

The DSM Diagnostic Criteria for Gender Identity Disorder in Children

Kenneth J. Zucker

© American Psychiatric Association 2009

Abstract In this article, I review the diagnostic criteria for Gender Identity Disorder (GID) in children as they were formulated in the DSM-III, DSM-III-R, and DSM-IV. The article focuses on the cumulative evidence for diagnostic reliability and validity. It does not address the broader conceptual discussion regarding GID as “disorder,” as this issue is addressed in a companion article by Meyer-Bahlburg (2009). This article addresses criticisms of the GID criteria for children which, in my view, can be addressed by extant empirical data. Based in part on reanalysis of data, I conclude that the persistent desire to be of the other gender should, in contrast to DSM-IV, be a necessary symptom for the diagnosis. If anything, this would result in a tightening of the diagnostic criteria and may result in a better separation of children with GID from children who display marked gender variance, but without the desire to be of the other gender.

Keywords Gender Identity Disorder · Children · DSM-V

...no one should mistake expert consensus for the truth (Hyman, 2003)

Introduction

Gender Identity Disorders entered the DSM nosological system with the publication of DSM-III (American Psychiatric

Association, 1980). In DSM-III, there were three relevant diagnostic entities: Gender Identity Disorder of Childhood (GIDC), Transsexualism (for adolescents and adults), and Psychosexual Disorder Not Elsewhere Classified. The last category was a residual diagnosis, “for disorders whose chief manifestations are psychological disturbances not covered by any of the other specific categories in the diagnostic class of Psychosexual Disorders” (American Psychiatric Association, 1980, pp. 282–283). One example pertained to “marked feelings of inadequacy related to self-imposed standards of masculinity or femininity...” (p. 283). In DSM-III-R (American Psychiatric Association, 1987), there were four relevant diagnostic entities: GIDC, Transsexualism, Gender Identity Disorder of Adolescence or Adulthood, Nontranssexual Type (GIDAANT), and Gender Identity Disorder Not Otherwise Specified (GIDNOS). The last category was a residual diagnosis and four examples were provided: (1) children with persistent cross-dressing without the other criteria for GIDC; (2) adults with transient, stress-related cross-dressing behavior; (3) adults with the clinical features of Transsexualism of less than 2 years’ duration; and (4) people who have a persistent preoccupation with castration or peotomy without a desire to acquire the sex characteristics of the other sex (American Psychiatric Association, 1987, p. 78). In DSM-IV and DSM-IV-TR (American Psychiatric Association, 1994, 2000), there were three relevant diagnostic entities: Gender Identity Disorder (GID) (with separate criteria sets for children versus adolescents/adults), Transvestic Fetishism (with Gender Dysphoria), and GIDNOS. The last category was a residual and three examples were provided: (1) intersex conditions with “accompanying gender dysphoria” (p. 582); (2) transient, stress-related cross-dressing behavior; and (3) persistent preoccupation with castration or penectomy without a desire to acquire the sex characteristics of the other sex. In DSM-IV, the previous categories of GIDC and Transsexualism were

K. J. Zucker (✉)
Gender Identity Service, Child, Youth, and Family Program,
Centre for Addiction and Mental Health, 250 College Street,
Toronto, ON M5T 1R8, Canada
e-mail: Ken_Zucker@camh.net

collapsed into one overarching diagnosis, GID, which had, as noted above, distinct criteria sets for children versus adolescent and adults. On the recommendation of the DSM-IV Subcommittee on Gender Identity Disorders (Bradley et al., 1991), elements of the GIDAANT diagnosis were also incorporated into the DSM-IV criteria for GID for adolescents and adults.

Over these three editions of the DSM, the Gender Identity Disorders have had different placements in the manual: in DSM-III, the diagnoses were in the section called Psychosexual Disorders; in DSM-III-R, the diagnoses were in the section called Disorders Usually First Evident in Infancy, Childhood, or Adolescence; and, in DSM-IV, the diagnoses were in the section called Sexual and Gender Identity Disorders.

This review paper will focus on the GID diagnostic criteria for children. It will examine the evolution of the criteria sets, evidence for their reliability and validity, criticisms of the current criteria, and then proposed options for reform of the criteria. In this review, I will not comment on the DSM-IV-TR GIDNOS diagnosis (or its predecessors in DSM-III and DSM-III-R), as this category will be discussed and considered by the entire Gender Identity Disorders subworkgroup. When I discuss below children who are subthreshold for the GID diagnosis, this is not meant to imply that they would meet criteria for GIDNOS as it has been formulated in the various editions of the DSM. The term “subthreshold” simply means that the child was not judged to meet the complete diagnostic criteria for GID.

Review of the Diagnostic Criteria (DSM-III, DSM-III-R, and DSM-IV)

DSM-III

Table 1 shows the DSM-III diagnostic criteria for GIDC. It should be noted that the criteria were somewhat different for females versus males (girls versus boys), a tradition that has continued through the DSM-IV and DSM-IV-TR. Although beyond the scope of this review, that the DSM has specified somewhat different criteria for boys versus girls is of interest in its own right, as there are very few DSM diagnoses that have sex-specific criteria. Some authors have, however, argued that they might be necessary for some conditions, such as Conduct Disorder (CD) (see, e.g., Crick & Zahn-Waxler, 2003; Zahn-Waxler, 1993; Zoccolillo, 1993; for a general overview, see Widiger, 2007; Widiger & First, 2007).

For the Point A criterion, both girls and boys were required to have a “strongly and persistently stated desire” to be of the other sex or to verbalize the “insistence” that one was a member of the other sex; for girls, there was the additional proviso that such a desire was not due to a perceived cultural advantage from being a boy. No such proviso was required for boys.

Table 1 DSM-III diagnostic criteria for Gender Identity Disorder of Childhood

For females

- A. Strongly and persistently stated desire to be a boy, or insistence that she is a boy (not merely a desire for any perceived cultural advantages from being a boy)
- B. Persistent repudiation of female anatomic structures, as manifested by at least one of the following repeated assertions
 - (1) that she will grow up to become a man (not merely in role)
 - (2) that she is biologically unable to become pregnant
 - (3) that she will not develop breasts
 - (4) that she has no vagina
 - (5) that she has, or will grow, a penis
- C. Onset of the disturbance before puberty (For adults and adolescents, see Atypical Gender Identity Disorder.)

For males

- A. Strongly and persistently stated desire to be a girl, or insistence that he is a girl
- B. Either (1) or (2)
 - (1) persistent repudiation of male anatomic structures, as manifested by at least one of the following repeated assertions
 - (a) that he will grow up to become a woman (not merely in role)
 - (b) that his penis and testes are disgusting or will disappear
 - (c) that it would be better not to have a penis or testes
 - (2) preoccupation with female stereotypical activities as manifested by a preference for either cross-dressing or simulating female attire, or by a compelling desire to participate in the games and pastimes of girls
- C. Onset of the disturbance before puberty. (For adults and adolescents, see Atypical Gender Identity Disorder.)

For the Point B criterion for girls, there was only one criterion: persistent repudiation of female anatomic structures (inferred from at least one of five indicators). For boys, there was an analogous persistent repudiation of male anatomic structures (inferred from at least one of three indicators), but there was a second criterion that could also be used. This criterion pertained to a “preoccupation with female stereotypical activities” (as manifested by at least one of two behavioral indicators) or by a “compelling desire” to participate in cross-gender activities.

For girls, then, GIDC was diagnosed based on two criteria: a persistent wish to be of the other sex and by the persistent negation of one’s sexual anatomy. For boys, GIDC was diagnosed based on a minimum of two criteria: a persistent wish to be of the other sex and by the persistent negation of one’s sexual anatomy *or* some manifestation of pervasive cross-gender role preferences/desires.

Comment and Critique

The criteria were formulated by a panel of experts, i.e., by at least some members of the Psychosexual Disorders Advisory Committee who had clinical and research experience with this population (e.g., Green, 1974; Stoller, 1968). The criteria were

not subject to any formal field trials for the purpose of establishing diagnostic reliability or validity.

Various descriptors in the criteria (“strongly,” “persistent/persistently,” “insistence,” “preoccupation,” and “compelling desire”) were all presumably used to differentiate children with potential gender identity problems from children who might, on a transitory or infrequent basis, verbalize a desire to be of the other sex or engage in cross-gender behavior (B2 for boys) (see, e.g., Lindsay, 1994). This point was also emphasized in the text portion of the DSM-III (American Psychiatric Association, 1980, pp. 264–265). Because the criteria for girls required the presence of “anatomic dysphoria” (Criterion B), but the criteria for boys did not (Criterion B2 was sufficient for this criterion), it could be argued that the criteria for girls were more conservative than they were for boys (for a historical documentation of this point using unpublished archival material, see Bryant, 2007).

In an early critique of the DSM-III criteria, Zucker (1982) argued that Criterion B for girls was, perhaps, overly stringent in that there was a strong emphasis on immature, if not “delusional,” statements (e.g., that “she has no vagina,” “that she has...a penis”). There was less of an emphasis on feelings of anatomic dysphoria or “fantasies” of having the sexual anatomy of the other sex (e.g., that one would, on a frequent basis, like to have a penis).

Reliability and Validity

Zucker, Finegan, Doering, and Bradley (1984) conducted the only study that attempted to establish the reliability of the DSM-III GIDC criteria in terms of agreement between two raters. From chart information reported by parents of gender-referred children during a clinical interview, Zucker et al. had two coders independently use this information to judge if the child met DSM criteria for GIDC ($N = 31$). For Criterion A, the two raters agreed in 34 of the 36 cases (19 present, 15 absent). Because this research team did not find that Criterion B “worked” for girls, Criterion B ratings were limited to boys ($N = 31$). For this criterion, the two raters agreed in 28 of the 31 cases (16 present, 12 absent).

Comparative studies of the sex-typed behavior of children referred for potential problems in their gender identity development versus various control groups (siblings, clinical controls, and non-referred controls) have been the most common method to establish the validity of the GID diagnosis (Zucker, 1992). Such studies have relied on a variety of measurement approaches: item analysis from questionnaires, standardized behavioral observations, projective tests or psychometrically sound questionnaires (for a summary review of commonly used measures, see Zucker, 2005a). As I will argue in more detail below, this line of research constitutes some of the strongest evidence for the validity of the GID diagnosis vis-à-vis the psychometric concept of “discriminant validity” (cf. Rutter, 1978).

Although not intended to be exhaustive, I will provide a couple of examples from the DSM-III era. In Green’s (1974, 1976, 1987) study of feminine boys and a control group of boys unselected for their degree of masculinity-femininity (both groups were recruited via advertisement), an array of questions (item analysis) answered by parents was used to test for significant between-groups differences. Using parent-interview data, Green (1987) reported a discriminant function analysis in which 6 of 16 sex-typed behaviors (e.g., cross-dressing, wish to be a girl, etc.) were able to classify correctly all boys as members of either the feminine group or the control group (see also Roberts, Green, Williams, & Goodman, 1987).

In DSM-III, one of the criteria for boys pertained to “a compelling desire to participate in the games and pastimes of girls.” Zucker, Doering, Bradley, and Finegan (1982) reported on the free play behavior of gender-referred children compared to that of their siblings and clinical controls on a 3-trial task in a “laboratory-like” situation. On Trial 1 (5 min), the children were exposed to stereotypical masculine and feminine toys; on Trial 2 (5 min), they were exposed to stereotypical masculine and feminine dress-up apparel; and on Trial 3 (10 min), they were exposed to both sets of stimuli simultaneously. On all three trials, the gender-referred children played significantly longer with the cross-sex stimuli than did the two control groups. Zucker (2005a) subsequently reported, using a conservative method, the effect sizes for this comparative analysis: the effect sizes ranged from 0.78 to 1.36.¹

In a subsequent study, Zucker et al. (1984) provided additional comparative analysis using several measures of sex-typed behavior by comparing gender-referred children who were judged by a clinician to meet the complete DSM-III criteria for GIDC versus those who were deemed subthreshold for the diagnosis. Of 10 specific measures, the threshold group showed more cross-gender behavior than the subthreshold group and, of these, six of the differences were statistically significant. This was the first in a series of studies that provided at least some evidence of discriminant validity within samples of gender-referred children, i.e., those threshold versus subthreshold for the diagnosis of GID (for further discussion, see below).

¹ As noted in Zucker (2005a), there is some debate in the literature regarding decision rules for calculating an effect size. In calculating Cohen’s d ($M_1 - M_2/SD$), one can use the pooled SD of two groups or the SD of the control group. In calculating normative gender differences for d , there is really no control group, so it is customary to use the pooled SD ; however, when comparing a group of probands with a control group, it can be argued that the SD of the control group is more appropriate than the pooled SD (see, e.g., Glass, McGaw, & Smith, 1981, pp. 106–107). If the variance for the control group is considerably smaller than the variance for the probands, d will be larger if only the control group SD is used, and this may well be important to consider with regard to clinical matters. In Zucker (2005a), the more conservative effect size formula was used because this was what most authors used in their own studies. Effect size calculations would be substantially higher if only the control group SD was used (see, e.g., Johnson et al., 2004).

From this first wave of empirical studies regarding discriminant validity (for reviews, see Zucker, 1992; Zucker & Bradley, 1995), there appeared to be reasonable evidence for diagnostic specificity.

DSM-III-R

Table 2 shows the DSM-III-R diagnostic criteria for GIDC. The structure of the criteria was similar to that in DSM-III, but there were some important changes in wording and in the content of the criteria. For both girls and boys, the phrase “[p]ersistent and intense distress about being a girl (boy)” was added to the Point A criterion. For girls, the phrase “strongly and persistently stated desire to be a boy” now read as “a stated desire to be a boy.” For boys, the phrase “strongly and persistently stated desire to be a girl” now read as “an intense desire to be a girl.” For the Point B criterion for girls, B1 was new and pertained to a girls’ marked rejection of the wearing of “normative feminine clothing” and an “insistence on wearing stereotypical masculine clothing...” and the wording for B2 (pertaining to anatomic dysphoria) was modified from how it was formulated in DSM-III. For the Point B criterion for boys, B1 introduced a

Table 2 DSM-III-R diagnostic criteria for Gender Identity Disorder of Childhood

For females

A. Persistent and intense distress about being a girl, and a stated desire to be a boy (not merely a desire for any perceived cultural advantages from being a boy), or insistence that she is a boy

B. Either (1) or (2)

- (1) persistent marked aversion to normative feminine clothing and insistence on wearing stereotypical masculine clothing, e.g., boys’ underwear and other accessories
- (2) persistent repudiation of female anatomic structures, as evidence by at least one of the following
 - (a) an assertion that she has, or will, grow a penis
 - (b) rejection of urinating in a sitting position
 - (c) assertion that she does not want to grow breasts or menstruate

C. The girl has not yet reached puberty.

For males

A. Persistent and intense distress about being a boy and an intense desire to be a girl, or, more rarely, insistence that he is a girl

B. Either (1) or (2)

- (1) preoccupation with female stereotypical activities, as shown by a preference for either cross-dressing or simulating female attire, or by an intense desire to participate in the games and pastimes of girls and rejection of male stereotypical toys, games, and activities
- (2) persistent repudiation of male anatomic structures, as manifested by at least one of the following repeated assertions
 - (a) that he will grow up to become a woman (not merely in role)
 - (b) that his penis and testes are disgusting or will disappear
 - (c) that it would be better not to have a penis or testes

C. The boy has not yet reached puberty

concomitant rejection of “male stereotypical toys, games, and activities.” The wording for B2 was identical to its wording in DSM-III.

Comment and Critique

The revised DSM-III-R criteria were formulated by a panel of experts, i.e., by at least some members of the Subcommittee on Gender Identity Disorders who had clinical and research experience with this population. The criteria were not subject to any formal field trials for the purpose of establishing diagnostic reliability or validity.

In one critique of the DSM-III-R criteria, Zucker (1992) noted that the addition of the distress passage in Criterion A was not accompanied by any formal guidelines regarding how it should be assessed nor was it made clear in what ways the distress was considered distinct from other operationalized components in the Point A criteria. It was also noted that the remainder of the descriptive material in Point A had a subtle difference between the two sexes. For girls, a “stated desire to be a boy” was required whereas, for boys, an “intense desire to be a girl” was required. In addition, the phraseology for girls did not address the issue of intensity or some other variable pertaining to duration (see Morgan, 2000). Why these distinctions appeared in the DSM-III-R was not clear because the phraseology in Criterion A for the two sexes was identical in the DSM-III. Langer and Martin (2004) noted that this distinction appeared to result in a lower threshold for boys than for girls in that boys did not have to verbalize the desire to be a girl. On the other hand, it could be argued that the threshold was, in fact, lower for girls than for boys because it lacked an intensity criterion.

As I have noted elsewhere (Zucker, 2006a), it is not clear why the DSM-III-R wound up changing the criteria for boys to “an intense desire” to be a girl from a “strongly and persistently stated desire.” The original was clearly more stringent. As a member of the DSM-III-R Subcommittee on Gender Identity Disorders, I have reviewed my own correspondence file and could find no indication that this distinction was either noted or commented upon by the Subcommittee at large. My conclusion, as stated in Zucker (2006a), was that “the committee just goofed.” It is also conceivable that the distinctions in phraseology were overlooked in the context of the added proviso of “persistent and intense distress.”

Reliability and Validity

To my knowledge, no formal studies examined the reliability of the DSM-III-R diagnostic criteria for children, i.e., inter-clinician agreement. As during the DSM-III era, the most common line of evidence for validity involved comparative studies of gender-referred children versus that of various control groups along with comparisons of gender-referred children deemed threshold versus subthreshold based on clinician diagnosis. I

will not review these studies here, as the same type of interpretive overview would apply to them as to the wave of DSM-III era studies (for references, see Zucker, 1992; Zucker & Bradley, 1995; Zucker, Bradley, & Sanikhani, 1997; Zucker, Lozinski, Bradley, & Doering, 1992; Zucker et al., 1993, 1999). Some of the more important studies using mixed samples of DSM-III-R and DSM-IV era patients are, however, reviewed in some detail below.

Predictive Validity

Wallien and Cohen-Kettenis (2008) reported psychosexual follow-up data on 77 gender-referred children (59 boys, 18 girls), originally assessed at a mean age of 8.4 years (range, 5–12). At the time of follow-up, the mean age was 18.9 years (range, 16–28). Regarding gender identity at follow-up, 21 children (12 boys, 9 girls) were classified as persisters, i.e., these children were still gender dysphoric and were seen clinically because of an ongoing desire for sex-reassignment (hormonal and surgical treatment); the remaining 56 children were classified as desisters (i.e., they were no longer gender-dysphoric), either based on a formal re-assessment or because they had not re-contacted the clinic requesting sex-reassignment. Of the 21 persisters, all had received a DSM-III-R diagnosis of GIDC at the time of assessment in childhood, compared to 37 (66.0%) of the desisters, a significant difference. On two dimensional measures of cross-gender identity, the Gender Identity Questionnaire for Children (GIQC) (Johnson et al., 2004) and the Gender Identity Interview for Children (GIIC) (Wallien et al., 2009; Zucker et al., 1993), the persisters showed significantly more cross-gender behavior and gender identity confusion than the desisters. Thus, using both categorical diagnosis and dimensional measures, Wallien and Cohen-Kettenis provided some evidence for predictive validity vis-à-vis persistence versus desistance. In my view, these data constitute an important addition to the empirical literature regarding the validity of the GIDC criteria.

DSM-IV

Table 3 shows the DSM-IV child criteria for GID. For the DSM-IV, the Subcommittee on Gender Identity Disorders (Bradley et al., 1991) reviewed the merit of altering the criteria for children to a polythetic format, in which various behavioral traits would be operationalized, from which a specified number would be required to meet the criteria for the diagnosis of GID. In its final form, there were two clinical indicator (symptom) criteria. As shown in Table 3, Criterion A was described as “[a] strong and persistent cross-gender identification (not merely a desire for any perceived cultural advantages of being the other sex)” (p. 537) and a child was deemed to meet this criterion if he or she manifested at least four of the five indicators. Criterion B was described as a “[p]ersistent discomfort with his or her sex or sense of inappropriateness in the gender role of that sex” (p. 537) and a

Table 3 DSM-IV diagnostic criteria for Gender Identity Disorder (for children)

A. A strong and persistent cross-gender identification (not merely a desire for any perceived cultural advantages of being the other sex)
In children, the disturbance is manifested by at least four (or more) of the following
(1) repeatedly stated desire to be, or insistence that he or she is, the other sex
(2) in boys, preference for cross-dressing or simulating female attire; in girls, insistence on wearing only stereotypical masculine clothing
(3) strong and persistent preferences for cross-sex roles in make-believe play or persistent fantasies of being the other sex
(4) intense desire to participate in the stereotypical games and pastimes of the other sex
(5) strong preference for playmates of the other sex
B. Persistent discomfort with his or her sex or sense of inappropriateness in the gender role of that sex
In children, the disturbance is manifested by any of the following: in boys, assertion that his penis or testes are disgusting or will disappear or assertion that it would be better not to have a penis, or aversion toward rough-and-tumble play and rejection of male stereotypical toys, games, and activities; in girls, rejection of urinating in a sitting position, assertion that she has or will grow a penis, or assertion that she does not want to grow breasts or menstruate, or marked aversion toward normative feminine clothing
C. The disturbance is not concurrent with a physical intersex condition
D. The disturbance causes clinically significant distress or impairment in social, occupational, or other important areas of functioning

child was deemed to meet this criterion if he or she manifested one of two indicators.

Compared to the diagnostic criteria for GIDC in DSM-III-R, there were five changes to the criteria set:

1. In contrast to both DSM-III and DSM-III-R, Criterion A contained the proviso “not merely a desire for any perceived cultural advantages of being the other sex” for both boys and girls, not just for girls.
2. The distress element of the Point A criterion in DSM-III-R (e.g., “[p]ersistent and intense distress about being a girl...”) was deleted and moved to the Point D clinical significance criterion. Note that the clinical significance criterion was added to about half of the DSM-IV diagnoses (see Spitzer & Wakefield, 1999; Wakefield & First, 2003).
3. For both boys and girls, the verbalized desire to be of the other sex was no longer a distinct criterion. Rather, in DSM-IV, it became one of five indicators for Criterion A. In contrast to DSM-III-R, this criterion was harmonized (equalized) for boys and girls. In DSM-III-R, girls only required a “stated desire” to be a boy, whereas boys were required to have “an intense desire” to be a girl (as noted earlier, the absence of an “intensity” qualifier for girls

was likely an oversight). Thus, in DSM-IV, Criterion A1 was written in a manner such that it could be applied equally to both boys and girls.

4. For boys, the other behavioral indicators of cross-gender identification (A2–A5) were better separated (in DSM-III-R, for example, two of these were given as examples of the B1 criterion) (see Table 2). For girls, three of these four behavioral indicators were new, as they were not explicitly required for girls in the DSM-III-R (A2, which pertained to cross-dressing, was extracted from the B1 criterion in DSM-III-R).
5. For Criterion B, for boys, there remained some similarity to the B1 criterion in DSM-III-R: the criterion was deemed met if a boy displayed signs of anatomic dysphoria *or* displayed an “aversion toward rough-and-tumble play *and* rejection of male stereotypical toys, games, and activities” (emphasis added). For girls, there also remained some similarity to the B1 criterion in DSM-III-R: the criterion was deemed met if a girl displayed signs of anatomic dysphoria *or* displayed a “marked aversion toward normative feminine clothing.”

Comment and Critique

Since the publication of DSM-IV, there have been various critiques leveled at the GID diagnosis as it applies to children. In this section, I will review the key conceptual and procedural criticisms. I will not, however, formally address the most fundamental criticism, namely that GID is not a mental disorder and should be removed from the DSM in its entirety. On this point, Meyer-Bahlburg (2009) has addressed the competing views on this broad philosophical debate (see also Bockting, 2009). I will also not address some of the putative sociopolitical criticisms of the GID diagnosis (see, e.g., Feder, 1997; Hegarty, 2009; Martin, 2008; Minter, 1999; Morgan, 2000; Sedgwick, 1991), such as the claim that it was introduced into the DSM-III as a backdoor maneuver to replace homosexuality (e.g., Ault & Bruzuzy, 2009), which had been delisted from the DSM-II in 1973. On this point, I have provided my own view elsewhere (Zucker & Spitzer, 2005; see also Meyer-Bahlburg, 2009; Zucker, Drummond, Bradley, & Peterson-Badali, 2009). Regarding sociopolitical issues in general, Drescher (2009) has provided an overview of this topic. In appraising the criticisms, I will attempt to address them, when possible, with empirical data.

The Cultural Consideration Proviso in Point A

This aspect of the Point A criteria has received little empirical attention. In a critique of this proviso as it appeared in DSM-III and DSM-III-R, I have previously argued that this exclusion rule

presented a couple of difficulties (Zucker, 1992). First, it made an etiological or motivational assumption about a girl’s desire to be a boy (e.g., that one type of desire was based on a perception of cultural disadvantage or bias, whereas at least one other unspecified type was not). A similar such assumption was not applied to boys. Although this distinction might have relevance for such parameters as natural history and response to treatment, I argued that it was unclear why it should be used diagnostically (Zucker, 1992). I also argued that the absence of this proviso for boys was puzzling because, in principle, a boy may wish to be a girl because of a similar, albeit inverted, perception of cultural disadvantage or bias (e.g., that girls get to wear dresses, are not yelled at as much, do not have to play rough, and so on). I concluded that, if such rationales were to be construed in cultural terms, the potential for bias should apply equally to both sexes (Zucker, 1992). I will provide a clinical vignette that illustrates how the cultural proviso could create a diagnostic dilemma:

A 5-year-old girl (IQ = 107) was referred for assessment by a relative, who was a health care professional. She lived with her mother and an older sibling. The biological father was deceased. At the time of assessment, the girl met all of the DSM-IV criteria for GID. Her mother was petrified to know why her daughter either insisted that she was a boy or that she had a strong desire to become one. Thus, she had never asked her daughter why she wanted to be a boy or thought that she was one. In the family interview, the mother was encouraged to do so and her daughter replied by stating: “Because I like boys’ underwear. Girls can’t wear boys’ underwear. That is why I want to be a boy.”

If one cast this remark under a cultural bias lens, one could make the argument that her perception that only boys can wear boys’ underwear would rule out the diagnosis of GID. Of course, one could always counter that the girl’s reasoning about why she desired to be a boy was not the “real” reason—that it was simply her own, idiosyncratic gendered social construction. This, however, moves into the realm of making causal assumptions, which the criteria are not intended (at least in theory) to address.

The DSM-IV Subcommittee on Gender Identity Disorders (Bradley et al., 1991) had taken the position that “it was inappropriate to place such an exclusion rule in the criteria themselves, as there may be many reasons why a child adopts a cross-gender identity, and that these issues should be dealt with in the text” (p. 326). Although differential diagnostic parameters continued to receive attention in the DSM-IV text, the decision adopted by the American Psychiatric Association was to harmonize (equalize) the cultural proviso for boys and girls in Criterion A.

Criteria for Cross-Dressing

Of the five indicators for GID in the Point A criterion (see Table 3), four were written in a manner such that they were

identical for boys and girls (A1, A3–A5). As noted by Langer and Martin (2004), the criterion for cross-dressing (A-2) has somewhat different wording for boys versus girls: “in boys, preference for cross-dressing or simulating female attire; in girls, insistence on wearing only stereotypical masculine clothing” (American Psychiatric Association, 2000, p. 581) (see also Morgan, 2000); however, Langer and Martin failed to note that there is an additional sex difference regarding the surface indicator of clothing preference. In the Point B criterion (see Table 3), there is the indicator, for girls, of a “marked aversion toward normative feminine clothing,” but there is no corresponding parallel indicator for boys (i.e., a marked aversion toward normative masculine clothing).

Regarding the Point A criterion, Langer and Martin interpreted the sex difference regarding the cross-dressing criterion as indicating a “lower diagnostic threshold for boys” (p. 8). In my view, the differences in the Point A criterion appear rather subtle and it is not clear why the wording is not similar for boys and girls (e.g., “in girls, preference for cross-dressing or simulating male attire”). My clinical hunch is that the slightly higher threshold for girls is related to the more general concern that the DSM makes clear that, for girls with GID, there is an extreme rejection of wearing culturally typical feminine clothing (as reflected in the Point B indicator). Clinically, there are some clear sex differences in the way that boys and girls manifest cross-dressing. Many boys with GID will not object to wearing culturally typical masculine clothing (e.g., pants and shirts) to school, but will resort to cross-dressing when the setting permits it (e.g., in the dress-up corner at nursery school, during fantasy play at home, etc.). In contrast, many girls with GID experience the wearing of culturally typical feminine clothing (e.g., dresses) as quite catastrophic and will refuse to wear them under any circumstances. Many parents of girls with GID report that one of its earliest indicators pertained to extreme anxiety and unease around clothing and hair-style. Thus, the preference for masculine clothing and the rejection of feminine clothing is often one of the most salient and emotionally charged surface indicators of gender dysphoria in young girls.

Revision of the Point A Criteria

Because the collapsing of the verbalized wish to be of the other sex with other behavioral indicators of cross-gender identification has received substantial criticism (see below), I will summarize here the rationale for it during the preparation phase for DSM-IV. Clinical opinion at that time was that some children who appeared to be struggling with their gender identity did not, at least at the time of a clinical evaluation, verbalize the desire to be of the other sex (Bradley et al., 1991). For example, it was argued that a boy who met the A2–A5 criteria and displayed an “aversion toward rough-and-tumble play and rejection of male stereotypical toys, games, and activities” (Criterion B) was unlikely to have a very positive sense of self as a boy. Clinical impression was that,

perhaps for social desirability reasons, such a boy might not overtly verbalize the desire to be a girl. The clinical opinion at that time was that this particular configuration might be particularly characteristic of older children.

Empirical evidence was then examined to justify the change for the Criterion A indicators. As reported in Zucker et al. (1998), factor analysis of 7 interviewer-rated items and 14 maternally-rated items pertaining to cross-gender identification from Green’s (1987) database of 66 feminine boys and 55 control boys identified a one-factor solution, containing 15 items with factor loadings $\geq .40$. One of these items, “Son states wish to be a girl,” had a factor loading of .61. The other 14 items had factor loadings ranging from .44–.84. It was thus argued that the wish to be of the other sex was simply one of a number of behaviors suggestive of cross-gender identification.

The conceptual notion that cross-gender identification has an underlying single-factor structure has received subsequent empirical support. Johnson et al. (2004), for example, factor-analyzed the 16-item GIQC (see above) in a large sample of gender-referred children ($N = 325$) and control children ($N = 504$). Johnson et al. found that two items pertaining to a cross-sex wish (Boy version: “He states the wish to be a girl or a woman” and “He states that he is a girl or a woman”) had factor loadings of .81 and .69, respectively, and loaded on a single, 14-factor solution. Eleven other items pertaining to cross-gender identification had factor loadings that ranged from .34–.91. A twelfth item (Boy version: “He talks about not liking his sexual anatomy (private parts)”) had a factor loading of .47.

In Green’s (1987) data set, there was empirical support for the hypothesis that the verbalized wish to be of the other sex was less common in older boys (9–12 years of age) than in younger boys (3–9 years of age) (Zucker et al., 1998). Zucker et al. then reexamined symptom ratings from parent interview data for 54 children seen at the Toronto Child and Adolescent Gender Identity Clinic who did not meet DSM-III criteria for GIDC. In this analysis, they assessed whether these children would meet the proposed Criterion A for DSM-IV with regard to the A2–A5 indicators (none of these 54 children had repeatedly verbalized the desire to be of the other sex). For the 54 children, the mean number of indicators rated as present was 2.36 ($SD = 1.33$; range, 0–4). Of the 54 children, 16 (29.6%) had all four indicators and thus would meet the proposed threshold for Criterion A. The subgroup that now met the threshold was compared with the subgroup that did not with regard to the demographic variables of age, IQ, and parent’s social class and marital status. There was a trend for the children who were at threshold for Criterion A to be younger than the children who were not at $p = .087$, two-tailed. None of the other demographic variables significantly distinguished the two subgroups.

Zucker et al. (1998) concluded that the revised criteria resulted in a modest increase in “diagnosed” cases; however, they also noted that this increase was likely an overestimate as the Criterion B indicators were not examined, including

the “aversion toward rough-and-tumble play and rejection of male stereotypical toys, games, and activities.” If anything, it would be likely that some of the children who met the threshold for Criterion A would not meet the threshold for Criterion B.

Congflation of Cross-Gender Behavior and Gender Dysphoria

In many respects, cross-gender identification and GID proper can only be understood in a social and phenomenological (i.e., subjective) context (Money, 1994; Zucker, 1999). Apart from any biological predisposition that underlies both normative and atypical gender development, children construct a gender identity based on information that they glean from the social environment. Cognitive-developmental gender theorists, for example, suggest that once children become aware of a “two-gendered” social world and develop the capacity for gender identity self-labeling they then scan their environment for information about “what boys do” and “what girls do” and then often adopt behavioral patterns that are consistent with their own gender identity (Martin & Ruble, 2004; Martin, Ruble, & Szkrybalo, 2002).

In my view, the A2–A5 behavioral indicators of cross-gender identification adopted in the DSM-IV were framed in relation to what is known about normative or typical gender development. The core behavioral attributes that constitute these indicators (dress-up play, fantasy role play, toy and activity preferences, and sex-of-playmate preference) rest on the assumption that they are, on average, sex-dimorphic, i.e., they show significant differences between typical boys and girls. As one example: boys with cross-gender identification or GID proper who adopt cross-gender roles in fantasy play (e.g., emulating various female characters—mother, sister, Snow White, The Little Mermaid, Batgirl, Princess Leia or Asajj Ventress from Star Wars, etc.) presumably do so, in part, because, on average, girls are more likely to adopt such role choices than boys. Johnson et al. (2004) found strong evidence for this on the parent-report GIQC. For the item, “In playing ‘mother/father,’ ‘house,’ or ‘school games,’” 92.6% (188/203) of control boys were judged to be “usually a boy or man” or “a boy or man at all times” whereas 95.5% (171/179) of controls girls were judged to be “usually a girl or woman” or “a girl or woman at all times.” In contrast, 61.9% of gender-referred boys were judged to be “usually a girl or woman” or “a girl or woman at all times” and 68.4% of gender-referred girls were judged to be “usually a boy or man” or “a boy or man at all times.” For the control boys versus girls, the effect size was 3.60 (my analysis).

Critics of the DSM-IV Point A criteria have argued that they inappropriately condense cross-gender identity (the desire to be of the other sex), as reflected in A1, and pervasive cross-gender role behaviors, as reflected in A2–A5 (Bartlett, Vasey, & Bukowski, 2000; Bockting & Ehrbar, 2005; Bryant, 2007; Corbett, 1996, 1998; Haldeman, 2000; Hill, Rozanski, Carfagnini, &

Willoughby, 2007; Moore, 2002; Richardson, 1996, 1999; Wilson, Griffin, & Wren, 2002).² Thus, it has been claimed that the Point A criterion blurs the distinction between a child who has both a cross-gender identity and pervasive cross-gender behavior and a child who merely shows signs of pervasive cross-gendered behavior (in descriptive terms, the “gender nonconforming” or “gender-variant” child). As a result, there is the concern that children might be inappropriately diagnosed with GID simply because they meet the A2–A5 criteria. (As an aside, in clinical practice, it is quite rare to assess a child who shows only signs of a cross-gender identity (A1) in the absence of pervasive cross-gender behavior (A2–A5).)³

Based on this criticism, Bartlett et al. (2000) suggested that the Point A criteria might capture two subgroups of children. In appraising the Zucker et al. (1998) data discussed earlier, Bartlett et al. surmised that “...the data might be better viewed as reflective of a common [co-occurrence] of cross-sex wishes and cross-gender behaviors, but not a complete overlap...perhaps those children who express cross-sex wishes may be expected to also exhibit cross-gender behaviors, though children who exhibit cross-gender behaviors may not necessarily be expected to also experience the desire to be the other sex” (p. 758). Bartlett et al.’s suggestion would lead to the following two hypotheses: (1) gender-referred children would, on average, show more cross-gender behavior than that of control children, regardless of whether or not they expressed the desire to be of the other sex; (2) the degree of cross-gender role behavior of gender-referred children would vary as a function of their verbalized desire to be of the other sex.

To test these hypotheses, I re-analyzed data from the GIQC for 438 gender-referred children (359 boys, 79 girls) and 807 control children (504 boys, 303 girls) seen in my clinic. For the gender-referred children, I partitioned them into five subgroups based on their stated desire to be of the other sex (GIQC Item 13), ranging from “every day” to “never.” Then, I calculated a

² Bryant (2006, pp. 31–33) has provided an interesting historical perspective on the GIDC diagnosis prior to its formal appearance in the DSM-III. Using unpublished correspondence (for details, see Bryant, 2007), much channeled through the office of Robert L. Spitzer, Bryant has shown that the debate regarding the distinction between cross-gender identity and cross-gender role behavior was apparently a key “behind the scenes” issue in the 1970s. Prior to the adoption of the GIDC name for the diagnosis, other naming options had been proposed, including Psychosexual Identity Disorder, Gender Role Disorder of Childhood, and Gender Identity or Role Disorder of Childhood. At that time, feedback given to the Psychosexual Disorders Committee (which apparently had a subcommittee called the Gender Role Disorders Committee or the Gender Identity/Role Disorders Committee) included the concern that the proposed diagnostic criteria did not adequately distinguish between cross-gender role behaviors and a cross-gender identity proper (see also Bryant, 2007, 2008).

³ Other terms to describe children who might meet the DSM-IV criteria for GID are “girlyboys” (e.g., Corbett, 1996; Ehrensaft, 2007), which seems to have supplanted the older term of “sissy boys” (Green, 1987), “gender-dissonant” (Vanderburgh, 2009), and “transgender” children (Brill & Pepper, 2008).

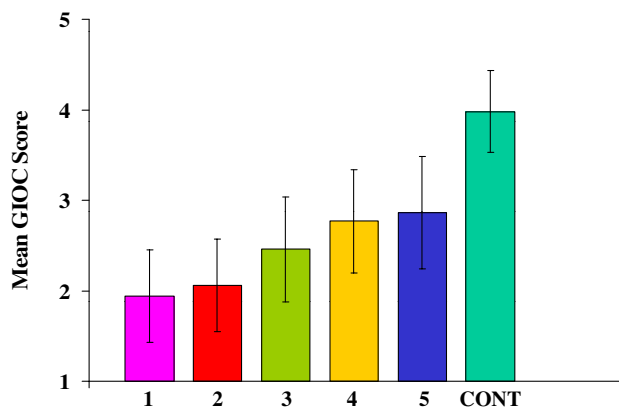


Fig. 1 Maternal ratings of cross-gender behavior on the GIQC as a function of Item 13 (stated desire to be of the other sex). A lower score indicates more cross-gender behavior. *Note:* On the horizontal axis, 1 = “every day” ($N = 23$); 2 = “frequently” ($N = 85$); 3 = “once-in-a-while” ($N = 158$); 4 = “very rarely” ($N = 83$); 5 = “never” ($N = 89$). For the controls, $N = 807$

revised GIQC mean score based only on the 11 GIQC items (Items 1–7, 9–12) pertaining to cross-gender role behavior (or same-gender role behavior) that had acceptable factor loadings on the factor described in Johnson et al. (2004).

Figure 1 shows the mean revised GIQC score as a function of the verbalized wish to be of the other sex. For reference purposes, the mean revised GIQC score of the control children is also shown in the figure. It can be seen in Fig. 1 that the gender-referred children had, on average, significantly more cross-gender role behavior than did the control children. Even the subgroup of gender-referred children who did not verbalize the wish to be of the other sex had significantly more maternally-rated gender-atypical behavior than did the control children, $t(894) = 20.78$, $p < .001$ ($d = 2.32$). Thus, this finding supports the first hypothesis advanced by Bartlett et al. (2000).

For the gender-referred children alone, I then calculated a 2 (Sex) \times 5 (Wish) analysis of covariance (ANCOVA). Age was covaried because the children who, by maternal report, “never” verbalized the wish to be of the other sex were significantly older than the children who “frequently” verbalized the wish ($p < .05$). None of the other paired age contrasts differed significantly. It can be seen in Fig. 1 that the degree of cross-gender role behavior showed a very clear linear relation to the frequency of the verbalized cross-sex wish. The formal statistical test via ANCOVA yielded a significant main effect for Wish, $F(4, 432) = 30.90$, $p < .001$. Duncan’s post hoc tests showed that children who verbalized the desire to be of the other sex “every day” or “frequently” had, on average, significantly more cross-gender role behavior on the GIQC than the children who verbalized a cross-sex wish “once-in-a-while,” “rarely,” or “never.” The children who verbalized a cross-sex wish “once-in-a-while” had, on average, significantly more cross-gender role behavior than the children who either “rarely” or “never” verbalized the wish (all $ps < .05$). The latter two subgroups

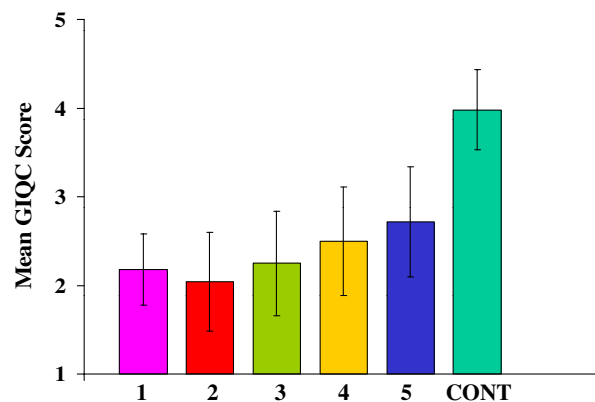


Fig. 2 Maternal ratings of cross-gender behavior on the GIQC as a function of Item 14 (insistence that one is of the other sex). A lower score indicates more cross-gender behavior. *Note:* On the horizontal axis, 1 = “every day” ($N = 14$); 2 = “frequently” ($N = 54$); 3 = “once-in-a-while” ($N = 83$); 4 = “very rarely” ($N = 69$); 5 = “never” ($N = 218$). For the controls, $N = 807$

did not differ significantly in their degree of cross-gender role behavior.

I conducted a similar analysis as a function of the GIQC item pertaining to the child’s insistence that he or she was a member of the other sex (Fig. 2). A 2 (Sex) \times 5 (Insistence) ANCOVA yielded a significant main effect for Insistence, $F(4, 432) = 19.61$, $p < .001$. Duncan’s post hoc tests showed that the children who insisted that they were the other sex either “every day” or “frequently” had, on average, significantly more cross-gender role behavior than children who “very rarely” or “never” insisted as such, but their mean score did not differ significantly from the children who verbalized such a statement “once-in-a-while.” The latter group did not differ significantly from the children who verbalized such a statement “very rarely” but had, on average, significantly more cross-gender role behavior than the children who “never” verbalized this remark. The “very rarely” and “never” subgroups did not differ significantly from each other (all significant p values $< .05$).

In my view, these two analyses support Bartlett et al.’s (2000) second hypothesis, namely that the degree of cross-gender role behavior among gender-referred children is related to the frequency with which they express the desire to be of the other sex (or the insistence that they are a member of the other sex).

In the analyses conducted so far, it could be argued that a methodological constraint is that the informant (i.e., the mother) was rating both the gender-role items and the two items pertaining to the verbalized desire to be, or insistence that one is, of the other sex. Thus, there is, perhaps, the problem of a “halo” effect, i.e., the higher one rates one class of behaviors, the higher one would rate the other class. In Johnson et al. (2004), it was, however, reported that the mother–father correlation for the total GIQC score was substantial, at $r = .90$, which, perhaps lends some confidence in the accuracy of the maternal ratings.

To explore the empirical issue further, I examined the relation between maternal ratings of the desire to be of the other sex and child report of gender identity confusion on the Gender Identity Interview for Children, a structured questionnaire schedule (Wallien et al., 2009; Zucker et al., 1993). The GIIC consists of 12-items, each rated on a 3-point response scale. The GIIC has been shown in two independent analyses, including one confirmatory factor analysis (CFA), to have a two-factor solution, consisting of 4 items labeled as Cognitive Gender Confusion and 8 items labeled as Affective Gender Confusion. For the purposes of this analysis, the unit-weighted sum score for all 12 items was calculated. In this analysis, there were GIIC scores for 332 gender-referred boys and 75 gender-referred girls.

Figure 3 shows the mean GIIC score as a function of the verbalized wish to be of the other sex. With age covaried, a 2 (Sex) \times 5 (Wish) ANCOVA yielded a significant main effect for Wish, $F(4, 401) = 22.97, p < .001$. Like the GIQC data, there was a clear linear relationship between these two parameters. For example, children who verbalized the wish to be of the other sex “every day” had a significantly higher GIIC mean score than the other four subgroups. All paired contrasts were statistically significant ($p < .05$), except the comparison between the children who “very rarely” verbalized the desire to be of the other sex and those who “never” verbalized such a desire.

I conducted a similar analysis as a function of the GIQC item pertaining to the child’s insistence that he or she was a member of the other sex (Fig. 4). A 2 (Sex) \times 5 (Insistence) ANCOVA yielded a significant main effect for Insistence, $F(4, 432) = 19.61, p < .001$. Children who insisted that they were of the other sex “frequently” or “every day” had a significantly higher GIIC mean score than the other three groups; the children who insisted they were of the other sex “once-in-a-while” had a significantly higher GIIC mean score than the children who “never” made such remarks (all $ps < .05$).

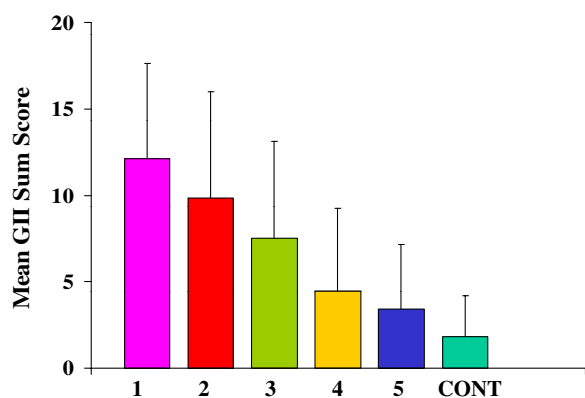


Fig. 3 Gender Identity Interview Sum Score as a function of Item 13 (stated desire to be of the other sex) on the GIQC. On the horizontal axis, 1 = “every day” ($N = 22$); 2 = “frequently” ($N = 80$); 3 = “once-in-a-while” ($N = 149$); 4 = “very rarely” ($N = 76$); 5 = “never” ($N = 80$). For the controls, $N = 173$ (GIIC data for the gender-referred probands from Wallien et al. (2009) and Zucker et al. (1993); control data from Wallien et al. (2009)). On the GIIC, absolute range is 0–24

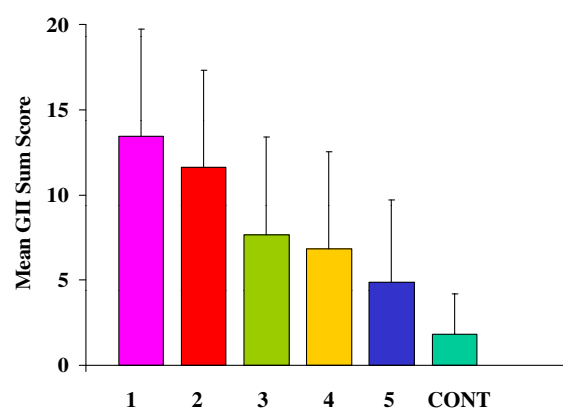


Fig. 4 Gender Identity Interview Sum Score as a Function of Item 14 (insistence that one is of the other sex) on the GIQC. On the horizontal axis, 1 = “every day” ($N = 13$); 2 = “frequently” ($N = 49$); 3 = “once-in-a-while” ($N = 81$); 4 = “very rarely” ($N = 66$); 5 = “never” ($N = 198$). For the controls, $N = 173$ (GIIC data for the gender-referred probands from Wallien et al. (2009) and Zucker et al. (1993); control data from Wallien et al. (2009)). On the GIIC, absolute range is 0–24

These results showed a convergence between the degree to which the mothers perceived their children to express the desire to be, or insistence that they are, a member of the other sex and the degree of child-reported gender identity confusion/dysphoria on the GIIC.

Reliability and Validity

Since the DSM-IV criteria for GID in children were published, there have been no formal reliability studies of the GID diagnosis for children. By this, I mean that there have been no studies that have reported inter-clinician agreement on the diagnosis. This is a serious deficiency in the literature. If the GID diagnosis for children is to remain in the DSM-V, it would be important to conduct field trials that establish the diagnostic reliability of the criteria, however they are formulated.

As was the case for pre-DSM-IV cohorts, a number of studies have demonstrated reasonable evidence of discriminant validity. There have been at least seven such studies in the DSM-IV era (Chiu et al., 2006; Cohen-Kettenis, Owen, Kaijser, Bradley, & Zucker, 2003; Cohen-Kettenis et al., 2006; Fridell, Owen-Anderson, Johnson, Bradley, & Zucker, 2006; Johnson et al., 2004; Wallien, Veenstra, Kreukels, & Cohen-Kettenis, 2009; Wallien et al., 2009). In some of these studies, data on specificity and sensitivity have been examined. In general, it has been argued that specific psychometric measures should have a high threshold for specificity, that is, to have a low rate of false positives for controls. In Johnson et al. (2004), the specificity rate was set at 95%, which yielded a sensitivity rate of 86.8% for the gender-referred probands (which improved to 96.3% when only probands who were threshold for the GID diagnosis were examined).

Threshold Versus Subthreshold Comparative Analyses

Within clinic-referred samples of gender-referred children, the majority have been deemed to meet the complete DSM criteria for GID based on clinician diagnosis. For example, in a cross-clinic, cross-national study of gender-referred children (total $N = 488$) in Toronto and Utrecht, The Netherlands (Cohen-Kettenis et al., 2003), the percentage who met the complete DSM criteria for GID was 67.0%. Clinically, it has been noted that the majority of subthreshold cases likely met the complete criteria at a younger age, but not at the time of assessment (Zucker & Bradley, 1995).

As noted earlier, some critics have expressed concern that the DSM criteria may not adequately differentiate children with GID from those children who merely show a pattern of extreme “gender nonconforming” behavior but who are not “truly” GID (e.g., Corbett, 1996; Haldeman, 2000; J alas, 2003; Richardson, 1996). Haldeman (2000), for example, claimed that “...it is conceivable that a child could be diagnosed with GID exclusively on the basis of preference for gender atypical activities or play objects” (p. 195) or that “any boy who, for example, displays an even passing interest in art, music or cooking could, conceivably, be diagnosed as GID...” (p. 198). Neither of these assertions are likely to occur and, to date, critics of the diagnostic criteria have not provided an empirical demonstration of systematic inaccurate diagnosis (Zucker, 2001). In contrast, one analogue-vignette study found that clinicians were prone to “profound underdiagnosis” of GID, i.e., they did not make the diagnosis even when the vignette included information that was consistent with the DSM-IV criteria as currently formulated (Ehrbar, Witty, Ehrbar, & Bockting, 2008).⁴

Comparative analysis of threshold versus subthreshold cases is important for two reasons. First, using external measures, it can indicate whether or not the DSM criteria reliably distinguish between these two diagnostic subgroups; in other words, the central issue is one of identifying the boundary of a psychiatric disorder (cf. Kendler, 1999). Second, if there is evidence that a valid distinction can be made, one can evaluate whether or not the subgroups differ in other ways, such as variation in long-term developmental trajectories, putative etiological factors, and so on.

⁴ At least one clinician (Pleak, 1999) indicates that he does not use the GID diagnosis because of its potential for stigma. The relation between stigma and psychiatric diagnosis for children is beyond the scope of this review; however, it is important to point out that a psychiatric “label” can have positive (or stigma-reducing) effects and is not uniformly negative (stigma-enhancing) (see, e.g., Walker, Coleman, Lee, Squire, & Friesen, 2008). I also suggest Clausen (1981) as an excellent review essay that articulates well the complex literature on psychiatric diagnosis and stigma. My own view on stigma runs something like this: When children with GID are socially ostracized by their peers, it is their overt behavior that elicits negative reactions (see, e.g., Fridell, 2001), not an abstract label (see, e.g., Law, Sinclair, & Fraser, 2007). There is considerable evidence that, even in normative samples of children, cross-gender behavior is appraised negatively by the peer group, and more so in boys than it is in girls (Zucker, Wilson-Smith, Kurita, & Stern, 1995).

In one study in which DSM-III-R criteria were used, Zucker and Bradley (1995) reported that the children who met the complete criteria for GID ($n = 113$) were significantly younger, of a higher social class background, and more likely to come from an intact, two-parent family than the children who were subthreshold for GID ($n = 80$). The two subgroups did not differ significantly with regard to sex composition and IQ. To test which variables, if any, contributed to the correct classification of the subjects in the two diagnostic groups, a discriminant function analysis was performed. Age, sex, IQ, and marital status contributed to the discriminant function, with age showing the greatest power. In the threshold group, 82.6% were correctly classified and, in the subthreshold group, 68.8% were correctly classified.

Several data sets have examined whether or not the two diagnostic subgroups differed on various measures of sex-typed behavior. As summarized in Zucker and Bradley (1995), the threshold group showed significantly more cross-gender behavior or less same-gender behavior than the subthreshold group on 11 of 17 measures, even after controlling for the demographic variables that also differed between the two subgroups (see also Zucker et al., 1984, summarized earlier).

More recent studies, largely using DSM-IV criteria, have continued to document significant differences between threshold and subthreshold cases. Johnson et al. (2004) found that the subthreshold group ($n = 109$) had a mean score on the GIQC that was intermediate between that of the threshold cases ($n = 216$) and the controls. There was, however, clear evidence that the subthreshold group was “gender nonconforming” in that the effect size between their mean score and that of the controls was substantial (Cohen’s d ranged from 1.44 to 3.28 when blocked by age groups [e.g., 3–5 years, 5–6 years, etc.]). In a sample of gender-referred children from Utrecht, Cohen-Kettenis et al. (2006) also found that the threshold cases ($n = 114$) had a significantly more deviant score on the GIQC than did the subthreshold cases ($n = 42$).

In my view, a particularly important study pertaining to the threshold–subthreshold distinction is that of Wallien et al. (2009). Wallien et al. conducted a CFA on the GIIC in a sample of 329 gender-referred children from Toronto, 228 gender-referred children from Amsterdam, and 173 control children from Toronto. The CFA documented the two-factor solution originally reported by Zucker et al. (1993). Both groups of gender-referred children had, on average, a significantly higher score on the GIIC than did the control children, indicating more gender identity confusion (Toronto-control effect size: 2.15; Amsterdam-control effect size: 3.46). More importantly for the present discussion, the threshold cases had a significantly higher GIIC sum score ($M, 9.58; SD = 5.70; N = 397$) than did the subthreshold cases ($M, 4.68; SD = 4.18; N = 160$). Not surprisingly, the sensitivity rates were higher for the clients who met the complete DSM criteria for GID than for the clients who were subthreshold for the diagnosis.

Taken together, these data suggest that, even within a population of gender-referred children, the DSM criteria, when used categorically (threshold versus subthreshold), significantly differentiate the behavior of the subgroups on external measures. There are, however, limitations to these kinds of analyses that should be acknowledged. For example, different combinations of the Point A and Point B criteria could (and probably did) result in a child meeting the complete criteria for GID (e.g., such combinations could include children who met A1 through A5 versus A2–A5 or even A1 and one of three combinations of A2–A5). As well, these studies did not report how many indicators of the criteria were met for the children who were judged subthreshold for the diagnosis. Nonetheless, the fact that in these various studies the subthreshold cases fall in-between that of the threshold cases and controls on external measures is exactly what one would have predicted (cf. Bartlett et al., 2000).

Anatomic Dysphoria

In adolescents and adults with GID, discomfort with the external, somatic indicators of one's phenotypic biological sex (both primary and secondary sex characteristics) are particularly salient with regard to the client's felt sense of gender dysphoria. In some respects, this is the *sine qua non* of the developmental end-state of gender dysphoria. Much less is known about the salience of anatomic dysphoria in children with GID (Coates, 1985; Lothstein, 1992). Other than one general item on the GIQC that pertains to anatomic dysphoria (Johnson et al., 2004), I am not aware of any other published empirical data on putative indicators of this construct. This represents a significant gap in validity research pertaining to the DSM criteria for GID in children.

Lambert (2009) assessed body image in 28 boys with GID, 23 clinical control boys, and 25 non-referred boys (M age, 8.34 years; $SD = 2.52$).⁵ The boys completed two self-report measures pertaining to general body image satisfaction. On the Body Esteem Scale for Children (BES), the GID boys endorsed, on average, significantly more body dissatisfaction than the non-referred boys ($p < .05$). The mean BES scores of the clinical control boys did not differ significantly from either the GID group or the non-referred group. On the Physical Feature Satisfaction Scale (PFSS), there was a borderline main effect for group. The GID boys endorsed significantly less body part satisfaction than the non-referred group. The mean PFSS scores of the clinical control boys did not differ significantly from either the GID group or the non-referred group.

Lambert (2009) also administered to the mothers of the three groups a 31-item Anatomic Satisfaction Scale (ASS). At the level of content or face validity, items were intended to reflect either general body image issues (e.g., "Does he say that he is ugly?") or gender-specific anatomic dysphoria (e.g., "Does he

say that he wants to get rid of his penis?"). Each item was rated on a 3-point response scale (No, Sometimes, or Yes) using the past 12 months as a time frame.

Because these data are novel and because they may provide leads in potential field trials, I provide the questionnaire in Appendix. Table 4 shows item level descriptive statistics as a function of group. At the item level, a preliminary analysis showed no significant difference on any of the items between the two control groups. Table 5 shows, for each item, the results of chi-square analyses that compared the GID group with the two control groups. Columns 3–4 show the results in which the response option of No was compared to the response options of Sometimes or Yes combined. Of the 31 items, there were 17 significant group effects, all of which showed greater endorsement of body image concerns in the GID group. Some of the significant contrasts pertained to general body image concerns and others pertained to gender-specific anatomic dysphoria. For example, regarding Item 5 ("Does he say that he is ugly?"), 42% of the mothers of GID boys endorsed either a Sometimes or a Yes response, compared to 15% of the mothers of the control boys ($p = .014$). Regarding Item 9 ("Does he say that he wants to get rid of his penis?"), 13% of the mothers of GID boys endorsed either a Sometimes or a Yes response, compared to 0% of the mothers of control boys ($p = .046$). Similarly, regarding Item 6 ("Does he say that he would like breasts?"), 16% of the mothers of GID boys endorsed either a Sometimes or a Yes response, compared to 0% of the mothers of the control boys ($p = .016$). Other items, however, pertaining to gender-specific anatomic dysphoria showed no significant differences among the three groups (e.g., Item 16: "Does he pretend that he has a vagina?").

A principal axis factor analysis with varimax rotation identified 12 items on the ASS that loaded on a general body image factor and 8 items that loaded on a gender-specific anatomic dysphoria factor. Unit-weighted factor scores significantly differentiated the GID boys from the control boys. For the general body image factor, Cohen's $d = 1.56$ and for the gender-specific anatomic dysphoria factor, Cohen's $d = 3.92$, using the SD of the control group.

Although preliminary, these data, particularly at the item level, may provide leads for further investigation in field trials regarding potential markers of gender-specific anatomic dysphoria in children.

Distress and Impairment

Critics who reject the GID diagnosis *in toto* have adopted alternative language to label children who display various degrees of cross-gender behavior and identity. One such label is to characterize such children as "gender nonconforming" (Pickstone-Taylor, 2003); another label that has received a fair bit of recent currency is to characterize them as "gender variant" (Lev, 2004; Menvielle, Tuerk, & Perrin, 2005). The *Oxford Dictionary* defines variant as "a form or version that varies from

⁵ These data come from a doctoral dissertation that I supervised.

other forms of the same thing.” Variation is defined as “a change or slight difference in condition, amount, or level.” Variance is defined as “the amount by which something changes or is different from something else.”

By definition, then, it is descriptively accurate to characterize children who meet the GID criteria as they are currently formulated as gender-variant (indeed, any child whose behavior or identity departs from some hypothetical mean gold standard could be characterized as gender variant). The deeper philosophical (and, perhaps, empirical) debate is whether or not one can demarcate a distinction between variance and disorder. Three decades ago, Meyer-Bahlburg (1985) characterized this distinction as the “zone of transition between clinically significant cross-gender behavior and mere statistical deviations from the gender norm” (p. 682).

Distress and impairment have come to occupy a critical position in the DSM. Originally, these constructs were used to formulate a working definition of mental disorder for the DSM-III (see Spitzer & Endicott, 1978), to set some kind of boundary between disorder and variation from the norm. As noted earlier, these constructs became part of the diagnostic criteria for about half of the DSM-IV diagnoses, in what has been called the clinical significance criterion. As I understand it, one reason this occurred was because there was a concern that the prevalence of some disorders as identified in epidemiological studies appeared to be “too high” and some researchers could document that prevalence was reduced if an impairment or distress criterion was required (for children, see, e.g., Canino et al., 2004).

Regarding the distress/impairment criterion for GID, there are two key issues: (1) How should these constructs be assessed? (2) Is the source of the distress or impairment “in the person” or is it simply secondary to social ostracism? Regarding the latter, critics of the diagnosis (op. cit.) have largely favored the latter interpretation (for a further discussion of this, see Zucker, 2005b).

If one considers the developmental end-state of GID, i.e., its mature form as expressed during adolescence and adulthood, I would argue that distress is manifested most acutely in the form of the disjunction between the client’s felt psychological gender identity and phenotypic sex (in children, per their awareness that they have an “assigned” sex). Many years ago, Fisk (1973) coined the term gender dysphoria to characterize the sense of awkwardness or discomfort in the anatomically congruent gender role and the desire to possess the body of the other sex, together with the negative affect associated with these feelings and desires. Clinically, it has been used to refer to the range of individuals who, at one time or another, experience sufficient discomfort with their assigned sex to form the wish for sex reassignment. It is this disjunction that often leads clients to seek out clinical care and treatment.

Although children with GID may experience some sense of discomfort with their sexual anatomy, as suggested by Lambert’s (2009) data, it is unlikely that this anatomic dysphoria is at

the core of distress, particularly in young children. In my opinion, the construct of distress is probably better understood, at least at the surface level, in relation to a child’s verbalized sense of unhappiness about being a boy or a girl, as expressed most concretely by remarks about wanting to be of the other sex. Regarding the clinical significance criterion (Table 3), it remains unclear how distress is to be inferred independently of the clinical indicators in Criterion A and Criterion B. I will note here that this is a conceptual problem that is not unique to GID. For example, the same problem is present for the diagnosis of Separation Anxiety Disorder in children.

Regarding impairment, the DSM-IV refers to difficulties in social, occupational, or other important areas of functioning. For some DSM diagnoses, evidence for impairment can be relatively easy to infer (e.g., a person with a dysthymic disorder who is unable to work). In other instances, impairment appears to be less clearly distinguishable from the clinical signs of disorder, as in the case of Conduct Disorder.

Regarding impairment and GID, one line of evidence might be to consider the presence of associated psychopathology (for a brief review, see Lawrence, 2008, pp. 437–439). If, for example, adolescents and adults have elevated rates of “other” forms of psychopathology, does this constitute evidence for impairment? On this point, there are different views. For example, is the associated psychopathology a result of the distress that accompanies GID (and its attendant impact on psychosocial well-being) or is it simply secondary to the experience of social ostracism (see, e.g., Nuttbrock et al., 2009; Zucker, 2008a)? If it is the result of the latter, then it would be arguable to consider this as satisfying an “in-the-person” definition of impairment.

The same interpretive matters apply to children with GID (Zucker, 2008a). There is reasonable evidence that children with GID have, on average, higher rates of behavior problems compared to non-referred children (Cohen-Kettenis et al., 2003; Zucker, 2008a; Zucker & Bradley, 1995). It has been demonstrated that such associated psychopathology can be predicted, in part, by social ostracism parameters (Cohen-Kettenis et al., 2003). Thus, it could be argued, quite reasonably, that this form of impairment is a by-product of stigma and not “in-the-person” per se. On the other hand, the DSM is not entirely clear with regard to the phrase “[t]he disturbance *causes* clinically significant distress or impairment in social, occupational, or other important areas of functioning” (my emphasis) (on this point, see also Bartlett et al., 2000). One could, for example, argue that the behaviors associated with GID “cause” impairment because of social ostracism, but I am not sure that this is what the DSM intends in its conceptual formulation of impairment.

As an aside, it should be pointed out that efforts to measure impairment, in general, are vulnerable to similar difficulties in interpretation. Consider, for example, items used to measure impairment by Canino et al. (2004): How much of a

Table 4 Descriptive statistics (in percent) for each item on the Anatomic Satisfaction Scale as a function of group

Item	GID						Clinical control						Community control					
	No		Sometimes		Yes		No		Sometimes		Yes		No		Sometimes		Yes	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
1	11	36	5	16	15	48	14	61	3	13	6	26	19	76	2	8	4	16
2	22	71	5	16	4	13	13	57	1	4	9	39	9	36	2	8	14	56
3	27	87	2	7	2	7	23	100	0	0	0	0	24	96	0	0	1	4
4	30	97	0	0	1	3	23	100	0	0	0	0	24	96	1	4	0	0
5	18	58	6	19	7	23	19	83	2	9	2	9	22	88	3	12	0	0
6	26	84	3	10	2	7	23	100	0	0	0	0	25	100	0	0	0	0
7	21	68	9	29	1	3	19	83	4	17	0	0	21	84	4	16	0	0
8	17	55	5	16	9	29	22	96	0	0	1	4	25	100	0	0	0	0
9	27	87	2	7	2	7	23	100	0	0	0	0	25	100	0	0	0	0
10	24	77	2	7	5	16	20	87	1	4	2	9	21	84	2	8	2	8
11	27	87	3	10	1	3	10	44	4	17	9	39	16	64	4	16	5	20
12	5	16	5	16	21	68	4	17	4	17	15	65	6	24	6	24	13	52
13	22	71	3	10	6	19	18	78	1	4	4	17	22	88	0	0	3	12
14	31	100	0	0	0	0	23	100	0	0	0	0	23	92	1	1	1	1
15	24	77	3	10	4	13	22	96	0	0	1	4	25	100	0	0	0	0
16	30	97	1	3	0	0	23	100	0	0	0	0	24	96	1	4	0	0
17	28	90	1	3	2	7	23	100	0	0	0	0	25	100	0	0	0	0
18	27	87	4	13	0	0	21	91	1	4	1	4	24	96	1	4	0	0
19	28	90	2	7	1	3	21	91	1	4	1	4	22	88	3	12	0	0
20	23	74	5	16	3	10	22	96	1	4	0	0	25	100	0	0	0	0
21	29	94	1	3	1	3	17	74	4	17	2	9	18	72	1	4	6	24
22	19	61	5	16	7	23	22	96	0	0	1	4	24	96	1	4	0	0
23	25	81	1	3	5	16	21	91	1	4	1	4	25	100	0	0	0	0
24	19	61	5	16	7	23	19	83	3	13	1	4	23	92	2	8	0	0
25	3	10	8	26	20	65	2	9	1	4	20	87	2	8	1	4	22	88
26	20	65	5	16	6	19	19	83	2	9	2	9	21	84	3	12	1	4
27	27	87	3	10	1	3	22	96	0	0	1	4	23	92	1	4	1	4
28	25	81	2	7	4	13	13	57	2	9	8	95	13	52	6	24	6	24
29	11	36	9	29	11	36	16	70	6	26	1	4	22	88	2	8	1	4
30	19	61	5	16	7	23	15	65	7	30	1	4	20	80	2	8	3	12
31	20	65	7	23	4	13	23	100	0	0	0	0	23	92	1	4	1	4

Note: Data from Lambert (2009)

problem does he/she have: (1) with feeling nervous or afraid? (2) getting along with his/her brothers/sisters? (3) getting along with other kids his/her age? (4) getting along with you [the mother]? (5) getting along with his/her father? (6) feeling unhappy or sad? In all instances, it is not entirely clear if positive responses to such questions would constitute evidence for “in-the-person” impairment or as secondary to social responses to deviant behaviors.

In summary, the constructs of distress and impairment require a great deal of further consideration in terms of how they should best be operationalized and measured in children with GID (and to children in general).

Harmonizing Descriptors

In the DSM, various adverbs or adjectives are used to emphasize for the clinician that an indicator or symptom represents an enduring pattern of behavior, not a transitory one. For the diagnosis of Attention-Deficit/Hyperactivity Disorder, for example, each of 18 possible symptoms is prefaced by the adverb “often.” The same adverb is used for each of the 8 possible symptoms of Oppositional Defiant Disorder. Thus, the clinician only needs to decide what counts as “often.” Of course, for these diagnoses, the clinician must also make a judgment about what counts as (the lack) of “close

Table 5 Results of chi-square analysis for each item on the ASS as a function of group and response choice

Item	No, Sometimes, or Yes ^a		No vs. Sometimes and Yes ^b	
	χ^2	<i>p</i>	χ^2	<i>p</i>
1	8.75	.013	7.15	.007
2	10.71	.005	3.85	.050
3	4.28	ns	2.11	ns
4	2.20	ns	<1	ns
5	8.57	.014	6.07	.014
6	8.27	.016	5.76	.016
7	3.48	ns	1.79	ns
8	22.83	.001	20.00	.001
9	6.52	.038	4.11	.046
10	1.15	.562	<1	ns
11	10.38	.006	7.81	.005
12	<1	ns	<1	ns
13	2.77	.250	1.05	ns
14	1.33	.516	<1	ns
15	9.01	.011	6.58	.010
16	<1	ns	<1	ns
17	4.83	.089	2.54	ns
18	2.63	ns	<1	ns
19	<1	ns	<1	ns
20	10.73	.005	8.28	.004
21	5.26	.072	3.95	.047
22	15.44	.001	13.13	.001
23	5.47	.065	3.25	.071
24	9.98	.007	5.94	.015
25	8.27	.016	<1	ns
26	4.20	ns	2.69	ns
27	2.28	ns	<1	ns
28	5.79	.055	4.67	.031
29	18.36	.001	13.46	.001
30	3.19	ns	<1	ns
31	13.51	.001	11.25	.001

Note: Data from Lambert (2009)

^a Three response choices

^b Two response choices

attention,” “difficulty sustaining attention” or “careless” mistakes, etc., which is a somewhat different matter.

In the case of GID, various qualifiers are employed (e.g., “repeatedly,” “insistence,” “strong and persistent,” “intense,” “strong,” “marked,” etc.). It is not entirely clear why, in the DSM-IV, these variations on the same theme were employed and if such semantic nuance weakens reliability in clinician judgment. There is, of course, also the translational problem for the clinician in deciding on what counts as “repeatedly,” “strong,” etc. Consider, for example, the A1

criterion, in which the child must be judged to have a “repeatedly” stated desire to be, or insistence that he or she is, the other sex. What counts as “repeatedly”? On the GIQC Item 13 reported by Johnson et al. (2004), 18.4% of gender-referred boys and 14.0% of gender-referred girls were rated by their mothers as verbalizing the wish to be of the other sex “very rarely.” The corresponding percentages for control boys and girls were 3.3% and 5.6%, respectively. It is likely that a clinician would not judge a “very rarely” response to be commensurate with “repeatedly.” In contrast, 22.1% of gender-referred boys and 44.0% of gender-referred girls were rated by their mothers as verbalizing the wish to be of the other sex “frequently” or “every day” in contrast to 0% of control boys and girls. One would suspect that such ratings would be deemed commensurate with the “repeatedly” descriptor. But what about the intermediate response option of “once-in-a-while”? This option was selected by 36.4% of the mothers of the gender-referred boys and 26.0% of the mothers of gender-referred girls; in contrast, only 1.7% of the mothers of the control boys and 2.0% of the mothers of the control girls selected this option. It is not entirely clear what “once-in-a-while” exactly means, but if the desire to be of the other sex is expressed on a once-in-a-while basis over, say, a 6-month period, does this count as repeatedly?

Recommendations

In this section, I advance three diagnostic options for consideration by the DSM-V Sexual and Gender Identity Disorders Workgroup.

Option 1

The first option would be to leave the criteria as they currently stand, other than consideration of some changes in wording (e.g., even greater harmonization in the criteria for boys and girls). An argument in favor of this option is that the current criteria have behaved reasonably well; for example, they show evidence of discriminant validity and, at least using the gold standard of clinician diagnosis, appear to reasonably distinguish between threshold and subthreshold cases. In my view, the main argument against retaining the criteria as they currently stand is that the ability to make the diagnosis in the absence of repeated verbal statements that one wishes to be of the other sex has led to confusion and the concern that the diagnosis is capturing children who are merely “gender variant.” Although I do not believe that this was the intent of the DSM-IV Subcommittee on Gender Identity Disorders, the concern about the A1 criterion runs across many of the critiques of the diagnosis as currently formulated.

Option 2

The second option would be to tighten the criteria by changing the Point A criterion to include all five parameters (A1–A5) as they are currently formulated in the DSM-IV. Inclusion of the verbalized desire to be of the other sex would make the diagnosis more transparent in its aim to identify children who are, without ambiguity, struggling with their gender identity (see, e.g., de Vries & Cohen-Kettenis, 2009). The analyses that I reported on above clearly show that children who more frequently state the desire to be of the other sex (by maternal report) also show more cross-gender surface behavior. Inclusion of A1 would likely constrict the net of children judged to meet the criteria for GID and this might be received by critics as responsive to concerns about misdiagnosis or overdiagnosis (even if this concern is incorrect).

A counter argument to this perspective is that children who meet the A2–A5 criteria and the B criterion may actually be struggling with their gender identity (for a clinical example, see Zucker, 2004). It has been suggested by some clinicians that there are children who may harbor a strong desire to be of the other sex, but do not verbalize it because of a coercive social environment (H. F. L. Meyer-Bahlburg, personal communication, May 26, 2009). If they do not receive a diagnosis, it may influence treatment options that ultimately might not be in the best interest of the child. One solution to this would be to use the residual diagnosis of GIDNOS, along with modification to the text that describes the clinical complexity in making a diagnosis, particularly for those children who, for whatever reason, do not verbally express their underlying gender dysphoria.

Option 3

A third, more radical option would be to eliminate from the criteria set all of the surface behaviors of possible cross-gender identity and relegate them to the text description of the diagnosis (e.g., Associated Descriptive Features). Here, one could point out that although these behaviors are often part of the GID phenomenology, they are also present among children who show pervasive cross-gender behavior but do not experience distress or unhappiness about their gender identity. In their place, one could recommend a largely new set of diagnostic criteria that focus more directly on different manifestations of gender dysphoria.

I favor Option 2. Option 2 would represent a reasonable response to criticisms of the criteria as currently formulated. It would, if anything, reduce the number of children who meet the criteria for GID. It would build on a history of studies that have already established reasonable evidence for the discriminant validity of the diagnosis and even some evidence of predictive validity (per Wallien & Cohen-Kettenis, 2008). Field trials would not have to start from scratch in terms of

Table 6 Proposed revision to the DSM-IV diagnostic criteria for Gender Identity Disorder in Children

-
- A. A strong discomfort with one's gender identity (in relation to the assigned sex at birth), of at least 6 months duration, as manifested by at least six of the following indicators (including A1)
- (1) a frequently stated desire to be the other sex or a frequently stated insistence that he or she is the other sex
 - (2) in boys, a strong preference for cross-dressing or simulating female attire; in girls, a strong preference for wearing only stereotypical masculine clothing and a strong rejection in the wearing of culturally normative feminine clothing
 - (3) a strong preference for cross-sex roles in make-believe or fantasy play
 - (4) a strong preference for the stereotypical toys, games, or activities of the other sex
 - (5) a strong preference for playmates of the other sex
 - (6) in boys, a strong rejection of stereotypical masculine toys, games, and activities and a strong avoidance of rough-and-tumble play; in girls, a strong rejection of stereotypical feminine toys, games, and activities
 - (7) a frequently stated or behaviorally represented dislike of one's sexual anatomy; in boys, manifested by one of the following: that he would like to have a vagina or to grow breasts; that he dislikes his penis or testes; simulation of female genitalia by sitting to urinate; in girls, manifested by one of the following: that she would like to have a penis or to grow one; that she dislikes the prospects of breast development or that she has a vagina; simulation of male genitalia by standing to urinate
- B. The disturbance is not concurrent with a physical intersex condition
- C. The disturbance causes clinically significant distress or impairment in social, occupational, or other important areas of functioning
-

Note: These proposed revisions represent my suggestions at the time I completed this review (December 29, 2008) for the Sexual and Gender Identity Disorders Work Group. They should not be read as reflecting any type of final consensus on the part of the Gender Identity Disorders subworkgroup

psychometrics. Field trials could thus focus on establishing the much-needed evidence of inter-clinician agreement in diagnosis and rely on already well-developed psychometric measures of external validity.

In Table 6, I provide a proposal for a revised criteria set that includes 7 indicators, which represent a combination of the A and B criteria in DSM-IV. It specifies that the desire to be of the other sex is necessary for the diagnosis to be given. The criteria are written in a manner that uses one of two consistent qualifiers (“frequently” or “strong”) across indicators. In addition, I suggest a lower-bound duration criterion of 6 months. The GID diagnosis has never had a formal duration criterion, unlike many other psychiatric diagnoses for children (e.g., 1 month: Selective Mutism; 4 weeks: Separation Anxiety Disorder [SAD]; 6 months: Attention-Deficit/Hyperactivity Disorder, Generalized Anxiety Disorder, and Oppositional Defiant Disorder [ODD]; 12 months: CD). Whereas there was good empirical evidence to justify the duration criterion for CD (Lahey et al., 1998), this was less

so the case, for example, for SAD and the duration criterion was modeled on the ICD-10 definition and an older empirical literature (Klein, Tancer, & Werry, 1997).

For GID, there is no formal empirical evidence for setting a specific lower-bound for duration, but this could, conceivably, be examined in a field trial. Clinically, with the exception of very young children (in the age range of 2–4 years), it is very common for the putative symptoms of GID to have been of substantial duration, as parents often do not seek out an evaluation until they, or a health professional, deem the behavior “no longer a phase” (see Zucker, 2000). The inclusion of a specific duration criterion would have the advantage of alerting the clinician to be attentive to chronicity and to be sensitive to instances of cross-gender behavior/identity that are transitory, perhaps in response to an acute or isolated stressor (e.g., the birth of a younger sibling) (Coates & Zucker, 1988).

Cultural Considerations

For DSM-V, there will likely be greater attention given to the interface between culture and psychopathology (Alarcón et al., 2002). There is certainly now a great deal of evidence to suggest that there are “non-Western” equivalents to GID in many different cultures and countries, both in children and in adults (see, e.g., Bartlett & Vasey, 2006; Newman, 2002; Tucker & Keil, 2001; Vasey & Bartlett, 2007). If cultural features are added to the DSM-V, it will be important to consider the applicability of the GID criteria, particularly in non-Western cultures.

Diagnostic Terminology

For DSM-IV, the Subcommittee on Gender Identity Disorders (Bradley et al., 1991) recommended that Gender Identity Disorder be used as an overarching term (collapsing the diagnoses of GIDC, Transsexualism, and GIDAANT from DSM-III-R). In part, this was argued because the term transsexualism was, at least in some circles, equated with a specific form of therapeutics, namely, contra-sex hormonal and surgical treatment. In addition, it was argued that GID in childhood versus adolescence and adulthood were, in effect, the same condition, but expressed differently as a function of developmental level.

Regarding the latter point, there is evidence for and against this argument. On the one hand, there is reasonable evidence for retrospective continuity, particularly when one examines the developmental histories of adolescents and adults who have a sexual orientation (attraction) to members of their birth sex (see, e.g., Singh et al., 2009; Zucker, 2006b). On the other hand, the evidence for prospective continuity is weaker, but still substantial if one relies on crude estimates of GID prevalence in adults (Drummond, Bradley, Peterson-Badali, & Zucker, 2008; Green, 1987; Wallien & Cohen-Kettenis,

2008; Zucker, 2008b). But the reasons for prospective discontinuity are likely to be multifactorial and, in and of itself, there is no compelling reason to contest *in toto* the relation between GID in childhood versus adolescence and adulthood. Indeed, the disjunction between retrospective and prospective continuity shares a similarity to the same kind of disjunction for CD and ODD (see Lahey, Loeber, Quay, Frick, & Grimm, 1997): classical CD is almost always preceded by ODD whereas the majority of children with ODD followed prospectively do not develop CD (see also Nock, Kazdin, Hiripi, & Kessler, 2007).

Over the years, there have been a myriad of terms used to label the phenomenology that is represented by the diagnostic label of GID. If GID is to remain in the DSM-V, should it retain the same name or should alternatives be considered? Di Ceglie (1998) has used the term Atypical Gender Identity Organization to “define an internal psychological configuration whose phenomenology is represented by the typical characteristics of a gender identity disorder” (p. 9). Vitale (2001) suggested the term Gender Expression Deprivation Anxiety Disorder, arguing that GID “may for treatment purposes be better described as a chronic anxiety disorder” (p. 121).

Others have suggested that the inclusion of the word “Disorder” in GID adds to the burden of stigma (see Meyer-Bahlburg, 2009). As an alternative, for example, Bancroft (2009), suggested the term “gender identity discordance” (p. 291).

On the matter of naming, I have no strong recommendation other than to consider the rule of parsimony. In DSM-IV, most diagnoses contain the word “Disorder,” but not all (e.g., Pica, Enuresis, Encopresis, Major Depressive Episode, Anorexia Nervosa, Bulimia Nervosa, all of the Paraphilias, etc.).

Secondary Data Analysis and Field Trials

Recommendations for secondary data analysis and field trials are as follows:

1. The re-analyses that I conducted on the GIQC can be examined in the cohort of child gender patients seen at the Amsterdam Gender Clinic, as both the GIQC and the GII are part of the Dutch assessment protocol. Secondary data analysis of the Dutch clinic data can provide a test of the consistency of the results reported here.
2. If the Gender Identity Disorders subworkgroup agrees on the merit of conducting field trials on a set of revised criteria, these should be studied on new clients seen in my own clinic and in the Amsterdam Gender Clinic. The subworkgroup should identify a target sample size for the probands and to collect clinical control data on a comparable sample size. The feasibility should be explored of enlisting other clinicians who assess children

with gender problems to test a revised set of criteria. The aim of the field trial would be to establish inter-clinician reliability and to conduct tests of discriminant validity.

Dimensional Diagnosis

The subworkgroup needs to explore possible methods for dimensional diagnosis. In my view, these could include at least two parameters:

1. A symptom count based on a polythetic approach to categorical diagnosis.

2. Identification of psychometric measures that match the reformulated diagnostic criteria.

Acknowledgments This review was prepared for the DSM-V Sexual and Gender Identity Disorders Work Group. I would like to thank the following individuals who provided feedback on the paper: Ray Blanchard, Peggy T. Cohen-Kettenis, Domenico di Ceglie, Jack Drescher, Heino F. L. Meyer-Bahlburg, Friedemann Pfäfflin, and Devita Singh. Elements of the paper were presented at the Annual Meeting of the Society for Sex Therapy and Research in Arlington, VA, April 2–5, 2009. Reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders V Workgroup Reports* (Copyright 2009), American Psychiatric Association.

Appendix

Anatomic Satisfaction Scale: Parent Report (ASS). *Instructions:* The following questions are about your child now or within the past year. Please answer each question by checking YES if you think your child is

like this, NO, if your child is not like this, or SOMETIMES if you think your child is somewhat like this

	Yes	Sometimes	No
1. Does ____ say that he wished he could change something about the way that he looked? If yes, what?			
2. Does ____ say that he likes his body?			
3. Does ____ say that he would like a vagina?			
4. Does ____ say that he dislikes his penis?			
5. Does ____ say that he is ugly?			
6. Does ____ say that he would like breasts?			
7. Does ____ pretend that he is pregnant (e.g., does he stuff his shirt)?			
8. Does ____ say that he wished he looked like a girl?			
9. Does ____ say that he wants to get rid of his penis?			
10. Does ____ express dislike for body hair, facial hair, or hair growth on any place aside from his head?			
11. Does ____ say he is proud of his body?			
12. Does ____ like to look at himself in the mirror?			
13. Does ____ think that he is overweight?			
14. Does ____ worry about the size of his penis?			
15. Does ____ talk about wanting to shave his legs when he is older?			
16. Does ____ pretend that he has a vagina?			
17. Does ____ say that he wishes his face were prettier?			
18. Does ____ pretend that he doesn't have a penis?			
19. Does ____ complain about his penis getting larger (e.g., when it gets erect)?			
20. Does ____ ever say that he wants surgery to change something about himself?			
21. Does ____ say that he wishes he were bigger (physically)?			
22. Does ____ pretend that he has breasts?			
23. Does ____ say that he wishes he were smaller (physically)?			
24. Does ____ fixate on a physical feature that he doesn't like about himself?			
25. Does ____ like what he looks like in pictures?			
26. Does ____ wish he were thinner?			
27. Does ____ pretend to shave his legs (or another part of his body)?			
28. Does ____ wish he had bigger muscles?			
29. Does ____ worry about the way that he looks?			
30. Does ____ wish he looked like someone else? If yes, who?			
31. Do ____'s looks upset him?			

Note: From Lambert (2009)

References

- Alarcón, R. D., Bell, C. C., Kirmayer, L. J., Lin, K.-M., Üstün, B., & Wisner, K. L. (2002). Beyond the funhouse mirrors: Research agenda on culture and psychiatric diagnosis. In D. J. Kupfer, M. B. First, & D. A. Regier (Eds.), *A research agenda for DSM-V* (pp. 219–289). Washington, DC: American Psychiatric Association.
- American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.
- American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.
- Ault, A., & Brzuzu, S. (2009). Removing gender identity disorder from the *Diagnostic and Statistical Manual of Mental Disorders*: A call for action. *Social Work, 54*, 187–189.
- Bancroft, J. (2009). *Human sexuality and its problems* (3rd ed.). Edinburgh: Churchill Livingstone (Elsevier).
- Bartlett, N. H., & Vasey, P. L. (2006). A retrospective study of childhood gender-atypical behavior in Samoan *Fa'afafine*. *Archives of Sexual Behavior, 35*, 659–666.
- Bartlett, N. H., Vasey, P. L., & Bukowski, W. M. (2000). Is gender identity disorder in children a mental disorder? *Sex Roles, 43*, 753–785.
- Bockting, W. (2009). Are gender identity disorders mental disorders? Recommendations for revision of the World Professional Association for Transgender Health's *Standards of Care*. *International Journal of Transgenderism, 11*(1), 53–62.
- Bockting, W. O., & Ehrbar, R. D. (2005). Commentary: Gender variance, dissonance, or identity disorder? *Journal of Psychology & Human Sexuality, 17*(3/4), 125–134.
- Bradley, S. J., Blanchard, R., Coates, S., Green, R., Levine, S. B., Meyer-Bahlburg, H. F. L., et al. (1991). Interim report of the DSM-IV subcommittee on gender identity disorders. *Archives of Sexual Behavior, 20*, 333–343.
- Brill, S., & Pepper, R. (2008). *The transgender child: A handbook for families and professionals*. San Francisco: Cleis Press.
- Bryant, K. (2006). Making Gender Identity Disorder of Childhood: Historical lessons for contemporary debates. *Sexuality Research and Social Policy: Journal of NRSC, 3*, 23–29.
- Bryant, K. (2008). In defense of gay children? 'Progay' homophobia and the production of homonormativity. *Sexualities, 11*, 455–475.
- Bryant, K. E. (2007). *The politics of pathology and the making of Gender Identity Disorder*. Unpublished doctoral dissertation, University of California at Santa Barbara.
- Canino, G., Shrout, P. E., Rubio-Stipec, B., Bird, H. R., Bravo, M., Ramirez, R., et al. (2004). The *DSM-IV* rates of child and adolescent disorders in Puerto Rico: Prevalence, correlates, service use, and the effects of impairment. *Archives of General Psychiatry, 61*, 85–93.
- Chiu, S. W., Gervan, S., Fairbrother, C., Johnson, L. L., Owen-Anderson, A. F. H., Bradley, S. J., et al. (2006). Sex-dimorphic color preference in children with gender identity disorder: A comparison to clinical and community controls. *Sex Roles, 55*, 385–395.
- Clausen, J. A. (1981). Stigma and mental disorder: Phenomena and terminology. *Psychiatry, 44*, 287–296.
- Coates, S. (1985). Extreme boyhood femininity: Overview and new research findings. In Z. DeFries, R. C. Friedman, & R. Corn (Eds.), *Sexuality: New perspectives* (pp. 101–124). Westport, CT: Greenwood Publishing.
- Coates, S., & Zucker, K. J. (1988). Gender identity disorders in children. In C. J. Kestenbaum & D. T. Williams (Eds.), *Handbook of clinical assessment of children and adolescents* (Vol. II, pp. 893–914). New York: New York University Press.
- Cohen-Kettenis, P. T., Owen, A., Kaijser, V. G., Bradley, S. J., & Zucker, K. J. (2003). Demographic characteristics, social competence, and behavior problems in children with gender identity disorder: A cross-national, cross-clinic comparative analysis. *Journal of Abnormal Child Psychology, 31*, 41–53.
- Cohen-Kettenis, P. T., Wallien, M., Johnson, L. L., Owen-Anderson, A. F. H., Bradley, S. J., & Zucker, K. J. (2006). A parent-report Gender Identity Questionnaire for Children: A cross-national, cross-clinic comparative analysis. *Clinical Child Psychology and Psychiatry, 11*, 397–405.
- Corbett, K. (1996). Homosexual boyhood: Notes on girlyboys. *Gender & Psychoanalysis, 1*, 429–461.
- Corbett, K. (1998). Cross-gendered identifications and homosexual boyhood: Toward a more complex theory of gender. *American Journal of Orthopsychiatry, 68*, 352–360.
- Crick, N. R., & Zahn-Waxler, C. (2003). The development of psychopathology in females and males: Current progress and future challenges. *Development and Psychopathology, 15*, 719–742.
- de Vries, A. L. C., & Cohen-Kettenis, P. T. (2009). Review of World Professional Association for Transgender Health's Standards of Care for children and adolescents with gender identity disorder: A need for change? *International Journal of Transgenderism, 11*, doi:10.1080/1553270903008040.
- Di Ceglie, D. (1998). Reflections on the nature of the "atypical gender identity organization". In D. Di Ceglie (Ed.), *A stranger in my own body: Atypical gender identity and mental health* (pp. 9–25). London: Karnac Books.
- Drescher, J. (2009). Queer diagnoses: Parallels and contrasts in the history of homosexuality, gender variance, and the *Diagnostic and Statistical Manual*. *Archives of Sexual Behavior*, doi:10.1007/s10508-009-9531-5.
- Drummond, K. D., Bradley, S. J., Peterson-Badali, M., & Zucker, K. J. (2008). A follow-up study of girls with gender identity disorder. *Developmental Psychology, 44*, 34–45.
- Ehrbar, R. D., Witty, M. C., Ehrbar, H. G., & Bockting, W. O. (2008). Clinician judgment in the diagnosis of gender identity disorder in children. *Journal of Sex and Marital Therapy, 34*, 385–412.
- Ehrensaft, D. (2007). Raising girlyboys: A parent's perspective. *Studies in Gender and Sexuality, 8*, 269–292.
- Feder, E. K. (1997). Disciplining the family: The case of gender identity disorder. *Philosophical Studies, 85*, 195–211.
- Fisk, N. (1973). Gender dysphoria syndrome (the how, what, and why of a disease). In D. Laub & P. Gandy (Eds.), *Proceedings of the second interdisciplinary symposium on gender dysphoria syndrome* (pp. 7–14). Palo Alto, CA: Stanford University Press.
- Fridell, S. R. (2001). *Sex-typed play behavior and peer relations in boys with gender identity disorder*. Unpublished doctoral dissertation, University of Toronto.
- Fridell, S. R., Owen-Anderson, A., Johnson, L. L., Bradley, S. J., & Zucker, K. J. (2006). The Playmate and Play Style Preferences Structured Interview: A comparison of children with gender identity disorder and controls. *Archives of Sexual Behavior, 35*, 729–737.
- Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.
- Green, R. (1974). *Sexual identity conflict in children and adults*. New York: Basic Books.
- Green, R. (1976). One-hundred ten feminine and masculine boys: Behavioral contrasts and demographic similarities. *Archives of Sexual Behavior, 5*, 425–446.
- Green, R. (1987). *The "sissy-boy syndrome" and the development of homosexuality*. New Haven, CT: Yale University Press.

- Haldeman, D. C. (2000). Gender atypical youth: Clinical and social issues. *School Psychology Review*, 29, 192–200.
- Hegarty, P. (2009). Toward an LGBT-informed paradigm for children who break gender norms: A comment on Drummond et al. (2008) and Rieger et al. (2008). *Developmental Psychology*, 45, 895–900.
- Hill, D. B., Rozanski, C., Carfagnini, J., & Willoughby, B. (2007). Gender identity disorders in childhood and adolescence: A critical inquiry. *International Journal of Sexual Health*, 19(1), 57–75.
- Hyman, S. E. (2003). Foreword. In K. A. Phillips, M. B. First, & H. A. Pincus (Eds.), *Advancing DSM: Dilemmas in psychiatric diagnosis* (pp. xi–xxi). Arlington, VA: American Psychiatric Association.
- Jalas, K. (2003). Between tomboys and butch lesbians: Gender non-conformity viewed through clinical psychoanalysis and lesbian and gay theory. *Psychoanalytic Review*, 90, 655–683.
- Johnson, L. L., Bradley, S. J., Birkenfeld-Adams, A. S., Radzins Kuksis, M. A., Maing, D. M., Mitchell, J. N., et al. (2004). A parent-report Gender Identity Questionnaire for Children. *Archives of Sexual Behavior*, 33, 105–116.
- Kendler, K. S. (1999). Setting boundaries for psychiatric disorders [Editorial]. *American Journal of Psychiatry*, 156, 1845–1848.
- Klein, R. G., Tancer, N. K., & Werry, J. S. (1997). Anxiety disorders of childhood or adolescence. In T. A. Widiger, A. J. Frances, H. A. Pincus, R. Ross, M. B. First, & W. Davis (Eds.), *DSM-IV sourcebook* (Vol. 3, pp. 221–239). Washington, DC: American Psychiatric Association.
- Lahey, B. B., Applegate, B., Barkley, R. A., Garfinkel, B. D., McBurnett, K., & Kerdyk, L. (1998). DSM-IV field trials for oppositional defiant disorder and conduct disorder in children and adolescents. In T. A. Widiger, A. J. Frances, H. A. Pincus, R. Ross, M. B. First, W. Davis, M. Kline (Eds.), *DSM-IV sourcebook* (Vol. 4, pp. 661–683). Washington, DC: American Psychiatric Association.
- Lahey, B. B., Loeber, R., Quay, H. C., Frick, P. J., & Grimm, J. (1997). Oppositional defiant disorder and conduct disorder. In T. A. Widiger, A. J. Frances, H. A. Pincus, R. Ross, M. B. First, & W. Davis (Eds.), *DSM-IV sourcebook* (Vol. 3, pp. 189–209). Washington, DC: American Psychiatric Association.
- Lambert, S. L. (2009). *Assessment of body image in boys with gender identity disorder: A comparison to clinical control boys and community control boys*. Unpublished doctoral dissertation, York University, Downsview, Ontario.
- Langer, S. J., & Martin, J. I. (2004). How dresses can make you mentally ill: Examining gender identity disorder in children. *Child and Adolescent Social Work Journal*, 21, 5–23.
- Law, G. U., Sinclair, S., & Fraser, N. (2007). Children's attitudes and behavioural intentions towards a peer with symptoms of ADHD: Does the addition of a diagnostic label make a difference? *Child Health Care*, 11, 98–111.
- Lawrence, A. A. (2008). Gender identity disorders in adults: Diagnosis and treatment. In D. L. Rowland & L. Incrocci (Eds.), *Handbook of sexual and gender identity disorders* (pp. 423–456). Hoboken, NJ: John Wiley & Sons.
- Lev, A. I. (2004). *Transgender emergence: Therapeutic guidelines for working with gender-variant people and their families*. New York: The Haworth Clinical Practice Press.
- Lindsay, L. A. (1994). Maternal reports of pregnancy, genital, and related fantasies in preschool and kindergarten children. *Journal of the American Academy of Child and Adolescent Psychiatry*, 33, 416–423.
- Lothstein, L. M. (1992). Clinical management of gender dysphoria in young boys: Genital mutilation and DSM IV implications. *Journal of Psychology & Human Sexuality*, 5, 87–106.
- Martin, C. L., & Ruble, D. N. (2004). Children's search for gender cues: Cognitive perspectives on gender development. *Current Directions in Psychological Science*, 13, 67–70.
- Martin, C. L., Ruble, D. N., & Szkrybalo, J. (2002). Cognitive theories of early gender development. *Psychological Bulletin*, 128, 903–933.
- Martin, J. I. (2008). Nosology, etiology, and course of gender identity disorder in children. *Journal of Gay & Lesbian Mental Health*, 12(1/2), 81–94.
- Menvielle, E. J., Tuerk, C., & Perrin, E. C. (2005). To the beat of a different drummer: The gender-variant child. *Contemporary Pediatrics*, 22(2), 38–39, 41, 43, 45–46.
- Meyer-Bahlburg, H. F. L. (1985). Gender identity disorder of childhood: Introduction. *Journal of the American Academy of Child Psychiatry*, 24, 681–683.
- Meyer-Bahlburg, H. F. L. (2009). From mental disorder to iatrogenic hypogonadism: Dilemmas in conceptualizing gender identity variants as psychiatric conditions. *Archives of Sexual Behavior*, doi:10.1007/s10508-009-9532-4.
- Minter, S. (1999). Diagnosis and treatment of gender identity disorder in children. In M. Rottnek (Ed.), *Sissies & tomboys: Gender nonconformity & homosexual childhood* (pp. 9–33). New York: New York University Press.
- Money, J. (1994). The concept of gender identity disorder in childhood and adolescence after 39 years. *Journal of Sex and Marital Therapy*, 20, 163–177.
- Moore, S. M. (2002). *Diagnosis for a straight planet: A critique of gender identity disorder for children and adolescents in the DSM-IV*. Unpublished doctoral dissertation, The Wright Institute, Berkeley, CA.
- Morgan, N. L. (2000). *Defining normal gender behavior: Therapeutic implications arising from psychologists' sex-role expectations and attitudes towards lesbians and gays*. Unpublished doctoral dissertation, American Schools of Professional Psychology, San Francisco Bay Area Campus, Point Richmond, CA.
- Newman, L. K. (2002). Sex, gender and culture: Issues in the definition, assessment and treatment of gender identity disorder. *Clinical Child Psychology and Psychiatry*, 7, 352–359.
- Nock, M. K., Kazdin, A. E., Hiripi, E., & Kessler, R. C. (2007). Lifetime prevalence, correlates, and persistence of oppositional defiant disorder: Results from the National Comorbidity Survey Replication. *Journal of Child Psychology and Psychiatry*, 48, 703–713.
- Nuttbrock, L., Hwahng, S., Bockting, W., Rosenblum, A., Mason, H., Macri, M., et al. (2009). Psychiatric impact of gender-related abuse across the life course of male to female transgender persons. *Journal of Sex Research*.
- Pickstone-Taylor, S. D. (2003). Children with gender nonconformity [Letter to the editor]. *Journal of the American Academy of Child and Adolescent Psychiatry*, 42, 266.
- Pleak, R. R. (1999). Ethical issues in diagnosing and treating gender-dysphoric children and adolescents. In M. Rottnek (Ed.), *Sissies & tomboys: Gender nonconformity & homosexual childhood* (pp. 34–51). New York: New York University Press.
- Richardson, J. (1996). Setting limits on gender health. *Harvard Review of Psychiatry*, 4, 49–53.
- Richardson, J. (1999). Response: Finding the disorder in gender identity disorder. *Harvard Review of Psychiatry*, 7, 43–50.
- Roberts, C. W., Green, R., Williams, K., & Goodman, M. (1987). Boyhood gender identity development: A statistical contrast of two family groups. *Developmental Psychology*, 23, 544–557.
- Rutter, M. (1978). Diagnostic validity in child psychiatry. *Advances in Biological Psychiatry*, 2, 2–22.
- Sedgwick, E. K. (1991). How to bring your kids up gay. *Social Text*, 9, 18–27.
- Singh, D., Deogracias, J. J., Johnson, L. J., Bradley, S. J., Kibblewhite, S. J., Owen-Anderson, A., et al. (2009). The Gender Identity/Gender Dysphoria Questionnaire for Adolescents and Adults: Further validity evidence. *Journal of Sex Research*, doi:10.1080/00224490902898728.

- Spitzer, R. L., & Endicott, J. (1978). Medical and mental disorder: Proposed definition and criteria. In R. L. Spitzer & D. F. Klein (Eds.), *Critical issues in psychiatric diagnosis* (pp. 15–39). New York: Raven Press.
- Spitzer, R. L., & Wakefield, J. C. (1999). DSM-IV diagnostic criterion for clinical significance: Does it help solve the false positives problem? *American Journal of Psychiatry*, *156*, 1856–1864.
- Stoller, R. J. (1968). Male childhood transsexualism. *Journal of the American Academy of Child Psychiatry*, *7*, 193–209.
- Tucker, J. B., & Keil, H. H. J. (2001). Can cultural beliefs cause a gender identity disorder? *Journal of Psychology & Human Sexuality*, *13*(2), 21–30.
- Vanderburgh, R. (2009). Appropriate therapeutic care for families with pre-pubescent transgender/gender-dissonant children. *Child and Adolescent Social Work Journal*, *26*, 135–154.
- Vasey, P. L., & Bartlett, N. H. (2007). What can the Samoan “Fa’afafine” teach us about the Western concept of Gender Identity Disorder in Childhood? *Perspectives in Biology and Medicine*, *50*, 481–490.
- Vitale, A. (2001). Implications of being gender dysphoric: A developmental review. *Gender and Psychoanalysis*, *6*, 121–141.
- Wakefield, J. C., & First, M. B. (2003). Clarifying the distinction between disorder and nondisorder: Confronting the overdiagnosis (false-positives) problem in DSM-V. In K. A. Phillips, M. B. First, & H. A. Pincus (Eds.), *Advancing DSM: Dilemmas in psychiatric diagnosis* (pp. 23–55). Arlington, VA: American Psychiatric Association.
- Walker, J. S., Coleman, D., Lee, J., Squire, P. N., & Friesen, B. J. (2008). Children’s stigmatization of childhood depression and ADHD: Magnitude and demographic variation in a national sample. *Journal of the American Academy of Child and Adolescent Psychiatry*, *47*, 912–920.
- Wallien, M. S. C., & Cohen-Kettenis, P. T. (2008). Psychosexual outcome of gender-dysphoric children. *Journal of the American Academy of Child and Adolescent Psychiatry*, *47*, 1413–1423.
- Wallien, M. S. C., Quilty, L. C., Steensma, T. D., Singh, D., Lambert, S. L., Leroux, A., et al. (2009). Cross-national replication of the Gender Identity Interview for Children. *Journal of Personality Assessment*, doi:10.1080/00223890903228463.
- Wallien, M. S. C., Veenstra, R., Kreukels, B. P. C., & Cohen-Kettenis, P. T. (2009). Peer group status of gender dysphoric children: A sociometric study. *Archives of Sexual Behavior*, doi:10.1007/s10508-009-9517-3.
- Widiger, T. A. (2007). DSM’s approach to gender: History and controversies. In W. E. Narrow, M. B. First, P. J. Sirovatka, & D. A. Regier (Eds.), *Age and gender considerations in psychiatric diagnosis: A research agenda for DSM-V* (pp. 19–29). Arlington, VA: American Psychiatric Association.
- Widiger, T. A., & First, M. B. (2007). Gender and diagnostic criteria. In W. E. Narrow, M. B. First, P. J. Sirovatka, & D. A. Regier (Eds.), *Age and gender considerations in psychiatric diagnosis: A research agenda for DSM-V* (pp. 127–137). Arlington, VA: American Psychiatric Association.
- Wilson, I., Griffin, C., & Wren, B. (2002). The validity of the diagnosis of gender identity disorder (child and adolescent criteria). *Clinical Child Psychology and Psychiatry*, *7*, 335–351.
- Zahn-Waxler, C. (1993). Warriors and worriers: Gender and psychopathology. *Development and Psychopathology*, *5*, 79–89.
- Zoccolillo, M. (1993). Gender and the development of conduct disorder. *Development and Psychopathology*, *5*, 65–78.
- Zucker, K. J. (1982). Childhood gender disturbance: Diagnostic issues. *Journal of the American Academy of Child Psychiatry*, *21*, 274–280.
- Zucker, K. J. (1992). Gender identity disorder. In S. R. Hooper, G. W. Hynd, & R. E. Mattison (Eds.), *Child psychopathology: Diagnostic criteria and clinical assessment* (pp. 305–342). Hillsdale, NJ: Erlbaum.
- Zucker, K. J. (1999). Commentary on Richardson’s (1996) “Setting Limits on Gender Health”. *Harvard Review of Psychiatry*, *7*, 37–42.
- Zucker, K. J. (2000). Gender identity disorder. In A. J. Sameroff, M. Lewis, & S. M. Miller (Eds.), *Handbook of developmental psychopathology* (2nd ed., pp. 671–685). New York: Kluwer Academic/Plenum Publishers.
- Zucker, K. J. (2001). Commentary on Haldeman’s (2000) “Gender Atypical Youth: Clinical and Social Issues.” Unpublished manuscript, Centre for Addiction and Mental Health, Toronto.
- Zucker, K. J. (2004). Gender identity disorder. In I. B. Weiner (Ed.), *Adult psychopathology case studies* (pp. 207–228). New York: Wiley.
- Zucker, K. J. (2005a). Measurement of psychosexual differentiation. *Archives of Sexual Behavior*, *34*, 375–388.
- Zucker, K. J. (2005b). Gender identity disorder in children and adolescents. *Annual Review of Clinical Psychology*, *1*, 467–492.
- Zucker, K. J. (2006a). Commentary on Langer and Martin’s (2004) “How Dresses Can Make You Mentally Ill: Examining Gender Identity Disorder in Children”. *Child and Adolescent Social Work Journal*, *23*, 533–555.
- Zucker, K. J. (2006b). Gender identity disorder. In D. A. Wolfe & E. J. Mash (Eds.), *Behavioral and emotional disorders in adolescents: Nature, assessment, and treatment* (pp. 535–562). New York: Guilford Press.
- Zucker, K. J. (2008a, October). Associated psychopathology in children and adolescents with gender identity disorder. In H. F. L. Meyer-Bahlburg (Chair), *From mental disorder to iatrogenic hypogonadism: Dilemmas in conceptualizing gender identity disorder (GID) as a psychiatric condition*. Symposium conducted at the meeting of the American Academy of Child and Adolescent Psychiatry, Chicago.
- Zucker, K. J. (2008b). On the “natural history” of gender identity disorder in children [Editorial]. *Journal of the American Academy of Child and Adolescent Psychiatry*, *47*, 1361–1363.
- Zucker, K. J., & Bradley, S. J. (1995). *Gender identity disorder and psychosexual problems in children and adolescents*. New York: Guilford Press.
- Zucker, K. J., Bradley, S. J., Kuksis, M., Pecore, K., Birkenfeld-Adams, A., Doering, R. W., et al. (1999). Gender constancy judgments in children with gender identity disorder: Evidence for a developmental lag. *Archives of Sexual Behavior*, *28*, 475–502.
- Zucker, K. J., Bradley, S. J., Lowry Sullivan, C. B., Kuksis, M., Birkenfeld-Adams, A., & Mitchell, J. N. (1993). A gender identity interview for children. *Journal of Personality Assessment*, *61*, 443–456.
- Zucker, K. J., Bradley, S. J., & Sanikhani, M. (1997). Sex differences in referral rates of children with gender identity disorder: Some hypotheses. *Journal of Abnormal Child Psychology*, *25*, 217–227.
- Zucker, K. J., Doering, R. W., Bradley, S. J., & Finegan, J. K. (1982). Sex-typed play in gender-disturbed children: A comparison to sibling and psychiatric controls. *Archives of Sexual Behavior*, *11*, 309–321.
- Zucker, K. J., Drummond, K. D., Bradley, S. J., & Peterson-Badali, M. (2009). Troubled meditations on psychosexual differentiation: Reply to Hegarty (2009). *Developmental Psychology*, *45*, 904–908.
- Zucker, K. J., Finegan, J. K., Doering, R. W., & Bradley, S. J. (1984). Two subgroups of gender-problem children. *Archives of Sexual Behavior*, *13*, 27–39.
- Zucker, K. J., Green, R., Bradley, S. J., Williams, K., Rebach, H. M., & Hood, J. E. (1998). Gender identity disorder of childhood: Diagnostic issues. In T. A. Widiger, A. J. Frances, H. A. Pincus, R. Ross, M. B. First, W. Davis, & M. Kline (Eds.), *DSM-IV sourcebook* (Vol. 4, pp. 503–512). Washington, DC: American Psychiatric Association.

- Zucker, K. J., Lozinski, J. A., Bradley, S. J., & Doering, R. W. (1992). Sex-typed responses in the Rorschach protocols of children with gender identity disorder. *Journal of Personality Assessment*, *58*, 295–310.
- Zucker, K. J., & Spitzer, R. L. (2005). Was the Gender Identity Disorder of Childhood diagnosis introduced into DSM-III as a backdoor maneuver to replace homosexuality? An historical note. *Journal of Sex and Marital Therapy*, *31*, 31–42.
- Zucker, K. J., Wilson-Smith, D. N., Kurita, J. A., & Stern, A. (1995). Children's appraisals of sex-typed behavior in their peers. *Sex Roles*, *33*, 703–725.