



Is Research on Transgender Children What It Seems? Comments on Recent Research on Transgender Children with High Levels of Parental Support

The Linacre Quarterly
2020, Vol. 87(1) 9-24
© Catholic Medical Association 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0024363919884799
journals.sagepub.com/home/lqr



Walter R. Schumm, PhD¹ , and Duane W. Crawford, PhD¹

Abstract

Recent research on transgender children who have had support from their parents for their transitioning has concluded that their mental health is virtually no different than that of nontransgender children. Such research has been extensively cited, over 370 times in the past three years. Most of the hundreds of reviews received the stated results of the studies with little caution. However, the research featured numerous statistical errors and omissions, the implications of which would likely lead neutral observers to conclude that the mental health of transgender children, even when supported by their parents, was poorer than that of the groups of control children. In particular, levels of anxiety as reported by both parents and their transgender children appear to be significantly higher, and the transgender children's reports of self-worth appear to be significantly lower. Although reports regarding depression are not as significantly different, the effect sizes were generally in a similar direction as the other outcomes, being less favorable for the transgender children. Such issues highlight the need for careful examination of statistical research, even when published in highly regarded medical journals. As with other research, findings from the early stages of controversial research may often be premature. Further research is needed to explore factors underlying these results.

Summary: Some scholars have believed that if transgender children were supported by their parents before the children reached puberty, the generally higher rates of mental illness experienced by many transgender persons might be prevented or alleviated. Dr. Kristina Olson of the Department of Psychology at the University of Seattle was the first scholar to have studied groups of transgender children who were being supported by their parents and to have compared them to a control group of children and to siblings of the same transgender children. Her conclusion was that there were minimal, if any, differences in anxiety, depression, and self-worth among the groups of children; her research has since been cited extensively as having found just that. We reanalyzed her raw data and found that, to the contrary, the transgender children, even when supported by their parents, had significantly lower average scores on anxiety and self-worth. Often, a significantly higher percentage of transgender children, compared to controls, featured preclinical or clinical levels of anxiety. Parental support of transgender children may temporarily reduce levels of poor mental health for some transgender children, but it does not appear to eliminate those problems for all transgender children. Our findings should serve as a warning against accepting research at a surface level, which can lead to acceptance of invalid information and pursuit of ineffective interventions.

¹ Applied Family Science Unit, School of Family Studies and Human Services, College of Human Ecology, Kansas State University, Manhattan, KS, USA

Corresponding Author:

Walter R. Schumm, PhD, School of Family Studies and Human Services, Kansas State University, 1700 Anderson Avenue, Manhattan, KS 66506, USA.
Email: schumm@ksu.edu

Keywords

Children, Parents of transgender children, Research methodology, Statistical errors, Transgender

Introduction

A recent article in *The Atlantic* magazine (Yong 2019) discussed the controversial issue of treatment of transgender children (Fitzgibbons 2015), citing the research of Dr. Kristina Olson of the Department of Psychology of the University of Seattle, Washington. While the article focused on factors related to a child's future transitioning, it also mentioned Olson's earlier studies that suggested "that children who are supported and affirmed in their transitions are just as mentally healthy as cisgender peers." Two of the articles published by Dr. Olson and her colleagues (Olson et al. 2016b; Durwood, McLaughlin, and Olson 2017) have been cited over 370 times in just two to three years. Most of those citations and literature reviews have accepted their results as having proven that the mental health of transgender children is on a par with that of cisgender children *if* the parents of the transgender children affirm the gender identity of their transgender children. In the same article in *The Atlantic* magazine, Yong cited Professor Aaron Devor (University of British Columbia) who hoped that Olson's seminal work would have an "Evelyn Hooker effect," (Hooker 1957, 1958), meaning that Olson's research would change the entire field of social science with respect to the treatment of transgender children as Hooker's research (Schumm 2012; Cameron and Cameron 2012) had done for homosexuality. However, the quality of literature reviews relies on their correct interpretation of the research they cite. Arriving at a correct interpretation is only as likely as the original authors' accurate interpretation of their own results. This issue boils down to whether or not Olson's research was accurately conducted and interpreted by herself and her colleagues. A number of statistical errors that were detected alerted us to question those matters (Schumm et al. 2019).

Objectives

Therefore, our plan here is (1) to explain what Olson and her colleagues reported in their research and (2) to show, with our reanalysis of their data, that their own conclusions about and interpretations of their data were not merely incorrect but led readers to assume conclusions about their findings that were the opposite of what their data actually imply. Furthermore, we will (3) evaluate whether Olson

and her colleagues used the best scientific procedures for their analyses, using a checklist from Du Prel, Rohrig, and Blettner (2009). Then, we will (4) observe how some scientific papers and literature reviews have gone on to report even more incorrect findings from the research of Olson and her colleagues.

Background: Olson's Research with Transgender Children

First Study

Olson et al. (2016b) compared seventy-three transgender children (ages 3 to 12 years, who had been supported by their parents for their transitioning gender identity) with a control group of seventy-three age- and gender-matched cisgender children and forty-nine nontransgender siblings of the transgender children. Most of the children were white, with average ages between 7.7 and 8.3 years. Most of their families, 81–90 percent, earned more than US\$75,000 annually. Specifically, Olson et al. (2016b) measured anxiety and depression for each of the children, as reported by their parents, and reported results for all children and results for each natal gender as subgroups of the children. They did not find significant results for the main effects of gender or group or for any interactions between gender and group. They found that the parents of the transgender children in their study reported lower internalizing (based on an average of anxiety and depression scores) scores for their children than had been found for transgender children in two other samples, from Canada and from the Netherlands (Olson et al. 2016b, 5).

The apparent conclusion was that if parents would only affirm their children's transgender status, then mental health problems would be prevented so strongly that the children would become essentially similar in mental health to their own siblings or to cisgender children from other comparable families. It is not clear what type of statistical analyses were used.

The expected positive correlation between the scores of the transgender children and their siblings would normally indicate that they used a repeated measures analysis of variance, while the independence of the scores between the transgender children and their control group of children suggests the use

of an independent samples analysis of variance. While we suspect they used the latter approach, using that approach would increase error rates in their statistics. Because they did not report standard deviations for their results, it was not possible to calculate effect sizes (the magnitude of their effects) as opposed to the statistical significance levels of their results. With small samples, such as those used by Olson et al., large effects may not be statistically significant.

Olson et al. (2016b) concluded that they had found no differences in depression and only marginally elevated levels of anxiety for the transgender children compared to those children from the other two groups. That interpretation was modified in their final conclusion section to “these results provide clear evidence that transgender children have levels of anxiety and depression no different from their nontransgender siblings and peers” (p. 7). McKean, Vande Voort, and Croarkin (2016) noted that nearly a third of the children in the Olson study were so young, the measures used had not been validated for such a young age-group; they also noted that the sample used was of very high socioeconomic status, whose results might not generalize to the average family.

Second Study

Durwood, McLaughlin, and Olson (2017, 117) included 63 transgender children, 63 age-matched controls, and 38 siblings aged 9 to 14 years, all of whom completed measures of depression and anxiety; parents also reported on their children’s apparent depression and anxiety. Some of the parents had participated in the earlier study (Olson et al. 2016a, b). In addition, 116 transgender children, 122 control children, and 72 siblings, ages 6 to 14 years of age, completed a measure of self-worth. The children were older than those in the Olson et al. (2016b) sample, with average ages from 10.6 to 10.9 for those who were measured on depression and anxiety. For those assessed on self-worth, average ages ranged between 9.1 and 9.3 years. The percentage of white children ranged between 50 percent and 66 percent, while the percentage of families earning more than US\$75,000 a year ranged between 71 percent and 82 percent. Mean scores and standard deviations, as well as the percent of children in a clinically high range for both depression and anxiety, were reported for all children and for those children from families earning US\$75,000 or less annually. Without explanation, scores for children from higher income families were not reported. Overall scores on

self-worth were not reported; however, Durwood, McLaughlin, and Olson (2017) broke the self-worth scores into three subgroups based on age of the children (youngest, oldest, in between) across the transgender, control, and sibling groups of children. With respect to comparisons of the mean scores across the three groups of children, the only statistically significant finding reported by Durwood et al. was from parents with respect to anxiety ($p = .002$).

Missing Information

Olson et al. (2016b) did not report clinical levels of anxiety or depression and did not report standard deviations. Without standard deviations, it is not possible for other scholars to calculate significance levels or effect sizes. Durwood, McLaughlin, and Olson (2017) did not report results for high-income families nor did they report overall mean scores and standard deviations for self-worth over their entire sample. Accordingly, we asked the authors to provide us with that information. Readers can read some of our back-and-forth discussion of these issues in the comment section associated with the Olson et al. (2016b) article, with dates between May 4 and August 8, 2018.

Research Questions

While we have questioned some of the details of their statistics elsewhere (Schumm et al. 2019), here the objective was to examine the validity of their major conclusions by assessing the accuracy of their statistical design and testing.

Thus, our primary research question was whether or not Olson et al. (2016b) and Durwood, McLaughlin, and Olson (2017) found, as many reviews have suggested, that there were no significant differences and only minimal, if any, substantial differences (interpreted as an effect size of .20 or greater being of substantive importance) in depression, anxiety, or self-worth in the two studies. At least one review of these two studies concluded that transgender children scored as well as other on both anxiety and depression (Allen, Watson, and VanMattson 2019, 3). The two studies have been cited over 370 times (Google Scholar), an indication of their impact on medical science concerning transgender children. We also wanted to consider whether they used the best methods available (Du Prel, Rohrig, and Blettner 2009) for conducting their research and/or reporting their results and the scholarly impact of their research.

Method

In the spring of 2018, the Alliance for Defending Freedom asked the author to review the research published by Olson et al. (2016b). The author agreed to take that article to the class he was teaching in basic statistics at the Wamego campus of Highland Community College and engage in a critique of its use of statistics as an applied exercise that might result in a publication for the students who were interested in participating in that project. Students were given course credit for their participation. Numerous statistical concerns were noted, as published elsewhere (Schumm et al. 2019). However, in many cases, Olson et al. (2016b) had not reported standard deviations or other data that were necessary to independently assess the statistical significance or the effect sizes of their findings. The author e-mailed Professor Olson and asked for the missing information, which was graciously provided.

Participants

Olson provided enough data in her reports or by inquiry to permit reconstruction of sample data for both groups. The sibling group was not included in the analyses because the sibling group came from the same families as the transgender children and the most appropriate statistical tests would have been paired samples *t*-tests, which cannot be calculated without knowing the correlation of results across the two related samples. Because the transgender and cisgender groups of children were independent of each other, it was possible to compare those two groups statistically with independent samples *t*-tests.

Analyses

Olson et al. (2016b) and Durwood, McLaughlin, and Olson (2017) used two types of comparison between their transgender children and the cisgender children. First, they compared mean scores between the two groups on depression and anxiety and self-worth; second, in some cases, they reported the percentages of children in each group that scored at or above certain clinical or preclinical levels of anxiety or depression.

Given a mean and standard deviation for two groups, along with sample sizes for each group, one can conduct independent samples *t*-tests from freely available websites (see Schumm et al. 2019). Given a percentage of children in each sample at or above clinical levels, it is possible to reconstruct the data and use binary logistic regression to obtain an odds

ratio that provides information on the relative odds of a child from one group versus the other group of scoring at or above the same clinical levels (for depression or anxiety). We used an α level of .05 to assess statistical significance and did not use Bonferroni procedures (dividing α by the number of tests) because their use inflates the chance of type II error. We took into consideration one- and two-tailed tests because most previous research has found that transgender children tend to score higher with respect to depression and anxiety than control groups of children. In addition to assessing statistical significance, the effect sizes of differences were calculated, with effect sizes of .20 or greater deemed of substantive significance and those of .24 or greater (per Cuijpers 2017) deemed of clinical significance. Effect sizes of .50 or greater will be deemed of sufficient magnitude to be observable to a careful observer, without using statistical methods (Cohen 1992). In order to provide a more conservative approach, where data were available, we performed Bayesian analyses (BF_{10}) and reported results when $BF > 3.0$. We also investigated the statistical power of Olson et al.'s analyses, using a power calculator at www.anzmtg.org/stats/PowerCalculator/PowerTest.

Hypotheses

The following hypotheses will be tested for statistical ($p < .05$) and for substantive significance (i.e., effect sizes $\geq .20$), using both one-tailed and two-tailed tests. Transgender children and/or their parents will report higher anxiety and higher depression scores for the transgender children than will children and/or their parents for cisgender children in the control group. When possible, results will be assessed for families above and below selected cut points on total family income. Differences by natal gender will be examined where data were available.

Transgender children, as reported by their parents or by the children themselves, will experience a higher odds ratio (≥ 1.5 deemed of substance) of reaching or exceeding clinical levels of depression or anxiety than parents or their children will report for cisgender children in the control group. Transgender children will report lower self-worth scores than will cisgender children, for the whole sample and for each of three different age groups in the overall sample.

Results

Raw data reported in Olson et al. (2016b) or Durwood, McLaughlin, and Olson (2017) as well as that

provided to us by the authors are presented in Table 1. Our analyses of the data in Table 1 are reported in Table 2. Tables 3 and 4 present data from Table 1 in a format that makes it easier to observe differences as a function of the natal and chosen gender identities of the children in terms of their scores on depression (Table 3) and anxiety (Table 4). Table 5 is a summary of our findings in Table 2. When odds ratios could not be calculated, we fit the results into one of three likely outcomes, of odds ratios of less than 1.5, 1.5–2.99, and 3.0 or higher, based on the effect size found with the *t*-tests. For depression, five results fell into the 1.5–2.99 range, with one above 3.0 and two below 1.5. For anxiety, one fell below 1.5 while seven were above 3.0. Table 3 shows that natal girls reported higher levels of depression than did natal boys, but the effect was about twice as strong for transgender children as for cisgender children. Table 4 shows that in terms of anxiety, transgender children reported higher levels, regardless of natal gender, but the difference was greater for transboys than for transgirls. Table 6 represents a power analysis of the samples used by Olson and her colleagues.

Even though depression was associated with fewer significant results (5.3 percent), most of the results with respect to depression favored cisgender children (78.9 percent, 15/19) in terms of having positive effect sizes while 52.6 percent (10/19) involved effect sizes of .20 or greater. If the underlying population results for the depression tests had been even (50/50), the chances of getting fifteen or more on one side out of the nineteen tests would be $p < .01$, $z = 2.29$. Anxiety outcomes were mostly in favor of cisgender children (94.7 percent, 18/19), with 84 percent involving effect sizes of .20 or greater with 53 percent (10/19) being significant statistically. The chances of getting eighteen of nineteen results for anxiety in favor of cisgender children, if the true chance per test was only 50 percent, were $p < .0001$, $z = 3.67$. In terms of self-worth, all (4/4) of the results favored cisgender children with 75 percent (3/4) involving effect sizes of .20 or greater and 50 percent being significant statistically. Combining the results for depression and anxiety together, the chances of getting thirty-three or more of thirty-eight tests to favor cisgender children would be $p < .00001$, $z = 4.37$. The chance of finding thirty-seven of all of the forty-two tests on the side of cisgender children would be $p < .000001$, $z = 4.78$. Altogether, 88 percent (37/42) of the tests favored cisgender children with over two-thirds (29/42) featuring effect sizes of .20 or greater, with 31 percent (13/42) being significant statistically by two-tailed

tests and nearly 48 percent (20/42) significant by one-tailed tests. In terms of effect sizes of .24 or greater, we found nearly 62 percent (26/42) of that size or larger. The issue of statistical power is important for studies with the range of sample sizes involved in Olson et al. (2016b) and Durwood, McLaughlin, and Olson (2017). For the *t*-tests, we correlated sample size, measured in terms of the degrees of freedom for each *t*-test, against the significance level obtained and found $r = -.44$ ($p < .03$) with Spearman's $\rho = -.46$ ($p < .02$), such that the larger the sample used, the lower the observed level of significance, with a large effect size for this calculation ($d > .80$). This indicates that sample size played a key role in whether or not the observed results, regardless of their actual substantive importance, were statistically significant.

Statistical Power

Table 6 contains information on the statistical power associated with many of the statistical tests conducted with Olson et al.'s data. For most of their analyses, statistical power was sufficient for a high chance of detecting effect sizes of .50 at $\alpha = .05$. However, at the same time, most of their analyses did not have sufficient ($>.50$) statistical power to detect effect sizes of .20 or smaller. That situation may account for the difference between having effect sizes of .20 or greater for 69 percent of the forty-two tests but two-sided significant results for only 31 percent of the results and one-sided significant results for only 48 percent of the forty-two tests.

Objections

The primary objection to our methodology might be that we did not use a Bonferroni correction—that we did not divide α (.05) by forty-two, yielding an α of .0012 as the new criterion for any of the forty-two test results to have been deemed significant (using that criterion would have yielded only one significant result, for parental reports of child's anxiety in Durwood, McLaughlin, and Olson 2017). If the results were entirely due to chance, we would expect 5 percent to be significant, not 31 percent—or 10 percent to be significant (α set to .10) rather than 48 percent. Clearly, there are more significant results than would have been expected by chance alone. Thus, the evidence appears to indicate that a Bonferroni approach would *over correct* for the risk or problem of getting significant results that were actually obtained by chance alone. A thought experiment can reveal the limitations of the Bonferroni

Table 1. Raw Data from Olson et al. (2016b) and Durwood, McLaughlin, and Olson (2017).

Article	Group Reporting	Outcome	Mean/SD for Transgender Children	Mean/SD for Control Group Children	Percentage of Preclinical/Clinical for Transgender Children	Percentage of Preclinical/Clinical for Control Children
Olson	Parents, for boys and girls	Depression	50.13/7.42 N = 73	48.36/7.31 N = 73	11.0/5.5	5.5/2.7
	Parents, for boys and girls	Anxiety	54.17/8.82 N = 73	50.87/6.97 N = 73	26.0/15.1	9.6/1.4
	Parents, transboys versus control boys	Depression	50.80/7.20 N = 21	48.04/8.31 N = 21		
	Parents, transboys versus control boys	Anxiety	55.17/7.23 N = 21	51.06/7.64 N = 21		
	Parents, transgirls versus control girls	Depression	49.84/7.56 N = 52	48.50/6.92 N = 52		
	Parents, transgirls versus control girls	Anxiety	53.70/9.44 N = 52	50.78/6.74 N = 52		
	Parents, transboys versus control girls	Depression	50.80/7.20 N = 21	48.50/6.92 N = 52		
	Parents, transboys versus control girls	Anxiety	55.27/7.23 N = 21	50.78/6.74 N = 52		
	Parents, transgirls versus control boys	Depression	49.84/7.56 N = 52	48.04/8.31 N = 21		
	Parents, transgirls versus control boys	Anxiety	53.70/9.44 N = 52	51.06/7.64 N = 21		
Durwood	All parents	Depression	50.2/8.8 N = 63	49.4/7.8 N = 63	6.3	3.2
	Low-income parents	Depression	53.4/8.6 N = 18	50.8/11.1 N = 13	5.6	7.7
	High-income parents	Depression	48.84/8.66 N = 45	49.01/6.83 N = 50	4.4	2.0
	All parents	Anxiety	54.9/9.0 N = 63	49.6/8.6 N = 63	22.2	4.8
	Low-income parents	Anxiety	56.2/8.4 N = 18	50.0/6.8 N = 13	22.2	0.0
	High-income parents	Anxiety	54.39/9.30 N = 45	49.94/9.06 N = 50	17.8	6.0
	All children	Depression	48.7/9.4 N = 63	46.4/8.0 N = 63	6.3	1.6
	Low-income children	Depression	46.7/9.3 N = 18	47.3/10.8 N = 13	0.0	7.7
	High-income children	Depression	49.56/9.45 N = 45	46.20/7.25 N = 50	8.9	0.0
	All children	Anxiety	52.0/9.6 N = 63	49.0/7.7 N = 63	12.7	3.2
	Low-income children	Anxiety	49.5/7.5 N = 18	48.5/10.5 N = 13	5.6	15.4
	High-income children	Anxiety	53.06/10.21 N = 45	49.19/6.96 N = 50	15.6	0.0
	All children	Self-worth	3.46/0.542 N = 116	3.61/0.415 N = 121		
	Younger children	Self-worth	3.50/0.54 N = 53	3.62/0.39 N = 59		
	Middle children	Self-worth	3.47/0.55 N = 49	3.68/0.35 N = 48		
	Older children	Self-worth	3.30/0.51 N = 14	3.37/0.64 N = 14		

Note: Data are reported on what Olson et al. reported to us by e-mail (two decimal points) or in their original reports (one decimal point). Results of clinical or subclinical levels did not always add from the high- and low-income groups to the total group, for unexplained reasons (we used what we were sent). Columns 6 and 7 report the results for transgender and cisgender children, respectively, in terms of both the percentage of children scored at preclinical and clinical levels of the mental health outcome variables. The larger percentage represents the preclinical level.

Table 2. Results for Analysis of Data from Olson et al. (2016b) and Durwood, McLaughlin, and Olson (2017).

Article	Group Reporting	Outcome	Test Used	Results <i>t/df</i> or Odds Ratio	Effect Size	<i>p</i> (two-tailed)	
Olson et al.	Parents, all children	Depression	<i>t</i> -test	1.45 (144)	0.23	<.15	
		Anxiety	<i>t</i> -test	2.51 (144)	0.42	<.02	
		Depression, preclinical	Odds ratio	2.12	0.20	<.24	
		Depression, clinical	Odds ratio	2.06	0.14	<.42	
		Anxiety, preclinical	Odds ratio	3.32	0.44	=.012	
		Anxiety, clinical	Odds ratio	12.8	0.51	=.016	
						BF = 3.38	
	Parents, transboys versus control boys	Depression	<i>t</i> -test	1.15 (40)	0.36	<.26	
		Anxiety	<i>t</i> -test	1.83 (40)	0.57	<.08	
		Depression	<i>t</i> -test	0.94 (102)	0.18	<.35	
		Anxiety	<i>t</i> -test	1.82 (102)	0.36	<.08	
		Depression	<i>t</i> -test	1.27 (71)	0.33	<.21	
		Anxiety	<i>t</i> -test	2.52 (71)	0.65	<.02	
		Depression	<i>t</i> -test	0.90 (71)	0.23	<.38	
Durwood et al.	Parents, transgirls versus control boys	Anxiety	<i>t</i> -test	1.14 (71)	0.29	<.26	
		All parents	<i>t</i> -test	0.54 (124)	0.10	<.60	
		Low-income parents	<i>t</i> -test	0.74 (29)	0.27	<.47	
		High-income parents	<i>t</i> -test	0.11 (93)	-0.02	<.92	
		All parents	Anxiety	<i>t</i> -test	3.38 (124)	0.60	=.001
		Low-income parents	Anxiety	<i>t</i> -test	2.19 (29)	0.80	<.04
		High-income parents	Anxiety	<i>t</i> -test	2.36 (93)	0.53	=.02
	All parents, clinical levels	Depression	Odds ratio	2.07	0.15	<.42	
		Low-income parents	Depression	Odds ratio	0.71	-0.09	<.82
		High-income parents	Depression	Odds ratio	2.28	0.14	<.51
		All parents, clinical levels	Anxiety	Odds ratio	5.71	0.53	<.005
		Low-income parents	Anxiety	Odds ratio	Cannot be calculated	0.14	<.08
		High-income parents	Anxiety	Odds ratio	3.39	0.38	<.09
						BF = 7.52	
Durwood et al.	All children	Depression	<i>t</i> -test	1.48 (124)	0.26	<.15	
	Low-income children	Depression	<i>t</i> -test	0.17 (29)	-0.06	<.87	
	High-income children	Depression	<i>t</i> -test	1.96 (93)	0.40	<.054	
	All children, clinical levels	Depression	Odds ratio	4.20	0.25	<.21	
	Low-income children	Depression	Odds ratio	Cannot be calculated	-0.44	<.42	

(continued)

Table 2. (continued)

Article	Group Reporting	Outcome	Test Used	Results <i>t/df</i> or Odds Ratio	Effect Size	<i>p</i> (two-tailed)
	High-income children	Depression	Odds ratio	Cannot be calculated	0.45	<.05
	All children	Anxiety	<i>t</i> -test	1.93 (124)	0.34	<.06
	Low-income children	Anxiety	<i>t</i> -test	0.31 (29)	0.11	<.76
	High-income children	Anxiety	<i>t</i> -test	2.18 (93)	0.45	<.04
	All children, clinical levels	Anxiety	Odds ratio	4.44	0.36	<.07
	Low-income children	Anxiety	Odds ratio	0.32	-0.33	=.38
	High-income children	Anxiety	Odds ratio	Cannot be calculated	0.62 BF = 9.17	<.005
	All children	Self-worth	<i>t</i> -test	2.40 (235)	0.31	<.02
	Younger children	Self-worth	<i>t</i> -test	1.36 (110)	0.26	<.18
	Middle children	Self-worth	<i>t</i> -test	2.24 (95)	0.45	<.03
	Older children	Self-worth	<i>t</i> -test	0.32 (26)	0.12	<.76

Note: Positive effect sizes indicate that parents of cisgender children or their children reported better mental health scores than did the transgender children or their parents. Even though our one-sided directional hypotheses would permit one-sided statistical tests, we used more conservative two-sided tests through Table 2. One-sided test results can be obtained by dividing the reported *p* values by 2. If one of the groups has no cases (0 percent) at or above clinical levels, then an odds ratio cannot be calculated; in those cases, effect sizes and significance levels were derived from Pearson zero-order correlations and/or a two-sided Fisher's Exact Test. BF_{10} = Bayes factor where scores from 3 to 10 represent moderate support for the alternative hypothesis and scores above 10 represent strong support.

Table 3. Raw Data (Mean/*SD*/*N*) from Olson et al. (2016b) on Depression as a Combination Pattern of Natal Gender and Transgender Status.

	Cisgender Children	Transgender Children
Natal boys	48.04 (8.31), <i>N</i> = 21	49.84 (7.56), <i>N</i> = 52
Natal girls	48.50 (6.92), <i>N</i> = 52	50.80 (7.20), <i>N</i> = 21

Table 4. Raw Data (Mean/*SD*/*N*) from Olson et al. (2016b) on Anxiety as a Combination Pattern of Natal Gender and Transgender Status.

	Cisgender Children	Transgender Children
Natal boys	51.06 (7.64), <i>N</i> = 21	53.70 (9.44), <i>N</i> = 52
Natal girls	50.78 (6.74), <i>N</i> = 52	55.27 (7.23), <i>N</i> = 21

correction. Let us suppose that we had five subscales for each of which the results were significant at $p = .01$. If the five subscales were combined to form a total scale, we might find it significant at $p = .01$. Even though all six of our tests would have been significant statistically, if we divide α (.05) by six for the six tests, then none of our tests would remain significant at the new Bonferroni level of α (.008). Thus, we think that Bonferroni corrections are too conservative, especially when the research objective is to not reject the null hypothesis.

A secondary objection might be that Olson's kindness in providing most of the information that was requested about her data (that had been

omitted in her published articles) was punished by contradicting her results in a published article. The intention is not to punish any attempt at transparency because transparency helps drive the proper functioning of science, which is to slowly, over time get us to a better understanding of reality. Results are results. The implications of results can vary. As is discussed shortly, the results in Tables 1 and 2 could be used to argue that transgender children need more support and/or that, for at least some transgender children, even with parental affirmation, their transitioning experience is somehow associated with lower conditions of mental health.

Table 5. Summary of Results from Table 2.

Article	Report from	Outcome	Number of Tests	Percentage of Positive	Percentage of $d \geq .20$	Percentage of $p < .05$	Percentage of $p < .10$
Olson et al.	Parents	Depression	7	100	71.4	None	None
	Parents	Anxiety	7	100	100	57.1	85.7
Durwood et al.	Parents	Depression	6	66.7	33.3	8.3	16.7
	Children	Depression	6	66.7	66.7	16.7	33.3
	Parents	Anxiety	6	100	83.3	66.7	100
	Children	Anxiety	6	83.3	66.7	33.3	66.7
	Children	Self-worth	4	100	75.0	50.0	50.0
Combined	Parents and children	Depression	19	78.9	52.6	5.3	10.5
Combined	Parents and children	Anxiety	19	94.7	84.2	52.6	84.2
Combined, all outcomes	Parents and children	All outcomes	42	88.1	69.0	31.0	47.6

Note: Percentage of negative are not counted in percentage for d and p .

Table 6. Statistical Power Calculations for Olson et al.'s Samples.

Transgender Sample Size	Control Group Sample Size	Power for $d = .50$, One-sided	Power for $d = .50$, Two-sided	Power for $d = .20$, One-sided	Power for $d = .20$, Two-sided
14	14	.58	.45	.18	.11
21	21	.73	.62	.23	.15
49	48	.97	.93	.40	.28
52	52	.97	.95	.42	.30
63	63	.99	.98	.48	.35
73	73	.996	.989	.52	.40
116	121	.999	.999	.70	.58

Note: Power calculations from www.anzmtg.org/stats/PowerCalculator/PowerTest with $\alpha = .05$.

A third objection might be my sources of funding, which have included some conservative organizations as well as government agencies such as the National Science Foundation. However, we have provided the data used in all of our calculations, so if our funding sources caused bias, it should be testable.

A fourth objection might be that the tests that were used were not independent as some were tests of subgroups of the main group. That is a reasonable concern, although Olson et al. (2016b) and Durwood, McLaughlin, and Olson (2017) did not find it necessary to adjust their methods accordingly. Because determining the exact effect of nonindependence is very complex (Schumm and Canfield 2011), we have deferred that sort of reanalysis of the data.

A fifth objection might relate to not using the sibling group in our analyses. We were concerned that the comparison between the transgender children and the control group of cisgender children was clearly an independent samples type of problem while the comparison between the transgender children and their siblings was *not* an independent samples type of problem. We wanted to focus on what we were sure of. However, when we treated the sibling group as an independent sample and ran a one-way analysis of variance on self-worth, the results, $F(2, 306) = 3.84$, were still significant ($p = .023$). It is unclear why Durwood, McLaughlin, and Olson (2017) did not report those results as statistically significant.

A sixth objection might be that we did not rely upon Bayesian statistics (Aczel et al. 2018). We tried

to use SPSS to obtain a Bayesian statistic for the oldest group of children regarding self-worth, but both the means and standard deviations failed tests for legitimacy (Heathers et al. 2018), preventing us from reconstructing the data, so we could perform a Bayesian analysis. It could be argued that if part of the data failed the Granularity-Related Inconsistency of Means (GRIM) and other tests, the legitimacy of the data and conclusions are in question (Brown and Heathers, 2017). We were able to run Bayesian analyses for the percentage comparisons; we found that the results for associations gave stronger results against the null hypothesis, so we used *t*-test Bayesian results in order to be even more conservative. Our four strongest Bayesian results were associated with effect sizes from 0.44 to 0.62, which would make sense, the strongest results yielding the best Bayesian results.

A seventh objection might be that we used control group values to compare with the values for transgender children, when it would be more appropriate to use national norms as comparisons. Some reviews of Olson et al.'s reports mentioned that the transgender children were doing well compared to national norms. However, for each of the five (*t*-test) anxiety comparisons from Olson et al. (2016b), when we compared the transgender results to a national norm of 50.0 (*SD* = 10.0, *N* = 100), instead of effect sizes of .42 ($p < .02$), .57 ($p < .08$), .36 ($p < .08$), .65 ($p < .02$), and .29 ($p < .26$) from Table 2, we obtained effect sizes of .44 ($p < .05$), .55 ($p < .05$), .38 ($p < .05$), .55 ($p < .05$), and .37 ($p < .05$), effects comparable in magnitude but usually of stronger statistical significance because we used a larger sample size for the simulated national comparison group. In other words, in terms of anxiety, Olson et al.'s (2016) results show the transgender children having higher levels of anxiety whether they are compared to the control group or to a national norm. In Durwood, McLaughlin, and Olson (2017), the results for anxiety are similar, regardless of the comparison used, because the control group scores are very close to the national norm of 50.0. For depression scores in both reports, the results are mixed, with some scores below national norms and others above them.

An eighth objection might be that we didn't compare the transgender children's scores to the comparison samples from Canada and the Netherlands that Olson et al. (2016b) used, based on clinic-referred children with possible gender identity disorder. We didn't focus on that because one might well expect clinic-referred children to score lower on mental health measures than children who were not referred. However, in the interests of completeness, we

compared anxiety scores for Olson's transgender children ($n = 73$) against the internalizing scores of Canada ($n = 343$) and from the Netherlands ($n = 123$). Using the original scores from Cohen-Kettenis et al. (2003), we obtained *t*-test results of 4.78 ($df = 414$, $d = 0.62$, $p < .0001$) and 6.78 ($df = 194$, $d = 1.00$, $p < .0001$). While those differences are substantial and statistically significant, they are not surprising given the selection effect differential between the samples. Interestingly, Olson et al. (2016b, 5) did not report any statistical tests across the three samples.

Discussion

"Comrade, your statement is factually incorrect."
"Yes, it is. But it is politically correct." (Codevilla 2016, 37)

Quality of Methodological Analysis

Du Prel, Rohrig, and Blettner (2009) provided several criteria for evaluating the quality of scientific articles published in medical journals. They indicated that higher quality studies would have statistical power $> .50$, that the statistical methods used would be clearly described, that the statistical results would be presented comprehensively and clearly, the effect sizes or confidence intervals would be presented, and that the conclusions would be supported by the study's findings. As we observed in Table 6, their data had insufficient statistical power for detecting small effect sizes, though adequate for detecting medium effect sizes. The statistical methods were not clearly described, particularly with respect to the fact that data from the transgender children and their siblings should have been positively correlated, lending itself to paired samples testing while the data for the transgender children and the control group of children lent itself to independent samples testing.

Some of these concerns about scientific quality and statistical clarity have been addressed elsewhere (Schumm et al. 2019). Effect sizes were not presented nor were confidence intervals. Moreover, as shown in Table 5, the study's conclusions of virtually no differences were not supported by the actual data. Among other issues, the participants were not randomly assigned to the tested groups, and it was not clear what proportion of the participants overlapped between the two studies and how dropouts may have differed between the two studies. Though the studies were "pioneering" (Kuvallanka, Gardner, and Munroe 2019), they had many substantial and

important limitations according to the criteria discussed by Du Prel, Rohrig, and Blettner (2009).

Scholarly Impact

Together, both articles have been cited over 370 times in the past two or three years. Chen et al. (2018, 76) found the two studies to be the *only* ones that had yet “explored psychosocial functioning in socially transitioned prepubertal children,” highlighting the critical importance of the two studies. As noted, Kuvalanka, Gardner, and Munroe (2019, 103) cited the research as “pioneering.” It is clear that the reported results of these two studies have had a huge impact on the field of social science and medicine.

Most of the scholars who have cited their articles have interpreted the findings in the same way as did Olson et al. (2016b) and Durwood, McLaughlin, and Olson (2017). Some articles repeated what the two articles claimed—that there are no differences in depression or self-worth between transgender children and control children with only slight or minimal differences in anxiety. Turban (2017, 101) stated clearly that “Transgender youth in this study showed only mildly increased levels of anxiety (below the subclinical range)” and that “child-report levels of self-worth were similar to those of matched non-transgender controls.” Chen et al. (2018, 76) noted that “Results show that transgender children did not differ from either control group on depression scores and had only marginally higher anxiety scores” and the two groups did not differ “on ratings of depression or self-worth and had marginally higher anxiety scores.” Other studies came to similar conclusions regarding either or both of the two articles (Alberse et al. 2019, 389; Alegria 2018, 132; Bonifacio et al. 2019, e72; Campo-Engelstein 2019, 85; Cartaya and Lopez 2018, 47; Chen, Hidalgo, and Garofalo 2017, 342; Deardorff et al. 2019, 143; Janicka and Forcier 2016, 33; Oswald and Lederer 2017, 7; Reilly et al. 2019; Saleem and Rizvi 2017, 3; Shumer 2018, 1; Toomey, Syvertsen, and Shramko 2018, 7; Valdiserri et al. 2019, 579; Wanta and Unger 2017, 126).

Other reports have concluded that mental health outcomes were similar between transgender children and age-matched controls and did not mention even minimal differences in anxiety symptoms (Becerra-Culqui et al. 2018, 8; Busa, Janssen, and Lakshman 2018, 28; Chodzen et al. 2019, 468; Cicero and Wesp 2017, 7; Ehrensaft et al. 2018, 255; Green 2017, 81; Nahata et al. 2017, 189;

Newhook et al. 2018, 333; Telfer et al. 2018, 134; Turban and Ehrensaft 2018, 1232).

Going further, some reviews concluded that being affirmed socially in their identified gender provided “substantial improvements in their mental health” (Riley 2018, 204) compared to transgender children not affirmed (even though there was no such comparison group in the studies) or that mental health disparities would be resolved “immediately” (Cicero and Wesp 2017, 6) if children were affirmed in their gender identity or that, if children were so affirmed, disparities in “emotional distress are reduced or eliminated” (Gower et al. 2018, 788). One review concluded that given parental support, transgender children would “thrive” (Ehrensaft 2017, 64), while another review concluded that anxiety and depression were *both* found to have decreased in Olson’s (2016b) study (Allen, Watson, and VanMattson 2019, 3). Another review (Kuvalanka, Gardner, and Munroe 2019, 103) mentioned that there were no differences in depression in Olson et al.’s (2016b) study, but said nothing about anxiety or self-worth, leaving the impression that there were probably no other variables of interest besides depression.

It is interesting that the seriousness of differences between transgender and cisgender children may be partly a function of how those differences are reported. While a difference in mean scores of 54 versus 51 may not seem like much (anxiety; Olson et al. 2016b), a difference of 26 percent versus less than 10 percent having preclinical levels of anxiety may seem more substantial. A parent may not care whether their child scores a point or two lower on some particular psychological test, but if asked whether they’d rather have a 26 percent risk of having a child with preclinical or clinical levels of significant depression or anxiety versus less than a 10 percent risk, it is presumed that most would choose the latter. It’s good that 74 percent of the transgender children didn’t show preclinical levels of anxiety, but that leaves an important question of how to help the other 26 percent of the transgender children. Do they need more protection in school from bullying? Do they need more peer support? Are their schools lacking in evidence-based policies to support transgender children? Are there ways in which their parents or other relatives could be more supportive than they have been? Are there other ways they could be helped? Unfortunately, the data at present don’t give us much guidance for those questions.

The scientific consensus would seem to be that transgender children are no different than cisgender children if they have parental support. However, our

reanalysis of Olson et al. (2016b) and Durwood, McLaughlin, and Olson (2017) would seem to indicate otherwise. While differences for depression were fewer, 79 percent favored cisgender children and over half (52.6 percent) involved effect sizes of .20 or greater in favor of cisgender children. Results for anxiety and self-worth were more notable in that nearly 95 percent (22/23) of those two outcomes favored cisgender children with over 82 percent (19/23) involving effect sizes of .20 or greater, with over 52 percent (12/23) being statistically significant (two-tailed) and over 78 percent (18/23) being significant (one-tailed). Olson et al. (2016b, 1) stated that the transgender children “had only marginally higher anxiety symptoms.” The effect size to which they referred was 0.42, nearly the 0.50 at which Cohen (1992) indicated an effect could be seen by a naked eye observer. In Durwood, McLaughlin, and Olson (2017) at least one effect size for anxiety reached the 0.80 level, which Cohen (1992) deemed “large,” well beyond what could be observed by the naked eye, without statistics. The results should have been interpreted as evidence that even with high levels of parental support, transgender children have lower levels of mental health, especially with respect to higher levels of anxiety and lower levels of self-worth, though marginally with respect to depression, supporting for the most part our three research hypotheses. It would seem that Ioannidis (2005) was correct, that much early research is taken too seriously, with major flaws being overlooked.

Clinical Implications

The most apparent implication would be to search for other sources of minority stress (Valentine and Shipherd 2018), such as discrimination or bullying from peers as an explanation for the higher levels of anxiety or depression observed among the transgender youth. Yet, if the bullying or discrimination from peers seems able to overcome the positive effects of parental support, school systems may be failing to adequately protect transgender children. However, if one accepts the scientific consensus viewpoint, those school systems may be getting an underserved “pass” in terms of their lack of effectiveness in protecting transgender children. In their response to a letter to the editor by McKean, Vande Voort, and Croarkin (2016), Olson et al. (2016a) reported that the mental health of their sample of transgender children had changed from a mean of 50.2 for the youngest to 56.9 for the oldest children (higher scores representing lower mental health).

Without standard deviations, it is not possible to know the exact effect sizes or significance levels involved in that change, but if we assumed both standard deviations to be 8.0, then the effect size of the decline would be 0.84, with $t(32) = 2.33$ ($p < .03$). Furthermore, examination of Durwood, McLaughlin, and Olson (2017) indicates that the self-worth of both transgender, effect size of 0.37, n.s., and cisgender, effect size of 0.56, $t(71) = 1.88$, $p < .07$, children appears to be declining with older age, which may suggest that school systems (or parents?) are not being as effective at supporting all children, transgender or cisgender, as they advance through higher grades (lacking the raw data, independent-samples t -tests were used in lieu of paired samples t -tests across times). However, if minority stress were the only explanation, it would not account for the parallel decline in self-worth reported by cisgender children, who presumably are not victims of minority stress in the same way that transgender children might be. Olson et al. (2016a, b) and Durwood, McLaughlin, and Olson (2017) did not offer any scientific tests of these more detailed hypotheses, so we remain in the dark as to why these observed differences seemed to occur.

Another clinical implication may be related to the higher anxiety and depression scores reported for the transboys in Olson et al.’s (2016b) sample (Durwood et al. did not break down their results by gender). It would seem that natal and transgender girls retained a depression differential associated with being female (Table 3) while acquiring a much higher anxiety score than those children in the other three groups (Table 4). Those unusual results may deserve further investigation. Our thought is that many cisgender boys have a hard time learning what it means to be a man, when they have the biological advantage of being natal males; how much more challenging would it be for a natal girl to figure out how to be a man, without the advantage of being a natal male? The threat of starting to menstruate or to develop breasts might add to the anxiety of trying to be a man. Conversely, natal males might have to worry less about developing more muscle as that would fit in with being a tomboy, so it might arouse less anxiety. While natal male transgender girls might develop a larger penis, unlike breasts, the penis may be easier to hide under clothing. Further research might clarify some of these issues.

Research Implications

Even though it is also untested, another hypothesis could be that transgender children’s concerns are not

being resolved through parental support or through social transitioning. This hypothesizes that even if it is assumed that the transgender children express a desire to transition and receive support for doing so, perhaps that transitioning experience is not as satisfying to them as they might have expected, leaving some of the transgender children with anxiety about having made the right decision (or not), or having associated questions of their own self-worth, if not other co-occurring mental health concerns (Bechard et al. 2017). It is also possible that some transgender children may not feel as much like their opposite sex as simply having sexual attractions to the same sex and feel that one way to resolve feeling “gay or lesbian” is to change their gender rather than accepting their sexual orientation (perhaps children who want to transition suddenly, without prior indications of being transgender, may be more likely to belong to this latter group). That doesn’t mean that all transgender children might feel that way, just enough of them to lower the average mean scores for transgender children as a whole. Future research should attempt to compare and test such competing hypotheses, though both or neither might be correct for some of the children.

Leaving the mental health of transgender children aside, the results raise serious questions about the validity of at least some medical or social science research (Ioannidis 2005). Results that are interpreted in one direction when the data actually speak in another direction have not been not an isolated phenomenon. It has been seen since the 1950s with Evelyn Hooker’s research (Hooker 1957, 1958; Schumm, 2012), later with same-sex parenting research (Schumm 2018; Schumm and Crawford 2019), and now with research on transgender children. In the case of Olson et al. (2016b) and Durwood, McLaughlin, and Olson (2017), not only were there numerous statistical errors (Schumm et al. 2019), but a great deal of data and results, including some significant results, were *not* reported until the authors were queried. Not reporting significant results may occur but when the apparent conclusion is that there weren’t any significant results, leaving out significant findings can be seen as self-serving to the idea of maintaining support for the null hypothesis regardless of the facts. Is good science being thrown under the bus for the sake of politically correct agendas? It’s difficult to escape a sense that such is not an uncommon occurrence in areas of considerable political controversy. One has to wonder what other areas of controversial science may have been infected with this type of problem.

It seems apparent that the methodological recommendations of Du Prel, Rohrig, and Blettner (2009) were not followed in these two studies. Outright errors were made. The issues we have brought up were significant enough to have caught the attention of peer reviewers and been corrected prior to publication; for that matter, the journal editors might have caught at least some of them on their own, prior to peer review. Furthermore, many of the scholars who have cited Olson et al. (2016b) and Durwood, McLaughlin, and Olson (2017) have also reported conclusions even less accurate than that reported by the original authors, raising concerns about the accuracy of the interpretation of literature in literature reviews.

Conclusion

Whereas Olson et al. (2016b) and Durwood, McLaughlin, and Olson (2017) concluded that transgender children with strong parental support had, at worst, only *slightly* higher levels of anxiety with no differences in self-worth or depression; a reanalysis of their findings suggests otherwise, with slightly higher levels of depression but significantly and substantively meaningful differences in anxiety and self-worth, and with results favoring cisgender children, *even when* the transgender children *had* high levels of parental support for their gender transitioning.

Such results leave open the possibility that discrimination from outside the families of the transgender children is having a corrosive effect on their mental health, especially as they get older, a possibility that should not be glossed over because of initially positive results and a possibility that if ignored could do further harm to transgender children by delaying preventive or remedial programs to prevent or ameliorate discrimination and bullying.

It is possible that one reason the two articles have been so highly cited is that they essentially let all other parts of society “off the hook” for the care of transgender children, assuming those children have parental support. It may also be possible that factors intrinsic to transgenderism or related to comorbid mental health concerns might be playing a role in mental health or self-worth. Further research is needed to sort out those different possibilities. Not only do we have to guard against science becoming little more than polemic (Green 2017), but we need to be sure that scientists remain dedicated to reporting their data and statistical testing fully and accurately.


Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: The senior author received an individual grant in the summer of 2018 from the Alliance Defending Freedom in order to conduct a literature review of issues regarding transgender children. Prior to that grant, the author with students in a basic statistics class reviewed the two articles discussed here for various statistical errors (see Schumm et al. 2019).

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Walter R. Schumm, PhD  <https://orcid.org/0000-0003-3097-3551>

References

- Aczel, Balazs, Bence Palfi, Aba Szollosi, Marton Kovacs, Barnabas Szaszi, Peter Szecsi, Mark Zrubka, et al. 2018. "Quantifying Support for the Null Hypothesis in Psychology: An Empirical Investigation." *Advances in Methods and Practices in Psychological Science* 1: 357–66.
- Alberse, Anne-Marie E., Annelou L. C. de Vries, Wieteke S. Elzinga, and Thomas D. Steensma. 2019. "Self-perception of Transgender Clinic Referred Gender Diverse Children and Adolescents." *Clinical Child Psychology and Psychiatry* 24:388–401.
- Alegria, Christine Aranburu. 2018. "Supporting Families of Transgender Children/Youth: Parents Speak on Their Experiences, Identity, and Views." *International Journal of Transgenderism* 19:132–43.
- Allen, Luke R., Laurel B. Watson, and Sarah B. VanMattson. 2019. "Trans Young Adults' Reflections on Adolescent Sources of Extra-familial Support." *Journal of LGBT Youth*.
- Becerra-Culqui, Tracy A., Yuan Liu, Rebecca Nash, Lee Cromwell, W. Dana Flanders, Darios Getahun, and Michael Goodman. 2018. "Mental Health of Transgender and Gender Nonconforming Youth Compared with Their Peers." *Pediatrics* 141:1–19.
- Bechard, Melanie, Doug P. VanderLaan, Hayley Wood, Lori Wasserman, and Kenneth J. Zucker. 2017. "Psychosocial and Psychological Vulnerability in Adolescents with Gender Dysphoria: A 'Proof of Principle' Study." *Journal of Sex & Marital Therapy* 43:678–88.
- Bonifacio, Joseph H., Catherine Maser, Katie Stadelman, and Mark Palmert. 2019. "Management of Gender Dysphoria in Adolescents in Primary Care." *Canadian Medical Association Journal* 191:e69–75.
- Brown, Nicholas J. L., and James A. J. Heathers. 2017. "The GRIM Test: A Simple Technique Detects Numerous Anomalies in the Reporting of Results in Psychology." *Social Psychological and Personality Science* 8:363–369.
- Busa, Samantha, Aron Janssen, and Mallika Lakshman. 2018. "A Review of Evidence Based Treatments for Transgender Youth Diagnosed with Social Anxiety Disorder." *Transgender Health* 3:27–33.
- Cameron, Paul, and Kirk Cameron. 2012. "Re-examining Evelyn Hooker: Setting the Record Straight with Comments on Schumm (2012) Reanalysis." *Marriage & Family Review* 48:491–523.
- Campo-Engelstein, Lisa. 2019. "A Two-pronged Approach to Minimizing Harms for Transgender Youth: Medical Interventions and Social Interventions." *American Journal of Bioethics* 19:85–87.
- Cartaya, Julia, and Ximena Lopez. 2018. "Gender Dysphoria in Youth: A Review of Recent Literature." *Current Opinion in Endocrinology, Diabetes, and Obesity* 25:44–48.
- Chen, Diane, Laura Edwards-Leeper, Terry Stancin, and Amy Tishelman. 2018. "Advancing the Practice of Pediatric Psychology with Transgender Youth: State of the Science, Ongoing Controversies, and Future Directions." *Clinical Practice in Pediatric Psychology* 6:73–83.
- Chen, Diane, Marco A. Hidalgo, and Robert Garofalo. 2017. "Parental Perceptions of Emotional and Behavioral Difficulties among Prepubertal Gender-nonconforming Children." *Clinical Practice of Pediatric Psychology* 5:342–52.
- Chodzen, Gia, Marco A. Hidalgo, Diane Chen, and Robert Garofalo. 2019. "Minority Stress Factors Associated with Depression and Anxiety among Transgender and Gender-nonconforming Youth." *Journal of Adolescent Health* 64:467–71.
- Cicero, Ethan C., and Linda M. Wesp. 2017. "Supporting the Health and Well-being of Transgender Students." *The Journal of School Nursing* 33:95–108.
- Codevilla, Angelo M. 2016. "The Rise of Political Correctness." *Claremont Review of Books* XVI:37–43.
- Cohen, Jacob. 1992. "A Power Primer." *Psychological Bulletin* 112:155–59.
- Cohen-Kettenis, Peggy T., Allison Owen, Vanessa G. Kaijser, Susan J. Bradley, and Kenneth J. Zucker. 2003. "Demographic Characteristics, Social Competence, and Behavior Problems in Children with Gender Identity Disorder: A Cross-national, Cross-clinic Comparative Analysis." *Journal of Abnormal Child Psychology* 31: 41–53.
- Cuijpers, Pim. 2017. "Four Decades of Outcome Research on Psychotherapies for Adult Depression: An Overview of a Series of Meta-analyses." *Canadian Psychology* 58:7–19.
- Deardorff, Julianna, Lindsay T. Hoyt, Rona Carter, and Elizabeth A. Shirtcliff. 2019. "Next Steps in Puberty

- Research: Broadening the Lens toward Understanding Populations." *Journal of Research on Adolescence* 29: 133–54.
- Du Prel, Jean-Baptist, Bernd Rohrig, and Maria Blettner. 2009. "Critical Appraisal of Scientific Articles." *Deutsches Arzteblatt International* 106:100–05.
- Durwood, Lily, Katie A. McLaughlin, and Kristina R. Olson. 2017. "Mental Health and Self-worth in Socially Transitioned Transgender Youth." *Journal of the American Academy of Child & Adolescent Psychiatry* 57: 116–23.
- Ehrensaft, Diane. 2017. "Gender Nonconforming Youth: Current Perspectives." *Adolescent Health, Medicine, and Therapeutics* 8:57–67.
- Ehrensaft, Diane, Shawn V. Giammattei, Kelly Storck, Amy C. Tishelman, and Colton Keo-Meier. 2018. "Prepubertal Social Gender Transitions: What We Know; What We Can Learn—A View from a Gender Affirmative Lens." *International Journal of Transgenderism* 19:251–68.
- Fitzgibbons, R. P. 2015. "Transsexual Attractions and Sexual Reassignment Surgery: Risks and Potential Risks." *The Linacre Quarterly* 82:337–50.
- Gower, Amy L., Nic Rider, Camille Brown, Barbara J. McMorris, Eli Coleman, Lindsay A. Taliaferro, and Maria E. Eisenberg. 2018. "Supporting Transgender and Gender Diverse Youth: Protection against Emotional Distress and Substance Use." *American Journal of Preventive Medicine* 55:787–94.
- Green, Richard. 2017. "To Transition or Not to Transition? That is the Question." *Current Sexual Health Reports* 9: 79–83.
- Heathers, James, Jordan Anaya, Tim van der Zee, and Nicholas J. L. Brown. 2018. "Recovering Data from Summary Statistics: Sample Parameter Reconstruction via Iterative Techniques (SPRITE)." *PeerJ Preprints* 6:e26968v1. doi: 10.7287/peerj.preprints.26968v1.
- Hooker, Evelyn. 1957. "The Adjustment of the Male Overt Homosexual." *Journal of Projective Techniques* 21: 18–31.
- Hooker, Evelyn. 1958. "Male Homosexuality in the Rorschach." *Journal of Projective Techniques* 22: 33–54.
- Ioannidis, John P. A. 2005. "Why Most Published Research Findings Are False." *PLoS Medicine* 2:e124.
- Janicka, Agnieszka, and Michelle Forcier. 2016. "Transgender and Gender Nonconforming Youth: Psychosocial and Medical Considerations." *Rhode Island Medical Journal* 99:31–34.
- Kuvalanka, Katherine A., Molly Gardner, and Cat Munroe. 2019. "All in the Family: How Extended Family Relationships Are Influenced by Children's Gender Diverse and Transgender Identities." In *Families in Transition: Parenting Gender Diverse Children*, edited by Andrew R. Gottlieb and Arlene I. Lev, 102–17. New York: Harrington Park Press.
- McKean, Alastair J., Jennifer L. Vande Voort, and Paul E. Croarkin. (2016). "Lack of Rating Scale Normalization and a Socioeconomically Advantaged Population Limits the Generalizability of Preadolescent Transgender Findings." *Pediatrics* 138:1203A.
- Nahata, Leena, Gwendolyn P. Quinn, Nicole M. Caltabelotta, and Amy C. Tishelman. 2017. "Mental Health Concerns and Insurance Denials among Transgender Adolescents." *LGBT Health* 4:188–94.
- Newhook, Julia Temple, Kelley Winters, Jake Pyne, Cindy Holmes, Stephen Feder, Sarah Pickett, and Marilynne Sinnott. 2018. "Teach Your Parents and Providers Well: Call for Refocus on the Health of Trans and Gender-diverse Children." *Canadian Family Physician* 64:332–35.
- Olson, Kristina R., Lily Durwood, Madeleine DeMeules, and Katie A. McLaughlin. 2016a. "Author Response to McKean, Vande Voort, and Croarkin (2016)." *Pediatrics* 138:1203B.
- Olson, Kristina R., Lily Durwood, Madeleine DeMeules, and Katie A. McLaughlin. 2016b. "Mental Health of Transgender Children Who Are Supported in Their Identities." *Pediatrics* 137:e20153223.
- Oswalt, Sara B., and Alyssa M. Lederer. 2017. "Beyond Depression and Suicide: The Mental Health of Transgender College Students." *Social Sciences* 6:1–10.
- Reilly, Marie, Vanessa Desousa, Alexandra Garza-Flores, and Ellen C. Perrin. 2019. "Young Children with Gender Nonconforming Behaviors and Preferences." *Journal of Developmental and Behavioral Pediatrics* 40:60–71.
- Riley, Elizabeth. 2018. "Bullies, Blades, and Barricades: Practical Considerations for Working with Adolescents Expressing Concerns Regarding Gender and Identity." *International Journal of Transgenderism* 19:203–11.
- Saleem, Fatima, and Syed W. Rizvi. 2017. "Transgender Associations and Possible Etiology: A Literature Review." *Cureus Journal of Medical Science* 9:e1984.
- Schumm, Walter R. 2012. "Re-examining a Landmark Research Study: A Teaching Editorial." *Marriage & Family Review* 48:465–89.
- Schumm, Walter R., and Kenneth R. Canfield. 2011. "Statistically Evaluating Multiple Comparisons among Correlated Measures: A Practical Example." *Psychology and Education—An Interdisciplinary Journal* 48: 51–55.
- Schumm, Walter R., and Duane W. Crawford. 2019. "Scientific Consensus on Whether LGBTQ Parents Are More Likely (or Not) to Have LGBTQ Children: An Analysis of 72 Social Science Reviews of the Literature Published Between 2001 and 2017." *Journal of International Women's Studies* 20:1–12.
- Schumm, Walter R., Duane W. Crawford, Mary M. Fawver, Nevada K. Gray, Zackery N. Niess, and Abigail D. Wagner. 2019. "Statistical Errors in Major Journals: Two Case Studies Used in a Basic Statistics Class to Assess Understanding of Applied Statistics." *Psychology and Education—An Interdisciplinary Journal* 56: 35–42.
- Schumm, Walter R. 2018. *Same-Sex Parenting Research: A Critical Assessment*. London, UK: Wilberforce, Press.

- Shumer, Daniel. 2018. "Health Disparities Facing Transgender and Gender Nonconforming Youth Are Not Inevitable." *Pediatrics* 141:e20174079.
- Telfer, Michelle M., Michelle A. Tollit, Carmen C. Pace, and Ken C. Pang. 2018. "Australian Standards of Care and Treatment Guidelines for Transgender and Gender Diverse Children and Adolescents." *Medical Journal of Australia* 209:132–36.
- Toomey, Russell B., Amy K. Syvertsen, and Maura Shramko. 2018. "Transgender Adolescent Suicide Behavior." *Pediatrics* 142:e20174218.
- Turban, Jack L. 2017. "Transgender Youth: The Building Evidence Base for Early Social Transition." *Journal of the American Academy of Child & Adolescent Psychiatry* 56:101–02.
- Turban, Jack L., and Diane Ehrensaft. 2018. "Research Review: Gender Identity in Youth: Treatment Paradigms and Controversies." *The Journal of Child Psychology and Psychiatry* 59:1228–43.
- Valdiserri, Ronald O., David R. Holtgrave, Tonia C. Poteat, and Chris Beyrer. 2019. "Unraveling Health Disparities among Sexual and Gender Minorities: A Commentary on the Persistent Impact of Stigma." *Journal of Homosexuality* 66:571–89.
- Valentine, Sarah E., and Jillian C. Shipherd. 2018. "A Systematic Review of Social Stress and Mental Health among Transgender and Gender Non-conforming People in the United States." *Clinical Psychology Review* 66:24–38.
- Wanta, Jonathon W., and Cecile A. Unger. 2017. "Review of the Transgender Literature: Where Do We Go From Here?" *Transgender Health* 2:119–28.

- Yong, Ed. 2019. "Young Trans Children Know Who They Are." *The Atlantic*, January 15, 2019. <https://www.theatlantic.com/science/archive/2019/01/young-trans-children-know-who-they-are/580366/>.

Biographical Notes

Walter R. Schumm, PhD, is a professor of Applied Family Science in the College of Health and Human Sciences at Kansas State University. He earned his PhD in family studies in 1979 from Purdue University and is also a retired colonel, US Army. He has published over 250 refereed articles and numerous other books chapters and technical reports. His major authored or coedited books include *Same-Sex Parenting Research: A Critical Assessment* (Wilberforce Press, 2018), *Transition to Parenthood* (Springer, 2014), and *Sourcebook of Family Theories and Methods* (Springer, 2009).

Duane W. Crawford, PhD, is a professor of Applied Family Science in the College of Health and Human Sciences at Kansas State University; he previously taught at Texas Tech University and was Associate Dean of the Graduate School, Kansas State University. He has published numerous journal articles and currently teaches a large number of graduate and undergraduate courses, including undergraduate statistics and research methods, family theory, family studies, and marital interaction, among others.