



January 11, 2023

The Dutch Studies and The Myth of Reliable Research in Pediatric Gender Medicine

A new study identifying flaws in gender medicine research demands urgent attention from the medical community

A new open-access publication, "[The Myth of Reliable Research in Pediatric Gender Medicine](#)," focuses on the two [Dutch studies](#) that gave rise to "gender-affirmative" care for youth worldwide. The authors convincingly demonstrate that rather than "solid prospective research" or even the "gold standard" in research, as these studies are frequently described by the proponents of "gender-affirmative care," the Dutch research suffers from profound, previously unrecognized problems. These problems range from **erroneously concluding that gender dysphoria disappeared** as a result of "gender-affirmative treatment," to **reporting only the best-case scenario outcomes** and **failing to properly examine the risks**, despite the fact that a significant proportion of the treated sample experienced adverse effects.

The authors note that the Dutch studies, while of unacceptably low quality by today's standards, were commensurate with the clinical and research practices during the era of expert opinion-led medicine widely practiced before the 1990s. The term "evidence-based medicine" and its focus on quality comparative clinical research to determine optimal treatment only emerged in the 1990s. The Dutch researchers began to medically transition gender dysphoric adolescents in the late 1980s and early 1990s—just as medicine was starting to undergo this major paradigm shift.

The authors assert that had the Dutch studies been published today for the first time, the "innovative practice" of using hormones and surgery to gender-transition children and young adults would never have been permitted to enter general medical settings due to the very low quality of the research, and problematic outcomes experienced by several of the young people. Unfortunately, since the publication of the final Dutch study in [2014](#), the practice of youth gender transitions underwent what's known as "runaway diffusion"—a problematic but not uncommon phenomenon whereby the medical community mistakes a small innovative experiment as a proven practice, and a potentially nonbeneficial or harmful practice "escapes the lab," rapidly spreading to general practice settings.

The authors note that the only way to curb the damage of ongoing "runaway diffusion" is to conduct systematic reviews of evidence, update treatment guidelines to reflect the lack of evidence, and then "de-implement" unproven or harmful practices—a process known as "practice reversal." They observe that such practice reversals of "gender-affirming" interventions for youth are already underway in [Finland](#), [Sweden](#), [England](#), and most recently the state of Florida.

The authors also address the plethora of new studies since the Dutch research. They note that these newer studies, which purport to have definitively proven that puberty blockers and cross sex hormones are "as benign as aspirin, as well-studied as penicillin and statins, and as essential to survival as insulin for childhood diabetes" are even more flawed

than the original Dutch research. To demonstrate the lack of rigor in more recent research, the authors focus on three recent studies, including the most [recent study by Tordoff et al.](#) That study made [headlines](#) for purportedly demonstrating that following 12 months of “treatment” with puberty blockers and cross-sex hormones, depression in gender-dysphoric youth all but disappeared. The authors demonstrate that this finding is both implausible and indefensible. They conclude that far from having perfected previously flawed research methods, these newer studies continue to suffer from serious problems but that researchers have indeed perfected the art of spin—misrepresenting weak, uncertain, or even negative findings as strong and positive.

The authors note that the question, “Just because we can, should we?” is not unique to pediatric gender medicine. What makes this area exceptional is the radical, irreversible nature of “gender-affirming” medical and surgical interventions desired by the exponentially growing numbers of youth in the Western world. New changes to the ICD-11 “gender incongruence” diagnosis, endorsed by WPATH “Standards of Care 8” as the diagnostic category that should guide treatment decisions, have resulted in a situation whereby virtually any body modification desired by a young person, including unlimited possibilities for “nonbinary” procedures, can be considered “medically necessary.” They also note the gender-clinicians’ relentless political activism. Rather than engaging in reflection and improving research methods, many gender-clinicians-turned-advocates are instead trying to [quell the ongoing scientific debate](#) by calling it “science denialism” motivated by ignorance, religious zeal, and transphobia.

The authors observe, “The key problem in pediatric gender medicine is not the lack of research rigor in the past—it is the field’s *present-day* denial of the profound problems in the existing research, and an unwillingness to engage in high quality research requisite in evidence-based medicine.” They caution that harm at scale will ensue if the gender medicine establishment does not acknowledge the problems with the research it relies on and continues to produce studies favoring spin rather than rigor. The authors remind readers that medicine is a double-edged sword with potential to help but also to harm, and that the history of medicine is replete with examples of “cures” that turned out to be far worse than the “diseases” they purported to treat.

[This new publication](#) represents an inflection point in the field of pediatric gender medicine. It is requisite reading for anyone following the ongoing scientific debate in the field of youth gender medicine. For those looking for highlights, below is a summary of the key points:

The Dutch Studies of Youth Gender Transitions are Deeply Flawed

Critics and [proponents](#) of youth gender transition agree that to date, the Dutch studies represent the best available evidence for pediatric gender transition, and indeed the entire model of “gender-affirming” care is based on the Dutch experience. The seminal importance of the Dutch studies is evidenced by the fact that the [Endocrine Society Guidelines](#), and WPATH “Standards of Care 7” under which the practice proliferated, refer only to the Dutch experience as proof of “benefits” of the practice.

The Dutch studies found that puberty blockers, cross-sex hormones, and surgery effectively transformed young bodies, leaving patients satisfied with their appearance shortly after the final surgery. However, the studies failed to show that these physical changes resulted in any clinically meaningful or reliable psychological benefits sufficient to justify the serious adverse effects that emerged.

Besides the small sample size, lack of control group, and only short-term follow-up—well-known issues with the Dutch research—the authors identify several serious methodological problems that have been overlooked.

1. Only best-case scenario outcomes were included in the studies’ results

The Dutch studies reported only the best-case scenarios at each stage of treatment: puberty blockers, cross-sex-hormones, and surgery. Those who did not fare as well, or experienced problems, were not included in the research results.

The [first](#) of the two Dutch studies, which focused specifically on the effects of puberty blockers, selected its 70 cases from a larger pool of treated cases focusing only those who successfully completed puberty blockade and were ready to start cross-sex hormones. The authors argue: "Using the start date of the *next phase* of treatment (cross-sex hormones) as the defining inclusion criterion for the study of the *prior phase* of the treatment (puberty blockers) introduced serious bias." Those who developed problems while on puberty blockers and / or discontinued them would not have made it into the "first 70" cases. This methodology also assured that the cases are most physically and mentally mature, as it favored older subjects.

The [second and final study](#) suffers from the same problem, but it is more visible, since the subjects were followed from the first into the second study. The 70 cases dropped down to 55. Those who developed medical problems (including 3 cases obesity and diabetes and 1 death) were reclassified as "nonparticipants," eliminating their negative outcomes from the studies' final results.

The Dutch studies' unusual subject recruitment and retention methods disqualify the research from the "quality longitudinal prospective cohort study" designation. Rather, the Dutch studies are best described as a "case series"—the lowest and least reliable level of evidence.

2. The conclusion that gender dysphoria disappeared after "affirmative" treatment is wrong

The linchpin finding of the final 2014 Dutch study (and only clinically significant improvement) was the reported "disappearance" of gender dysphoria post-surgery. However, this finding is compromised by misuse of the UGDS "gender dysphoria" measurement scale. Not only did the wording of the gender dysphoria scale questions switch from female to male (and vice versa) before and after treatment, but the entire scoring mechanism was effectively reversed. As a result, a gender-incongruent patient answering the same question in the same way would register an instant "drop" of "gender dysphoria scores" as soon as the scale version was switched—independent of any treatment. They explain:

The following hypothetical scenario clearly demonstrates the problem. A severely gender dysphoric, cross-sex identified female patient is asked to answer two of the UGDS questions: "Every time someone treats me like a girl I feel hurt" and "Every time someone treats me like a boy I feel hurt" (Items 2 on the "female" and the "male" versions of the UGDS scale, respectively). It is likely that the patient would strongly agree with the first statement, and strongly disagree with the second. The first answer would lead to the score of "5" on the UGDS gender dysphoria scale, indicating the highest possible level of gender dysphoria. The second answer—which is effectively the same answer—would result in the score of "1" indicating the lowest possible gender dysphoria. This is because unlike the first question, which belongs to the "female" battery of questions, the second question belongs to the "male" battery of questions and effectively assumes the subject to be male—hence, the lack of distress of being associated with "maleness" receives the minimum "gender dysphoria" score.

If we now consider that only the "female" scale was used for gender dysphoric females at baseline but was then switched to the "male" scale after the final surgery (and vice-versa for male subjects), it becomes clear that the remarkable drop in "gender dysphoria" the UGDS scale registered after surgery entirely results from switching the scale.

The authors note that it is vitally important to recognize that inversion of the UGDS scale after gender reassignment was not merely “not ideal” as lead author, Dr. de Vries acknowledged—but that it entirely invalidated the final Dutch study’s main finding. They explain:

When defending the choice to reverse the UGDS scale (de Vries, 2022), de Vries pointed out—and we agree—that it would make no sense to ask postoperative natal males to answer a question such as “I dislike having erections” (Table 1, UGDS-M, item 11), since they no longer have penises. We empathize with the Dutch researchers’ plight, as they found themselves without a valid tool to measure the construct of “gender dysphoria” after treatment. It is equally nonsensical, however, to ask natal males to answer questions such as, “I hate menstruating because it makes me feel like a girl” (Table 1, UGDS-F, item 10)—and it makes even less sense to report “resolution of gender dysphoria” because they don’t “hate menstruating.”

The problem with the claim of “disappearance of gender dysphoria” which relies on reversing the scale was evident in the Dutch team’s presentation at the WPATH Symposium in late 2022. Presenting long-term follow-up data on male-to-female transitioners, the Dutch team acknowledged that while subjects continued to score low on “gender dysphoria,” over half reported genital shame, and nearly a quarter continued to feel inadequate about their bodies, which they still deemed too masculine.

3. Psychotherapy was offered to all subjects making it impossible to determine which intervention “worked”

Contrary to the notion that psychotherapy to explore the causes of gender dysphoria or to ameliorate it noninvasively is akin to “conversion therapy, the Dutch protocol required ongoing psychotherapy and the studies appropriately acknowledged that psychotherapy may have contributed to the high postoperative functioning of the treated youth. However, commingling psychotherapy with hormones and surgery made it impossible to determine whether medical and surgical interventions, psychotherapy, or simply the passage of time allowed these already high-functioning youth to retain their high function 1.5 years after surgery.

4. The studies only focused on potential benefits and failed to evaluate risks

Despite their longstanding hypothesis that hormones impair bone and brain development, the Dutch researchers did not evaluate these risks. The failure to consider all relevant outcomes—including physical health effects—is a key methodological flaw in the Dutch research. Even without setting out to evaluate physical health risks, the Dutch studies revealed that up to 7% of subjects suffered severe morbidity and mortality: several subjects developed diabetes and obesity during treatment, one died, and 100% became sterile since the protocol required removal of reproductive organs.

Many [risks](#) and [uncertainties associated](#) with puberty blockers and cross-sex hormones have emerged in more recent research. Long-term follow-up data presented at the WPATH Symposium suggest that reproductive regret and significant problems with sexual function affected a significant proportion of those transitioned in the Dutch clinic as adolescents. The authors observe: “Patients and their families cannot make informed decisions about a treatment when the physical health risks are assumed to be minimal and not reported, and only the potential psychological benefits are considered.”

5. The results are not generalizable to most currently presenting cases of gender-dysphoric youth

Currently, most youth with gender dysphoria suffer from post-pubertal onset of gender dysphoria, significant mental illness, and increasingly present with non-binary gender identities—clinical presentations the Dutch explicitly disqualified from their studies. Today, the Dutch protocol is administered to the very youth who would have been

disqualified from this foundational research. This is alarming since up to [1 in 10-20 youth](#) in the US (with similar trends in the rest of the Western world) currently claim a transgender identity, placing them at risk for treatment with a protocol explicitly contraindicated for them.

6. Conducting research with the preconceived notion of the superiority of medical gender reassignment

Equipose, a key concept in clinical research, requires that investigators maintain a genuine sense of uncertainty and curiosity about which interventions are most effective. This allows design of studies that do not favor one intervention over another but instead generate reliable evidence about which intervention is most likely to provide benefits while minimizing risks. The Dutch researchers started with the preconceived but untested notion that the only way to help youth with early childhood onset of gender dysphoria that worsened in adolescence was to transition them using hormones and surgery.

The danger and inaccuracy of assuming that nothing but hormones and surgery “works” is highlighted by findings from an earlier, [often overlooked study](#) from the same Dutch gender clinic that reported the outcomes of gender-dysphoric youth who were rejected from gender reassignment in adolescence. Remarkably, 80% of these youth were no longer interested in gender transition as adults and found other ways to address their gender dysphoria. This calls into question the assumption that gender dysphoric adolescents will invariably want to live as transgender-identified adults.

7. The availability of the Dutch protocol may have created the crisis of growing demand for youth gender reassignment in the absence of objective medical necessity criteria or quality research

The authors point to an [earlier study of Dutch youth](#) conducted before the advent of the Dutch protocol. This prospective population study of 879 Dutch children reported that 6% of the young children were considered “gender variant,” but 24 years later none had chosen to undergo medical transition despite being eligible for gender reassignment as mature adults.

These “gender-variant” children were coming of age just a few years before the Dutch protocol was created, and before the Dutch clinicians devised the notion of “juvenile transsexuals”—which today we term “transgender youth.” Thus, these children were able to experience identity development without the pressure of labels and importantly, without puberty blockers and cross-sex hormones.

Proponents of treating youth with hormones and surgery assert that these interventions are essential for “those who need it.” The authors pose two key questions: “First, has the availability of the Dutch protocol create this “need?” Second, absent clear criteria to separate a young person’s “wish” from a “need,” will rigorous research be required to demonstrate that the benefits outweigh the risks?”

8. It is time to conduct a re-evaluation of the Dutch experience using rigorous research methodologies

In view of the thousands of referrals to the Amsterdam gender clinic since the practice of gender-transition of youth was launched, the authors urge the Dutch research team to use quality methodologies to re-analyze the quarter-century of outcomes of youth gender reassignment. The authors issue specific recommendations about how to begin: by accessing health registries and reviewing the records of all patients diagnosed with gender dysphoria, regardless of the treatment chosen. They encourage the Dutch to design studies that compare various subgroups, and to pay attention to objective long-term mental health and physical outcomes, analyzing them separately for each type of intervention and by sex at birth.

Importantly, the authors note that the primary outcome of youth gender reassignment must be clearly articulated and justified. At first, the Dutch researchers stated that the goal of transition was to improve psychological functioning. However, when studies failed to show convincing benefits of improved psychological functioning, the goal was redefined; the Dutch researchers most recent assertion is that psychological outcomes are not the correct outcomes to measure, the preferred outcomes are quality of life and satisfaction with treatment. Outcomes should not be changed based on what the researchers see “works” after the treatment, because such methodology introduces a serious risk of finding an association by chance. Rather, the outcomes should be clearly articulated in advance and consistently tracked and reported for all “intent-to-treat” subjects, regardless of whether their conditions improve, worsen, or are unchanged.

The Research that Followed the Dutch Studies is Even More Flawed

Newer studies purportedly proving the benefits of gender transition suffer from even more significant limitations than do the Dutch studies. The authors illustrate this deficit by analyzing three recent studies that have been widely cited by proponents of youth gender reassignment as proof that early gender transition is beneficial, illustrating how the studies’ weak or even negative findings are spun into favorable results.

A striking recent example is the [study by Tordoff et al.](#), which has been touted as proof that the benefits of gender reassignment rapidly appear during the first year of treatment and lead to dramatic reductions in depression. In fact, the rates of depression in the treated group *remained unchanged before and after treatment*. The optimistic conclusion of benefits came from the fact that the *untreated group worsened*, and this finding was incorrectly perceived as a signal that treatments must have helped the “treated” group remain stable and avoid deterioration. However, Tordoff et al. failed to notice that they lost a whopping 80% of “untreated” participants by the end of the study, and that this dropout rate entirely invalidated the study’s methodology and conclusions:

However, by basing their conclusion about the relative success of the “treated” on the finding of lack of success among the “untreated” cases, the researchers failed to consider that they lost an astounding 80% of their “untreated” cohort by the end of the study (28 of 35); in contrast, over 80% of the “treated” cohort (57 of 69) remained enrolled. The high dropout rate in “untreated” subjects makes intuitive sense: the study took place in a gender clinic setting, the primary purpose of which is provision of gender transition services. Youth whose distress was ameliorated without the use of hormones would have little reason to stay enrolled in the clinic and participate in the ongoing research. However, what this also suggests is that the highest functioning “untreated” youth dropped out of the study. Thus, the entire conclusion that because “untreated” cases fared so poorly on measures of depression, anxiety, or suicidality, it must be that hormones given to the “treated” cases “worked,” is invalid. There are other problems in the study, including the fact that the use of psychiatric medications was not accounted for in the analysis. The university was aware of the problems with this research but chose to remain silent because the study’s optimistic conclusions were so well received by national news media outlets (Rantz, [2022](#)).

Much research conducted by gender clinics suffers from similar problems, and few frontline clinicians responsible for recommending treatments have the time and skill to critically evaluate methodologies of these deeply flawed studies, which somehow make it past peer-reviewers in even the most prestigious journals.

SEGM Thoughts

[“The Myth of Reliable Research in Pediatric Gender Medicine”](#) mounts a formidable challenge to the claim that the Dutch studies’ conclusion that psychosocial benefits arise from “solid prospective research.” The article demands urgent attention from the medical community. It not only challenges the notion that the rapidly growing numbers of gender

dysphoric youths can be safely and effectively treated with hormones and surgery but also questions whether even those with childhood-onset are best served with hormones and surgery.

The authors highlight how far the field of gender-medicine has drifted from the core principles of evidence-based medicine. They advocate enlisting methodology experts without intellectual or other conflicts of interest in any new research, rather than continuing to rely on research from online surveys or the gender clinics themselves, which, fighting for the survival of their practices, have increasingly assumed a politicized and activist stance.

Insistence by the gender medicine community that questioning science and demanding rigor in research is akin to “science denialism” or “transphobia” places the field at risk of becoming one of the biggest medical epidemics of harm. As the authors point out, currently as many as 1 in 10-20 high school and college youth in the U.S. self-identify as transgender. This self-identification, “affirmed” by healthcare professionals, puts them at risk for what has been demonstrated to be a largely irreversible “gender-affirming” medical treatment pathway which results in irreversible changes and is expected to lead to infertility or sterility. Thousands of harmed young detransitioners are already [speaking out](#), and recent studies indicate [medical detransition rates of up to 30%](#). Not all of these individuals identify as “detransitioners” and not all will conceptualize their experiences as “regret,” [but many have](#), and many more likely will. Prior research suggests there is roughly an [8-11](#) year gap between gender reassignment and the emergence of long term regret and markedly increased morbidity and mortality.

Medical societies and scientific journals that suppress debate in gender medicine, while uncritically promoting the fallacious “settled science” narrative by publishing deeply flawed studies, are contributing to this crisis. The gender medicine field has a limited time to self-correct before public health authorities, and increasingly, elected officials who do not understand medicine but do understand the risks of harm to youth, step in to curb the damage.