

**Gender Dysphoria and Psychological Functioning in Adolescents Treated with GnRHa:  
Comparing Dutch and English Prospective Studies**

Michael Biggs

Department of Sociology and St Cross College, University of Oxford

42 Park End Street, Oxford OX1 1JD, UK; email: michael.biggs@sociology.ox.ac.uk

*Archives of Sexual Behavior*, accepted 5 June 2020

The number of children and adolescents presenting with gender dysphoria (previously labeled as gender identity disorder) has increased rapidly in Western countries. Over the last 15 years, referrals to the Tavistock and Portman NHS Foundation Trust's Gender Identity Development Service in London multiplied by a factor of 60 (Di Ceglie, 2018; Gender Identity Development Service, 2019), while those to the Center of Expertise on Gender Dysphoria in Amsterdam increased ten-fold (Arnoldussen et al., 2020). It has become standard to administer gonadotropin-releasing hormone analogs (GnRHa) to young adolescents diagnosed with gender dysphoria, in order to suppress puberty. Pioneered in the Netherlands, this treatment is known as the Dutch model (Cohen-Kettenis & Goozen, 1998; Delemarre-van de Waal & Cohen-Kettenis, 2006). One aim is to prevent the development of unwanted secondary sex characteristics and thus to facilitate subsequent physical transition with cross-sex hormones and surgery. Another aim is diagnostic, "to provide time to make a balanced decision regarding actual gender reassignment" (de Vries, Steensma, Doreleijers, & Cohen-Kettenis, 2011, p. 2276). This treatment is still experimental, as GnRHa is not licensed for gender dysphoria, though it is to treat precocious puberty (Thornhill, 2020). The GIDS has administered GnRHa to around 300 adolescents aged under 15 since 2011 (Tavistock and Portland NHS Foundation Trust, 2019). The Amsterdam clinic provided this treatment to almost 200 adolescents between 2012 and 2015 (van der Miesen, Steensma, de Vries, Bos, & Popma, 2020).

Despite the increasing prevalence of GnRHa treatment for gender dysphoria, evidence for its effects is thin. There is a particular lack of prospective or longitudinal studies, which take measures from the same subjects before treatment and afterwards. (The Appendix describes how the literature was searched.) The most important is a study of gender dysphoria and psychological functioning among Dutch adolescents (de Vries et al., 2011). The sample numbered 70, though

not all measures were available for all subjects, and so the number of observations is as few as 41. A recent study examined depression and quality of life for 23 American adolescents whose puberty was suppressed (by GnRHa or other drugs); almost half the subjects who entered the study did not complete the questionnaire, but the high rate of attrition is unexplained (Achille et al., 2020). Given the scarcity of prospective studies, it is worth analyzing unpublished data from one started by the GIDS in 2011: “Early pubertal suppression in a carefully selected group of adolescents with gender identity disorder,” later known as the Early Intervention Study (Viner et al., 2010). This introduced the Dutch model to the United Kingdom, lowering the age for GnRHa treatment from 16 to 12 (Biggs, 2019b). Longitudinal data on 30 subjects were reported to the Tavistock’s Board of Directors (Gender Identity Development Service, 2015), but have not been published in the scientific literature. The study was designed to enable comparison with the Dutch results, specifying that “the entry criteria will be consistent with the protocol used at the Amsterdam gender clinic” (Viner et al., 2010, p. 7). It also used many of the same instruments to measure psychological functioning and gender dysphoria.

This Letter’s comparison of Dutch and English samples is necessarily derived from tabulations rather than subject-level data, and so the results are tentative. The lack of any control group makes it impossible to identify the treatment’s causal effect (Zucker, 2019). Psychological measures are highly sensitive to placebo response (Kirsch, 2019). An additional complication with this treatment is that the Dutch model combines GnRHa with psychological support, and so the two effects are inevitably conflated (Biggs, 2019a; van der Miesen et al., 2020). Given the paucity of information on this treatment for gender dysphoria, however, any additional evidence has value. This Letter compares the two samples at baseline and then compares changes after GnRHa treatment. Two salient findings emerge. Firstly, the improvement in psychological

functioning seen in the Dutch sample was not repeated among the English. Following treatment, the English females experienced more internalizing problems (“female” and “male” will consistently be used to refer to sex recorded at birth). Secondly, both samples show similar divergence in bodily image, with dissatisfaction easing among males but worsening among females.

## **Data**

The Dutch sample comes from adolescents aged under 16 who were referred to the Amsterdam clinic from 2000 to 2008 (de Vries et al., 2011). Of the 111 prescribed GnRHa, the sample comprised the first 70 who then subsequently took cross-sex hormones, 37 females and 33 males. They started GnRHa at a mean age of 14.8 years (SD, 1.9). They were assessed at the outset and then on the eve of their progression to cross-sex hormones. Their time on GnRHa ranged from 0.4 to 5.1 years, with a mean of 1.9; variation in duration was not considered in the original analysis. Not all the measures were collected from all the subjects; the number of observations—with measures recorded both at baseline and at follow-up—ranged from 41 to 57. Note that later analysis of this same cohort, after cross-sex-hormones and surgery, reported lower numbers due to subsequent attrition (de Vries et al., 2014). The English sample came from 44 subjects aged between 12 and 15 years recruited to the Tavistock’s study from 2011 to 2014 (Biggs, 2019b; Gunn et al., 2015). They commenced GnRHa at a mean age of 13.2 years (SD, 1.1). “Preliminary Results from the Early Intervention Research” reported to the Tavistock’s Board of Directors provided longitudinal data on 30 subjects, 16 females and 14 males (Gender Identity Development Service, 2015). They were assessed at the outset and then after 12 months on GnRHa. Few subjects had missing values; the minimum number of observations was 28. (The

last 14 subjects in the study would not have completed 12 months on GnRHa by June 2015, when results were reported.)

These 30 English adolescents must have been incorporated into a published longitudinal study from the GIDS, in which GnRHa was administered to 60 subjects (Biggs, 2019a; Costa et al., 2015). That study, however, cannot be used for comparison. Its sample also included many older adolescents: the mean age of commencing GnRHa was 16.5 years. It reported only a single psychological outcome, the Children's Global Assessment Scale (CGAS). The figures aggregate females and males, even though there was a statistically significant difference between them at baseline. The baseline measure was calculated from 60 subjects, while the follow-up—after 12 months of GnRHa treatment—came from only 35. This attrition was unexplained by the authors. It cannot be explained by subjects stopping GnRHa because this is so rare; at the GIDS, only 1% of young adolescents discontinue GnRHa treatment (Carmichael, 2016; Gunn et al., 2015). Due to all these problems, the English data will be drawn solely from the Early Intervention Study.

Table 1 summarizes the outcomes to be compared along with the number of observations for each measure. There were three instruments for psychological functioning. CGAS reflects a clinician's judgment of the child's functioning compared to the overall population (Shaffer & Ambrosini, 1983). The Youth Self-Report (YSR) enables the adolescent to describe their problems, while the Child Behavior Checklist (CBCL) provides one parent's assessment. YSR and CBCL each yield three *T*-scores: one for Internalizing Problems like anxiety; one for Externalizing Problems like anger; and a Total Problem score, combining these two along with other problems such as social isolation (Aschenbach & Rescorla, 2001). *T*-scores are normalized relative to the child's cohort of the same age and sex. When interpreting how these psychological measures change after treatment, we should remember that youths and their parents—and

arguably clinicians too—believe in the benefits of GnRHa, and so might be inclined to discern positive changes (Kirsch, 2019).

Two instruments were available for gender dysphoria. The Body Image Scale (BIS) measures the youth's dissatisfaction with their body. It is divided into primary sex characteristics, like penis or clitoris; secondary sex characteristics, like hips; and neutral characteristics, like eyebrows (Lindgren & Pauly, 1975). The Utrecht Gender Dysphoria Scale (UGDS) captures the child's dissatisfaction with their body and with gender roles. For example, a question given to females is “I feel unhappy because I have to behave like a girl” (Schneider et al., 2016, p. 557). The English figures for UGDS took the mean of the responses to the 12 questions while the Dutch figures took the sum; the latter are rescaled by dividing by 12.

Altogether, 11 separate measures can be compared. Higher values represent worse outcomes, except for CGAS. For each measure, we know the mean and SD for females and for males at the baseline, and then following GnRHa treatment—after a year for the English sample and an average of two years for the Dutch. The original English tabulation reported repeated-measures ANOVA separately for each sex, and so the  $F$ -statistic can be used to calculate the SD of each subject's change from baseline to follow-up. (Denoting within-subject change as  $Y_{i1} - Y_{i0}$ , its SD  $= |\bar{Y}_1 - \bar{Y}_0| \sqrt{n}/\sqrt{F}$ .) The original Dutch tabulation does not provide comparable information, unfortunately, and we have to estimate the number of male and female subjects for each measure by applying the overall sexual composition (53% female).<sup>1</sup> Supplement Table S1 details the data

---

<sup>1</sup> The authors were unable to provide the original data (P. Cohen-Kettenis, personal communication, April 11, 2020; A. L. C. de Vries, personal communication, April 15, 2020).

underlying the graphs; all results can be replicated from the raw data file and Stata do-file provided as an online supplement.

## Results

We begin by comparing the two samples at baseline. For psychological functioning, Fig. 1 depicts each sample's mean score, along with the 95% confidence interval. CBCL and YSR *T*-scores are normalized with a mean of 50 and a SD of 10. There were no statistically significant differences between the Dutch and English samples for any of these scores (two-sample *t*-test, variance estimated separately for each sample). CGAS is scored from 1 to 100. The English adolescents were rated as functioning worse than the Dutch, and this difference was statistically significant ( $p = .05$  for females,  $.04$  for males). Scores in the 60s indicate "some difficulty in a single area" while those in the 70s reflect "no more than slight impairment in functioning." The English scores were worse by, on average, 7 or 8 points (for females and males, respectively). This difference echoes a more general pattern among all adolescents admitted to these two clinics: the English have worse psychological functioning than the Dutch (de Graaf et al., 2018). Fig. 2 compares gender dysphoria across the samples. Each measure averages questions scored from 1 to 5. There were no significant differences for females. Males, however, were clearly different. The English had a worse BIS score for primary sex characteristics ( $p = .002$ ), and likewise their UGDS was substantially worse ( $p = .00006$ ).

We are primarily interested in how these measures changed after GnRHa was administered. The duration of treatment differed between the two samples, of course, with the Dutch being treated for, on average, twice as long. It should be emphasized that, in the absence of a control group, changes cannot necessarily be attributed to the treatment itself. We compute how each measure changes between baseline and after the administration of GnRHa. Change is expressed

in a uniform metric by dividing by SD at baseline; the sign is reversed for CGAS, so that positive change consistently indicates improvement. It is crucial to separate the results by sex, because there is only minimal correlation between each measure's change for females and for males ( $r = .14$  for Dutch,  $-.07$  for English,  $n = 11$ ). The original Dutch tabulation did not differentiate all the results by sex, complicating comparison. Nevertheless, we know that a difference between the two samples was statistically significant if change in one sample was positive and in the other was negative and both were statistically significant.

Fig. 3 depicts how psychological functioning changed after treatment, along with 95% confidence intervals for the English sample. The uniform improvement seen in the Dutch sample was not found in the English. Indeed, there is no correlation between the two samples, considering all changes of psychological functioning for each sex ( $r = -.05$ ,  $n = 14$ ). The contrast is especially marked for females. Deterioration in the CBCL's Internalizing  $T$ -score was statistically significant for the English ( $p = .03$ ), while improvement for the Dutch was statistically significant ( $p = .001$ , both sexes combined; de Vries et al, 2011, p. 2280). Similarly, deterioration in the YSR's Internalizing  $T$ -score among the English females was almost statistically significant ( $p = .052$ ). One component of this score did show a statistically significant change: after treatment, female subjects were more likely to state that they sometimes “deliberately try to hurt or kill myself” ( $p = .01$  from the Wilcoxon test; Gender Identity Development Service, 2015, p. 53). (Affirmative answers also increased for the other question on self-harm—“think about killing myself”—but this was not statistically significant.) For females, we can be confident that the divergence between samples reflects real differences between the two patient groups, treatment regimes, or social context.



Fig. 4 shows how gender dysphoria changed after treatment. Very similar changes occurred in each sample ( $r = .89, n = 8$ ). For males, physical dissatisfaction diminished. For the English males, the improvement in BIS for primary sex characteristics was statistically significant ( $p = .05$ ). For females, by contrast, their physical dissatisfaction was exacerbated. In the English sample, the decline in BIS for neutral and secondary sex characteristics was statistically significant ( $p = .01$  and  $.03$  respectively). This deterioration was also statistically significant in the Dutch sample (de Vries et al., 2011, p. 2280).

## Conclusion

The crudity of this comparison is justified only by the scarcity of longitudinal data. More than two decades after the Dutch model was introduced, the strongest evidence for treating gender dysphoria with GnRHa comes from observations of between 41 and 57 subjects—lacking any control group (de Vries et al., 2011). The comparison with the Early Intervention Study reiterates a neglected aspect of the original Dutch findings, that female adolescents treated with GnRHa become more dissatisfied with their bodies. This fact might help to explain why this treatment almost invariably leads to further physical interventions. At the Leiden gender clinic, 96% of adolescents treated with GnRHa continued to cross-sex hormones, and only 4% discontinued (Brik, Vrouenraets, de Vries, & Hannema, 2020). Although puberty suppression is still described as an “extended diagnostic phase” (Steensma, Wensing-Kruger, & Klink, 2017, p. 765), in practice it becomes the first stage of irreversible physical transition.

The comparison also shows the hazard of extrapolating the psychological improvement found in the Dutch case to other populations. The results were more negative than positive for the English females. This might be explained by initial differences between the two samples, though the measure that deteriorated most for the English (YSR) did not differ from the Dutch at

baseline. The English started at a younger age and were treated for less time. The divergence could also reflect differences in the wider social context. It is premature to attempt explanation before further research confirms these differences in larger samples. The differences do, however, question the widespread assumption that outcomes from the Netherlands can be generalized to other countries (van der Miesen et al., 2020). In England, for example, the National Health Service's (2015) specification for gender dysphoria recommends suppressing puberty with GnRHa, citing Dutch research. It is puzzling that so many countries have adopted the Dutch model without publishing comparable prospective results; one wonders whether clinics whose findings are negative or contradictory are reluctant to publish. The most important conclusion is that clinicians who use experimental treatments for gender dysphoria should systematically collect data on outcomes and publish them.

## **Appendix: Prospective studies on the psychological outcomes of suppressing puberty with GnRHa**

Prospective studies were identified in two ways. Citations in Hembree et al. (2017) on the treatment of adolescents (section 2) yielded 52 articles. Searching PubMed (on May 24, 2020) for all articles published since 2016 with the search string “(transsexualism or transgender) and (gonadotropin or gnrrha or gnrrh)” yielded 84 articles. For each potential reference, I examined the title, and where necessary the abstract, to see whether it met the key criteria. Studies of physiological outcomes, such as bone density, were excluded.

Aside from de Vries et al. (2011)—cited by 65 articles in PubMed—and Costa et al. (2015)—cited by 14—two other studies meet these criteria. Schneider et al. (2017)—cited by 2 articles—reported various cognitive outcomes for one male; after 2 years and 4 months of GnRHa treatment, IQ had dropped by 10 points. Achille et al. (2020) reported outcomes for 15 males and 8 females after puberty suppression either using GnRHa or using antiandrogens (for males) or Medroxyprogesterone (for females). After 12 months of puberty suppression, males reported lower depression, controlling for psychiatric medication and engagement with counselling, as measured by the Center for Epidemiologic Studies Depression Scale ( $p = .008$ ). Other improvements were not statistically significant. The study, which included another group of subjects treated with cross-sex hormones, had an overall attrition rate of 47%. There was no explanation for why almost half the subjects who entered the study failed to complete the three waves of questionnaires and were thus excluded from the analysis.

## REFERENCES

- Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms and profiles*. Burlington: University of Vermont, Research Center for Children, Youth, and Families.
- Achille, C., Taggart, T., Eaton, N. R., Osipoff, J., Tafuri, K., Lane, A., & Wilson, T. A. (2020). Longitudinal impact of gender-affirming endocrine intervention on the mental health and well-being of transgender youths: Preliminary results. *International Journal of Pediatric Endocrinology*. doi:10.1186/s13633-020-00078-2
- Arnoldussen, M., Steensma, T. D., Popma, A., van der Miesen, A. I. R., Twisk, J. W. R., & de Vries, A. L. C. (2020). Re-evaluation of the Dutch approach: Are recently referred transgender youth different compared to earlier referrals? *European Child and Adolescent Psychiatry*. doi:10.1007/s00787-019-01394-6
- Biggs, M. (2019a). A letter to the editor regarding the original article by Costa et al: Psychological support, puberty suppression, and psychosocial functioning in adolescents with gender dysphoria. *Journal of Sexual Medicine*, 16, 2043. doi:10.1016/j.jsxm.2019.09.002
- Biggs, M. (2019b). Britain's experiment with puberty blockers. In M. Moore and H. Brunsell-Evans (Eds.), *Inventing transgender children and young people* (pp. 40–55). Newcastle, England: Cambridge Scholars Press.
- Brik, T., Vrouenraets, L. J. J. J., de Vries, M. C., & Hannema, S. E. (2020). Trajectories of adolescents treated with gonadotropin-releasing hormone analogues for gender dysphoria. *Archives of Sexual Behavior*. doi:10.1007/s10508-020-01660-8

- Carmichael, P. (2016, June). *Gender dysphoria in children and adolescents: Defining best practice in a fast changing context*. Paper presented at the World Professional Association for Transgender Health, Amsterdam. Video retrieved from <http://av-media.vu.nl/VUMedia/Play/581e58c338984dafb455c72c56c0bfa31d?catalog=2d190891-4e3f-4936-a4fa-2e9766ae0d0d>
- Cohen-Kettenis, P. T., & van Goozen, S. H. M. (1998). Pubertal delay as an aid in diagnosis and treatment of a transsexual adolescent. *European Child and Adolescent Psychiatry*, 7, 246–248.
- Costa, R., Dunsford, M., Skagerberg, E., Holt, V., Carmichael, P., & Colizzi, M. (2015). Psychological support, puberty suppression, and psychosocial functioning in adolescents with gender dysphoria. *Journal of Sexual Medicine*, 12, 2206–2214. doi:10.1111/jsm.13034
- de Graaf, N. M., Cohen-Kettenis, P. T., Carmichael, P., de Vries, A. L. C., Dhondt, K., Laridaen, J., ... Steensma, T. D. (2018). Psychological functioning in adolescents referred to specialist gender identity clinics across Europe: A clinical comparison study between four clinics. *European Child and Adolescent Psychiatry*, 27, 909–919. doi:10.1007/s00787-017-1098-4
- de Vries, A. L. C., McGuire, J. K., Steensma, T. D., Wagenaar, E. C. F., Doreleijers, T. A. H., & Cohen-Kettenis, P. T. (2014). Adult psychological outcome after puberty suppression and gender reassignment. *Pediatrics*, 134, 696–704. doi:10.1542/peds.2013-2958
- de Vries, A. L. C., Steensma, T. D., Doreleijers, T. A. H., & Cohen-Kettenis, P. T. (2011). Puberty suppression in adolescents with gender identity disorder: A prospective follow-up study. *Journal of Sexual Medicine*, 8, 2276–2283. doi:10.1111/j.1743-6109.2010.01943.x
- Delemarre–van de Waal, H. A. & Cohen-Kettenis, P. T. (2006). Clinical management of gender identity disorder in adolescents: A protocol on psychological and paediatric endocrinology

aspects. *European Journal of Endocrinology*, 155(suppl 1), S131–137.

doi:10.1530/eje.1.02231

Di Ceglie, D. (2018). The use of metaphors in understanding atypical gender identity development and its psychosocial impact. *Journal of Child Psychotherapy*, 44, 5–28.

doi:10.1080/0075417X.2018.1443151

Gender Identity Development Service. (2015). *Preliminary results from the Early Intervention Research*. Tavistock and Portman NHS Foundation Trust, Board of Directors Part One:

Agenda and Papers ... 23<sup>rd</sup> June 2015, 50–55. Retrieved from

<http://tavistockandportman.nhs.uk/documents/142/board-papers-2015-06.pdf>

Gender Identity Development Service. (2019). *Referrals to GIDS, 2014–15 to 2018–19*.

Retrieved from <http://gids.nhs.uk/number-referrals>

Gunn, H. M., Goedhart, C., Butler, G., Khadr, S. N., Carmichael, P. A., & Viner, R. M. (2015).

Early medical treatment of gender dysphoria: Baseline characteristics of a UK cohort beginning early intervention. *Archives of Disease in Childhood*, 100(suppl 3), A198.

doi:10.1136/archdischild-2015-308599.424

Hembree, W. C., Cohen-Kettenis, P., Gooren, L. J., Hannema, S. E., Meyer, W. J. III, Murad, M.

H., ... T'Sjoen, G. G. (2017). Endocrine treatment of gender-dysphoric/gender-incongruent persons: An Endocrine Society clinical practice guideline. *Journal of Clinical Endocrinology and Metabolism*, 102, 3869–3903. doi:10.1210/jc.2017-01658

Kirsch, I. (2019). Placebo effect in the treatment of depression and anxiety. *Frontiers in Psychiatry*, 10. doi:10.3389/fpsyt.2019.00407

Lindgren, T. W., & Pauly, I. B. (1975). A Body Image Scale for evaluating transsexuals.

*Archives of Sexual Behavior*, 4, 639–656.

- National Health Service England. (2015). *NHS standard contract for Gender Identity Development Service for children and adolescents*. Service specification E13/S(HSS)/e. Retrieved from <http://www.england.nhs.uk/wp-content/uploads/2017/04/gender-development-service-children-adolescents.pdf>
- Shaffer, D., & Ambrosini, P. (1983). A Children's Global Assessment Scale (CGAS). *Archives of General Psychiatry*, 40, 1228–1231.
- Schneider, C., Cerwenka, S., Nieder, T. O., Briken, P., Cohen-Kettenis, P. T., de Cuypere, G., ... Richter-Appelt, H. (2016). Measuring gender dysphoria: A multicenter examination and comparison of the Utrecht Gender Dysphoria Scale and the Gender Identity/Gender Dysphoria Questionnaire for Adolescents and Adults. *Archives of Sexual Behavior*, 45, 551–558. doi:10.1007/s10508-016-0702-x
- Schneider, M. A., Spritzer, P. M., Soll, B. M. B., Fontanari, A. M. V., Carneiro, M., Tovar-Moll, F., ... Lobato M. I. R. (2017). Brain maturation, cognition and voice pattern in a gender dysphoria case under pubertal suppression. *Frontiers in Human Neuroscience*, 11. doi: 10.3389/fnhum.2017.00528
- Steensma, T. D., Wensing-Kruger, S. A. & Klink, D. T. (2017). How should physicians help gender-transitioning adolescents consider potential iatrogenic harms of hormone therapy? *AMA Journal of Ethics*, 19, 762–70. doi:10.1001/journalofethics.2017.19.8.ecas3-1708
- Tavistock and Portman NHS Foundation Trust. (2019). *Gender dysphoria referrals 2012–2018*. Retrieved from [http://tavistockandportman.nhs.uk/documents/1431/FOI\\_18-19333\\_Gender\\_Dysphoria\\_Treatment\\_2012-18.pdf](http://tavistockandportman.nhs.uk/documents/1431/FOI_18-19333_Gender_Dysphoria_Treatment_2012-18.pdf)

Thornhill, J. (2020). *A letter to the editor regarding the original article by Giordano and Holm: Is puberty delaying treatment “experimental treatment”? Manuscript submitted for publication.*

van der Miesen, A. I. R., Steensma, T. D., de Vries, A. L. C., Bos, H., & Popma, A. (2020). Psychological functioning in transgender adolescents before and after gender-affirmative care compared with cisgender general population peers. *Journal of Adolescent Health*. doi:10.1016/j.jadohealth.2019.12.018

Viner, R., Carmichael, P., Di Ceglie, D., Butler, G., Brain, C., Holt, V., ... Skagerberg, E. (2010). *An evaluation of early pubertal suppression in a carefully selected group of adolescents with gender identity disorder*. Research Ethics Committee 10/H0718/62. Retrieved from [http://tavistockandportman.nhs.uk/documents/1845/FOI\\_19-20011\\_GIDS\\_Research\\_Information\\_with\\_attachments.pdf](http://tavistockandportman.nhs.uk/documents/1845/FOI_19-20011_GIDS_Research_Information_with_attachments.pdf)

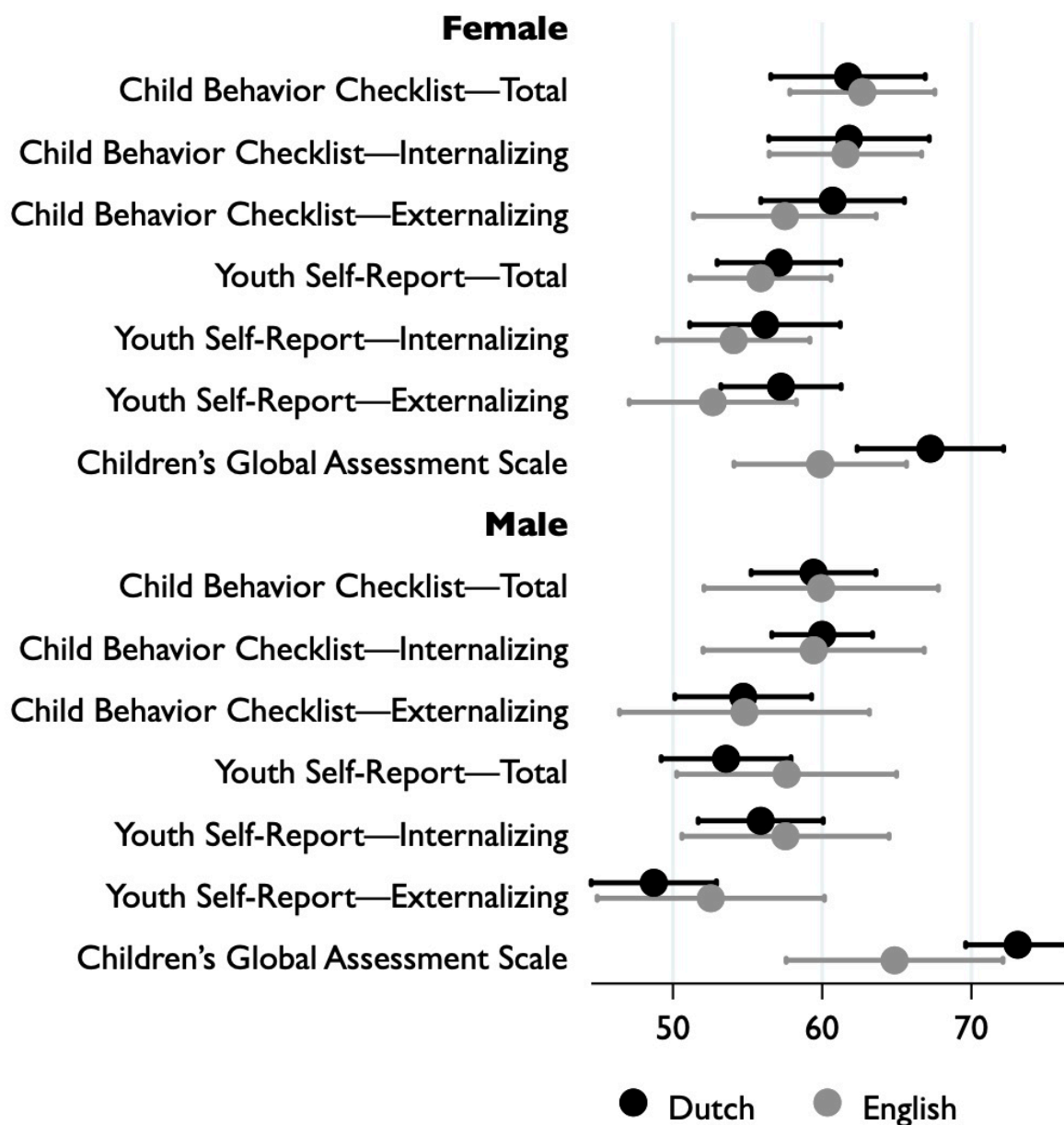
Zucker, K. J. (2019). Adolescents with gender dysphoria: Reflections on some contemporary clinical and research issues. *Archives of Sexual Behavior*, 48, 1983–1992. doi:10.1007/s10508-019-01518-8



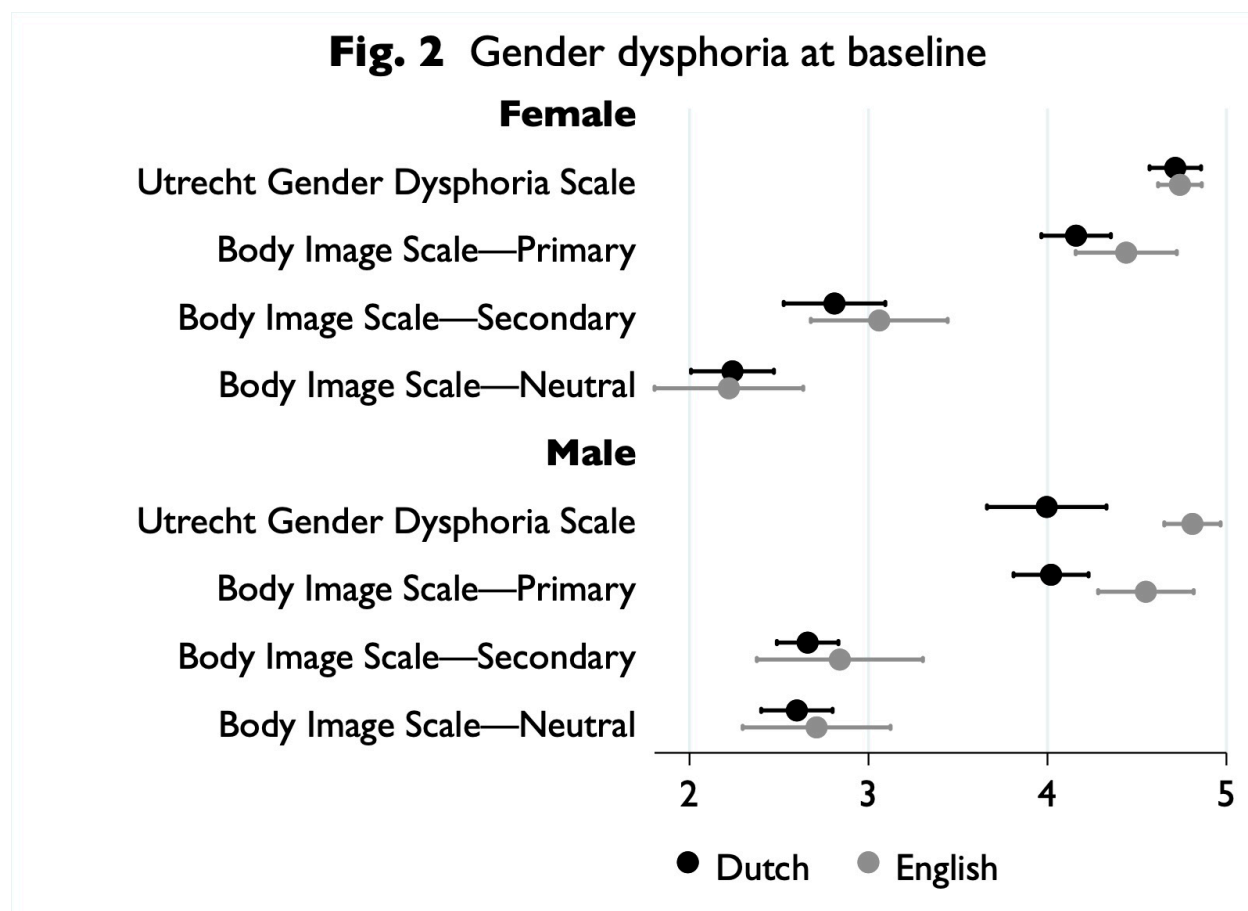
**Table 1** Subjects measured at baseline and after GnRHa treatment

Measure	Dutch total	total	English females	males
Children's Global Assessment Scale	41	26	14	12
Youth Self-Report—Externalizing	54	26	14	12
Youth Self-Report—Internalizing	54	26	14	12
Youth Self-Report—Total	54	26	14	12
Child Behavior Checklist—Externalizing	54	28	15	13
Child Behavior Checklist—Internalizing	54	28	15	13
Child Behavior Checklist—Total	54	28	15	13
Body Image Scale—Neutral	57	27	15	12
Body Image Scale—Secondary	57	27	15	12
Body Image Scale—Primary	57	27	15	12
Utrecht Gender Dysphoria Scale	41	28	15	13

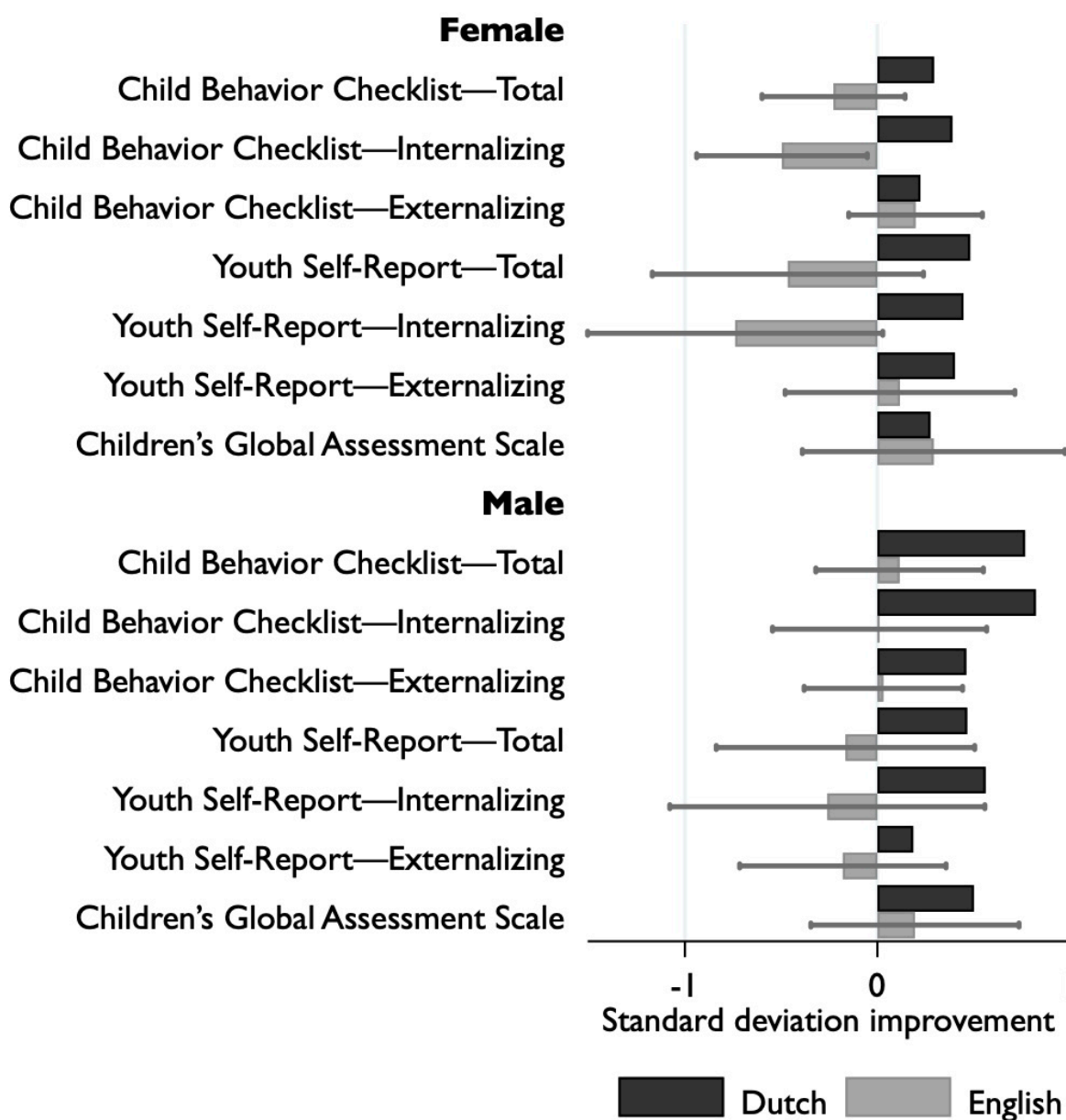
Dutch sample from de Vries et al. (2011), English sample from Gender Identity Development Service (2015).

**Fig. 1** Psychological functioning at baseline

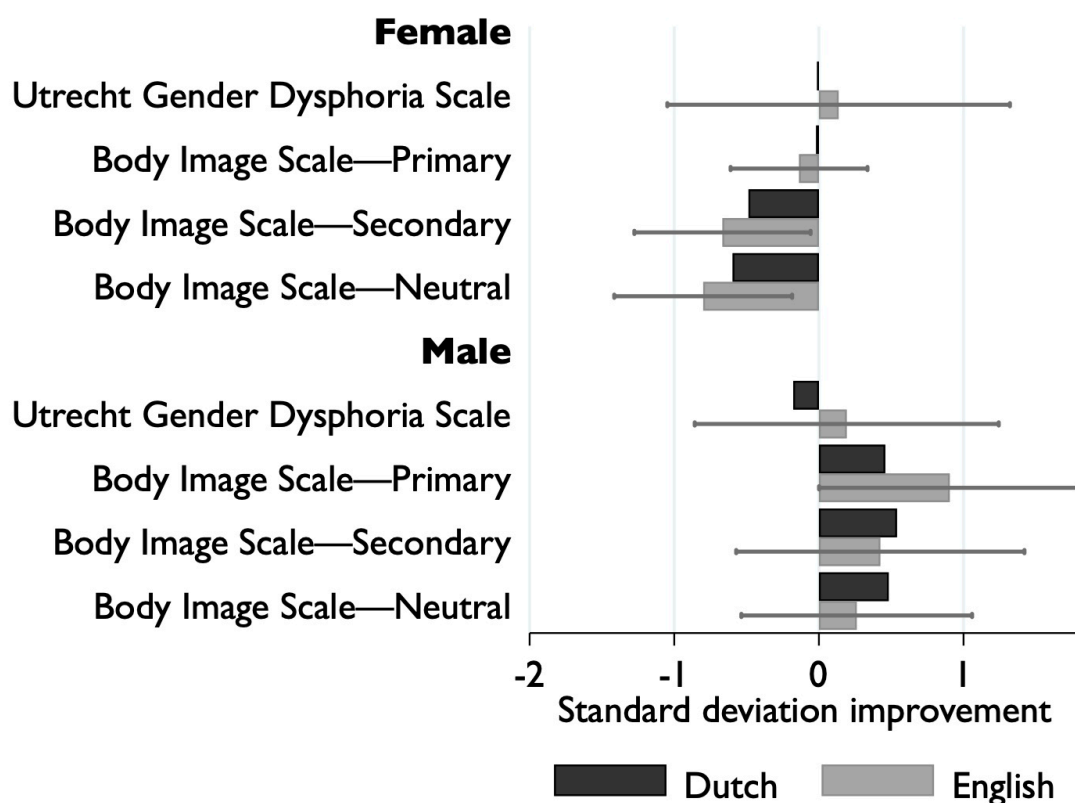
For each measure, differentiated by sex and country, the circle indicates the mean before GnRHa treatment and the line traces the 95% confidence interval.



For each measure, differentiated by sex and country, the circle indicates the mean before GnRHa treatment and the line traces the 95% confidence interval.

**Fig. 3** Change in psychological functioning after treatment

The magnitude of the bar represents change after GnRHa treatment divided by SD at baseline; for the English sample, the line traces the 95% confidence interval. Positive values indicate better functioning.

**Fig. 4** Change in gender dysphoria after treatment

The magnitude of the bar represents change after GnRHa treatment divided by the SD at baseline; for the English sample, the line traces the 95% confidence interval. Positive values indicate lessening dysphoria.

## ONLINE SUPPLEMENT

**Table S1** Measures at baseline ( $t_0$ ) and change after GnRHa treatment ( $t_1-t_0$ )

Measure	Sample	Sex	n	$t_0$ : mean	$t_0$ : SD	$t_1-t_0$ : mean	$t_1-t_0$ : SD
CBCL-T	Dutch	F	29	61.73	13.60	4.00	.
CBCL-T	Dutch	M	33	59.42	11.78	9.04	.
CBCL-T	English	F	15	62.69	8.77	-2.00	5.88
CBCL-T	English	M	13	59.93	12.99	1.50	9.36
CBCL-I	Dutch	F	29	61.80	14.12	5.50	.
CBCL-I	Dutch	M	33	60.00	9.51	7.83	.
CBCL-I	English	F	15	61.56	9.22	-4.57	7.38
CBCL-I	English	M	13	59.43	12.26	0.14	11.29
CBCL-E	Dutch	F	29	60.70	12.64	2.83	.
CBCL-E	Dutch	M	33	54.71	12.91	5.96	.
CBCL-E	English	F	15	57.50	11.04	2.19	6.92
CBCL-E	English	M	13	54.79	13.85	0.43	9.44
YSR-T	Dutch	F	29	57.10	10.87	5.24	.
YSR-T	Dutch	M	33	53.56	12.26	5.72	.
YSR-T	English	F	14	55.87	8.17	-3.80	9.97
YSR-T	English	M	12	57.62	11.59	-1.92	12.25
YSR-I	Dutch	F	29	56.17	13.25	5.93	.
YSR-I	Dutch	M	33	55.88	11.81	6.64	.
YSR-I	English	F	14	54.07	8.84	-6.53	11.74
YSR-I	English	M	12	57.54	10.92	-2.84	14.07
YSR-E	Dutch	F	29	57.24	10.59	4.27	.
YSR-E	Dutch	M	33	48.72	11.83	2.20	.
YSR-E	English	F	14	52.67	9.70	1.14	10.03
YSR-E	English	M	12	52.54	11.98	-2.15	10.11
CGAS	Dutch	F	22	67.25	11.06	3.05	.
CGAS	Dutch	M	25	73.10	8.44	4.23	.
CGAS	English	F	14	59.87	10.01	2.93	11.83
CGAS	English	M	12	64.85	11.43	2.23	9.73
UGDS	Dutch	F	22	4.71	0.32	-0.00	.
UGDS	Dutch	M	25	4.00	0.81	-0.14	.
UGDS	English	F	15	4.74	0.22	0.03	0.47
UGDS	English	M	13	4.81	0.26	0.05	0.45
BIS-I	Dutch	F	30	4.16	0.52	-0.01	.
BIS-I	Dutch	M	35	4.02	0.61	0.28	.
BIS-I	English	F	15	4.44	0.51	-0.07	0.44
BIS-I	English	M	12	4.55	0.42	0.38	0.60
BIS-2	Dutch	F	30	2.81	0.76	-0.37	.
BIS-2	Dutch	M	35	2.66	0.50	0.27	.
BIS-2	English	F	15	3.06	0.69	-0.46	0.76
BIS-2	English	M	12	2.84	0.73	0.31	1.14
BIS-0	Dutch	F	30	2.24	0.62	-0.37	.
BIS-0	Dutch	M	35	2.60	0.58	0.28	.
BIS-0	English	F	15	2.22	0.75	-0.60	0.83
BIS-0	English	M	12	2.71	0.65	0.17	0.82

For the Dutch sample, n for each sex is estimated from the proportion in the total sample of 70.