IN PRESS: *PSYCHOLOGICAL SCIENCE*

Predicting Early Childhood Gender Transitions

James R. Rae[1,2]

Selin Gülgöz[3]

Lily Durwood[3]

Madeleine DeMeules[3]

Riley Lowe[3]

Gabrielle Lindquist[3]

Kristina R. Olson[3†]

[1] University of Massachusetts Amherst

[2] University of Oxford

[3] University of Washington

[†] Please address correspondence regarding this research to Kristina Olson, krolson@uw.edu.

Acknowledgments

The authors thank Paul Bloom, Susan Gelman, and Michael Norton for comments on an earlier draft of this manuscript, as well as Fabian Dablander, John Kruschke, Amy Orben, Jessica Pearlman, and Nils Reimer for helpful insights on the statistical analyses reported in this manuscript. Data collection and author time was supported by NSF Grants #1523632, #1837857, #1715068, NIH Grant # HD092347 and the Arcus Foundation.

Abstract

Increasing numbers of gender nonconforming children are socially-transitioning—changing pronouns to live as their identified genders. We studied a cohort of gender nonconforming children (*N*=85). When re-contacted approximately 2 years later, 36 of the children had socially-transitioned. We found that stronger cross-sex identification and preferences expressed by gender nonconforming children at initial testing predicted whether they later socially-transitioned. We then compared the gender nonconforming children to groups of transitioned transgender children (*N*=84) and gender conforming controls (*N*=85). Children from our longitudinal cohort who would later transition were highly similar to transgender children and control children of the gender to which they would eventually transition, while gender nonconforming children who would not go on to transition were different from these groups. These results suggest that (a) social transitions may be predictable from gender identification and preferences and (b) gender identification and preferences may not meaningfully differ before and after social transitions.

Keywords: transgender, gender nonconformity, social transitions, gender development

Predicting Early Childhood Gender Transitions

In any given classroom, one will likely find many children who disregard some gender norms, such as boys who like pink and girls who engage in rough-and-tumble play (e.g., Sandberg, Meyer-Bahlburg, Ehrhardt, & Yager, 1993). Less common are children who consistently show a preference for opposite-sex[1] peers, prefer toys and clothing that are culturally-associated with the opposite sex, or express a wish to be members of the opposite gender group—children whom we term "gender nonconforming." Parents, scientists, and clinicians have often wondered about these children's later outcomes. Longitudinal data suggest that most gender nonconforming children do *not* end up identifying as transgender[2] (i.e., identifying as a gender opposite their assigned sex[3]) later in life, though in every study at least some do (Green, 1987; Steensma, McGuire, Kreukels, Beekman, & Cohen-Kettenis, 2013; Wallien & Cohen-Kettenis, 2008; Zucker & Bradley, 1995).

Interest in early childhood gender nonconformity and later transgender identity has recently become especially pronounced (Dreger, 2009; Green, 2017) as some families are supporting their prepubescent children through social transitions (Edwards-Leeper, Leibowitz, &

---

[1] Our use of the term "opposite" implies that boy and girl or male and female contrast one another. We use this term for ease of comprehension and linguistic simplicity. We agree with the many scholars who point out that sex and gender are nonbinary and are likely better conceptualized continuously (e.g., Bem, 1974; Ehrensaft, 2010).

[2] Many of these studies used "gender dysphoria" as an outcome. Gender dysphoria is a medical term for experiencing distress related to one's assigned sex and a desire to be a member of the other gender group.

[3] We use the term "assigned sex" to refer to the categorization made at birth based on their external genitalia and/or chromosomes in line with the recommendations of the World Professional Association for Transgender Health (Bouman et al., 2017), while we use "gender" or "gender identity" to refer to a person's self-categorization. We use the term "gender nonconformity" rather than "sex nonconformity" because colloquially and in past research this term is used to refer to behaviors/identities not typically associated with one's assigned sex.

Sangganjanavanich, 2016; Sherer, 2016; Turban, 2017). Social transitions, as they apply to young pre-pubescent gender nonconforming children, involve changing a child's pronouns, hairstyle, clothing, and sometimes name to align with the child's identity, rather than his/her assigned sex (Malpas, 2011; Steensma et al., 2013). Relatively unheard of ten years ago, early childhood social transitions are contentious within the clinical, scientific, and broader public communities (Edwards-Leeper et al., 2016; Green, 2017; Steensma & Cohen-Kettenis, 2011). Despite their increasing occurrence, we know little about who does and does not transition, the predictors of social transitions, and whether transitions impact children's views of their own gender. These are the central questions of the current paper.

Past longitudinal work has found that those gender nonconforming children who identify as transgender in adolescence and adulthood tended to show more extreme childhood gender nonconformity than gender nonconforming children who did not later identify as transgender (Singh, 2012; Steensma et al, 2013; Wallien & Cohen-Kettenis, 2008). Most or all participants in these studies had not completed a full social transition as prepubescent children (i.e., had not changed pronouns), as social transitions are a newer practice. Therefore, it is an open question whether those who socially transition in early childhood also show systematic differences in gender nonconformity from those who do not.

On one hand, if the findings from post-pubertal transitions apply to those undertaken earlier, children showing more extreme gender nonconforming identities and preferences should be more likely to socially transition in childhood. Supporters of social transitioning as a practice often argue that social transitions should be considered for children exhibiting particularly strong and consistent cross-gender identification for extended periods of time (e.g., assigned males who identify most strongly as girls, Malpas, 2011; Ehrensaft, 2017). These supporters suggest that a child's degree of gender identification (as perceived by parents, therapists, etc.) may contribute to their family's and/or clinician's support for their decision to transition. On the other hand, there is no standardized protocol or measure to help families or clinicians decide which children to support through a transition, or any battery of tasks to describe how extreme a given child's cross-gender behavior is relative to other gender nonconforming children. Thus, social transitions may be occurring randomly or without regard to variation in children's identities or preferences.

To test these possibilities, we recruited a group of gender nonconforming children who had *not* socially transitioned and assessed their gender identification and preferences. An average of two years later, we asked their parents whether each child had socially transitioned. This approach allowed us to prospectively investigate whether children who went on to socially transition (hereafter, *future transitioners*) differed from those that did not socially transition (*non-transitioners*) in terms of earlier extremity in cross-sex gender identification and preferences. Importantly, parents were not told how their child's results compared to other children in the study, nor were they given an individual report of their children's results.

As a second research question, we also investigated whether social transitions are associated with changes in the degree to which children express their gender identity and preferences. That is, would an assigned male who is living as a girl (i.e., a transgender girl) be more feminine than an assigned male, who has not transitioned, but later does? On one hand, after a transition, a child is more likely to be treated as a member of his/her identified gender in everyday interactions because the child now appears (i.e., through clothing, pronouns, etc.) to be a member of that gender group. This treatment may reinforce the child's sense of identity, thereby leading to more extreme preferences and identity expression. In this case, a child tested before transitioning might not show as extreme preferences and identity expression as a child who has already transitioned. On the other hand, perhaps the child's gender identification and preferences were already very strong, before the transition. In this case, children tested before transitioning may not differ from children tested after transitioning. To assess this question, we recruited a comparison group of transgender children (i.e., those who had already socially transitioned), had them complete the same assessments of gender identity and preferences, and compared the two groups.

Finally, we asked whether children tested before transitioning (future transitioners) and children tested after transitioning (transgender children) differ in terms of their gender identity and preferences from control children assigned the opposite sex at birth. Past work has suggested that after transitions, transgender children show comparable gender identity and preferences to peers with the opposite assigned sex (Fast & Olson, 2018; Olson, Key, & Eaton 2015), so the inclusion of a comparison group of control children provided an opportunity for replication (for comparison to transgender children) and possible extension (to future transitioners).

Method

Participants

Recruitment. Gender nonconforming and transgender children were recruited through a wide range of community groups. Controls were recruited through a university database of families interested in participating in research. All children completed measures individually with an experimenter, beyond earshot of their parents. See Table 1 for demographics of each participant group and the Supplemental Material for more details on recruitment and testing sessions.[4]

Gender nonconforming children (future transitioners and non-transitioners). Every gender nonconforming child who participated in research between the start of the project in July 2013 and December 2016 is included in this research except one child who did not complete any of the measures reported in this paper. Thus, our sample size was determined by the number of participants we could recruit in this 3.5-year period, rather than a target sample size. As gender nonconforming children are rare and hard-to-reach and were not the primary participants recruited during this time by the research team (they signed up in the course of recruitment for a study of transgender children), estimating a sample size in advance was impossible. A social transition was defined as having changed one's pronouns to align with the child's identified gender (i.e., an assigned male going by "she"; Fast & Olson, 2018; what is called a "complete transition" by Steensma et al., 2013). By the time children change their pronouns, or at the same time, they typically change their first name (if their original first name was gendered), hairstyle, and clothing. To assess later transition status, we contacted parents for confirmation, or we received an update via an in-person follow-up visit or parent online survey. If there had been multiple contacts after the initial data collection, the most recent contact before the paper submission was used to determine whether the child had socially transitioned or not (all gender nonconforming children who met the criteria for a social transition at one point continued to do so for all subsequent points). The average time from original testing to follow-up was 25 months

[4] Data from 26 transgender children included in the present work were also included in past published work (19 from Fast & Olson, 2018; 7 from Olson et al, 2015). None of the current controls or gender nonconforming children were reported in past work.

(*SD* = 10 months). Of the original sample, 36 had transitioned (i.e., changed pronouns to those opposite their assigned sex) and 49 had not.

Table 1. Demographic characteristics for all participants.

Variable	Gender Nonconforming		Transgender		Control	
	Future Transitioners (*n* = 36)	Non-Transitioners (*n* = 49)	Matched to Future Transitioners (*n* = 35)	Matched to Non-Transitioners (*n* = 49)	Matched to Future Transitioners (*n* = 36)	Matched to Non-Transitioners (*n* = 49)
Assigned Sex (% male)	83%	61%	86%	61%	17%	39%
Age (*M* months, *SD*)	83.4 (27.6)	95.1 (30.2)	84.4 (27.1)	94.2 (31.0)	84.6 (27.4)	95.5 (30.0)
Race (% White)	75%	65%	49%	63%	67%	73%
Time Between Testing and Transition Check-In (*M* months, *SD*)	28.8 (8.8)	24.2 (10.0)	–	–	–	–
Parent Political Orientation (*M* on 1-7 scale, *SD*)	6.1 (1.0)	6.4 (0.8)	6.7 (0.5)	6.3 (0.9)	5.8 (1.3)	5.6 (1.3)
Parent Income (*M* on 1-5 scale, *SD*)	3.7 (1.1)	3.7 (1.2)	4.2 (0.8)	4.2 (0.9)	4.4 (1.0)	4.2 (1.1)

Note: We only followed up with gender nonconforming families; thus, time since initial test is not reported for transgender and control participants. Three parents did not report their political orientation and one parent did not report their household income.

Transgender children. A transgender comparison group (i.e., a group of children who had socially transitioned *before* completing the battery) was recruited from an ongoing longitudinal study of transgender youth. For each gender nonconforming participant, a transgender child who had the same assigned sex and was within four months of age at time of testing was included in the transgender comparison group. Matching was completed using a master file that included the child's assigned sex and age on day of first testing but lacked any responses from the child to ensure that responses could not inform participant selection. Matches were available for all children except one (no one in the database met the matching criteria for one child).

Control children. Gender nonconforming children were also matched to control participants of the same age (within four months) but with the opposite assigned sex. This matching approach has been utilized in related past work with transgender children (e.g., Fast & Olson, 2018; Olson et al., 2015).

Measures and Data Preparation

Gender identity and preferences. The present analyses focused on a composite of five gender development measures, which were selected because they are the general battery of measures given to all children in this line of work in our research group. The contributing measures were:

Peer preference. Peer preferences were assessed on six trials in which children were presented with the pictures of a boy and a girl and were asked whom they would prefer to be friends with (from Olson et al., 2015). The proportion of trials on which they selected girls was recorded.

Toy preference and outfit preference measures. Toy and outfit preferences were each assessed via four trials (from Fast & Olson, 2018). On each trial, children were presented with five images of toys or outfits at a time. Pictures were pilot-tested with a separate set of children to represent very feminine, slightly feminine, gender-neutral, slightly masculine, or very masculine items. Responses were coded on a 5-point scale with higher scores indicating more feminine responses. Within each measure, responses from the four trials were averaged and then re-scaled to range from 0 to 1 (Cohen, Cohen, Aiken, & West, 1999). Different toys and clothes were used for 5-7 vs. 8-11 year olds as children generally play with different toys and wear different clothes at these ages.

Similarity measure. Children indicated how similar they felt to boys and girls on five items using a visual 5-point scale (Martin, Andrews, England, Zosuls, & Ruble, 2017). Following Fast and Olson (2018), a difference score was computed by subtracting the average of the five boy items from the average of the five girl items. The scores were re-scaled to range from 0 (most similar to boys and most dissimilar to girls) to 1 (most similar to girls and most dissimilar to boys).

Identity measure. Children were asked whether they (1) are currently and (2) will in the future be boys, girls, both, neither, it changes over time, or they are not sure (Fast & Olson, 2018). Each “girl” response was assigned +1 point, each “boy” response was assigned −1 point, and all other answers were given a score of 0. The two items were added together and again, re-scaled to a 0 to 1 scale.

As all five measures were scaled between 0 and 1, we created a gender identity and preferences composite score by taking the average. The gender identity and preferences composite variable demonstrated acceptable reliability ($\alpha = .74$). Masculine children’s (assigned female gender nonconforming and transgender children, and assigned male controls) scores were then reverse-scored so that higher numbers indicated more extreme cross-sex responding for gender nonconforming and transgender children and more extreme same-sex responding for controls, as has been done in related work (e.g., Fast & Olson, 2018).

Demographics. We collected several demographic variables (see Table 1). We recorded participants' age (months), assigned sex (female=0; male=1), race (0=non-White; 1= White), and time between the initial testing session and follow-up (months). We further recorded information about their family, including participating parents' political orientation (1= least liberal; 7= most liberal) and household income (1=lowest income; 7=highest income). Participating children tended to be White, and their parents tended to be high-income and politically liberal.

Missing Data

Missing data on items ranged from 0% to 15.7%, which we addressed via multiple imputation by chained equations (MICE; White, Royston, & Wood, 2011). Using the mice package (Van Buuren & Groothuis-Oudshoorn, 2011) in R (R Core Team, 2016), we used predictive mean matching (i.e., imputed values are draws from observed values; Vink, Frank, Pannekoek, & Buuren, 2014) to generate 20 "complete" datasets - a rule of thumb for the degree of missingness in our dataset (White et al., 2011). We then created the gender identity and preferences composite (see above). Statistical analyses were conducted on each imputed dataset and estimates were obtained by pooling results across these analyses. R code for all analyses are available on the Open Science Framework (OSF): https://osf.io/m6zac/.

Analytic Strategy

Consistent with "the New Statistics" (Cumming, 2014), we employed Bayesian estimation for our statistical analyses (Kruschke & Liddell, 2018). Bayesian methods offer numerous advantages over frequentist methods. Indeed, as Bayesian statistics do not rely on large sample sizes in the same way as frequentist methods, they may be better equipped to model data with small samples like those used here (Depaoli & van de Schoot, 2017). As an overview of the Bayesian estimation framework (for detailed coverage of Bayesian methods, see Gelman et al., 2014; Kruschke, 2015; McElreath, 2016), before observing the data, uncertainty about the value of each model parameter is encoded in a probability distribution of possible values - a *prior distribution*. After obtaining the data, the prior distribution is combined with the *likelihood* (the probability of the observed data given the parameter values), yielding a *posterior distribution* that reflects the updated uncertainty about the value of each parameter. In practice, the posterior distribution is typically approximated (rather than obtained analytically) using simulation methods, such as Markov Chain Monte Carlo (MCMC). We used medians and 95% highest density intervals (HDIs) to summarize the central tendency and uncertainty of the posterior distribution, respectively. Unlike frequentist confidence intervals (Hoekstra, Morey, Rouder, & Wagenmakers), HDIs have a natural interpretation; if 95% of the most credible values for a parameter are between .50 and .75 (i.e., the 95% HDI = [.50–.75]), we are 95% sure that the population value lies between .50 and .75 (Kruschke & Liddell, 2017). Critically, HDIs can also quantify evidence for null values, such that a parameter is zero for practical purposes if only values deemed to be functionally equivalent to zero (i.e., in the *region of practical equivalence*; ROPE) are contained within an HDI (Kruschke, 2015).

We assessed whether gender nonconformity predicted social transition status by fitting a logistic regression model in which transition status (no = 0; yes =1) was regressed on gender identity and preference scores. We estimated the model using the brms package (Bürkner, 2017) as a “front end” to the probabilistic programming language Stan (Carpenter et al., 2017). Details of our analyses (i.e., the priors we used, a sensitivity analysis comparing our results using more/less informative priors, our results using a model comparison approach, etc.) are presented in the Supplemental Material. Next, we re-estimated our model after controlling for covariates (recommended by Simmons, Nelson, & Simonsohn, 2011). To facilitate comparisons of coefficients for binary and continuous variables, regression inputs were scaled by 2 standard deviations (Gelman, 2008). Finally, coefficients were transformed into odds ratios; values greater than 1 provide evidence of a positive association between the predictor and socially transitioning (and vice versa for values less than 1).

We used a different approach to ask whether future transitioners differed in their gender identity and preferences from transgender and control participants. As these scores were bounded (i.e., between 0 and 1), we applied a recommended transformation that prevents 0's or 1's (Smithson & Verkuilen, 2006) and then fit a beta regression (Ferrari & Cribari-Neto, 2004) with a logit link function to estimate gender identity and preferences scores of future transitioners and matched transgender and control participants. Bayesian multilevel modeling alleviates concerns of multiple comparisons (Gelman, Hill, & Yajima, 2012), such as the between-group comparisons examined here. Accordingly, we estimated a multilevel model that included a unique intercept for each group of participants (i.e., varying intercept model). We then calculated posterior differences in the parameter estimates across groups (Kruschke, 2015). Finally, we re-estimated our initial model after including covariates to obtain covariate-adjusted mean differences. MCMC samples of the posteriors are available on the OSF: https://osf.io/m6zac/.[5]

[5] We cannot share the raw data due to issues of identifiability in this rare sample. Figure 1 shows the data at the individual level in a way that has been approved by our IRB. As recommended by Kruschke (2015), MCMC samples of the posteriors are available, which allows interested readers to (1) explore posterior comparisons not reported in this manuscript and (2) use our results as a prior for future analyses that use a similar design and model.

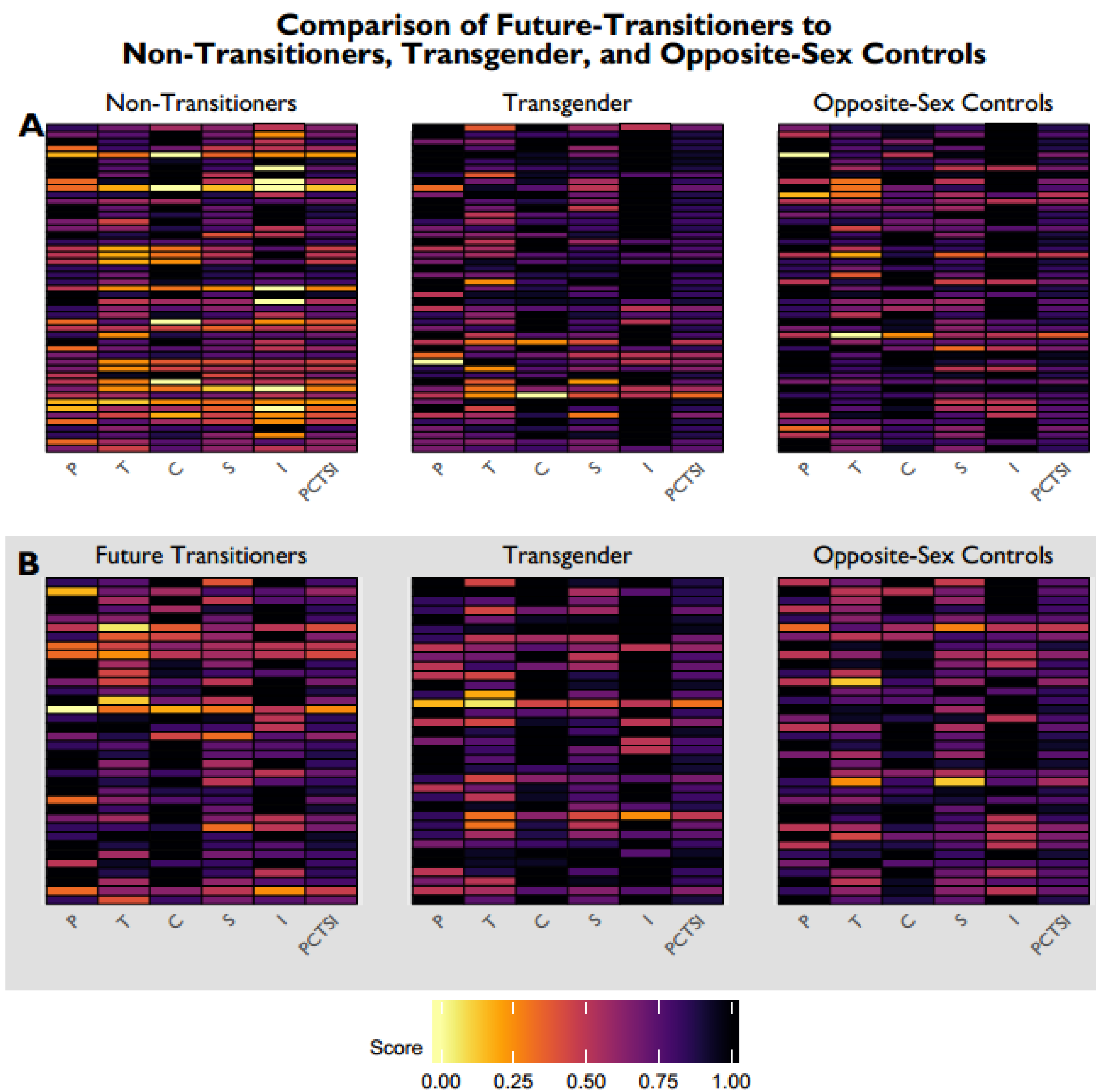
Results

Figure 1 shows how individual participants responded to the gender identity and preferences measures (missing values were replaced with the average score across imputed datasets). Figure 1a shows responses for future transitioners with their matched transgender and control groups, whereas Figure 1b shows responses for non-transitioners with their matched transgender and control group. The Supplemental Material contains (a) group means for the scores presented in Figure 1; (b) descriptive statistics; and (c) zero-order correlations among measures.

Analysis 1: Do Gender Identity and Preferences Predict Social Transitions?

Pooled results from logistic regression models containing no covariates suggested that participants expressing greater gender nonconformity in the initial testing session were more likely to socially transition before follow-up, odds ratio=4.22, 95% HDI = [1.55–12.20]. That is, assigned males that tended to have more extremely feminine preferences and gender identities were more likely to socially transition to live as girls after testing than assigned males who exhibited less extremely feminine identities and preferences. Our model predicted that a child with a gender nonconformity score of .50 would roughly have a .30 probability (95% HDI = [0.17–0.42]) of socially transitioning. By contrast, a child with a gender nonconformity score of .75 would roughly have a .48 probability (95% HDI = [0.37–0.60]) of transitioning.

Figure 1. Scores on the gender identity and preferences measures and the gender composite score by participant. The five measures (peer preferences [P], toy preferences [T], clothing preferences [C], gender similarity [S], gender identity [I]) and the composite (average of all 5 measures) are each represented as a column within each cluster of data. Each row within a cluster represents one child's responses on all measures (when data were missing we used the mean score across 20 imputed data sets). The darker the color (i.e., the Score) the more the gender nonconforming (GNC) or transgender child's answer was stereotypically associated with the opposite assigned sex. For control participants, the darker the color the more the child's answer was stereotypically associated with their assigned sex. Panel A (white background) shows non-transitioners with their matched transgender and control groups; Panel B (gray background) shows future transitioners with their matched transgender and control groups.



We next introduced covariates to our model to examine if they accounted for the association between gender nonconformity and socially transitioning. One possible explanation for the association between social transition status and gender nonconformity is that future transitioners may have more politically liberal parents who could be more likely to support or encourage social transitions. However, we found that the coefficient for parent political orientation was not credibly different from 1.0, odds ratio=0.93, 95% HDI = [0.33–2.67]. That is, because 1.0 was among the most plausible values for the parent political orientation coefficient, we have little confidence that there is a meaningful association between parent political orientation and transition status in this sample. Another possible explanation is that future transitioners could be further along a path toward transitioning than non-transitioners when initial testing occurred, in which case future transitioners may be older or have had a longer period between the initial testing session and follow-up. While the coefficient for age was not credibly different from 1.0, odds ratio=0.43, 95% HDI =[.15–1.24], months between the initial testing session and follow-up was associated with higher levels of socially transitioning, odds ratio=3.51, 95% HDI = [1.14–11.01].[6] Critically, more extreme gender nonconformity continued to predict whether a child socially transitioned even after these and other covariates were added to the model, odds ratio=5.20, 95% HDI = [1.60–17.11], suggesting that differences on the covariates we measured did not explain the association between gender nonconformity and socially transitioning.

[6] Assigned males were more likely to transition than assigned females, odds ratio= 4.26, 95% HDI = [1.34–14.13]. All other covariates had 95% HDIs that contained zero. See Supplemental Material for full results.

Analysis 2: Do Future Transitioners, Transgender Children, and Controls Differ in their Gender Identity and Preferences?

Pooled results from our multilevel beta regression models yielded median gender identity and preferences estimates of .74 for future transitioners (95% HDI = [0.69–0.78]), .77 for transgender children (95% HDI = [0.73–0.81]), and .76 for controls (95% HDI = [0.72–0.81]). As shown in Figure 2a, the median differences between groups (represented as the effect size for proportions Cohen's *h*) were ≤ |.07|. To examine whether group differences were functionally equivalent, we identified a ROPE value around zero of ± .20 (a "small" effect; Cohen, 1988). We estimated both the 95% HDI of the effect size for the difference between groups and the proportion of the posterior inside the ROPE values.[7] The 95% HDI for the difference between (a) control and transgender participants and (b) control and future transitioners fell entirely inside (or bordered the upper bound) the ROPE (see Figure 2a). Thus, we are 95% sure that the differences between these groups are smaller than "small". Similarly, at least 97% of the posterior density for these differences fell inside the ROPE values. For the difference between transgender participants and future transitioners, the 95% HDI was not completely contained within the ROPE values (see Figure 2a). Nonetheless, over 95% of the posterior density for the difference between future transitioners and transgender participants fell inside the ROPE values, which provides some evidence that most of the plausible values for the difference are smaller than a small effect.

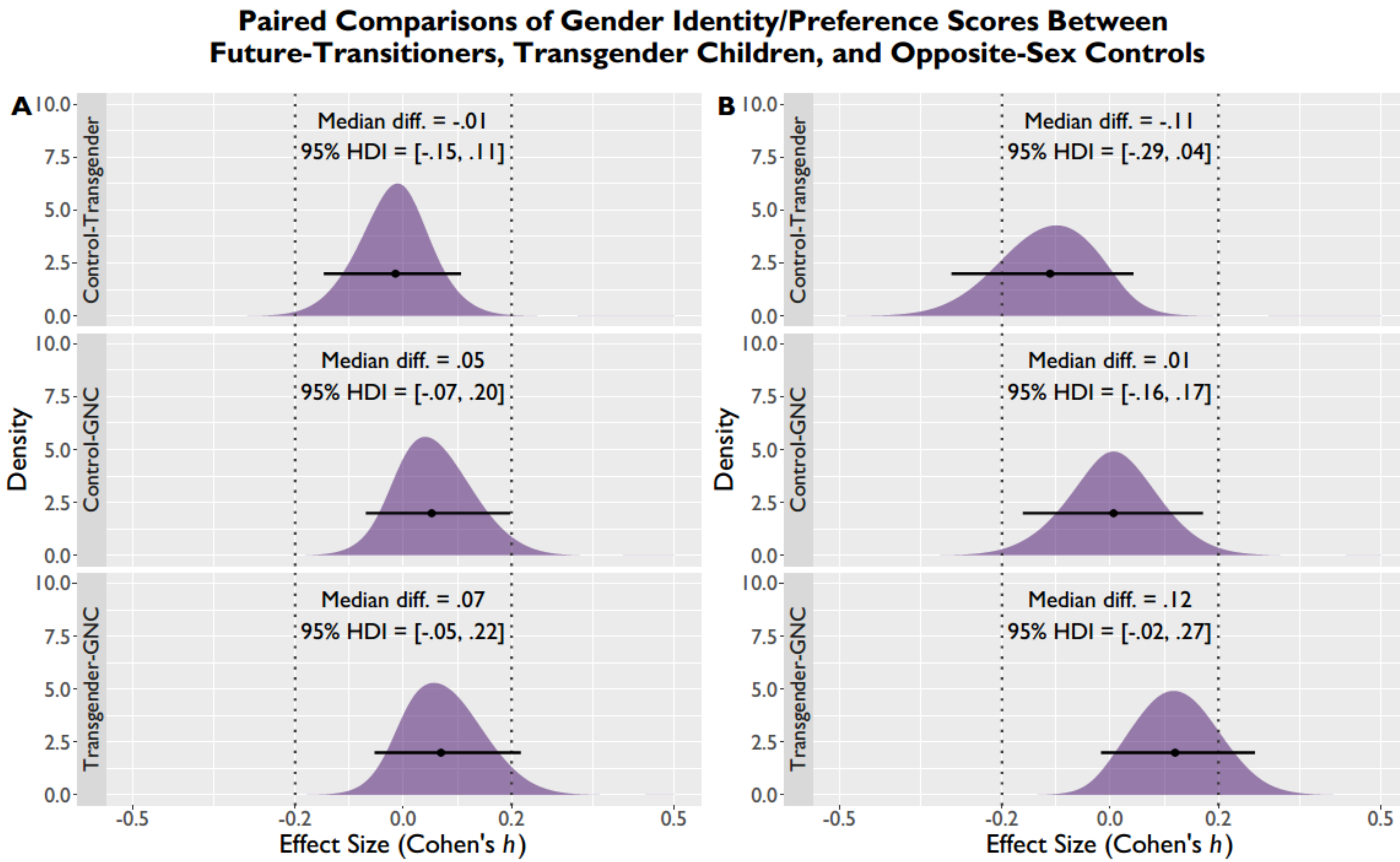
The estimated covariates-adjusted mean was .75 for future transitioners (95% HDI = [0.70–0.79]), .80 for transgender participants (95% HDI = [0.75–0.84]), and .75 for controls

[7] Unlike HDIs, ROPE values were constrained to be symmetrical around zero. Thus, the proportion of the posterior in the ROPE may be different than the HDI – especially for non-symmetrical posteriors.

(95% HDI = [0.70–0.80]). The median differences between groups were ≤ |.12| (see Figure 2b). The 95% HDI (and over 97% of the posterior) for the difference between control and future transitioners fell inside the ROPE value, whereas the 95% HDIs for the difference between transgender participants and both control participants and future transitioners did not fall completely inside the ROPE (see Figure 2b). So, how confident are we that the differences between groups were functionally equivalent to zero? For the difference between control and transgender participants, we found that 83% of the posterior distribution was in the ROPE. For the difference between transgender participants and future transitioners, 85% of the posterior distribution fell within the ROPE cutoffs. Thus, we are at least 83% sure that the covariate-adjusted differences between future transitioners and both control and transgender participants are smaller than a “small” effect (the Supplemental Material displays the proportion of the posterior inside the ROPE when using smaller/larger effect sizes to define ROPE cutoffs).

Discounting the possibility that covariate matching would equate gender identity and preference scores for *non-transitioners* and matched control and transgender participants, the Supplemental Material presents evidence that non-transitioners had lower scores than their matched control and transgender peers. Thus, our analytic approach itself did not make all differences null, rather, future transitioners look quite similar to their comparison groups while the non-transitioners look substantially different.

Figure 2. Posterior distributions along with 95% HDIs for the differences (represented as the effect size for proportions Cohen's *h*) in gender identity and preference scores between controls, transgender children, and future transitioners (labeled GNC) without covariates (panel a) and after controlling for covariates (panel b). The vertical dotted lines correspond to a small negative effect size (Cohen's *h*= −.20) and small positive effect size (Cohen's *h* = .20), respectively. The area between these lines is the ROPE (a region for which values are practically equivalent to zero) and evidence that the groups are not different comes from (1) 95% HDIs that fall completely inside the ROPE and/or (2) a large portion of the posterior distribution falling inside the ROPE.



Multiverse Analyses

A multiverse analysis repeats a statistical analysis across all plausible combinations of data processing decisions (e.g., how data are selected, cleaned, or coded) to quantify the extent to which these decisions influence a result (Steegen, Tuerlinckx, Gelman, & Vanpaemel, 2016). We conducted a multiverse analyses to explore how our results differed across three data processing decisions that were not explored in the analyses above. First, the analyses above used a gender identity and preference composite created by averaging five gender development measures: peer preferences (P), clothing preferences (C), toy preferences (T), gender similarity (S), and gender identity (I). However, there are 31 variables that we could constructed from assessing the impact of each measure by itself or different combinations of these 5 measures (e.g., using different combinations of four of the five measures, three of the five measures, etc.). Second, while we chose to address missing data via multiple imputation, we could have instead ignored missingness and used only the data provided by participants (e.g., if a child completed five of the six peer preference items, instead of imputing a value for the sixth item, we could have just averaged the five items they completed). Third, while we elected to retain all respondents in our analyses, an alternative strategy would have been to exclude observations deemed to be influential (by comparing the full-data predictive distribution and the predictive distributions obtained when each observation is left out of the analysis; see Gabry, Simpson, Vehtari, Betancourt, & Gelman, 2017). Combining these data analytic decisions should have resulted in 31 (all possible combinations of the five gender developmental measures) × 2 (missing data approach: ignore missingness vs. multiple imputation) ×2 (influential observations: ignore vs. exclude influential cases) = 124 data sets to analyze. However, we did not identify any cases that were highly influential (see Supplemental Material for details) leaving us 31 × 2 = 62 data sets for each analysis. For each data set, we fit statistical models from Analysis 1 and Analysis 2 that excluded covariates (see above).

Multiverse Analysis 1: Do Gender Identity and Preferences Predict Social Transitions? Figure 3 presents the estimates (as odds ratios) and 95% HDIs from our multiverse analysis testing the association between social transition status and gender nonconformity. Figure 3 provides four take-aways. First, most of the possible datasets we could have analyzed would have supported the conclusion that greater gender nonconformity predicts later social transition status. Indeed, 50 of the 62 analyses (81%) yielded 95% HDIs that entirely exceeded 1.0. Second, even in data sets where we are less confident that there is a positive association between social transition status and gender nonconformity (i.e., the 95% HDI included 1.0), it was always true that the majority of every 95% HDI was greater than 1.0 (i.e., the most plausible values of the slope were positive). Third, more 95% HDIs exceeded 1.0 as additional gender development measures were included in the composite variable. Indeed, only two of five gender development measures (clothing preferences and gender identity) predicted social transition status in isolation. In contrast, in all cases but one (i.e., 31 of 32), composite variables consisting of three or more gender development measures always predicted social transition status. Lastly, our results were extremely consistent across both missing data approaches.

Multiverse Analysis 2: Do Future Transitioners, Transgender Children, and Controls Differ in their Gender Identity and Preferences? Figure 4 presents the median differences between groups along with 95% HDIs of the differences (represented as the effect sizes) from our multiverse analysis comparing future transitioners to matched control and transgender participants. Across the 186 comparisons (62 data sets × 3 between-group comparisons per data set), 130 of the comparisons fell completely inside the ROPE cutoffs

(70%). However, the proportion of comparisons in the ROPE varied as a function of whether the comparison was between future transitioners and control participants (48/62=77%), control and transgender participants (52/62=84%), or future transitioners and transgender participants (30/62=48%). While there was variability in terms of which comparisons strictly fell inside the ROPE cutoffs, the more striking conclusion from Figure 4 is that “small” or “smaller than small” differences between groups were almost always the most credible (especially for composite variables containing four or more gender development measures). Indeed, many of the between-group comparisons narrowly exceeded the ROPE boundaries, which was apparent in our examination of the proportion of the posterior distribution of the ROPE for each of the 186 comparisons shown in Figure 4. We found that (with few exceptions) most of the posterior distribution (often near 100%) was inside the ROPE for all between group-comparisons. Indeed, 162 of the 186 comparisons (87%) had more than 95% of the posterior distribution inside the ROPE (see the Supplemental Material for a figure displaying these results). Finally, we found that missing data approach had little impact on our results (see Figure 4).

In stark contrast to the results presented in Figure 4, a multiverse analysis comparing *non-transitioners* and covariate-matched control and transgender participants showed consistent evidence for differences between groups, demonstrating again that the major conclusions were not tied to a particular analytic decision or two (see Supplemental Material).

Figure 3. Multiverse analysis predicting social transition status across all gender development measures and combinations of measures (31 columns) and missing data approaches (2 rows). Each estimate (dot) is the odds ratio from a simple logistic regression model predicting transition status from different combinations of gender development measures in data sets where missingness was ignored (upper row) and addressed via multiple imputation (bottom row). Intervals are 95% HDIs. P=peer preferences, T=Toy preferences, C=Clothing Preferences, S=Gender Similarity, I=Gender Identity.

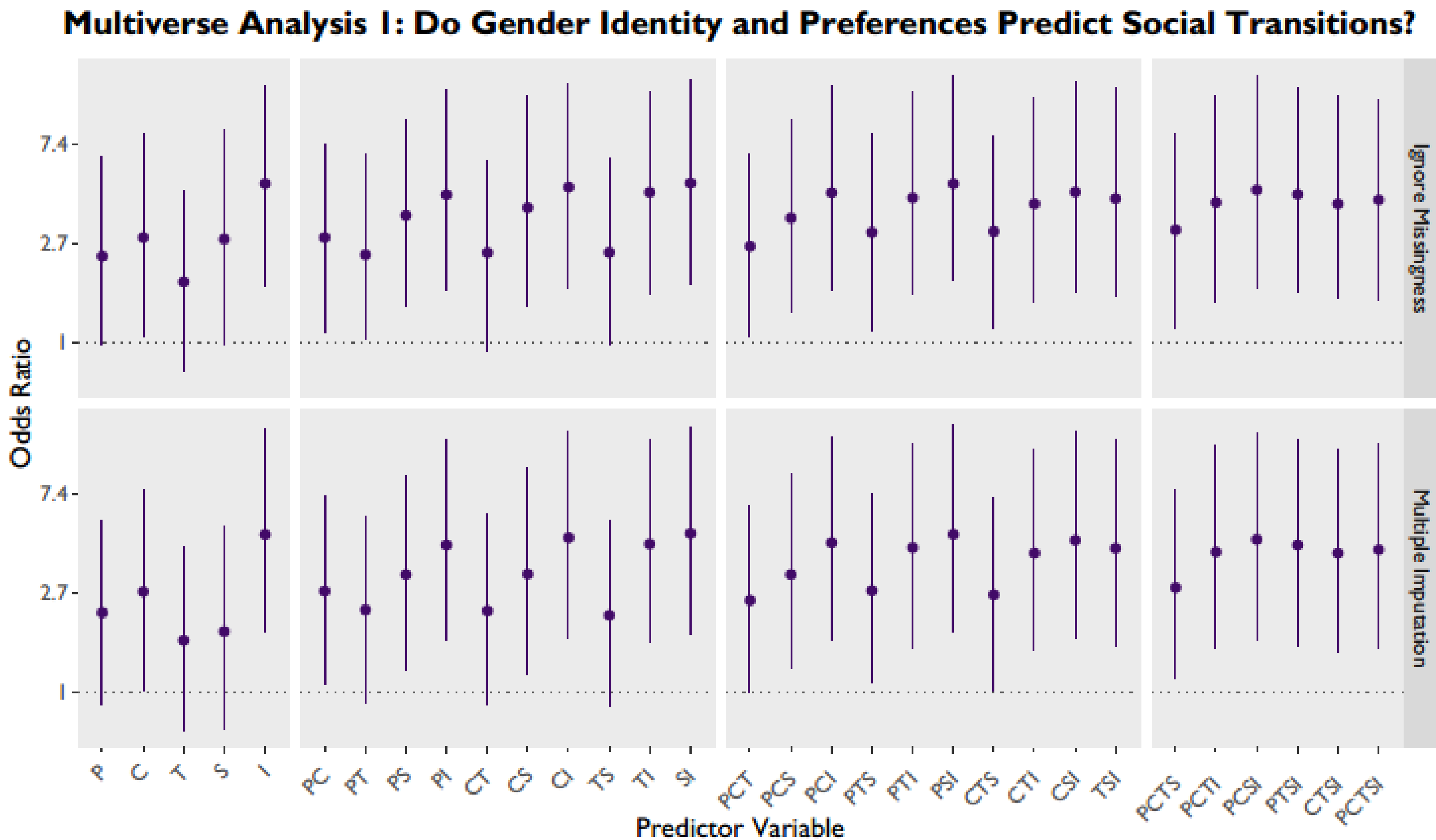
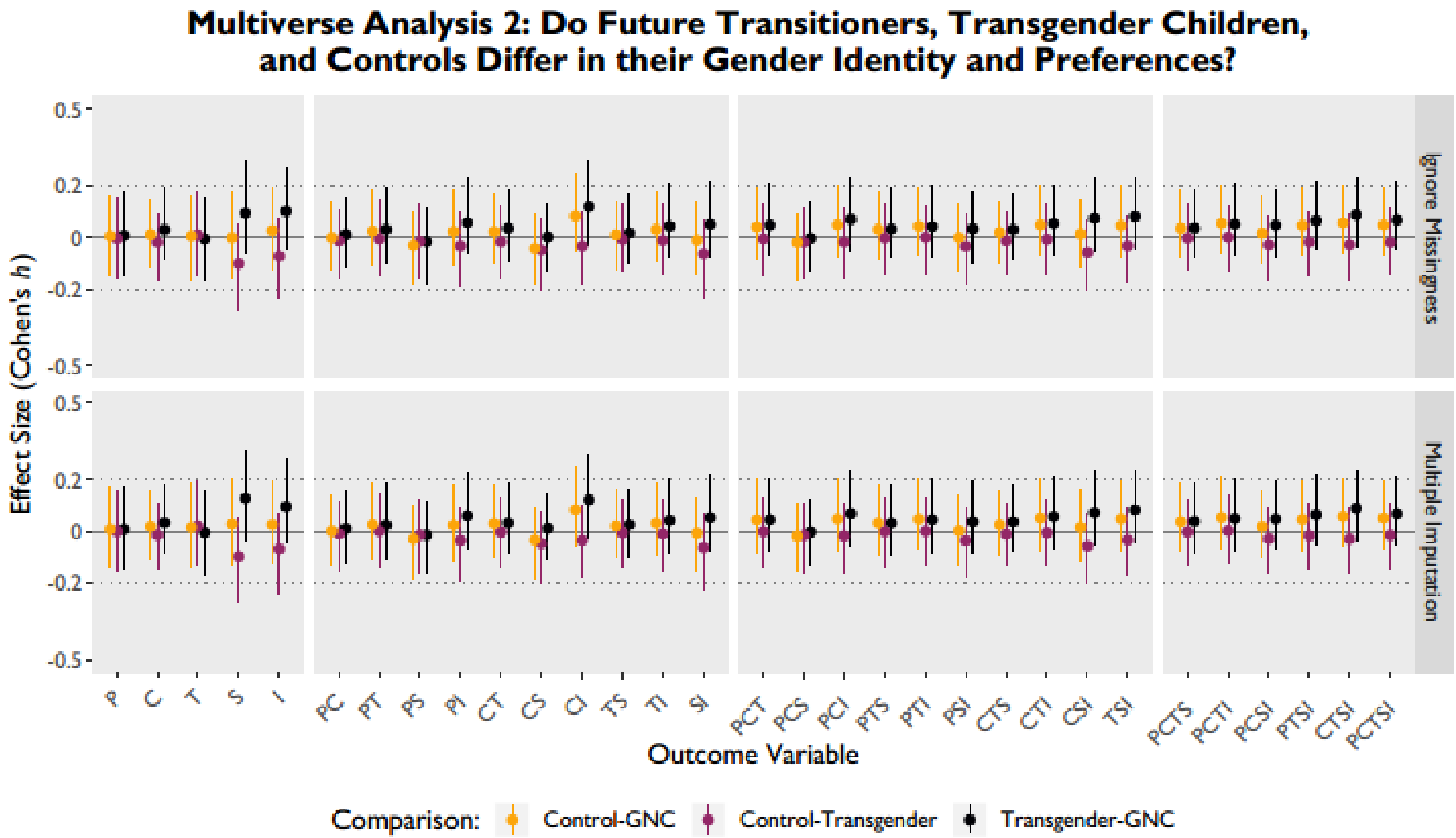


Figure 4. Multiverse analysis in which all gender development measures (31 columns) were used as outcomes in multilevel beta regression models with unique intercepts for each group using both missing data approaches (2 rows). Between-group differences were created and each estimate (dot) is the median between-group difference and intervals are 95% HDIs. Estimates and intervals are represented as effect size measures (Cohen's *h*). Dashed lines correspond to small negative (−.20) and small positive (.20) effect sizes, respectively. The ROPE is the area between these dashed lines. P=peer preferences, T=Toy preferences, C=Clothing Preferences, S=Gender Similarity, I=Gender Identity.



Discussion

The degree of gender identification and preferences expressed by gender nonconforming children predicted which children later socially transitioned. For example, assigned males that had stronger feminine gender identities and preferences were more likely to be living as girls two years later than assigned males who exhibited less feminine identities and preferences. This pattern was observed even though there are no agreed-upon standards or measures to determine whether to support a given child through a social transition. As past work has linked extremity of gender nonconformity in the absence of early social transitions to transgender identification later in life (e.g., Singh, 2012; Steensma et al., 2013; Wallien & Cohen-Kettenis, 2008), these findings could suggest that the children transitioning at early ages may also be more likely to identify as transgender later in life. Critically, these results were robust to a large number of analytic (e.g., inclusion of covariates or using different prior distributions for our Bayesian analyses) and data processing (e.g., how we combined the five gender development measures or handled missing data) decisions.

Children who went on to socially transition showed gender identification and preferences comparable in magnitude to children who had already transitioned (i.e., transgender participants) and those whose assigned sex and gender identity had aligned for their entire life (i.e., control participants). Said differently, an assigned male who will later transition to live as a girl is roughly as feminine before transition as a transgender girl is after a transition, and both are comparable in degree of feminine identity and preferences to a non-transgender girl. Again, this effect was remarkably robust across different analytic and data processing decisions. While replication of this effect is needed, preferably from a longitudinal study comparing a single group of children before and after transition, this finding could reduce worries that the transition itself is leading children to identify as or behave in ways more stereotypically associated with the opposite assigned sex.

One previous study examined the relation between early social transitions and later transgender identity (Steensma et al., 2013). All four children in that study who had socially transitioned in childhood identified as transgender in adolescence, while only 35% of the 123 children who did not “completely” socially transition (i.e., children who did not change pronouns) in childhood identified as transgender later. Green (2017) identified two explanations for this finding. First, children who socially transition could differ from those who do not even before transitioning. Second, transitioning could change a child’s sense of identity, making them identify more with the opposite sex group. Consistent with the first explanation, we found that the children who transitioned showed more extreme cross-sex identification and preferences before transitioning. In contrast, we found evidence that children tested after transitioning (i.e., transgender participants) did *not* differ meaningfully from those tested before transitioning (i.e., future transitioners) in terms of identification and preferences.

Limitations

A primary limitation of this work is the small sample size. We tried to address this concern by utilizing a Bayesian approach, which may be better suited to model data with small samples (Depaoli & van de Schoot, 2017). Further, we tested a sample skewed by race, class, parental education, and political orientation. This may or may not reflect the set of children who are socially transitioning now or in the future. Thus, replication with a larger and more diverse sample would increase confidence in our conclusions and suggest that these results are not sample specific. Another limitation is that follow-up occurred only two years after testing. Some of the 49 children who had not transitioned when the present study ended could transition in the future and some of the 36 children who did transition could transition again to the gender aligning with their assigned sex. Therefore, re-analysis of data at later points will be necessary. Finally, as this research was exploratory in that we did not pre-register our analytic plan before collecting the data, one concern could be that we used “researcher degrees of freedom” to obtain a desired pattern of results (Simmons et al., 2011). However, we examined the sensitivity of our results to a variety of data processing and analytic decisions (e.g., by conducting multiverse analyses), which demonstrated that the results from this (small) sample were robust to many researcher degrees of freedom.

We found that 41% of our sample of gender nonconforming children had transitioned roughly two years after initial testing sessions. We believe this percentage is likely an over-estimate of how many gender nonconforming children in the general population will socially transition. We recruited through listservs and events serving transgender children and gender nonconforming children, and the word *transyouth* was widely utilized in recruitment materials. The families responding to our recruitment may have already been questioning whether their child could be transgender, while parents of children showing less extreme gender nonconformity might be less likely to have reached out. As evidence, Figure 1 shows that nearly all participants showed cross-sex identification and preferences (i.e., these were not simply “less masculine” boys). We therefore caution against using this work as a broad reference point for rates of social transitions.

Finally, as in all studies reporting means of groups, care should be taken in extending group-level results to individuals. Some children who showed high levels of identification and preferences opposite their sex at birth did not transition. There were also children who did transition but did not show especially extreme gender identification and preferences opposite their assigned sex.

Conclusion

Despite limitations and a need for future replications, these results provide preliminary evidence that extremity of identification with the gender “opposite” one’s assigned sex predicts childhood social transitions. Moreover, differences in gender extremity likely exist prior to — and not because of — social transitions.

References

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology, 42,* 155-162. http://dx.doi.org/10.1037/t00748-000

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

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1). http://dx.doi.org/10.18637/jss.v080.i01

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software, 76*. https://doi.org/10.18637/jss.v076.i01

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, P., Cohen, J., Aiken, L. S., & West, S. G. (1999). The problem of units and the circumstance for POMP. *Multivariate Behavioral Research*, *34*, 315-346. http://dx.doi.org/10.1207/S15327906MBR3403_2

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*, 7-29. http://dx.doi.org/10.1177/0956797613504966

Depaoli, S., & van de Schoot, R. (2017). Improving transparency and replication in Bayesian statistics: The WAMBS-Checklist. *Psychological Methods*, *22*, 240-261. http://dx.doi.org/10.1037/met0000065

Dreger, A. (2009). Gender identity disorder in childhood: Inconclusive advice to parents. *Hastings Cent. Rep. 39*, 26–29. https://doi.org/10.1353/hcr.0.0102

Edwards-Leeper, L., Leibowitz, S., & Sangganjanavanich, V. F. (2016). Affirmative practice with transgender and gender nonconforming youth: Expanding the model. *Psychology of Sexual Orientation and Gender Diversity*, *3*, 165-172. http://dx.doi.org/10.1037/sgd0000167

Ehrensaft, D. (2010). "I'm a Prius": A child case of a gender/ethnic hybrid. *Journal of Gay & Lesbian Mental Health, 15*, 46-57. https://doi.org/10.1080/19359705.2011.530571

Ehrensaft, D. (2016). *The gender creative child: Pathways for nurturing and supporting children who live outside gender boxes*. New York, NY: Workman Publishing.

Fast, A. A., & Olson, K. R. (2018). Gender development in transgender preschool children. *Child Development*, *89*, 620-637. http://dx.doi.org/10.1111/cdev.12758

Ferrari, S., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, *31*, 799-815. https://doi.org/10.1080/0266476042000214501

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2017). Visualization in Bayesian workflow. arXiv preprint arXiv:1709.01449.

Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine*, *27*, 2865-2873. https://doi.org/10.1002/sim.3107

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (Vol. 2). Boca Raton, FL: CRC press.

Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness, 5*, 189-211. https://doi.org/10.1080/19345747.2011.618213

Green, R. (1987). *The "sissy boy syndrome" and the development of homosexuality*. New Haven, CT: Yale University Press.

Green, R. (2017). To transition or not to transition? That is the question. *Current Sexual Health Reports, 9*, 79-83. http://dx.doi.org/10.1007/s11930-017-0106-5

Hoekstra, R., Morey, R. D., Rouder, J. N., & Wagenmakers, E. J. (2014). Robust misinterpretation of confidence intervals. *Psychonomic Bulletin & Review, 21*, 1157-1164. http://dx.doi.org/10.3758/s13423-013-0572-3

Kruschke, J. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Burlington, MA: Academic. 2nd ed.

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review, 25*, 178-206. http://dx.doi.org/10.3758/s13423-016-1221-4

Malpas, J. (2011). Between pink and blue: A multi-dimensional family approach to gender nonconforming children and their families. *Family Process, 50*, 453-470. http://dx.doi.org/10.1111/j.1545-5300.2011.01371.x

Martin, C. L., Andrews, N. C., England, D. E., Zosuls, K., & Ruble, D. N. (2017). A dual identity approach for conceptualizing and measuring children's gender identity. *Child Development, 88*, 167-182. http://dx.doi.org/10.1111/cdev.12568

McElreath, R. (2016). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Boca Raton, FL: Chapman & Hall/CRC Press.

Olson, K. R., Key, A. C., & Eaton, N. R. (2015). Gender cognition in transgender children. *Psychological Science, 26*, 467-474. https://doi.org/10.1177/0956797614568156

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Sandberg, D. E., Meyer-Bahlburg, H. F., Ehrhardt, A. A. & Yager, T. J. (1993). The prevalence of gender-atypical behavior in elementary school children. *Journal of the American Academy of Child & Adolescent Psychiatry, 32*, 306–314. http://dx.doi.org/10.1097/00004583-199303000-00010

Sherer, I. (2016). Social transition: Supporting our youngest transgender children. *Pediatrics, 137*, e20154358. http://dx.doi.org/10.1542/peds.2015-4358

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359-1366. http://dx.doi.org/10.1177/0956797611417632

Singh, D. (2012). *A follow-up study of boys with gender identity disorder*. (Doctoral dissertation). Department of Human Development and Applied Psychology, University of Toronto, Toronto, Ontario, Canada.

Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods, 11*, 54-71. https://doi.org/10.1037/1082-989X.11.1.54

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*, 702-712. https://doi.org/10.1177/1745691616658637

Steensma, T. D., McGuire, J. K., Kreukels, B. P., Beekman, A. J., & Cohen-Kettenis, P. T. (2013). Factors associated with desistence and persistence of childhood gender dysphoria: A quantitative follow-up study. *Journal of the American Academy of Child & Adolescent Psychiatry, 52*, 582-590. https://doi.org/10.1016/j.jaac.2013.03.016

Van Buren, S., & Groothuis-Oudshoorn, C. (2011). MICE: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software, 45*, 1-67. https://doi.org/10.18637/jss.v045.i03

Vink, G., Frank, L. E., Pannekoek, J., & Buuren, S. (2014). Predictive mean matching imputation of semicontinuous variables. *Statistica Neerlandica*, *68,* 61-90. https://doi.org/10.1111/stan.12023

Wallien, M. S., & Cohen-Kettenis, P. T. (2008). Psychosexual outcome of gender-dysphoric children. *Journal of the American Academy of Child & Adolescent Psychiatry*, *47*, 1413-1423. https://doi.org/10.1097/CHI.0b013e31818956b9.

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: issues and guidance for practice. *Statistics in Medicine*, *30*, 377-399. https://doi.org/10.1002/sim.4067

Zucker, K. J., & Bradley, S. J. (1995). *Gender identity disorder and psychosexual problems in children and adolescents*. New York, NY: Guilford Press.