.

Of those who had undergone reassignment surgery, 29 (36.7%) were female-to-male; 50 (63.3%) were male-to-female; ratio 1:1.7. The number of patients not on hormones and who had undergone at least 1 gender-related surgical procedure (genital or non-genital) was small (n=2). The number of female-to-male patients on hormones who had undergone such surgery (mastectomy, hysterectomy, or phalloplasty) was 28 (77.8%). The number of male-to-female patients on hormones who had undergone such surgery (mammoplasty, facial feminization, buttock feminization, vaginoplasty, orchiectomy, and vocal feminization (thyroid chondroplasty) was 49 (58.3%).

Analysis of the data revealed that although the mean scores HAD-Anxiety, HAD-Depression, and SADS were statistically lower (better) in those on hormone therapy than in those not on hormone therapy, the mean scores for HAD-Depression and SADS were in the normal range for gender dysphoric patients not using hormones. The HAD-Anxiety score was borderline-elevated (9) in non-treated patients and normal (6.4) in treated patients. Meta-analysis of the HAD-A score showed 8 was the optimal cutoff and was 80% sensitive and specific as a screening tool for
anxiety consistent with a possible mood disorder. Thus while the in patients not using hormones. The mean scores for HAD-Anxiety, HAD-Depression, and SADS were in the normal range for gender dysphoric patients using hormones, HAD-Anxiety was significantly diminished compared with non-treated patients. In hormone treated patients 58% of the male-to-female transsexuals and 78% of the female-to-male patients. ANOVA revealed that results did not differ by whether the patient had undergone at least one gender reassignment surgery. To assure the comparability of these two groups with different frequency of having undergone surgery, ANOVA was performed and found no difference related surgical procedure or not.


Gómez-Gil et al. conducted a prospective, non-blinded observational study using a non-specific quality of life tool. There were no formal controls for this mixed population ± non-genital reassignment surgery. The investigators assessed quality of life in the context of culture in patients with gender dysphoria who were from a single-center, specialty and gender identity clinic. The clinic used WPATH guidelines. Patients were required to have met diagnostic criteria during evaluations by both a psychologist and psychiatrist. Patients could have undergone non-genital surgeries, but NOT genital reassignment surgeries (e.g., orchiectomy, vaginoplasty, or phalloplasty).

The Spanish version of the World Health Organization Quality of Life-Abbreviated version of the WHOQOL-100 (WHOQOL-BREF) was used to evaluate quality-of-life, which has 4 domains (environmental, physical, psychological, and social) and 2 general questions. Family dynamics were assessed with the Spanish version of the Family Adaptability, Partnership Growth, Affection, and Resolve (APGAR) test. Regression analysis was used to explore effects of surgical treatment.

All consecutive of the 277 patients presenting at clinic (277) were recruited and agreed to participate. However some patients did not meet inclusion criteria because of prior genital surgery, or did not ultimately complete the survey. Ultimately, of this number, 193 were included in the study (the mean age of this group was 31.2±9.9 (range 16-67). Of these, 74 (38.3%) were female-to-male patients; 119 (61.7%) were male-to-female patients; ratio 1:1.6. 120 (62.2%) were on hormone therapy; 29 (39.2%) of female-to-male patients had undergone at least 1 non-genital, surgical procedure (hysterectomy n=19 (25.7%); mastectomy n=29 (39.2%)); 51 (42.9%) of male-to-female patients had undergone at least 1 non-genital surgical procedure with mammoplasty augmentation being the most common procedure, n=47 (39.5%), followed by facial feminization, n=11 (9.2%), buttocks feminization, n=9 (7.6%), and vocal feminization (thyroid chondroplasty), n=2 (1.7%).

WHOQOL-BREF domain scores for gender dysphoric patients with and without non-genital surgery were: “Environmental” 58.81±14.89 (range 12.50-96.88), “Physical” 63.51±17.79 (range 14.29-100), “Psychological” 56.09±16.27 (range 16.67-56.09), “Social” 60.35±21.88 (range

---

8.33-100), and “Global QOL and Health” 55.44±27.18 (range 0-100). The mean APGAR family score was 7.23±2.86 (range 0-10).

Regression analysis, which was used to assess the relative importance of various factors to WHOQOL-BREF domains and general questions, revealed that family support was an important element for all 4 domains and the general health and quality-of-life questions. Hormone therapy was an important element for the general questions and for all of the domains except “Environmental.” Having undergone non-genital reassignment surgery, like age, educational levels, and partnership status, did not impact domain and general question results related to quality of life. However as the authors noted, given that 94% of their sample had undergone non-genital gender reassignment surgery, the ability to ascertain a difference between the majority of participants and the 6% who had not undergone non-genital gender reassignment surgery was significantly diminished. In addition, all types of non-genital GRS were analyzed together. This would further dilute results if certain surgeries were more beneficial than others. Thus this study was underpowered to determine a difference between transgender patients who had undergone non-genital GRS and those who had not.


Mate-Kole et al. conducted a prospective non-blinded, observational study using a cross-sectional design and 2 psychological tests (1 with some normative data). Concurrent controls were used in this study design. The investigators assessed neuroticism and sex role in natal males with gender dysphoria. Patients at various stages of management, (i.e., under evaluation, using cross-sex hormones, or post reassignment surgery [6 months to 2 years]) were matched by age of cross-dressing onset, childhood neuroticism, personal psychiatric history, and family psychiatric history. Both a psychologist and psychiatrist conducted assessments. The instruments used were the Crown Crisp Experiential Index (CCEI) for psychoneurotic symptoms and the Bem Sex Role Inventory. ANOVA was used to identify differences between the three treatment cohorts.

For each cohort, investigators recruited 50 male-to-female patients from a hospital-based Gender Identity clinic in London. The mean ages of the three cohorts were as follows: 34 years for patients undergoing evaluation, 35 years for wait-listed patients, and 37 years for post-operative patients. Of the groups under evaluation or postsurgical, 16% (8 each) were unemployed; 8% of the wait-listed patients were unemployed. For the cohorts, 22% of those under evaluation, 24% of those on hormone treatment only, and 30% of those post-surgery had prior psychiatric histories, and 24%, 24%, and 14% in each cohort, respectively, had a history of attempted suicide. More than 30% of patients in each cohort had a first degree relative with a history of psychiatric disease.

The scores for the individual CCEI domains for depression and somatic anxiety were relatively higher (worse) for patients under evaluation than those on hormone treatment alone. The scores for all of the individual CCEI domains (free floating anxiety, phobic anxiety, somatic anxiety, depression, hysteria, and obsessionality) were lower (better) in the post-operative cohort than in the other 2 cohorts.

Eldh et al. conducted a non-blinded, observational study using a prospective cross-sectional design with a self-designed questionnaire and retrospective acquisition of pre-operative data. The investigators assessed economic circumstances, family status, satisfaction with surgical results, and sexual function in patients who had undergone gender reassignment surgery. Of the 175 patients who underwent reassignment surgery in Sweden between 1965 and 1995, 136 were treated at Karolinska Hospital and sent surveys for the study. 90 responded (response rate of 66%). Of this number, 50 were female-to-male and 40 were male-to-female (ratio: 1:0.8). Patients generally satisfied with the appearance of the reconstructed genitalia, with greater satisfaction amongst those (no numbers provided). Of the patients who underwent surgery after 1985 prior to 1986, seven (14%) were dissatisfied with the size or shape of the neophallus; eight (16%) declined comment. There were 14 (35%), with 12 having surgery prior to 1986 and two between 1986 and 1995 inclusive, were moderately satisfied because of insufficient vaginal volume; eight (20%) declined comment. A neoclitoris was not constructed until the later surgical cohort. Three of 33 reported no sensation or sexual sensation. Eight had difficulties comprehending the question and did not respond.

Of the patients who underwent female-to-male genital surgery, 35 (70%) reported being satisfied with the size and shape of the neophallus and eight (16%) declined to comment. Of those who underwent surgery prior to 1986, seven (14%) reported dissatisfaction with the size or shape of the neophallus. None of the patients who underwent surgery after 1985 reported dissatisfaction. Of the patients who underwent male-to-female genital surgery, 12 who underwent surgery prior to 1986 and two who underwent surgery after 1985 were “moderately satisfied” with the depth and width of the neovagina (35% moderately satisfied). 18 (45%) were fully satisfied. Eight patients (20%) did not respond. 33 male-to-female surgeries included construction of a neoclitoris. Three (11%) reported no sensation or sexual sensation. Eight (24%) did not understand the survey question and did not respond.

Comment [RNG8]: This should have been interpreted contextually. A subject matter expert in transgender health care would have known this was very likely because at the time of the study, transgender women had to present hyperfeminine in order to get referral for HRT and SRS. So it is not surprising that people who completed all necessary treatment feel less pressure to present themselves in an overly feminine fashion.

Comment [RNG9]: Sweden keeps public records of those who have had legal sex change (which requires genital SRS to obtain). Eldh notes that 175 people in Sweden had sex reassignment from 1965 to 1995 (a matter of public records). Of these 136 had their treatment at Karolinska Hospital, Stockholm (where the authors practice and where the study was done). Patients who underwent treatment at the Karolinska Hospital are the only group that had surveys sent to them. So they actually had a 66% response rate (90 of 136). Because of good public records, the authors were able to provide some information about the 175 total, so when available they did (such as surgical complications). However the only point where response rate applicable in the entire study was to the surveys.
Nine patients (10%) reported doubts about their sexual orientation (which should not be confused with gender identity); 13 (14%) declined to answer the question; 44 (27.1%) female-to-male and 17 (42.5%) male-to-female patients had doubts about their gender orientation; 13 (26%) declined to answer the question; 44 (27.1%) female-to-male and 17 (38.6%) male-to-female were unmarried or without a steady partner; 19 (38.0%) female-to-male patients reported the absence of a sex life (14.0% declined to answer this particular question); 15 (37.5%) male-to-female reported dissatisfaction with their sex lives and Additionally, 3 (7.6%) reported absence of sexual activity post-operatively. Ten patients (11.1%) were dissatisfied with their life situation (17.8% declined to answer this question). Two patients who had attempted suicide pre-surgery also had post-surgery suicide attempts. The study found that 2 female-to-male patients and two male-to-female patients regretted their reassignment surgery and continued to live as the natal gender. Sixty-five of 90 (72%) patients were fully accepted by their families, friends, and other people. An additional 10 (11%) were partially accepted. 64 of the 74 patients surveyed (86%) were satisfied with their overall life situation, and two patients attempted suicide.


Hepp et al. conducted a prospective, non-blinded, observational study using a cross-sectional design. There was some acquisition of retrospective data. The investigators assessed current and lifetime psychiatric co-morbidity using structured interviews for diagnosis of Axis 1 disorders (clinical syndromes) and Axis 2 disorders (developmental or personality disorders) and HADS for dimensional evaluation of anxiety and depression. Statistical description of the cohort and intra-group comparisons was performed. Continuous variables were compared using t-tests and ANOVA.

A total of 31 patients with gender dysphoria participated in the study: 11 (35.5%) female-to-male; 20 (64.5%) male-to-female (ratio 1:1.8). The overall mean age was 32.2±10.3. Of the participants, seven had undergone reassignment surgery, 10 pre-surgical patients had been prescribed hormone therapy, and 14 pre-surgical patients had not been prescribed hormone therapy. Forty five and one half percent of female-to-male and 20% of male-to-female patients did not carry a lifetime diagnosis of an Axis 1 condition. Sixty three and six tenths percent of female-to-male and 60% of male-to-female patients did not carry a current diagnosis of an Axis 1 condition. Lifetime diagnosis of substance abuse and mood disorder were more common in male-to-female patients (50% and 55% respectively) than female-to-male patients (36.4% and 27.3% respectively). Current diagnosis of substance abuse and mood disorder were present in male-to-female patients (15% and 20% respectively) and absent in female-to-male patients. One or more personality disorders were identified 41.9%, but whether this was a current or lifetime condition was not specified. Of the patients, five (16.1%) had a Cluster A personality disorder (paranoid-schizoid), seven (22.6%) had a Cluster B personality disorder (borderline, anti-social, histrionic, narcissistic), six (19.4%) had a Cluster C personality disorder (avoidant, dependent, obsessive-compulsive), and two (6.5%) were not otherwise classified.
The HADS test revealed non-pathologic results for depression (female-to-male: 6.64±5.03; male-to-female: 6.58±4.21) and borderline results for anxiety (female-to-male: 7.09±5.11; male-to-female: 7.74±6.13, where a result of 7-10 = possible disorder). There were no differences by natal gender. HADS scores were missing for at least 1 person. The investigators reported a trend for less anxiety and depression as measured by HADS in the patients who had undergone surgery.

b. Prospective, non-blinded, observational, cross-sectional studies with patients serving as their own controls


Rakic et al. conducted a prospective, non-blinded, observational study using a cross-sectional design and an investigator-designed quality of life tool that asked longitudinal (pre- and post-treatment) questions. Patients served as their own controls. The authors state that the study was not designed to assess the predictors of poor outcomes.

The investigators assessed global satisfaction, body image, relationships, employment status, and sexual function in patients with gender dysphoria who underwent reassignment surgery between 1989 and 1993 and were at least 6 months post-operative. The criteria to qualify for gender surgery were delineated (1985 standards from the Harry Benjamin International Gender Dysphoria Association) and included cross-gender behavior for at least 1 year and heterosexual sexual orientation (based on the post-transition sex). The questionnaire consisted of 10 questions using yes/no answers or Likert-type scales. Findings were descriptive without statistical analysis. As such, changes or differences considered to be biologically significant were not pre-specified, and there were no adjustments for multiple comparisons.

Of the 38 patients who had undergone reassignment surgery, 34 were eligible for the study and 32 participated in the study. 10 (31.2%) female-to-male and 22 (68.8%) male-to-female (ratio 1:2.2). The duration of follow-up was 21.8 ±13.4 months (range 6 months to 4 years). The age was female-to-male 27.8±5.2 (range 23-37) and male-to-female 26.4±7.8 (range 19-47).

Using an investigator-designed quality of life tool, 100% of all patients reported satisfaction with having undergone the surgery. Of the total participants, 65% were satisfied: four (12.5%) (all male-to-female) and eight (25%) (87.5% male-to-female) reported complete dissatisfaction or partial satisfaction with their body, 25% were satisfied to some extent, and 13% were not satisfied at all appearance. Regarding relationships, 80% of female-to-male and 100% of male-to-female patients were dissatisfied with their relationships with others prior to surgery; whereas, no female-to-male patients and 18.1% of male-to-female patients were dissatisfied with relationships after surgery. Regarding sexual partners, 60% of female-to-male and 72.7% of male-to-female patients reported not having a sexual partner prior to surgery; whereas, 20% of female-to-male patients and 27.3% of male-to-female patients did not have a sexual partner after surgery. Of those with partners at each time interval, 100% of female-to-male and 50% of male-to-female patients reported not experiencing orgasm prior to surgery; whereas, 75% of female-to-male and 37.5% of male-to-female patients reported not experiencing orgasm after surgery.
Fifty percent of female-to-male and 54.5% of male-to-female patients reported being either unemployed or not being a student full-time prior to surgery. After surgery, no female-to-male patients and 7 (31.8%) male-to-female patients reported being either unemployed or not being a student full-time. The change was due to student status as 10-Six (60%) of the 32 subjects who were unemployed and not attending school full time prior to surgery returned to full time schooling after. No patients who were employed or full time students before surgery were unemployed after, but 6 of the female-to-male patients and 15 (68.2%) of the male-to-female remained unemployed before and after surgery.

### c. Prospective, non-blinded, observational, cross-sectional studies with controls

Wolfradt and Neumann conducted a controlled, prospective, non-blinded, observational study using a cross-sectional design. The investigators assessed aspects of personality in male-to-female patients who had undergone vocal cord surgery for voice feminization and in healthy non-transgender volunteers from the region. The patients had undergone gender reassignment surgery 1 to 5 years prior to voice surgery. The volunteers were matched by age and occupation. The primary hypothesis was that depersonalization, with the sense of being detached from one’s body or mental processes, would be more common in male-to-female patients with gender dysphoria. German versions of the Scale for Depersonalization Experiences (SDPE), the Body Image Questionnaire (BIQ), a Gender Identity Trait Scale (GIS), and the Self-Esteem Scale (SES) were used in addition to a question regarding global satisfaction. Three of the assessments used a 5 point scale (BIQ, GIS, and SDPE) for questions. One used a 4 point scale (SES). Another used a 7 point scale (global satisfaction). The study consisted of 30 male-to-female patients, 30 healthy female volunteers, and 30 healthy male volunteers. The mean age of study participants was 43 (range 29-67).

Results of the study revealed that there were no differences between the three groups for the mean scores of measures assessing depersonalization, global satisfaction, the integration of masculine traits, and body-image-rejected (subset). Also, the sense of femininity was equivalent for male-to-female patients and female controls and higher than that in male controls. The levels of self-esteem and body image-dynamic (subset) were equivalent for male-to-female patients and male controls and higher than that in female controls, and none of the numeric differences between means exceeded 0.61 units. While this study was limited in that there was no control group of transgender women who had not undergone vocal surgery, it should be noted that the comparison groups were non-clinical samples (i.e. men and women without any referring mental or medical condition). In the studies reviewed by CMS transgender patients even post treatment still have significant medical and mental health morbidity. We note then that this study demonstrated that transgender women postoperatively from vocal surgery were not distinguishable from non-clinical controls by any of the measures assessed.

Beatrice conducted a prospective, non-blinded, observational study using a cross-sectional design and control cohorts in the U.S. The investigator assessed psychological adjustment and functioning (self-acceptance) in male-to-female patients with gender dysphoria (with and without gender reassignment surgery [GRS]), transvestites from two university specialty clinics, and self-identified heterosexual males recruited from the same two universities. The criteria to qualify for the study included being known to the clinic for at least one year, cross-dressing for at least one year without arrest, attendance at 10 or more therapy sessions, emotionally self-supporting, and financially capable of payment for reassignment surgery, and all of these criteria were met by the pre-operative cohort as well as the post-operative cohort. The cohorts were matched to the post-operative cohort (age, educational level, income, ethnicity, and prior heterosexual object choice). The post-operative cohort was selected not on the basis of population representation, but on the basis of demographic feasibility for a small study. The instruments used were the Minnesota Multiphasic Personality Inventory (MMPI) and the Tennessee Self-Concept Scale (TSCS).

Changes or differences considered to be biologically significant were not pre-specified. The scales on the MMPI are not indicative of specific diagnoses, but assess personality structure and function. The three scales on which there were differences between groups were paranoia, schizophrenia, and masculine/feminine (M/F). The scale paranoia measures interpersonal sensitivity, moral self-righteousness and suspiciousness. The schizophrenia scale measures bizarre thoughts, peculiar perceptions, social alienation, poor familial relationships, difficulties in concentration and impulse control, sense of self-worth and self-identity. The scale M/F measures interests in vocations and hobbies, aesthetic preferences, activity-passivity and personal sensitivity and in general how a person conforms to masculine or feminine roles. Moreover, neither the retired MMPI-1 used in this analysis nor the current MMPI-2 have been normed to transgender patient populations. While scoring transgender patients according to their gender identity rather than birth-assigned sex improves the validity of scoring, true norms for these populations are unknown. Unfortunately Beatrice et al. did not use appropriate female scoring for transgender participants in their study.

Of the initial 54 recruits, ten subjects were left in each of the cohorts because of exclusions identified due to demographic factors. The mean age of each cohort were as follows: pre-operative gender dysphoric patients 32.5 (range 27-42) years, postoperative patients 35.1 (30-43) years old, transvestite 32.5 (29-37) years old, and heterosexual male 32.9 (28-38) years old. All were Caucasian. The mean age for cross-dressing in pre-operative patients (6.4 years) and post-operative patients (5.8 years) was significantly lower than for transvestites (11.8 years).

The scores for self-acceptance did not differ by diagnostic category or surgical status as measured by the TSCS instrument. As measured by the T-scored MMPI instrument (50±10), levels of paranoia and schizophrenia were higher for post-operative (GRS) patients (63.0 and 68.8) than transvestites (55.6 and 59.6) and heterosexual males (56.2 and 51.6). Levels of schizophrenia were higher for pre-operative patients (65.1) than heterosexual males (51.6). There were no statistically significant differences between patients with gender dysphoria by surgical status. Scores for the Masculine-Feminine domain were equivalent in those with transvestitism and gender dysphoria with or without surgery, but higher than in heterosexual males - but is should be remembered that. The analysis revealed that despite the high level of socio-economic
functioning in these highly selected subjects, the MMPI is scored based on the gender of the patient and in this case, transgender women. Categories with the highest scores were scored by masculine norms. While the notable for antisocial personality, emotionally unstable personality, paranoid schizophrenia, and paranoia scales were elevated, schizoid personality in the transgender pre-operative GRS patients compared with a non-clinical. By contrast, the same MMPI profiling in heterosexual control group, this is not indicative that these patients had higher levels of schizophrenia or psychosis. The scales measure as described above, traits such as social and family alienation, interpersonal sensitivity, self-worth, males and identity. Contextualizing the study to the 1980s when transvestism was performed in the U.S. it should not be surprising that despite relatively high socioeconomic functioning, transgender patients both preoperatively and post-operatively had mental health concerns in these areas. In addition, as the M/F scale was not scored based on female norms, it would be surprising if transgender participants did not score outside of the normal range. Wearing a dress and having stereotypically feminine interests would not be considered unusual for women, but would for men on the historically rigidly gender binary MMPI-1 notable for the absence of psychological dysfunction.


Kraemer et al. conducted a prospective, non-blinded, observational study using a cross-sectional design comparing pre- and post-surgical cohorts. The investigators assessed body image, and patients were required to meet DSM III or DSM IV criteria as applicable to the time of entry into the clinic. Post-surgical patients were from a long-term study group (Hepp et al., 2002). Pre-surgical patients were recent consecutive referrals. The assessment tool was the Fragebogen zur Beurteilung des eigenen Körpers (FBeK) which contained 3 domains.

There were 23 pre-operative patients: 7 (30.4%) female-to-male and 16 (69.6%) male-to-female (ratio 1:2.3). There were 22 post-operative patients: 8 (36.4%) female-to-male and 14 (63.6%) male-to-female (ratio 1:1.8). The mean ages of the cohorts were as follows: pre-operative 33.0±11.3 years; post-operative 38.2±9.0 years. The mean duration after reassignment surgery was 51±25 months (range 5-96).

The pre-operative groups had statistically higher insecurity scores compared to normative data for the natal sex: female-to-male 9.0±3.8 versus 5.1±3.7; male-to-female 8.1±4.5 versus 4.7±3.1 as well as statistically lower self-confidence in one’s attractiveness: female-to-male 3.1±2.9 versus 8.9±3.1; male-to-female 7.0±2.9 vs 9.5±2.6. Scores for insecurity and self-confidence in the post-operative cohort were not inferior to the normative values. Insecurity decreased statistically from 9.0±3.8 in the female-to-male pre-operative cohort and from 8.1±4.5 in the male-to-female pre-operative cohort to 4.4±3.8 and 3.4±2.3 in the respective post-operative cohorts. Self-confidence increased statistically from 3.1±2.9 in the female-to-male pre-operative cohort and 7.0±2.9 male-to-female pre-operative cohort to 9.29±1.98 and 10.29±2.0 in the respective post-operative cohorts.

d. Prospective, non-blinded, observational, cross-sectional studies with semi-matched controls

Kuhn et al. conducted a prospective, non-blinded, observational study using a cross-sectional design and semi-matched control cohort. The investigators assessed global satisfaction in patients who were from gynecology and endocrinology clinics, and who had undergone some aspect of gender reassignment surgery in the distant past, but were still receiving cross-sex hormones from the clinic. The quality-of-life assessment tools included a VAS and the King’s Health Questionnaire (KHQ) with its eight domains including one for incontinence. The KHQ questionnaire and the numerical change/difference required for clinical significance (≥5 points in a given domain, with higher scores being more pathologic) were included in the publication.

Twenty healthy female controls from the medical staff who had previously undergone an abdominal or pelvic surgery were partially matched by age and body mass index (BMI), but not sex.

Of the 55 participants, three (5.4%) were female-to-male and 52 (94.5%) were male-to-female (ratio 1:17.3). Reassignment surgery had been conducted 8 to 23 years earlier (median 15 years). The median age of the patients at the time of the study was 51 years (range 39-62 years). The patients had undergone a median of 9 surgical procedures in comparison to the 2 undergone by controls. Patients were less likely to be married (23.6% versus 65%; p=0.002), and partnership status was unknown in 5 patients. The scores of VAS global satisfaction (maximal score 8) were lower for surgically reassigned patients (4.49±0.1 SEM) than controls (7.35±0.26 SEM) (p<0.0001).

There were statistically and biologically significant differences for 4 of the 8 domains between the patients and controls: physical limitation: 37.6±2.3 versus 20.9±1.9 (p<0.0001), personal limitation: 20.9±1.9 versus 11.6±0.4 (p<0.001), role limitation: 27.8±2.4 versus 34.6±1.7 (p<0.055), and general health: 31.7±2.2 versus 41.0±2.3 (p<0.02). Information as to whether a high or low score was positive for the various domains was not provided. Wording from the abstract suggests that these 4 differences all reflected lower quality-of-life.

e. Prospective, blinded, observational, cross-sectional studies with no concurrent controls


Hess et al. conducted a prospective, blinded observational study using a cross-sectional design and a self-designed questionnaire. The investigators assessed post-operative satisfaction in male-to-female patients with gender dysphoria who were followed in a urology specialty clinic. Patients had met the ICD-10 diagnostic criteria, undergone gender reassignment surgeries including penile inversion vaginoplasty, and a Likert-style questionnaire survey with 11 elements. Descriptive statistics were provided.

There were 254 consecutive eligible patients who had undergone surgery between 2004 and 2010 identified and sent surveys, of whom 119 (46.9%) responded anonymously. Of the
participants, 13 (10.9%) reported dissatisfaction with outward appearance and 16 (13.4%) did not respond; three (2.5%) reported dissatisfaction with surgical aesthetics and 25 (21.0%) did not respond; eight (6.7%) reported dissatisfaction with functional outcomes of the surgery and 26 (21.8%) did not respond; 16 (13.4%) reported they could not achieve orgasm and 28 (23.5%) did not respond; four (3.4%) reported feeling completely male/more male than female and 28 (23.5%) did not respond; six (5.0%) reported not feeling accepted as a woman, two (1.7%) did not understand the question, and 17 (14.3%) did not respond; and 16 (13.4%) reported that life was harder and 24 (20.2%) did not respond.


Lawrence conducted a prospective, blinded observational study using a cross-sectional design and a partially self-designed quality of life tool using yes/no questions or Likert scales. The investigator assessed sexual function, urinary function, and other pre/post-operative complications in patients who underwent male-to-female gender reassignment surgery. Questions addressed core reassignment surgery (neo-vagina and sensate neo-clitoris) and related reassignment surgery (labiaplasty, urethral meatus revision, vaginal deepening/widening, and other procedures), use of electrolysis, and use of hormones.

Questionnaires were designed to be completed anonymously and mailed to 727 eligible patients. Of those eligible, 232 (32%) returned valid questionnaires. The age at the time reassignment surgery was 44±9 (range 18-70) years and mean duration after surgery was 3±1 (range 1-7) years.

Happiness with sexual function and the reassignment surgery was reported to be lower when permanent vaginal stenosis, clitoral necrosis, pain in the vagina or genitals, or other complications such as infection, bleeding, poor healing, other tissue loss, other tissue necrosis, urinary incontinence, and genital numbness were present. Quality-of-life (QOL) was impaired when pain in the vagina or genitals was present.

Satisfaction with sexual function, gender reassignment surgery, and overall QOL was lower when genital sensation was impaired and when vaginal architecture and lubrication were perceived to be unsatisfactory. Intermittent regret regarding reassignment surgery was associated with vaginal hair and clitoral pain. Vaginal stenosis was associated with surgeries performed longer ago; whereas, more satisfaction with vaginal depth and width was present in more recent surgeries. Given the finding that vaginal hair is associated with intermittent regret, preoperative hair removal through electrolysis or laser therapy may be necessary to diminish intermittent regret rates.

f. Prospective, non-blinded, observational, longitudinal and patients served as their own controls
Heylens et al. conducted a prospective, non-blinded observational study using a longitudinal design in which patients served as their own controls. They used a non-specific psychiatric test with normative data along with two self-designed questionnaires. The investigators assessed psychosocial adjustment and psychopathology in patients with gender identity disorders. Patients were to be sequentially evaluated prior to institution of hormonal therapy, then 3 to 6 months after the start of cross-sex hormone treatment, and then again one to 12 months after reassignment surgery. The Dutch version of the SCL-90R with 8 subscales (agoraphobia, anxiety, depression, hostility, interpersonal sensitivity, paranoid ideation/psychoticism, and sleeping problems) and a global score (psycho-neuroticism) was used serially. A seven parameter questionnaire was used serially to assess changes in social function. Another cross-sectional survey assessed emotional state. The cohorts at each time point consisted of patients who were in the treatment cohort at the time and who had submitted survey responses.

Ninety of the patients who applied for reassignment surgery between June 2005 and March 2009 were recruited. Fifty seven entered the study. Forty six (51.1% of the recruited population) underwent reassignment surgery. Baseline questionnaire information was missing for 3 patients. Baseline SCL-90 scores were missing for 1 patient but included SCL-90 scores from some of the 11 recruits who had not yet undergone reassignment surgery. Time point 2 (after hormone therapy) SCL-90 information was missing for 10, but included SCL-90 scores from some of the 11 recruits who had not yet undergone reassignment surgery. At time point 3, 42 (91.3% of those who underwent reassignment surgery) patients completed some part of the SCL-90 survey and the psychosocial questionnaires. Some questionnaires were incomplete. The investigators reported response rates of 73.7% for the psychosocial questionnaires and 82.5% rates the SCL-90.

Of those who responded at follow-up after surgery, 88.1% reported having good friends; 52.4% reported the absence of a relationship; 47.6% had no sexual contacts; 42.9% lived alone; 40.5% were unemployed, retired, students, or otherwise not working; 2.4% reported alcohol abuse; and 9.3% had attempted suicide. The frequency of these parameters reportedly did not change statistically during the study interval, but there was no adjustment for the inclusion of patients who did not undergo surgery.

In a cross-sectional, self-report mood survey, of the 42 study entrants who completed the entire treatment regimen including reassignment surgery and the final assessment (refers to the initial 57) reported improved body-related experience (97.6%), happiness (92.9%), mood (95.2%), and self-confidence (78.6%) and reduced anxiety (81.0%). Of participants, 16.7% reported thoughts of suicide. Patients also reported on the intervention phase that they believed was most helpful: hormone initiation (57.9%), reassignment surgery (31.6%), and diagnostic-psychotherapy phase (10.5%).

The global “psycho-neuroticism” SCL-90R score, along with scores of 7 of the 8 subscales, at baseline were statistically more pathologic than the general population. After hormone therapy,
the score for global “psycho-neuroticism” normalized and remained normal after reassignment surgery. More specifically the range for the global score is 90 to 450 with higher scores being more pathologic. The score for the general population was 118.3±32.4. The respective scores for the various gender dysphoric cohorts were 157.7±49.8 at initial presentation, 119.7±32.1 after hormone therapy, and 127.9±37.2 after surgery. The scores for the general population and the scores after either hormone treatment or surgical treatment did not differ.

The authors noted that this population was not representative of the entire population of patients with gender dysphoria and may be thus less applicable to U.S. populations. Specifically they noted that once patients were accepted for treatment and began hormone replacement therapy, they would be assured of getting appropriate surgical treatment which is, along with hormonal and mental health care reimbursed in Belgium as it is not universally in the U.S. The authors note that patients were assured of surgical care and that “[t]his perspective might certainly have an influence on the level of psychoneurotic distress. If there had been less certainty, at the end of the diagnostic phase and after initiation of hormonal treatment, about results could have been different.”


Smith et al. conducted a prospective, non-blinded, observational study using a longitudinal design and psychological function tools. Patients served as their own control prior to and after reassignment surgery. The investigator assessed gender dysphoria, body dissatisfaction, physical appearance, psychopathology, personality traits, and post-operative function in patients with gender dysphoria. All subjects underwent standard clinic protocol including mental health assessment, hormone replacement therapy, real life experience, and Patients underwent some aspect of reassignment surgery. The test instruments included the Utrecht Gender Dysphoria Scale (12 items), the Body Image Scale adapted for a Dutch population (30 items), Appraisal of Appearance Inventory (3 observers, 14 items), the Dutch Short MMPI (83 items), the Dutch version of the Symptom Checklist (SCL)(90 items), and clinic-developed or modified questionnaires. The Utrecht Gender Dysphoria Scale while developed by this research group has undergone subsequent validation as a measure of gender dysphoria. Pre-treatment data was obtained shortly after the initial interview. Post-surgery data were acquired at least 1 year post reassignment surgery.

Altogether 325 consecutive adult and adolescent patients presented to the program for diagnosis and treatment. Of these 222 were eligible and started hormone treatment, but only 188 continued treatment which were the subjects of the study. Of these 188, 136 agreed to partial participation in the study and 158 agree to complete participation (the two parts being a questionnaire and an interview). The size of the pool of available patients was not identified. Overall 325 consecutive adolescents and adults initially were “involved.” Of these, 103 (29 [28.2%] female to male patients and 74 [71.8%] male to female patients [ratio 1:2.6]) never started hormone therapy, 222 (76 [34.2%] female-to-male patients and 146 [65.8%] male-to-female patients [ratio 1:1.9]) initiated hormone therapy. Of the patients who started hormone therapy, 31 (5 [14.7%] female-
to male patients and 29 (85.3%) male to female patients [ratio 1:5.8]) discontinued hormone therapy. After discontinuation of hormone therapy, the study was limited to adults. Of adults, 162 (58 [35.8%] female-to-male and 104 [64.2%] male-to-female [ratio 1:1.8]) were eligible and provided pre-surgical test data, and 126 (77.8% of eligible adults) (49 [38.9%] female-to-male and 77 [61.1%] male-to-female [ratio 1:1.6]) provided post-surgical data. For those patients who completed reassignment, the mean age at the time of surgical request was 30.9 years (range 17.7-68.1) and 35.2 years (range 21.3-71.9) years at the time of follow-up. The intervals between hormone treatment initiation and surgery and surgery and follow-up were 20.4 months (range 12 to 73) and 21.3 months (range 12 to 47) respectively.

Of the 126 adults who provided post-surgical data, 50 (40.0%) reported having a steady sexual partner, three (2.3%) were retired, and 58 (46.0%) were unemployed. Regarding regret, six patients expressed some regret regarding surgery, but did not want to resume their natal gender role, and one male-to-female had significant regret and would not make the same decision.

Gender dysphoria as measured by the subsequently validated UGDS was significantly diminished postoperatively (14.8±3.0) when compared with preoperative values (54.3±7.1). Post-surgery Utrecht dysphoria scores dropped substantially and approached reportedly normal values. The patients’ appearance better matched their new gender. No one was dissatisfied with his/her overall appearance at follow-up and 98 (91.6%) were very satisfied with their appearance. Satisfaction with primary sexual, secondary sexual, and non-sexual body traits improved over time. Male-to-female patients, however, were more dissatisfied with the appearance of primary sex traits than female-to-male patients. Eleven (28.9%) were completely satisfied with their mastectomy, 22 (57.9%) were not completely satisfied, and 5 (12.2%) were dissatisfied due to the visibility of the scars. Regarding mastectomy, 27 of 38 (71.1%) female-to-male respondents (not including 11 non-respondents) reported incomplete satisfaction with their mastectomy procedure. For five of these patients, the incomplete satisfaction was because of scarring. Regarding vaginoplasty, 20 of 67 (29.8%) male-to-female respondents (not including 10 non-respondents) reported incomplete satisfaction with their vaginoplasty.

Psychological functioning was measured by the Dutch Short MMPI and the Dutch version of the Symptom Check List which together had 14 subscales such as depression, anxiety, hostility, shyness, somatization and sleeping problems. While most of the pre-treatment means for these subscales were in the normal range, there was a statistically significant decrease in all 14 scales save for hostility which was diminished but did not reach statistical significance (p=0.147).

Most of the MMPI scales were already in the normal range at the time of initial testing. SCL global scores for psycho-neuroticism were minimally elevated before surgery 143.0±40.7 (scoring range 90 to 450) and normalized after surgery 120.3±31.4. (An analysis using patient level data for only the completers was not conducted.)

Megeri D, Khoosal D. Anxiety and depression in males experiencing gender dysphoria. Sexual and Relationship Therapy. 2007 Feb; 22(1):77-81. (Not in PubMed) and Udeze B, Abdelmawla
Udeze et al. conducted a prospective, non-blinded, longitudinal study assessing a randomized subset of patients who had completed a non-specific psychological function tool prior to and after male-to-female reassignment surgery. Patients served as their own controls. The investigators used the WPATH criteria for patient selection. Psychiatric evaluations were routine. All patients selected for treatment were routinely asked to complete the self-administered SCL-90R voluntarily on admission to the program and post-operatively. A post-operative evaluations (psychiatric and SCL-90R assessment) were conducted within 6 months to minimize previously determined loss rates. The patient pool was domestic and international. There were 546 gender dysphoric patients from all over the United Kingdom and abroad, of whom 318 (58.2%) progressed to surgery. Of these, 127 were from the local Leicester area in the United Kingdom and 38 (29.9%) progressed to surgery. The mean age for the selected male-to-female patients at the time of study was 47.33±13.26 years (range 25 to 80) and reflected an average wait time for surgery of 14 months (range 2 months to 6 years). For this investigation, 40 male-to-female subjects were prospectively selected.

The raw SCL-90 global scores for psychosocial adjustment were unchanged over time: 48.33 prior to surgery and 49.15 after surgery. If the scale was consistent with T-scoring, the results were non-pathologic. A statistical trend in the anger/hostility subscale was reported. No psychiatric disorders were otherwise identified prior to or after surgery.

Investigators from the same clinical group (Megeri, Khoosal, 2007) conducted additional testing to specifically address anxiety and depression with the Beck Depression Inventory, General Health Questionnaire (with 4 subscales), HADS, and Spielberger State and Trait Anxiety Questionnaire (STAI-X1 and STA-X2). The test population and study design appear to be the same. No absolute data were presented, only changes in scores were presented. There were no statistically significant changes.


Kockott and Fahrner conducted a prospective, observational study using a longitudinal design. Treatment cohorts were used as controls, and patients served as their own controls. The investigators assessed psychosocial adjustment in patients with gender identity issues. Patients were to have met DSM III criteria. Trans-sexuality, transvestitism, and homosexuality were differentiated. The criteria required for patients to receive hormone therapy and/or reassignment surgery were not delineated. After receiving hormone therapy, patients were later classified by surgical reassignment status (pre-operative and post-operative) and desire for surgery (unchanged desire, hesitant, and no longer desired).

The first investigative tool was a semi-structured in-person interview consisting of 125 questions. The second investigative tool was a scale that organized the clinical material into nine domains which were then scored on a scale. The Psychological Integration of Transsexuals (PIT) instrument was not otherwise described in the publication or in other citations. There were
15 interviews and two separate interviewers. There were 80 patients identified, but 58 (72.5%) patients (26 pre-operative; 32 post-operative) were ultimately included in the analysis. The duration of follow-up was longer for post-operative patients (6.5 years) than for pre-operative patients (4.6 years) (including time for one patient subsequently excluded). The mean age of the post-operative patients was 35.5±13.1 years, and the age of the patients who maintained a continued desire for surgery was 31.7±10.2 years. The age of the patients who hesitated about surgery was somewhat older, 40.3±9.4 years. The age of the patients who were no longer interested in surgery was 31.8±6.5 years. All were employed or in school at baseline. Patients with hesitation were financially better-off, had longer-standing relationships even if unhappy, and had a statistical tendency to place less value on sex than those with an unchanged wish for surgery.

Post-operative patients more frequently reported contentment with the desired gender and the success of adaption to the gender role than the pre-operative patients with a persistent desire for surgery. Post-operative patients more frequently reported sexual satisfaction than pre-operative patients with a continuing desire for surgery. Post-operative patients also more frequently reported financial sufficiency and employment than pre-operative patients with a persistent desire for surgery. Suicide attempts were stated to be statistically less frequent in the post-surgical cohort.

Psychosocial adjustment scores were in the low end of the range with “distinct difficulties” (19-27) at the initial evaluation for the post-operative patients (19.7), the pre-operative patients with a persistent wish for surgery (20.2), and the hesitant patients (19.7). At initial evaluation, psychosocial adjustment scores for patients no longer wanting surgery were at the high end of the range with “few difficulties” (10-18). At the final evaluation, Psychosocial adjustment scores were at the high end of the range “few difficulties” (10-18) for the post-operative patients (13.2) and the patients no longer wanting surgery (16.5). Psychosocial adjustment scores at the final evaluation were in the borderline range between “few difficulties” (10-18) and “distinct difficulties” (19-27) for both the pre-operative patients with a persistent desire for surgery (18.7), and the hesitant patients (19.1).

The changes in the initial score and the final follow-up score within each group were tracked and reported to be statistically significant for the post-operative group, but not for the other groups. Statistical differences between groups were not presented. Moreover, the post-operative patients had an additional test immediately prior to surgery. The first baseline score (19.7) would have characterized the patients as having “distinct difficulties” in psychosocial adjustment while the second baseline score (16.7) would have categorized the patients as having “few difficulties” in psychosocial adjustment despite the absence of any intervention except the prospect of having imminent reassignment surgery. No statistics reporting on the change between scores of the initial test and the test immediately prior to surgery and the change between scores of the test immediately prior to surgery and the final follow-up were provided.

g. Prospective, non-blinded, observational, longitudinal study with retrospective baseline data
Meyer JK, Reter DJ. Sex reassignment. Follow-up. Arch Gen Psychiatry. 1979 Aug;36(9):1010-5. (United States study)

Meyer and Reter conducted a prospective, non-blinded, observational study using a longitudinal design and retrospective baseline data. Interview data were scored with a self-designed tool. There were treatment control cohorts, and patients served as their own controls. The investigators assessed patients with gender dysphoria. The 1971 criteria for surgery required documented cross-sex hormone use as well as living and working in the desired gender for at least 1 year in patients subsequently applying for surgery. Clinical data including initial interviews were used for baseline data. In follow-up, the investigators used extensive 2 to 4 hour interviews to collect information on (a) objective criteria of adaptation, (b) familial relationships and coping with life milestones, and (c) sexual activities and fantasies. The objective criteria, which were the subject of the publication, included employment status (Hollingshead job level), cohabitation patterns, and need for psychiatric intervention. The investigators designed a scoring mechanism for these criteria and used it to determine a global adjustment score. In addition to being a non-validated score, there were substantive problems with this scale that call the validity of the study into question. In specific homosexual transgender people are inappropriately penalized for being in a relationship. For example a lesbian transgender woman in a relationship with a female partner would have lost 2 to 4 points simply by remaining in a long term relationship through and after sex reassignment surgery. In contrast, a heterosexual person who had no partner at the beginning or end of the study would not be penalized.

The clinic opened with 100 patients, but in follow-up, 50 of the 100 patients were excluded, 50 were interviewed and 48 of the interviewees gave consent for publication. Of the 50 who were not excluded, 15 (4 female-to-male, 11 male-to-female; ratio 1:2.8) were part of the initial operative cohort, 14 (1 female-to-male; 13 male-to-female; ratio 1:13) later underwent reassignment surgery at the institution or elsewhere, and 21 (5 female-to-male; 16 male-to-female; ratio 1:3.2) did not undergo surgery. The mean ages of these cohorts were 30.1, 30.9, and 26.7 years respectively. The mean follow-up time was 62 months (range 19-142) for those who underwent surgery and 25 months (range 15-48) for those who did not. Socioeconomic status was lowest in those who subsequently underwent reassignment surgery.

Of patients initially receiving surgery, 8% had some type of later psychiatric contact, which was approximately 3.5 times higher in those who had not undergone surgery or who had done so later. There was a single female-to-male patient with multiple surgical complications who sought partial reassignment surgery reversal.

The adjustment scores improved over time with borderline statistical significance for the initial operative group and with statistical significance for the never operated group. However as has been noted in a critique of this study, the operated group was followed for 62 months, residual unoperated 27, and subsequently operated for 21. Given that many of these events in the author’s self-designed score (e.g. psychiatric hospitalization or being arrested) were events that accumulate over time, the fact that Meyer did not correct for this follow-up time difference would inevitably cause any group followed for a significantly longer period of time to have a higher incidence (and thus lower score) even if the rate of event per period of time were the same in both groups. It should be noted that Meyer counted each negative event towards the score, so
someone arrested and jailed twice would have a -4. This was not directly stated in the article, but can be inferred from one patient having a baseline score of -18, a greater score than possible if events were counted only once. So if in all three groups, the average number of arrests per 6 month period was 1, the operated group would have an average score of -10, residual unoperated -5, and subsequently operated -4. In addition, the authors did not indicate if arrests were appropriate or inappropriate (e.g. a transgender woman being arrested post-operatively for appropriately using the women's restroom). The fact that the operated group had a lower score (albeit one that did not reach statistical significance) despite having a follow up time of double or triple the other two groups with resultant increased chance to accumulate negative points for accessing mental health care or legal complications suggests not only was the study underpowered, but also negatively biased against the operated group.

Both the absolute score value at follow-up and the magnitude of change were the same. By contrast, the adjustment scores did not improve for those who were not in the cohort initially approved for surgery, but who subsequently underwent surgery later. This was particularly true if the surgery was performed elsewhere.

h. Prospective, non-blinded, observational, semi-cross sectional with no controls


Johansson et al. conducted non-blinded, prospective longitudinal cohort observational study using a semi cross-sectional design (albeit over an extended time interval) using a self-designed tool and Axis V assessment. The study was prospective except for the acquisition of baseline Axis V data. There were no formal controls in this mixed population with and without surgery.

The investigators assessed satisfaction with the reassignment process, employment, partnership, sexual function, mental health, and global satisfaction in gender-reassigned persons from two clinics in the north and south of Sweden disparate geographic regions. No other information regarding the sites of care was provided. Surgical candidates were required to have met National Board of Health and Welfare criteria including initial and periodic psychiatric assessment, ≥1 year of real-life experience in preferred gender, and ≥1 year of subsequent hormone treatment. In addition, participants were required to have been approved for reassignment 5 or more years prior and/or to have completed surgical reassignment (e.g., sterilization, genital surgery) 2 or more years prior. The investigators employed semi-structured interviews covering a self-designed list of 55 pre-formulated questions with a 3 or 5 point ordinal scale. Clinician assessment of Global Assessment of Functioning (GAF; Axis V) was also conducted and compared to initial finding during the study. An increase in GAF points of five or more was selected to be considered improved and a decrease of five points was considered as worse. In addition, the authors combined GAF with 4 other measures to determined a “Global Outcome” score which included: SES, work/study, relationships, psychiatric care. “Global improvement” was pre-defined to be improved in at least 2 areas of the “Global Outcome” and worsened in no

areas. In addition to the “Global Outcome” score they determined patients’ self-assessment as to global outcome. There was no stratification by specific types of changes or differences considered to be biologically significant were not pre-specified. Diagnostic cut points were not provided. Statistical corrections for multiple comparisons were not included. There was no stratification by treatment.

Of the pool of 60 eligible patients, there were 21 (35.0%) female-to-male and 39 (65%) male-to-female (ratio 1:1.9); 32 (53.3% of eligible) (14 [43.8%] female-to-male; 18 [56.2%] male-to-female [ratio 1:1.3]) had completed genital gender reassignment surgery (not including 1 post mastectomy). 5 were still in the process of completing surgery, and 5 (1 female-to-male; 4 male-to-female; ratio 1:4) had discontinued the surgical process prior to castration and genital surgery.

The ages of the patients (ranges) at entry into the program, reassignment surgery, and follow-up were 27.8 (18-46), 31.4 (22-49), and 38.9 (28-53) in the female-to-male group respectively and 37.3 (21-60), 38.2 (22-57), and 46.0 (25.0-69.0) in the male-to-female group respectively. The differences in age by cohort group were statistically significant. Of participants, 88.2% of all enrolled female-to-male versus 44.0% of all enrolled female-to-male patients had cross-gender identification in childhood (versus during or after puberty) (p<0.01).

Although 95.2% of all enrolled patients self-reported improvement in their global function GAF, in contrast, clinicians determined GAF-improvement in the “Global Function” score in 61.9% of patients. Clinicians observed improvement in the “Global Function” score in 47% of female-to-male patients and 72% of male-to-female patients. A ≥5 point improvement in the GAF score was present in 18 (42.9%). Of note, three of the five patients who were in the process of reassignment and five of the five who had discontinued the process were rated in the “Global Function” score by clinicians as having improved.

Of all enrolled 95.2% (with and without surgery) reported satisfaction with the reassignment process. Of these patients, 33 (79%) identified themselves by their preferred gender and nine (21%) identified themselves as transsexuals. None of these nine (eight male-to-female) had completed reassignment surgery. This was either because they were not yet that far in the process or because of ambivalence secondary to lack of acceptance by others and dissatisfaction with their appearance. Of the patients who underwent genital surgery (n=32) and mastectomy only (n=1), 22 (66.7%) were satisfied while four (three female-to-male) were dissatisfied with the surgical treatment.

Regarding relationships after surgery, 16 (38%) (41% of female-to-male; 36% of male-to-female) were reported to have a partner. Yet more than that number commented on partner relationships: 62.2% of the 37 who answered (50.0% of female-to-male; 69.6% of male-to-female) reported improved partner relationships (5 [11.9%] declined to answer); 70.0% of the 40 who answered (75.0% of female-to-male; 66.7% of male-to-female) reported an improved sex life. Investigators observed that reported post-operative satisfaction with sex life was statistically more likely in those with early rather than late cross-gender identification.

In addition 55.4% self-reported improved general health; 16.1% reported impaired general health; 11.9% were currently being treated with anti-depressants or tranquilizers; 44.7% of the
who answered (53.3% of female-to-male; 39.1% of male-to-female) reported improved work circumstances (4 [9.5%] declined to answer.); 61.9% were students or employed. The remainder (38.1%) were living on disability pensions (28.6%), unemployed (4.8%), or retired (4.8%).

i. **Prospective, cross sectional, observational, internet self-report survey, with unknown blinding, no formal controls**


Newfield et al. conducted a prospective, observational internet self-report survey of unknown blinding status using a cross-sectional study of transgender males who had or had not undergone hormone therapy design and mastectomy which used also used U.S. population health norms (obtained from the 1998 National Survey of Functional Health Status) using the SF36v2 as a measure of general health care non-specific quality of life tool in a mixed, population with and without reassignment surgery. There were no formal controls.

The investigators recruited natal female to male transgender participants identifying as male using email, internet bulletin boards, and flyers/postcards distributed in the San Francisco Bay Area. Reduction of duplicate entries by the same participant was limited to the use of a unique user name and password.

The investigators employed the SF-36 Version 2 using U.S. normative data. They reported using both male and female normative data for the comparator SF-36 cohort. Data for the 8 domains were expressed as normative scoring. The Bonferroni correction was used to adjust for the risk of a Type 1 error with analyses using multiple comparisons.

A total of 379 U.S. respondents classified themselves as males or females to males with or without therapeutic intervention. The mean age of the respondents who classified themselves as male or female-to-male was 32.6±10.8 years. 89% were Caucasian, 3.6% Latino, 1.8% African American, 1.8% Asian, and 3.8% other. 254 (67.0%) reported any testosterone use in the past or currently; and 242 (63.8%) reported current testosterone use. In addition, 136 (36.7%) reported having had “top” surgery and 11 (2.9%) reported having “bottom” surgery. The Physical Summary Score (53.45±9.42) was statistically higher (better) than the natal gender unspecified SF-36 normative score (50±10) (p=<0.001), but was within 1 standard deviation of the normative mean. The Mental Summary Score (39.63±12.2) was statistically lower (worse) than the natal gender unspecified SF-36 normative score (50±10) (p<0.001), but was well within 2 standard deviations of the normative mean. Subcomponents of this score: Mental Health (42.12±10.2), Role Emotional (42.42±11.6), Social Functioning (43.14±10.9), and Vitality (46.22±9.9) were statistically lower (worse) than the SF-36 normative sub-scores, but well within 1 standard deviation of the normative sub-score means. Interpretive information for these small biologic differences in a proprietary assessment tool was not provided.

Additional intragroup analyses were conducted, although the data were not stratified by type of therapeutic intervention (hormonal, as well as, surgical). Outcomes of hormone therapy were considered separately and dichotomously from reassignment surgery. The Mental Summary
Score was statistically higher (better) in those who had “Ever Received Testosterone” (41.22±11.9) than those with “No Testosterone Usage” (36.08±12.6) (p=0.001). There were statistically significant differences (p<0.01) in the Vitality, Social Functioning, Role Emotional, and Mental Health sub-scores. Participants who had mastectomy had higher QOL scores than those who had not received surgery, with statistically significant findings (p<0.01) for the General Health, Social Functioning, and all three mental health concepts, and the total. The Mental Summary Scores showed a trend towards statistical difference between those who “Ever Received Top Surgery” (41.21±11.6) and those without “Top Surgery” (38.01±12.5) (p=0.067). These differences were well within 1 standard deviation of the normative mean. Interpretive information for these small biologic differences in a proprietary assessment tool was not provided.

j. Partially prospective, non-blinded, observational studies with longitudinal designs and patients served as their own controls


Ruppin and Pfafflin conducted a partially prospective, non-blinded, observational study using a longitudinal design and non-specific psychometric tests and a self-designed interview tool and questionnaire. Patients served as their own controls.

The investigators assessed psychological symptoms, interpersonal difficulties, gender role stereotypes, personality characteristics, societal function, sexual function, and satisfaction with new gender role in patients with gender dysphoria. Patients were required to have met the ICD-10 criteria for trans-sexualism, been seen by the clinic by prior to 2001, and completed an official change in gender including name change prior to 2001. Assessment tools included German versions of standardized surveys with normative data: the SCL 90R, the Inventory of Interpersonal Problems (IIP), Bem Sex Role Inventory (BSRI), and the Freiburg Personality Inventory (FPI-R), along with semi-structured interviews with self-designed questionnaires. The prospective survey results were compared to retrospective survey results.

Overall, 140 patients received recruitment letters then 71 (50.7%) agreed to participate. Of these participants, 36 (50.7%) were female-to-male; 35 (49.3%) were male-to-female (ratio 1:0.97). The ages of the patients were: 41.2±5.78 years female-to-male and 52.9±10.82 years male-to-female 52.9±10.82 years. The intervals for follow-up were 14.1±1.97 years and 13.7±2.17 years respectively.

All female-to-male patients had undergone mastectomy; 91.7% had undergone oophorectomy and/or hysterectomy; 61.1% had undergone radial forearm flap phalloplasty or metaiodiplasty; 94.3% of male-to-female patients had undergone vaginoplasty and perhaps an additional procedure (breast amplification, larynx surgery, or vocal cord surgery). Two male-to-female patients had not undergone any reassignment surgery, but were still included in the analyses.

A total of 68 patients ranked their well-being as 4.35±0.86 out of five (three patients did not respond to this question). Of respondents, 40% reported not in a steady relationship. Regular
sexual relationships were reported by 57.1% of 35/36 female-to-male respondents and 39.4% of 33/35 of male-to-female respondents (three patients did not respond to this question). A total of 11 patients reported receiving out-patient psychotherapy. No patients expressed a desire for gender role reversal (although two of the 71 did not respond to this question). The response rate was less than 100% for most of the self-designed survey questions. Of participants, 78.6% were employed full- or part time or were self-employed, 14.3% received a pension, and 7.1% were unemployed. Only 2.8% (2) participants reported problems with alcohol or illegal drugs.

Changes from the initial visit to the follow-up visit were assessed using three psychometric instruments: SCL-90, IIP, and FPI-R. For the SCL-90 in 62 of 71 patients Participants scores improved at follow-up on all scales and except for the scale Somatization, all were statistically significant. The effect sizes were small for the scales Phobic anxiety and Paranoid ideation, medium for Obsessive–compulsive, Depression, Anxiety, Hostility, and Psychoticism, and large for Interpersonal sensitivity. Changes from the initial visit to the follow-up visit were assessed for the IIP in 55 of 71 patients. Participants’ scores improved (were lower) at follow-up, and was statistically significance on all scales although effect sizes were small for the scale Domineering/controlling, medium for Vindictive/self-centered, Cold/distant, Socially inhibited, Nonassertive, Self-sacrificing, and Intrusive/needy, and large only for Overly accommodating. Changes from the initial visit to the follow-up visit were assessed for the FPI-R in 58 of 71 patients. This comparison showed a significant increase in Life satisfaction with a large effect size. The decreases on the scales Irritability, Openness, and Emotionality were also statistically significant but with a medium effect size. The effect size was large only for the “Life Satisfaction” scale. Changes from the initial visit to the follow-up visit were assessed for the BSRI in 16 of 36 female-to-male patients and 19 of 35 male-to-female patients. The “Social Desirability” score increased for the female-to-male respondents. At endpoint, both categories of respondents reported androgynous self-images.

k. Partially prospective, non-blinded, observational studies with cross-sectional designs that had control groups but were not concurrent


Haraldsen and Dahl conducted a partially prospective, non-blinded, observational study using a cross-sectional design and a non-specific psychometric test. There was a control group, but it was not concurrent.

In the germane sub-study, the investigator assessed psychopathology in patients with gender dysphoria. Patients, who were independently evaluated by 2 senior psychiatrists, were required to meet DSM III-R or DSM IV diagnostic criteria and the Swedish criteria for reassignment surgery. The Norwegian version of the SCL-90 was used. The testing was conducted from 1987 to 1989 for those who had undergone reassignment surgery between 1963 and 1987 and from 1996 to 1998 for pre-surgical patients who had applied for reassignment surgery between 1996 and 1998. In addition, Axis I, Axis II, and Axis V (Global Functioning) was assessed.
Of 65 post–surgical and 34 pre-surgical patients, 59 post-surgical and 27 pre-surgical patients ultimately entered the study. The combined cohorts consisted of 35 (40.7%) female-to-male patients and 51 (59.3%) male-to-female patients (ratio 1:1.5). The ages were female-to-male 34±9.5 years and female-to-male 33.3±10.0 years. The other control group consisted of patients with personality disorder. 101 (27 men (33.9±7.3 years) and 74 women (31.6±8.2) were tested during a treatment program. One year later, 98% were evaluated.

A total of 28 (32.5%) of the pre- and post- reassignment surgery patients had an Axis I diagnosis compared to 100 (99.0%) of those with personality disorders. Depression and anxiety were the most common diagnoses in both groups, but were approximately three to four times more common in the personality disorder cohort. Seventeen (19.8%) of the pre- and post- reassignment surgery patients had an Axis II diagnosis whereas the mean number of personality disorders in the personality disorder cohort was 1.7±1. The Global Assessment of Function was higher (better) in the gender dysphoric groups 78.0±8.9 than in the personality disorder cohort (53.0±9.0).

Global Severity Indices (GSI) were highest for those with personality disorder regardless of gender and exceeded the cut-point score of 1.0. The GSI scores for females-to-males and males-to-females were 0.67±0.57 and 0.56±0.45. Although they were nominally higher than the healthy normative controls (males: 0.32±0.36 and females 0.41±0.43), they were well within the non-pathologic range. The same was true for the subscales.

SCL-90 GSI scores did not differ substantively between pre- and post-surgical patients, nor did the SCI subscale scores differ substantially between pre- and post-surgical patients. Any small non-significant differences tracked with the age and sex differences.

I. Partially prospective, non-blinded, observational studies with cross-sectional designs that had no control groups


Leinung et al. conducted a partially prospective, non-blinded, observational study using a cross-sectional design and descriptive statistics. There were no formal controls. This study in the U.S. is not specifically useful to determine the usefulness of GRS as no comparisons were made between pre and post treatment patients. However this is a representative description of typical care in the U.S. in a population where the majority of patients have insurance exclusions specifically prohibiting payment for any care related to gender dysphoria (including hormonal and surgical treatments). While it cannot be directly compared due to other population differences this is one of the few studies in the U.S. that corresponds to the more numerous follow-up studies in European centers. None of the breast augmentations and <5% of the vaginoplasties and orchiectomies were paid for by insurance for male to female patients and none of the surgical procedures were covered by insurance, using transsexualism as the diagnosis for female to male patients. The investigators assessed employment, substance abuse, psychiatric disease, mood disorders, Human Immunodeficiency Virus (HIV) status in patients who had met WPATH guidelines for therapy, and who had initiated cross-sex hormone treatment.
A total of 242 patients treated for gender identity disorder in the clinic from 1992 through 2009 inclusive were identified. The number of those presenting for therapy almost tripled over time. Of these patients, 50 (20.7%) were female-to-male; 192 (79.3%) male-to-female (ratio 1:3.8).

The age of female-to-male and male-to-female patients with gender dysphoria at the time of clinic presentation was 29.0 and 38.0 years respectively.

The female-to-male and male-to-female patients with gender dysphoria at the time of hormone initiation were young: 27.5 and 35.5 years old respectively (p<0.5). Of the male-to-female cohort, 19 (7.8%) had received hormone therapy in the absence of physician supervision; 91 (37.6%) had undergone gender-reassignment surgery (32 female-to-male [64.0% of all female-to-male; 35.2% of all surgical patients]; 59 male-to-female [30.7% of all male-to-female; 64.8% of all surgical patients]; ratio 1:1.8).

Psychiatric disease was more common in those who initiated hormone therapy at an older age (>32 years) 63.9% versus 48.9% at a younger age and by natal gender (48.0% of female-to-male; 58.3% male-to-female). Mood disorders were more common in those who initiated hormone therapy at an older age (>32 years) 52.1% versus 36.0% at a younger age and this finding did not differ by natal gender (40.0% of female-to-male; 44.8% male-to-female). The presence of mood disorders increased the time to reassignment surgery in male-to-female patients. Of participants 36.4% were employed in jobs requiring a high school degree or less; 28.1% (excluding students) were on disability and/or unemployed. Rates of disability and unemployment were higher in male-to-female patients (31.8%) than female-to-male patients (14.0%). Mental health diagnoses reportedly were the major reason for disability. HIV infection and substance abuse were higher in male-to-female patients than female-to-male patients (8.3% versus 0% and 12.5% versus 6.0% respectively). This population has a high proportion of mental health problems, unemployment, and HIV positivity. These high levels of comorbidities and adverse outcomes are disproportionate to the rates among patients enrolled in the European studies of the same time period.

m. Retrospective, non-blinded, observational, longitudinal studies


Asscheman et al. conducted a retrospective, non-blinded, observational study of mortality using a longitudinal design of population treated with hormones, as well as, reassignment surgery and a population-based cohort. The investigators assessed mortality in patients who (a) were from a single-center, unspecified, university specialty clinic, (b) initiated cross-sex hormone treatment prior to July 1, 1997, and (c) had been followed by the clinic for at least 1 year or had expired during the first year of treatment. The National Civil Record Registry (Gemeentelijke Basis Administratie) was used to identify/confirm deaths of clinic patients. Information on the types or hormones used was extracted from clinic records, and information on the causation of death was extracted from medical records or obtained from family physicians. Mortality data for the general population was obtained through by the Central Bureau of Statistics of the Netherlands.
Mortality data from Acquired Immune Deficiency Syndrome (AIDS) and substance abuse were extracted from selected Statistics Netherlands reports. The gender of the general Dutch population comparator group was the natal sex of the respective gender dysphoric patient groups.

A total of 1,331 patients who met the hormone treatment requirements were identified (365 female-to-male [27.4%]; 966 male-to-female [72.6%]; ratio 1:2.6). Of these, 1,177 (88.4%) underwent reassignment surgery (343 [94.0% of female-to-male entrants]; 834 [86.3% of male-to-female entrants]; ratio 1:2.4; p<0.0001). The mean age at the time of hormone initiation in female-to-male and male-to-female patients was young: 26.1±7.6 (range 16–56) years and 31.4±11.4 (range 16–76) years respectively, although the male-to-female subjects were relatively older (p<0.001). The mean duration of hormone therapy in female-to-male and male-to-female patients was 18.8±6.3 and 19.4±7.7 years respectively.

There were a total of 134 deaths in the clinic population using hormone therapy as well as reassignment surgery. Of the patients, 12 (3.3%) of the 365 female-to-male patients and 122 (12.6%) of the 966 male-to-female patients died. All-cause mortality was 51% higher and statistically significant (Standardized Mortality Ratio [SMR] 95% confidence interval [CI]) 1.47-1.55 for males-to-females when compared to females-males in the general Dutch population. The small increase in all-cause mortality (12%) for females-to-males when compared to males-males in the general Dutch population was not statistically significant; 95% CI 0.87-1.42.

The major known contributors to the mortality difference between male-to-female patients and the Dutch population at large were ischemic heart disease (n=18, SMR 1.64 [95% CI 1.43-1.87]), completed suicide (n=17, SMR 5.70 [95% CI 4.93-6.54]), AIDS (n=16, SMR 30.20 [95% CI 26.0-34.7]), hematological cancer (n=6, SMR 2.58 [95% CI 1.97-3.30]), and illicit drug use (n=5, SMR 13.20 [95% CI 9.70-17.6]). An additional major contributor was “unknown cause” (n=21, SMR 4.00 [95% CI 3.52-4.51]). Of the 17 male-to-female hormone treated patients who committed suicide, 13 (76.5%) had received prior psychiatric treatment and 6 (35.3%) had not undergone reassignment surgery because of concerns about mental health stability.

The mean age for ischemic heart disease was a major disparate contributor to excess mortality in male-to-female patients was 59.7 years in older patients (n=18, SMR 1.64 [95% CI 1.43-1.87], mean age [range]: 42-79) years. Current use of a particular type of estrogen, ethinyl estradiol, was found to contribute to death from myocardial infarction or stroke (Adjusted Hazard Ratio 3.12 [95% CI 1.28-7.63], p=0.01). There was a smaller, but statistically significant increase in lung cancer that was thought to possibly be related to higher rates of smoking in this cohort.

As the authors noted, comparison of a clinical population (transgender patients) to a non-clinical population increases bias and is confounded by lifestyle factors, prone to associated pathology and other factors specific for the transsexual population that are not referable to cross-sex hormone treatment. In particular in this study, male to female patients had an increased history of suicide attempts, more psychopathology, and substance abuse, which although probably
associated with the psychological burden of gender dysphoria was not related to hormonal treatment. They also had an increased prevalence of HIV infection prior to entering the study.

Although overall mortality was not increased in the hormone-treated female-to-male patients, there were more deaths due to illicit drug use than expected (SMR 25 [6.00–32.5]).

n. Retrospective, non-blinded, observational, longitudinal studies using matched national data


Dhejne et al. conducted a retrospective, non-blinded, observational study of nation-wide mortality using a longitudinal and a population-based matched cohort. The investigators assessed mortality, suicide attempts, psychiatric morbidity, accidents hospitalization, and crime following sex reassignment including: (1) all-cause mortality, (2) death by definite/uncertain suicide, (3) death by cardiovascular disease, and (4) death by tumour. Morbidity included (5) any psychiatric disorder (substance abuse in gender identity disorders excluded), (6) alcohol/drug misuse and dependence, (7) definite/uncertain suicide attempt, and (8) accidents, convictions for (9) any criminal offence and (10) any violent offence reassigned persons and randomly selected unexposed non-clinical controls matched by birth year and natal sex (1:10) as well as by birth year and the reassigned gender (1:10). Data were extracted from national databases including the Total Population Register (Statistics Sweden), the Medical Birth Register, the Cause of Death Register (Statistics Sweden), the Hospital Discharge Register (National Board of Health and Welfare), the Crime Register (National Council of Crime), and those from the Register of Education for highest educational level. The criteria required to obtain the initial certificate for reassignment surgery and change in legal status from the National Board of Health and Welfare were not delineated, but included evaluation and treatment by one of 6 specialized teams, name change, a new national identity number, continued use of hormones, and sterilization/castration. Descriptive statistics with hazard ratios were provided. There were 804 patients identified with Any 302.XX gender identity disorder according to ICD-9 (or related disorders) in Sweden during the period from 1973 to 2003 inclusive. Of these patients, 324 (40.3%) underwent gender-reassignment surgery (133 female-to-male [41.0%]; 191 male-to-female [59.0%]; ratio 1:1.4). The 480 persons that did not shift gender variable comprise persons who either did not apply, or were not approved, for sex reassignment surgery or comprise persons that with sexual disorders other than transsexualism as 302.XX is not specific to gender dysphoria. The average follow-up time for all-cause mortality was 11.4 years (median 9.1). The average follow-up time for psychiatric hospitalization was 10.4 years (median 8.1).

The mean ages in female-to-male and male-to-female reassigned patients were: 33.3±8.7 (range 20–62) and 36.3±10.1 (range 21–69) respectively. Immigrant status was twice times higher in reassigned patients (n=70, 21.6%) than in either type of control (birth [natal] sex matched n=294 [9.1%] or reassigned gender matched n=264 [8.1%]). Educational attainment (10 or more years) was somewhat lower for reassigned patients (n=151 [57.8%]) than in either type of control (birth sex matched n=1,725 [61.5%] or reassigned gender matched n=1,804 [64.3%]) (cohort data were
incomplete). The biggest discordance in educational attainment was for female-to-male reassigned patients regardless of the control used. Prior psychiatric morbidity (which did not include hospitalization for gender dysphoria) was more than four times higher in reassigned patients (n=58, 17.9%) than in either type of control (birth sex matched n=123 [3.8%] or reassigned gender matched n=114 [3.5%]).

All-cause mortality was higher for patients who underwent gender reassignment surgery (n=27 [8.3%]) than in controls (hazard ratio 2.8 [1.8-4.3]) even after adjustment for covariants (prior psychiatric morbidity and immigration status). Divergence in the survival curves began at 10 years. The major contributor to this mortality difference was completed suicide (n=10 [3.1%]; adjusted hazard ratio 19.1 [5.8-62.9]). Mortality due to cardiovascular disease was modestly higher for reassigned patients (n=9 [2.8%]) than in controls (hazard ratio 2.5 [1.2-5.3]).

Suicide attempts were more common in patients who underwent gender reassignment surgery (n= 29 [9.0%] than in controls (adjusted hazard ratio 4.9 [2.9-8.5]). Male- to-female patients were at higher adjusted risk for attempted suicide than either control whereas female-to-male patients were at higher adjusted risk compared to only male controls and maintained the female pattern of higher attempted suicide risk. Hospitalizations for psychiatric conditions (not related to gender dysphoria) were more common in reassigned persons n= 64 [20.0%] than in controls (hazard ratio 2.8 [2.0-3.9]) even after adjusting for prior psychiatric morbidity. Hospitalization for substance abuse was not greater than either type of control. The increased risk for conviction of any crime or violent crime observed during the 1973-1988 interval was not seen later.

It is crucial to remember though that this was a study of postoperative patients compared to a general population sample. This does not speak at all to the efficacy of the treatments provided to transgender patients in the study. As the authors state in the paper: “It is therefore important to note that the current study is only informative with respect to transsexuals persons health after sex reassignment; no inferences can be drawn as to the effectiveness of sex reassignment as a treatment for transsexualism. In other words, the results should not be interpreted such as sex reassignment per se increases morbidity and mortality. Things might have been even worse without sex reassignment. As an analogy, similar studies have found increased somatic morbidity, suicide rate, and overall mortality for patients treated for bipolar disorder and schizophrenia. This is important information, but it does not follow that mood stabilizing treatment or antipsychotic treatment is the culprit.”

Dhejne provided additional analysis by dividing their 30 year cohort in half and assessing morbidity and mortality differences between the early and later group. They found that the statistically significant higher mortality, and specifically suicide rate, was confined to those who had surgery in 1973-1988, and was not found in those having surgery in 1989-2003. That is, transgender persons operated on after 1989 had an overall mortality rate and suicide rate indistinguishable from the non-clinical control sample. This is despite the fact that a higher mortality and suicide rate would generally be expected in any clinical sample of patients with a significant mental health diagnosis who are well known from numerous studies in the literature.
to already have high suicidality pre-treatment. In addition there was an increased risk for conviction of any crime or violent crime observed during the 1973-1988 interval was not seen later.


Dhejne et al. conducted a non-blinded, observational study that was longitudinal for the capture of patients with “regret” in a national database. This same group (Landen et al., 1998) conducted a similar study along with retrospective acquisition of clinical data to explore the differences between the cohorts with and without regret. There were no external controls; only intra-group comparisons for this surgical series.

The investigators assessed the frequency of regret for gender reassignment surgery. Data were extracted from registries at the National Board of Health and Welfare to which patients seeking reassignment surgery or reversal of reassignment surgery make a formal application and which has maintained such records since a 1972 law regulating surgical and legal sex reassignment. The investigators reviewed application files from 1960 through 2010. The specific criteria to qualify for gender surgery were not delineated. Patients typically underwent diagnostic evaluation for at least 1 year. Diagnostic evaluation was typically followed by the initiation of gender confirmation treatment including hormonal therapy and real-life experience. After 2 years of evaluation and treatment, patients could make applications to the national board. Until recently sterilization or castration were the required minimal surgical procedures. (Dhejne et al., 2011) Secular changes in this program included consolidation of care to limited sites, changes in accepted diagnostic criteria, and provision of non-genital surgery, e.g., mastectomy during the real-life experience phase, and family support.

Of the 767 applicants for legal and surgical reassignment (289 [37.7%] female-to-male and 478 [62.3%] male-to-female; ratio 1:1.6]. The number of applicants doubled each ten year interval starting in 1981.

Of the applicants, 88.7% or 681 (252 [37.0%] female-to-male and 429 [63.0%] male-to-female; ratio 1:1.7] had undergone surgery and changed legal status by June 30, 2011. This number included eight (four [50.0%] female-to-male and four [50.0%] male to female; ratio 1:1) people who underwent surgery prior to the 1972 law. (This number [6.0%] appears to include 41 (two [4.9%] female-to-male and 39 [95.1%] male-to-female; ratio 1:19.5) people who underwent surgery abroad at their own expense [usually in Thailand or the U.S.]. This cohort includes one person who was denied reassignment surgery by Sweden.)

Twenty-five (3.3%) of the applications were denied with the two most common reasons being an incomplete application or not meeting diagnostic criteria. An additional 61(8.0%) withdrew their application, were wait-listed for surgery, postponed surgery (perhaps in hopes of the later revocation of the sterilization requirement), or were granted partial treatment.
The formal application for reversal of the legal gender status, the “regret rate”, was 2.2%. No one who underwent sex-reassignment surgery outside of Sweden (36 of 41 after 1991) has requested reversal. The authors noted, however, that this preliminary number may be low because the median time interval to reversal request was eight years-only three of which had elapsed by publication submission-and because it was the largest serial cohort. This number did not include other possible expressions of regret including suicide (Dhejne et al., 2011).

Dhejne et al. in 2014 reported that the female-to-male: male-to-female ratio among those who made formal applications for reversal was 1:2. But this was due to the higher proportion of male to female patients to female to male patients. The regret rate was 2.0% transgender men and 2.3% in transgender women. The investigators also reported that the female-to-male applicants for reversal were younger than the entire female-to-male cohort (median age 22 versus 27 years) while the male-to-female applicants for reversal were older than the entire male-to-female cohort (median age 35 versus 32 years). Other clinical data to explore the differences between the cohorts with and without regret were not presented in this update publication.

In their earlier publication, in addition to determining a regret rate (3.8%), Landen et al. extracted data from medical records and government verdicts. Logistic regression analyses were used identify relationships between variables. They observed that: (a) 25.0% of the cohort with regrets and 11.4% of the cohort without regrets were unemployed, (b) 16.7% of the cohort with regrets and 15.4% of the cohort without regrets were on “sick benefit”, (c) 15.4% of the cohort with regrets and 13.9% of the cohort without regrets had problems with substance abuse, (d) 69.2% of the cohort with regrets and 34.6% of the cohort without regrets had undergone psychiatric treatment, (e) 15.4% of the cohort with regrets and 8.8% of the cohort without regrets had a mood disorder, and (f) 15.4% of the cohort with regrets and 1.5% of the cohort without regrets had a psychotic disorder.

The putative prognostic factors that were statistically different (albeit without Bonferroni correction) between the cohorts with and without regret included prior psychiatric treatment, a history of psychotic disorder, atypical features of gender identity, and poor family support. Of these family acceptance and support was of greatest influence. Factors that trended towards statistical difference included having an unstable personality, sexual orientation and transvestitism. These variables were tested with logistic regression. Initial modeling included the variable “history of psychotic disorder”. Although this variable was predictive, it was excluded from future analyses because it was already a contraindication to reassignment surgery. Additional analyses identified poor family support as the most predictive variable and atypical features of gender identity as the second most important variable. Presence of both variables has a more than additive effect.


o. Randomized, non-blinded, longitudinal, some patients served as their own controls

Mate-Kole at al. conducted a prospective, non-blinded, controlled, randomized, longitudinal study using investigator-designed patient self-report questionnaires and non-specific psychological tests with some normative data. The investigators assessed neuroticism and sex role in natal males with gender dysphoria who had qualified for male-to-female reassignment surgery at a single-center specialty clinic. Forty sequential patients were alternately assigned to early reassignment surgery or to standard wait times for reassignment surgery. Patients were evaluated after acceptance and 2 years later. The criteria used to qualify for gender surgery were the 1985 standards from the Harry Benjamin International Gender Dysphoria Association. These included a ≥2 year desire to change gender, a ≥1 year demonstrable ability to live and be self-supporting in the chosen gender, and psychiatric assessment for diagnosis and reassessment at 6 months for diagnostic confirmation and exclusion of psychosis. Reassignment surgery was defined as orchidectomy, penectomy, and construction of a neo-vagina. The instruments used were the CCEI for psychoneurotic symptoms and the Bem Sex Role Inventory along with an incompletely described investigator-designed survey with questions about social life and sexual activity. The mean age and range of the entire cohort was 32.5 years (21-53).

Members of the early surgery cohort had a history of attempted suicide (one patient), psychiatric treatment for non-gender issues (six patients), and first degree relatives with psychiatric histories (four patients). Members of the standard surgery cohort were similar, with a history of attempted suicide (two patients), psychiatric treatment for non-gender issues (five patients), and first degree relatives with psychiatric histories (six patients). The early surgery group had surgery approximately 1.75 years prior to the follow-up evaluation. In both groups, cross-dressing began at about age 6.

At baseline, the Bem Sex Role Inventory femininity scores were slightly higher than masculinity scores for both cohorts and were similar to Bem North American female normative scores. The scores did not change in either group over time.

At baseline, the scores for the CCEI individual domains (free floating anxiety, phobic anxiety, somatic anxiety, depression, hysteria, and obsessional anxiety) were similar for the cohorts. The total CCEI scores for the two cohorts were consistent with moderate symptoms. Over the 2 year interval, total CCEI scores increased for standard wait group and approached the relatively severe symptom category. During the same interval, scores dropped into the asymptomatic range for the post-operative patients.

The investigator-designed survey assessed changes in social and sexual activity of the prior 2 years, but the authors only calculated comparisons of compared patients in a given cohort to themselves. Though the researchers did not report conduct statistical studies to compare the differences between the 2 cohorts, they did report increased participation in some, but not all, types of social activities such as sports (solo or group), dancing, dining out, visiting pubs, and visiting others. Sexual interest also increased. By contrast, pre-operative patients did not increase their participation in these activities. Work status remained the same for post-operative patients.
while which unemployment increased in the standard wait pre-operative cohort. The authors allude to comparing the two cohorts in stating that the post-operative group was significantly more active than the pre-operative group in sports, dancing, dining out, and visiting others, but they do not include the figures.

2. External Technology Assessments

a. CMS did not request an external technology assessment (TA) on this issue.

b. There were no AHRQ reviews on this topic.

c. There are no Blue Cross/Blue Shield Health Technology Assessments written on this topic within the last three years.

d. There were two publications in the COCHRANE database, and both were tangentially related. Both noted that there are gaps in the clinical evidence base for gender reassignment surgery.


“Findings supported that LGBT issues have been neglected by public health research and that research unrelated to sexually transmitted diseases is lacking.”

A systematic review of lesbian, gay, bisexual and transgender health in the West Midlands region of the UK compared to published UK research. West Midlands Health Technology Assessment Collaboration. Health Technology Assessment Database. Meads, et al., 2009. No.3.

“Further research is needed but must use more sophisticated designs with comparison groups. This systematic review demonstrated that there are so many gaps in knowledge around LGBT health that a wide variety of studies are needed.”

e. There were no National Institute for Health and Care Excellence (NICE) reviews/guidance documents on this topic.

f. There was a technology assessment commissioned by the New Zealand Ministry of Health and conducted by New Zealand Health Technology Assessment (NZHTA) (Christchurch School of Medicine and the University of Otago).


The research questions included the following: (1) Are there particular subgroups of people with transsexualism who have met eligibility criteria for gender reassignment surgery (GRS) where evidence of effectiveness of that surgery exists? And (2) If there is evidence of effectiveness,
what subgroups would benefit from GRS?” Based upon the research, “Some 593 possibly relevant articles in abstract form were identified of which 70 articles were retrieved in full text.”

The NZHTA stated, “The reviewed studies may indicate that early, rather than delayed, sex reassignment surgery is of greater benefit to transsexual people who have gone through rigorous assessment procedures and have been accepted for surgery. Also, the reviewed studies identify characteristics of groups defined as core and non-core transsexual people, but these characteristics are heterogeneous and anecdotal.”

The NZHTA also stated, “Gender reassignment surgery may benefit some carefully assessed and selected transsexual people who have satisfied recognized diagnostic and eligibility criteria, and have received recognized standards of care for surgery. More research is required to improve the evidence base identifying the subgroups of transsexual people most likely to benefit from sex reassignment surgery.”

Given that the majority of research published regarding transgender patients has been published after the date of this analysis, the utility of the NZHTA is questionable. In addition, the NZHTA describes it’s Tech Briefs as “rapidly produced assessments of the best available evidence for a topic of highly limited scope.”

3. Medicare Evidence Development & Coverage Advisory Committee (MEDCAC) Meeting

CMS did not convene a MEDCAC meeting.

4. Evidence-Based Guidelines

a. American College of Obstetricians and Gynecologists (ACOG)

Though ACOG did not have any evidence-based guidelines on this topic, they did have the following document:

Health Care for Transgender Individuals: Committee Opinion


“Questions [on patient visit records] should be framed in ways that do not make assumptions about gender identity, sexual orientation, or behavior. It is more appropriate for clinicians to ask their patients which terms they prefer. Language should be inclusive, allowing the patient to decide when and what to disclose. The adoption and posting of a nondiscrimination policy can also signal health care providers and patients alike that all persons will be treated with dignity and respect. Assurance of confidentiality can allow for a more open discussion, and confidentiality must be ensured if a patient is being referred to a different health care provider. Training staff to increase their knowledge and sensitivity toward transgender patients will also help facilitate a positive experience for the patient.”

Comment [RNG23]: While not evidence based guidelines, the lack of a description of consultations with relevant experts should have been performed and included. We are also globally concerned that professional consensus from medical organizations was not well represented.
b. American Psychiatric Association

Report of the American Psychiatric Association Task Force on Treatment of Gender Identity Disorder

The American Psychiatric Association (APA) was unable to identify any Randomized Controlled Trials (RCTs) regarding mental health issues for transgender individuals.

"There are some level B studies examining satisfaction/regret following sex reassignment (longitudinal follow-up after an intervention, without a control group); however, many of these studies obtained data retrospectively and without a control group (APA level G). Overall, the evidence suggests that sex reassignment is associated with an improved sense of well-being in the majority of cases, and also indicates correlates of satisfaction and regret. No studies have directly compared various levels of mental health screening prior to hormonal and surgical treatments on outcome variables; however, existing studies suggest that comprehensive mental health screening may be successful in identifying those individuals most likely to experience regrets."

Relevant Descriptions of APA Evidence Coding System/Levels:

[B] Clinical trial. A prospective study in which an intervention is made and the results of that intervention are tracked longitudinally. Does not meet standards for a randomized clinical trial.”

[G] Other. Opinion-like essays, case reports, and other reports not categorized above.”

c. Endocrine Society

Endocrine Treatment of Transsexual Persons: an Endocrine Society Clinical Practice Guideline.

This guideline primarily addressed hormone management and surveillance for complications of that management. A small section addressed surgery and found the quality of evidence to be low.

“This evidence-based guideline was developed using the Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) system to describe the strength of recommendations and the quality of evidence, which was low or very low.”

d. World Professional Association for Transgender Health (WPATH)

The WPATH is “an international, multidisciplinary, professional association whose mission is to promote evidence-based care, education, research, advocacy, public policy, and respect in transsexual and transgender health.”

WPATH reported, “The standards of care are intended to be flexible in order to meet the diverse health care needs of transsexual, transgender, and gender-nonconforming people. While flexible, they offer standards for promoting optimal health care and guiding the treatment of people experiencing gender dysphoria—broadly defined as discomfort or distress that is caused by a discrepancy between a person’s gender identity and that person’s sex assigned at birth (and the associated gender role and/or primary and secondary sex characteristics) (Fisk, 1974; Knudson, De Cuypere, & Bockting, 2010b).”

The WPATH standards of care (SOC) “acknowledge the role of making informed choices and the value of harm-reduction approaches.”

The SOC noted, “For individuals seeking care for gender dysphoria, a variety of therapeutic options can be considered. The number and type of interventions applied and the order in which these take place may differ from person to person (e.g., Bockting, Knudson, & Goldberg, 2006; Bolin, 1994; Rachlin, 1999; Rachlin, Green, & Lombardi, 2008; Rachlin, Hansbury, & Pardo, 2010). Treatment options include the following:

- Changes in gender expression and role (which may involve living part time or full time in another gender role, consistent with one’s gender identity);
- Hormone therapy to feminize or masculinize the body;
- Surgery to change primary and/or secondary sex characteristics (e.g., breasts/chest, external and/or internal genitalia, facial features, body contouring);
- Psychotherapy (individual, couple, family, or group) for purposes such as exploring gender identity, role, and expression; addressing the negative impact of gender dysphoria and stigma on mental health; alleviating internalized transphobia; enhancing social and peer support; improving body image; or promoting resilience.”

The SOC were carefully reviewed because they are frequently cited as the basis of management by clinicians, including some of the clinical groups with whom CMS spoke used it as a flexible guide. In the WPATH’s SOC Appendix D titled “Evidence for Clinical Outcomes of Therapeutic Approaches,” WPATH noted, “One of the real supports for any new therapy is an outcome analysis. Because of the controversial nature of sex reassignment surgery, this type of analysis has been very important. Almost all of the outcome studies in this area have been retrospective.” They further reported, “More studies are needed that focus on the outcomes of current assessment and treatment approaches for gender dysphoria.”
e. American Psychological Association


“The purpose of the Guidelines for Psychological Practice with Transgender and Gender Nonconforming People (hereafter Guidelines) is to assist psychologists in the provision of culturally competent, developmentally appropriate, and trans-affirmative psychological practice with TGNC people.”

“These Guidelines refer to psychological practice (e.g., clinical work, consultation, education, research, training) rather than treatment.”

5. Other Reviews

a. Institute of Medicine (IOM)

The Health of Lesbian, Gay, Bisexual, and Transgender People: Building a Foundation for Better Understanding

“To advance understanding of the health needs of all LGBT individuals, researchers need more data about the demographics of these populations, improved methods for collecting and analyzing data, and an increased participation of sexual and gender minorities in research. Building a more solid evidence base for LGBT health concerns will not only benefit LGBT individuals, but also add to the repository of health information we have that pertains to all people.”

“Best practices for research on the health status of LGBT populations include scientific rigor and respectful involvement of individuals who represent the target population. Scientific rigor includes incorporating and monitoring culturally competent study designs, such as the use of appropriate measures to identify participants and implementation processes adapted to the unique characteristics of the target population. Respectful involvement refers to the involvement of LGBT individuals and those who represent the larger LGBT community in the research process, from design through data collection to dissemination.”

b. National Institutes of Health (NIH)

National Institutes of Health Lesbian, Gay, Bisexual, and Transgender (LGBT) Research Coordinating Committee. Consideration of the Institute of Medicine (IOM) report on the health
In response to the IOM report, the NIH LBGT research Coordinating Committee noted that most of the health research for this set of populations is “focused in the areas of Behavioral and Social Sciences, HIV (human immunodeficiency virus)/AIDS, Mental Health, and Substance Abuse. Relatively little research has been done in several key health areas for LGBT populations including the impact of smoking on health, depression, suicide, cancer, aging, obesity, and alcoholism.”

6. Pending Clinical Trials

ClinicalTrials.gov
There is one currently listed and recently active trial directed at assessment of the clinical outcomes pertaining to individuals who have had gender reassignment surgery. The study appears to be a continuation of work conducted by investigators cited in the internal technology assessment.

NCT01072825 (Ghent, Belgium sponsor) European Network for the Investigation of Gender Incongruence (ENIGI) is assessing the physical and psychological effects of the hormonal treatment of transgender subjects in two years prior to reassignment surgery and subsequent to surgery. This observational cohort study started in 2010 and is still in progress.

7. Consultation with Outside Experts

Consistent with the authority at 1862(l)(4) of the Act, CMS consulted with outside experts on the topic of treatment for gender dysphoria and gender reassignment surgery.

Given that the majority of the clinical research was conducted outside of the United States, and some studies took place in a or suggested continuity-of-care and coordination-of-care were beneficial to health outcomes, we conducted expert interviews with centers across the U.S. that provided some form of specialty-focused or coordinated care for transgender patients. These interviews informed our knowledge about the current healthcare options for transgender people, the qualifications of the professionals involved, and the uniqueness of treatment options. We are very grateful to the organizations that made time to discuss treatment for gender dysphoria with us.

From our discussions with the all of the experts we spoke with, we noted the following practices in some centers: (1) specialized cultural competency and basic medical training for all staff about transgender healthcare and transgender cultural issues; (2) use of an intake assessment by either a social worker or health care provider that addressed physical health, mental health, and other life factors such as housing, relationship, personal safety, domestic violence risk and employment status; (3) offering primary care services for transgender people in addition to services not related to gender-affirming therapy/treatments; (4) navigators who connected patients with
name-change information or other legal needs related to gender; (5) counseling for individuals, groups, and families; (6) an informed-consent model whereby individuals were often referred to as “clients” instead of “patients,” and (7) an awareness of and screening for depression and suicidality among transgender people (often measured with tools such as the Adult Outcomes Questionnaire and the Patient Health Questionnaire) and how, in some instances, with hormone treatment for gender dysphoria, the depression lessens.

8. Public Comments

Initial Comment Period: 12/03/2015 – 01/02/2016

During the initial comment period, we received 103 comments. Of those, 78% supported coverage of gender reassignment surgery, 15% opposed, and 7% were neutral. The majority of comments supporting coverage were from individuals and advocacy groups. All of the initial public comments are available at: https://www.cms.gov/medicare-coverage-database/details/nca-view-public-comments.aspx?NCAId=282&ExpandComments=n&bc=ACAAAAAAA%3d%3d

VIII. CMS Analysis

National coverage determinations are determinations by the Secretary with respect to whether or not a particular item or service is covered nationally under § 1862(l)(6) of the Act. In general, in order to be covered by Medicare, an item or service must fall within one or more benefit categories contained within Part A or Part B and must not be otherwise excluded from coverage. Moreover, in most circumstances, the item or service must be reasonable and necessary for the diagnosis or treatment of illness or injury or to improve the functioning of a malformed body member (§1862(a)(1)(A)). The Supreme Court has recognized that “[t]he Secretary’s decision as to whether a particular medical service is ‘reasonable and necessary’ and the means by which she implements her decision, whether by promulgating a generally applicable rule or by allowing individual adjudication, are clearly discretionary decisions.” Heckler v. Ringer, 466 U.S. 602, 617 (1984). See also, 78 Fed. Reg. 48,164, 48,165 (August 7, 2013)

When making national coverage determinations, we consider whether the evidence is relevant to the Medicare beneficiary population. In considering the generalizability of the results of the body of evidence to the Medicare population, we carefully consider the demographic characteristics and comorbidities of study participants as well as the provider training and experience. This section of the proposed decision provides an analysis of the evidence, which included the published medical literature and guidelines pertaining to gender dysphoria, that we considered during our review to answer the question:

Is there sufficient evidence to conclude that gender reassignment surgery improves health outcomes for Medicare beneficiaries with gender dysphoria?

A. Analysis

1. Study Demographics
These studies were conducted in a total of 13 countries. Most were conducted in Europe (a total of 24 in Europe: Belgium four, Germany four, Holland two, Norway one, Spain two, Sweden four, Switzerland three, the United Kingdom three [not including the Barrett, 1998 study and the duplicative Megeri, Khoosal, 2007 study], and Yugoslavia one). One was in Asia (Singapore); one in South America (Brazil). Seven were conducted in North America (U.S. six, Canada one). One of the North American studies was a U.S.-conducted internet survey with non-U.S. and U.S. participants with a sub-analysis of the U.S. patients (Newfield et al., 2006). As noted earlier in the introduction, the greater number of European studies and the dearth of U.S. studies reflects the fact that until the last 10-15 years there was near universal exclusion of coverage for transgender surgery under public and private health insurance. Despite the fact that DHHS finds exclusion of transgender care may be unlawful sex discrimination on the health insurance exchanges, it is still the case that in a majority of U.S. states, coverage of transgender care under private insurance often excludes transgender care. Only 10 states and the District of Columbia have explicit laws preventing exclusion of transgender care and most of these were enacted in the past five years.

All of the studies, with the exception of a national-wide mortality study (Dhejne et al., 2011) and a prospective longitudinal study of long term functional improvements in patients undergoing GRS (Johansson et al., 2010), the international internet survey (Newfield et al., 2006), and the internet/convention site survey (Ainsworth, Spiegel, 2010), were conducted with patient populations from single sites. Many of these clinical centers cited in these studies were specialized tertiary referral centers in Europe offering comprehensive, integrated (psychiatric-psychological, endocrine, and surgical) care and whose staff could have been involved in both the patient care and the study. Of the studies reviewed, the Lawrence, 2006 study was conducted by a physician psychologist who surveyed the patient population of a single U.S. surgeon. The Ainsworth, Spiegel, 2010 study was conducted by a U.S. otolaryngologist with extensive surgery training who assessed the impact of facial feminization on transgender patients. The Hess et al. 2014 study was undertaken at a German university urologic specialty clinic. The Wolfradt, Neumann, 2001 study was conducted in Germany by a university otorhinolaryngologist and psychologist on patients who had undergone vocal cord surgery after reassignment surgery. The Ruppin, Pfafflin 2015 study was undertaken by investigators who had seen the patients in a German clinic for Psychosomatic Medicine and Psychotherapy Forensic Psychotherapy Clinic.

2. Patient Population

Demographic assessments of the studies revealed that the mean ages of participants were in the 20s and 30s. (See Appendix C and Appendix D). Even when including standard deviation, most patients included in the study were under the age 60. Age of participants in the reviewed studies is important to assess generalizability to the Medicare population which is comprised predominantly of adults’ age 65 years and older. This may be a function of the fact that most studies are from European centers where GRS is provided under national health insurance programs. Because of this, patients may apply at a much younger age to have GRS. In the U.S. because such care has been most often excluded from insurance there is a population of older

Comment [RNG24]: The vast majority of what they presented was on the US participants, so specifying “international internet survey” is unnecessary.

---

5 https://www.healthcare.gov/transgender-health-care/
6 http://www.transequity.org/issues/resources/map-state-health-insurance-rules
transgender patients who have not undergone surgery. Therefore there is a population of older transgender individuals who may be undergoing surgery at a much older age than those represented in prior studies and as such the results in this population have not been adequately studied before making ongoing research in these populations in the U.S. even more important to study prospectively. While certain younger disabled adults are included in Medicare, generalizability of studies performed outside in the U.S. is likely reduced further since criteria to determine disability is unique to Medicare. When reporting ages of patients participating in studies, studies included mean age of population, but often failed to reveal standard deviation of the population. Most studies reported pre and post gender reassignment surgery ages, though some studies only reported post-surgery ages (Dehjne, 2011; Kuhn et al., 2009; Rakic et al., 1996; Ruppin, Pfafflin, 2015; Udeze et al., 2008; Megeri, Khoosal, 2007; Wolfradt, Neumann, 2001; Blanchard et al., 1985; Weyrs et al., 2009; Wierckx et al., 2011; Eldh et al., 1997; Hess et al., 2014; Lawrence, 2006; Salvador et al., 2012; Tsoi, 1993).

There was extensive lack of study participation and loss to follow-up in the published studies. (See Appendix C and Appendix G). This suggests that the population that seeks evaluation/treatment for gender dysphoria and/or applies for reassignment surgery is not the same population that undergoes reassignment surgery without hesitation or regret. The notable numbers of incomplete questionnaires similarly raises questions. This selection bias limits generalizability of any results.

3. Study Design

As noted earlier, a number of research designs were found when exploring the question, “Does gender reassignment surgery improve health outcomes for Medicare beneficiaries with gender dysphoria?” (See Appendix C). The vast majority of studies found were observational in nature though there was a single randomized trial performed 26 years ago (Mate-Kole et al., 1990) (see Figure 1). This reflects the fact that currently, it would be unethical to perform an RCT, and IRB approval for this type of study would be impossible to obtain. Two of the studies were blinded. (Hess, 2014; Lawrence, 2006) A total of 29 studies were not blinded. The blinding status of the two internet surveys is unknown (Ainsworth, Spiegel, 2010; Newfield et al., 2006).

Observational studies can be prospective, retrospective, or have components of both. But each observational study design has limitations, and may not be able to show the true association between gender/reassignment surgery and improved health outcomes. Limitations of observational studies include that they frequently generate unreliable findings, and they also generate bias; because of confounding, causal inferences cannot reliably be drawn. Thus these types of studies are limited in terms of evidentiary weight. Only a true experimental study (e.g., randomized clinical trial) has the potential to demonstrate a definitive causal relationship between two factors. However clinical decisions about individual patient care can be informed by the evidence currently available, which is done routinely in many cases where the gold standard true placebo controlled RCT cannot be performed as is the case for many well accepted treatments, for example as use of N-acetylcysteine for acetaminophen overdose.
In general, one of the advantages of prospective studies is that they could potentially help determine factors associated with improved outcomes due to their longitudinal observation over time, and the collection of results at regular time intervals minimizes recall error. However, retrospective studies have problems including: some key statistics cannot be measured, significant biases including selection bias, recall bias, and information bias may limit a retrospective study’s applicability. Another problem with retrospective studies is that the temporal relationship between variables is frequently difficult to assess. Finally, it is difficult to control exposure or outcome assessment in a retrospective study design.

Studieds that use controls as part of its research design have higher evidentiary weight than studies that lack controls. That is because the use of controls can help to eliminate the possibility of confounding. But controls by themselves are no guarantee of complete validity. In terms of the use of controls in these studies that we evaluated some studies had no concurrent controls; some studies used control groups, but they were not concurrent; some studies used semi-matched controls; and in other studies patients served as their own controls.

Seventeen observational studies, of which 10 used longitudinal and 7 used cross-sectional study designs, had formal control groups. In this group of studies, the cross-sectional studies used various controls including healthy volunteers and patients with other disorders or treatments. In this same group of studies, the longitudinal studies used various controls including the patients as their own serial control, other treatment groups in addition to having patients serve as their own controls, and control cohorts derived from national databases. Among the longitudinal studies with used patients as their own controls, 4 used self-report test instruments that were validated in large populations. Of these 4, 1 had more than 100 subjects, self-reported and others, or other cohorts using either national data or national registries. Some observational studies included in this analysis had surgery-only populations and used no controls other than patients as their own controls, or used indirect controls, or incorporating normative testing. The remainder of the observational studies mixed populations that included surgical patients and patients using other treatments or patients treated with non-genital gender reassignment surgical procedures. The studies that included groups with mixed populations either had no controls, or used patients as their own controls or indirect controls (statistical methods included ANOVA, correlation, or regression).

Our review included 25 prospective studies. Of these prospective studies, one used a retrospective approach to acquire data for a single parameter (Eldh et al., 1997; Johansson et al., 2009); one prospective study used a retrospective approach to acquire data for several parameters (Ruppin, Pfaflin, 2015); and one study used a prospective approach beginning in 2003, but used a retrospective approach for data accumulated prior to that year (Leinung et al., 2013).

We found three retrospective studies (Asscheman et al., 2011; Dhejne et al., 2011; Landen et al., 1998). One study had at least a partially retrospective component, but with insufficient information to determine whether any of the data were obtained prospectively (Haraldsen, Dahl, 2000).

There were 12 longitudinal studies (Asscheman et al., 2011; Dhejne et al., 2011; Heylens et al., 2014; Johansson et al., 2010; Kockott, Fahrner, 1987; Landen et al., 1998; Mate-Kole et al., 2000).
1979; Rakic et al., 1996; Ruppin, Pfafflin, 2015; Smith et al., 2005; Udeze et al., 2008). Ten of
the longitudinal studies occurred in the group of studies with a designated control group (all of
the above with the exception of Asscheman et al., 2011). In seven of the 11 longitudinal studies,
the patients served as their own control over time before and after surgery (Heylens et al., 2014;
Kockott, Fahrner, 1987; Meyer, Reter, 1979; Rakic et al., 1996; Ruppin, Pfafflin, 2015; Smith et
al., 2005; Udeze et al., 2008).

There were 19 cross-sectional studies (Ainworth, 2010; Haraldsen, Dahl, 2000; Beatrice, 1985;
Kraemer et al., 2008; Kuhn et al., 2009; Mate-Kole et al., 1988; Wolfradt, Neumann, 2001;
Blanchard et al., 1985; Weyers et al., 2009; Wierckx et al., 2011; Eldh et al., 1997; Hess et al.,
2014; Lawrence, 2006; Salvador et al., 2012; Tsoi, 1993; Gómez-Gil et al., 2012, Hepp et al.,
2005; Motmans et al., 2012; Newfield et al., 2006; Gómez-Gil et al., 2013; Johansson et al.,
2009; Leinung et al., 2013). Of this number, one was two were cross-sectional with the exception
of data collection for aspects of a single parameter that had occurred in the past (Eldh et al.,
1997; Johansson et al., 2009), and one study asked participants to recall the status of a parameter
prior to treatment (Wierckx et al., 2011a).

Seventeen of the studies had explicit control groups. Of the studies with explicit control
groups, two studies derived controls from national databases (Dhejne et al., 2011 and 2014; Landen
et al., 1998); six five studies used the patients themselves as longitudinal controls (Heylens 2014a;
Johansson et al., 2010; Rakic et al. 1996; Ruppin, Pfafflin, 2015; Smith et al., 2005a; Udeze et
al., 2008; Megeri 2007); eight used various other controls (Ainsworth, Spiegel, 2010; Beatrice
1985; Haraldsen, Dahl, 2000; Kraemer et al., 2008; Kuhn et al., 2009; Mate-Kole et al., 1988 and
1990; Wolfradt, Neumann, 2001); and two studies used both treatment-type cohorts and patients
themselves as controls (Kockott, Fahrner, 1987; Meyer, Reter 1979).

Five A number of studies consisted of surgical series without concurrent controls (Wierckx et al., 2011; pansalador et al., 2012; Blanchard et al., 1985; Tsoi, 1993; Eldh et al., 1997; Hess et al., 2014; Lawrence, 2006; Weyers, 2009a). In three surgical series normative data from psychometric instruments were used as the control (Blanchard et al., 1985a; Weyers 2009a; Wierckx et al., 2011b). In five surgical series, controls were lacking (except for the use of serial employment data in the Eldh et al. 1997 study) (Eldh et al., 1997; Hess 2014; Lawrence 2006; Salvador 2012; Tsoi, 1993).

Patients underwent a variety of surgical interventions in five studies. There were no controls. The role of surgical intervention was in part assessed indirectly post hoc by statistical techniques (analysis of variance and regression) (Gomez-Gil et al., 2012 and 2014; Hepp et al., 2005; Motmans et al., 2011; Newfield et al., 2006).

As mentioned in previous paragraphs, some prospective studies included in this analysis were
cross-sectional in nature, and consisted of treated cohorts using a normative test, or a treatment
cohort along with volunteer healthy cohorts. However, as we have noted, cross-sectional studies
also have their limitations, including inability to determine temporal relationship between
exposure and outcome (lacks time element). In other words, findings noted in a cross-sectional
design cannot be inferred, because only current health and exposure to interventions are being
studied. Also measurement error is an issue. Longitudinal studies with controls, when ethically feasible are most appropriate for determining this relationship between exposure and outcomes.

Observational studies have limitations. The lack of blinding has the potential to interfere with patient reported outcomes, which by their nature are subjective. Observational studies are prone to selection bias. Patients who seek treatment may not be the same as those who complete treatment-particularly if there are serial steps in the treatment process. (See Appendix G) Cross-sectional studies are prone to confounding. The impact of a particular step in a multi-faceted treatment process cannot be ascertained with as much certainty. The lack of a control group limits the certainty of does not permit attribution of any outcome change to a specific intervention. There were seven studies where the patients themselves serve as longitudinal controls. The lack of an ideal control makes it difficult to assess the results because there is not an untreated group to make comparisons however in some cases patients serving as their own controls is the only possible study type - especially in rare diseases such as gender dysphoria. As an example, pegademase bovine (Adagen) for ADA deficient Severe Combined Immunodeficiency (SCID) was FDA approved based on a study of less than a dozen patients all of which served as their own control. Because this is a deadly rare disease, large gold standard placebo controlled RCTs are simply impossible and unethical to perform. While gender dysphoria is certainly more prevalent than ADA deficient SCID the same type of limitations exist to a lesser extent with gender dysphoria. The lack of a control makes it difficult to assess the results because there is not another group to make comparisons.

4. Psychometric Measurement Tools

There is also myriad use of measurement tools to assess patients suffering with gender dysphoria. (See Appendix E for a list of Psychometric Measurement tools.)

Some of the domains addressed in psychometric measurement tools measure the degree of depression and anxiety, body imagery, quality of life, identity traits, general wellbeing, physical and psychological function, self-concept, and others. Some of these measurement tools have been validated for patients with this condition, while others have been validated for other medical and/or mental health conditions. Some of the measurement tools found in this assessment were self-developed however only two have undergone subsequent validation and there is no mention of validity when trying to determine if the Utrecht Gender Dysphoria Scale (UGDS) and the Gender Identity/Gender Dysphoria Questionnaire for Adolescents and Adults (GIDYQ-AA) test reliably measures what it is intended to measure.

In most of the studies non-specific psychometric tests (and non-specific quality of life indicators) were used. Given that gender dysphoria is a rare disease (as defined in the United States), the lack for many years of disease specific psychometric and quality of life measures is not surprising. In the last few years new scores have been proposed and preliminarily validated, like

7 https://www.accessdata.fda.gov/drugsatfda_docs/label/2008/019818s042lbl.pdf
the Utrecht Gender Dysphoria Scale (UGDS) and the Gender Identity/Gender Dysphoria Questionnaire for Adolescents and Adults (GIDYQ-AA)\(^9\). However, it has been noted that in the instance of rare diseases, lacking a prospectively validated disease specific instrument, the next best option is a combination of non-specific instruments combined with measures determined by patients or clinical experts in the field often as a self-designed instrument\(^10\). It is this approach that is taken by many of the studies we reviewed. While this combination of measures is not as optimal as measures that are now available like the UGDS and the GIDYQ-AA, especially when interpreting studies from as long ago as 1979, lack of specific tools is expected. Given that “Evidence based medicine is defined as the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients”\(^11\), it we must acknowledge that the lack of specific measures for gender dysphoria is a function of the difficulties of performing research on small patient populations.

5. Study Endpoints

A wide variety of study endpoints were used. Endpoints were collected from a number of sources, including self-reporting, clinician assessment, and medical records as well national databases. Some of the endpoints included patient reported quality of life (QOL) as manifest by psychometric testing, sense of well-being, body imagery, anxiety and depression, sexual function and satisfaction, and social function. Objective endpoints included employment status, psychiatric function, and morbidity and mortality as well as adverse events.

Thirty of the studies employed 31 psychometric tools or investigator designed self-report surveys. (See Appendix E) Because of the aforementioned limitations in the literature only two of the specific tools (UGDS and GIDYQ-AA) for assessing gender dysphoria have published multicenter validation information. Twenty investigators designed their own measurement tools or modified those of others (including the initial publication of the two subsequently validated tools).-.

External information on test validity, the size/composition of the reference population(s), diagnostic cut-points, and scoring was often not available because some of it was unpublished, proprietary, or in a non-English language. Six of the instruments, all non-specific, (the European QOL Survey, MMPI-1 and MMPI-2, SF-36, SCL-90, TSCS, and WHO-QOL-BREF), appear to have substantive normative data for comparative scoring (i.e., reference populations (≥1000) and obtained through representative sampling). Although these tools had been validated in a reference population, none had been validated in populations with gender dysphoria, and in the case of the MMPI may overestimate pathology depending on the sex for which the test is scored.\(^9\)


Furthermore the investigators sometimes did not provide diagnostic cut-points and did not pre-
specify the magnitude of test score change or test score difference considered to be biologically
significant so the clinical importance could not be easily ascertained.

Only four investigator groups used only these psychometric tools validated in other large
populations as their test instrument (Beatrice, 1985; Haraldsen, Dahl, 2000; Motmans et al.,
2012; Newfield et al., 2006). Nine investigator groups used a mix of psychometric tools
validated in large normative populations, less well validated tools, and/or self-designed tools
(Ainsworth, Spiegel, 2010; Blanchard et al., 1985a; Gomez-Gil et al., 2014; Heylens 2014a;
Ruppin, Pfafflin, 2015; Smith et al., 2005a (Udeze et al., 2008; Megeri 2007; Weyers 2009a;
Wierckx et al., 2011b). Nine investigators used self-designed tools as their only test instrument
(Eldh et al., 1997; Hess 2014; Johansson et al, 2009; Kockott, Fahmner, 1987; Lawrence, 2006;
Meyer, Reter 1979; Rakic 1996; Salvador 2012; Tsai 1993). A single investigator did not use
any type of testing tool and provided only descriptive statistics (Leinung et al., 2013). However,
it has been noted that in the instance of rare diseases, lacking a prospectively validated disease
specific instrument, the next best option is a combination of non-specific instruments combined
with measures determined by patients or clinical experts in the field often as a self-designed
instrument. It is this approach that is taken by the 9 studies that used a mixture of validated
non-specific tools and self-designed tools. While this combination of measures is not as optimal
as validated measures that are recently available like the UGDS and the GIDYQ-AA, especially
when interpreting studies from as long ago as 1979, such tools were unavailable. Given that
“Evidence based medicine is defined as the conscientious, explicit, and judicious use of current
best evidence in making decisions about the care of individual patients,” it should be noted that
lacking the ideal evidence should not preclude making individual evidence-based decisions about
individual patients which is the basis of current treatment for patients with gender dysphoria
which we feel should continue in the absence of an NCD.

Three studies reported on complications linked or possibly linked to hormone treatment
(Asscheman et al., 2011; Dhejne et al., 2011; Leinung et al., 2013), six studies reported on
complications from reassignment surgery (Eldh et al., 1997; Lawrence, 2006; Ruppin, Pfafflin,
2015; Smith et al., 2005; Weyers et al., 2009; Wierckx et al., 2011). One study reported on
serious and formalized regret for undergoing reassignment surgery (Landen et al., 1998), and one
study reported on a single patient with suicidal ideation who requested phalloplasty due to
complications of phalloplasty, although did not wish to detransition or have any chance in the
removal of sex assignment results (i.e. mastectomy and the masculinizing effects of hormone
replacement therapy). (Meyer, Reter, 1979). Others reported on less severe or less formalized
levels of regret. Five studies reported on the treatment or diagnosis of psychiatric disease
(Dhejne et al., 2011; Haraldsen, Dahl, 2000; Hepp et al., 2005; Landen et al., 1998; Leinung et
al., 2013; Meyer, Reter, 1979; Ruppin, Pfafflin, 2015; Udeze et al., 2008). Three two studies
reported on the history of psychiatric disease in their patient populations (Matte-Kole, 1988;

Comment [RNG29]: This should be noted as it was in Meyer. The patient didn’t want to transition back or regret mastectomy, but just had such complications that he wished reversal of phalloplasty.

Comment [RNG30]: Maybe others as well, but Dhejne 2011 comes to mind.

---


Four studies reported on suicide attempts (Dehjne et al., 2011; Eldh et al., 1997; Heylens et al., 2013; Kockott, Fahrner, 1987), two studies reported on the history of suicide attempts in their patient population (Matte-Kole, 1988; Matte-Kole, 1990). Three studies reported on suicide, of which one of them occurred incidentally (Asscheman et al., 2011; Blanchard et al., 1985; Dhejne et al., 2011). Two studies also reported on overall mortality (Asscheman et al., 2011; Dhejne et al., 2011).

There was a great degree of inconsistency in endpoints. Also endpoints were collected from a number of sources, including self-reporting, clinician assessment, and medical records as well national databases. Some endpoints lacked operational definitions thus making their applicability difficult. CMS is interested in knowing what patients diagnosed with gender dysphoria believe are important endpoints that should be studied.

**Mortality and Regret as Endpoints**

Certain kinds of objective outcomes can be assessed by other types of study designs albeit somewhat less robust. These include mortality and regret when the data are rigorously prospectively collected in a comprehensive registry for all patients who have met specified entry criteria and treatment criteria. Because regret is a complicated outcome falling on a range of severity from regret due to complications but with overall satisfaction with the transition process, to severe regret manifested as request for surgical or medical detransition or even suicide, criteria for measuring severity of regret should be developed. In addition, because severe regret is relatively rare in studies to date, multicenter trials and research cooperatives such as the ENIGI trial in Europe should be encouraged in the United States.

More specifically, Swedish investigators extracted data from registries at the National Board of Health and Welfare to which all patients seeking reassignment surgery or reversal of reassignment surgery must make formal application. In the initial 1998 study, of the 233 applicants for reassignment surgery between July 1972 and June 1992, 20 were denied surgery, and subsequently 13 (3.8%) surgical patients requested return to the natal sex (Landen et al., 1998). In the 2014 follow-up study, of the 767 applicants for reassignment surgery or a change in legal status after surgery between 1960-2010, 86 were denied, and subsequently 15 (2.2%) requested reversal to the natal gender (Dhejne et al., 2014). Although the data from the two studies are not directly comparable because of the much shorter follow-up period in the latter study and although the analyses also did not consider other possible expressions of regret including suicide, the studies suggest that the majority of highly vetted patients in a structured care system do not express regret as defined by a formal request for return to natal gender status (Dhejne et al., 2011).

While the study cannot assess the impact of gender reassignment surgery alone because of the confounding introduced by the other interventions, because the vast majority of transgender patients undertake hormonal, surgical, and psychological care as part of their treatment, the combined therapy which most transgender patients undertake has this positive impact. In addition, a trial of surgery without hormonal or psychological treatment is not ethically possible given the demonstrated benefits for hormonal treatment and the fact that

---

without such treatments transgender patients would be subject to increased risks of interpersonal violence due to increased visibility as a transgender person. A national survey of transgender Americans found that visual non-conformity was associated with eliciting anti-transgender bias, increased risk of suicide, and increased risk of homelessness.\textsuperscript{15} The study, however, cannot assess the impact of gender reassignment surgery per se because of the confounding introduced by the other interventions.

Swedish investigators also conducted the most comprehensive study with functional endpoints of the 33 studies reviewed. This study relied on compulsory national databases (Dhejne et al., 2011) tracked all patients who had undergone reassignment surgery (at a mean age 35.1 years) over a 30 year interval and compared them to 6480 matched controls from the general population. They identified both increased mortality and increased psychiatric hospitalization. The mortality was primarily due to completed suicides (19.1-fold greater), but death due to neoplasm and cardiovascular disease was increased 2 to 2.5 times as well. The divergence in mortality from the Swedish population did not become apparent until after 10 years. The risk for psychiatric hospitalization was 2.8 times greater than in control Swedes even after adjustment for prior psychiatric disease (18%). The risk for attempted suicide was greater in male-to-female patients regardless of the sex of the control. For the same reasons as delineated above, this study cannot assess the impact of gender reassignment surgery per se because of the confounding introduced by the other interventions. The finding of this study demonstrated that reassignment surgery does not return patients to a normal level of morbidity risk and that the morbidity risk is significant even in highly vetted patients in a structured care system.

B. Discussion

The question addressed in this NCD is whether there is sufficient evidence to conclude that gender reassignment surgery improves health outcomes for Medicare beneficiaries with gender dysphoria.

Gender dysphoria by the latest and prior nomenclature is a state in which there is incongruence between the gender assigned at birth and the gender\textsuperscript{15} with which the person identifies. This incongruence may result in varying degrees of dysphoria - the primary symptom of the diagnosis. However this can be severed content and results in a high rate of suicidality in untreated patients which is ameliorated significantly with treatment, but does not return to general population levels of distress. Satisfaction and quality-of-life are well recognized as “latent variables” (hypothetical constructs) which cannot be measured directly (Borsboom et al., 2003; Newsom, 2015). As such, observable entities are used to infer or approximate satisfaction and/or quality-of-life. It may be challenging to identify parameters that truly reflect the nature and the magnitude of dysphoria in the individual. This challenge is followed by the need to know to what extent a specific test measures that which it purports to measure (test validity) and whether repeat testing will yield a comparable answer (test reliability).

The investigators of the clinical research reviewed in this NCD have attempted to measure dysphoria levels by objective data elements and by use of various psychometric and function scales/surveys. The objective data elements include a number of variables such as employment, mortality due to homicide and suicidality, and formal requests for surgical reversal.

The psychometric tools used to assess outcomes have limitations. Many of the instruments that are most specific for the condition were designed by the investigators themselves or by other investigators in the field. In addition, the relevant diagnostic cut-points for scores and changes in scores that are clinically significant should be delineated to permit adequate interpretation of test results. As such, these studies were not definitive in nature.

Other factors might impact the utility of a given test. Patients undergo serial evaluations and a sequence of treatments (Bockting et al., 2011). These other interventions may reduce internal validity of the test. The affirmation and support obtained in psychotherapy-psychiatric care, the adjustment confidence gained in real life cross-gender behavior, and/or the physical and mental changes from hormone therapy contribute to the improvement in symptoms and make it difficult to ascertain what specific components of the process have the greatest effect. Some have suggested that alternative causes (e.g., depression) may be responsible for the noted changes. Several studies suggest that there is a major therapeutic benefit from hormone therapy (Colizzi et al., 2013; Gómez-Gil et al., 2011; Gorin-Lazard et al., 2011; Heylens et al., 2014; Dubois, 2012). Another suggests that there is therapeutic benefit from time in the preferred gender role without other intervention (Greenberg, Laurence, 1981). As such, results from cross-sectional studies may be misleading. None of the studies used matched controls over time. We believe more longitudinal studies with serial assessment of the same patients will make the existing evidence more robust. We note that even from the results from the six four studies in which patients served as their own controls and which used an instrument known to be validated in large populations were negative (i.e., there was no improvement in psychometric or quality of life outcomes when patients were tested just prior to and at some point after the reassignment surgical intervention). (Blanchard, 1985; Heylens, 2014; Johansson 2010; Ruppin, Pfafflin, 2015; Smith et al., 2005; Udeze et al., 2008). Further, rigorous studies with the use of appropriate comparison patients could better clarify the specific benefits and harms of each of the interventions. However as described above, ethical considerations preclude withholding of psychotherapy or medical therapy, so the best possible study would be a rigorous analysis of currently accepted treatment algorithms which include multiple treatments at the same time. For example, WPATH criteria state that mastectomy can be undertaken at the same time as hormone therapy in transgender men - which may occur before or after real life experience depending on the patient’s social situation and his therapist’s recommendations.

CMS reviewed and considered potential objective measures of function including mortality, psychiatric treatment, and attempted suicide. None of the longitudinal studies in which patients served as their own control, however, comprehensively tracked changes in these events as objective measures of function before and after surgery. Events post treatment such as suicide and institutionalization were so few in number that they were not statistically analyzable with the small numbers of patients mentioned incidentally when describing patients in studies resulting excluded from the rarity of this condition. Even suicide which is overall more prevalent in some studies has such a low incidence that statistical analysis is challenging. Follow-up study or impossible.
during the study (Heylens et al., 2014; Ruppin, Pfafflin, 2015). Other times investigators tracked these functional outcomes (e.g., psychiatric out-patient treatment, psychiatric in-patient treatment, and substance abuse) for the most current prior year (Ruppin, Pfafflin, 2015).

The most comprehensive study with functional endpoints, the Swedish study that followed all patients who had undergone reassignment surgery (at mean age 35.1 years) over a 30 year interval and compared them to 6480 matched non-clinical controls, identified increased mortality and increased psychiatric hospitalization (Dhejne et al., 2011). The mortality was primarily due to completed suicides which in the entire thirty year sample was 19.1-fold greater than in control Swedes. However, death due to neoplasm and cardiovascular disease was increased 2 to 2.5 times as well. The divergence in mortality from the Swedish population did not become apparent until after 10 years. The risk for psychiatric hospitalization was 2.8 times greater than in controls even after adjustment for prior psychiatric disease (18%). The risk for attempted suicide was greater in male-to-female patients regardless of the gender of the control. Unfortunately, the study was not constructed to assess the impact of gender reassignment surgery per se. As the authors of the study explain: “no inferences can be drawn as to the effectiveness of sex reassignment as a treatment for transsexualism. In other words, the results should not be interpreted such as sex reassignment per se increases mortality risk and morbidity. Things might have been even worse without sex reassignment. As an analogy, similar studies have found increased somatic morbidity, suicide rate, and overall mortality for patients treated for bipolar disorder and schizophrenia. This risk is important information, but it does not follow that mood stabilizing treatment or antipsychotic treatment is the culprit.” In addition, the study also looked at the cohort divided into 2 groups each spanning 15 years. When the second (later) cohort was compared to control Swedes, suicide rates and overall mortality were not statistically significantly different, however rates of inpatient hospitalization for psychiatric concerns was still increased. As Dhejne et al stated, any cohort of patients treated for a serious mental health diagnosis would be expected to have greater morbidity and mortality when compared to a non-clinical control sample. The fact that in the latter half of the study suicidality and overall morbidity did not differ to a statistically significant extent from a non-clinical control sample demonstrates improved results for overall treatment, although it is impossible to ascertain which parts of treatment had the greatest effect, because of its clinical importance, its persistence over the interval of data collection and the increase in risk over time for the individual.

1. Patient Care

Additional questions regarding the care of patients with gender dysphoria remain. The specific type(s) of gender/sex reassignment surgery (genital, non-genital) that could improve health outcomes in adults remain(s) uncertain because most studies included patients who had undertaken one or more of a spectrum of surgical procedures or did not define the specific surgical procedures under study. Furthermore, most studies did not assess specific surgical procedures except for technical aspects. Surgical techniques have changed significantly over the last 60 years, with diminished but significant complication rates for certain procedures. (Bjerrome Ahlin et al., 2014; Doornaert, 2011; Green, 1998; Pauly, 1968; Selvaggi et al., 2007; Selvaggi, Bellringer, 2011; Tugnet et al., 2007; Doornaert, 2011).
The WPATH care recommendations presented a general framework and guidance on the care of transgender individual. The standards of care are often cited by entities that perform gender reassignment surgery. WPATH noted: “More studies are needed that focus on the outcomes of current assessment and treatment approaches for gender dysphoria.” Appendix D in the WPATH Standards of Care acknowledged the historical problems with evidentiory standards, the preponderance of retrospective data, and the confounding impact of multiple interventions, specifically distinguishing the impact of hormone therapy from surgical intervention. However in Appendix D WPATH also states “The vast majority of follow-up studies have shown an undeniable beneficial effect of sex reassignment surgery on postoperative outcomes such as subjective well being, cosmetics, and sexual function, although the specific magnitude of benefit is uncertain from the currently available evidence”, and concluded that “Overall, studies have been reporting a steady improvement in outcomes as the field becomes more advanced. Outcome research has mainly focused on the outcome of sex reassignment surgery. In current practice there is a range of identity, role, and physical adaptations that could use additional follow-up or outcome research.”

The surgical expertise and care setting(s) required to improve health outcomes in adults with gender dysphoria remain(s) uncertain. The selection of a particular surgeon could become an important variable if subjective outcomes depend on functional surgical results (Ross 1989). Many of these procedures involve complicated gynecologic, urologic surgical techniques accompanied by significant risk (Goddard et al., 2007a; Kuhn et al., 2011; Lawrence, 2003; Leclere et al., 2015; Rachlin, 1999; Ruppin, Pfafflin, 2015). Most of the studies for reassignment surgery have been conducted in northern Europe at select centers with integrated care (psychological, psychiatric, primary care, endocrinologic, and surgical). In the U.S. such centers are only now being developed and are mainly bicoastal. For example, the Kaiser Gender Pathways Clinic in San Francisco, the GeMS clinic at Boston Children’s Hospital, and the University of California, San Francisco Center of Excellence in Transgender Care have over the past 5 years developed integrated models similar to larger European centers.16,17,18 Unfortunately, such integrated care models are not available to the majority of transgender Americans because of geographic limitations. Many transgender patients receive hormonal treatment in primary care settings and travel long distances to receive surgical care from the small number of competent surgeons for some surgeries (although others are commonly done by general surgical specialists such as hysterectomy by gynecologists, or chest surgery by plastic surgeons already competent in such procedures for non-transgender patients) endocrinologic, and surgical in which there is sequential evaluation of patients for progressively more invasive interventions.

Additionally, CMS met with several stakeholders and conducted several interviews with centers that focus on healthcare for transgender individuals in the U.S. Primary care was often the centers’ main focus rather than gender reassignment surgery. Few of the U.S. based

18
reassignment surgeons we could identify work as part of an integrated practice, and few provide the most complex procedures.

2. Generalizability

With the variability in the study participants, providers and settings, the generalizability of the studies reviewed to the Medicare population is unclear. Many of the studies are old since they were conducted more than 10 years ago. Many of the programs were single-site centers without replication elsewhere. Most of these studies were conducted outside of the U.S. with far different medical systems for treatment and follow-up. The study populations were young and without significant physical or additional psychiatric co-morbidity. As noted above psychiatric co-morbidity may portend poor outcomes (Asscheman et al., 2011; Landen et al., 1998).

For the above reasons, it is difficult to generalize these studies to the Medicare population.

3. Knowledge Gaps

This patient population faces complex and unique challenges. The medical science in this area is evolving. There are, however, many gaps in the evidentiary base. These gaps have been delineated because they represent areas in which patient care can be optimized and which are opportunities for much needed research.

The Institute of Medicine, the National Institutes of Health, and others have delineated many of the gaps in the data. (Boehmer, 2002; HHS-HP, 2011; IOM, 2011; Kreukels-ENIGI, 2012; Lancet, 2011; Murad et al., 2010; NIH-LGBT, 2013) The current or completed studies listed in ClinicalTrials.gov are not structured to assess these gaps.

The currently available evidence has limitations:

- There were design deficiencies. All but one of the studies were observational in nature. All but two were non-blinded. The accompanying loss to follow-up suggests that there is selection bias and that the population that seeks treatment for gender dysphoria is not the same population that undergoes reassignment surgery without hesitation or regret.
- The psychometric and psychosocial function endpoints are limited by lack of well validated measures of dysphoria, however two measures that have been developed are supported by some validation research and further studies are ongoing: Utrecht Gender Dysphoria Scale (UGDS) and the Gender Identity/Gender Dysphoria Questionnaire for Adolescents and Adults (GIDYQ-AA)19.
- There were limitations of the psychosocial endpoints and of the data collection of other hard functional outcomes. Evidence on mortality and especially suicide was stronger. There is evidence of improved overall mortality and diminished suicide rates in European centers over time, but these outcomes and psychiatric hospitalization rates even after treatment setting in coordinated care highly structured programs are of concern.

---

Ethical limitations preclude the ideal trial (a randomized controlled trial at least single blinding of evaluators) and low incidence rates of gender dysphoria (a rare disease by U.S. definition) limit the functional size of trials. In addition, the preponderance of European research and the greater development of integrated centers in Europe likely is the result of such care being covered under public and private insurance in Europe for many decades while similar care in the U.S. was almost universally excluded by public and private insurance until the last decade.

There are insufficient data to select optimal candidates for surgery. Therefore, this should be done on a case by case basis.

The results were inconsistent, but negative in the best studies, i.e., those that reduced confounding by testing patients prior to and after surgery and which used psychometric tests with some established validation in other large populations. (Atkins et al., 2004; Balshem et al., 2011; Chan, Altman, 2005; Deeks et al., 2003; Guyatt et al., 2008a-e; 2011a-e; Kunz, Oxman, 1998; Kunz et al., 2007 and 2011; Odgaard-Jensen et al., 2011).

Data on reassignment surgery performed on geriatric patients or follow-up data in geriatric patients who had reassignment surgery in the distant past is anecdotal (Orel, 2014). Moreover, there is a large population of transgender elders in the U.S. who were unable to access surgery in their youth because of insurance exclusions. Therefore data about surgery in elderly patients is generally rare because in Europe such patients generally receive care in their youth.

C. Health Disparities

Four studies included information on racial or ethnic background. The participants in the 3 U.S. based studies were predominantly Caucasian (Beatrice, 1985; Meyer, Reter, 1979; Newfield et al., 2006). All of the participants in the single Asian study were Chinese (Tsoi, 1993). Additional research is needed in this area. In the U.S. a 2010 survey of transgender Americans demonstrated significant ethnic and racial disparities in not only access to care, but in rates of discrimination, unemployment, homelessness, HIV, sexual assault, and suicide. In particular, 49% of black transgender respondents had attempted suicide, 34% had incomes of less than $10,000 annually, 20% were HIV positive, and 49%, 27%, and 15% reported harassment, physical assault, and sexual assault respectively in K-12 education. Transgender people of color are at particularly high risk for negative health outcomes and often have even less access to transgender healthcare. The compounding effect of membership in two or more minority populations cannot be underestimated and future efforts to support improved health outcomes for this subgroup should be championed.

D. Summary

Based on a thorough review of the clinical evidence available at this time, there is not enough evidence to determine whether gender reassignment surgery as a whole improves health outcomes for elderly Medicare beneficiaries with gender dysphoria. Study results were not always comparable, and while overall benefits seem to exceed harms, the available literature does not allow for firm conclusions about sex reassignment surgery as a single group of treatments. That is, sex reassignment surgery is not a single procedure. It is a group of disparate surgeries designed to treat a diagnosis with a range of presentations and severity. Some of these surgeries have significant risks for complications while others (such as chest surgeries and facial feminization surgeries) have fewer and generally more minor complications. Taking these procedures as a whole results in the combination of some procedures that likely have more benefits and fewer risks with procedures that may have the opposite. As an analogy, it would be similar to performing an analysis of all medical treatments for type 2 diabetes. Such an analysis would conflate the greater evidence for benefit of using metformin with the lesser evidence for use of thiazolidinediones. One would be able to conclude that overall treatment of type 2 diabetes is beneficial, but the relative contributions of different treatments would be difficult to ascertain. In clinical practice patients may need zero to multiple surgeries to address their gender dysphoria. Studies have reported outcomes on patients with multiple or single surgeries. While the available literature supports the overall benefit of surgery as part of comprehensive treatment for gender dysphoria in the general transgender population, it does not provide adequate information to determine whether specific surgeries benefit specific sub-populations of the transgender community (such as elderly patients likely to receive Medicare benefits), or improves health outcomes for Medicare beneficiaries with gender dysphoria. There were conflicting (inconsistent) study results — of the best designed studies, some reported benefits while others reported harms. The quality and strength of evidence were low due to the mostly observational study designs with no comparison groups, potential confounding and small sample sizes, however this must be interpreted with the limitations of studying a rare disease and ethical limitations of research on human subjects. Many studies that reported positive outcomes were exploratory type studies (case-series and case-control) with no confirmatory follow-up. Due in part to the generally younger and healthier study participants, the generalizability of the studies to the Medicare population is also unclear. Additional research is needed. WPATH also noted the need for further research while recognizing that the existing research does support the benefit of sex reassignment surgery in many patients: “The vast majority of follow-up studies have shown an undeniable beneficial effect of sex reassignment surgery on postoperative outcomes such as subjective well being, cosmesis, and sexual function (De Cuypere et al., 2005; Garaffa, Christopher, & Ralph, 2010; Klein & Gorzalka, 2009), although the specific magnitude of benefit is uncertain from the currently available evidence.” This proposed conclusion is consistent with the West Midlands Health Technology Assessment Collaboration (2009) that reported “Further research is needed but must use more sophisticated designs with comparison groups.” WPATH also noted the need for further research: “More studies are needed that focus on the outcomes of current assessment and treatment approaches for gender dysphoria.” Further, as mentioned earlier, patient preference is an important aspect of any treatment. With that in mind, CMS is interested in knowing from the patients with gender dysphoria what is important to them as a result of a successful gender reassignment surgery.
Knowledge on gender reassignment surgery for individuals with gender dysphoria is rapidly evolving. The specific role for various surgical procedures is less well understood than the role of hormonal intervention. Much of the available research has been conducted in highly vetted patients at select care programs integrating psychotherapy, endocrinology, and various surgical disciplines and operating under European medical management and regulatory structures. Standard psychometric tools currently used need to be refined and developed and tested in the patients with gender dysphoria to validly assess long term outcomes. As such, further evidence as demonstrated via efficacy studies as well as effectiveness studies in this area would contribute to the question of whether gender reassignment surgeries in what specific sub-populations improve health outcomes in adults with gender dysphoria.

Because CMS is mindful of the unique and complex needs of this patient population and because CMS seeks sound data to guide proper care of the Medicare subset of this patient population, CMS strongly encourages robust clinical studies with adequate patient protections that will fill the evidence gaps delineated in this decision memorandum. As the Institute of Medicine (IOM, 2011) importantly noted: “Best practices for research on the health status of LGBT populations include scientific rigor and respectful involvement of individuals who represent the target population. Scientific rigor includes incorporating and monitoring culturally competent study designs, such as the use of appropriate measures to identify participants and implementation processes adapted to the unique characteristics of the target population. Respectful involvement refers to the involvement of LGBT individuals and those who represent the larger LGBT community in the research process, from design through data collection to dissemination.”

IX. Proposed Decision

Currently, the local Medicare Administrative Contractors (MACs) determine coverage of gender reassignment surgery on an individual claim basis. The Centers for Medicare & Medicaid Services (CMS) proposes to continue this practice and not issue a National Coverage Determination (NCD) at this time on gender reassignment surgery for Medicare beneficiaries with gender dysphoria.

Our review of the clinical evidence for gender reassignment surgery was inconclusive for the specific Medicare population largely due to the at large. The low number of clinical studies about gender reassignment surgery in elderly patients in general and specifically about Medicare beneficiaries’ health outcomes for gender reassignment surgery and small sample sizes inhibited our ability to create clinical appropriateness criteria for cohorts of Medicare beneficiaries. In addition, taking these disparate surgeries as a whole and in context of comprehensive treatment of transgender patients which is necessitated by ethical and epidemiological realities makes it difficult to determine which of these very different surgeries might be of a differential benefit to the Medicare population.

This should not be taken to indicate that there is insufficient evidence to support gender reassignment surgery in the transgender patient population as a whole or in other patient sub-populations. Indeed, the increased coverage of treatment for transgender Americans under
insurance in the last decade is supported by the American Medical Association\textsuperscript{21} and has resulted in the development of American centers as described above and will hopefully spur continued research to better promote evidence based medical decision making and higher quality care for transgender Americans.

Our conclusion that an NCD is not warranted does not dispute the medical necessity of transition-related care on a case-by-case basis in accordance with accepted standards of care. In the absence of a NCD, initial coverage determinations under section 1862(a)(1)(A) of the Social Security Act (the Act) and any other relevant statutory requirements will be made by the local Medicare Administrative Contractors (MACs) on an individual claim basis, pursuant to Department of Health and Human Services Departmental Appeals Board Docket No. A-13-87, Decision No. 2576.

While we are not issuing a NCD, CMS supports and encourages the relatively nascent efforts of U.S.-based researchers to provide evidence-based clinical guidance on how to best advocate and provide improved health outcomes for the transgender community. This includes our vigorous support of federally-funded quality research for this cohort. In the absence of a NCD, initial coverage determinations under section 1862(a)(1)(A) of the Social Security Act (the Act) and any other relevant statutory requirements will be made by the local Medicare Administrative Contractors (MACs) on an individual claim basis.

While we are not issuing a NCD, CMS encourages robust clinical studies that will fill the evidence gaps and help inform the answer to the question posed in this proposed decision memorandum. Based on the gaps identified in the clinical evidence, these studies should focus on which patients are most likely to achieve improved health outcomes with gender reassignment surgery, which types of surgery are most appropriate, and what types of physician criteria and care setting(s) are needed to ensure that patients achieve improved health outcomes.

We are requesting public comments on this proposed decision memorandum pursuant to section 1862(l)(3)(a) of the Act. We are specifically interested in public comments on the evidence we cited in this decision, comments containing any new evidence that has not been considered, and comments on whether a study could be developed that would support coverage with evidence development (CED), which would only cover gender reassignment surgery for beneficiaries who choose to participate in a clinical study.

X. Appendices

A. Appendix A

Diagnostic & Statistical Manual of Mental Disorders (DSM) Criteria for Disorders of Gender Identity since 1980

<table>
<thead>
<tr>
<th>DSM Version</th>
<th>Condition</th>
<th>Criteria</th>
<th>Criteria</th>
<th>Comments</th>
</tr>
</thead>
</table>

\textsuperscript{21} H.185.950 Removing Financial Barriers to Care for Transgender Patients. Our AMA supports public and private health insurance coverage for treatment of gender identity disorder as recommended by the patient’s physician. (Res. 122; A-08)
<table>
<thead>
<tr>
<th>Name</th>
<th>Required A (cross-gender identification) and B (aversion to one’s natal gender) criteria</th>
<th>Sense of discomfort and inappropriateness about one’s anatomic sex. Wish to be rid of one’s own genitals and to live as a member of the other sex. The disturbance has been continuous (not limited to periods of stress) for at least 2 years.</th>
<th>Further characterization by sexual orientation Distinguished from Atypical Gender Identity Disorder 302.85</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSM III 1980</td>
<td>Transsexualism 302.5x [Gender Identity Disorder of Childhood (302.6)]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chapter: Psychosexual Disorders</td>
<td>Sense of discomfort and inappropriateness about one’s anatomic sex. Wish to be rid of one’s own genitals and to live as a member of the other sex. The disturbance has been continuous (not limited to periods of stress) for at least 2 years.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DSM III-Revised 1987</td>
<td>Transsexualism (TS) (302.50) [GID of C]</td>
<td>Persistent discomfort and sense of inappropriateness about one’s assigned sex. Persistent preoccupation for at least 2 years with getting rid of one’s 1st and 2nd sex characteristics and acquiring the sex characteristics of the other sex. Has reached puberty</td>
<td>Further characterization by sexual orientation Distinguished from Gender Identity Disorder of Adolescence or Adulthood, Non-trans-sexual Type • e.g., cross-dressing not for the purposes of sexual excitement Gender Identity Disorder Not Otherwise Specified 302.6 • e.g., intersex conditions Gender Identity Disorder Not Otherwise Specified 302.85 • e.g., persistent preoccupation with castration or penectomy w/o desire to acquire the sex traits of the other sex</td>
</tr>
</tbody>
</table>

| GID of                                   |                                                                                     |                                                                                     |                                                                                                         |
| DSM IV 1994 | Gender Identity Disorder in Adolescents and Adults (302.85) (Separate criteria & code for children, but same name) | Required A and B criteria | Cross-gender identification:
- e.g., Stated desire to be another sex
- e.g., Desire to live or be treated as a member of the other sex
- e.g., conviction that he/she has the typical feelings and reactions of the other sex
- e.g., frequent passing as the other sex
- Persistent discomfort with his/her sex or sense of inappropriateness in the gender role of that sex.
- e.g., belief the he/she was born the wrong sex
- e.g., preoccupation with getting rid of 1st and 2nd sex characteristics &/or acquiring sexual traits of the other sex
- Clinically significant distress or impairment in social, occupational, or other important areas of functioning
| Further characterization by sexual orientation
Distinguished from Gender Identity Disorder Not Otherwise Specified 302.6
- e.g., intersex conditions
- e.g., stress related cross-dressing
- e.g., persistent preoccupation with castration or penectomy w/o desire to acquire the sex traits of the other sex |

| DSM IV-Revised 2000 | Gender Identity Disorder (Term transsexual-ism eliminated) | Required A & B criteria | Cross-gender identification:
- e.g., stated desire to be the other sex
- e.g., desire to live or be treated as the other sex
- e.g., conviction that he/she has the typical feelings & reactions of the other sex |
| Outcome may depend on time of onset
Further characterization by sexual orientation
Distinguished from Gender Identity Disorder Not Otherwise Specified 302.6 |
<table>
<thead>
<tr>
<th>DSM V 2013 Separate Chapter from Sexual Dysfunctions &amp; Paraphilic Disorders</th>
<th>Gender Dysphoria (302.85)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender nonconformity itself not considered to be a mental disorder</td>
<td>Marked discordance between natal 1° and 2° sex characteristics* and experienced/expressed gender</td>
<td>Includes diagnosis for post transition state to permit continued treatment access</td>
</tr>
<tr>
<td>The dysphoria associated with the gender incongruence is</td>
<td>Conviction that he/she has the typical feelings &amp; reactions of the other sex (or some alternative gender)</td>
<td>Includes disorders of sexual development such as congenital hyperplasia and androgen insensitivity syndromes</td>
</tr>
<tr>
<td>Eliminates A &amp; B criteria</td>
<td>Marked desire to be the other sex (or some alternative gender)</td>
<td></td>
</tr>
<tr>
<td>Considers gender incongruence to be a spectrum</td>
<td>Marked desire to desire be treated as the other sex (or some alternative gender)</td>
<td></td>
</tr>
<tr>
<td>Considers intersex/“disorders of sex development” to be a subsidiary and not exclusionary to dx of GD</td>
<td>Marked desire to be rid of natal 1° and 2° sex characteristics**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Marked desire to acquire 1° and 2° sex characteristics of the other sex (or some alternative gender)</td>
<td></td>
</tr>
<tr>
<td>Clinical significance</td>
<td>Clinically significant distress or impairment in social, occupational, or other important areas of functioning</td>
<td></td>
</tr>
<tr>
<td>Includes diagnosis for post transition state to permit continued treatment access</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Includes transsexual disorders such as gender identity disorders and gender dysphoria

**Includes gender identity disorder not otherwise specified (GID-NOS) and gender identity disorder due to medical condition (GID-MC)
distress or impairment in social, occupational, or other important areas of functioning

* or in young adolescents, the anticipated 2\textsuperscript{o} sex characteristics

** or in young adolescents, prevent the development of the anticipated 2\textsuperscript{o} sex characteristics

≥ 6 month marked discordance between natal gender & experienced/expressed gender as demonstrated by ≥ 6 criteria:

- Strong desire to be of the other gender or an insistence that one is of another gender.
- Strong preference for cross-gender roles in make-believe play.
- Strong preference for the toys, games, or activities of the other gender.
- Strong preference for playmates of the other gender.
- In boys, strong preference for cross-dressing; in girls, strong preference for wearing masculine clothing
- In boys, rejection of masculine toys, games, activities, avoidance of rough and tumble play; in girls, rejection of feminine toys, games,
This category applies to presentations in which sx c/w gender dysphoria that cause clinically significant distress or impairment, but do not meet the full criteria for gender dysphoria & the reason for not meeting the criteria is not provided.

If the reason that the presentation does not meet the full criteria is provided then this dx should be used.

C/W=consistent with Dx=diagnosis GD=gender dysphoria Sx=symptoms TS=transsexual

1°=primary 2°=secondary

B. Appendix B

1. General Methodological Principles of Study Design

When making national coverage determinations, CMS evaluates relevant clinical evidence to determine whether or not the evidence is of sufficient quality to support a finding that an item or service is reasonable and necessary. The overall objective for the critical appraisal of the evidence is to determine to what degree we are confident that: 1) the specific assessment questions can be answered conclusively; and 2) the intervention will improve health outcomes for patients.

We divide the assessment of clinical evidence into three stages: 1) the quality of the individual studies; 2) the generalizability of findings from individual studies to the Medicare population; and 3) overarching conclusions that can be drawn from the body of the evidence on the direction and magnitude of the intervention’s potential risks and benefits.

The methodological principles described below represent a broad discussion of the issues we consider when reviewing clinical evidence. However, it should be noted that each coverage determination has its unique methodological aspects.

Assessing Individual Studies

Methodologists have developed criteria to determine weaknesses and strengths of clinical research. Strength of evidence generally refers to: 1) the scientific validity underlying study findings regarding causal relationships between health care interventions and health outcomes;
and 2) the reduction of bias. In general, some of the methodological attributes associated with stronger evidence include those listed below:

- Use of randomization (allocation of patients to either intervention or control group) in order to minimize bias.
- Use of contemporaneous control groups (rather than historical controls) in order to ensure comparability between the intervention and control groups.
- Prospective (rather than retrospective) studies to ensure a more thorough and systematic assessment of factors related to outcomes.
- Larger sample sizes in studies to demonstrate both statistically significant as well as clinically significant outcomes that can be extrapolated to the Medicare population. Sample size should be large enough to make chance an unlikely explanation for what was found.
- Masking (blinding) to ensure patients and investigators do not know to which group patients were assigned (intervention or control). This is important especially in subjective outcomes, such as pain or quality of life, where enthusiasm and psychological factors may lead to an improved perceived outcome by either the patient or assessor.

Regardless of whether the design of a study is a randomized controlled trial, a non-randomized controlled trial, a cohort study or a case-control study, the primary criterion for methodological strength or quality is the extent to which differences between intervention and control groups can be attributed to the intervention studied. This is known as internal validity. Various types of bias can undermine internal validity. These include:

- Different characteristics between patients participating and those theoretically eligible for study but not participating (selection bias).
- Co-interventions or provision of care apart from the intervention under evaluation (performance bias).
- Differential assessment of outcome (detection bias).
- Occurrence and reporting of patients who do not complete the study (attrition bias).

In principle, rankings of research design have been based on the ability of each study design category to minimize these biases. A randomized controlled trial minimizes systematic bias (in theory) by selecting a sample of participants from a particular population and allocating them randomly to the intervention and control groups. Thus, in general, randomized controlled studies have been typically assigned the greatest strength, followed by non-randomized clinical trials and controlled observational studies. The design, conduct and analysis of trials are important factors as well. For example, a well-designed and conducted observational study with a large sample size may provide stronger evidence than a poorly designed and conducted randomized controlled trial with a small sample size. The following is a representative list of study designs (some of which have alternative names) ranked from most to least methodologically rigorous in their potential ability to minimize systematic bias:

Randomized controlled trials
Non-randomized controlled trials
Prospective cohort studies
Retrospective case control studies
Cross-sectional studies
Surveillance studies (e.g., using registries or surveys)
Consecutive case series
Single case reports

When there are merely associations but not causal relationships between a study’s variables and outcomes, it is important not to draw causal inferences. Confounding refers to independent variables that systematically vary with the causal variable. This distorts measurement of the outcome of interest because its effect size is mixed with the effects of other extraneous factors. For observational, and in some cases randomized controlled trials, the method in which confounding factors are handled (either through stratification or appropriate statistical modeling) are of particular concern. For example, in order to interpret and generalize conclusions to our population of Medicare patients, it may be necessary for studies to match or stratify their intervention and control groups by patient age or co-morbidities.

Methodological strength is, therefore, a multidimensional concept that relates to the design, implementation and analysis of a clinical study. In addition, thorough documentation of the conduct of the research, particularly study selection criteria, rate of attrition and process for data collection, is essential for CMS to adequately assess and consider the evidence.

**Generalizability of Clinical Evidence to the Medicare Population**

The applicability of the results of a study to other populations, settings, treatment regimens and outcomes assessed is known as external validity. Even well-designed and well-conducted trials may not supply the evidence needed if the results of a study are not applicable to the Medicare population. Evidence that provides accurate information about a population or setting not well represented in the Medicare program would be considered but would suffer from limited generalizability.

The extent to which the results of a trial are applicable to other circumstances is often a matter of judgment that depends on specific study characteristics, primarily the patient population studied (age, sex, severity of disease and presence of co-morbidities) and the care setting (primary to tertiary level of care, as well as the experience and specialization of the care provider). Additional relevant variables are treatment regimens (dosage, timing and route of administration), co-interventions or concomitant therapies, and type of outcome and length of follow-up.

The level of care and the experience of the providers in the study are other crucial elements in assessing a study’s external validity. Trial participants in an academic medical center may receive more or different attention than is typically available in non-tertiary settings. For example, an investigator’s lengthy and detailed explanations of the potential benefits of the intervention and/or the use of new equipment provided to the academic center by the study sponsor may raise doubts about the applicability of study findings to community practice.
Given the evidence available in the research literature, some degree of generalization about an intervention’s potential benefits and harms is invariably required in making coverage determinations for the Medicare population. Conditions that assist us in making reasonable generalizations are biologic plausibility, similarities between the populations studied and Medicare patients (age, sex, ethnicity and clinical presentation) and similarities of the intervention studied to those that would be routinely available in community practice.

A study’s selected outcomes are an important consideration in generalizing available clinical evidence to Medicare coverage determinations. One of the goals of our determination process is to assess health outcomes. These outcomes include resultant risks and benefits such as increased or decreased morbidity and mortality. In order to make this determination, it is often necessary to evaluate whether the strength of the evidence is adequate to draw conclusions about the direction and magnitude of each individual outcome relevant to the intervention under study. In addition, it is important that an intervention’s benefits are clinically significant and durable, rather than marginal or short-lived. Generally, an intervention is not reasonable and necessary if its risks outweigh its benefits.

If key health outcomes have not been studied or the direction of clinical effect is inconclusive, we may also evaluate the strength and adequacy of indirect evidence linking intermediate or surrogate outcomes to our outcomes of interest.

**Assessing the Relative Magnitude of Risks and Benefits**

Generally, an intervention is not reasonable and necessary if its risks outweigh its benefits. Health outcomes are one of several considerations in determining whether an item or service is reasonable and necessary. CMS places greater emphasis on health outcomes actually experienced by patients, such as quality of life, functional status, duration of disability, morbidity and mortality, and less emphasis on outcomes that patients do not directly experience, such as intermediate outcomes, surrogate outcomes, and laboratory or radiographic responses. The direction, magnitude, and consistency of the risks and benefits across studies are also important considerations. Based on the analysis of the strength of the evidence, CMS assesses the relative magnitude of an intervention or technology’s benefits and risk of harm to Medicare beneficiaries.

**Appendix C**

**Patient Population: Enrolled & Treated with Sex Reassignment Surgery Loss of Patients & Missing Data**

**Panel A (Controlled Studies)**

<table>
<thead>
<tr>
<th>Author</th>
<th>Study Type</th>
<th>Recruitment Pool</th>
<th>Enrolled</th>
<th>% GRS</th>
<th>Completion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dhejne 2011</td>
<td>Longitudinal Controlled</td>
<td>480 w GID who did not apply or were not approved for SRS were excluded</td>
<td>324</td>
<td>324 (100%)</td>
<td>-</td>
</tr>
<tr>
<td>Study</td>
<td>Type</td>
<td>Comparison</td>
<td>Application Status</td>
<td>Surgical Reversal</td>
<td>Psychosocial Survey</td>
</tr>
<tr>
<td>--------------------</td>
<td>--------------------</td>
<td>------------</td>
<td>--------------------</td>
<td>-------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>Dhejne 2014 Landen</td>
<td>Longitudinal</td>
<td>Controlled</td>
<td>767 applied for SRS 25 applications denied. 61 not granted full legal status 15 formal applications for surgical reversal</td>
<td>681 (100%)</td>
<td>NA: Clinical data extracted retrospectively in earlier paper</td>
</tr>
<tr>
<td>Heylens</td>
<td>Longitudinal</td>
<td>Controlled</td>
<td>90 applicants for SRS 33 excluded 11 later excluded had not yet received SRS by study close.</td>
<td>57 (46)</td>
<td>46 (80.7%) Only those w SRS evaluated</td>
</tr>
<tr>
<td>Kockott</td>
<td>Longitudinal</td>
<td>Controlled</td>
<td>80 applicants for SRS 21 excluded</td>
<td>59</td>
<td>32 (54.2%) went to surgery</td>
</tr>
<tr>
<td>Mate-Kole 1990</td>
<td>Longitudinal</td>
<td>Controlled</td>
<td>40 sequential patients of accepted patients.</td>
<td>40</td>
<td>20 (50%) went to surgery</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Study</th>
<th>Design</th>
<th>Recruitment pool: 100</th>
<th>Excluded</th>
<th>Underwent surgery</th>
<th>Lost post SRS</th>
<th>Questionnaire completed</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meyer</td>
<td>Longitudinal</td>
<td>50</td>
<td>50 (30%)</td>
<td>14 (28%)</td>
<td></td>
<td></td>
<td>The assessments of all were complete</td>
</tr>
<tr>
<td>Rakic</td>
<td>Longitudinal</td>
<td>32</td>
<td>32 (100%)</td>
<td></td>
<td></td>
<td></td>
<td>Questionnaire completed by all.</td>
</tr>
<tr>
<td>Ruppin</td>
<td>Longitudinal</td>
<td>71</td>
<td>69 (97.2%)</td>
<td></td>
<td></td>
<td></td>
<td>The SCL-90, BSRI, FPI-R, &amp; IPP tests were not completed by 9, 34, 13, &amp; 16 respectively. Questions about romantic relationships, sexual relationships, friendships, &amp; family relationships were not answered by 1, 3, 2, &amp; 23 respectively. Questions regarding gender security &amp; regret were not answered by 1 &amp; 2 respectively.</td>
</tr>
</tbody>
</table>
| Smith   | Longitudinal    | 162                   | 162 (100%)|                   |               |                         | 36 to 61 (22.2%--)}
<table>
<thead>
<tr>
<th>Study</th>
<th>Design</th>
<th>Description</th>
<th>N (Pre-SRS)</th>
<th>N (Post-SRS)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Udeze Megeri</td>
<td>Longitudinal Controlled</td>
<td>International patient w GD 546 &amp; post SRS 318. 40 M to F subjects were prospectively selected.</td>
<td>40</td>
<td>40 (100%)</td>
<td>-</td>
</tr>
<tr>
<td>Ainsworth</td>
<td>Internet/convention Survey Cross-sectional Controlled</td>
<td>Number of incomplete questionnaires not reported</td>
<td>247</td>
<td>72 (29.1%) 75 (30.6%) facial 147 (59.5%) had received neither facial nor reassignment surgery</td>
<td>-</td>
</tr>
<tr>
<td>Beatrice</td>
<td>Cross-sectional Controlled</td>
<td>14 excluded for demographic matching reasons</td>
<td>40</td>
<td>10 (25%)</td>
<td>The assessments were completed by all</td>
</tr>
<tr>
<td>Haraldsen</td>
<td>Cross-sectional Controlled</td>
<td>Recruitment pool: 99</td>
<td>86</td>
<td>59 (68.6%)</td>
<td>-</td>
</tr>
<tr>
<td>Kraemer</td>
<td>Cross-sectional Controlled</td>
<td>The number in the available patient pool was not specified.</td>
<td>45</td>
<td>22 (48.9%)</td>
<td>-</td>
</tr>
<tr>
<td>Kuhn</td>
<td>Cross-sectional Controlled</td>
<td>The number in the available patient pool was not specified.</td>
<td>75</td>
<td>55 (73.3%)</td>
<td>-</td>
</tr>
<tr>
<td>Mate-Kole 1988</td>
<td>Cross-sectional Controlled</td>
<td>150 in 3 cohorts. Matched on select traits. The number in the available patient pool was not specified.</td>
<td>150</td>
<td>50 (66.7%)</td>
<td>-</td>
</tr>
</tbody>
</table>
Panel B (Surgical Series: No Concurrent Controls)

<table>
<thead>
<tr>
<th>Author</th>
<th>Study Type</th>
<th>Recruitment Pool</th>
<th>Enrolled</th>
<th>% GRS</th>
<th>Completion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blanchard et al.</td>
<td>Cross-sectional Control: Normative test data</td>
<td>294 clinic patients w GD had completed study questionnaire 116 authorized for GRS. 103 completed GRS &amp; 1yr post-operative. 24 excluded</td>
<td>79</td>
<td>79(100%)</td>
<td>-</td>
</tr>
<tr>
<td>Weyers et al.</td>
<td>Cross-sectional Control: Normative test data</td>
<td>&gt;300 M to F patients had undergone GRS 70 eligible patients recruited 20 excluded</td>
<td>50</td>
<td>50 (100%)</td>
<td>SF-26 not completed by 1</td>
</tr>
<tr>
<td>Wierckx et al.</td>
<td>Cross-sectional except for recall questions Control: Normative test data</td>
<td>79 F to M patients had undergone GRS &amp; were recruited. 3 additional non-clinic patients were recruited by other patients. 32 excluded initially; 1 later.</td>
<td>49</td>
<td>49 (100%)</td>
<td>SF-36 test not completed by 2. Questions regarding sexual relationship, sex function, &amp; surgical satisfaction were answered by as few as 27, 28, 32 respectively.</td>
</tr>
<tr>
<td>Eldh et al.</td>
<td>Cross-sectional except for 1 variable Control: Self for 1 variable: employment</td>
<td>136 were identified. 46 excluded</td>
<td>90</td>
<td>90 (100%)</td>
<td>Questions regarding gender identity, sex life, acceptance, &amp; overall satisfaction were not answered by 13, 14, 14 &amp; 16 respectively. Employment data missing for 11.</td>
</tr>
<tr>
<td>Hess et al.</td>
<td>Cross-sectional</td>
<td>254 consecutive eligible patients post GRS</td>
<td>119</td>
<td>119 (100%)</td>
<td>Questions regarding the...</td>
</tr>
</tbody>
</table>
Panel C (Mixed Treatment Series: No Direct Control Groups)

<table>
<thead>
<tr>
<th>Author</th>
<th>Study Type</th>
<th>Recruitment Pool</th>
<th>Enrolled</th>
<th>% GRS</th>
<th>Completion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gómez-Gil et al. 2012</td>
<td>Cross-sectional No direct control: Analysis of variance</td>
<td>200 consecutive patients were recruited. 13 declined participation or were excluded for incomplete questionnaires.</td>
<td>187</td>
<td>79 (42.2%)</td>
<td>See prior box.</td>
</tr>
<tr>
<td>Hepp et al.</td>
<td>Cross-sectional No direct control: Analysis of variance</td>
<td>The number in the available patient pool was not specified.</td>
<td>31</td>
<td>7 (22.6%)</td>
<td>HADS test not completed by 1</td>
</tr>
</tbody>
</table>
| Motmans et al. | Cross-sectional No direct control: Analysis of variance & regression | 255 with GD were identified. 77 were excluded. | 148 (→140) | Not clearly stated. At least 103 underwent some form of GRS. | 8 later excluded for incomplete SF-36 tests. 37 w recent GRS or hormone initiation were
<table>
<thead>
<tr>
<th>Study</th>
<th>Methodology</th>
<th>Design</th>
<th>Control</th>
<th>Analysis</th>
<th>Number of respondents</th>
<th>U.S. respondents excluded</th>
<th>% excluded</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newfield et al.</td>
<td>Internet survey</td>
<td>Cross-sectional</td>
<td>No direct control</td>
<td>Analysis of variance</td>
<td>376 (U.S.)</td>
<td>139 to 150 (37.0-39.9%) in U.S.</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gomez-Gil et al. 2014</td>
<td>Cross-sectional</td>
<td>No direct control</td>
<td>Analysis with regression</td>
<td>The number in the available patient pool was not specified. 277 were recruited. 25 excluded</td>
<td>252 (→193)</td>
<td>80 (41.4%) non-genital surgery</td>
<td>59 were excluded for incomplete questionnaires. See prior box.</td>
<td></td>
</tr>
<tr>
<td>Asscherman</td>
<td>Longitudinal</td>
<td>No analysis by tx status</td>
<td></td>
<td></td>
<td>1331</td>
<td>1177 (88.4%)</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Johansson et al.</td>
<td>Cross-sectional</td>
<td>except for 1 variable</td>
<td>No analysis by tx status except for 1 question</td>
<td>60 eligible patients 18 excluded.</td>
<td>42</td>
<td>32 (76.2% of enrolled &amp; 53.3% of eligible) (genital surgery)</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Leinung et al.</td>
<td>Cross-sectional</td>
<td>No analysis by tx status</td>
<td></td>
<td></td>
<td>242 total clinic patients</td>
<td>242</td>
<td>91 (37.6%)</td>
<td>Employment status data missing for 81 of all patients</td>
</tr>
</tbody>
</table>

*Data obtained via a survey on a website and distributed at a conference
B/C=because
BSRI=Bem Sex Role Inventory
F=Female
FP-R=Freiberg Personality Inventory
GD=Gender dysphoria
GID=Gender identity disorder
**Appendix D**

**Demographic Features of Study Populations**

**Panel A (Controlled Studies)**

<table>
<thead>
<tr>
<th>Author</th>
<th>Age (years; mean, S.D., range)</th>
<th>Gender</th>
<th>Race</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ainsworth</td>
<td>Only reassignment surgery: 50 (no S.D.)&lt;br&gt;Only facial surgery: 51 (no S.D.)&lt;br&gt;Both types of surgery: 49 (no S.D.)&lt;br&gt;Neither surgery: 46 (no S.D.)</td>
<td>247 M to F</td>
<td>-</td>
</tr>
<tr>
<td>Beatrice</td>
<td>Pre-SRS M to F: 32.5 (27–42), Post-SRS: 35.1 (30–43)</td>
<td>20 M to F plus 20 M controls</td>
<td>100% Caucasian</td>
</tr>
<tr>
<td>Dehjne 2011</td>
<td>Post-SRS: all 35.1±9.7 (20–69), F to M 33.3+8.7 (20–62), M to F 36.3+ 10.1(21–69)</td>
<td>133 (41.0%) F to M, 191 (59.0%) M to F; ratio 1:1.4</td>
<td>-</td>
</tr>
<tr>
<td>Dhejne 2014 Landen</td>
<td>F to M SRS cohort: median age 27, M to F SRS cohort: median age 32, F to M applicants for reversal: median age 22, M to F applicants for reversal: median age 35</td>
<td>767 applicants for legal/surgical reassignment&lt;br&gt;289 (37.7%) F to M, 478 (62.3%) M to F; ratio 1:1.6&lt;br&gt;681 post SRS &amp; legal change&lt;br&gt;252 (37.0%) F to M, 429 (63.0%) M to F; ratio 1:1.7&lt;br&gt;15 applicants for reversal&lt;br&gt;5 (33.3%) F to M, 10 (66.7%) M to F; ratio 1:2</td>
<td>-</td>
</tr>
<tr>
<td>Haraldsen</td>
<td>Pre-SRS &amp; Post-SRS: F to M 34±9.5, F to M 33.3±10.0 Post-SRS cohort reportedly older. No direct data provided.</td>
<td>Pre &amp; Post SRS 35 (40.7%) F to M, 51 (59.3%) M to F; ratio 1:1.5</td>
<td>-</td>
</tr>
<tr>
<td>Heylens</td>
<td>-</td>
<td>11 (19.3% of 57) F to M, 46 (80.7%); ratio 1:4.2 (80.7% underwent surgery)</td>
<td>-</td>
</tr>
<tr>
<td>---------------</td>
<td>--------------------------------------</td>
<td>-----------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td>Kockott</td>
<td>31.7±10.2</td>
<td>35.5±13.1</td>
<td>3 (25%) F to M, 9 (75%) M to F; ratio 1:3</td>
</tr>
<tr>
<td>Kraemer</td>
<td>33.0±11.3, Post-SRS: 38.2±9.0</td>
<td></td>
<td>Pre-SRS 7 F to M (30.4%), 16 M to F (69.6%); ratio 1:2.3</td>
</tr>
<tr>
<td>Kuhn</td>
<td>All post SRS: median (range): 51 (39-62) (long-term follow-up)</td>
<td>3 (5.4%) F to M, 52 (94.5%) M to F; ratio 1:17.3.</td>
<td>-</td>
</tr>
<tr>
<td>Mate-Kole 1988</td>
<td>Initial evaluation: 34, Pre-SRS: 35, Post-SRS: 37</td>
<td>150 M to F</td>
<td>-</td>
</tr>
<tr>
<td>Mate-Kole 1990</td>
<td>Early &amp; Usual wait SRS: 32.5 years (21-53)</td>
<td>40 F to M</td>
<td>-</td>
</tr>
<tr>
<td>Meyer</td>
<td>Pre-SRS: 26.7 Delayed, but completed SRS: 30.9 Post-SRS: 30.1</td>
<td>Pre-SRS: 5 (23.8%) F to M, 16 (76.2%) M to F; ratio 1:3.2 Delayed, but completed SRS: 1 (7.1%) F to M, 13 (92.9%) M to F; ratio 1:1.3 Post-SRS: 4 (26.7%) F to M, 11 (73.3%) M to F; ratio 1:2.8</td>
<td>-</td>
</tr>
<tr>
<td>Rakic</td>
<td>All: 26.8±6.9 (median 25.5, range 19-47), F to M: 27.8±5.2 (median 27, range 23-37), M to F: 26.4±7.8 (median 24, range 19-47).</td>
<td>10 (31.2%) F to M, 22 (68.8%) M to F; ratio 1:2.2</td>
<td>-</td>
</tr>
<tr>
<td>Ruppin</td>
<td>All: 47.0±10.42 (but 2 w/o SRS) (13.8±2.8 yrs post legal name change) (long-term follow-up) F to M: 41.2±5.78, M to F: 52.9±10.82</td>
<td>36 (50.7%) F to M, 35 (49.3%) M to F; ratio 1:0.97</td>
<td>-</td>
</tr>
<tr>
<td>Smith</td>
<td>Time of surgical request for post-SRS: 30.9 (range 17.7-68.1) Time of follow-up for post-SRS: 35.2 (range 21.3-71.9)</td>
<td>Pre-SRS: 162; 58 (35.8%) F to M, 104 (64.2%) M to F; ratio 1:1.8 Post-SRS: 126; 49 (38.9%) F to M, 77 (61.1%) M to F; ratio 1:1.6</td>
<td>-</td>
</tr>
<tr>
<td>Udeze Megeri</td>
<td>M to F: 47.33±13.26 (range 25-80).</td>
<td>40 M to F</td>
<td>-</td>
</tr>
<tr>
<td>Wolfradt</td>
<td>Patients &amp; controls: 43 (range 29-67).</td>
<td>30 M to F plus 30 F controls plus 30 M controls.</td>
<td>-</td>
</tr>
</tbody>
</table>

*Data obtained via a survey on a website and distributed at a conference SD=Standard deviation

Panel B (Surgical Series: No Concurrent Controls)
<table>
<thead>
<tr>
<th>Author</th>
<th>Age (years; mean, S.D., range)</th>
<th>Gender</th>
<th>Caucasian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blanchard et al.</td>
<td>F to M: 32.6, M to F w M partner preference: 33.2, F to M w F partner preference: 47.7 years</td>
<td>Post-GRS: 47 (45.6%) F to M, 56 (54.4%) M to F; ratio 1:1.19. In study: 38 (48.1%) F to M, 32 (40.5%) M to F w M partner preference, 9 (11.4%) M to F w F partner preference; ratio 1:0.8: 0.2</td>
<td>-</td>
</tr>
<tr>
<td>Weyers et al.</td>
<td>Post-GRS M to F: 43.1 ±10.4 (long-term follow-up)</td>
<td>50 M to F</td>
<td>-</td>
</tr>
<tr>
<td>Wierckx et al.</td>
<td>Time of GRS: 30±8.2 years (range 16 to 49)</td>
<td>49 M to F</td>
<td>-</td>
</tr>
<tr>
<td>Eldh et al.</td>
<td>-</td>
<td>50 (55.6%) F to M, 40 (44.4%) M to F; ratio 1:0.8 There is 1 inconsistency in the text suggesting that these should be reversed.</td>
<td>-</td>
</tr>
<tr>
<td>Hess et al.</td>
<td>-</td>
<td>119 M to F</td>
<td>-</td>
</tr>
<tr>
<td>Lawrence</td>
<td>Time of GRS: 44±9 (range 18-70)</td>
<td>232 M to F</td>
<td>-</td>
</tr>
<tr>
<td>Salvador et al.</td>
<td>Time of follow-up for post-GRS: 36.28±8.94 (range 18-58) (Duration of follow-up: 3.8±1.7 [2-7])</td>
<td>52 M to F</td>
<td>-</td>
</tr>
<tr>
<td>Tsoi</td>
<td>Time of initial visit: All: 24.0±4.5, F to M: 25.4±4.4 (14-36), M to F: 22.9±4.6 (14-36). Time of GRS: All: 25.9±4.14, F to M: 27.4±4.0 (20-36), M to F: 24.7±4.3 (20-36).</td>
<td>36 (44.4%) F to M, 45 (55.6%) M to F; ratio 1:1.25 0% 100% Asian</td>
<td>-</td>
</tr>
</tbody>
</table>

**Panel C (Mixed Treatment Series: No Direct Control Groups)**

<table>
<thead>
<tr>
<th>Author</th>
<th>Age (years; mean, S.D., range)</th>
<th>Gender</th>
<th>Caucasian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gómez-Gil et al. 2012</td>
<td>W &amp; W/O GRS: All: 29.87±9.15 (range 15-61), W/O hormone tx: 25.9±7.5, W current hormone tx: 33.6±9.1. (At hormone initiation:</td>
<td>W/O hormone tx: 38 (56.7%) F to M, 29 (43.3%) M to F; ratio 1:0.8. W hormone tx: 36 (30.0%) F to M, 84 (70.0%) M to F; ratio</td>
<td>-</td>
</tr>
<tr>
<td>----------------------</td>
<td>---------------</td>
<td>----------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Hepp et al.</td>
<td>24.6±8.1</td>
<td>1:2.3, Post-GRS: 29 (36.7%) F to M, 50 (63.3%) M to F; ratio 1:1.7.</td>
<td>W &amp; W/O GRS: 11 (35.5%) F to M; 20 (64.5%) M to F; ratio 1:1.8.</td>
</tr>
<tr>
<td>Motmans et al.</td>
<td>W &amp; W/O GRS: All (n=140): 39.9±10.2, F to M: 37.0±8.5, M to F: 42.3±10.4</td>
<td>Time of hormone tx: F to M: 26.1±7.6 (16–56), M to F: 31.4±11.4 (16–76)</td>
<td>W &amp; W/O GRS: N=140 63(45.0%) F to M, 77 (55.0%) M to F; ratio 1:1.2 N=103 49 (47.6%) F to M, 54 (52.4%) M to F; ratio 1:1.1</td>
</tr>
<tr>
<td>Gomez-Gil, et al. 2014</td>
<td>W &amp; W/O Non-genital GRS: 31.2±9.9 (range 16-67).</td>
<td>Approved for GRS: 21 (35%) F to M, 39 (65%) M to F; ratio 1:1.9</td>
<td>W &amp; W/O Non-genital GRS: 74 (38.3%) F to M, 119 (61.7%) M to F; ratio1:1.6.</td>
</tr>
<tr>
<td>Asscherman</td>
<td>Time of hormone tx: F to M: 26.1±7.6 (16–56), M to F: 31.4±11.4 (16–76)</td>
<td>Approved for GRS: 21 (35%) F to M, 39 (65%) M to F; ratio 1:1.9</td>
<td>Met hormone tx requirements: 365 (27.4%) F to M, 966 (72.6%) M to F; ratio 1:2.6.</td>
</tr>
<tr>
<td>Johanssen</td>
<td>Time of initial evaluation: F to M: 27.8 (18-46), M to F 37.3 (21-60). Time of GRS: F to M: 31.4 (22-49), M to F 38.2 (22-57). Time of follow-up for post-GRS: F to M: 38.9 (28-53), M to F 46.0 (25-69) (Long-term follow-up)</td>
<td>Approved for GRS: 21 (35%) F to M, 39 (65%) M to F; ratio 1:1.9</td>
<td>Post-GRS: 14 (43.8%) F to M; 18 (56.2%) M to F; ratio 1:1.3</td>
</tr>
<tr>
<td>Leinung et al.</td>
<td>Time of hormone initiation: F to M: 27.5, M to F 35.5</td>
<td>Approved for GRS: 21 (35%) F to M, 39 (65%) M to F; ratio 1:1.9</td>
<td>W &amp; W/O GRS: 50 (20.7%) F to M, 192 M to F (79.3%); ratio 1:3.8. Post-GRS: 32 F to M (35.2%); 59 (64.8%) M to F; ratio 1:1.8.</td>
</tr>
</tbody>
</table>

**Appendix E**

**Psychometric and Satisfaction Survey Instruments**

<table>
<thead>
<tr>
<th>Instrument Name and Development and Validation</th>
<th>Instrument Name and Development and Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Developer</td>
<td>Information</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
</tr>
</tbody>
</table>
| APGAR Family Adaptability, Partner-ship Growth, Affection, and Resolve Smilkstein | Published in 1978  
Initial data: 152 families in the U.S.  
A “friends” component was added in 1983.  
Utility has challenged by many including Gardner 2001 |
| Beck Depression Inventory Beck, Ward, Mendelson, Mock, & Erbaugh | Published initially in 1961 with subsequent revisions  
It was initially evaluated in psychiatric patients in the U.S.A.  
Salkind (1969) evaluated its use in 80 general outpatients in the UK.  
It is copyrighted and requires a fee for use |
| Bem Sex Role Inventory Bem | Published 1974  
Initial data: 100 Stanford Undergraduates  
1973 update: male 444; female 279  
1978 update: 470; female 340 |
| Body Image Questionnaire Clement & Lowe | Validity study published 1996 (German)  
Population: 405 psychosomatic patients, 141 medical students, 208 sports students |
| Body Image Scale Lindgren & Pauly (Kuiper, Dutch adaptation 1991) | 1975  
Initial data: 16 male and 16 female transsexual patients in Oregon |
| Crown Crisp Experiential Index (formerly Middlesex Hospital Questionnaire) Crown & Crisp | Developed circa 1966  
Manual published 1970  
Initial data: 52 nursing students while in class in the UK |
| (2nd) European Quality of Life Survey Anderson, Mikalić, Vermeulen, Lyly-Yrjanainen, & Zigante, | Published in 2007  
The pilot survey was tested in the UK and Holland with 200 interviews. The survey was revised especially for non-response questions. Another version was tested in 25 persons of each of the 31 countries to be surveyed. Sampling methods were devised. 35,634 Europeans were ultimately surveyed. Additional updates |
<table>
<thead>
<tr>
<th>Test Description</th>
<th>Publication Details</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Female Sexual Function Index</strong></td>
<td>Published in 2000</td>
</tr>
<tr>
<td><strong>Fragebogen zur Beurteilung des eigenen Körpers</strong></td>
<td>Published 1996 (German)</td>
</tr>
<tr>
<td><em>Strauss</em></td>
<td></td>
</tr>
<tr>
<td><strong>Freiberg Personality Inventory</strong></td>
<td>7th edition published 2001, 8th edition in 2009</td>
</tr>
<tr>
<td><em>Fahrenberg, Hampel, &amp; Selg</em></td>
<td>(Not in PubMed)</td>
</tr>
<tr>
<td></td>
<td>German equivalent of MMPI</td>
</tr>
<tr>
<td><strong>“gender identity disorder in childhood”</strong></td>
<td>11 items derived from the Biographical Questionnaire for Trans-sexuals</td>
</tr>
<tr>
<td><em>Smith, van Goozen, Kuiper, &amp; Cohen-Kettenis</em></td>
<td>(Verschoor Poortinga 1988)</td>
</tr>
<tr>
<td></td>
<td>(Modified by authors of the Smith study)</td>
</tr>
<tr>
<td><strong>Gender Identity Trait Scale</strong></td>
<td>Published 1989 (German)</td>
</tr>
<tr>
<td><em>Altstotter-Gleich</em></td>
<td></td>
</tr>
<tr>
<td><strong>General Health Questionnaire</strong></td>
<td>Initial publication 1970</td>
</tr>
<tr>
<td><em>Goldberg &amp; Williams (manual)</em></td>
<td>Initial data: 553 consecutive adult patients in a single UK primary care practice were assessed. Sample of 200 underwent standardized psychiatric interview. Developed to screen for hidden psychological morbidity. Proprietary test. Now 4 versions.</td>
</tr>
<tr>
<td><strong>Hospital Anxiety &amp; Depression Scale</strong></td>
<td>Published in 1983</td>
</tr>
<tr>
<td><em>Zigmond &amp; Snaith</em></td>
<td>Initial data: Patients between 16 &amp; 65 in outpatient clinics in the UK &gt;100 patients; 2 refusals. 1st 50 compared to 2nd 50.</td>
</tr>
<tr>
<td><strong>Inventory of Interpersonal Problems</strong></td>
<td>Published 1988</td>
</tr>
<tr>
<td><em>Horowitz</em></td>
<td>Initial data: 103 patients about to undergo psychotherapy; some patients post psycho-therapy (Kaiser)</td>
</tr>
<tr>
<td>Test Name</td>
<td>Description</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>King’s Health Questionnaire</strong></td>
<td>Published in 1997&lt;br&gt;Initial data: 293 consecutive women referred for urinary incontinence evaluation in London&lt;br&gt;Comparison to SF-36</td>
</tr>
<tr>
<td><strong>Minnesota Multi-Phasic Personality Inventory</strong></td>
<td>Published in 1941&lt;br&gt;Updated in 1989 with new, larger, more diverse sample.&lt;br&gt;MMPI-2: 1,138 men &amp; 462 women from diverse communities &amp; several geographic regions in the U.S.A.&lt;br&gt;The test is copyrighted.</td>
</tr>
<tr>
<td><strong>Modified Androphilia-Gynephilia Index</strong></td>
<td>Neither the underlying version or the Blanchard modified version could be located in PubMed (Designed by the author of the Blanchard et al. study)</td>
</tr>
<tr>
<td><strong>“post-operative functioning 13 items”</strong></td>
<td>Published 1996 (Dutch) (Not in PubMed) (Designed by 1 of the authors of the Smith study)</td>
</tr>
<tr>
<td><strong>“post-operative functioning 21 items”</strong></td>
<td>Published 1996 (Dutch) (Not in PubMed) (Designed by 1 of the authors of the Smith study)</td>
</tr>
<tr>
<td><strong>Scale for Depersonalization Experiences</strong></td>
<td>Unpublished manuscript 1998 (University of Halle) (Designed by 1 of the authors of the Wolfpradt study)</td>
</tr>
<tr>
<td><strong>“sex trait function”</strong></td>
<td>Published 1997&lt;br&gt;Assessed in 22 adolescents (Designed by 1 of the authors of the Smith Study)</td>
</tr>
<tr>
<td><strong>Self-Esteem Scale</strong></td>
<td>Published 1965 (Not in PubMed) &lt;br&gt;Initial data: 5,024 high-school juniors &amp; seniors from 10 randomly selected New York schools</td>
</tr>
<tr>
<td><strong>Short-Form 36</strong></td>
<td>Originally derived from the Rand Medical Outcomes Study (n=2471 in version 1; 6742 in version 2 1989). The earliest test version is free.</td>
</tr>
<tr>
<td>Year</td>
<td>Instrument</td>
</tr>
<tr>
<td>----------</td>
<td>-------------------------------------------------</td>
</tr>
<tr>
<td>1993</td>
<td>Alternative scoring has been developed.</td>
</tr>
<tr>
<td></td>
<td><strong>Social Anxiety &amp; Distress Scale</strong></td>
</tr>
<tr>
<td></td>
<td><em>Watson &amp; Friend</em></td>
</tr>
<tr>
<td></td>
<td><strong>Social Support Scale</strong></td>
</tr>
<tr>
<td></td>
<td><em>Van Tilburg 1988</em></td>
</tr>
<tr>
<td></td>
<td><strong>Spielberger State &amp; Trait Anxiety Questionnaire</strong></td>
</tr>
<tr>
<td></td>
<td><em>Spielberger, Gorsuch, Lashene, Vagg, &amp; Jacobs</em></td>
</tr>
<tr>
<td></td>
<td><strong>Symptom Checklist-90</strong></td>
</tr>
<tr>
<td></td>
<td><em>Derogatis, Lipman, Covi</em></td>
</tr>
<tr>
<td></td>
<td><em>Derogatis &amp; Cleary</em></td>
</tr>
<tr>
<td></td>
<td><strong>Tennessee Self-Concept Scale</strong></td>
</tr>
<tr>
<td></td>
<td><em>Fitts &amp; Warren</em></td>
</tr>
<tr>
<td></td>
<td><strong>Utrecht Gender Dysphoria Scale</strong></td>
</tr>
<tr>
<td></td>
<td><em>Cohen-Kettenis &amp; van Goozen</em></td>
</tr>
<tr>
<td></td>
<td><strong>WHO-Quality of Life</strong></td>
</tr>
<tr>
<td></td>
<td><em>(abbreviated version)</em></td>
</tr>
<tr>
<td></td>
<td><em>Harper for WHO group</em></td>
</tr>
</tbody>
</table>

Althof et al., 1983; Greenberg, Frank, 1965; Gurtman, 1996; Lang, Vernon, 1977; Paap et al., 2012; Salkind et al., 1969; Vacchiano, Strauss, 1968.

Comment [RNG40]: This has been somewhat validated: Schneider, Catharina, et al. "Measuring gender dysphoria: a multicenter examination and comparison of the Utrecht Gender Dysphoria Scale and the Gender Identity/Gender Dysphoria Questionnaire for Adolescents and Adults." Archives of sexual behavior 45.3 (2016): 551-558.
Appendix F

Endpoint Data Types and Sources

Panel A (Controlled Studies)

<table>
<thead>
<tr>
<th>Author</th>
<th>National Data</th>
<th>Instrument w Substantive Normative Data</th>
<th>Instrument w/o Substantive &amp;/or Accessible Normative Data</th>
<th>Investigator-designed</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dhejne 2011</td>
<td>Yes</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Criminality, Mortality (Suicide, Cardiovascular Disease [possible adverse events from Hormone Tx], Cancer), Psych hx &amp; hospitalization, Suicide attempts</td>
</tr>
<tr>
<td>Dhejne</td>
<td>Yes</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Includes demographics*</td>
</tr>
<tr>
<td>Landen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Criminality, Education, Employment, Formal application for reversal of status, Psych dx &amp; tx, Substance abuse** More elements in earlier paper</td>
</tr>
<tr>
<td>Beatrice</td>
<td>-</td>
<td>MMPI form R, TSCS</td>
<td>-</td>
<td>1</td>
<td>Demographic Education, Income, Relationships</td>
</tr>
<tr>
<td>Haraldsen</td>
<td>-</td>
<td>SCL-90/90R</td>
<td>-</td>
<td>1</td>
<td>Demographic DSM Axis I, II, V (GAF), Substance abuse</td>
</tr>
<tr>
<td>Last Name</td>
<td>Measure(s)</td>
<td>Campaign</td>
<td>Scores</td>
<td>Check</td>
<td>Demographic</td>
</tr>
<tr>
<td>-----------</td>
<td>------------</td>
<td>----------</td>
<td>--------</td>
<td>-------</td>
<td>-------------</td>
</tr>
<tr>
<td>Heylens</td>
<td>SCL-90</td>
<td>-</td>
<td>-</td>
<td>Yes-2</td>
<td>Demographic</td>
</tr>
<tr>
<td>Ainsworth</td>
<td>Likely SF-36v2*</td>
<td>-</td>
<td>-</td>
<td>Yes-1</td>
<td>Demographic</td>
</tr>
<tr>
<td>Ruppin</td>
<td>SCL-90R</td>
<td>BSRI, FPI-R, IIP</td>
<td>Yes-2</td>
<td>Demographic</td>
<td>Adverse events from surgery, Employment, Psych tx, Relationships, Substance abuse</td>
</tr>
<tr>
<td>Smith</td>
<td>MMPI-short, SCL-90?R</td>
<td>BIS, UGDS, ?Cohen-Kettenis’, Doorn’s x2, (Gid-c, SSS)</td>
<td>Yes-1 or 2</td>
<td>Demographic</td>
<td>Adverse events from surgery, Employment, Relationships</td>
</tr>
<tr>
<td>Udeze Megeri</td>
<td>SCL-90R</td>
<td>BDI, GHQ, HADS,STAI-X1, STAI-X2</td>
<td>-</td>
<td>-</td>
<td>Psych eval &amp; ICD-10 dx</td>
</tr>
<tr>
<td>Kuhn</td>
<td>-</td>
<td>KHQ</td>
<td>Yes-1</td>
<td>Demographic</td>
<td>Relationships</td>
</tr>
<tr>
<td>Mate-Kole 1990</td>
<td>-</td>
<td>BSRI, CCEI</td>
<td>Yes-1</td>
<td>Demographic</td>
<td>Employment (relative change), Psych hx, Suicide hx</td>
</tr>
<tr>
<td>Wolfradt</td>
<td>-</td>
<td>BIQ, GITS, SDE, SES</td>
<td>Yes-1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Kraemer</td>
<td>-</td>
<td>FBeK</td>
<td>-</td>
<td>Demographic</td>
<td>-</td>
</tr>
<tr>
<td>Mate-Kole 1988</td>
<td>-</td>
<td>BSRI, CCEI</td>
<td>-</td>
<td>Demographic</td>
<td>Employment, Psych hx, Suicide hx,</td>
</tr>
<tr>
<td>Kockott</td>
<td>-</td>
<td>-</td>
<td>Yes-1</td>
<td>Demographic</td>
<td>Employment, Income, Relationships, Suicide attempts</td>
</tr>
<tr>
<td>Meyer</td>
<td>-</td>
<td>-</td>
<td>Yes-1</td>
<td>Demographic</td>
<td>Education, Employment, Income, Psych tx, Phallus</td>
</tr>
<tr>
<td>Author</td>
<td>National Data</td>
<td>Instrument w Substantive Normative Data</td>
<td>Instrument w/o Substantive &amp;/or Accessible Normative Data</td>
<td>Investigator-designed</td>
<td>Other</td>
</tr>
<tr>
<td>----------</td>
<td>---------------</td>
<td>-----------------------------------------</td>
<td>----------------------------------------------------------</td>
<td>-----------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Rakic</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Yes-1</td>
<td>Demographic</td>
</tr>
<tr>
<td>Weyers</td>
<td>-</td>
<td>SF-36</td>
<td>FSFI</td>
<td>Yes-2</td>
<td>Demographic</td>
</tr>
<tr>
<td>Blanchard</td>
<td>-</td>
<td>SCL-90R</td>
<td>(AG)</td>
<td>Yes-1</td>
<td>Demographic</td>
</tr>
<tr>
<td>Wierckx</td>
<td>-</td>
<td>SF-36</td>
<td>-</td>
<td>Yes-3</td>
<td>Demographic</td>
</tr>
<tr>
<td>Eldh</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Yes-1</td>
<td>-</td>
</tr>
<tr>
<td>Hess</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Yes-1</td>
<td>-</td>
</tr>
<tr>
<td>Lawrence</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Yes-4</td>
<td>Demographic</td>
</tr>
</tbody>
</table>

Panel B (Surgical Series: No Concurrent Controls)
<table>
<thead>
<tr>
<th>Author</th>
<th>National Data</th>
<th>Instrument w/ Substantive Normative Data</th>
<th>Instrument w/o Substantive &amp;/or Accessible Normative Data</th>
<th>Investigator-designed</th>
<th>Other</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salvador</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Yes-1</td>
<td>Demographic</td>
<td>Relationships</td>
</tr>
<tr>
<td>Tsoi</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Yes-1</td>
<td>Demographic</td>
<td>Education, Employment, Relationships (relative change)</td>
</tr>
</tbody>
</table>

Panel C (Mixed Treatment Series: No Direct Control Groups)

<table>
<thead>
<tr>
<th>Author</th>
<th>National Data</th>
<th>Instrument w/ Substantive Normative Data</th>
<th>Instrument w/o Substantive &amp;/or Accessible Normative Data</th>
<th>Investigator-designed</th>
<th>Other</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asscheman et al.</td>
<td>Yes</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Demographic</td>
<td>Mortality (HIV, Possible adverse events from Hormone Tx, Substance abuse, Suicide)</td>
</tr>
<tr>
<td>Motmans et al.</td>
<td>-</td>
<td>SF36 EQOLS (2nd)</td>
<td>-</td>
<td>-</td>
<td>Demographic</td>
<td>Education, Employment, Income, Relationships</td>
</tr>
<tr>
<td>Newfield et al.</td>
<td>-</td>
<td>SF-36v2</td>
<td>-</td>
<td>-</td>
<td>Demographic</td>
<td>Income</td>
</tr>
<tr>
<td>Gómez-Gil et al.2014</td>
<td>-</td>
<td>WHOQOL-BREF</td>
<td>APGAR</td>
<td>Yes-1</td>
<td>Demographic</td>
<td>Education, Employment, Relationships</td>
</tr>
<tr>
<td>Gómez-Gil et al.2012</td>
<td>-</td>
<td>-</td>
<td>HADS, SADS</td>
<td>-</td>
<td>Demographic</td>
<td>Education, Employment, Living arrangements</td>
</tr>
<tr>
<td>Hepp et al.</td>
<td>-</td>
<td>-</td>
<td>HADS</td>
<td>-</td>
<td>Demographic</td>
<td>DSM Axis 1&amp; II Psych dx</td>
</tr>
<tr>
<td>Johansson</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Yes-1</td>
<td>Demographic</td>
<td>Axis V</td>
</tr>
<tr>
<td>et al.</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Demographic Employment, Disability, DVT, HIV status, Psych dx</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>----------------------------------------------------------------</td>
<td></td>
</tr>
</tbody>
</table>

*Listed as San Francisco-36 in manuscript
** From medical charts & verdicts ?=Possibly self-designed
AG=Androphilia-Gynephilia Index (investigator designed 1985) (used more for classification)
APGAR=Family Adaptability, Partnership growth, Affection, and Resolve
BDI=Beck Depression Inventory
BIQ=Body Image Questionnaire
BSRI=Bem Sex Role Inventory
CCEI=Crown Crisp Experiential Index
Cohen-Kettenis’= Sex trait function (An author helped design)
Dorn’s x2= Post-operative functioning 13 items (An author helped design)
  Post-operative functioning 21 items (An author helped design)
EQOLS (2nd)=2nd European Quality of Life Survey
FBeK=Fragebogen zur Beurteilung des eigenen Korpers
FPI-R=A version of the Freiberg Personality Inventory
FSFI+Female Sexual Function Index
GHQ=General Health Questionnaire
Gid-c=Gender identity disorder in childhood (used more for predictors) (An author helped design)
GITS=Gender Identity Trait Scale
HADS=Hospital Anxiety Depression Scale
IIP=Inventory of Interpersonal Problems
KHQ=King’s Health Questionnaire
MMPI=Minnesota Multi-phasic Personality Inventory
SADS=Social Anxiety & Distress Scale
SCL-90 (cR)=A version of the Symptom Checklist 90
SDE=Scale for Depersonalized Experiences (An author designed)
SES=Self-Esteem Scale
SF-36 (v2)=Short Form-36(version2)
SSS=Social Support Scale (used more for predictors)
STAI-X1, STAI-X2=Spielberger State and Trait Anxiety Questionnaire
TSCS=Tennessee Self-Concept Scale
UGDS=Utrecht Gender Dysphoria Scale (An author helped design)
WHOQOL-BREF=World Health Organization-Quality of Life (abbreviated version)

**Appendix G.**
Longitudinal Studies Which Used Patients as Their Own Controls and Which Used Psychometric Tests with Extensive Normative Data or Longitudinal Studies Which Used National Data Sets

<table>
<thead>
<tr>
<th>Author</th>
<th>Test</th>
<th>Patient and Data Loss</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heylens et al.</td>
<td>SCL-90R</td>
<td>90 applicants for SRS were recruited.</td>
<td>At t=0, the mean global “psychoneuroticism” SCL-90R score, along with scores of 7 of 8 subscales, were statistically more pathologic than the general population.</td>
</tr>
<tr>
<td>Belgium 2014</td>
<td></td>
<td>• 8 (8.9%) declined participation.</td>
<td>After hormone tx, the mean score for global “psychoneuroticism” normalized &amp; remained normal after reassignment surgery.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• 12 (13.3%) excluded b/c GID-NOS dx.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• 12 (13.3%) did not complete the treatment sequence b/c of psychiatric/physical co-morbidity, personal decision for no tx, or personal decision for only hormone tx.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• 1 (1.1%) committed suicide during follow-up.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>57 (63.3% of recruited) entered the study.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• 1 (12.2% of initial recruits) had not yet received SRS by study close.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>→46 (51.1% of recruited) underwent serial evaluation</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The test was not completed by 1 at t=0, 10 at t=1 (after hormone tx), &amp; 4 at t=2 (after SRS)</td>
<td>→missing data for</td>
</tr>
<tr>
<td>Source</td>
<td>Test</td>
<td>Description</td>
<td>Results</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Ruppin, Pfafflin, Germany 2015</td>
<td>SCL-90R</td>
<td>The number in the available patient pool was not specified. 140 received recruitment letters. • 2 (1.4% of those with recruitment letters) had died. • 1 (0.7%) was institutionalized. • 5 (3.6%) were ill. • 8 (5.7%) did not have time. • 8 (5.7%) stated that GD was no longer an issue. • 8 (5.7%) provided no reason. • 28 (20.0%) declined further contact. • 9 (6.4%) were lost to follow-up. → 71 (50.7%) agreed to participate. • 2 (1.4%) had not undergone SRS • The test was not completed by 9. → missing data for another 6.4%.</td>
<td>At t=0, the “global severity index “SCL-90R score was 0.53±0.49. At post-SRS follow-up the score had decreased to 0.28±0.36. The scores were statistically different from one another, but are of limited biologic significance given the range of the score for this scale: 0-4. In the same way, all of the subscale scores were statistically different, but the effect size was reported as large only for “interpersonal sensitivity”: 0.70±0.67 at t=0 and 0.26±0.34 post-SRS.</td>
</tr>
<tr>
<td>Smith et al. Holland 2005</td>
<td>MMPI SCL-90</td>
<td>The number in the available adult patient pool was not specified. 325 adult &amp; adolescent applicants for SRS were recruited. • 103 (31.7%) were not eligible to start</td>
<td>Most of the MMPI scales were already in the normal range at the time of initial testing. At t=0, the global “psychoneuroticism” SCL-90 score, which included the drop-</td>
</tr>
</tbody>
</table>
hormone tx & real-life experience.
- 34 (10.7%) discontinued hormone tx 162 (an unknown percentage of the initial recruitment) provided pre-SRS test data.
- 36 to 61 (22.2% - 37.6% of those adults w pre-SRS data) did not complete post-SRS testing.

Udeze, et al. 2008
Megeri, Khoosal 2007
UK

<table>
<thead>
<tr>
<th>National Databases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dehjne Sweden 2011</td>
</tr>
<tr>
<td>804 with GID in Sweden 1973 to 2003 were identified.</td>
</tr>
<tr>
<td>- 480 (59.7%) did not apply or were not approved for SRS 324 (40.3%) underwent SRS.</td>
</tr>
<tr>
<td>- All were followed.</td>
</tr>
<tr>
<td>All cause mortality was higher (n=27[8%]) than in controls (H.R 2.8 [1.8-4.3]) even after adjustment for covariants.</td>
</tr>
<tr>
<td>Divergence in survival curves was observed after 10 years. The major</td>
</tr>
</tbody>
</table>

At post SRS-follow-up, the score had decreased to 120.3±31.4.
The scores were statistically different from one another, but are of limited biologic significance given the range of the score for this scale: 90 to 450, with higher scores consistent with more psychological instability.

At t=0, the mean raw global score was 48.33. At post-SRS follow-up, the mean score was 49.15.
There were no statistically significant changes in the global score or for any of the subscales.
3240 controls of the natal sex and 3240 controls of the reassigned gender were randomly selected from national records. A contributor was completed suicide (n=10 [3%]; adjusted H.R. 19.1 [5.8–62.9]).

Suicide attempts were more common (n= 29 [9%]) than in controls (adjusted H.R. 4.9 [2.9–8.5]).

Hospitalizations for psychiatric conditions (not related to gender dysphoria) were more common n= 64 [20%] than in controls (H.R. 2.8 [2.0–3.9]) even after adjusting for prior psychiatric morbidity.

<table>
<thead>
<tr>
<th>Dhejne et al. 2014</th>
<th>Swedish National Registry</th>
<th>767 applied for SRS/legal status (1960-2010)</th>
<th>15 formal applications for reversal to natal/original gender (2.2% of the SRS population) were identified thus far (preliminary number). (Does not reflect other manifestations of regret such as suicide.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Landen et al. 1998</td>
<td>Sweden</td>
<td></td>
<td>25 (3.3%) applications denied. 61 (8.0%) not granted full legal status 681 (88.7%) underwent SRS. All were followed.</td>
</tr>
</tbody>
</table>
Ahmed SF, Morrison S, Hughes IA. Intersex and gender assignment; the third way? Arch Dis Child. 2004 Sep;89(9):847-50. PMID: 15321864.


Gates GJ. How many people are lesbian, gay, bisexual, and transgender? 2011. (Not in PubMed) williamsinstitute@law.ucla.edu.


Lothstein LM. The aging gender dysphoria (transsexual) patient. Arch Sex Behav. 1979 Sep;8(5):431-44. PMID: 496624.


PRO Guidance 2009: U.S. Department of Health and Human Services; Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics
Evaluation and Research (CBER), Center for Devices and Radiological Health (CDRH). 

Quirós C, Patrascioiu I, Mora M, Aranda GB, Hanzu FA, Gómez-Gil E, Godás T, Halperin I. 


Stephens SC, Bernstein KT, Philip SS. Male to female and female to male transgender persons have different sexual risk behaviors yet similar rates of STDs and HIV. *AIDS Behav.* 2011 Apr;15(3):683-6. PMID: 20694509.


