



A Follow-Up Study of Boys With Gender Identity Disorder

Devita Singh¹, Susan J. Bradley² and Kenneth J. Zucker^{2*}

¹ Department of Human Development and Applied Psychology, Ontario Institute for Studies in Education, University of Toronto, Toronto, ON, Canada, ² Department of Psychiatry, University of Toronto, Toronto, ON, Canada

OPEN ACCESS

Edited by:

Maria Inês Rodrigues Lobato,
Clinical Hospital of Porto Alegre, Brazil

Reviewed by:

Ray Blanchard,
University of Toronto, Canada
Lanna Petterson,
University of Lethbridge, Canada
Scott William Semenyina,
University of Lethbridge, Canada

*Correspondence:

Kenneth J. Zucker
ken.zucker@utoronto.ca

Specialty section:

This article was submitted to
Public Mental Health,
a section of the journal
Frontiers in Psychiatry

Received: 24 November 2020

Accepted: 18 February 2021

Published: 29 March 2021

Citation:

Singh D, Bradley SJ and Zucker KJ
(2021) A Follow-Up Study of Boys
With Gender Identity Disorder.
Front. Psychiatry 12:632784.
doi: 10.3389/fpsy.2021.632784

This study reports follow-up data on the largest sample to date of boys clinic-referred for gender dysphoria ($n = 139$) with regard to gender identity and sexual orientation. In childhood, the boys were assessed at a mean age of 7.49 years (range, 3.33–12.99) at a mean year of 1989 and followed-up at a mean age of 20.58 years (range, 13.07–39.15) at a mean year of 2002. In childhood, 88 (63.3%) of the boys met the DSM-III, III-R, or IV criteria for gender identity disorder; the remaining 51 (36.7%) boys were subthreshold for the criteria. At follow-up, gender identity/dysphoria was assessed via multiple methods and the participants were classified as either persisters or desisters. Sexual orientation was ascertained for both fantasy and behavior and then dichotomized as either biphilic/androphilic or gynephilic. Of the 139 participants, 17 (12.2%) were classified as persisters and the remaining 122 (87.8%) were classified as desisters. Data on sexual orientation in fantasy were available for 129 participants: 82 (63.6%) were classified as biphilic/androphilic, 43 (33.3%) were classified as gynephilic, and 4 (3.1%) reported no sexual fantasies. For sexual orientation in behavior, data were available for 108 participants: 51 (47.2%) were classified as biphilic/androphilic, 29 (26.9%) were classified as gynephilic, and 28 (25.9%) reported no sexual behaviors. Multinomial logistic regression examined predictors of outcome for the biphilic/androphilic persisters and the gynephilic desisters, with the biphilic/androphilic desisters as the reference group. Compared to the reference group, the biphilic/androphilic persisters tended to be older at the time of the assessment in childhood, were from a lower social class background, and, on a dimensional composite of sex-typed behavior in childhood were more gender-variant. The biphilic/androphilic desisters were more gender-variant compared to the gynephilic desisters. Boys clinic-referred for gender identity concerns in childhood had a high rate of desistance and a high rate of a biphilic/androphilic sexual orientation. The implications of the data for current models of care for the treatment of gender dysphoria in children are discussed.

Keywords: gender dysphoria, gender identity disorder, gender non-conformity, sexual orientation, DSM-5

INTRODUCTION

Gender identity is considered to be, for most people, a central aspect of one's sense of self (1–6).¹ By around 3 years of age, if not earlier, most children can self-label themselves as either a boy or a girl (11–14) although cognitive-developmental gender theory suggests that the understanding of gender as an “invariant” aspect of the self does not occur until early to middle childhood, with the achievement of concrete operational thought (12, 15, 16). Gender differences in the adoption of gender role behavior, i.e., behavior associated with cultural definitions of masculinity and femininity, also emerge during the preschool years, if not earlier. These behaviors span various domains, including peer, toy, role play, and activity preferences [e.g., (3, 17, 18)]. Normative developmental research has long documented that, on average, both gender identity and gender role behaviors show significant and substantial between-sex differences (19–21). Later in development, sexual orientation also shows a substantial between-sex difference, i.e., most males are sexually attracted to females and most females are sexually attracted to males (19, 22).

In the 1950s and 1960s, a small clinical literature began to describe the phenomenology of children who displayed marked gender-variant behavior, including the strong desire to be of the other gender [e.g., (23–27)]. Subsequent volumes by Stoller (28) and Green (29) provided more comprehensive descriptions of such children. These early works were the sequel to the introduction of the diagnostic term Gender Identity Disorder (GID) of Childhood to the psychiatric nomenclature in the third edition of the *Diagnostic and Statistical Manual of Mental Disorders* [DSM-III; (30)], currently termed Gender Dysphoria (GD) in the DSM-5 (31). Since 1980, empirical research has examined a number of parameters pertaining to GID/GD: epidemiology, diagnostic and assessment methods, associated psychopathology, causal mechanisms, and therapeutic approaches [for reviews, see, e.g., (32–39)].

An additional parameter (the focus of the present study) pertains to the developmental course of GID in children. In the early literature, it was posited by some that pervasive gender-variant behavior in children might be a predictor of GID in adulthood (termed Transsexualism in the DSM-III) [e.g., (26, 40)]. At the same time, it was also recognized that gender-variant behavior in childhood was associated with sexual orientation (in males, androphilia, i.e., sexual attraction to men; in females, gynephilia, i.e., sexual attraction to women), but without co-occurring gender dysphoria [see, e.g., (41, 42); for a meta-analytic review, see (43)].

To date, there have been at least 10 follow-up studies of children whose behavior was consistent with the DSM diagnosis

of GID (or GD per DSM-5) (44–53). Across these studies, the year at the time of first evaluation in childhood ranged from 1952 (49) to 2008 (51). For the 9 studies that included boys, the sample sizes (excluding those lost to follow-up) ranged from 6 to 79 (Mean age, 26 years). Most of these studies also provided the age at the time of first evaluation in childhood, which ranged from a mean of 7 years (47) to a mean of 9 years (48), with an age range from 4 to 12 years.

At the time of follow-up, using different metrics (e.g., clinical interview, maternal report, dimensional measurement of gender dysphoria, a DSM diagnosis of GID, etc.), these studies provided information on the percentage of boys who continued to have gender dysphoria (herein termed “persisters”) and the percentage of boys who did not (herein termed “desisters”).² Of the 53 boys culled from the relatively small sample size studies (Bakwin, Davenport, Kosky, Lebovitz, Money and Russo, Zuger), the percentage classified as persisters was 9.4% (age range at follow-up, 13–30 years). In Green (47), the percentage of persisters was 2% (total $n = 44$; Mean age at follow-up, 19 years; range, 14–24); in Wallien and Cohen-Kettenis (52), the percentage of persisters was 20.3% (total $n = 59$; Mean age at follow-up, 19.4 years; range, 16–28); and in Steensma et al. (51), the percentage of persisters was 29.1% (total $n = 79$; Mean age at follow-up, 16.1 years; range, 15–19). Across all studies, the percentage of persisters was 17.4% (total $N = 235$), with a range from 0 to 29.1%.³

These studies also provided information on the sexual orientation of the boys at the time of follow-up. In the early studies, sexual orientation was ascertained from various sources (e.g., open-ended interviews with the patient, parent-report, chart information, etc.). In the more recent studies, sexual orientation was assessed in a more systematic manner, such as the use of a structured interview to assign a Kinsey-based rating of sexual orientation in fantasy and a rating of sexual orientation in behavior, dummy coded where a 0 = gynephilia and a 6 = androphilia [e.g., (47)].

Of the 53 boys culled from the relatively small sample size studies (op. cit.), 13 (34.2%) of the patients were classified as gynephilic and 25 (65.8%) were classified as biphilic/androphilic.⁴ In the remaining 15 patients (28.3% of the combined samples), their sexual orientation was either uncertain or unknown.

²The terms persistence and desistance have been used for a long time in clinical developmental psychiatry and psychology [e.g., (54)]. Zucker (55) was the first to apply these terms to describe the developmental psychosexual trajectories of children diagnosed with GID.

³The percentages provided here differ somewhat from other summary reviews [(39), pp. 285–286, (56, 57)] because we have excluded patients who were seen for the first time in adolescence [for this reason, data from Zuger (58) are also not included]. One other follow-up study was conducted by Nakamura (59). Unfortunately, this dissertation is not available for purchase at ProQuest (Ann Arbor, MI) and is only available for loan at the University of Essex library. Due to COVID-19 restrictions, it is currently inaccessible (K. Clarke, personal communication to G. Rieger, June 15, 2020). The director of the clinic at the time when the data were collected does not have a copy of the dissertation (D. Di Ceglie, personal communication, June 15, 2020).

⁴As pointed out by Reviewer 1, biphilic is a dubious neologism, combining Latin and Greek derivatives. Diphilic would be the more accurate derivative. However, introducing this term would probably confuse many readers, so we have retained the term biphilic (see https://en.wikipedia.org/wiki/Androphilia_and_gynephilia).

¹In one study, Turner and Brown (7) found that school-age children rarely mentioned their gender when providing open-ended self-descriptions; the most frequent descriptor pertained to activities and preferences. Turner and Brown suggested that it might be the case that gender is so central to one's self-concept that it “goes without saying” (p. 709). In contemporary times in the West, a very small number of parents choose to not “gender” their children (“theybies”) by not referring to them as boys or girls (and, at times, not even announcing to others the child's biological sex), dressing them in gender-neutral ways, etc. Little is known about the gender identity and gender role patterns of these children (8–10).

In Green's (47) study, 11 (25%) of the boys were classified as gynephilic (Kinsey ratings of 0–1) and 33 (75%) were classified as biphilic/androphilic in fantasy (Kinsey ratings of 2–6). For behavior, 6 (20%) were classified as gynephilic and 24 (80.0%) were classified as biphilic/androphilic. The remaining 14 boys (31.8% of the total sample) could not be classified with regard to behavior because they had had no interpersonal sexual experiences. In Green's study, the sexual orientation of a comparison group of boys, who had been recruited from the community, was also assessed: 100% of these boys ($n = 35$) were classified as gynephilic in fantasy and 96% ($n = 25$) were classified as gynephilic in behavior.

In the Wallien and Cohen-Kettenis (52) study, sexual orientation was assessed for attraction (2 items), fantasy (2 items), behavior (4 items), and sexual identity (1 item) using a self-developed Sexual Orientation Questionnaire. As in Green, Kinsey-type ratings were used in the analysis. Depending on the metric, data on sexual orientation were not available for anywhere between 22 and 40 (27.2–67.7%) patients. For attraction, 32% were classified as gynephilic and 68% were classified as androphilic (total $N = 37$); for fantasy, 19% were classified as gynephilic, 19% were classified as biphilic, and 62% were classified as androphilic (total $N = 21$); for behavior, 21% were classified as gynephilic, 16% were classified as biphilic, and 63% were classified as androphilic (total $N = 19$); lastly, for sexual identity, 19% were classified as gynephilic ("heterosexual"), 19% were classified as biphilic ("bisexual"), and 62% were classified as androphilic ("homosexual") (total $N = 27$). Steensma et al. (51) used the same metrics as Wallien and Cohen-Kettenis. Depending on the metric, data on sexual orientation were not available for anywhere between 25 and 40 (31.6%–50.6%) patients. For attraction, 19.2% were classified as gynephilic, 15.4% were classified as biphilic, and 65.4% were classified as androphilic (total $N = 52$); for fantasy, 14% were classified as gynephilic, 22% were classified as biphilic, and 64% were classified as androphilic (total $N = 50$); for behavior, 35.9% were classified as gynephilic, 12.8 were classified as biphilic, and 51.3% were classified as androphilic (total $N = 39$); lastly, for sexual identity, 13% were classified as gynephilic ("heterosexual"), 27.8% were classified as biphilic ("bisexual"), and 59.3% were classified as androphilic ("homosexual") (total $N = 54$).

In recent years, there have been various criticisms of these follow-up studies [see, e.g., (60–63); for a rebuttal, see (64)], particularly with regard to the putatively high percentage of desistance. It has been questioned, for example, to what extent the patients in these studies truly had GID/GD. For example, in the early studies, prior to the publication of DSM-III, one could reasonably argue that the diagnostic status of the patients was unclear because there were no formal diagnostic criteria to rely upon. However, one could argue in return that the behavior of these boys was phenomenologically consistent with the subsequent DSM criteria.

Consider, for example, the systematic study by Green [(47), Figure 1.2]. Green reported that 15% of the feminine boys, per parent-report, had "never" expressed the desire to be a girl or a woman at the time of the baseline assessment, 60% "occasionally" had such a desire, and only 25% had such a desire

"frequently." Thus, a conservative critic might argue that only the last group would have met one of the key indicators for the GID/GD diagnosis in the DSM.⁵ On the other hand, suppose a boy "occasionally" voiced the desire to be a girl over a period of several years. One might want to make the case that this would be consistent with the DSM descriptors of "persistently" or "repeatedly," etc. Of course, one could debate what would genuinely count as "occasionally" (in Green's trichotomous metric, it would be anything more than "never" and less than "frequently"). In any case, it is probably reasonable to argue that, in Green's study, some boys were threshold and some boys were subthreshold for the equivalent of a DSM diagnosis. Given that in Green's study only one boy persisted with gender dysphoria at the time of follow-up, the threshold-subthreshold distinction would not really matter.

Studies that employed DSM criteria for GID/GD allow for a more formal examination of the "No True Scotsman" argument (https://en.wikipedia.org/wiki/No_true_Scotsman).

In the Wallien and Cohen-Kettenis (52) study, the DSM-III-R criteria were used to diagnose GID. Of the 12 persisters, all met the criteria for GID at the time of the baseline assessment; in contrast, only 68% of the 47 desisters met the criteria for GID; the remainder were deemed subthreshold for the diagnosis. Thus, in their study, the threshold-subthreshold distinction appears to have been an important one in predicting outcome; nonetheless, it should be noted that 68% of the desisters had been threshold for the diagnosis in childhood—perhaps a strong rebuttal to the No True Scotsman argument. In Steensma et al. (51), the DSM-IV-TR criteria were used. Of the 23 persisters, 21 (91.3%) met the criteria for GID; in contrast, only 22 (39.3%) of the 56 desisters were threshold for the diagnosis, suggesting an even more substantial difference in the threshold-subthreshold distinction than was found in Wallien and Cohen-Kettenis. Although the latter percentage was lower than what was found in Wallien and Cohen-Kettenis, that almost 40% of the desisters met the criteria for GID in childhood still argues in favor that the children were desisting from something.⁶

From Wallien and Cohen-Kettenis (52) and Steensma et al. (51), one predictor of outcome, therefore, was the distinction between being threshold or subthreshold for the GID diagnosis in childhood. Dimensional measures of gender-variant behavior have also proven useful. In both Wallien and Cohen-Kettenis and Steensma et al., dimensional measures of sex-typed behavior in childhood also significantly discriminated between the persisters and desisters, with the former group having, on average, more severe gender-variant behavior at the time of the childhood

⁵The situation is compounded even further because in the DSM-IV, unlike in the DSM-III and DSM-III-R (65), the stated desire to be of the other gender was not a necessary criterion for the diagnosis [for the rationale, see (66), pp. 483–486]. In DSM-5, the desire to be of the other gender does not require explicit verbalization; the clinician is allowed leeway in drawing inferences based on other sources of information [see (67), pp. 904–905].

⁶In the follow-up study by Drummond et al. (46) of 25 girls from our clinic, the desistance rate was 88%. Of the 22 desisters, 13 (59.0%) met the DSM-III, III-R or IV criteria for GID. In Wallien and Cohen-Kettenis (52), of the 9 girls who desisted, 55.5% met the DSM-III-R criteria for GID. In Steensma et al. (51), of the 24 girls who desisted, 58.3% met the DSM-IV criteria for GID.

assessment. Steensma et al. found two other predictors of persistence: boys who were assessed at an older age and boys who had made either a partial or complete gender “social transition” [see (68–70)]. Of the 12 boys who had partially or completely transitioned prior to puberty, 10 (83.3%) were classified as persisters. In contrast, of the 67 boys who had not socially transitioned, only 13 (19.4%) were classified as persisters.

In the present study, we provide follow-up data with regard to both gender identity (persistence vs. desistance) and sexual orientation (gynephilia vs. biphilia/androphilia) on the largest sample of boys studied to date. Apart from providing percentage data on these two variables, which will be discussed in a comparative perspective in relation to the prior studies and the epidemiological literature, we also examine the predictors of outcome in relation to both demographic and sex-typed behavior measures (including whether or not the boys were threshold or subthreshold for GID) collected at the time of the baseline assessment in childhood.

METHOD

Participants

The participants were 139 boys (“birth-assigned males”)⁷ who, in childhood, had been referred to and then assessed in the Gender Identity Service, Child, Youth, and Family Program at the Centre for Addiction and Mental Health (CAMH) in Toronto, Ontario between 1975 and 2009 (Mean year of assessment, 1989.36) and were adolescents or adults at follow-up (Mean year at follow-up, 2002.35).⁸

Participants entered the follow-up study through two methods of recruitment. The majority of participants (77%) were recruited for research follow-up. There were two main waves of participant recruitment through research contact, from 1986 to 1993 ($n = 32$) and then from 2009 to 2011 ($n = 71$). During the period of data collection, 32 patients re-contacted the service for clinical reasons (eight for gender dysphoria, six for sexual orientation, and 18 for heterogeneous concerns) [for details, see (77), Appendix E]. They were informed about the opportunity to participate in the follow-up study and subsequently completed the study protocol. The majority of the patient-initiated participants had contacted the clinic between the two main waves of research recruitment. Thus, from 1994 to 2008, the participants who entered the study were primarily those who had contacted the service for clinical reasons.

In the early wave of follow-up, a lower-bound age for participation was set at 14 years, but by the mid-1990s this was

changed to a lower-bound age of 16 years. In total, 110 (79.1%) participants were at least 16 years of age and 29 (20.9%) were younger than 16. Across the entire period of data collection, eligible participants, after review of the medical chart, were contacted at random (other than the participants who had returned to the service for clinical reasons). Due to lack of study resources and time constraints, contact with 162 other eligible participants was not attempted.

In total, 145 patients were approached about the follow-up study, either through research contact ($n = 113$) or following their clinical involvement with the Gender Identity Service ($n = 32$). Six patients declined, which yielded a participation rate of 95.9%. For those recruited for research purposes, initial contact, by telephone, letter or email, was first made with the parents because the patients were minors at the time of the childhood assessment and may have had no recollection of their clinic attendance. A total of 19 (14.3%) potential participants could not be reached/traced through previous addresses, registrars, and personal contacts.

Of the 139 participants, 110 were seen for a face-to-face assessment. For various reasons, the remaining 29 patients could not be seen for the face-to-face assessment (e.g., lived in another province or country, “too busy,” severe mental health issues). For some patients, they provided some information over the phone or information was provided by the parents; thus, for these patients, it was possible to obtain some follow-up data about their gender identity and sexual orientation.

The demographic characteristics of the participants, including their age at assessment in childhood and at the time of follow-up, are shown in **Table 1**. The GID diagnosis in childhood was based on the DSM-III ($n = 53$), DSM-III-R ($n = 46$), or DSM-IV ($n = 40$) criteria applicable at the time of assessment.⁹ A total of 88 (63.3%) boys met complete DSM criteria for GID in childhood. The remaining 51 (36.7%) boys were subthreshold for a DSM diagnosis, but all had some indicators of GID, and, based on the historical information provided during the assessment, some would have met the complete DSM criteria at some point in their lives prior to their assessment in childhood.¹⁰ The percentage who met the complete DSM criteria for GID did not differ significantly as a function of DSM edition, $\chi^2_{(2)} < 1$.

Procedure

The majority of participants who completed the face-to-face assessment were evaluated on a single day. Three participants were seen twice. In these instances, the participants completed the self-report measures during their second visit as the complexity of their clinical presentation extended the duration of the assessment. Participants were provided a stipend for their participation in the follow-up assessment and reimbursement for travel expenses. For participants followed-up prior to 2009 ($n = 68$), the data were collected by the third author; for those followed-up between 2009 and 2011, the data were collected

⁷Two reviewers asked why we chose to use the noun “boys” instead of the noun “males.” In our view, the question was reasonable but also a matter of semantics and taste. The third edition of *The Oxford Dictionary of Current English* (71) defines boy as “a male child...” Thus, we believe that the two words can be used synonymously. Males can refer to any age in the life-span whereas boys connote childhood. The participants in our study were coded as male at the time of their birth in the hospital delivery record, of which we had the actual birth records for the majority of the participants in the current study (72). As per Bouman et al. (73), one would say that the participants were “assigned male at birth” and then declared socially to be “boys” (74).

⁸The clinic was established in 1975 at the Clarke Institute of Psychiatry (75, 76), which became part of the CAMH in 1998.

⁹For boys seen prior to the publication of DSM-III in 1980, the draft criteria were used.

¹⁰In DSM-III, termed Atypical Gender Identity Disorder; in DSM-III-R and DSM-IV, termed Gender Identity Disorder Not Otherwise Specified.

TABLE 1 | Demographic characteristics ($N = 139$).

Characteristic	<i>M</i>	<i>SD</i>	Range	%
From childhood				
Age (in years)	7.49	2.66	3.33–12.99	
Year of birth	1981.87	7.50	1966–1996	
Year of assessment	1989.36	7.50	1975–2004	
IQ ^a	105.93	15.47	69–138	
Social class ^b	40.74	15.15	8.0–66.0	
Marital status ^c				
Two-parent family				64.7
Other				35.3
Caucasian				84.9
At follow-up				
Age (in years)	20.58	5.22	13.07–39.15	
Year of follow-up	2002.35	9.08	1986–2011	
Follow-up interval (in years) ^d	12.88	6.07	2.77–29.29	
IQ ^{e,f}	105.88	16.03	65–138	

^aFull-Scale IQ was obtained with age-appropriate Wechsler intelligence scales.

^bHollingshead's (78) Four Factor Index of Social Status (absolute range, 8–66).

^cOther included the following family constellations: single parent, separated, divorced, living with relatives, or in the care of a child protection agency.

^dInterval denotes the time between childhood assessment and follow-up assessment.

^eFull Scale IQ estimated using four subtests: Vocabulary, Comprehension, Block Design, and Object Assembly.

^fAn IQ score was available only for participants who completed the face-to-face assessment. Of these, scores were not available for one participant.

by the first author ($n = 71$). The study was approved by the Institutional Review Boards at the Clarke Institute of Psychiatry (subsequently the Centre for Addiction and Mental Health; Protocol #198/2008–2011) and the University of Toronto.

Measures

Below, we describe the measures from assessment and follow-up of relevance for this article. A list of all measures used in the follow-up study can be found in Singh [(77), Table 4].

Childhood Assessment

Cognitive Functioning

Based on the child's age at the time of assessment, the appropriate version of the Wechsler Intelligence Scale for Children was administered (WPPSI-R or the WISC-R/WISC-III/WISC-IV). Full scale IQ scores were used to characterize level of cognitive functioning.

Behavioral and Emotional Problems

Parents completed the Child Behavior Checklist (CBCL), a measure of behavioral and emotional problems (79). Although not the focus of the present study, it is noted here because we used three CBCL indices (sum of all behavior problems and Internalizing and Externalizing T scores) as part of an internal validity analysis when comparing participants vs. non-participants (see Results).

Sex-Typed Behavior

Five child informant and two parent informant measures were used to assess the participants' sex-typed behavior in childhood: (1) Draw-a-Person [DAP] test (80); (2) a free-play task (81); (3) the Playmate and Playstyle Preferences Structured Interview (PPPSI) (82, 83); (4) sex-typed responses on the Rorschach test (84); (5) the Gender Identity Interview for Children (GIIC) (85–87); (6) the Gender Identity Questionnaire for Children (GIQC) (88–90); and (7) a measure of activity level/extraversion [(39); see also (91)]. These child and parent informant measures all have established discriminant validity, that is, they significantly differentiated the boys clinic-referred for gender identity concerns from control boys [for reviews, see (18, 92)]. A Childhood Sex-Typed Behavior Composite was subsequently computed for each participant (see below).

Follow-Up Assessment

Cognitive Functioning

Four subtests from the age-appropriate version of the Wechsler Intelligence Scales were administered (Vocabulary, Comprehension, Block Design, and Object Assembly). The standard scores from the subtests were averaged to form a prorated IQ score for cognitive functioning (93).

Concurrent Gender Identity

Concurrent gender identity was evaluated using a semi-structured interview and self-report questionnaires. During an audiotaped interview, each participant was asked to describe their current feelings about being a biological male. They were also asked to describe positive and negative aspects about their gender identity. For example, participants who reported a "male" gender identity were asked to describe positive and negative aspects of being male. The semi-structured interview also included questions based on the adolescent and adult GID criteria outlined in the DSM-III-R or DSM-IV (65, 94). Participants were asked to respond to these questions according to the last 12 months with *No*, *Sometimes*, or *Yes* [for details, see (77), Appendix G].

Two self-report measures were also used to assess current gender identity and gender dysphoria: (1) The Gender Identity/Gender Dysphoria Questionnaire for Adolescents and Adults (GIDYQ-AA) (95–97) or (2) the Gender Dysphoria/Identification questionnaire (GDIQ) (98). The GDIQ was developed prior to the GIDYQ-AA. As such, the GIDYQ-AA was introduced to the protocol subsequent to the GDIQ and, as a result, the more recent participants completed the GIDYQ-AA while earlier participants completed the GDIQ.

The male version of the GIDYQ-AA was completed. This 27-item questionnaire measures gender identity and gender dysphoria in adolescents or adults; participants over the age of 17 completed the adult version and younger participants completed the adolescent version. The adolescent and adult versions are identical except that, in the adult version, the words "man" and "woman" are used instead of "boy" and "girl." Each item was rated on a 1–5 point response scale with verbal anchor points ranging from *Never* to *Always* based on a time frame of the past 12 months. Coding was such that a "lower" score signified more gender dysphoria. Item examples include the following:

“In the past 12 months, have you felt unhappy about being a man?” and “In the past 12 months, have you had the wish or desire to be a woman?” Principal axis factor analysis identified a one-factor solution that accounted for 61.3% of the variance. All factor loadings were ≥ 0.30 (median, 0.86; range, 0.34–0.96). The GIDYQ-AA has strong evidence for discriminant validity and a high threshold for specificity (i.e., low false positive rate for non-GID individuals) [see (95, 96, 99–102)].

The GDIQ (98) contains 8 items pertaining to gender identity and gender dysphoria. Factor analysis identified two factors, accounting for 31.4 and 12.5% of the variance, respectively (all factor loadings ≥ 0.45). Factor 1 consisted of five items pertaining to gender dysphoria and Factor 2 consisted of three items pertaining to gender role identification. For the present study, only the questions for Factor 1 were used. Each item was rated on a 3-point or 5-point scale for the past 12 months (see **Appendix 1** in **Supplementary Material**).

Participants were classified as having persistent gender dysphoria if their mean score on the GIDYQ-AA was ≤ 3.00 , in line with sensitivity and specificity analyses from other data sets (95, 96). For participants who did not complete the GIDYQ-AA, the GDIQ was used. A participant was classified as a persister if two or more of the following five items on the GDIQ were endorsed: wish to have been born a girl (Item 1), wish to have surgery to change body (Item 2), feel more like a girl than a boy (Item 3), wonder if would be happier as a girl (Item 4), and somewhat or very dissatisfied with being a boy (Item 5).

Information regarding participants' gender identity/gender dysphoria was also obtained during the semi-structured clinical interview and, therefore, allowed for cross-validation of these questionnaire data. For those participants who did not complete the face-to-face interview, clinical information regarding gender identity/gender dysphoria was obtained through self- or parent-report or chart review. Across the entire sample, the GIDYQ-AA was used to classify persistence or desistence for 64 participants, the GDIQ for 42 participants, and interview/chart data/parent report for 33 cases.

Sexual Orientation

Sexual orientation in fantasy was assessed with specific questions from an audiotaped face-to-face interview and the self-report Erotic Response and Orientation Scale (EROS) (103).

The interview asked about four types of sexual fantasy over the past 12 months: (1) crushes on other people; (2) sexual arousal to visual stimuli (e.g., acquaintances, partners, and individuals from movies, television, etc.); (3) sexual content of night dreams; and (4) sexual content of masturbation fantasies. During the interview, participants were not asked directly about the gender of the person or persons who elicited sexual arousal, thus allowing time for the participant to provide this information spontaneously. Directed questions about the gender of the person(s) who elicited sexual arousal were asked only if the participant did not volunteer specific information about whether their arousal was directed to same-sex or opposite-sex individuals, or both. By the end of the interview, each participant provided information about sexual arousal to both same-sex and opposite-sex individuals. Using the Kinsey scale criteria

(104), the interviewer assigned Kinsey ratings that ranged from 0 (exclusively gynephilic in fantasy) to 6 (exclusively androphilic in fantasy) for each question. A dummy score of 7 denoted that the participant did not experience or report any fantasies. A global fantasy score was also derived based on ratings from the four questions. Kinsey ratings for sexual orientation in fantasy were available for 129 participants.

Inter-rater reliability on Kinsey ratings for sexual orientation in fantasy was examined for 29 participants, selected at random. The second scorer listened to the audio recordings of the semi-structured interview, with specific attention to the information collected on sexual orientation. The inter-rater agreement on the Kinsey global fantasy rating was very good ($\kappa = 0.95$) and the kappa values for the four specific components ranged from 0.81 to 1.00.

The EROS is a 16-item self-report measure assessing sexual orientation in fantasy over the past 12 months. Half of the questions pertained to gynephilic fantasy (e.g., “How often have you noticed that you had sexual feelings [even the slightest] while looking at a woman?”) and the other half pertained to androphilic fantasy (e.g., “How often have you noticed that you had sexual feelings [even the slightest] while looking at a man?”). Participants who were 18 years and older completed the adult version and younger participants completed the adolescent version. The adolescent and adult versions are identical except that, in the adult version, the words “man” and “woman” were used instead of “boy” and “girl.” Each item was rated on a 5-point scale for frequency of occurrence, ranging from 1 (“none”) to 5 (“almost every day”). Mean androphilic and gynephilic fantasy scores were derived for each participant. In the present study, we calculated a difference score between the participants' mean androphilic and gynephilic scores. Previous use of the EROS has shown good evidence of discriminant validity (98, 101).

Sexual orientation in behavior was assessed with specific questions during the face-to-face interview and with a modified version of the Sexual History Questionnaire (SHQ) (105). In the interview, questions asked about five types of sexual behavior: (1) dating; (2) holding hands in a romantic manner; (3) kissing; (4) genital fondling or touching a woman on the breasts, and (5) intercourse (penile-vaginal and anal). Kinsey ratings for behavior in the past 12 months were made in the same manner as fantasy ratings. Kinsey ratings for sexual orientation in behavior were available for 108 participants. Inter-rater reliability on Kinsey ratings for sexual orientation in behavior was examined for the same 29 participants. There was perfect inter-rater agreement on the Kinsey global behavior rating ($\kappa = 1.0$) and the kappa values for the five specific components ranged from 0.91 to 1.00.

The modified SHQ consists of 20 questions. Ten questions pertained to gynephilic experiences (e.g., “How many women have you kissed on the lips in a romantic way?”) and 10 questions pertained to androphilic experiences (e.g., “How many men have you kissed on the lips in a romantic way?”). Participants who were 18 years and older completed the adult version and younger participants completed the adolescent version. The adolescent and adult versions are identical except that, in the adult version, the words “man” and “woman” were used instead of “boy” and “girl.” Each item was rated on a 5-point scale for frequency

of occurrence, ranging from 1 (“none”) to 5 (“11 or more”), based on a time frame of the past 12 months. Mean total scores for gynephilic and androphilic experiences were derived. In the present study, we calculated a difference score between the participants’ mean androphilic and gynephilic scores.

On the basis of Kinsey ratings, participants who completed the face-to-face interview were classified, similar to Green (47), into the following three sexual orientation groups for both fantasy and behavior: (1) gynephilic (Kinsey global ratings of 0–1); (2) biphilic/androphilic (Kinsey global ratings of 2–6), and (3) no sexual fantasy or behavior.

Social Desirability

Social desirability refers to the desire to cast a favorable impression on others. It can threaten the validity of self-report scales if in answering questions respondents seek social approval or try to represent themselves in a favorable manner (106). People scoring high on social desirability tend to provide socially acceptable answers regardless if their response accurately describes them. Participants 18 years and older completed the Marlow-Crowne Social Desirability Scale (M-CSDS) (107), which consists of 33 true-false items. The scale contains 18 culturally acceptable but unlikely statements keyed in the true direction and 15 socially undesirable but probable statements keyed in the false direction for a maximum possible score of 33. Participants 17 years and under were given a shorter version of the M-CSDS (108), containing 20 items that consist of 12 culturally acceptable but improbable statements keyed in the true direction and eight socially undesirable but probable statements keyed in the false direction for a maximum possible score of 20. For the present study, the percentage of endorsed socially desirable items was calculated for each participant. In order to integrate the data from both versions of the M-CSDS, participants’ percentage score on each measure was converted to a proportion score which ranged from 0 to 1, which was used in all analyses. A higher proportion score indicates a greater propensity to give socially desirable responses. Several studies have found that the MCSDS is a reliable and valid measure of social desirability (107, 109, 110).

RESULTS

Preliminary Analyses

Participants vs. Non-participants

Given that not all eligible participants were seen for follow-up, it is important to see to what extent the participants vs. non-participants were similar with regard to baseline characteristics, in part to gauge the internal validity of the sample (111).

The non-participants consisted of three subgroups: (1) patients who were eligible to participate in the study but were not contacted ($n = 163$), (2) patients who declined to participate ($n = 6$), and (3) patients who were not successfully traced ($n = 19$). Two sets of analyses were conducted to compare study participants vs. non-participants. First, the participants were compared to the patients who were eligible but not contacted. Second, the participants were compared to those who declined to participate and to those where contact was attempted but not successfully traced. Group comparisons were conducted on

five demographic variables (age at assessment in childhood, IQ, ethnicity, and parents’ marital status and social class), parent-report of behavior problems on the CBCL (three indices), and nine measures of childhood sex-typed behavior.

Of these 17 variables, there was only one significant difference between the 139 boys in the study compared to the 163 boys who were eligible to participate but were not contacted: participants had a higher IQ than non-participants, $t_{(289)} = 2.01$, $p = 0.046$.¹¹ The effect size for this comparison was small (unpooled $d = 0.22$) [for details, see (77), Tables 5, 6]. When compared to the six cases where participation in the study was declined and to the 19 cases where the families could not be traced, there was also only one significant difference: parent’s marital status, $\chi^2_{(2)} = 9.02$, $p = 0.011$. The participants did not differ significantly from the non-participants who refused; however, they differed significantly from the cases that could not be traced, $\chi^2_{(1)} = 6.39$, $p = 0.012$. The participants were more likely to have originated within a two-parent household than those who could not be traced. The comparison between the non-participants who refused and those who could not be traced approached significance ($p = 0.056$, Fisher’s exact test). Again, the non-participants who could not be traced were more likely to have come from a family composition that was not two-parent. A further summary of comparisons between the participants and those who declined or could not be traced can be found in the **Supplementary Material**.

Participants: Method of Recruitment

Using t -tests or chi-square tests, the 107 participants who entered the study through research contact were compared to the 32 participants who were recruited into the study after they had re-contacted the clinic for clinical reasons on the demographic variables, CBCL behavior problems in childhood, and the measures of childhood sex-typed behavior. There were no significant differences between the two groups on the demographic variables of age at assessment, ethnicity or parents’ social class and marital status ($ps > 0.05$). The comparison on childhood IQ approached significance, $t_{(137)} = 1.97$, $p = 0.051$, with the research entry participants having, on average, a higher IQ than the clinical entry participants. On the CBCL, there was a significant difference on Internalizing problems only, $t_{(137)} = -2.02$, $p = 0.046$, with the clinical entry participants rated by their parents as having more internalizing problems compared to the research entry participants. Of the nine measures of childhood sex-typed behavior, the two groups differed significantly on three: (1) free play, $t_{(119)} = -2.11$, $p = 0.037$, (2) the Gender Identity Interview for Children, $t_{(83)} = -2.09$, $p = 0.04$, and (3) the Gender Identity Questionnaire for Children, $t_{(95)} = 2.39$, $p = 0.019$, with the clinical entry participants having, on average, more childhood cross-gender behavior than the research entry participants. The percentage of clinical entry participants who were threshold for the diagnosis of GID in childhood did not differ significantly from the research entry participants (75.8 vs. 59.8%), $\chi^2_{(1)} = 1.83$. Of the 32 clinical entry participants, 8 had re-contacted the clinic because

¹¹IQ data were not available for 11 of the 163 boys who were eligible for the study but were not contacted.

of gender dysphoria. The above-described comparisons were repeated to compare the research and clinical entry participants but with these 8 participants excluded. With the eight participants who contacted the clinic for gender dysphoria removed, there were no significant group differences on demographic variables, CBCL behavior problems, and measures of childhood sex-typed behavior (all $ps > 0.05$).

Gender Identity at Follow-Up

Appendix 2 in Supplementary Material shows the follow-up data for gender identity and sexual orientation for each participant. Of the 139 participants, 17 (12%) were classified as persisters and the remaining 122 (88%) were classified as desisters. The age at the time of follow-up did not differ significantly between the persisters (Mean, 20.12 years; SD = 5.54) and desisters (Mean, 20.64 years; SD = 5.19), $t_{(137)} < 1$. Of the 107 participants who, for research purposes only, were contacted for the follow-up study, 10 (9%) were classified as persisters; of the 32 participants who were recruited into the study after they were seen for some type of clinical concern, 7 (22%) were classified as persisters. The difference in persistence rate as a function of recruitment entry type was not significant, $\chi^2_{(1)} = 2.53$, $p = 0.112$. The difference in persistence rate between those patients seen for the face-to-face assessment vs. those who were not (14.5 vs. 3.4%) was also not significant, $\chi^2_{(1)} = 1.70$, $p = 0.192$. **Supplementary Table 1** summarizes information on some domains of gender role outcome for the 17 participants classified as having persistent gender dysphoria.

For the 42 participants where the GDIQ was used to determine gender identity status at follow-up, four were classified as persisters and 38 were classified as desisters. Of the 38 desisters, three endorsed one item and the remainder endorsed none of the items.¹² The four participants classified as persisters endorsed between three and five items.

For the 64 participants where the GIDYQ-AA was used to determine gender identity status at follow-up, 12 were classified as persisters and 52 were classified as desisters. All 52 desisters had a mean score >3.00 on the GIDYQ-AA. Of the 12 persisters, 10 had a mean score ≤ 3.00 and two had mean scores that were >3.00 . In spite of having mean scores on the GIDYQ-AA that were above the recommended cutoff for caseness (95), these two participants were considered persisters because their clinical interview data indicated that they were experiencing significant gender dysphoria. Thus, clinical judgment was used to make the final classification for these two participants.

For the remaining 33 participants, clinical interview, parent-report or chart data were used to classify the percentage who were persisters ($n = 1$; 3%) or desisters ($n = 32$; 97%).

The persistence rate of gender dysphoria was examined as a function of participants' GID diagnostic status in childhood (threshold vs. subthreshold). Of the 88 participants who met the full diagnostic criteria for GID in childhood, 12 (13.6%) were classified as persisters and the remaining 76 (86.4%) were

not. Of the 51 participants who were subthreshold for the GID diagnosis in childhood, 5 (9.8%) were classified as persisters and the remaining 46 (90.2%) were not. A chi-square analysis indicated that the rate of persistence did not differ significantly between the threshold and subthreshold groups, $\chi^2_{(1)} < 1$.

Over the years, prevalence rates for gender dysphoria in adults have varied considerably. The variation is likely a function of many factors, including definition, time period, and source of ascertainment. For example, in the Standards of Care of the World Professional Association for Transgender Health (112), probably relying on an estimate given in the DSM-IV-TR, the prevalence of gender dysphoria in adult males was estimated to be 1 in 30,000. In the meta-analysis by Arcelus et al. (113), the prevalence in adult males was estimated at 1 in 14,705. Lastly, Zhang et al.'s (114) review of recent population-based surveys estimated the prevalence of a self-reported transgender identity in adults to range between 0.33 and 0.53% (males and females combined). Regardless of which base rate figure one might choose to use as a point of comparison, the persistence rate of 12% (while low in an absolute sense) would be considerably higher than what one would detect in the general population.

Sexual Orientation at Follow-Up

Table 2 shows the Kinsey ratings for sexual orientation in fantasy. Data were not available for 10 participants, all of whom were desisters with regard to gender dysphoria. Based on the global rating for sexual orientation in fantasy, 43 (33.3%) participants were classified as gynephilic in fantasy and 82 (63.6%) were classified as biphilic/androphilic in fantasy. In the remaining four (3.1%) cases, the participants were classified as having no sexual fantasies and, therefore, a Kinsey rating could not be assigned.¹³ In all four cases, the participants were desisters. Of the 17 participants classified as persisters, 1 (5.9%) was gynephilic in fantasy and 16 (94.1%) were biphilic/androphilic in fantasy. For participants assigned a Kinsey rating between 0 and 6 in fantasy, we correlated the interviewer's Kinsey rating with the participants' responses on the EROS in which their mean gynephilic score was subtracted from their mean androphilic score. This yielded an $r(101) = 0.86$, $p < 0.001$.

Table 2 also shows the Kinsey ratings for sexual orientation in behavior. Data were available for 108 participants. Based on the global rating for sexual orientation in behavior, 29 (26.9%) participants were classified as gynephilic and 51 (47.2%) were classified as biphilic/androphilic. The remaining 28 (25.9%) participants did not report any sexual behaviors in the 12 months preceding the follow-up assessment. For participants assigned a Kinsey rating between 0 and 6 in behavior, we correlated the

¹²By "endorsed," we mean that the participants answered other than "never" on Items 1–4 or response options d–e for Item 5 (see **Appendix 1** in Supplementary Material).

¹³For 104 participants, the Kinsey rating in fantasy was based on the information provided in the face-to-face interview. For 21 other participants, the Kinsey rating in fantasy was based on self-report (by telephone), information available in the participant's health record, or parent-report. Participants were assigned a Kinsey rating of 6 if the participant self-identified as "gay" or if the health record indicated that the patient was "homosexual" or gay, etc. Participants were assigned a Kinsey rating of 0 if the patient self-identified as "straight" or "heterosexual," etc. A chi-square test showed that the percentage of participants who were classified as Kinsey 0–1 vs. 2–6 did not differ significantly as a function sexual orientation ascertainment method, $\chi^2_{(1)} = 1.49$.

TABLE 2 | Kinsey ratings for sexual orientation in fantasy and behavior.

Variable	Kinsey rating (fantasy) ^a															
	0		1		2		3		4		5		6		No fantasy	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Crush	36	36.7	0	0	2	2.0	4	4.1	2	2.0	11	11.2	29	29.6	14	14.3
Visual	31	31.6	1	1.0	2	2.0	10	10.2	3	3.1	12	12.2	29	29.6	10	10.2
Dreams	13	13.3	1	1.0	1	1.0	4	4.1	3	3.1	3	3.1	27	27.6	46	46.9
Masturbation	21	21.9	2	2.1	3	3.1	6	6.3	2	2.1	7	7.3	33	34.4	22	22.9
Global fantasy rating	40	31.0	3	2.3	3	2.3	8	6.2	2	1.6	14	10.9	55	42.6	4	3.1
	Kinsey rating (behavior) ^a															
	0		1		2		3		4		5		6		No sexual behavior	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Holding hands	26	26.3	0	0	0	0	5	5.1	1	1.0	1	1.0	35	35.4	31	31.3
Kissing	21	21.2	0	0	0	0	6	6.1	2	2.0	2	2.0	34	24.3	34	34.3
Genital/breast contact	13	13.1	0	0	0	0	3	3.0	2	2.0	1	1.0	35	35.4	45	45.5
Intercourse	8	8.2	0	0	0	0	3	3.1	2	2.0	0	0	27	27.6	58	59.2
Global behavior rating	28	25.9	1	0.9	0	0	4	3.7	3	2.8	1	0.9	43	39.8	28	25.9

^a0 = Exclusively gynephilic to 6 = Exclusively androphilic.

interviewer's Kinsey rating with the participants' responses on the SHQ in which their mean gynephilic score was subtracted from their mean androphilic score. This yielded an $r(75) = 0.79, p < 0.001$.

For those participants who could be assigned a Kinsey rating (i.e., excluding those participants who did not report any sexual fantasies or behavior or for whom data were not available), the correlation between Kinsey global fantasy and global behavior ratings was very strong, $r(78) = 0.92, p < 0.001$.

Group Classification as a Function of Gender Identity and Sexual Orientation in Fantasy at Follow-Up¹⁴

Combining gender identity (i.e., persistor or desister) and sexual orientation in fantasy (i.e., gynephilic or biphilic/androphilic) at follow-up, the participants were classified into one of four outcome groups (for which we had all of the relevant data): (1) persistence of gender dysphoria with a biphilic/androphilic sexual orientation ($n = 16$); (2) desistance of gender dysphoria with a biphilic/androphilic sexual orientation ($n = 66$); (3) desistance of gender dysphoria with a gynephilic sexual orientation ($n = 42$); and (4) persistence of gender dysphoria with a gynephilic sexual orientation ($n = 1$). The participants who reported no sexual fantasies ($n = 4$) could not be included in this outcome classification. Given that only one participant was classified as gender dysphoric with a co-occurring gynephilic sexual orientation (Group 4), this category was excluded from subsequent analyses that compared these outcome groups.

¹⁴Given the strong correlation between Kinsey fantasy and behavior ratings and that there were fewer missing data on the Kinsey fantasy variable, participants were classified into one of the four outcome groups based on their fantasy ratings.

Demographic Characteristics in Childhood as a Function of Gender Identity and Sexual Orientation in Fantasy

Table 3 shows the demographic variables in childhood as a function of group. One-way ANOVAs and chi-square were conducted to evaluate whether the outcome groups differed on these variables. The groups differed significantly on four of the five childhood demographic variables. Duncan's multiple range test for unequal Ns showed that the biphilic/androphilic persisters were, on average, significantly older at the time of the childhood assessment than both the gynephilic desisters and the biphilic/androphilic desisters, who did not differ significantly from each other. The biphilic/androphilic desisters had, on average, a higher IQ than the biphilic/androphilic persisters but did not differ significantly from the gynephilic desisters. There was no significant difference in childhood IQ score between biphilic/androphilic persisters and gynephilic desisters. The biphilic/androphilic persisters were significantly more likely to come from a lower social class background compared to the gynephilic desisters and the biphilic/androphilic desisters, who did not differ significantly from each other (see also Figure 1). The biphilic/androphilic desisters were more likely to be living with both parents compared to the biphilic/androphilic persisters. There was no significant difference on marital status between the two desister groups.

The demographic variables from childhood on which the three groups differed—age at assessment, IQ, social class, and marital status—were significantly correlated (r s ranged from $|0.32-0.58|$) [see Table 12 in (77)]. To evaluate the predictive status of these variables on group outcome at follow-up, a multinomial logistic regression was performed. Table 4 shows the results. For these analyses, the biphilic/androphilic desisters served as the reference

TABLE 3 | Demographic characteristics as a function of group.

Variable		Group			F or χ^2	p	η^2 or Cramer's V
		Persisters Biphilic/ Androphilic (n = 16)	Desisters Biphilic/ Androphilic (n = 66)	Desisters Gynephilic (n = 42)			
Childhood							
Age (in years)	M	8.85	6.96	7.49	3.57	0.031	0.06
	SD	1.67	2.69	2.62			
IQ ^a	M	101.63	110.20	103.18	3.77	0.026	0.06
	SD	14.81	14.56	15.16			
Social class ^b	M	23.76	44.97	39.44	15.30	<0.001	0.20
	SD	10.22	13.64	15.91			
Marital status^c							
Two-parent	N (%)	7 (43.8)	49 (74.2)	24 (57.1)	6.74	0.034	0.23
Other	N (%)	9 (56.3)	17 (25.8)	18 (42.9)			
Ethnicity							
Caucasian	N (%)	14 (87.5)	58 (87.9)	32 (76.2)	2.77	0.250	0.14
Other	N (%)	2 (12.5)	8 (12.1)	10 (23.8)			
Follow-up							
Age at follow-up (in years) ^d	M	20.32	22.13	17.85	10.41	<0.001	0.15
	SD	5.67	4.97	3.95			
IQ at follow-up ^{a,e,f}	M	99.07	110.47	104.19	3.82	0.025	0.07
	SD	16.29	13.54	17.50			
Follow-up interval (in years)	M	11.47	15.17	10.36	9.63	<0.001	0.04
	SD	6.77	6.03	4.85			
Social desirability ^g	M	0.44	0.43	0.52	3.07	0.051	0.07
	SD	0.17	0.18	0.19			

^aFull-Scale IQ was obtained with age-appropriate Wechsler intelligence scales.

^bHollingshead's (78) Four Factor Index of Social Status (absolute range, 8–66).

^cOther included the following family constellations: single parent, separated, divorced, living with relatives, or in the care of a child protection agency.

^dInterval denotes the time between childhood assessment and follow-up assessment.

^eFull Scale IQ was estimated using four subtests: Vocabulary, Comprehension, Block Design, and Object Assembly.

^fAn IQ score was available only for participants who completed the face-to-face assessment.

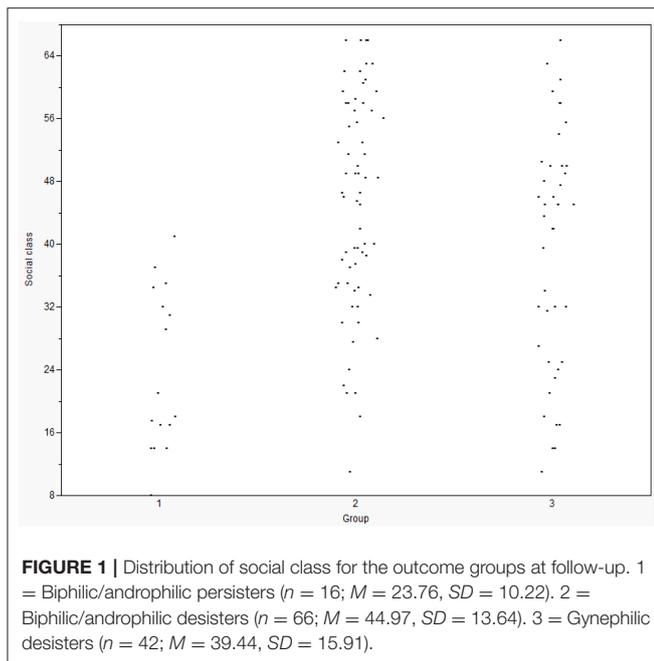
^gAbsolute range, 0.00–1.00. Higher score indicates a greater propensity to give socially desirable responses. Age at follow-up, IQ at follow-up, social class, and parent's marital status were co-varied.

group. Each coefficient, B , represents the change in the log odds for Group for a 1-unit increase in the corresponding predictor, controlling for all other predictors in the model. The next column presents the standard error (SE) for each B . The Wald statistic was the quantity used to determine the significance level of each predictor variable. The quantity, e^B , is the multiplicative change in the odds of being classified as a biphilic/androphilic persister (Model 1) or a gynephilic desister (Model 2) for a 1-unit increase in the corresponding predictor, and thus $100 \times (e^B - 1)$ represents the percentage change in the odds ratio for a 1-unit increase in that predictor (115).

It can be seen from **Table 4** that only social class had a significant contribution to the prediction of group outcome at follow-up (see also **Figure 1**). The biphilic/androphilic persisters had a 13% increase in odds of coming from a lower social class background compared to the biphilic/androphilic desisters.

However, social class did not predict outcome when the two desister groups were compared.

Table 3 also shows the variables of age, IQ, and social desirability scores at follow-up as a function of group. One-way ANOVAs revealed that both age and IQ differed significantly among the three groups ($ps < 0.01$), but social desirability scores did not. Duncan's multiple range test for unequal Ns showed that the gynephilic desisters were, on average, younger than both the biphilic/androphilic persisters and the biphilic/androphilic desisters (both $ps < 0.05$), who did not differ significantly from each other. Regarding IQ at follow-up, the results were similar to those for IQ in childhood. The biphilic/androphilic desisters had, on average, a higher IQ than the biphilic/androphilic persisters ($p < 0.05$) but did not differ significantly from the gynephilic desisters. There was no significant difference in IQ between the biphilic/androphilic persisters and the gynephilic desisters.



Childhood Sex-Typed Behavior as a Function of Gender Identity and Sexual Orientation at Follow-Up Supplementary Table 2 shows the means or percentage scores (for dichotomous measures) of the nine sex-typed measures obtained at the assessment in childhood as a function of the three outcome groups. ANCOVAs (with age at assessment, IQ, social class, and marital status covaried) or chi-square were used to examine whether the groups differed on any of these variables.¹⁵ There was a significant difference between the groups on four child-report measures (first drawn person on the Draw-a-Person, free play, Gender Identity Interview, and cross-sex peer preference on the Playmate and Play Style Preferences Structured Interview, and one parent-report measure (Gender Identity Questionnaire for Children). A statistical summary of these individual measures can be found in the **Supplementary Text** and the data are shown in **Supplementary Table 2**.

The childhood sex-typed behavior measures on which the groups differed were all significantly correlated (r s ranged from $|0.30-0.76|$) [reported in (77), Table 15].¹⁶ From these six measures (first drawn person on the Draw-a-Person, free play, Gender Identity Interview, cross-sex peer preference on the Playmate and Play Style Preferences Structured Interview, cross-sex toy preference on the Playmate and Play Style Preferences Structured Interview, and the Gender Identity Questionnaire for Children), a composite score of childhood sex-typed behavior was derived for each participant by taking the average of the

six variables (each expressed as z -scores).¹⁷ A higher composite z -score indicates more cross-gender behavior at the assessment in childhood.

To evaluate the influence of childhood sex-typed behavior and demographic variables on group outcome at follow-up, a multinomial logistic regression was performed using the composite score and the demographic variables on which the groups differed—age at assessment, IQ, and social class—as predictor variables. It can be seen from **Table 5** that both social class and the composite score of childhood sex-typed behavior were significant predictors of group outcome at follow-up in the first model, which compared the biphilic/androphilic persisters to the biphilic/androphilic desisters.

The biphilic/androphilic persisters had a 274% increase in odds of having a higher composite score (i.e., more childhood cross-gender behavior) and an 11% reduction in the odds of coming from a higher social class compared to the biphilic/androphilic desisters. Age at childhood assessment and IQ did not have a significant effect on group outcome (both p s > 0.05). In the second model, which compared the gynephilic desisters to the biphilic/androphilic desisters, the only significant predictor of group outcome was the composite measures of sex-typed behavior. The biphilic/androphilic desisters had a 48% increase in odds of having a higher composite score compared to the gynephilic desisters.

DISCUSSION

Methodological Issues

We were not able to recruit into the study all eligible patients; however, our analyses which compared the participants vs. the non-participants did not show any substantive or pervasive differences with regard to the baseline assessment characteristics, suggesting that the internal validity of the sample was not grossly compromised (111). The majority of follow-up participants were recruited for research purposes; however, a minority entered the study after having been seen in adolescence for some clinical issue. There was some evidence that the patients who were enrolled in the study after recontacting the clinic were, on average, more extreme in their gender-variant behavior in childhood; however, the percentage who were threshold for the GID diagnosis in childhood did not differ significantly between the two subgroups. Although the percentage of persisters was higher in the subgroup that had recontacted the clinic than the subgroup recruited for research purposes only (22% vs. 9%), the difference was also not statistically significant. If anything, the direction of the difference would suggest that the overall rate of persistence may have been slightly overestimated had we relied entirely on a “research-only” follow-up sample.

Another methodological issue is that we relied on different metrics to assess gender identity and gender dysphoria at follow-up. For example, we replaced the GDIQ with the GIDYQ-AA as we viewed the latter as a better measure; in some instances,

¹⁵The ANCOVA model was adjusted to accommodate a categorical covariate.

¹⁶Although the groups did not differ significantly on cross-sex toy preference on the PPPSI, this measure is included here because there was a trend in the direction of a significant group difference.

¹⁷For some participants, data were not available on all six measures. In these cases, the composite score was the average of the number of variables for which there were data.

TABLE 4 | Multinomial logistic regression of group outcome at follow-up.

Predictor	Biphilic/Androphilic persisters					Gynephilic desisters				
	B	SE	Wald	p	e ^B	B	SE	Wald	p	e ^B
Age at assessment	0.11	0.14	0.62	0.433	1.12	-0.02	0.09	0.03	0.856	0.98
IQ	0.02	0.03	0.85	0.358	1.02	-0.02	0.02	1.91	0.167	0.98
Social class	-0.14	0.04	13.66	<0.001	0.87	-0.01	0.02	0.13	0.716	0.99
Marital status	0.76	0.80	0.88	0.349	0.47	-0.43	0.52	0.70	0.402	1.54

Reference group is the Biphilic/Androphilic Desisters. This group was chosen as the reference because it had the largest group size.

TABLE 5 | Multinomial logistic regression predicting group outcome at follow-up.

Predictor	Biphilic/Androphilic persisters					Gynephilic desisters				
	B	SE	Wald	p	e ^B	B	SE	Wald	p	e ^B
Age at assessment	0.26	0.16	2.90	0.09	1.30	-0.14	0.11	1.55	0.21	0.87
IQ	0.02	0.03	0.58	0.45	1.02	-0.03	0.01	2.77	0.10	0.97
Social class	-0.12	0.03	12.28	<0.001	0.89	-0.01	0.01	0.51	0.47	0.99
Composite z-score	1.32	0.55	5.82	0.02	3.74	-0.66	0.31	4.38	0.04	0.52

Reference group is the Biphilic/Androphilic Desisters. This group was chosen as the reference because it had the largest group size. A preliminary analysis with marital status included as a predictor variable showed that it did not have a significant effect and was, therefore, excluded in the final regression model. As suggested by Reviewer 3, per Benjamin et al. (116), for the "discovery of new effects," p-values between 0.05 and 0.005 should be viewed as "suggestive" (i.e., informative, but cautiously interpreted), and p-values < 0.005 as "significant" (i.e., stronger evidence for the implausibility of a difference merely by chance).

we relied solely on interview data or information available in the patient's medical chart. However, we did not detect any substantive difference in the percentage of persisters across these different sources of information and thus do not believe that such method variance challenges the validity of the findings.

Although a minority of participants were seen on more than one occasion for follow-up, the majority were not. Thus, our results and interpretation of the follow-up data are largely limited to one "moment in time," at a mean age of 20.58 years. It would, of course, be of value to have additional follow-up of the patients as they move further into adulthood in order to assess the stability (or lack thereof) of the data with regard to both gender identity and sexual orientation. In our own clinical experience, for example, we have observed that some of the patients seen during adolescence "fluctuated" between self-identifying as transgender and self-identifying as gay. Others have noted that a small number of apparent or presumed desisters during adolescence subsequently identified as transgender when seen at a later point in time (117).

Summary of Key Findings

The present study provided follow-up data with regard to gender identity and sexual orientation in boys referred clinically for gender dysphoria. There were three key findings: (1) the persistence of gender dysphoria was relatively low (at 12%), but obviously higher than what one would expect from base rates in the general population; (2) the percentage who had a biphilic/androphilic sexual orientation was very high (in fantasy: 65.6% after excluding those who did not report any sexual fantasies; in behavior: 63.7% after excluding those who did not have any interpersonal sexual experiences), markedly higher than what one would expect from base rates in the general

population; (3) we identified some predictors (from childhood) of long-term outcome when contrasting the persisters with a biphilic/androphilic sexual orientation with the desisters with a biphilic/androphilic sexual orientation and when contrasting the desisters with a biphilic/androphilic sexual orientation and the desisters with a gynephilic sexual orientation.

The 12% persistence rate was somewhat lower than the overall persistence rate of 17.4% from the prior follow-up studies of boys combined. When compared to the three most methodologically sound follow-up studies, the persistence rate was higher than the 2.2% rate found by Green (47), but lower than the 20.3% rate found by Wallien and Cohen-Kettenis (52) and the 29.1% rate found by Steensma et al. (51). There is one methodological caveat regarding the Steensma et al. study that is worth noting. In their study, the mean interval between assessment and follow-up was relatively short (7.21 years). The patients were eligible for follow-up if they were at least 15 years of age. Given the relatively short interval between the assessment in childhood and the follow-up assessment in adolescence, this meant that patients who had been assessed at younger ages in childhood would not have been old enough to participate in the follow-up assessment. Given that Steensma et al. found that (older) age at the time of the assessment in childhood was a significant predictor of persistence, it is conceivable that their persistence rate was an overestimate. Nonetheless, in the broadest sense, our data were quite consistent with the general finding from the prior follow-up studies that desistance from gender dysphoria is by far the more common outcome.

In our study, we did not find that persistence was more common among boys who were threshold for the diagnosis of GID when compared to the boys who were subthreshold (13.6% vs. 9.8%) although the pattern was in the same direction

as that found by Wallien and Cohen-Kettenis (52) and Steensma et al. (51). We would, therefore, argue that the threshold-subthreshold distinction should not be abandoned in future follow-up studies although such studies might profit from using a symptom count of DSM indicators in addition to the dichotomous coding of the diagnosis as threshold vs. subthreshold. Consistent with both Wallien and Cohen-Kettenis and Steensma et al., our composite measure of sex-typed behavior in childhood was a significant predictor of outcome in that the patients classified as persisters with a biphilic/androphilic sexual orientation had more severe gender-variant behavior than the patients classified as desisters with a biphilic/androphilic sexual orientation; in addition, desisters with a biphilic/androphilic sexual orientation had more gender-variant behavior than the desisters with a gynephilic sexual orientation. Thus, dimensional measurement of gender identity and gender role behaviors from childhood provides added nuance in characterizing longer term trajectories with regard to both gender identity and sexual orientation.

With regard to sexual orientation at follow-up, the percentage of patients with a biphilic/androphilic sexual orientation in either fantasy or behavior was reasonably similar to those reported on in the prior follow-up studies which included standardized assessment measures (47, 51, 52). This finding also converges with three representative, general population prospective studies (118–120) and many retrospective studies (43) which document a significant association between patterns of gender-typed behavior in childhood and later sexual orientation.

The multinomial logistic regression analysis (**Table 4**) also showed a trend for the persisters with a biphilic/androphilic sexual orientation to be older at the time of the assessment in childhood compared to the desisters with a biphilic/androphilic sexual orientation; however, when the composite measure of sex-typed behavior in childhood was added to the equation (**Table 5**), age at assessment in childhood no longer showed such a trend [cf. Steensma et al. (51)]. In our smaller study of girls with GID (46), the persisters were, on average, 2.5 years older than the desisters at the time of the assessment in childhood (11.08 vs. 8.59 years) although the difference was not significant. It is our view that age at the time of a childhood assessment in relation to long-term outcome should continue to be examined in future follow-up studies.

Social class was a significant predictor of outcome: the persisters with a biphilic/androphilic sexual orientation were from a lower social class background compared to the desisters with a biphilic/androphilic sexual orientation (even after controlling for the other demographic variables). Why might this be the case? Because we had not made formal a priori predictions of outcome regarding any of our demographic variables, it is, of course, important to see whether or not it will be replicated in new follow-up studies. At present, our interpretation of the social class effect reflects on its relationship to other literatures.

One possibility pertains to the notion that acceptance of a gay or homosexual sexual identity is less in “working class” subculture (121). If this is, in fact, the case, it has been argued that transitioning from male to female—the so-called “homophobic” hypothesis with regard to gender dysphoria in adults (122)—would allow an androphilic sexual orientation to be more

acceptable. Future studies would need to systematically examine whether boys with persistent GID first attempt to live as gay men before transitioning to the female gender role and whether or not this temporal sequence, when it occurs, is related to social class background.

In the present study, it could be hypothesized that the parents of persisters held less favorable views of androphilia (homosexuality) compared to the desisters and thus predisposed to persistence in order to “normalize” one’s sexual orientation. However, this is simply a conjecture as parental attitudes toward homosexuality were not measured in the study sample. Indeed, none of the follow-up studies to date on boys with gender dysphoria have specifically examined attitudes toward homosexuality as a predictor of outcome.

Social class could also be a proxy for other explanatory factors. For example, in the present study, a lower social class background was significantly correlated with age at assessment in childhood ($r = 0.44$) and families where there had been a separation/divorce, etc. ($r = 0.58$). If one wanted to make the case that a later age at assessment might be associated with persistence (for a variety of reasons), perhaps social class is associated with a “delay” in seeking out an assessment and possible treatment (e.g., family stress, various other mental health challenges in the child and/or the family, etc.). In one study comparing the demographic characteristics of children vs. adolescents clinic-referred for gender dysphoria, it was found that the adolescents were more likely than the children to come from a lower social class background and from families in which there had been a separation/divorce, etc. (123).

Clinical Implications

What clinical implications might be drawn from our data on the persistence and desistence rates of gender dysphoria in children? First, it should be recognized that the boys in the current study were seen during a period of time when treatment recommendations, if such were made, often aimed to reduce the gender dysphoria between the child’s felt gender identity and biological sex. If one peruses the treatment literature, such recommendations were carried out using many therapeutic modalities: psychotherapy or psychoanalysis, behavior therapy, group therapy, parent-counseling, and interventions in the naturalistic environment, such as encouragement of same-sex peer relations [see, e.g., (124–126); for reviews, see (127, 128)].¹⁸

¹⁸This “broad stroke” summary of therapeutic goals is not meant to minimize the complexity of ethical issues regarding how treatment has been conceptualized over the years [see, e.g., (129–133)]. In the early years, treatment recommendations included other goals: for example, Bakwin (44) wrote that “Suggestions for management... [were]... designed to encourage gender appropriate behavior and to prevent homosexuality” [p. 620, emphasis added; see also (134)]. Rekers (135) was subsequently quite transparent regarding the influence of his own religious beliefs in formulating treatment goals, sometimes congruent with parents’ religious beliefs (see p. 131). Prayer appears to have guided Rekers’ selection of behavior therapy as a treatment modality for the treatment of his patients with childhood GID (p. 131). Money and Russo (50) wondered what the course of psychosexual differentiation might be if “a group of boys with discordance of gender identity/role [were] transferred from the home of origin to, say, a children’s recovery center or foster home... as happens in the case of child-abuse dwarfism...” (p. 40). In our own clinic, although some parents might have desired or requested that treatment be designed in order to prevent homosexuality, this was a goal that we never endorsed [see (136), pp. 391–393]. Over the years, many secular-minded

In our own sample, the kinds of treatments that the boys received, if any, were quite variable but it is beyond the scope of this article to describe them in general [however, for examples, see (136, 140, 141)]. It can, however, be said with certainty that the vast majority of boys were seen during a particular period of time when the therapeutic approach of recommending or supporting a gender social transition prior to puberty was not made. Indeed, in the current study, there was only one patient who had socially transitioned prior to puberty (at the suggestion and support of the professionals involved in this individual's care) and this particular patient was one of the persisters with a biphilic/androphilic sexual orientation. Second, it should also be recognized that, for the boys seen in the current study, none who were in late childhood and had (likely) entered puberty (Tanner Stage 2) had received puberty-blocking hormone treatment (GnRH analogs) to suppress somatic masculinization (142, 143) until sometime during adolescence.

In contrast, in recent years, it has become more common for some clinicians to recommend a gender social transition prior to puberty [e.g., (69, 144–147); for discussion, see (148–150)]. It has also become more common for parents to have already implemented a gender social transition on their own, without any formal input from a health professional (151). As argued by Zucker (64, 152), this is a very different type of psychosocial treatment designed to reduce gender dysphoria when compared to the other kinds of treatments noted above that have been recommended over the years.

The study by Steensma et al. (51), which found the highest rate of persistence, included some patients who had made a partial or complete gender social transition prior to puberty and this variable proved to be a unique predictor of persistence (see the Introduction). Rae et al. (153) recruited from a variety of community groups a sample of 85 markedly gender non-conforming children (Mean age, 7.5 years), none of whom had socially transitioned at a baseline assessment. At the time of follow-up, at a mean of 2.1 years later, 36 (42.3%) had socially transitioned and 49 (57.6%) had not. Using a composite of various metrics of gender identity and gender role behaviors, Rae et al. found that those who subsequently socially transitioned had more extreme gender-variant behavior at baseline than those who had not. Thus, this short-term follow-up study was consistent

with the longer-term findings reported on by Wallien and Cohen-Kettenis (52), Steensma et al. (51), and the present study.

To date, however, there are no long-term follow-up studies of clinic-referred samples of children who had all socially transitioned prior to puberty. Future follow-up studies should be able to capture a much larger subgroup of such children and compared to those who have not with regard to long-term outcome with regard to persistence and desistance [e.g., (154)]. The persistence-desistance rates found in this study and the ones preceding it can be used as a comparative benchmark for samples in which a social transition took place prior to puberty.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The research protocol was reviewed and approved by Clarke Institute of Psychiatry (subsequently the Centre for Addiction and Mental Health) and the University of Toronto. All participants who completed the face-to-face assessment gave written informed consent.

AUTHOR CONTRIBUTIONS

DS contributed to the conceptualization, data collection, data analysis, interpretation, and writing of the paper. SB contributed to the conceptualization and interpretation of the study. KZ contributed to the conceptualization, data collection, data analysis, interpretation, and writing of the paper. All authors contributed to the article and approved the submitted version.

FUNDING

DS was supported by an Ontario Graduate Scholarship (2008–2009, 2009–2010, 2010–2011) and the Social Sciences Humanities Research Council (2010–2011). Funding for this study was provided, in part, by the Laidlaw Foundation and internal research funds from the Clarke Institute of Psychiatry.

ACKNOWLEDGMENTS

This article is based on DS's doctoral dissertation at the University of Toronto. Preliminary versions of this article were presented at the 1989 meeting of the Society for Research in Child and Adolescent Psychopathology, Miami Beach, Florida and the 2010 Gender Development Research Conference, San Francisco, California.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsy.2021.632784/full#supplementary-material>

clinicians—although clearly opposed to any type of preventive efforts with regard to sexual orientation—argued in favor of reducing gender dysphoria vis-à-vis natal sex, if that was feasible. Meyer-Bahlburg (125), for example, wrote: "...we cannot rule out the possibility that early successful treatment of childhood GID will diminish the role of a continuation of GID into adulthood. If so, successful treatment would also reduce the need for the long and difficult process of sex reassignment which includes hormonal and surgical procedures with substantial medical risks and complications" (p. 362). Along similar lines, Cohen-Kettenis and Pfäfflin (33) remarked: "Relatively little dispute exists regarding the prevention of transsexualism, though evidence about the effectiveness of treatment in preventing adult transsexualism is also virtually nonexistent" (p. 120). In more recent years, what the best-practice should be for the treatment of gender dysphoria in children has been widely discussed and debated, which highlight the various limitations of treatment effectiveness studies (137–139).

REFERENCES

1. Lewis M. Self-knowledge: a social cognitive perspective on gender identity and sex-role development. In: Lamb ME, Sherrod LR, editors. *Infant Social Cognition: Empirical and Theoretical Considerations*. Hillsdale, NJ: Lawrence Erlbaum Associates (1981). p. 395–414.
2. Maccoby EE. Gender as a social category. *Dev Psychol.* (1988) 24:755–65. doi: 10.1037/0012-1649.24.6.755
3. Ruble DN, Martin CL, Berenbaum SA. Gender development. In: Damon W, Lerner RM, Eisenberg M, editors. *Handbook of Child Psychology: Vol. 3. Social, Emotional, and Personality Development*, 6th ed. New York, NY: Wiley (2006). p. 858–932.
4. Stoller RJ. The sense of maleness. *Psychoanal Q.* (1965) 34:207–18. doi: 10.1080/21674086.1965.11926345
5. Stoller RJ. The sense of femaleness. *Psychoanal Q.* (1968) 37:42–55. doi: 10.1080/21674086.1968.11926450
6. Zucker KJ, VanderLaan DP. The self in gender dysphoria: a developmental perspective. In: Kyrios M, Moulding R, Doron G, Bhar SS, Nedeljkovic M, Mikulinic M, editors. *The Self in Understanding and Treating Psychological Disorders*. Cambridge: Cambridge University Press (2016). p. 222–32. doi: 10.1017/CBO9781139941297.023
7. Turner KL, Brown CS. The centrality of gender and ethnic identities across individuals and contexts. *Soc Dev.* (2007) 16:701–19. doi: 10.1111/j.1467-9507.2007.00403.x
8. Matei A. Raising a Theybie: The Parent Who Wants Their Child to Grow Up Gender-Free. *The Guardian* (2020). Retrieved from: <https://www.theguardian.com/lifeandstyle/2020/jul/08/parent-raising-gender-free-child>
9. Myers K. *Raising Them: Our Adventure in Gender Creative Parenting*. New York, NY: TOPPLE Books/Little A (2020).
10. Witterick K. Dancing in the eye of the storm: the gift of gender diversity to our family. In: Green FG, Friedman M, editors. *Chasing Rainbows: Exploring Gender Fluid Parenting Practices*. Bradford, ON: Demeter Press (2013). p. 21–42.
11. Kleeman JA. The establishment of core gender identity in normal girls. I (a) Introduction; (b) Development of the ego capacity to differentiate. *Archiv Sexual Behav.* (1971) 1:103–16. doi: 10.1007/BF01541055
12. Martin CL, Ruble DN, Szkrybalo J. Cognitive theories of early gender development. *Psychol Bull.* (2002) 128:903–33. doi: 10.1037/0033-2909.128.6.903
13. Paluszny M, Beit-Hallahmi B, Catford JC, Cooley RE, Dull CY, Guiora AZ. Gender identity and its measurement in children. *Compreh Psychiatry.* (1973) 14:281–90. doi: 10.1016/S0010-440X(73)80022-7
14. Slaby RG, Frey KS. Development of gender constancy and selective attention to same-sex models. *Child Dev.* (1975) 46:849–56. doi: 10.2307/1128389
15. Kohlberg L. A cognitive-developmental analysis of children's sex-role concepts and attitudes. In: Maccoby EE, editor. *The Development of Sex Differences*. Stanford, CA: Stanford University Press (1966). p. 82–173.
16. Zucker KJ, Bradley SJ, Kuski M, Pecore K, Birkenfeld-Adams A, Doering RW, et al. Gender constancy judgments in children with gender identity disorder: evidence for a developmental lag. *Archiv Sexual Behav.* (1999) 28:475–502. doi: 10.1023/A:1018713115866
17. Fagot BI, Leinbach MD, Hagan R. Gender labelling and adoption of sex-typed behaviors. *Dev Psychol.* (1986) 22:440–3. doi: 10.1037/0012-1649.22.4.440
18. Zucker KJ. Measurement of psychosexual differentiation. *Archiv Sexual Behav.* (2005) 34:375–88. doi: 10.1007/s10508-005-4336-7
19. Hines M. *Brain Gender*. Oxford: Oxford University Press (2004).
20. Owen Blakemore JE, Berenbaum SA, Liben LS. *Gender Development*. New York, NY: Taylor & Francis Group (2009).
21. Weisgram ES, Dinella LM, editors. *Gender Typing of Children's Toys: How Early Play Experiences Impact Development*. Washington, DC: American Psychological Association (2018).
22. Lippa RA. Gender-related traits in gay men, lesbian women, and heterosexual men and women: the virtual identity of homosexual-heterosexual diagnosticity and gender diagnosticity. *J Pers.* (2000) 68:899–926. doi: 10.1111/1467-6494.00120
23. Bakwin H. Transvestism in children. *J Pediatr.* (1960) 56:294–8. doi: 10.1016/S0022-3476(60)80128-X
24. Friend MR, Schiddel L, Klein B, Dunaeff D. Observations on the development of transvestitism in boys. *Am J Orthopsychiatry.* (1954) 24:563–75. doi: 10.1111/j.1939-0025.1954.tb06128.x
25. Green R, Money J. Incongruous gender role: nongenital manifestations in prepubertal boys. *J Nerv Ment Dis.* (1960) 131:160–8. doi: 10.1097/00005053-196008000-00009
26. Stoller RJ. Male childhood transsexualism. *J Am Acad Child Psychiatry.* (1968) 7:193–209. doi: 10.1016/S0002-7138(09)62167-1
27. Zuger B. Effeminate behavior present in boys from early childhood I. The clinical syndrome and follow-up studies. *J Pediatr.* (1966) 69:1098–107. doi: 10.1016/S0022-3476(66)80301-3
28. Stoller RJ. *Sex and Gender (Vol. I). The Development of Masculinity and Femininity*. New York, NY: Jason Aronson (1968).
29. Green R. *Sexual Identity Conflict in Children and Adults*. New York, NY: Basic Books (1974).
30. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 3rd ed. Washington, DC: American Psychiatric Association (1980).
31. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed. Arlington, VA: American Psychiatric Press (2013).
32. Coates S. Ontogenesis of boyhood gender identity disorder. *J Am Acad Psychoanal.* (1990) 18:414–38. doi: 10.1521/jaap.1.1990.18.3.414
33. Cohen-Kettenis PT, Pfäfflin F. *Transgenderism and Intersexuality in Childhood and Adolescence: Making Choices*. Thousand Oaks, CA: Sage (2003).
34. de Vries ALC, Kreukels BPC, Steensma TD, McGuire JK. Gender identity development: a biopsychosocial perspectives. In: Kreukels BPC, Steensma TD, de Vries ALC, editors. *Gender Dysphoria and Disorders of Sex Development: Progress in Care and Knowledge*. New York, NY: Springer (2014). p. 53–80. doi: 10.1007/978-1-4614-7441-8_3
35. Spivey LA, Edwards-Leeper L. Future directions in affirmative psychological interventions with transgender children and adolescents. *J Clin Child Adolesc Psychol.* (2019) 48:343–56. doi: 10.1080/15374416.2018.1534207
36. Turban JL, Ehrensaft D. Gender identity in youth: treatment paradigms and controversies. *J Child Psychol Psychiatry.* (2018) 59:1228–43. doi: 10.1111/jcpp.12833
37. Zucker KJ. Gender identity disorder in children and adolescents. *Annu Rev Clin Psychol.* (2005) 1:467–92. doi: 10.1146/annurev.clinpsy.1.102803.144050
38. Zucker KJ. Gender dysphoria. In: Lewis M, Rudolph KD, editors. *Handbook of Developmental Psychopathology*, 3rd ed. New York, NY: Springer (2014). p. 683–702. doi: 10.1007/978-1-4614-9608-3_35
39. Zucker KJ, Bradley SJ. *Gender Identity Disorder and Psychosexual Problems in Children and Adolescents*. New York, NY: Guilford Press (1995).
40. Green R. The behaviorally feminine male child: pretranssexual? Pretransvestic? Prehomosexual? Preheterosexual? In: Friedman RC, Richart RM, Vande Wiele RL, editors. *Sex Differences in Behavior*. New York, NY: John Wiley & Sons (1974). p. 301–14.
41. Bell AP, Weinberg MS, Hammersmith SK. *Sexual Preference: Its Development in Men and Women*. Bloomington: Indiana University Press (1981).
42. Whitam F. Childhood indicators of male homosexuality. *Archiv Sexual Behav.* (1977) 6:89–96. doi: 10.1007/BF01541701
43. Bailey JM, Zucker KJ. Childhood sex-typed behavior and sexual orientation: a conceptual analysis and quantitative review. *Dev Psychol.* (1995) 31:43–55. doi: 10.1037/0012-1649.31.1.43
44. Bakwin H. Deviant gender-role behavior in children: relation to homosexuality. *Pediatrics.* (1968) 41:620–9.
45. Davenport CW. A follow-up study of 10 feminine boys. *Archiv Sexual Behav.* (1986) 15:511–7. doi: 10.1007/BF01542316
46. Drummond KD, Bradley SJ, Peterson-Badali M, Zucker KJ. A follow-up study of girls with gender identity disorder. *Dev Psychol.* (2008) 44:34–45. doi: 10.1037/0012-1649.44.1.34
47. Green R. *The "Sissy Boy Syndrome" and the Development of Homosexuality*. New Haven, CT: Yale University Press (1987).
48. Kosky RJ. Gender-disordered children: does inpatient treatment help? *Med J Aust.* (1987) 146:565–9. doi: 10.5694/j.1326-5377.1987.tb120415.x
49. Lebovitz PS. Feminine behavior in boys: aspects of its outcome. *Am J Psychiatry.* (1972) 128:1283–9. doi: 10.1176/ajp.128.10.1283

50. Money J, Russo AJ. Homosexual outcome of discordant gender identity/role in childhood: longitudinal follow-up. *J Pediatr Psychol.* (1979) 4:29–41. doi: 10.1093/jpepsy/4.1.29
51. Steensma TD, McGuire JK, Kreukels BPC, Beekman AJ, Cohen-Kettenis PT. Factors associated with desistance and persistence of childhood gender dysphoria: A quantitative follow-up study. *J Am Acad Child Adolesc Psychiatry.* (2013) 52:582–90. doi: 10.1016/j.jaac.2013.03.016
52. Wallien MSC, Cohen-Kettenis PT. Psychosexual outcome of gender dysphoric children. *J Am Acad Child Adolesc Psychiatry.* (2008) 47:1413–23. doi: 10.1097/CHI.0b013e31818956b9
53. Zuger B. Effeminate behavior present in boys from childhood: ten additional years of follow-up. *Compreh Psychiatry.* (1978) 19:363–9. doi: 10.1016/0010-440X(78)90019-6
54. August GJ, Realmuto GM, Joyce T, Hektner JM. Persistence and desistance of oppositional defiant disorder in a community sample of children with ADHD. *J Am Acad Child Adolesc Psychiatry.* (1999) 38:1262–70. doi: 10.1097/00004583-199910000-00015
55. Zucker KJ. Persistence and desistance of gender identity disorder in children [Discussant]. In: *Paper Presented at the Meeting of the Harry Benjamin International Gender Dysphoria Association*, Gent, Belgium (2003).
56. Cantor JM. Transgender and gender diverse children and adolescents: fact-checking of AAP policy. *J Sex Marit Ther.* (2020) 46:307–13. doi: 10.1080/0092623X.2019.1698481
57. Ristori J, Steensma TD. Gender dysphoria in childhood. *Int Rev Psychiatry.* (2016) 28:13–20. doi: 10.3109/09540261.2015.1115754
58. Zuger B. Early effeminate behavior in boys: outcome and significance for homosexuality. *J Nerv Ment Dis.* (1984) 172:90–7. doi: 10.1097/00005053-198402000-00005
59. Nakamura H. *Follow-up study of children and adolescents with gender identity development issues who attended the specialist Gender Identity Development Unit (GIDU) and who are now 18 or older* (Unpublished dissertation), University of Essex, Essex, United Kingdom (2007).
60. Rafferty J, Committee on Psychosocial Aspects of Child and Family Health, Committee on Adolescence, and Section on Lesbian, Gay, Bisexual, and Transgender Health and Wellness. Ensuring comprehensive care and support for transgender and gender-diverse children and adolescents. *Pediatrics.* (2018) 142:e20182162. doi: 10.1542/peds.2018-2162
61. Tannehill B. The End of the Desistance Myth. *Huffpost* (2016). Retrieved from: https://www.huffingtonpost.com/brynn-tannehill/the-end-of-the-desistance_b_8903690.html
62. Temple Newhook J, Pyne J, Winters K, Feder S, Holmes C, Tosh J, et al. A critical commentary on follow-up studies and “desistance” theories about transgender and gender-nonconforming children. *Int J Transgen.* (2018) 19:212–24. doi: 10.1080/15532739.2018.1456390
63. Winters K. The “80% desistance dictum: is it science? In: Lev AI, Gottlieb AR, editors. *Families in Transition: Parenting Gender Diverse Children, Adolescents, and Young Adults*. New York, NY: Harrington Park Press (2019). p. 88–101.
64. Zucker KJ. The myth of persistence: Response to “A Critical Commentary on Follow-Up Studies and Desistance Theories about Transgender and Gender Non-Conforming Children” by Temple Newhook et al. (2018). *Int J Transgen.* (2018) 19:231–45. doi: 10.1080/15532739.2018.1468293
65. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 3rd ed. rev. Washington, DC: American Psychiatric Association (1987).
66. Zucker KJ. The DSM diagnostic criteria for gender identity disorder in children. *Archiv Sexual Behav.* (2010) 39:477–98. doi: 10.1007/s10508-009-9540-4
67. Zucker KJ, Cohen-Kettenis PT, Drescher J, Meyer-Bahlburg HFL, Pfäfflin F, Womack WM. Memo outlining evidence for change for gender identity disorder in the DSM-5. *Archiv Sexual Behav.* (2013) 42:901–14. doi: 10.1007/s10508-013-0139-4
68. Fast AA, Olson KR. Gender development in transgender preschool children. *Child Dev.* (2018) 89:620–37. doi: 10.1111/cdev.12758
69. Meadow T. *Trans Kids: Being Gendered in the Twenty-First Century*. Berkeley, CA: University of California Press (2018).
70. Olson KR. Prepubescent transgender children: what we do and do not know. *J Am Acad Child Adolesc Psychiatry.* (2016) 55:155–6. doi: 10.1016/j.jaac.2015.11.015
71. Soares C, editor. *The Oxford Dictionary of Current English*, 3rd ed. Oxford: Oxford University Press (2001).
72. Blanchard R, Zucker KJ, Cavacas A, Allin S, Bradley SJ, Schachter DC. Fraternal birth order and birth weight in probably prehomosexual feminine boys. *Horm Behav.* (2002) 41:321–7. doi: 10.1006/hbeh.2002.1765
73. Bouman WP, Suess Schwend A, Motmans J, Smily A, Safer JD, Deutsch MB, et al. Language and trans health. *Int J Transgen.* (2017) 18:1–6. doi: 10.1080/15532739.2016.1262127
74. Intons-Peterson MJ, Reddel M. What do people ask about a neonate? *Dev Psychol.* (1984) 20:358–9. doi: 10.1037/0012-1649.20.3.358
75. Bradley SJ, Steiner BW, Zucker K, Doering RW, Sullivan J, Finegan JK, et al. Gender identity problems of children and adolescents: the establishment of a special clinic. *Can Psychiatr Assoc J.* (1978) 23:175–83. doi: 10.1177/070674377802300309
76. Zucker KJ. Comment on “Serving Transgender Youth: Challenges, Dilemmas, and Clinical Examples” by Tishelman et al. (2015). *Prof Psychol Res Pract.* (2015) 46:306. doi: 10.1037/pro0000030
77. Singh D. *A follow-up study of boys with gender identity disorder* (Unpublished doctoral dissertation), University of Toronto, Toronto, ON, Canada (2012).
78. Hollingshead AB. *Four Factor Index of Social Status*. Unpublished manuscript. New Haven, CT: Department of Sociology, Yale University (1975).
79. Achenbach TM. *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Burlington, VT: University of Vermont Department of Psychiatry (1991).
80. Zucker KJ, Finegan JK, Doering RW, Bradley SJ. Human figure drawings of gender-problem children: a comparison to siblings, psychiatric, and normal controls. *J Abnorm Child Psychol.* (1983) 11:287–98. doi: 10.1007/BF00912092
81. Zucker KJ, Doering RW, Bradley SJ, Finegan JK. Sex-typed play in gender-disturbed children: a comparison to sibling and psychiatric controls. *Archiv Sexual Behav.* (1982) 11:309–21. doi: 10.1007/BF01541592
82. Alexander GM, Hines M. Gender labels and play styles: their relative contribution to children’s selections of playmates. *Child Dev.* (1994) 65:869–79. doi: 10.2307/1131424
83. Fridell SR, Owen-Anderson A, Johnson LL, Bradley SJ, Zucker KJ. The playmate and play style preferences structured interview: a comparison of children with gender identity disorder and controls. *Archiv Sexual Behav.* (2006) 35:729–37. doi: 10.1007/s10508-006-9085-8
84. Zucker KJ, Lozinski JA, Bradley SJ, Doering RW. Sex-typed responses in the Rorschach protocols of children with gender identity disorder. *J Pers Assess.* (1992) 58:295–310. doi: 10.1207/s15327752jpa5802_9
85. Zucker KJ, Bradley SJ, Lowry Sullivan CB, Kuksis M, Birkenfeld-Adams A, Mitchell JN. A gender identity interview for children. *J Pers Assess.* (1993) 61:443–56. doi: 10.1207/s15327752jpa6103_2
86. Wallien MSC, Quilty LC, Steensma TD, Singh D, Lambert SL, Leroux A, et al. Cross-national replication of the gender identity interview for children. *J Pers Assess.* (2009) 91:545–52. doi: 10.1080/00223890903228463
87. Zucker KJ. Gender identity interview for children. In: Milhausen RR, Sakaluk JK, Fisher TD, Davis CM, Yarber WL, editors. *Handbook of Sexuality-Related Measures*, 4th ed. New York, NY: Routledge (2020). p. 325–8.
88. Cohen-Kettenis PT, Wallien M, Johnson LL, Owen-Anderson AFH, Bradley SJ, Zucker KJ. A parent-report gender identity questionnaire for children: a cross-national, cross-clinic comparative analysis. *Clin Child Psychol Psychiatry.* (2006) 7:433–56. doi: 10.1177/1359104506059135
89. Johnson LL, Bradley SJ, Birkenfeld-Adams AS, Kuksis MAR, Maing DM, Mitchell JN, et al. A parent-report gender identity questionnaire for children. *Archiv Sexual Behav.* (2004) 33:105–16. doi: 10.1023/B:ASEB.0000014325.68094.f3
90. Zucker KJ. Gender identity questionnaire for children. In: Milhausen RR, Sakaluk JK, Fisher TD, Davis CM, Yarber WL, editors. *Handbook of Sexuality-Related Measures*, 4th ed. New York, NY: Routledge (2020). p. 329–34.
91. Spencer D, Pasterski V, Neufeld S, Glover V, O’Connor TG, Hindmarsh PC, et al. Prenatal androgen exposure and children’s

- aggressive behavior and activity level. *Horm Behav.* (2007) 96:156–65. doi: 10.1016/j.yhbeh.2017.09.012
92. Zucker KJ, Wood H. Assessment of gender variance in children. *Child Adolesc Psychiatr Clin N Am.* (2011) 20:665–80. doi: 10.1016/j.chc.2011.07.006
 93. Sattler JM. *Assessment of Children: Cognitive Applications*, 4th ed. San Diego, CA: Jerome M. Sattler Publisher, Inc. (2001).
 94. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed. Washington, DC: American Psychiatric Association (1994).
 95. Deogracias JJ, Johnson LL, Meyer-Bahlburg HFL, Kessler SJ, Schober JM, Zucker KJ. The gender identity/gender dysphoria questionnaire for adolescents and adults. *J Sex Res.* (2007) 44:370–9. doi: 10.1080/00224490701586730
 96. Singh D, Deogracias JJ, Johnson LL, Bradley SJ, Kibblewhite SJ, Meyer-Bahlburg HFL, et al. The gender identity/gender dysphoria questionnaire for adolescents and adults: further validity evidence. *J Sex Res.* (2010) 47:49–58. doi: 10.1080/00224490902898728
 97. Zucker KJ, Meyer-Bahlburg HFL, Kessler SJ, Schober J. Gender identity/gender dysphoria questionnaire for adolescents and adults. In: Milhausen RR, Sakaluk JK, Fisher TD, Davis CM, Yarber WL, editors. *Handbook of Sexuality-Related Measures*, 4th ed. New York, NY: Routledge (2020). p. 343–50.
 98. Zucker KJ, Bradley SJ, Oliver G, Blake J, Fleming S, Hood J. Psychosexual development of women with congenital adrenal hyperplasia. *Horm Behav.* (1996) 30:300–18. doi: 10.1006/hbeh.1996.0038
 99. Schneider C, Cerwenka S, Nieder TO, Briken P, Cohen-Kettenis PT, De Cuypere G, et al. Measuring gender dysphoria: a multicenter examination and comparison of the Utrecht gender dysphoria scale and the gender identity/gender dysphoria questionnaire for adolescents and adults. *Archiv Sexual Behav.* (2016) 45:551–8. doi: 10.1007/s10508-016-0702-x
 100. Siegmann EM, Müller T, Dziadek I, Mühle C, Lenz B, Kornhuber J. Digit ratio (2D:4D) and transgender identity: new original data and a meta-analysis. *Sci Rep.* (2020) 10:19326. doi: 10.1038/s41598-020-72486-6
 101. Zucker KJ, Bradley SJ, Owen-Anderson A, Kibblewhite SJ, Wood H, Singh D, et al. Demographics, behavior problems, and psychosexual characteristics of adolescents with gender identity disorder or transvestic fetishism. *J Sex Marital Ther.* (2012) 38:151–89. doi: 10.1080/0092623X.2011.611219
 102. Zucker KJ, Bradley SJ, Owen-Anderson A, Singh D, Blanchard R, Bain J. Puberty-blocking hormonal therapy for adolescents with gender identity disorder: a descriptive clinical study. *J Gay Lesbian Mental Health.* (2011) 15:58–82. doi: 10.1080/19359705.2011.530574
 103. Storms MD. Theories of sexual orientation. *J Pers Soc Psychol.* (1980) 38:783–92. doi: 10.1037/0022-3514.38.5.783
 104. Kinsey AC, Pomeroy WB, Martin CE. *Sexual Behavior in the Human Male*. Philadelphia, PA: W. B. Saunders (1948).
 105. Langevin R. *Sexual Strands: Understanding and Treating Sexual Anomalies in Men*. Hillsdale, NJ: Erlbaum (1985).
 106. King MF, Brunner GC. Social desirability bias: a neglected aspect of validity testing. *Psychol Market.* (2000) 17:79–103. doi: 10.1002/(SICI)1520-6793(200002)17:2<79::AID-MAR2>3.0.CO;2-0
 107. Crowne DP, Marlowe D. A new scale of social desirability independent of psychopathology. *J Consult Psychol.* (1960) 24:349–54. doi: 10.1037/h0047358
 108. Strahan R, Gerbasi KC. Short, homogeneous versions of Marlowe-Crowne social desirability scale. *J Clin Psychol.* (1972) 28:191–3. doi: 10.1002/1097-4679(197204)28:2<191::AID-JCLP2270280220>3.0.CO;2-G
 109. Holden RR, Fekken GC. Three common social desirability scales: friends, acquaintances, or strangers? *J Res Pers.* (1989) 23:180–91. doi: 10.1016/0092-6566(89)90022-6
 110. Silverthorn NA, Gekoski WL. Social desirability effects on measures of adjustment to university, independence from parents, and self-efficacy. *J Clin Psychol.* (1995) 51:244–251. doi: 10.1002/1097-4679(199503)51:2<244::AID-JCLP2270510214>3.0.CO;2-Q
 111. Campbell DT, Stanley JC. *Experimental and Quasi-Experimental Designs for Research*. Chicago, IL: Rand McNally & Company (1969).
 112. Coleman E, Bockting W, Botzer M, Cohen-Kettenis P, DeCuypere G, Feldman J, et al. Standards of care for the health of transsexual, transgender, and gender-nonconforming people, version 7. *Int J Transgen.* (2011) 13:165–232. doi: 10.1080/15532739.2011.700873
 113. Arcelus J, Bouman WP, Van Den Noortgate W, Claes L, Witcomb G, Fernandez-Aranda F. Systematic review and meta-analysis of prevalence studies in transsexualism. *Eur Psychiatry.* (2015) 30:807–15. doi: 10.1016/j.eurpsy.2015.04.005
 114. Zhang Q, Goodman M, Adams N, Corneil T, Hashemid L, Kreukels B, et al. Epidemiological considerations in transgender health: a systematic review with focus on higher quality data. *Int J Transgen.* (2020) 21:125–37. doi: 10.1080/26895269.2020.1753136
 115. DeMaris A. *Logit Modeling: Practical Applications*. Newbury Park, CA: Sage (1992).
 116. Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers EJ, Berk R, et al. Redefine statistical significance. *Nat Hum Behav.* (2018) 2:6–10. doi: 10.1038/s41562-017-0189-z
 117. Steensma TD, Cohen-Kettenis PT. More than two developmental pathways in children with gender dysphoria? *J Am Acad Child Adolesc Psychiatry.* (2015) 54:147–8. doi: 10.1016/j.jaac.2014.10.016
 118. Li G, Kung KTE, Hines M. Childhood gender-typed behavior and adolescent sexual orientation: a longitudinal population-based study. *Dev Psychol.* (2017) 53:764–77. doi: 10.1037/dev0000281
 119. Steensma TD, van der Ende J, Verhulst FC, Cohen-Kettenis PT. Gender variance in childhood and sexual orientation in adulthood: a prospective study. *J Sex Med.* (2013) 10:2723–33. doi: 10.1111/j.1743-6109.2012.02701.x
 120. Xu Y, Norton S, Rahman Q. Childhood maltreatment, gender nonconformity, and adolescent sexual orientation: a prospective birth cohort study. *Child Dev.* (2020) 91:e984–94. doi: 10.1111/cdev.13317
 121. Nyberg KL, Alston JP. Analysis of public attitudes towards homosexuality. *J Homosex.* (1976) 2:99–107. doi: 10.1300/J082v02n02_01
 122. Hellman RE, Green R, Gray JL, Williams K. Childhood sexual identity, childhood religiosity, and ‘homophobia’ as influences in the development of transsexualism, homosexuality, and heterosexuality. *Archiv Gen Psychiatry.* (1981) 38:910–5. doi: 10.1001/archpsyc.1981.01780330068007
 123. Zucker KJ, Owen A, Bradley SJ, Ameeriar L. Gender-dysphoric children and adolescents: a comparative analysis of demographic characteristics and behavior problems. *Clin Child Psychol Psychiatry.* (2002) 7:398–411. doi: 10.1177/1359104502007003007
 124. Green R, Newman LE, Stoller RJ. Treatment of boyhood “transsexualism.” An interim report of four years’ experience. *Archiv Gen Psychiatry.* (1972) 26:213–7. doi: 10.1001/archpsyc.1972.01750210021003
 125. Meyer-Bahlburg HFL. Gender identity disorder in young boys: a parent- and peer-based treatment protocol. *Clin Child Psychol Psychiatry.* (2002) 7:360–76. doi: 10.1177/1359104502007003005
 126. Newman LE. Treatment for the parents of feminine boys. *Am J Psychiatry.* (1976) 133:683–7. doi: 10.1176/ajp.133.6.683
 127. Zucker KJ. Cross-gender-identified children. In: Steiner BW, editor. *Gender Dysphoria: Development, Research, Management*. New York, NY: Plenum Press (1985). p. 75–174. doi: 10.1007/978-1-4684-4784-2_4
 128. Zucker KJ. Gender identity disorder in children and adolescents. In: Gabbard GO, editor. *Treatments of Psychiatric Disorders*, 3rd ed., Vol. 2. Washington, DC: American Psychiatric Press (2001). p. 2069–94.
 129. Drescher J, Pula J. Ethical issues raised by the treatment of gender-variant prepubescent children. *Hast Center Rep.* (2014) 44:S17–22. doi: 10.1002/hast.365
 130. Giordano S. *Children With Gender Identity Disorder: A Clinical, Ethical, and Legal Analysis*. New York, NY: Taylor & Francis Group (2013).
 131. Pleak RR. Ethical issues in diagnosing and treating gender-dysphoric children and adolescents. In: Rottnek M, editor. *Sissies and Tomboys: Gender Nonconformity and Homosexual Childhood*. New York, NY: New York University Press (1999). p. 34–51.
 132. Stein E. Commentary on the treatment of gender variant and gender dysphoric children and adolescents: common themes and ethical reflections. *J Homosex.* (2012) 59:480–500. doi: 10.1080/00918369.2012.653316
 133. Zucker KJ. Treatment of gender identity disorders in children. In: Blanchard R, Steiner BW, editors. *Clinical Management of Gender Identity Disorders in Children and Adults*. Washington, DC: American Psychiatric Press (1990). p. 25–45.

134. Rekers GA, Lovaas OL. Behavioral treatment of deviant sex role behaviors in a male child. *J Appl Behav Anal.* (1974) 7:173–90. doi: 10.1901/jaba.1974.7-173
135. Rekers GA. *Shaping Your Child's Sexual Identity.* Grand Rapids, MI: Baker Book House (1982).
136. Zucker KJ, Wood H, Singh D, Bradley SJ. A developmental, biopsychosocial model for the treatment of children with gender identity disorder. *J Homosex.* (2012) 59:369–97. doi: 10.1080/00918369.2012.653309
137. American Academy of Child and Adolescent Psychiatry. Practice parameter on gay, lesbian, or bisexual sexual orientation, gender nonconformity, and gender discordance in children and adolescents. *J Am Acad Child Adolesc Psychiatry.* (2012) 51:957–74. doi: 10.1016/j.jaac.2012.07.004
138. American Psychological Association. Guidelines for psychological practice with transgender and gender nonconforming people. *Am Psychol.* (2015) 70:832–864. doi: 10.1037/a0039906
139. Byne W, Bradley SJ, Coleman E, Eyler AE, Green R, Menvielle EJ, et al. Report of the American Psychiatric Association Task Force on treatment of gender identity disorder. *Archiv Sexual Behav.* (2012) 41:759–96. doi: 10.1007/s10508-012-9975-x
140. Zucker KJ. Gender identity disorder. In: Weiner IB, editor. *Adult Psychopathology Case Studies.* New York, NY: Wiley (2004). p. 207–28.
141. Zucker KJ. “I’m half-boy, half-girl”: play psychotherapy and parent counseling for gender identity disorder. In: Spitzer RL, First MB, Williams JBW, Gibbons M, editors. *DSM-IV-TR® Casebook, Volume 2. Experts Tell How They Treated Their Own Patients.* Washington, DC: American Psychiatric Publishing (2006). p. 321–34.
142. Cohen-Kettenis PT, van Goozen SHM. Pubertal delay as an aid in diagnosis and treatment of a transsexual adolescent. *Eur Child Adolesc Psychiatry.* (1998) 7:246–8. doi: 10.1007/s007870050073
143. Gooren L, Delemarre-van de Waal H. The feasibility of endocrine interventions in juvenile transsexuals. *J Psychol Hum Sex.* (1996) 8:69–84. doi: 10.1300/J056v08n04_05
144. Ehrensaft D. Found in transition: our littlest transgender people. *Contemp Psychoanal.* (2014) 50:571–92. doi: 10.1080/00107530.2014.942591
145. Ehrensaft D. Treatment paradigms for prepubertal children. In: Forcier M, Van Schalkwyk G, Turban JL, editors. *Pediatric Gender Identity: Gender-Affirming Care for Transgender & Gender Diverse Youth.* New York, NY: Springer (2020). p. 171–85. doi: 10.1007/978-3-030-38909-3_13
146. Wong W, Chang SCH. Social transitioning for gender dysphoric children. In: Lev AI, Gottlieb AR, editors. *Families in Transition: Parenting Gender Diverse Children, Adolescents, and Young Adults.* New York, NY: Harrington Park Press (2019). p. 356–73.
147. Wong WI, van der Miesen AIR, Li TGF, MacMullin LN, VanderLaan DP. Childhood social gender transition and psychosocial well-being: a comparison to cisgender gender-variant children. *Clin Pract Pediatr Psychol.* (2019) 7:241–53. doi: 10.1037/cpp0000295
148. Chen D, Edwards-Leeper L, Stancin T, Tishelman A. Advancing the practice of pediatric psychology with transgender youth: state of the science, ongoing controversies, and future directions. *Clin Pract Pediatr Psychol.* (2018) 6:73–83. doi: 10.1037/cpp0000229
149. Dreger A. Gender identity disorder in childhood: inconclusive advice to parents. *Hast Center Rep.* (2009) 39:26–9. doi: 10.1353/hcr.0.0102
150. Green R. To transition or not to transition? That is the question. *Curr Sexual Health Rep.* (2017) 9:79–83. doi: 10.1007/s11930-017-0106-5
151. Olson KR, Blotner C, Alonso D, Lewis K, Edwards D, Durwood L. Family discussions of early childhood social transitions. *Clin Pract Pediatr Psychol.* (2019) 7:22–40. doi: 10.1037/cpp0000289
152. Zucker KJ. Different strokes for different folks. *Child Adolesc Mental Health.* (2020) 25:36–7. doi: 10.1111/camh.12330
153. Rae JR, Gulgoz S, Durwood L, DeMeules M, Lowe R, Lindquist G, et al. Predicting early-childhood gender transitions. *Psychol Sci.* (2019) 30:669–81. doi: 10.1177/0956797619830649
154. Tollit MA, Pace CC, Telfer M, Hoq M, Bryson J, Fulkoski N, et al. What are the health outcomes of trans and gender diverse young people in Australia? Study protocol for the Trans20 longitudinal cohort study. *BMJ Open.* (2019) 9:e032151. doi: 10.1136/bmjopen-2019-032151

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer RB declared a past co-authorship with one of the authors KZ to the handling Editor.

Copyright © 2021 Singh, Bradley and Zucker. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.