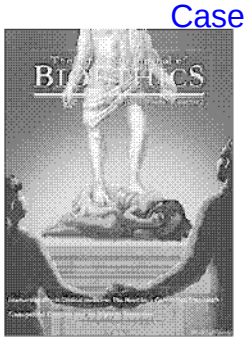


Taylor & Francis
Taylor & Francis Group

The American Journal of Bioethics

ISSN: 1526-5161 (Print) 1536-0075 (Online) Journal homepage: https://www.tandfonline.com/loi/uajb20

The Right to Best Care for Children Does Not Include the Right to Medical Transition

Michael Laidlaw, Michelle Cretella & Kevin Donovan

To cite this article: Michael Laidlaw, Michelle Cretella & Kevin Donovan (2019) The Right to Best Care for Children Does Not Include the Right to Medical Transition, The American Journal of Bioethics, 19:2, 75-77, DOI: 10.1080/15265161.2018.1557288

To link to this article: https://doi.org/10.1080/15265161.2018.1557288

Pl. Trial Ex. 144

The Right to Best Care for Children Does Not Include the Right to Medical Transition

Michael Laidlaw, Private Practice
Michelle Cretella, American College of Pediatricians
G. Kevin Donovan, Georgetown University Medical Center

Contrary to the suggestion in the article, by Priest (2019) watchful waiting with support for gender-dysphoric children and adolescents up to the age of 16 years is the current standard of care worldwide, not gender affirmative therapy (de Vries and Cohen-Kettenis 2012). The treatment of children and adolescents with sex hormones for a mismatch between their mind's perceived gender and their biological sex—otherwise known as gender dysphoria (GD)—results in unique medical and ethical challenges not present in adults. The main challenge results from stopping normal pubertal development by the use of puberty blocking agents (PBA). These are given as part of a treatment paradigm known as gender affirmative therapy (GAT). After some period of time on PBA, cross-sex hormones are introduced and dosages are increased, and gonads and breasts may be surgically removed. The consequences of PBA/GAT are not trivial and include potential sterility, sexual dysfunction, thromboembolic and cardiovascular disease, osteoporosis, and malignancy (Hembree et al. 2017). The adolescent does not have the intellectual or emotional maturity or judgment to make the decision to undergo PBA/GAT without parental approval.

PUBERTY BLOCKING AGENTS

Gonadotropin-releasing hormone (GnRH) analogues such as leuprolide are being used off label for suppression of normal puberty. These powerful hormones act directly on the pituitary–gonadal axis to inhibit the production of luteinizing hormone (LH)/follicle-stimulating hormone (FSH) and consequently lower testosterone and estrogen to subnormal levels. This pathologic state is known as hypogonadotropic hypogonadism. There are no randomized controlled studies for the use of PBA—including safety—for stopping normal puberty. The Endocrine Society has published revised clinical guidelines in 2017 on the treatment of gender dysphoric persons including adolescents (Hembree et al. 2017). The quality of evidence for PBA is noted to be low. In fact, all of the evidence in the guidelines with regard to treating children/adolescents by GAT is low to very low because of the absence of proper studies.

These same guidelines, however, recommend arresting normal puberty at Tanner stage 2. This is highly significant, because it is the pubertal stage occurring before menarche in girls and before spermarche in boys. Continued suppression of the pituitary gonadal axis by PBA will maintain a state of immaturity of the male and female gonads. As a result, though the child will likely continue to grow in stature, the gonads and entire pelvic genitalia will remain stunted at Tanner stage 2. The addition of cross sex hormones will not change this condition. As a result, the patient will be infertile as an adult. The continued administration of cross-sex hormones may lead to permanent sterility. Gonadectomy, of course, will ensure sterility.

INFERTILITY

Infertility is a major health problem. The literature shows it to be a source of significant psychological distress that reduces quality of life, including "emotional well being, relationships, and sexuality" (Carter et al. 2005). Infertile women even without other health conditions have "depression at twice the rate of the normal population and report levels of psychological distress comparable to the experience of a life threatening illness" (Carter et al. 2005). This is obviously a complex problem that a child, particularly at Tanner stage 2, which may be as young as 8 or 9 years old, does not have the capacity to fully comprehend.

LOW RATES OF FERTILITY PRESERVATION

What about fertility preservation (FP)? An obvious parallel can be found in child and adolescent cancer patients who may lose fertility because of gonadotoxic treatments and surgeries. "From an ethical standpoint, the key

Address correspondence to Michael Laidlaw, Endocrinology, Private Practice, 4770 Rocklin Rd, Ste #1, Rocklin, CA 95677, USA. E-mail: docdrlaidlaw@gmail.com

reason for pursuing fertility protection [in these patients] is to restore personal autonomy to those who might, in the future, become unable to conceive" (Patrizio and Caplan 2010). Unfortunately, studies show that fewer than 5% of adolescents receiving GAT even attempt this process (Chen et al. 2017). Those adolescents who receive PBA at a late stage of puberty may be able to preserve fertility by sperm or oocyte cryopreservation. However, ovarian tissue cryopreservation is still considered experimental in most centers and testicular tissue cryopreservation remains entirely experimental. These experimental forms of FP would be the only options in children blocked prior to spermarche and menarche and are high in cost and limited to specialized centers. Even with FP there is no guarantee of having a child.

IMPAIRED SEXUAL FUNCTION

Another consequence of PBA is impairment of adult sexual function (SF). Again, early blockade of puberty will arrest the genitalia at an early developmental stage with limited to absent functioning as an adult. Normal SF in the male, including erection, orgasm, and ejaculation, will be significantly impaired to absent. In the female, PBA induce a menopausal state with consequent low estradiol and progesterone levels. The pelvic genitalia will acquire a menopausal yet undeveloped state. Consequences will include decreased blood flow to the vagina and vulva, thinning of the vaginal epithelium, and vaginal dryness leading to atrophy. Conditions such as hypoactive sexual desire (HSDD) have been found in naturally menopausal females and to be of even higher prevalence in 20- to 49-year-olds with surgically induced menopause. HSDD is associated with a significant decrease in general health, including aspects of mental and physical health (Leiblum et al. 2006).

Both the male and female will have significantly impaired to absent SF as adults because of the effects of PBA and due to the addition of cross-sex hormones. These effects will be compounded by genital affirmation surgery. The child or adolescent lacks the knowledge, foresight, and in most cases experience of SF to be able to fully comprehend the loss or impairment of SF resulting from the initial step of PBA.

DISRUPTION OF NORMAL BONE DEVELOPMENT

Abnormal bone growth and development is another unique consequence of PBA. Skeletal growth and development is an exquisitely timed and coordinated event involving multiple hormonal interactions, particularly the simultaneous increase in sex hormones and growth hormone. PBA results in very low to absent sex hormone levels during the normal time of increased growth hormone levels. The results and implications of causing this hormonal mismatch are unknown. It has been shown that PBA lead to a decline in the bone mineral density z score (Hembree et al. 2017). This has lifelong implications that may include early-onset osteopenia or osteoporosis.

THE PROBLEM OF ACCURATE DIAGNOSIS AND DESISTANCE

In spite of the previous discussion of risks, one might argue that the benefits of PBA outweigh these risks, and that children and adolescents should have the legal right to access these agents without parental approval. But as in any disease process, how can we be certain that a condition has been properly diagnosed before undergoing life-altering therapies? In the case of cancer, for example, physical evidence and confirmation of the condition would be warranted to both justify and tailor therapy accordingly.

There are no objective tests—laboratory, imaging, or otherwise—to diagnose a "true transgender" child. By adulthood 61–98% will outgrow GD (Ristori and Steensma 2016). There is no known way to predict who will desist and who will remain dysphoric. Furthermore, the recent phenomenon has been described of adolescents, primarily teenage girls, many with neurodevelopmental disorders, suddenly developing GD without prior history, through social contagion (Littman 2018). Therefore, a large number of children and adolescents diagnosed with GD who would otherwise be predicted to desist or who have developed the condition due to social pressures would be irreversibly harmed by PBA/GAT.

PBA compound the problem of natural desistance in a profound way. In a study of 70 adolescents who were followed after receiving PBA, 100% desired to continue on to cross-sex hormones (de Vries et al. 2011). The natural pattern of desistance has been broken, either because of physiologic or psychological effects of PBA or a combination of the two. Adolescents who would have otherwise desisted instead retain the notion that their gender identity and sex do not match and are thereby guaranteed to risk developing the problems addressed in the preceding.

COMORBID PSYCHIATRIC CONDITIONS

An additional issue related to the adolescent's decision to take PBA without parental involvement concerns associated psychological conditions that may affect judgement. A study of the Finnish gender identity service showed that "75% of adolescents [assessed] had been or were currently undergoing child and adolescent psychiatric treatment for reasons other than gender dysphoria" (Kaltiala-Heino et al. 2015). A recent parental survey showed that 62.5% of the affected adolescents had "a psychiatric disorder or neurodevelopmental disability preceding the onset of gender dysphoria" (Littman 2018). These psychiatric conditions compound the problem of immaturity in contemplating the very complex issues that we have addressed.

THE CURRENT STANDARD OF CARE

As stated, watchful waiting with support for GD children and adolescents is the current standard of care worldwide until the age of 16 years, not GAT (de Vries and Cohen-Kettenis 2012). Children referred for psychological therapy or simple watchful waiting have been able to alleviate their GD without the damaging health consequences of GAT. These methods are the obvious and preferred therapy for GD, as they do the least harm with the most benefit for the greatest number.

CONCLUSION

Children and adolescents have neither the cognitive nor the emotional maturity to comprehend the consequences of receiving a treatment for which the end result is sterility and organs devoid of sexual pleasure function. To argue that all children who are self-declared as transgendered will be harmed psychologically and physically without puberty blocking treatments is false; the greatest number will be seen to not require this at all. To further argue that these adolescents should receive hormonal therapy without parental approval betrays a poor understanding of adolescent psychology and the role of parents in the family dynamic. Evidence of severe and permanent harm from an appropriate delay for the psychological evaluation and treatment of such children, prior to permanently altering them, does not exist. To argue that such supposed harm rises to the level of denying parental involvement in the care of their gender-dysphoric child is grossly overreaching, and should not be suggested as the standard of care. Rather, it would constitute an unmonitored, experimental intervention in children without sufficient evidence of efficacy or safety, for which informed consent therefore would not be possible. ■

ORCID

Michael Laidlaw http://orcid.org/0000-0001-6849-7285
Michelle Cretella https://orcid.org/0000-0002-4264-9911
G. Kevin Donovan https://orcid.org/0000-0002-6115-7840

REFERENCES

Carter, J., K. Rowland, D. Chi, et al. 2005. Gynecologic cancer treatment and the impact of cancer-related infertility. *Gynecologic Oncology* 97(1): 90–95.

Chen, D., L. Simons, E. K. Johnson, B. A. Lockart, and C. Finlayson. 2017. Fertility preservation for transgender adolescents. *The Journal of Adolescent Health* 61(1): 120–123.

de Vries, A. L. C., T. D. Steensma, T. A. H. Doreleijers, and P. T. Cohen-Kettenis. 2011. Puberty suppression in adolescents with gender identity disorder: A prospective follow-up study. *The Journal of Sexual Medicine* 8(8): 2276–2283. doi: 10.1111/j.1743-6109.2010.01943.x.

de Vries, A. L., and P. T. Cohen-Kettenis. 2012. Clinical management of gender dysphoria in children and adolescents: The Dutch approach. *Journal of Homosexuality* 59(3): 301–320.

Hembree, W. C., P. T. Cohen-Kettenis, L. Gooren, et al. 2017. Endocrine treatment of gender-dysphoric/gender-incongruent persons: An endocrine society clinical practice guideline. *The Journal of Clinical Endocrinology & Metabolism* 102(11): 3869–3903. doi: 10.1210/jc.2017-01658.

Kaltiala-Heino, R., M. Sumia, M. Työläjärvi, and N. Lindberg. 2015. Two years of gender identity service for minors: Overrepresentation of natal girls with severe problems in adolescent development. *Child and Adolescent Psychiatry and Mental Health* 9: 9.

Leiblum, S. R., P. E. Koochaki, C. A. Rodenberg, I. P. Barton, and R. C. Rosen. 2006. Hypoactive sexual desire disorder in postmenopausal women: US results from the women's international study of health and sexuality (WISHeS)). *Menopause* 13(1): 46–56.

Littman, L. 2018. Rapid-onset gender dysphoria in adolescents and young adults: A study of parental reports. *PLoS One* 13(8): e020233016.

Patrizio, P., and A. Caplan. 2010. Ethical issues surrounding fertility preservation in cancer patients. *Clinical Obstetrics and Gynecology* 53(4): 717–726.

Priest, M. 2019. Transgender children and the right to transition: Medical ethics when parents mean well but cause harm. *American Journal of Bioethics* 19(2): 45–59.

Ristori, J., and T. D. Steensma. 2016. Gender dysphoria in childhood. *International Review of Psychiatry (Abingdon, England)* 28(1): 13–20.

Commentary

CATHOLIC MEDICAL ASSOCIATION

The Linacre Quarterly
2020, Vol. 87(2) 131-137

DOI: 10.1177/0024363919875383
journals.sagepub.com/home/lqr
SAGE

Strangers in a Strange Land: How Our Founding Principles and a Bitter Pill Undo the Assimilation of US Catholics

G. Kevin Donovan, MD, MA[1], and Claudia Sotomayor, MD, DBe[1]

Abstract

Most Catholic physicians work with the comfortable assumption that we can practice our profession and our faith, fully assimilated into modern American culture and society. Increasingly, we have come to realize that to be a Catholic Christian is by nature to be countercultural. American culture, ordered by the founding fathers in concepts of liberty and freedom, has been profoundly affected by the introduction and reliance on a contraceptive pill. This has changed the mores and sexual behaviors of society in ways that are antithetical to Catholic values. The consequences of contraception have directly led to an acceptance of a broad number of behaviors and attitudes that society insists must be tolerated. This challenges the commitments of Catholic physicians both personally and professionally.

Keywords

Bioethics, Catholic, Contraception, Happiness, Liberalism, Morality, Physician, Utilitarianism

Despite historical opposition to Catholic emigration in America, the past two or three generations have seen an adjustment of societal attitudes and greater acceptance of Catholics in all walks of life including politics, business, and the professions. At times, we seemed to have achieved a near seamless integration into modern American culture and society. We have grown increasingly comfortable, adopting American principles and traditions as our own. However, in the recent past, cultural flash points have arisen, many of which seem to be a divisive challenge for faithful Catholics in particular. Specifically, we have encountered problems in the practice of medicine more attributable to our role of being both a Catholic and a physician, problems that can make us feel like we are again strangers in a strange land. Why should this be true—why should we not be totally at ease in our society? Perhaps the seeds of our discontent, our present disassociation from mainstream American values, may have been present since our Founding Fathers began the American experiment. It would seem that, like in the parable, they may have sown bad seed with the good (13, Matthew 24:32 RSV). This is not to say they set out to produce anything bad; rather, they certainly aimed for both the good and the true. They began with a description of self-evident truths, "All men are created equal . . . endowed by their Creator with certain inalienable rights, that among these are life, liberty, and the pursuit of happiness" (Declaration of Independence).

How then did our nation devolve from this lofty declaration to present day denigrations, often leaving us feeling like outsiders again? In this essay, we will

[1] Pellegrino Center for Clinical Bioethics, Georgetown University, Washington, DC, USA

Corresponding Author:
G. Kevin Donovan, MD, MA, Pellegrino Center for Clinical Bioethics, Georgetown University, 4000 Reservoir Rd NW, Bldg D #236, Washington, DC 2000
Email: donovangk@georgetown.edu

argue that there is a link to a succession of present-day actions and behaviors that begins with something small, a contraceptive pill. Temporally, this was juxtaposed to a reworking of the meaning of liberty and happiness that our founding fathers had introduced under the influence of the Enlightenment philosophies but led to a much different activity than they would have envisioned. These behaviors and these actions would have been labeled in previous generations as sins by a majority of our fellow Americans. This can be seen as entirely in keeping with our own church teaching.

Moral Challenges of the Notion of "Liberty and the Pursuit of Happiness" in Today's Context

From the perspective of the Catechism of the Catholic Church (1999, n.1849), sin is seen as "an offense against reason, truth and right conscience . . . love of oneself, even to contempt of God." It is, as it always has been, a preference for our own desires, or our own will, over that of God. There are indeed great many kinds of sin; Scripture provides several lists of them. The letter to the Galatians contrasts the works of the flesh with the fruit of the spirit: "Now the works of the flesh are plain: fornication, impurity, licentiousness . . . selfishness, drunkenness, carousing, and the like" (5 Galatians, 19:21). To most, this would seem like an unequivocal listing of things to avoid, even if they have never been completely avoided throughout history. To some today, it might also look like the unapologetic agenda for many people's weekend. Things seen in the past as moral failings are increasingly practiced openly, even celebrated in contemporary culture. This is where our moral fault lines begin. As a society, we have become increasingly more tolerant—tolerant of contraception, sexual promiscuity, pornography, oral and anal sex, both for heterosexual as well as homosexual couples, and as a consequence, same-sex marriage. Of course, the field of bioethics has matched these issues with moral challenges of its own: abortion, embryo destruction for research, transgender surgeries, physician-assisted suicide, and euthanasia. Although exploring the morality of each of these is beyond our scope, we can look at a few of them as exemplars. If we can identify a common theme that unites them, we may be able to perceive a common explanation for our present predicament. Perhaps this can be identified in those who intended to found our society and its culture on twin pillars of *liberty* and *the pursuit of happiness*.

A contemporary concept for the pursuit of happiness could be defined as "freedom for self-actualization." This was famously stated in a form that pushes the limits of credulity, by Supreme Court Justice Anthony Kennedy: "At the heart of liberty is the right to define one's concept of existence, of meaning, of the universe, and of the mystery of human life" (Planned Parenthood v. Casey 1992, p. 833). The egotism of this stark assertion of the independence of an individual's values from outside moral influences could be traced to the individual rights language of philosophers such as John Locke that was echoed in the Declaration of Independence. Other philosophers, such as Mill (2011) and Comte (1817), emphasized the concept that the only real freedom is to pursue our own good as long as we do not harm others. Old virtues could be swept away in favor of the new, especially that foundational virtue, tolerance. The *liberty* to pursue one's will should then lead to ever-expanding personal *happiness*. In a secular morality, the sin does not disappear, it just gets redefined, but it is actually a tightly constrained freedom, limited to only certain acceptable choices. The tolerance being advocated finds its limits here. Wrong choices don't just threaten the new social order; they signal a problem with the chooser who clearly must be disordered. Those who would oppose the prevalent secular values cannot be considered as merely being in principled opposition. They must be seen as mentally deranged and labeled with a corresponding phobia. An individual can no longer express a principled opposition to homosexual behavior; the only acceptable interpretation of such a position is that it represents an irrational fear of homosexual persons, a homophobia. Such a viewpoint is immediately discounted as impermissible, without further consideration or discussion. Tolerance is thus a one-way street and a narrow one at that.

Such a moral shift has implications for how we live our lives and what we find acceptable or unacceptable. Right and wrong become subjective interpretations, not standards, and are entirely based on personal feelings. One should be free to pursue whatever makes one happy, as long as one doesn't hurt anyone. The fundamentally American pursuit of happiness leads to a moral practice motivated by a form of narcissism, which some studies have shown as the hallmark of this generation of millennials (Stinson, Dawson, and Grant 2008). External authority, particularly moral authority, is largely rejected in favor of the authority of the autonomous self. The primary goal is to seek happiness in order to feel good.

Consequences of the Pill

Mankind's original sin was one of ego, of pride, of wanting to be one's own God. This resulted in grasping a forbidden fruit, with all its deadly consequences. The present generations also want to define their own knowledge of good and evil, but the focal point of sinful rebellion is not a contraband apple, but a contraceptive pill. We argue that the taking of the contraceptive pill, to avoid fruitfulness, has led to as many of the subsequent dire consequences as did the original fruit.

For some, this may be seen as too bizarre, too extreme, proposing a culprit that is too tangential to the more serious transgressions of the modern age. Most millennials and even their parents cannot remember the time when contraception wasn't seen by society as a positive good from which many blessings flow. It is understood as a given that regulation of the timing and number of children in a marriage allows women to pursue careers outside the home and allows unmarried women to prevent having their lives ruined by unexpected and unwanted pregnancies. We can even reassure ourselves that we are limiting excessive growth of world population and helping to save the planet. Given these apparently obvious benefits, many young adults would be surprised if not shocked to find that contraception was not always viewed in this favorable light. Throughout the ages, various attempts at contraception focused on condoms, pessaries, herbals, and coitus interruptus. Although Christian churches traditionally discouraged these, the low level of controversy about their use reflected their low level of effectiveness. Even when feminists in the 1870s argued for "voluntary motherhood" in their desire for women's emancipation, they disapproved of contraception, arguing instead for periodic abstinence, and engaging in sex only for purposes of procreation! (Gordon 2002). In America, the Comstock Act of 1870 made it a federal offense to distribute contraceptives, abortifacients, sex toys and erotica, or information regarding them. Note how the use of contraceptives was linked to the use of erotica and sex toys and other "obscenities" of that time. Both later became closely linked with support for abortion—more about that link later. After the turn of the century, a free love movement arose in opposition, spearheaded by Margaret Sanger in the United States and Marie Stopes in England. Both advocated birth control to liberate women and to decrease overall population, especially of the poor and "inferior" races (Sanger 1919). The Anglican Lambeth Conference of 1930 made them the first major Christian church to break ranks in the opposition to contraception, but many more "mainline" Protestant denominations eventually followed. What is little noted is that the previous Lambeth conference had strongly rejected this position, considering contraception "hostile to the family." Instead, they had emphasized the purposes of marriage to be the begetting of children and "deliberate and thoughtful self-control" (Gore 1930).

This controversy did not effectively explode until effective contraception in the form of a hormone pill became available. In 1957, a hormonal pill was approved by the US Food and Drug Administration for serious menstrual disorders, and by 1960, it was approved as a contraceptive. Even then there were problems including some serious side effects (Nikolev 2010) and resistance by African American activists who charged Planned Parenthood with genocidal intent by pushing it in their neighborhoods. The "pill" has since been joined by a variety of implantable hormonal devices and adjuncts to diaphragms.

When a study of contraception was completed by the Catholic Church, an encyclical letter, *Humanae Vitae*, was issued by Pope Paul VI in 1968. While supporting the concept of birth control, insofar as it applies to the desire or need to space or limit the number of children born to a married couple, it insisted that each marital act must remain "open" to both unitive love and possible procreation. It rejected artificial interference in the latter, including artificial hormonal contraception. A moral decision to limit conception would have to be in accord with natural law, precluding both hormonal drugs and other barriers including sterilization. Of course, abortion also remained unacceptable. Taking advantage of the natural rhythms to both conceive, and to prevent conception, was still seen as morally licit. Humanae Vitae challenged the ideas of "free love" and "sexual liberation" by promoting marital love, chastity, and openness to life. In this Encyclical, social issues such as infidelity, the degradation of morality, the loss of sight of what is right and wrong regarding sexual behavior, and the political uses of contraceptives are foreseen (Klaus 2018). The Pope finished by stating "not everyone will easily accept what has been said."

Only the last statement found a common point of agreement. A storm of controversy ensued from laypersons as well as many clergyman and theologians (Harris 1968). It was rejected by many Catholics "in the pews" who were supported by some of their pastors and teachers. The consequences have been far ranging, with some self-identified Catholics feeling that they should also be able to pick and choose

which other moral teachings of the church they will accept. In a recent study of Catholic college students, not only do a large majority use contraception, but 57 percent support abortion, 71 percent homosexual marriage, and 49 percent casual sex (Gray and Cicade 2010). These discrepancies between what the church teaches as morally acceptable compared to what contemporary Catholics practice are not a random assortment of issues. They flow from one to another, with their own internal logical consistency and their own cultural constituency. Less obvious is the fact that contraception can be seen as the linchpin. The contraceptive pill had become the "sine qua non" of the sexual revolution of the 1960s and 1970s. Cultural attitudes of the time were packed with generational conflict, rebellion toward authority, an unpopular war, and the early onset of radical feminism and the drug culture. The keg was already primed, but it took the pill to light the fuse. The youth of America declared, "If it feels good, do it." Freedom and the pursuit of happiness were taking a sharp turn, but were still on the same road begun by our founding fathers, and subsequently endorsed by Justice Kennedy. It was no surprise that few things fit the "feel good" category better than sex. The pill gave us the ability to indulge ourselves without fear of the consequences. If two people want to do this, who should be able to tell them no, especially if no one gets hurt?

In order to fully examine the myth of sex without harm, we must trace out some of the obvious and inevitable consequences. Because the pill separates sex for pleasure from sex for procreation, young women are expected to be able to exercise their liberty to pursue such happiness. When in the past, they could avoid a decision to relinquish their virginity and shun sexual activity due to a very real fear of pregnancy, any preference for chastity now had no automatic justification or irrefutable argument. The free love of the Boomer generation devolved into the hookup culture of the Gen X and millennials. The pill first separated sex from pregnancy; it now separates sex from love, at least a true and committed love. Sex is sought as something for one's own pleasure or, at best, reciprocal pleasure. Denying one's own pleasure is to deny one's own happiness. The possibility of sexual activity becomes the expectation of sexual activity, with the bar set lower and lower. No longer is the quaint question of kissing on the first date an issue; young women can claim it as a sign of their integrity if they withhold sex until a second or third date, or even later. Dating more than one person at a time with these new expectations of presumptive sexual activity makes a young man apparently a "player" or a young girl presumably promiscuous. This dilemma is eclipsed by forgoing all these quaint rules in order to seek hookups, where nearly complete strangers can couple for pleasure without any supposed emotional involvement, and no supposed harm done. Because these occasions usually arise at parties where large amounts of alcohol are consumed, the only harm to be feared appears to be sex without consent. As a consequence, rather than discouraging the hookups, or the alcohol use that leads to it, college campuses have devised elaborate protocols to assure that consent has occurred ("Code of Student Conduct" 2018–2019). Occasionally, egregious violations are publicized and condemned (Stanford Rape Case 2016). Mostly they are not recognized as such or not investigated (Maloney 2016). The ones who suffer the most are often the young women that the pill was supposed to liberate and make happy.

Clearly, the pill does not solve all the problems that flow from increased sexual activity. While it may prevent pregnancy, there are other consequences from which the pill cannot provide protection. Most obvious, sexually transmitted infections (STIs) have concomitantly taken a sharp upturn in recent years. Antibiotic-resistant infections are an increasing danger, as are viral illnesses such as herpes and genital warts, and chlamydia, for which treatments are imperfect and consequences (such as infertility or chronic HIV) can be devastating. Teenagers make up one-third of the United States population but carry 50 percent of STIs. Depression related to loveless and perfunctory teen sex can also be as devastating as any STI (Meeker 2017). Although teen birth rates and abortion have declined over this time, it is due more to a trend away from condom use toward hormonal birth control. As we have seen, hormonal birth control does not lead to avoidance of sexual activity or avoidance of STIs (Green 2016).

Although the increased sexual activity encouraged by the availability of the pill could lead to significant collateral damage, at least it wasn't supposed to lead to pregnancy. Used with perfect compliance, it is an effective contraceptive in more than 90 percent to 95 percent of cases. Unfortunately, we live in an imperfect world with imperfect people, and unintended pregnancies do occur. The cultural response to this unavoidable fact took two directions. One can indulge in nonprocreative sex, anal or oral, or one can seek an abortion for those

who adamantly refused to accept the results of potentially procreative sex. The former was seen as aberrant and even repugnant in previous generations, particularly by females. However, it became more common after President Clinton finally and publicly admitted his own "nongenital" activity with a young White House intern. He justified his previous denials by asserting that because it wasn't mutually genital, it wasn't really sex (Sanders and Reinisch 1999). This led to numbers of high school girls claiming "technical virginity" while acceding to similar demands from their male companions. This change in sexual practice among heterosexual couples led inevitably to a more tolerant attitude toward those for whom nongenital or nonprocreative sex was the preferred or only option. Negative attitudes toward homosexual activity weakened, in part, because it could be seen that they were only practicing what many heterosexuals were also doing, and how could that then be construed as bad or unnatural?

When homosexuality became more acceptable as an "alternate lifestyle," then the twin principles of toleration and pursuit of pleasure would dictate that related issues should be made culturally acceptable as well. These issues included homosexual marriage, transgender conversion surgeries, and a myriad of bathroom issues and controversies. If we cannot condemn those who merely do what we sometimes do, should we not support them for being like ourselves?

Those who followed the other path and became pregnant despite their use of the pill would not accept such an intolerable consequence. The increased demand for abortion became inevitable for those who felt betrayed by the failure of their contraceptive. Indeed 60 percent of abortions are sought by those who were using contraception at the time they became pregnant (Furendi 2017; British Pregnancy Advisory Service 2017). When your cultural norm assures you that no pregnancy should ever be unintended, then some reliable backup for failed contraception will be seen as both a necessity and a right. **This is the inevitable linkage: a contraceptive culture *requires* a concomitant culture of abortion on demand**. An unrestricted commitment to liberty and the pursuit of happiness has its price. This price is the embryo or fetus. As a result, the field of bioethics has had a hard time objecting to embryo experimentation, genetic manipulation, or any creation or destruction of embryos. If it is acceptable to destroy an embryo by a woman in the pursuit of happiness, it is difficult to deny the liberty to pursue its destruction in the pursuit of science.

Final Remarks

While the chief remaining argument against contraception in our culture has been theological, we have eschewed that approach for the sociological. There appear to be a long list of actions and behaviors that might be considered objectionable or sinful by many who would have *no* objection to the use of contraception. Yet, without making a direct religious objection to the pill and other contraceptives, it can be seen that these behaviors that might be termed sinful can be directly linked to the use of contraception in our culture. Fehring, Bouchard, and Meyers (2018) showed additional correlations between contraceptive use in adolescents and negative sexual outcomes. There are additional arguments that can be made against it, including the deleterious effects of cohabitation on the subsequent marriage and divorce rate, the contamination of the water supply by hormonal drugs, and declining fertility rates. Moreover, some of the most challenging and troublesome ethical questions for Catholic physicians in the current age, such as assisted suicide, euthanasia, and restriction of nutrition/hydration leading to a patient's death, all would seem to have a tenuous connection to the contraception–abortion continuum. However, a cheapening of life, and diminishing respect for life in general, is the necessary substrate for these other attitudes and developments. Their justification often depends on similar arguments based on maximizing liberty and happiness. It appears to be no accident that the most ardent supporters of assisted suicide are frequently ardently pro-choice at the other end of the life continuum as well. These considerations are worthy of exploration but also are beyond the scope of this article. What was intended was to recognize a link to a succession of actions and behaviors that begins with the use of contraception. Contraception is seen by its proponents as justified and necessary when procreative sex is severed from sex for pleasure. Seeking unencumbered sexual pleasure is justified by a misunderstood or misapplied sense of liberty and pursuit of happiness. I do not believe the founding fathers would recognize what we have done with their self-evident truths.

If our modern moral morass and overemphasis on individual rights can be interpreted as a devolution from the founding fathers' principles, perhaps a reinterpretation of their intent may suggest a way out. This interpretation would need to rely on a strong justification for the respect shown to individual liberty. Such a justification could be construed from natural law arguments, buttressed by a Christian concept of the value of the individual. Natural rights and basic human equality are most strongly supported by a

Christian justification, as seen in the belief that all individuals have inherent dignity and value. This value and dignity are found in the individual's status as God's beloved creation, as a child of God made in his image and likeness. When the emphasis on the rights and liberty of the individual is once again linked to this religious principle, rather than enlightenment perspectives, not only are the rights of liberty and pursuit of happiness strengthened, they are appropriately redirected. A more profound understanding of happiness then points us away from shallower material and sexual pleasures to deeper and more satisfying metaphysical ones. Thus, those practices that we examined previously would be superseded by those fostered through the practice of continence, self-discipline, charity toward others, and a true understanding of purity. As Chaput (2017) recently put it,

> Given the hyper sexualized nature of today's culture, when we think of purity, we usually think of sexual purity. And thinking of purity, we typically focus on abstinence. So purity somehow transforms into not experiencing a thing we want to experience. This is a distortion. Purity is about wholeness or integrity. It means that the body, mind, heart, and soul are rightly ordered toward God. Every element of who we are is doing its part to bring us to union with God, which is our ultimate happiness. Given the strength of the sexual desires we all feel, rightly acting on those desires is a key part of maintaining purity. For single people and celibates, it means offering those desires up to God and seeking to channel them in our love and service for others. (p. 126)

Finally, if we accept this chain of events and its consequences, where does that leave us as physicians? We may have started with the comfortable assumption that we can practice our profession and our faith, fully assimilated into modern American culture and society. Increasingly, we have come to realize an age-old truth: to be a Catholic Christian is by nature to be countercultural. Christ did not come to help us assimilate into the world but to help us to seek a better one together. Faced with this, what can we do for our patients and what can we do for ourselves? For our patients, there are measures that can be taken, some simple, some daunting. Can we imagine rejecting the assumption that all our young unmarried patients are sexually active or intend to be? Patients rely on their physicians for contraceptive advice, as well as for referrals for abortion, and instructions in safe sex, the condom cautionary, and so on. How much more time will it take to try to warn them of the adverse effects of a sexually libertine lifestyle, just as we try to warn them against smoking? It might seem to be an impossible task to find ourselves in opposition to a pervasive culture that no longer respects our values. It might seem like trying to empty the ocean with a thimble. But let me remind you of another oceanside parable: after a major storm, when hundreds of starfish were washed up on the beach, a little boy walked along throwing them back into the water before they were fried by the sun. An older man told him to forget about it, the task was too great, he couldn't make a difference. The boy replied, as he threw it back into the water, "Well, it will make a difference to this one." Perhaps developing a one-on-one relationship with our patients has always been the only chance of making a difference for the practitioner as well as for the patient and parents. And in attempting to do this for our patients, for as many patients as possible who are willing to accept it, we are doing something more. We are doing something for ourselves, for the maintenance of our own values, for the strengthening of our own moral character and virtues, and for the preservation of what we find best in our profession. This understanding, this approach, would seem more in keeping with the founding fathers' original intent, who did not appear to seek separation of individual liberty from its Christian foundation or our true happiness.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

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

ORCID iD

G. Kevin Donovan, MD, MA https://orcid.org/0000-0002-6115-7840

References

BPAS (British Pregnancy Advisory Center). 2017. *Women Cannot Control Fertility through Contraception Alone: BPAS Data Shows 1 in 4 Women Having an Abortion Were Using Most Effective Contraception*. https://www.bpas.org/about-our-charity/press-office/press-releases/women-cannot-control-fertility-through-contraception-alone-bpas-data-shows-1-in-4-women-having-an-abortion-were-using-most-effective-contraception/.

Chapman, G. 1999. "Sin. Catechism of the Catholic Church (CCC)." Part Three, Section 1, Chapter 1, Article 8.
Chaput, C. 2017. *Strangers in a Strange Land*. New York: Henry Holt.
Comte, C. 1817. *Du nouveau projet de loi sur la presse*. Paris, France: Au Bureau du Censeur européen.
Code of Student Conduct, Sexual Misconduct Policies and Procedures. 2018–2019. Georgetown University, Section VI, C3, 9–10.
Fehring, R., T. Bouchard, and M. Meyers. 2018. "Influence of Contraception Use on the Reproductive Health of Adolescents and Young Adults." *The Linacre Quarterly* 85:167–77. doi: 10.1177/0024363918770462.
Furendi, A. 2017, July 7. "Women Seeking Abortion after Failed Contraception." *British Pregnancy Advisory Service*.
Gordon, L. 2002. *The Moral Property of Women: A History of Birth Control Politics in America*. University of Illinois press, 55–57.
Gore, C. 1930. *Lambeth on Contraceptives*. London, UK: Mombay.
Gray, M., and M. Cicade. 2010. CARA Report: Catholicism on Campus. *Catholic Education: A Journal of Inquiry and Practice* 14:212–237.
Green, E. 2016. "Condoms Don't Necessarily Help Teen Girls Avoid Pregnancy." *The Atlantic* 46–49.
Harris, P. 1968. *On Humanae Vitae: An Examination of Humanae Vitae*. London, UK: Burns and Oatos.
Klaus, H. 2018. "Rejecting Humanae Vitae: The Social Costs of Denying the Obvious." *The Linacre Quarterly* 85:322–26. doi: 10.1177/0024363918817319.
Maloney, A. 2016. "What the Hookup Culture has Done to Women. *Crisis Magazine* June 14.
Meeker, M. 2017, May 31. Teenagers and the STD Nobody is Talking about. *Blog Entry*.
Mill, J. 2011. *On Liberty*. Bedfordshire, UK: Andrews U.K.
Nikolev, A. 2010. "A Brief History of the Birth Control Pill." *Need to Know*. https://www.pbs.org/wnet/need-to-know/health/a-brief-history-of-the-birth-control-pill/480/.
Pope Paul IV. 1968. *On the Regulation of Birth = Humanae Vitae: Encyclical Letter*. Washington, DC: United States Catholic Conference.
Sanders, S. A., and J. M. Reinisch. 1999. "Would You Say You "Had Sex" If . . . ?" *Journal of the American Medical Association* 281:275–77.
Sanger, M. 1919. "Birth Control and Racial Betterment." In *Society for Constructive Birth Control and Racial Progress, edited by M.* Stopes.
Stanford Rape Case: Inside the Court Documents 2016, June 11, CNN. http://cnn.com/2016/06/10/us/Stanford-rape-case-court-documents/Index.html.
Stinson, F., D. Dawson, and B. Grant. 2008. "Prevalence, Correlates, Disability, and Comorbidity of DSM-IV Narcissistic Personality Disorder: Results from the Wave 2 National Epidemiologic Survey on Alcohol and Related Conditions." *Primary Care Companion to the Journal of Clinical Psychiatry* 69:1033–45.
US Supreme Court. 1992. *Planned Parenthood vs. Casey*.-West Supreme Court Report, vol. 112, 2791–885.

Biographical Notes

G. Kevin Donovan, MD, MA, is the Director of the Pellegrino Center for Clinical Bioethics at Georgetown University Medical School and a professor in the Department of Pediatrics. He is a clinician ethicist with over thirty years experience in the field. He began his training as a visiting scholar with Dr. Edmund Pellegrino at the Kennedy Institute of Ethics of Georgetown University in 1989–1990. It was during this time that he began his studies that led to his earning a master's degree in bioethics from University of Oklahoma. At the request of Dr. Pellegrino, he returned to Georgetown in 2012, as director of the Pellegrino Center. Prior to his return to Georgetown, he had served as section chief, vice chair, interim chair, and professor of pediatrics at the University of Oklahoma College of Medicine—Tulsa, where he was the founding director of the Oklahoma Bioethics Center. He received his undergraduate degree from Notre Dame, his MD from the University of Oklahoma, and his master's in bioethics. He trained in pediatrics at Baylor College of Medicine, completed fellowships in pediatric gastroenterology at the Children's Hospital of Oklahoma and the NIH in Bethesda, Maryland, and is board certified in pediatric gastroenterology. He completed a three-year term as chair of the bioethics section of the American Academy of Pediatrics and was appointed as the first person to serve as liaison from the bioethics section to the Committee on Bioethics of the AAP. He also served on the bioethics committee for the North American Society for pediatric gastroenterology, hepatology, and nutrition, the ethics committee of the Oklahoma State Medical Association and was medical ethics consultant to the Roman Catholic Diocese of Tulsa. He served on the local board of directors for the organ sharing network, the Genetics Advisory Council, and was a founding member and first vice president of the Oklahoma Association for Healthcare Ethics. He also served as chair of the Institutional Review Board at St. Francis Hospital for seventeen years. He has published articles on both pediatrics and bioethics and has spoken extensively on both subjects at the local, national, and international level on four continents.

Claudia Sotomayor, MD, DBe, currently works as clinical ethicist at the Pellegrino Center for Clinical Bioethics and adjunct assistant professor of internal medicine of GUMC. She holds an MD from Universidad Autonoma de Chihuahua in Chihuahua, Mexico. She also graduated with a master's degree in bioethics from Anahuac University in Mexico City, and she graduated with a Doctorate in bioethics from Loyola University in Chicago, Illinois. She also completed a fellowship in clinical bioethics at MD Anderson Cancer Center in Houston, Texas. She has been a research scholar for UNESCO chair in bioethics and human rights since 2012 where she has worked in the area of multiculturalism, bioethics, and religion. She has also served as a member of the Ethics Committee in different hospitals in the United States. Before coming to the United States, she worked in different hospitals in Mexico as a primary care physician and was the health committee coordinator for FUNDESPEN, a nonprofit that provides medical care to Mayan communities in rural areas of Quintana Roo, Mexico.

INTERNATIONAL JOURNAL OF UROLOGY

International Journal of Urology (2020) 27, 39–46 doi: 10.1111/iju.14111

Review Article

Effect of antidepressant medications on semen parameters and male fertility

Lauren A Beeder[1] and Mary K Samplaski[2]

[1]Keck School of Medicine, and [2]Institute of Urology, University of Southern California, Los Angeles, California, USA

Abbreviations & Acronyms
BMI = body mass index
FSH = follicle-stimulating hormone
LH = luteinizing hormone
MAOI = monoamine oxidase inhibitor
NDRI = norepinephrine–dopamine reuptake inhibitor
SNRI = serotonin–norepinephrine reuptake inhibitor
SSRI = selective serotonin reuptake inhibitor
TCA = tricyclic antidepressant
TUNEL = transferase dUTP nick end labeling

Correspondence: Lauren A Beeder M.D., Keck School of Medicine, University of Southern California, 1975 Zonal Avenue, Los Angeles, CA 90033, USA. Email: beeder@usc.edu

Received 23 April 2019; accepted 25 August 2019. Online publication 21 September 2019

Abstract: Antidepressant medications are commonly used in males of reproductive age for long-term treatment of depression, as well as other disorders. Although antidepressants are known to be associated with sexual side-effects, their effects on semen parameters and other markers of male fertility have been less thoroughly described. The majority of available studies have focused on selective serotonin reuptake inhibitors, which have been shown to negatively impact semen quality in *in vitro*, animal and human studies. Fluoxetine, in particular, has been the subject of multiple studies and has been associated with gonadotoxic effects, including decreased sperm concentration and motility, increased deoxyribonucleic acid fragmentation, and decreased reproductive organ weights. Studies of several other selective serotonin reuptake inhibitors have yielded similar results. Reassuringly, this effect does seem to be reversible. The data regarding serotonin–norepinephrine reuptake inhibitors, norepinephrine–dopamine reuptake inhibitors, tricyclic antidepressants, monoamine oxidase inhibitors and atypical antidepressants are sparse, varied and conflicting. Given the widespread and often long-term use of antidepressant medications, there is a clear need for further data regarding their impact on semen quality and male fertility.

Key words: antidepressant, effects, male infertility, medications, semen.

Introduction

Antidepressants are one of the most commonly used therapeutic drug classes in the USA. While the majority of these medications are taken to treat depression, antidepressants can also be taken to treat other conditions, such as anxiety disorders. The major classes of these are SSRIs, SNRIs, NDRIs, TCAs, MAOIs and atypical antidepressants. Each of these have slightly different mechanisms of action, and therefore can affect sperm in different ways.

According to the National Health and Nutrition Examination Survey, 8.6% of males between the ages of 12 and 18 in the USA took antidepressants between 2011 and 2014.[1] The number of people taking these medications has increased nearly 65% over a 15-year time frame, from 7.7% in 1999–2002 to 12.7% in 2011–2014. Although females are more likely to use these mediations, males have shown a parallel rise in their antidepressant use. Antidepressant use increases with age, from 3.4% among persons aged 12–19 years to 19.1% among persons aged ≥60 years. In addition, these medications are typically taken long term, and between 2011 and 2014, 68.0% of persons aged ≥12 years who took antidepressant medications had been taking these for ≥2 years, and 25% had been taking them for ≥10 years.

There is limited information on the frequency of antidepressant use globally. The Organization for Economic Co-operation and Development has examined this to some degree. Countries with the highest rates of antidepressant use are the USA (110 per 1000 people), Iceland (106 per 1000 people), Australia (89 per 1000 people), Canada (86 per 1000 people) and Denmark (85 per 1000 people). Countries with the lowest reported antidepressant use are Korea (13 per 1000 people), Chile (13 per 1000 people), Estonia (18 per 1000 people), Hungary (27 per 1000 people) and the Slovak Republic (31 per 1000 people).[2] For Japan specifically, although large epidemiological studies are lacking, it is estimated that up to 6 million Japanese suffer from depression, similar to that seen in Western countries.[3]

All classes of antidepressants are known to be associated with some de function in both men and women. In men, the most notable sexual side- impaired libido, erectile dysfunction, delayed ejaculation or anejaculati

antidepressant medications on semen parameters have been less thoroughly studied, although data do exist for some of the medications in each of the classes of antidepressants. We review all available data (*in vitro*, animal and human studies) regarding the use of antidepressants on semen parameters and male fertility (Table 1).

SSRIs

SSRIs act by inhibiting the reuptake of serotonin, and include citalopram, escitalopram, fluvoxamine, paroxetine, fluoxetine and sertraline. These medications are currently considered first line for the treatment of depression and anxiety disorders. However, SSRIs in particular are known to be associated with significant sexual side-effects, including decreased libido, increased ejaculation latency, alteration of circulating hormones and erectile dysfunction.[4,5] Studies estimate that 25–73% of people treated with an SSRI will experience some type of sexual dysfunction, higher than that of other antidepressants.[4] Studies looking at the impact of these medications on male reproduction and semen parameters are not as robust, but there are some available data. This is the class of antidepressants with the most available data for their effects on semen parameters and male fertility (Table 2).

In vitro studies

There is a single *in vitro* study investigating SSRIs and human sperm. Kumar *et al*. incubated human semen with varying doses of paroxetine, fluoxetine, sertraline, citalopram and fluvoxamine *in vitro*. All SSRIs showed some degree of spermicidal activity, whereas serotonin showed no negative effect on sperm counts. Fluoxetine, which showed the highest spermicidal activity, had a minimal effective concentration comparable to nonoxynol-9, a contraceptive utilized for its spermicidal properties.[6]

Animal studies

Multiple studies carried out in animal models have shown the negative effects of SSRIs on male fertility. Most studies of SSRIs are on fluoxetine, and most do show some degree of gonadotoxic effect. Male rats treated with oral fluoxetine for 60 days were found to have decreased spermatogenesis on histology, as well as significantly decreased sperm density and motility. Treated male rats also had decreased pregnancy and implantation rates with untreated female rats, and had decreased reproductive organ weights, including testes, epididymes, prostates and seminal vesicles.[7] A similar study of rats treated with varying doses of oral fluoxetine for 5 days found a dose-dependent decrease in sperm counts and motility in rats exposed to fluoxetine. At the highest dose (13 mg/kg), sperm count and motility were fourfold less that that observed in controls.[8] Long-term administration of fluvoxamine has also been shown to negatively affect semen parameters, and induce oxidative stress and apoptosis in the testes of rats, which were treated with both low therapeutic doses (9 mg/kg) and high therapeutic doses (27 mg/kg) for 8 weeks.[9] Oxidative stress might mediate and enhance the negative impact that these drugs seem to have on semen parameters. Rats exposed to chronic unpredictable mild stress showed signs of oxidative stress, including increased levels of malondialdehyde and corticosterone, and decreased anti-oxidants, sperm count and motility. These effects were exaggerated in animals treated with

Table 1 Studies carried out on different antidepressant classes

Antidepressant class	*In vitro* studies	Case studies	Animal studies	Human studies
SSRIs	Kumar *et al*.[6]	–	Batalneh and Daradka[7]	Tanrikut and Schlegel[16]
			Alzahrani[8]	Elnazer and Baldwin[18]
			Galal *et al*.[9]	Tanrikut *et al*.[19]
			Sakr *et al*.[10]	Akasheh *et al*.[20]
			Atli *et al*.[11]	Koyuncu *et al*.[21]
			Attia and Bakheet[12]	Safarinejad[22]
			Ilgin *et al*.[13]	Relwani *et al*.[23]
			Ayala *et al*.[14]	
			Vieira *et al*.[15]	
SNRIs	Bandegi *et al*.[24]	–	Bandegi *et al*.[24]	–
NDRIs	–	–	Urra *et al*.[25]	–
			Cavariani *et al*.[26]	
			Fazelipour *et al*.[28]	
			Cansu *et al*.[29]	
			Adriani *et al*.[30]	
			Bellentani *et al*.[31]	
TCAs	Levin *et al*.[32]	–	Bandegi *et al*.[24]	Levin *et al*.[32]
			Chowdary and Rao[34]	Padrón and Nodarse[33]
			Hassanane *et al*.[35]	
MAOIs	–	–	Kalász *et al*.[36]	–
			Mihalik *et al*.[37]	
Atypical antidepressants	Cassidy and Pearson[39]	Elnazer and Baldwin[18]	Ilgin *et al*.[38]	–
			El-Sisi *et al*.[40]	

Table 2 SSRI effects on semen and other fertility parameters

Drug	*In vivo* effects on semen parameters	Other effects on fertility
SSRIs		
Fluoxetine	*Rats:* Decreased spermatogenesis, sperm density and motility	*Rats:* Decreased pregnancy and implantation rates, decreased reproductive organ weight
Fluvoxamine	*Rats:* Decreased sperm concentration and motility, increased abnormal forms	*Rats:* Oxidative stress and apoptosis in testes, decreased FSH, LH, testosterone and estrogen
Sertraline	*Human case study:* Decreased sperm concentration and motility *Human prospective:* Decreased sperm count, increased abnormal morphology and DNA fragmentation *Rats:* Increased DNA damage and abnormal forms, decreased sperm count	*Rats:* Testicular degeneration, oxidative stress
Citalopram	*Human case study:* Decreased sperm count and motility, increased abnormal morphology *Rats:* Increased DNA strand breaks and oxidative damage, increased abnormal forms	*Rats:* Decreased seminal vesicle mass, decreased volume of seminiferous tubules
Paroxetine	*Human prospective:* Increase in DNA fragmentation	–
Escitalopram	*Human prospective:* Decreased sperm concentration and motility, increased abnormal morphology	–

fluoxetine (10 mg/kg/day for 28 days), and mitigated in groups treated with either resveratrol (20 mg/kg/day for 28 days) or fluoxetine plus resveratrol.[10]

One study examined the effects of sertraline on the reproductive system of rats treated with 5, 10 or 20 mg/kg for 4 weeks. There was a dose-dependent increase in DNA damage (measured using the Comet assay), testicular degeneration and abnormal sperm forms, as well as a decrease in sperm count in treated animals. Decreased levels of glutathione and increased levels of malondialdehyde suggested oxidative stress as a main mechanism for the testicular toxicity observed.[11]

Citalopram has also been shown to exert toxic effects on sperm and testicles, largely through oxidative stress. Attia and Bakheet saw a dose- and duration-dependent increase in DNA strand breaks, oxidative DNA damage, and abnormal primary spermatocytes in male rats treated with citalopram.[12] A study by Ilgin *et al.* found similar results, with citalopram-administered rats showing reduced sperm counts, increased abnormal sperm morphology and increased DNA damage.[13]

These drugs might affect pre- and postpubertal animals differently, and a study examining pre- and postpubertal rats treated with fluoxetine for 30–53 days found that these two populations had different responses. Although all animals had a decrease in LH, FSH, progesterone and testosterone, prepubertal rats were more likely to show decreased sperm membrane integrity, density, motility and morphology versus their adult counterparts.[14] This might have implications for males taking these medications beginning at young ages.

Maternal use of SSRIs during pregnancy and lactation might also impact semen quality in male offspring. Rat offspring exposed to fluoxetine *in utero* and while nursing have been shown to have decreased seminal vesicle mass and sperm counts, as well as reduced height and diameter of seminiferous tubules.[15]

Human studies

Human data consistently support an association between male infertility (semen parameters and sperm DNA fragmentation) and SSRI use. In 2007, Tanrikut and Schlegel described cases of oligospermia, impaired motility and abnormal morphology in two patients taking SSRIs for depression. The first patient presented on citalopram with "marked oligospermia and 1% motility." Semen analysis 1 month after citalopram discontinuation showed a marked improvement in all parameters to within the normal range. Bupropion was started for depression shortly thereafter, and a semen analysis while on bupropion again showed a decrease in sperm concentration to 21 million/mL with 10% motility. After two failed *in vitro* fertilization attempts, the patient was reassessed (still on bupropion). His DNA fragmentation (tested by the sperm chromatin structural assay) was 76%. He was weaned from the bupropion and his follow-up semen analysis 1 month after bupropion discontinuation showed a normal sperm concentration of 41 million/mL with 75% motility. A second semen analysis carried out 2 months after bupropion discontinuation showed normal sperm concentration and motility. A similar pattern (impaired semen parameters on sertraline [sperm concentration of 20 000 with 0% motility]), with dramatic improvement after SSRI discontinuation (3 months after discontinuation 40 million motile sperm) was seen for the second patient described taking sertraline.[16] Human spermatogenesis takes 72 days, and therefore this marked improvement within weeks of antidepressant discontinuation suggests that SSRIs might exert their effects on post-testicular processes rather than spermatogenesis itself.[17] Similarly, Elnazer and Baldwin described a patient with markedly improved sperm concentration, progressive motility and morphology after discontinuation of citalopram.[18]

In a subsequent prospective study, Tanrikut *et al.* examined the effects of paroxetine on semen parameters and DNA fragmentation in 35 healthy male volunteers with normal baseline semen parameters and DNA fragmentation (measured by the TUNEL assay). Study participants (mean age 34 years, range 19–58 years) were treated with therapeutic paroxetine for 5 weeks. Semen parameters and sperm DNA fragmentation were tested before treatment and again post-treatment after a 1-month washout period. Use of paroxetine was associated with a significant increase in DNA fragmentation, from 14% at baseline to 30% post-treatment. In

addition, the number of men having elevated sperm DNA fragmentation of >30% increased from 10% at baseline to 50% post-treatment (odds ratio 9, confidence interval 2.3–38). In contrast to some other studies, these authors did not identify a change in semen parameters with SSRI use.[19] This suggests that although raw semen parameters might be affected by SSRI use, sperm DNA fragmentation might also be affected even in the absence of changes in semen parameters and could represent an alternative means for impaired male reproductive potential.

Prospective data have also supported a relationship between SSRI use and markers of male infertility. In a randomized, single-blinded clinical trial, 60 men were treated for primary premature ejaculation with either sertraline or non-pharmacological behavioral therapy. The sertraline group was treated with sertraline 25 mg/day for 1 week, followed by 50 mg/day for 3 months. Both sperm concentration (reduction by 10^5/mL) and percent normal morphology were significantly decreased in the sertraline group versus controls. DNA fragmentation (sperm chromatin dispersion method) was also increased in the treatment group (31% *vs* 16%).[20] Another prospective study by Koyuncu *et al.* showed decreased sperm concentration (26.4×10^6/mL *vs* 68.9×10^6/mL), motility (23.4% *vs* 58.2%) and morphology (23.4% *vs* 58.2%) after 3 months of exposure to escitalopram for the treatment of premature ejaculation.[21]

Other factors, including duration of SSRI use and BMI, might synergistically adversely affect semen parameters. One cross-sectional study compared semen parameters and sperm DNA fragmentation in men taking SSRIs versus those of healthy men, and also included an evaluation of the duration of antidepressant use. Men taking SSRIs were found to have significantly lower sperm counts (61 million *vs* 184 million), motility (49% *vs* 66%) and normal morphology (8% *vs* 20%), as well as significantly increased amounts of fragmented sperm DNA versus controls (43% *vs* 21%). All differences in semen parameters and sperm DNA fragmentation correlated with the duration of antidepressant use (6–12 months *vs* 1–2 years), although no differences were observed between specific antidepressants within the SSRI class.[22] Another study of 530 men aged 18–50 years examining the effect of BMI found that use of combination SSRIs was associated with a significant decrease in sperm motility, independent of BMI.[23]

In vitro, animal and human studies all showed a decline in semen quality with SSRI use, as manifested by both impaired semen parameters and increased DNA fragmentation rates. The duration of recovery (<73 days, the time required for spermatogenesis) to baseline semen parameters and DNA fragmentation suggests that these effects might be due to some type of post-testicular process. Given the wide prevalence of the use of this class of medications, there is a clear need for further large-scale, randomized, placebo-controlled trials to further characterize the role of SSRIs in infertility, and their effect on semen parameters and other markers of male fertility.

SNRIs and NDRIs

SNRIs exert their effects by inhibiting the reabsorption of both serotonin and norepinephrine. This class of medications includes desvenlafaxine, duloxetine, levomilnacipran and venlafaxine. The prevalence of sexual dysfunction is 58–70% in patients treated with SNRIs,[4] in general slightly less than that seen for SSRIs. There has only been a single study investigating the effects of any of these medications on semen parameters (Table 3). This group examined 40 adult male mice given oral venlafaxine (2 mg/kg) or venlafaxine (2 mg/

Table 3 Non-SSRI antidepressant effects on semen and other fertility parameters

Drug	*In vivo* effects on semen parameters	Other effects on fertility
SNRIs		
Venlafaxine	*Human prospective:* Improved normal morphology and sperm viability, increased non-progressive motility	–
NDRIs		
Bupropion	*Rats:* Decreased motility at high doses (30 mg/kg)	*Rats:* Increased epididymal duct contractility
Methylphenidate	*Rats:* Increased abnormal sperm tail morphology; increased spermatogonia; reduced round spermatids; increased sperm count	*Rats:* Increased testicular interstitial tissue; decreased germinal epithelium thickness, increased gonadotropins; decreased testicular weight, increase in apoptosis; increased testicular weight
Sibutramine	*Rats:* Decreased sperm number in epididymis, decreased transit time within epididymis	*Rats:* Decreased reproductive organ weight
TCAs		
Desipramine	*Human:* No change in sperm count or motility	–
Amitriptyline	*Human prospective:* Increased ejaculate volume, sperm count and normal morphology; decreased sperm concentration and viability; decreased sperm count and normal forms	*Human prospective:* Increased germ cell mutations
MAOIs		
Selegiline	*Rats:* Increased sperm count and viability	*Rats:* Increase in testis mass
Atypical antidepressants		
Trazodone	*Rats:* Decreased sperm concentration, motility and normal morphology, increased DNA damage	–
Mirtazapine	*Rats:* Protective effect against nitrofurazone-induced decrease in sperm count and viability	–
Agomelatine	*Human case study:* No effect on semen parameters	–

kg) plus vitamin C (10 mg/kg) for 35 days. Mice treated with venlafaxine alone had better sperm morphology (58.50% *vs* 43.71%), non-progressive motility (25.50% *vs* 16.25%) and sperm viability (80.25% *vs* 64.62%) compared with controls. This effect is thought to be a result of the anti-oxidant properties of venlafaxine in protecting against lipid peroxidation. There were no significant differences between semen parameters in mice treated with venlafaxine alone and those treated with combination venlafaxine and ascorbic acid.[24]

NDRIs act by blocking the reuptake of norepinephrine and dopamine from the synaptic terminal, thereby increasing their bioavailability. This class includes bupropion, dexmethylphenidate, diphenylprolinol, ethylphenidate, methylenedioxypyrovalerone, methylphenidate, pipradrol, prolintane and sibutramine. In general, the limited available data for these medications show varied effects on semen parameters (Table 3).

Bupropion is commonly used in combination with other medications in the treatment of depression, as well as for smoking cessation. Although the role of dopamine in reproductive physiology has not been clearly established, there are limited data implicating some role in male reproductive function. Urra *et al.* first identified the presence of functional dopamine transporters in equine sperm. In the present study, high levels of dopamine were associated with decreased total and progressive sperm motility, and this effect was partially reversed by the addition of bupropion. Blocking the dopamine transporter reduced uptake of a dopamine analog, thereby decreasing accumulation of the catecholamine in equine sperm.[25] Another study evaluated the effects of bupropion on semen parameters and epididymal duct contractility in rats. At lower doses (15 mg/kg), bupropion increased epididymal duct contractility, but had no effect on semen parameters. At higher doses (30 mg/kg), the drug was shown to impair sperm motility.[26]

Methylphenidate is a psychostimulant that inhibits norepinephrine and dopamine reuptake. It is currently most commonly used in the treatment of attention-deficit/hyperactivity disorder in children and adolescents. In the past it was used as an antidepressant, and there are some conflicting data on its effect on semen parameters. Motagnini *et al.* studied the effects of methylphenidate administration on rats during childhood and young adult development. An increase in abnormal sperm tail morphology was observed, as well as an increase in testicular interstitial tissue in treated animals.[27] A different rat study found that treatment with methylphenidate was associated with decreased germinal epithelium thickness, as well as an increase in the number of spermatogonia, likely secondary to increases in serum gonadotropin levels.[28] Finally, a rat study by Cansu *et al.* showed a dose-dependent association between 90-day exposure to methylphenidate and reduced numbers of round spermatids, decreased testicular weight, and an increase in apoptosis (TUNEL method) and expression of p53.[29] Conversely, Adriani *et al.* found that adolescent rats exposed to methylphenidate had increased testicular weights and increased sperm count as adults.[30] These data are conflicting, and there are no human data, making it difficult to truly know the effect of methylphenidate on semen parameters and male fertility.

Sibutramine, initially developed for use in the treatment of depression, is a monoamine reuptake inhibitor commonly used today for weight loss. There are no human studies for sibutramine, and only a single animal study. Bellentani *et al.* found that exposure to 10 mg/kg of sibutramine for 28 days decreased weights of reproductive organs in male rats, including the ventral prostate and epididymis, although there were no histological changes noted in these organs. The sperm number within the epididymis (180.98×10^6/organ *vs* 276.16×10^6/organ) and transit time within the epididymis (4.73 days *vs* 7.85 days) were also significantly decreased. There was no change in spermatid number within testes, daily sperm production, sperm motility or morphology between groups.[31]

There is a clear lack of data for many of these medications. No studies exist for duloxetine, desvenlafaxine, levomilnacipran, dexmethylphenidate, diphenylprolinol, ethylphenidate, methylenedioxypyrovalerone, pipradrol or prolintane. Scant data exist for venlafaxine, bupropion, methylphenidate and sibutramine. No prospective clinical studies have yet been carried out exploring the effects of SNRIs and NDRIs on semen quality. Given the contradictory results found in preliminary animal studies, there is a clear need for additional research in this area.

TCAs

TCAs, including amitriptyline, nortriptyline, amoxapine, desipramine, doxepin, imipramine, protriptyline and trimipramine, were one of the earliest medications used to treat depression. However, they are generally no longer used as first-line medications because of significant side-effect profiles. The estimated prevalence of sexual dysfunction in men and women taking TCAs is comparatively low, at approximately 30%.[4] Evidence regarding the effects of TCAs on semen quality is scant (Table 3). There is a small number of studies on desipramine and amitriptyline, but no studies on nortriptyline, amoxapine, doxepin, imipramine, protriptyline or trimipramine.

A 1981 study by Levin *et al.* described both *in vitro* and clinical studies examining the effect of desipramine on semen parameters. *In vitro*, desipramine was associated with dose-dependent inhibition of sperm motility. However, *in vivo* clinical evaluation found no difference in sperm count or motility between treatment and control groups. Treatment with desipramine was associated with decreased sperm viability (defined as the percentage of motile spermatozoa, no true viability testing was carried out). Other semen parameters did not significantly differ between treatment and control groups.[32]

There have been four studies examining amitriptyline, yielding conflicting results. A small study on the effects of amitriptyline in 20 infertile men with oligospermia found increased ejaculate volume, sperm count and normal morphology after treatment with amitriptyline. Sperm count was increased in 50% of patients, and motility was increased in 35% of patients.[33] In contrast, Bandegi *et al.* found negative effects of amitriptyline on semen parameters in rats treated with amitriptyline alone versus amitriptyline and ascorbic acid. Rats treated with amitriptyline alone had lower sperm

concentration (18.11 million *vs* 22.41 million) and viability (31.25% *vs* 64.62%) compared with controls. These results were also seen when comparing the group receiving amitriptyline plus ascorbic acid and controls, and the addition of ascorbic acid did not seem to mitigate this effect.[24] Two studies suggest that amitriptyline has a mutagenic effect on sperm. A study by Chowdary and Rao showed mutagenic effects of amitriptyline in germ cells of mice treated with various oral doses of the antidepressant.[34] Similar results were obtained in an animal study in which amitriptyline was found to increase chromosome abnormalities, and decrease sperm count and normal morphology.[35]

MAOIs

MAOIs are typically reserved for patients who have failed other first-line medications for the treatment of depression. Medications belonging to this class include selegiline, isocarboxazid, phenelzine and tranylcypromine. Approximately 40% of male and female patients taking MAOIs will experience some degree of sexual dysfunction.[4] Data on semen parameters are scant (Table 3). There have been two rat studies examining the effects of selegiline, but no studies examining the effects of other MAOIs, including isocarboxazid, phenelzine and tranylcypromine, on semen parameters or other markers of male fertility.

Selegiline, used to treat major depressive disorder and parkinsonism, might actually have a favorable effect on male fertility. A small rat study found an increase in testis mass, sperm number and viability (defined as the ratio of live to dead sperm) in rats treated with oral selegiline for 4 weeks.[36] These findings were corroborated in a study of rats treated with intraperitoneal selegiline. Treated rats had significantly higher sperm counts (137.73×10^6/mL *vs* 115.09×10^6/mL on semen analysis) than those receiving intraperitoneal saline.[37] The mechanism for this increase in sperm counts and viability is unclear.

Atypical antidepressants

Atypical antidepressants are those that act by mechanisms separate from those discussed above. These medications include mirtazapine, trazodone, nefazodone, tianeptine, agomelatine, vilazodone and vortioxetine. There are limited data regarding the effects of these medications on fertility through semen parameters, although the available data do suggest a negative effect on semen quality (Table 3).

One study of rats receiving vehicle (control), 5, 10 or 20 mg/kg/day of trazodone for 28 consecutive days found decreased sperm concentration (4.68×10^6/mL, 3.04×10^6/mL, 2.84×10^6/mL and 2.68×10^6/mL, respectively), sperm motility (86.49%, 80.06%, 78.85% and 76.23%, respectively) and normal morphology (18.00%, 28.90%, 31.20% and 37.08% abnormal forms, respectively), as well as increased DNA damage in treated rats. Increased malondialdehyde levels suggested that oxidative stress contributed to the testicular toxicity in these animals.[38] Similarly, Cassidy and Pearson showed that trazodone had an inhibitory effect on motility in *in vitro* samples of human sperm.[39]

Mirtazapine has been shown in one study to have a protective effect against oxidative stress and testicular damage. In that study, testicular damage in rats was induced by administration of nitrofurazone. Rats exposed prophylactically to mirtazapine 1 week before the initiation of nitrofurazone had a significantly less pronounced decline in sperm count and viability than those receiving only nitrofurazone, as well as decreased indicators of oxidative damage.[40]

One case study suggested that agomelatine (*vs* citalopram) does not negatively impact semen quality (at least in one patient). Elnazer and Baldwin described a case of decreased sperm concentration, motility, progressive motility and normal morphology in a patient treated with citalopram for mixed depression and anxiety. These effects resolved after withdrawal of citalopram. The patient was subsequently treated with agomelatine, which was not associated with a decline in semen parameters.[18]

There have been no studies to date regarding the effects of nefazodone, tianeptine, vilazodone or vortioxetine on semen parameters or other markers for male fertility.

Conclusions

Given the relatively common use of antidepressant medications, the limited data on their use is worrisome. This is particularly concerning, given that these medications are taken by young men, generally on a long-term basis. We do not have pregnancy or live birth data on any of these medications. All of them have the potential to affect sexual performance to varying degrees.

The existing data show that SSRIs exert a harmful effect on semen quality and rates of DNA fragmentation, as well as increase oxidative stress within reproductive organs. Most of these effects do seem to be reversible on cessation of treatment with SSRIs, although this might not be possible for all patients who need these medications to control their depression. There is contrasting evidence regarding whether bupropion (an NDRI), negatively or positively impacts sperm motility. Similarly, the available data for methylphenidate is animal only, and conflicting, making it difficult to know its true effect. Studies on amitriptyline have yielded conflicting results, with one small clinical study suggesting improved semen parameters in patients with baseline oligospermia, but three animal studies finding the opposite effect. The only MAOI that has been studied in this arena is selegiline, which has been shown to increase sperm counts and viability in two rat studies.

Evaluations of the effects of treatment with antidepressant medication on semen parameters should consider the effects of untreated depression and anxiety on fertility as well. Shiraishi and Matsuyama showed that comorbid medical conditions negatively impacted spermatogenesis and that treatment of medical comorbidities, including hypertension, hyperlipidemia, hyperuricemia and skin disease, was associated with a significant increase in motile sperm count.[41] In particular, psychosocial stress has been shown to have an inverse effect on testosterone levels.[42] Furthermore, it is not surprising that a diagnosis of major depression appears to affect gonadal

function. Men with depressive episodes of moderate-to-severe severity have been shown to have lower levels of testosterone[43] as well as lower sex hormone-binding globulin and DHEA-S, higher secretion of cortisol and prolactin, and lower semen volume and sperm density.[44]

Given the widespread and often long-term use of antidepressant medications, there is a clear need for further data regarding their impact on semen quality and male fertility. The existing data are often based on animal studies or human studies with low numbers of patients. There is a stark absence of prospective data. At this point, it is difficult for clinicians to counsel patients on the effect that these medications might have on their fertility. We would recommend an informed discussion with patients attempting parenthood and taking these medications. Checking a baseline semen analysis and sperm DNA fragmentation might provide some level of guidance. If possible, a trial of discontinuation of antidepressants is advised, although this should be carried out in conjunction with the patient's mental health provider.

Conflict of interest

None declared.

References

1 Pratt LA, Brody DJ, Gu Q. *Antidepressant Use among Persons Aged 12 and over: United States, 2011–2014*. NCHS data brief, no 283. Hyattsville, MD: National Center for Health Statistics, 2017.

2 Organisation for Economic Co-operation and Development. *OECD Health Statistics*. Paris: Organisation for EconomicCo-operation and Development, 2015.

3 Berger D. Antidepressant drug development in Japan: current & future horizons. *Clin. Res. Focus* 2005; **16**: 32–5.

4 Higgins A, Nash M, Lynch AM. Antidepressant-associated sexual dysfunction: impact, effects, and treatment. *Drug Healthc. Patient Saf.* 2010; **2**: 141–50.

5 Montejo AL, Llorca G, Izquierdo JA, Rico-Villademoros F. Incidence of sexual dysfunction associated with antidepressant agents: a prospective multicenter study of 1022 outpatients. Spanish Working Group for the study of psychotropic-related sexual dysfunction. *J. Clin. Psychiatry* 2001; **62** (Suppl 3): 10–21.

6 Kumar VS, Sharma VL, Tiwari P *et al*. The spermicidal and antitrichomonas activities of SSRI antidepressants. *Bioorg. Med. Chem. Lett.* 2006; **16**: 2509–12.

7 Batainch HN, Daradka T. Effects of long-term use of fluoxetine on fertility parameters in adult male rats. *Neuro. Endocrinol. Lett.* 2007; **28**: 321–5.

8 Alzahrani HA. Sister chromatid exchanges and sperm abnormalities produced by antidepressant drug fluoxetine in mouse treated in vivo. *Eur. Rev. Med. Pharmacol. Sci.* 2012; **16**: 2154–61.

9 Galal AA, Alam RT, Abd El-Aziz RM. Adverse effects of long-term administration of fluvoxamine on haematology, blood biochemistry and fertility in male albino rats: a possible effect of cessation. *Andrologia* 2016; **48**: 914–22.

10 Sakr HF, Abbas AM, Elsamanoudy AZ, Ghoneim FM. Effect of fluoxetine and resveratrol on testicular functions and oxidative stress in a rat model of chronic mild stress-induced depression. *J. Physiol. Pharmacol.* 2015; **66**: 515–27.

11 Atli O, Baysal M, Aydogan-Kilic G *et al*. Sertraline-induced reproductive toxicity in male rats: evaluation of possible underlying mechanisms. *Asian J. Androl.* 2017; **19**: 672–9.

12 Attia SM, Bakheet SA. Citalopram at the recommended human doses after long-term treatment is genotoxic for male germ cell. *Food Chem. Toxicol.* 2013; **53**: 281–5.

13 Ilgin S, Kilic G, Baysal M *et al*. Citalopram induces reproductive toxicity in male rats. *Birth Defects Res.* 2017; **109**: 475–85.

14 Ayala ME, Gonzáles A, Olivarez RM, Aragón-Martínez A. Fluoxetine treatment of prepubertal male rats uniformly diminishes sex hormone levels and in a subpopulation of animals, negatively affects sperm quality. *Reprod. Fertil. Dev.* 2018; **30**: 1329–41.

15 Vieira ML, Hamada RY, Gonzaga NI *et al*. Could maternal exposure to the antidepressants fluoxetine and St. John's Wort induce long-term reproductive effects on male rats? *Reprod. Toxicol.* 2013; **35**: 102–7.

16 Tanrikut C, Schlegel PN. Antidepressant-associated changes in semen parameters. *Urology* 2007; **69**: 185.e5–e7.

17 Drobnis EZ, Nangia AK. Psychotropics and male reproduction. In: Drobnis EZ, Nangia AK (eds). *Impacts of Medications on Male Fertility*, 1034th edn. Springer, 2017; 64–78.

18 Elnazer HY, Baldwin DS. Treatment with citalopram, but not with agomelatine, adversely affects sperm parameters: a case report and translational review. *Acta Neuropsychiatr.* 2014; **26**: 125–9.

19 Tanrikut C, Feldman AS, Altemus M, Paduch DA, Schlegel PN. Adverse effect of paroxetine on sperm. *Fertil. Steril.* 2010; **94**: 1021–6.

20 Akasheh G, Sirati L, Noshad Kamran AR, Sepehrmanesh Z. Comparison of the effect of sertraline with behavioral therapy on semen parameters in men with primary premature ejaculation. *Urology* 2014; **83**: 800–4.

21 Koyuncu H, Serefoglu EC, Yencilek E, Atalay H, Akbas NB, Sarica K. Escitalopram treatment for premature ejaculation has a negative effect on semen parameters. *Int. J. Impot. Res.* 2011; **23**: 257–61.

22 Safarinejad MR. Sperm DNA damage and semen quality impairment after treatment with selective serotonin reuptake inhibitors detected using semen analysis and sperm chromatin structure assay. *J. Urol.* 2008; **180**: 2124–8.

23 Relwani R, Berger D, Santoro N *et al*. Semen parameters are unrelated to BMI but vary with SSRI use and prior urological surgery. *Reprod. Sci.* 2011; **18**: 391–7.

24 Bandegi L, Anvari M, Vakili M, Khoradmehr A, Mirjalili A, Talebi AR. Effects of antidepressants on parameters, melondiadehyde, and diphenyl-2-picryl-hydrazyl levels in mice spermatozoa. *Int. J. Reprod. Biomed.* 2018; **16**: 365–72.

25 Urra JA, Villaroel-Espíndola F, Covarrubias AA, Rodríguez-Gil JE, Ramírez-Reveco A, Concha II. Presence and function of dopamine transporter (DAT) in stallion sperm: dopamine modulates sperm motility and acrosomal integrity. *PLoS One* 2014; **9**: e112834.

26 Cavariani MM, de Almeida Kiguti LR, de Lima RJ *et al*. Bupropion treatment increases epididymal contractility and impairs sperm quality with no effects on the epididymal sperm transit time of male rats. *J. Appl. Toxicol.* 2015; **35**: 1007–16.

27 Montagnini BG, Silva LS, dos Santos AH *et al*. Effects of repeated administration of methylphenidate on reproductive parameters in male rats. *Physiol. Behav.* 2014; **133**: 122–9.

28 Fazelipour S, Tootian Z, Saremi ZG *et al*. Evaluation of histopathologic and histomorphometric changes of testicular tissue and gonadotropin levels following consumption of methylphenidate in male mice. *Turk. J. Med. Sci.* 2014; **44**: 554–9.

29 Cansu A, Ekinci O, Serdaroglu A, Erdogan D, Coskun ZK, Gürgen SG. Methylphenidate has dose-dependent negative effects on rat spermatogenesis: decreased round spermatids and testicular weight and increased p53 expression and apoptosis. *Hum. Exp. Toxicol.* 2011; **30**: 1592–600.

30 Adriani W, Leo D, Guarino M *et al*. Short-term effects of adolescent methylphenidate exposure on brain striatal gene expression and sexual/endocrine parameters in male rats. *Ann. N. Y. Acad. Sci.* 2006; **1074**: 52–73.

31 Bellentani FF, Fernandes GS, Perobelli JE *et al*. Acceleration of sperm transit time and reduction of sperm reserves in the epididymis of rats exposed to sibutramine. *J. Androl.* 2011; **32**: 718–24.

32 Levin RM, Amsterdam JD, Winokur A, Wein AJ. Effects of psychotropic drugs on human sperm motility. *Fertil. Steril.* 1981; **36**: 503–6.

33 Padrón RS, Nodarse M. Effects of amitriptyline on semen of infertile men. *Br. J. Urol.* 1980; **52**: 226–8.

34 Chowdary PS, Rao MS. Cytogenetic effects of amitriptyline hydrochloride in somatic and germ cells of mice. *Toxicol. Lett.* 1987; **39**: 199–204.

35 Hassanane MS, Hafiz N, Radwan W, El-Ghor AA. Genotoxic evaluation for the tricyclic antidepressant drug, amitriptyline. *Drug Chem. Toxicol.* 2012; **35**: 450–5.

36 Kalász H, Thuróczy J, Karvaly G *et al*. Effects on sperms' quality of selegiline in aged rats. *Open Med. Chem. J.* 2017; **11**: 138–45.

37 Mihalik J, Mašlanková J, Solár P *et al*. The effect of R-(-)-deprenyl administration on reproductive parameters of rat males. *Eur. J. Pharmacol.* 2015; **754**: 148–52.
38 Ilgın S, Aydoğan-Kılıç G, Baysal M *et al*. Toxic effects of trazodone on male reproductive system via disrupting hypothalamic-pituitary-testicular axis and inducing testicular oxidative stress. *Oxid. Med. Cell Longev.* 2018; **2018**: 7196142.
39 Cassidy SL, Pearson RM. Effects of trazodone and nadolol upon human sperm motility. *Br. J. Clin. Pharmacol.* 1986; **22**: 119–21.
40 El-Sisi AE, El-Sayad ME, Abdelsalam NM. Protective effects of mirtazapine and chrysin on experimentally induced testicular damage in rats. *Biomed. Pharmacother.* 2017; **95**: 1059–66.
41 Shiraishi K, Matsuyama H. Effects of medical comorbidity on male infertility and comorbidity treatment on spermatogenesis. *Fertil. Steril.* 2018; **110**: 1006–11.e2.
42 Nilsson PM, Møller L, Solstad K. Adverse effects of psychosocial stress on gonadal function and insulin levels in middle-aged males. *J. Intern. Med.* 1995; **237**: 479–86.
43 Schweiger U, Deuschle M, Weber B *et al*. Testosterone, gonadotropin, and cortisol secretion in male patients with major depression. *Psychosom. Med.* 1999; **61**: 292–6.
44 Wdowiak A, Bień A, Iwanowicz-Palus G, Makara-Studzińska M, Bojar I. Impact of emotional disorders on semen quality in men treated for infertility. *Neuro. Endocrinol. Lett.* 2017; **38**: 50–8.

Editorial Comment

Editorial Comment to Effect of antidepressant medications on semen parameters and male fertility

The paper by Beeder and Samplaski is a review article of antidepressant medications and male fertility.[1]

Antidepressant medications are widely applied for patients in the psychological field. As the authors mentioned, it is well known that antidepressant medications cause male sexual dysfunction, including erectile dysfunction and late-onset hypogonadism syndrome.

A few previous studies have reported a relationship between male fertility factors and antidepressants, mainly in basic research besides the references in this article.[2–4] The authors concluded that the existing data show that selective serotonin reuptake inhibitors exert harmful effects on semen quality and rates of DNA fragmentation. This conclusion might almost be acceptable. However, at present, there are insufficient data to clearly show relationships between male fertility factors and antidepressants, because we did not find adequate and decisive results from previous reports. The effects of antidepressants on male infertility are still equivocal and the mechanisms are unclear. In addition, there are so many variations of antidepressants, and each selective serotonin reuptake inhibitor might have their own mechanisms and characteristics, as mentioned in this review.

Although there is still not enough evidence in the content of this review to generally and widely accept the influence of antidepressants on male infertility, this article is significant in that it raises urologists' concerns about the risks of antidepressants in regard to male infertility.

Another interesting point is that this harmful effect seems to be reversible upon cessation. In clinical practice, we sometimes experience reversible testicular function (spermatogenesis) in patients with male hypogonadotropic hypogonadism. Spermatogenesis is often observed in male hypogonadotropic hypogonadism patients when luteinizing hormone and follicle-stimulating hormone treatment are commenced, even after testosterone replacement therapy. A normal to nearly normal testis itself seems to have a high potential for spermatogenesis function.

It is almost impossible to clarify which antidepressants affect which level of spermatogenesis. It is most important that we pay attention to the risks posed to male infertility by antidepressants, and avoid overprescribing antidepressants.

Toshiyasu Amano M.D., Ph.D.
Department of Urology, Nagano Red Cross Hospital, Nagano, Japan
amanot@nagano-med.jrc.or.jp

DOI: 10.1111/iju.14126

Conflict of interest

None declared.

References

1 Beeder LA, Samplaski MK. Effect of antidepressant medications on semen parameters and male fertility. *Int. J. Urol.* 2020; **27**: 39–46.
2 Câmara ML, Almeida TB, de Santi F *et al*. Fluoxetine-induced androgenic failure impairs the seminiferous tubules integrity and increases ubiquitin carboxyl-terminal hydrolase L1 (UCHL1): Possible androgenic control of UCHL1 in germ cell death? *Biomed. Pharmacother.* 2019; **109**: 1126–39.
3 Prasad P, Ogawa S, Parhar IS. Serotonin reuptake inhibitor citalopram inhibits GnRH synthesis and spermatogenesis in the male zebrafish. *Biol. Reprod.* 2015; **93**: 102.
4 Erdemir F, Atilgan D, Firat F, Markoc F, Parlaktas BS, Sogut E. The effect of sertraline, paroxetine, fluoxetine and escitalopram on testicular tissue and oxidative stress parameters in rats. *Int. Braz. J. Urol.* 2014; **40**: 100–8.

Original Research ajog.org

GYNECOLOGY

The effect of antidepressants on fertility

Marianne M. Casilla-Lennon, BS; Samantha Meltzer-Brody, MD; Anne Z. Steiner, MD, MPH

BACKGROUND: Information on the effects of different pharmaceuticals on fertility is sparse. Human and animal models indicate that antidepressant use could have a negative effect on fertility through alteration of levels of the neurosteroid, allopregnanolone.

OBJECTIVE: The objective of this study is to assess the effects of antidepressants on the natural fertility in women.

STUDY DESIGN: A secondary analysis of data from Time to Conceive, a prospective cohort study, was conducted. Women ages 30 to 44 years without a history of infertility, early in their attempts to conceive, were followed with standardized pregnancy testing until pregnancy was detected. Medication use was assessed at enrollment, daily for up to 4 months, and then monthly. For this analysis, discrete time regression models were created to calculate the association between antidepressant use and fecundability. Potential confounders—age, body mass index, caffeine, alcohol use, and education—were included in all models.

RESULTS: Ninety-two (9.6%) of 957 women reported antidepressant use while attempting to conceive. Women taking antidepressants were more likely to be non-Hispanic Caucasian (91% vs 75%, $P < .01$) and to consume alcoholic beverages (74% vs 61%, $P < .01$). Antidepressant use at enrollment had an adjusted fecundability ratio (FR) of 0.86 (95% confidence interval [CI], 0.63—1.20). However, time-varying analyses suggested that antidepressant use in a given cycle is associated with a reduced probability of conceiving in that cycle (adjusted FR, 0.75; 95% CI, 0.53—1.06). After adjusting for history of depression or restricting the analysis to women who reported a history of depression, the association between antidepressant use and decreased fecundability remained [adjusted FR, 0.66 (95% CI, 0.45—0.97) and (adjusted FR, 0.64; 95% CI, 0.43—0.94), respectively].

CONCLUSION: Our data suggest that antidepressants may reduce the probability of a woman with a history of depression to conceive naturally. Future studies are needed to differentiate the extent to which this association is due to the antidepressant itself versus the underlying depression.

Key words: antidepressants, depression, fertility, fecundability

Antidepressants are the first line of treatment for unipolar major depression[1] and a number of other psychiatric disorders, such as obsessive compulsive disorder.[2] In 2011, antidepressants were the most dispensed drug in the United States, accounting for over 260 million prescriptions.[3] Antidepressants were disproportionately used by women,[4] the majority of reproductive age, compared with men.[5] Although there has been significant research on the teratogenicity of antidepressants, little is known about their direct impact on fertility.

The three most commonly prescribed classes of antidepressants increase allopregnanolone,[6] a progesterone derivative considered a neurosteroid that is produced and is neuroactive in the brain. Allopregnanolone is a positive allosteric modulator of the gamma aminobutyric acid ($GABA_A$) receptor. Its presence enhances the activity of GABA, which is the main inhibitory neurotransmitter in the central nervous system, including the hypothalamus.

Research on rodents suggests that an antidepressant-induced increase in allopregnanolone levels could lead to dysregulation of the hypothalamic-pituitary-ovarian (HPO) axis. Incubation of hypothalamic tissue with allopregnanolone suppresses GnRH release in a concentration-dependent manner. The effects are blocked with the addition of a GABA-antagonist, suggesting that allopregnanolone increases GABA activity, suppressing GnRH release.[7] Accordingly, injection of allopregnanolone into the hypothalamus of rats has been shown to decrease circulating LH levels and result in fewer oocytes at oestrus.[6] The results of animal models have been consistent in humans. Intravenous administration of allopregnanolone in healthy, fertile women decreases follicle-stimulating hormone (FSH) and LH levels[8] and subsequently decreases rates of ovulation. This dysregulation in allopregnanolone could inhibit the pulsatile action of GnRH needed to sustain proper synchronization of ovulation, resulting in infertility.

The above evidence suggests that antidepressants may contribute to changes in the HPO axis through their GABAergic action. HPO axis dysfunction can manifest in numerous ways, including anovulation or luteal phase defects, which can negatively affect a woman's ability to conceive. We hypothesized that antidepressant use would impair natural fertility, which we examined by analyzing the effect of antidepressants on fecundability, the probability of conceiving in a menstrual cycle.

Materials and Methods

This study is a secondary analysis of Time to Conceive (TTC), an ongoing prospective time-to-pregnancy cohort study approved by the institutional review board at the University of North Carolina. A detailed description of the TTC study has been published previously.[9] In brief, women were recruited to the study via community-based fliers, informational emails, internet, radio, television and print advertising, and community blogs. English-speaking women 30 to 44 y
been attempting to c
3 months, were elig

Cite this article as: Casilla-Lennon MM, Meltzer-Brody S, Steiner AZ, et al. The effect of antidepressants on fertility. Am J Obstet Gynecol 2016;215:314.e1-5.

0002-9378/$36.00
Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.ajog.2016.01.170

Women with a history of infertility, polycystic ovary syndrome, pelvic inflammatory disease, endometriosis, pelvic radiation, or a partner with a history of infertility were excluded. Our analysis includes women enrolled between April 2008 and July 2015. Information was collected via web-based questionnaires and daily diaries. At enrollment, women provided demographic information, medical and surgical history, obstetric, gynecologic, and menstrual history, information on behaviors, height and weight, and pregnancy history, as well as partner demographics. Women were asked if they had a history of anxiety or depression (yes/no). The question was not restricted to any specific type of depression or anxiety. The Cerner Multum drug database system was embedded to identify and record over-the-counter and prescription medication use. This comprehensive database included dosing information and drug names, both generic and brand names.

While attempting to conceive, women completed an online daily diary for 4 months or until pregnancy was detected. The daily diary collected information on vaginal bleeding, intercourse, pregnancy test results, and medication use. If the woman did not conceive in the first 4 months of enrollment, she completed an online questionnaire monthly thereafter, which also collected information on medication use. The women were followed for twelve months or until pregnancy was achieved. Pregnancy tests (sensitivity = 20 mIU hCG/mL) were provided with standardized instructions. For this analysis, a positive pregnancy test was our primary outcome of interest.

Antidepressants were categorized by class via mechanism of action. A comprehensive list of traditional antidepressants was constructed. Our list included selective serotonin reuptake inhibitors (SSRIs), serotonin/norepinephrine reuptake inhibitors, serotonin antagonist and reuptake inhibitors, norepinephrine reuptake inhibitors, norepinephrine/dopamine reuptake inhibitors, tri and tetracyclic antidepressants (TCAs), monoamine oxidase inhibitors, and other antidepressants. Adjunct antidepressants and other psychotropic medications with antidepressant action were excluded.

Statistical analysis

Women were categorized by antidepressant use (yes/no). Antidepressant use was categorized by 1) use at enrollment, 2) use at any given time during the study, and 3) use in a given menstrual cycle. Bivariate analyses, with χ^2 tests, were used to compare demographics and potential covariates (age, race, education level, pregnancy history, menstrual cycle regularity and length, frequency of intercourse, previous contraceptive use, body mass index [BMI], partner age, smoking status, alcohol use, and caffeine use) between the antidepressant users (at any given time in the study) and non-users.

Discrete-time survival models were constructed to assess the relationship between 1) antidepressant use at enrollment and fecundability and, as antidepressant use of an individual woman could vary over time, 2) antidepressant use in a given cycle and fecundability in that given cycle. All analyses were restricted to the first nine cycles of attempt, due to few women remaining after that time. All models accounted for both the right censoring and left truncation (due to women enrolling in cycles 1, 2, 3, or 4 of their pregnancy attempt) present in the data. In these models, a fecundability ratio (FR) less than 1.0 suggested reduced fecundability.

In all adjusted models we included both the covariates strongly associated with fecundability or antidepressants and those identified in multiple prior studies as related to fecundability. These covariates were age, Caucasian race (yes/no), BMI, education level, and alcohol use (yes/no). Maternal age was collapsed into three categories (30−34, 35−37, and 38−44 years of age), and BMI was categorized into 4 groups ($\leq$18.4, 18.5−24.9, 25.0−29.9, and $\geq$30 kg/m^2). Additional analyses adjusted for history of anxiety/depression as reported in the baseline questionnaire and restricted analysis to those women, who reported a history of anxiety/depression. In an attempt to explore the extent to which underlying anxiety/depression may play a role in the relationship between antidepressants and fertility, we analyzed the relationship between history of anxiety/depression, as reported in the baseline questionnaire and fecundability.

Results

A total of 957 women and 3355 menstrual cycles were included in this analysis. Of the analysis cohort, 70% of women were 30 to 34 years old, 18% were 35 to 37 years old, and 12% were 38 to 44 years old. Participants tended to be Caucasian (76%), highly educated (72% with some graduate education or more), and normal weight (62%). Six hundred twenty-two (65%) of subjects conceived during their first 9 cycles of attempt. Analysis was restricted to the first nine cycles due to high levels of drop-out past this point.

Ninety-two (9.6%) women reported antidepressant use at some point during enrollment. Antidepressant use was reported by 7.6% of women in the first cycle of attempt, 9.0% in the second, 8.9% in the third, 8.2% in the sixth, and 11.0% in the eighth. The most frequently used antidepressants were SSRIs and all antidepressants, except one report of TCAs, were reuptake inhibitors. Women taking antidepressants were more likely to be non-Hispanic Caucasian (91% vs 75%, $P < .01$) and to consume alcoholic beverages (74% vs 61%, $P < .01$). There were no differences in other variables analyzed (Table 1). Menstrual cycle length and history of regular menstrual cycles did not differ at baseline by antidepressant use. Antidepressant users at enrollment reported on average 2.18 [standard deviation (SD) 1.25] acts of intercourse per week, which did not differ significantly from non-users at that time with 2.38 (SD 1.59) act per weeks ($P = .27$. Women reporting a history of depression had similar fecundability to those who did not (adjusted FR, 1.03; 95% CI, 0.82−1.24).

Initial models suggested minimal to no effect of antidepressants on fecundability (Figure 1) when analyzing baseline antidepressant use; women who were taking antidepressants at enrollment had an adjusted FR of 0.86 (95%

CI, 0.63–1.20). When analyzing cycles individually, time varying analyses suggested that antidepressant use in a given cycle was associated with a reduced probability of conceiving in that cycle (adjusted FR, 0.75; 95% CI, 0.53–1.06). After adjusting for a history of anxiety/depression as reported on the baseline questionnaire, the association between antidepressant use and fecundability remained [adjusted FR, 0.66 (95% CI, 0.45–0.97)]. When the analyses were restricted to women with a history of anxiety/depression, the association also remained (adjusted FR, 0.64; 95% CI, 0.43–0.94).

Discussion

In this prospective cohort study we found that 9.6% of women took antidepressants while trying to conceive. Almost all antidepressants used in our study were reuptake inhibitors, sharing a similar mechanism of action. Women in our study who took antidepressants tended to be non-Hispanic Caucasian and more likely to consume alcoholic beverages. History of depression or anxiety was not associated with fecundability. Antidepressant use within a given menstrual cycle was associated with a lower probability of conceiving among women who reported a history of anxiety or depression.

Antidepressant use in our cohort was similar to that in previous studies. A cross-sectional study of 12,637 Americans found antidepressant use in women ages 18 to 39 years to be 9.2%.[4] Another study found similar rates of antidepressant use in this age group at 9.9%.[10] In addition, a study of antidepressant use during pre-conception and early pregnancy reported a rate of 10.0% and was similar in regard to age and race.[11] Retrospective studies relying on reported antidepressant use in IVF recipients reported significantly lower rates than prospective studies.[11-14]

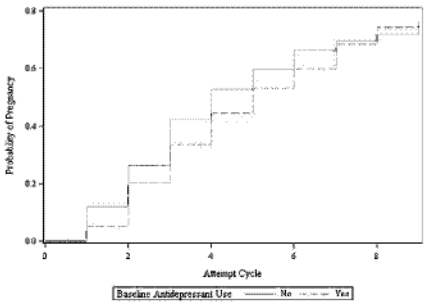
Prior studies reported higher rates of antidepressant use in older women and women of Caucasian race.[4,10] In our study antidepressant users did not differ from non-users by race or age. The age distribution did not significantly differ by antidepressant use. Regarding race distribution, 79% of antidepressant users were Non-Hispanic Caucasian, whereas 76% of non-users were non-Hispanic Caucasian. Our inability to detect previously observed associations may be due to the relatively narrow age range and underrepresentation of minorities in this study.

TABLE 1
Baseline characteristics of women taking antidepressants

Characteristic	Category	Antidepressant taken during study: No (%) *n* = 865	Antidepressant taken during study: Yes (%) *n* = 92	*P*
Age (years)	30–34	611 (70)	64 (69)	.46
	35–37	157 (18)	14 (15)	
	38 and over	97 (11)	14 (15)	
Race	Non-Hispanic Caucasian	652 (75)	84 (91)	<.01
	Other	213 (25)	8 (8)	
Education level	Less than college degree	71 (8)	5 (5)	.77
	College degree	174 (20)	17 (18)	
	Some graduate/masters	79 (9)	11 (12)	
	Completed postgraduate	541 (63)	59 (64)	
Nulligravid	Previous pregnancy	445 (51)	53 (58)	.26
	Never been pregnant	420 (49)	30 (42)	
Regular menstrual cycles	No	122 (14)	13 (14)	1.00
	Yes	742 (86)	79 (86)	
Body mass index (kg/m^2)	Less than 18.5	23 (3)	1 (1)	.50
	18.5–24.9	538 (62)	54 (59)	
	25–29.9	175 (20)	24 (26)	
	30 and over	128 (15)	13 (14)	
Current smoker	No	852 (99)	91 (99)	.75
	Yes	13 (1)	1 (1)	
Partner age (years)	<50	853 (99)	92 (100)	.28
	50 or more	11 (1)	0 (0)	
Past year hormonal contraceptive use	No	474 (55)	44 (48)	.20
	Yes	391 (45)	48 (52)	
Caffeinated beverage consumption	No	163 (19)	15 (16)	.04
	Yes	701 (81)	77 (84)	
Alcohol consumption	No	335 (39)	24 (26)	.01
	Yes	526 (61)	67 (74)	
Average menstrual cycle length (days)	25 or fewer	75 (9)	12 (13)	.28
	26–30	618 (72)	61 (67)	
	30–35	126 (15)	16 (18)	
	36 or more	42 (5)	2 (2)	

Casilla-Lennon et al. The effect of antidepressants on fertility. Am J Obstet Gynecol 2016.

FIGURE 1
Unadjusted Kaplan-Meier curves of time to pregnancy, according to baseline antidepressant use



Casilla-Lennon et al. The effect of antidepressants on fertility. Am J Obstet Gynecol 2016.

We did not find any differences in the reported regularity or length of menstrual cycles between women using and not using antidepressants. If menstrual cycle characteristics had been different between groups, this may have suggested HPO axis dysfunction. However, menstrual cycles tend to be highly variable,[15] and a lack of these findings does not exclude the possibility of HPO axis dysregulation leading to anovulation or luteal phase defects.

Our study found lower fecundability in cycles in which a woman took an antidepressant. This finding was statistically significant after adjusting for history of anxiety or depression or restricting the analysis to those that reported a history of anxiety or depression. To our knowledge, this is the first prospective study to look at the effect of antidepressants on fertility. In a case control study, Grodstein et al[19] found that women who took antidepressants for more than 6 months had 2.9 times the odds of infertility (95% CI, 0.9−8.3). This finding was based on 5 cases who took antidepressants.[19] Klock et al[12] performed a retrospective study to analyze the effect of SSRIs on the success of IVF treatments, reporting that the group of women using SSRIs had a lower pregnancy rate than non-users (40% vs 51%). Their findings were limited by a small sample size and were not statistically significant.[12]

The observed association between antidepressants and fecundability could be due to the antidepressants itself or the underlying fertility. This study was not designed to look at the relationship between depression and fecundability. There was a single question on the enrollment questionnaire that queried about history of underlying depression or anxiety. This simple measure was not associated with fecundability. However, we are unable to determine the extent to which antidepressants users or non-users suffered from depression or residual depressive symptoms. Therefore, this study is unable to truly differentiate the effect of the underlying depression from the antidepressant. Future studies could use validated questionnaires to assess anxiety and depression symptoms.

The true relationship between depression and fertility is unknown. Odegard found in a study of 30,438 married or previously married women in Norway hospitalized for psychiatric disorders, that women with affective psychoses had a similar number of observed children as expected for the general population.[16] This study assessed severe psychiatric conditions and did not specifically stratify the differences in mood disorders alone. Harlow et al[17] found, in a prospective cohort of women going through menopause, that women with a history of depression reported fewer live births. Neither of these studies account for desires to conceive. In a prospective time-to-pregnancy study, Lynch et al[18] used the Hospital Anxiety and Depression Scale to assess the association between depression and fecundability. They did not find an association between depression scores and fecundability. However, depression scores for their cohort were within the normal range; thus they were not able to assess the impact of depressive disorders.[18]

Confounding by indication, in which a disease is associated with the use of a medication and the outcome, is a problem in may many pharmacological studies.[20] A recent study by Andersen et al[21] examined the relationship between antidepressant use and miscarriage using a pharmacy database. The authors found an increased risk of miscarriage in those women taking antidepressants during the first trimester (1.27 HR; 95% CI, 1.22−1.33) compared with women who did not. However, they found a similarly increased risk of miscarriage in those women who stopped the antidepressant prior to conceiving (1.24 HR; 95% CI, 1.18−1.30). This would suggest a potential for confounding by indication with this class of medication. To try to account for the possibility of confounding by indication, we restricted our analysis to women who reported a history of anxiety/depression. However, we did not account for current disease or severity of disease. Future studies may need to collect more information to allow the use of propensity scores (to predict probability of using antidepressants) in analyses.

Our study is not without limitations. By design it includes older reproductive-aged women. While this increases the prevalence of antidepressant use, it may decrease generalizability. Although we identified numerous variables known to affect both depression and fertility, the nature of an observational study precludes us from adjusting for all possible confounders. The outcome we used was an epidemiological measure of fecundability, not incidence of infertility, and the outcome was pregnancy, not live birth. One would presume, however, that decreased fecundability would result in increased rates of infertility. Because our study was not originally designed to assess depression, we relied on a self-report of a history of depression/anxiety. As discussed above, this study cannot differentiate between active or past depression nor between anxiety and depression.

Our study design was the most important strength of our study. A prospective cohort provides the strongest evidence when assessing pharmaceuticals on women attempting to conceive, as randomized controlled trials are not done in this context.[20] Cross-sectional studies are limited by their inability to determine temporal causality. The online self-reported daily diaries allowed for accuracy, as well as privacy, compared with face-to-face interviews, which may be especially important when

assessing sensitive information.[22] To further improve accuracy by avoiding misclassification and monitoring adherence of antidepressants, the Center Multum drug database was used.

In conclusion, antidepressants are among the most widely used medications in reproductive-aged women in the United States. Our study suggests that antidepressant use in women with a history of anxiety or depression diminishes natural fertility. While this is concerning, it is possible that the indication for the antidepressant is the causal factor. Thus, this study alone cannot be used as justification for women to stop their antidepressants when attempting to conceive.

References

1. Davidson JRT. Major depressive disorder treatment guidelines in America and Europe. J Clin Psychiatry 2010;71(Suppl E1):e04.
2. Kellner M. Drug treatment of obsessive-compulsive disorder. Dialogues Clin Neurosci 2010;12:187-97.
3. Lindsley CW. The top prescription drugs of 2011 in the United States: antipsychotics and antidepressants once again lead CNS therapeutics. ACS Chem Neurosci 2012;3:630-1.
4. Pratt L, Brody D, Gu Q. Antidepressant use in persons aged 12 and over: United States, 2005–2008. NCHS data brief, No. 76. Hyattsville, MD: National Center for Health Statistics. 2011 [Internet]. [cited 2014 Jul 15]. Available from: http://www.cdc.gov/nchs/data/databriefs/db76.htm.
5. Farr SL, Bitsko RH, Hayes DK, Dietz PM. Mental health and access to services among US women of reproductive age. Am J Obstet Gynecol 2010;203:542.e1-9.
6. Uzunova V, Wrynn AS, Kinnunen A, Ceci M, Kohler C, Uzunov DP. Chronic antidepressants reverse cerebrocortical allopregnanolone decline in the olfactory-bulbectomized rat. Eur J Pharmacol 2004;486:31-4.
7. Calogero AE, Palumbo MA, Bosboom AM, et al. The neuroactive steroid allopregnanolone suppresses hypothalamic gonadotropin-releasing hormone release through a mechanism mediated by the gamma-aminobutyric acid A receptor. J Endocrinol 1998;158:121-5.
8. Timby E, Hedström H, Bäckström T, Sundström-Poromaa I, Nyberg S, Bixo M. Allopregnanolone, a GABAA receptor agonist, decreases gonadotropin levels in women: a preliminary study. Gynecol Endocrinol Off J Int Soc Gynecol Endocrinol 2011;27:1087-93.
9. Steiner AZ, Herring AH, Kesner JS, et al. Antimüllerian hormone as a predictor of natural fecundability in women aged 30-42 years. Obstet Gynecol 2011;117:798-804.
10. Olfson M, Marcus SC. National patterns in antidepressant medication treatment. Arch Gen Psychiatry 2009;66:848-56.
11. Cooper WO, Willy ME, Pont SJ, Ray WA. Increasing use of antidepressants in pregnancy. Am J Obstet Gynecol 2007;196:544.e1-5.
12. Klock SC, Sheinin S, Kazer R, Zhang X. A pilot study of the relationship between selective serotonin reuptake inhibitors and in vitro fertilization outcome. Fertil Steril 2004;82:968-9.
13. Friedman BE, Rogers JL, Shahine LK, Westphal LM, Lathi RB. Effect of selective serotonin reuptake inhibitors on in vitro fertilization outcome. Fertil Steril 2009;92:1312-4.
14. Van Gelder MMHJ, van Rooij IALM, de Walle HEK, Roeleveld N, Bakker MK. Maternal recall of prescription medication use during pregnancy using a paper-based questionnaire: a validation study in the Netherlands. Drug Saf 2013;36:43-54.
15. Münster K, Schmidt L, Helm P. Length and variation in the menstrual cycle: a cross-sectional study from a Danish county. Br J Obstet Gynaecol 1992;99:422-9.
16. Odegård O. Fertility of psychiatric first admissions in Norway 1936-1975. Acta Psychiatr Scand 1980;62:212-20.
17. Harlow BL, Wise LA, Otto MW, Soares CN, Cohen LS. Depression and its influence on reproductive endocrine and menstrual cycle markers associated with perimenopause: the Harvard Study of Moods and Cycles. Arch Gen Psychiatry 2003;60:29-36.
18. Lynch CD, Sundaram R, Buck Louis GM, Lum KJ, Pyper C. Are increased levels of self-reported psychosocial stress, anxiety, and depression associated with fecundity? Fertil Steril 2012;98:453-8.
19. Grodstein F, Goldman MB, Ryan L, Cramer DW. Self-reported use of pharmaceuticals and primary ovulatory infertility. Epidemiol Camb Mass 1993;4:151-6.
20. Koren G, Nickel S. Sources of bias in signals of pharmaceutical safety in pregnancy. Clin Investig Med Online 2010;33:E349-55.
21. Andersen JT, Andersen NL, Horwitz H, Poulsen HE, Jimenez-Solem E. Exposure to selective serotonin reuptake inhibitors in early pregnancy and the risk of miscarriage. Obstet Gynecol 2014;124:655-61.
22. Perlis TE, Des Jarlais DC, Friedman SR, Arasteh K, Turner CF. Audio-computerized self-interviewing versus face-to-face interviewing for research data collection at drug abuse treatment programs. Addict Abingdon Engl 2004;99:885-96.

Author and article information

From the Department of Obstetrics and Gynecology, University of North Carolina School of Medicine, Chapel Hill, North Carolina (Ms. Casilla-Lennon and Dr. Steiner); Department of Psychiatry, University of North Carolina, Chapel Hill, North Carolina (Dr. Meltzer-Brody).

Received Oct. 26, 2015; revised Jan. 4, 2016; accepted Jan. 20, 2016.

The authors report no conflict of interest.

Supported by grants T35-DK007386 and R01067683 from the NIH/NICHD.

Presented at the 63rd Annual Meeting of the Pacific Coast Reproductive Society; Palm Springs, California, March 2015.

Corresponding author: Anne Z. Steiner, MD, MPH, University of North Carolina, CB#7570, 4001 Old Clinic Bldg., Chapel Hill, NC 27599. asteiner@med.unc.edu

OPEN

PRS GLOBAL OPEN

ORIGINAL ARTICLE

Reconstructive

Trends in Gender-affirming Surgery in Insured Patients in the United States

Megan Lane, BA
Graham C. Ives, BA
Emily C. Sluiter, BS
Jennifer F. Waljee, MD, MPH, MS
Tsung-Hung Yao, BS
Hsou Mei Hu, PhD, MBA, MHS
William M. Kuzon, MD, PhD

Background: An estimated 0.6% of the U.S. population identifies as transgender and an increasing number of patients are presenting for gender-related medical and surgical services. Utilization of health care services, especially surgical services, by transgender patients is poorly understood beyond survey-based studies. In this article, our aim is 2-fold; first, we intend to demonstrate the utilization of datasets generated by insurance claims data as a means of analyzing gender-related health services, and second, we use this modality to provide basic demographic, utilization, and outcomes data about the insured transgender population.

Methods: The Truven MarketScan Database, containing data from 2009 to 2015, was utilized, and a sample set was created using the Gender Identity Disorder diagnosis code. Basic demographic information and utilization of gender-affirming procedures was tabulated.

Results: We identified 7,905 transgender patients, 1,047 of which underwent surgical procedures from 2009 to 2015. Our demographic results were consistent with previous survey-based studies, suggesting transgender patients are on average young adults (average age = 29.8), and geographically diverse. The most common procedure from 2009 to 2015 was mastectomy. Complications of all gender-affirming procedures was 5.8%, with the highest rate of complications occurring with phalloplasty. There was a marked year-by-year increase in utilization of surgical services.

Conclusion: Transgender care and gender confirming surgery are an increasing component of health care in the United States. The data contained in existing databases can provide demographic, utilization, and outcomes data relevant to providers caring for the transgender patient population. *(Plast Reconstr Surg Glob Open 2018;6:e1738; doi: 10.1097/GOX.0000000000001738; Published online 16 April 2018.)*

INTRODUCTION

Identifying as transgender can have varying meanings to different individuals, but is generally understood to describe discordance between one's biologic sex assigned at birth and the gender with which one identifies. An estimated 0.6% of the U.S. population identifies as transgender or gender-nonconforming.[1] Transgender patients face many unique health risks, including an increased risk of suicide, mental health issues, and HIV compared with those in the general population.[2,3] In addition, trans patients are less likely to be insured[2,3] and often find health insurance coverage for transgender medical care to be lacking. The paucity of insurance coverage for hormone therapy and gender-affirming surgery, even among those who are able to secure health insurance, contributes to a high financial burden. Of the respondents in the 2015 U.S. Transgender Survey, 55% of patients who sought surgery and 25% of those on hormones reported difficulty obtaining insurance coverage for these services.[3] In addition to cost, transgender patients report delaying health care for urgent needs and preventative care due to discrimination and disrespect from providers.[3]

Although there are financial and social barriers to health care access for transgender patients in the United States, the need for medical and surgical gender-related care is increasing. The 2015 US Transgender Survey re-

From the Department of Surgery, Division of Plastic Surgery, University of Michigan, Ann Arbor, Mich.

Received for publication November 28, 2017; accepted February 7, 2018.

Ms. Lane and Mr. Ives contributed equally to this work.

DOI: 10.1097/GOX.0000000000001738

Disclosure: *The authors have no financial interest to declare in relation to the content of this article. The Article Processing Charge was paid for by the authors.*

Supplemental digital content is available for this article. Clickable URL citations appear

ported in a sample of over 27,000 transgender and gender-nonconforming Americans, 25% had undergone 1 or more gender-affirming surgeries.[3] Beyond large survey studies such as the U.S. Transgender Survey, there has been little investigation into the utilization of gender-related health services over time, which is a particularly pertinent area of investigation, given both increasing visibility of transgender issues and the ongoing debate regarding health care reform in the United States.

We aim to shed light on basic demographics and trends in utilization of gender-affirming surgeries through the use of insurance claims data available on the Truven MarketScan Database, and provide a model for future investigation into gender-related health care utilizing this study modality. The Truven MarketScan Database is composed of health information from a large cohort of patients with employer-sponsored health insurance between 2009 and 2015. Demographics, types of surgical procedures performed, and utilization over time will be described and information not easily elicited in survey-based studies such as the rates of major complications.

METHODS

Data Source

The Truven MarketScan Database, which includes inpatient and outpatient claims from prescriptions, procedures, and laboratory results from over 122 million unique individuals beginning in 1999, was analyzed.[4] Specifically, the MarketScan Commercial Claims and Encounters Database, which collects information on several million patients annually from 2009 to 2015, was used.[4] These data are collected from over 150 employers and 20 health plans.[4] The database is deidentified, and information about providers and select demographic information is withheld from the database to protect patient privacy.

Database Analysis

A dataset was initially created by pulling patients in the MarketScan Commercial Claims and Encounters Database with the International Code of Diseases, 9th Edition (ICD-9) diagnosis code for Gender Identity Disorder (GID, 302.85) within the 2009–2015 date range (Fig. 1). Patient information collected included age at diagnosis, date of diagnosis within the database range, gender marker, employment status, and location (**see pdf, Supplemental Digital Content 1**, which displays table of variables provided for all patients with GID ICD-9 code, ***http://links.lww.com/PRSGO/A762***). Gender marker was not included in the analysis, given the ability to change gender marker on insurance plans, and is not necessarily reflective of one's biologic sex or gender. To isolate patients within this dataset who underwent gender-affirming surgery, patients with ICD-9 and Current Procedural Terminology, 4th Edition (CPT-4) codes associated with mastectomy, breast augmentation, vaginoplasty, free flaps (for phalloplasty), urethroplasty, orchiectomy, and hysterectomy from 2009 to 2015 in the GID dataset were pulled (Fig. 1). The date of the first instance of any code listed for each procedure was used as the date of surgery. Patients who had more than 1 surgery were only counted once in the analysis for demographic information. As there are no universal CPT codes for phalloplasty, ICD-9 procedure codes for construction of penis, the CPT code for male-to-female intersex surgery, and CPT codes for free flap were used to represent a phalloplasty in this dataset. Free flaps associated with trauma based on ICD-9 diagnosis codes were excluded. Similarly, ICD-9 procedure codes for vaginal construction, the CPT code for female-to-male intersex surgery, and any CPT code related to operations on the vagina were assumed to represent vaginoplasty in this dataset (**see pdf, Supplemental Digital Content 2**, which displays ICD-9 and CPT-4 codes utilized to determine patient procedures, ***http://links.lww.com/PRSGO/A763***; **see pdf, Supplemental Digital Content 3**, which displays table of variables provided for GID patients who underwent gender-affirming surgery, ***http://links.lww.com/PRSGO/A764***). All cases of possible phalloplasty and vaginoplasty were reviewed manually and, in cases of discrepant or incomplete data, a final determination of whether to include a procedure in a particular category was made based on the associated ICD-9 diagnosis codes. All other procedures were identified by their respective ICD-9 and CPT procedure codes. We did not use the ICD-9 procedure code for "sex transformation operation NEC" (645) due to its ambiguity and lack of gender-specificity, and the majority of patients in the dataset also had encounters with the CPT codes associated with genital surgery noted above. From the patients who underwent gender-affirming surgery, information about date of operation, length of stay, and venous thromboembolism (VTE) was collected (**Supplemental Digital Content 3**). Paired *t* tests were utilized to provide descriptive analysis.

RESULTS

Transgender Patient Demographics in the MarketScan Database

A total of 7,905 transgender patients in the Truven MarketScan Commercial Claims and Encounters Database were identified using the GID diagnosis code. The first diagnosis in the dataset is recorded on January 2, 2009, and the last is October 20, 2015 (Fig. 2).

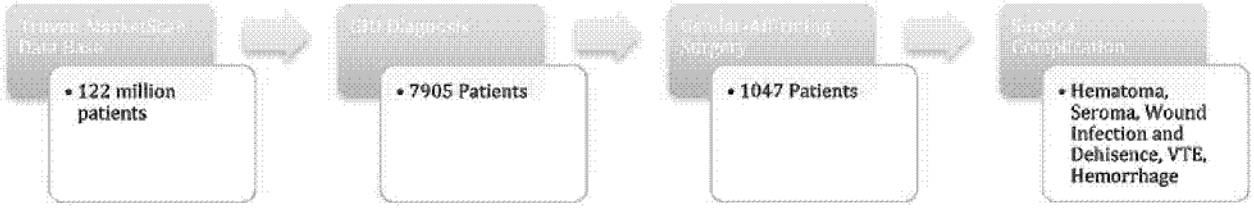


Fig. 1. Outline of dataset creation and isolation of surgical variables.

Age, Income, Employment Status, and Location

The average age of transgender patients within the database was 29.8, and the largest number of transgender patients was in the young adult age range (Fig. 3). The largest income bracket represented in this dataset earned greater than $70,000 a year (50% of patients with available income information), with only 4 patients earning less than $40,000 a year (Table 1). Of the patients where employment data were available, 90% were listed as full-time employees (Table 2), but a variety of employment statuses,

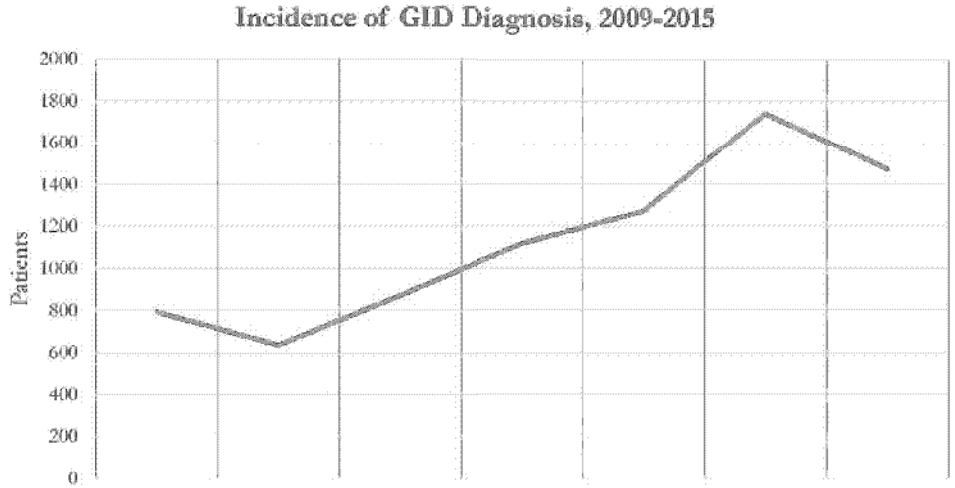
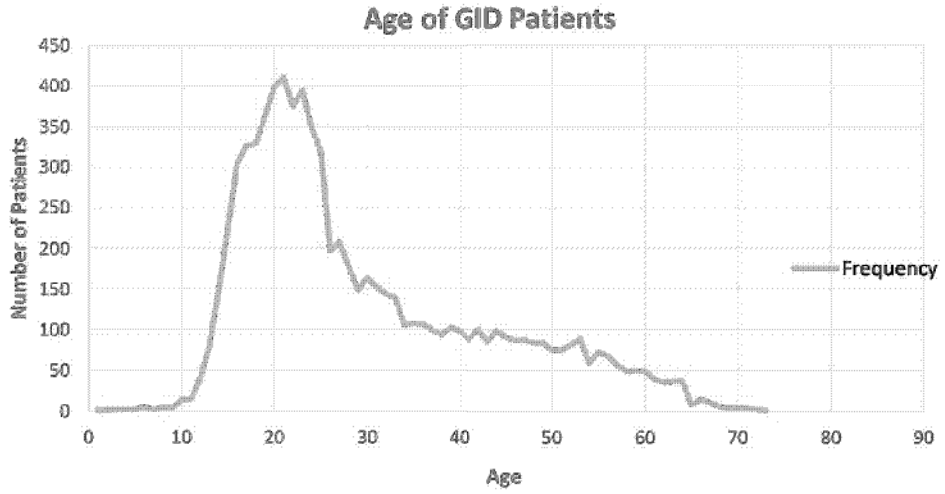


Fig. 2. Number of GID diagnoses from 2009 to 2015.



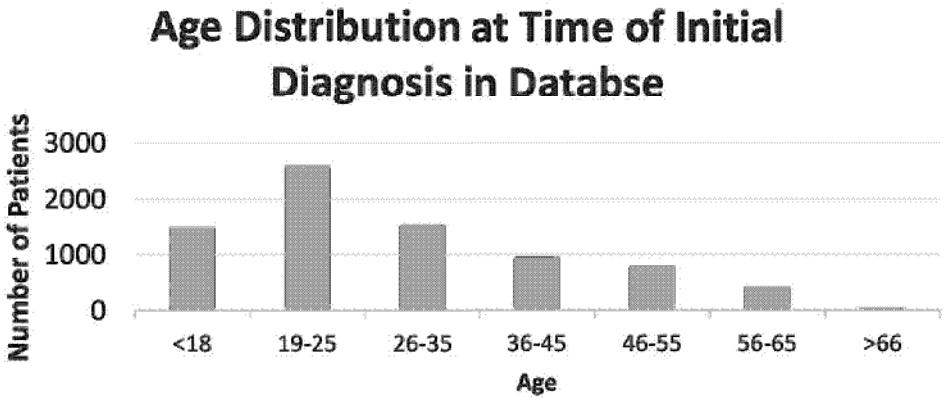


Fig. 3. Age distribution of patients at first GID diagnosis in the MarketScan Database.

including retirees, unemployed, and those relying on COBRA Consolidated Omnibus Budget Reconciliation Act (COBRA) were represented in the dataset. Patients from 49 states were represented in the sample—with the largest number of patients residing in New York, California, and Texas, with 765, 1,285, and 457 patients, respectively. There were no patients from Hawaii in the dataset.

Insured Patients Undergoing Gender-affirming Surgery

A total of 1,047 patients, or 13.2% of the sample, underwent 1 or more gender-affirming surgeries from 2009 to 2015, with the number of patients undergoing surgeries increasing over the sample period (Fig. 4). There were a total of 401 mastectomies, 62 breast augmentations, 60 phalloplasties, 193 vaginoplasties, 189 hysterectomies, and 93 orchiectomies in the sample. Mastectomy was the most common procedure represented within the sample, accounting for 11.7% of all procedures. Phalloplasty was the least common procedure, accounting for 5.7% of all surgical cases.

Age

The average age for any procedure was 29.5. The youngest subset of patients were those undergoing mastectomy,

Table 1. Income of Patients with GID Diagnosis and Patients who Underwent a Gender-affirming Procedure

Median Income	GID Diagnosis	Percentage	GID Diagnosis and Procedure	Percentage
< 40K	4	0.1	0	0.0
40K ≤ Median Inc < 50K	46	0.6	3	0.2
50K ≤ Median Inc < 60K	867	11.0	72	6.8
60K ≤ Median Inc < 70K	2,598	32.9	255	24.4
Median Inc ≥ 70K	3,479	44.0	479	45.7
Missing Median Inc	913	11.5	238	22.7
Total	7,907	100.0	1,047	

Table 2. Employment Status of Patients with GID Diagnosis and Patients who Underwent a Gender-affirming Procedure

Employment Status	GID Diagnosis	Percentage	GID Diagnosis and Procedure	Percentage
Active full time	4,099	51.8	557	53.2
Active part time or seasonal	119	1.5	23	2.2
Early retiree	152	1.9	16	1.5
Medicare eligible retiree	56	0.7	8	0.7
Retiree, employment unknown	54	0.7	2	0.2
COBRA	55	0.7	9	0.9
Long-term disability	11	0.1	2	0.2
Surviving spouse/dependent	8	0.1	1	0.1
Other/unknown	3,353	42.4	429	41.0
Total	7,907	100	1,047	

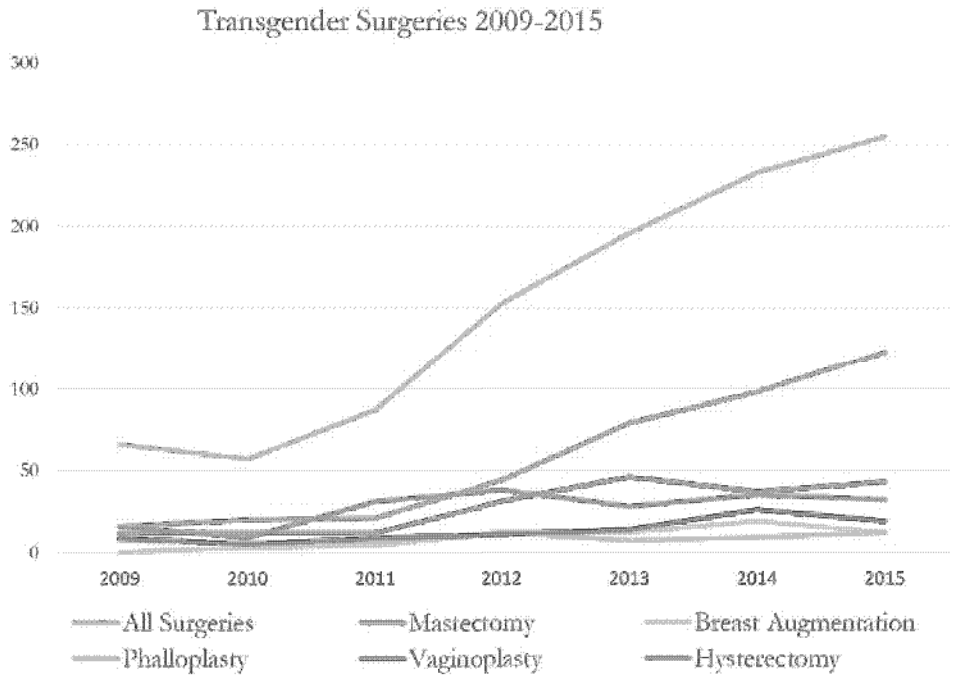


Fig. 4. Number of gender-affirming surgeries by year.

Table 3. List of Procedures by Average Age

Procedure	Average Age
Mastectomy	28.1
Breast augmentation	42.4
Phalloplasty	38.1
Vaginoplasty	40.2
Hysterectomy	31.0
Orchiectomy	37.3

with an average age of 28.1, and those undergoing breast augmentation were the oldest, with an average age of 42.4. With the exception of mastectomy, the average age of patients undergoing all individual procedures were older than the average age of patients with the GID diagnosis in the dataset (Table 3). The youngest patient undergoing gender-affirming surgery in the sample was age 14 at the time of mastectomy and the oldest patient was 76 at the time of vaginoplasty.

Income, Employment Status, and Location

Fifty-nine percentage of individuals, undergoing gender-affirming procedures earned greater than $70,000 a year (Table 1). Like the greater sample, the majority of patients who underwent gender-affirming surgery were employed full time (n = 557, 90% of patients with employment information; Table 2). Surgical patients were located in 45 states and were not present in Montana, Wyoming, North Dakota, Alaska, and Hawaii. California had the largest number of surgical patients (n = 211).

Complications

VTE, hematoma, seroma, wound infection, and wound dehiscence were recorded as complications and found in patients who underwent gender-affirming surgery. In total, 62 complications (5.8% of all procedures) were recorded in the sample, with the most common being wound infection (n = 16; Table 4, Fig. 5). With the exception of mastectomy and breast augmentation where hematoma and hemorrhage were the most common complications respectively, wound infection was the most common complication among all groups undergoing gender-affirming surgery. The rate of complications was the highest in phalloplasty at 0.22 and lowest in orchiectomy at 0.4 (Table 4). There were a total of 4 incidences of VTE in the sample set, 2 following mastectomy and 1 following hysterectomy and vaginoplasty. The rate of VTE in all procedures was 0.003.

DISCUSSION

This article is intended to describe both the demographics of the transgender population in the Unites States and trends in gender-affirming surgery utilizing an insurance claims database. Our results suggest the majority of patients who have a GID diagnosis are young adults with an average age of 29.8. Although the age of transgender patients in the MarketScan dataset were consistent with previous survey-based demographic results,[6] the economic demographics in our sample were skewed toward higher-earning employed patients. Previous survey-based studies report a 35% full-time employment rate among trans individuals with 76% of respondents earning $50,000 or less.[3] Given the use of commercial and employee-based insurance plans and the absence of Medicaid within the MarketScan database, these differences in demographics are expected compared with large, survey-based studies of the general population.

Unlike questionnaire-based studies, the use of insurance claims data allowed for the survey of specific gender-affirming surgeries and associated complication rates. In total, 13.2% of trans patients within the sample underwent gender-affirming procedures from 2009 to 2015, with an increasing number of patients seeking surgical interventions annually. Mastectomy was the most common gender-affirming surgical procedure within the sample. The demographics of patients undergoing gender-affirming surgery were consistent with those in the general transgender population in the MarketScan Database.

The use of a large database also allows for the analysis of surgical complications and outcomes. Of the patients who underwent gender-affirming surgery, 5.9% had complications, with the most common being wound infection. Phalloplasty had the highest rate of complication among gender-affirming procedures, which is consistent with the clinical experience of the authors and their institution. The rate of VTE among the study population was 0.003, which is less than the estimated VTE rate associated with other plastic surgery procedures of 0.5–2%.[8] Complication rates within the Marketscan database were significantly lower than those reported from single-center studies in gender mastectomy,[9,10] and systematic reviews in vaginoplasty[11,12] and phalloplasty.[13] This may represent variation in coding practices leading to an underestimation of complications. Many minor complications may also be treated conservatively and not be captured by diagnosis codes.

Most importantly, our results suggest that the number of patients presenting for gender-related medical and surgical care is on the rise. Because the incidence of gender dysphoria is likely stable, these increasing numbers reflect increasing access to gender-related health services, given social and political shifts. This increase in the number of patients with a GID diagnosis and those seeking medical

Table 4. Complications by Procedure Type and Rates (in Parentheses)

	VTE	Hematoma	Hemorrhage	Seroma	Infection	Wound Disruption	Delayed Wound Healing
Mastectomy	2 (0.005)	8 (0.012)	0	4 (0.010)	4 (0.010)	2 (0.005)	0
Breast augmentation	0	1 (0.016)	2 (0.032)	0	0	0	0
Phalloplasty	0	1 (0.017)	1 (0.017)	0	6 (0.100)	0	5 (0.08)
Vaginoplasty	1 (0.005)	1 (0.005)	1 (0.005)	0	1 (0.005)	2 (0.010)	3 (0.016)
Hysterectomy	1 (0.005)	0	3 (0.016)	1 (0.005)	5 (0.026)	2 (0.010)	0
Orchiectomy	0	1 (0.010)	1 (0.010)	0	0	2 (0.022)	0

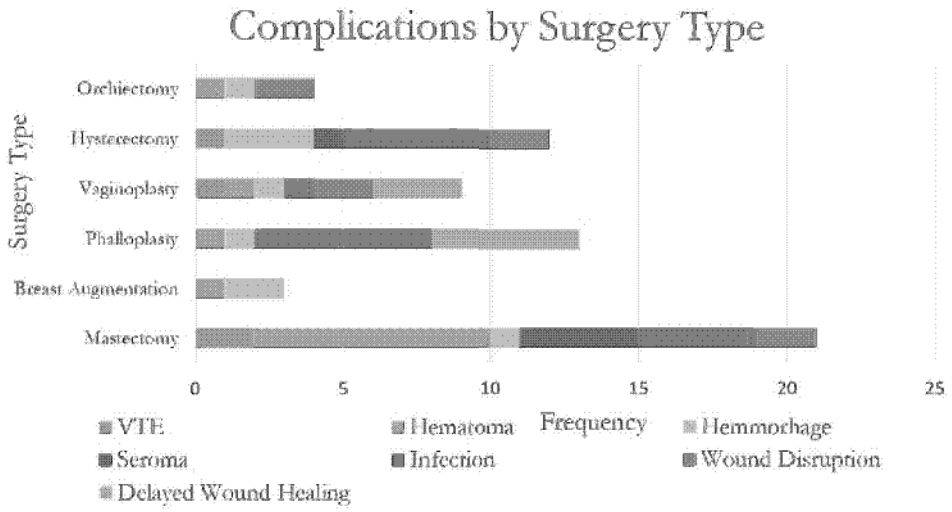


Fig. 5. Complications by procedure.

and surgical intervention is in part thought to be the result of section 1557 of the Affordable Care Act, which bans gender discrimination in health care coverage and was previously interpreted to include transition-related care for transgender patients until December 2016.[7] This possible effect is observed in the increase in patients with the GID diagnosis and number of gender-affirming surgeries in the MarketScan dataset from 2009. With an increasing number of transgender patients presenting for care, an understanding and partnership with the transgender community at both the patient and population level is paramount.

There are multiple limitations to this study stemming from the MarketScan database. The first is the use of the formal diagnosis of GID. Patients who identify as transgender or gender-nonconforming may either not feel comfortable sharing this information with their provider, or may not agree with the concept of the transgender identity as medicalized disorder. The MarketScan database does not capture individuals on Medicaid, and only includes Medicare Advantage plans, which skews both the income and employment of the sample. The data also do not account for individuals who chose to pay for visits and procedures out of pocket. Another major limitation is the variety of CPT and ICD-9 codes that are utilized to bill for gender-affirming procedures, specifically phalloplasty and vaginoplasty. This was controlled for by both utilizing a large number of codes, including those that are nonspecific such as "free-flap," and manually confirming diagnosis by using other associated ICD-9 and CPT codes within a clinical encounter. We have attempted to capture all patients in the dataset by utilizing a broad range of billing codes (**Supplemental Digital Content 2**), but this collection may underestimate the number of surgeries in this dataset. The utilization of an insurance database may underestimate complication rates, given conservative management of minor complications and limitations, given coding variation as mentioned above. MarketScan also does not include information about conditions present on admission, which may also skew estimated complication rates; however, our chosen complications in this analysis are acute in nature rather than chronic sequelae. The study only encompasses a 6-year period and does not allow for long-term follow-up.

This study is the first use of a large, commercially available insurance database to examine the transgender patient population and the use of surgical gender-related services. This article is intended to be a first step into a research modality that could be utilized to study multiple areas of gender-related care, including trends in hormone prescribing patterns, rates of chronic illness in transgender patients, and outcomes of surgical and medical transgender care.

CONCLUSIONS

This study is the first use of a large, commercial database to examine demographics of insured transgender patients and trends in gender-related surgical care. Our results suggest that the number of patients interacting with health care system and those pursuing gender-affirming surgeries has steadily increased over the intended study period. Mastectomy was the most common procedure performed over the 2009–2015 period, and the rates of complication were the highest with phalloplasty. Although there are limitations, such as skewed income and employment status compared with previously surveyed transgender patient populations, database-based studies can be utilized to provide a powerful tool to observe and analyze transgender health trends and outcomes.

William M. Kuzon, MD, PhD
Division of Plastic Surgery
University of Michigan
1201 Taubman Center
1500 E. Medical Center Drive
Ann Arbor, MI 48109
E-mail: wkuzon@med.umich.edu

REFERENCES

1. Flores AR, Herman JL, Gates GJ, et al. *How Many Adults Identify as Transgender in the United States? The Williams Institute.* 2016.

Available at https://williamsinstitute.law.ucla.edu/wp-content/uploads/How-Many-Adults-Identify-as-Transgender-in-the-United-States.pdf. Accessed September 1, 2017.

2. Blechinger D. Understanding the LGBT communities. In: Eckstrand KL, Ehrenfeld JM, eds. *Lesbian, gay, Bisexual, and Transgender Healthcare: A Clinical Guide to Preventative, Specialist, and Preventive Care.* New York, N.Y.: Springer; 2016:3–20.
3. James S, Herman J, Rankin S, et al. 2015 US Transgender survey report on health and healthcare. *National Center for Transgender Equality.* 2016. Available at http://www.transequality.org/sites/default/files/docs/usts/USTS%20Full%20Report%20-%20FINAL%201.6.17.pdf. Accessed March 30, 2017.
4. Hansen LG, Chang S. Health research for the real world: the marketscan databases. Truven Health Analytics White Paper. July 2011. Available at http://truvenhealth.com/portals/0/assets/PH_11238_0612_TEMP_MarketScan_WP_FINAL.pdf. Accessed March 30, 2017.
5. Herman JL, Flores AR, Brown TN, et al. Age of individuals who identify as transgender in the United States. *The Williams Institute.* 2017. Available at https://williamsinstitute.law.ucla.edu/wp-content/uploads/TransAgeReport.pdf. Accessed September 1, 2017.
6. Baker KE. The future of transgender coverage. *New Engl J Med.* 2015. 376:1801–1804. Accessed March 30, 2017. Available at http://www.nejm.org/doi/full/10.1056/NEJMp1702427.
7. Hernandez S, Valdes J, Salama M. Venous thromboembolism prophylaxis in plastic surgery: a literature review. *AANA J.* 2016;84:167–172.
8. Berry MG, Curtis R, Davies D. Female-to-male transgender chest reconstruction: a large consecutive, single-surgeon experience. *J Plast Reconstr Aesthet Surg.* 2012;65:711–719.
9. Cregten-Escobar P, Bouman MB, Buncamper ME, et al. Subcutaneous mastectomy in female-to-male transsexuals: a retrospective cohort-analysis of 202 patients. *J Sex Med.* 2012;9:3148–3153.
10. Lawrence AA. Patient-reported complications and functional outcomes of male-to-female sex reassignment surgery. *Arch Sex Behav.* 2006;35:717–727.
11. Horbach SER, Bouman MB, Maerten JS, et al. Outcome of vaginoplasty in male-to-female transgenders: a systematic review of surgical techniques. *J Sexual Med.* 2015;12:1499–1512.
12. Frey JD, Poudrier G, Chiodo MV, et al. A systematic review of metoidioplasty and radial forearm flap phalloplasty in female-to-male transgender genital reconstruction: is the "ideal" neophallus an achievable goal? *Plast Reconstr Surg Glob Open.* 2016;4:e1131.

Archives of Sexual Behavior (2023) 52:483–494
https://doi.org/10.1007/s10508-022-02472-8

LETTER TO THE EDITOR

Gender-Affirming Treatment of Gender Dysphoria in Youth: A Perfect Storm Environment for the Placebo Effect—The Implications for Research and Clinical Practice

Alison Clayton[1]

Received: 22 March 2022 / Revised: 3 November 2022 / Accepted: 3 November 2022 / Published online: 14 November 2022

Introduction

In the last decade, there has been a rapid increase in the numbers of young people with gender dysphoria (GD youth) presenting to health services (Kaltiala et al., 2020). There has also been a marked change in the treatment approach. The previous "common practice" of providing psychosocial care only to those under 18 or 21 years (Smith et al., 2001) has largely been replaced by the gender affirmative treatment approach (GAT), which for adolescents includes hormonal and surgical interventions (Coleman et al., 2022). However, as a recent review concluded, evidence on the appropriate management of youth with gender incongruence and dysphoria is inconclusive and has major knowledge gaps (Cass, 2022). Previous papers have discussed that the weaknesses of the studies investigating the efficacy of GAT for GD youth mean they are at high risk of bias and confounding and, thus, provide very low certainty evidence (Clayton, 2022a, b; Levine et al., 2022). To date, however, there has been little discussion of the inability of these studies to differentiate specific treatment effects from placebo effects. Of note, the term "placebo effect" is no longer used to just simply refer to the clinical response following inert medication; rather, it describes the beneficial effects attributable to the brain-mind responses evoked by the treatment context rather than the specific intervention (Wager & Atlas, 2015). This Letter argues that the current treatment approach for GD youth presents a perfect storm environment for the placebo effect. This raises complex clinical and research issues that require attention and debate.

✉ Alison Clayton
alclayton@student.unimelb.edu.au

[1] School of Historical and Philosophical Studies, University of Melbourne, Parkville 3010, Australia

A Brief Introduction to the Contemporary Concept of the Placebo Effect

The term "placebo effect" can be used variably by different authors. As recently defined in a consensus statement, placebo (beneficial) and nocebo (deleterious) effects occur in clinical or research contexts and are due to psychobiological mechanisms evoked by the treatment (or research) context rather than any specific effect of the intervention. Importantly, placebo and nocebo effects not only occur during the prescription of placebo (inert) pills, but they can also substantially modulate the efficacy and tolerability of active medical treatments (Evers et al., 2018).

The therapeutic ritual, the encounter between a sick person and a clinician, is a powerful psychosocial event. Clinicians, particularly physicians, are our society's designated healers and their prestige, status, and authority help engender patients' trust and expectations of relief from suffering (Benedetti, 2021a). Positive clinician–patient interactions are associated with decreased anxiety and increased hope. Complex neurobiological mechanisms are implicated in the placebo effect, including release of neurotransmitters (e.g., endorphins, cannabinoids, dopamine, and oxytocin) and activation of specific areas of the brain (e.g., the prefrontal cortex, anterior insula, rostral anterior cingulate cortex, and the amygdala) (Colloca & Barsky, 2020; Kaptchuk & Miller, 2015). These changes are associated with an increased sense of well-being. They also impact on cardiovascular, respiratory, immune, and endocrine functioning, all of which may contribute to patients' clinical improvement (Enck et al., 2013; Wager & Atlas, 2015).

Several unconscious psychological mechanisms, including classical conditioning and social learning, play a role in the placebo effect (Benedetti, 2021a). In clinical trials, where patients communicate with each other, a process of social observational learning may be associated with emotional contagion and, thus, placebo and nocebo effects (Benedetti,

2013). The media and social media may also foster these effects and contribute to the dissemination of symptoms and illness throughout the general population (Colloca & Barsky, 2020).

Expectation of outcome is a principal mechanism of the placebo effect and anything that increases patients' expectations is potentially capable of boosting placebo effects (Evers et al., 2018). Although research has demonstrated that changes in physiological parameters may occur following placebo administration (Wager & Atlas, 2015), these response expectations have been particularly noted in patient-reported outcomes, such as anxiety, pain, life satisfaction, and mood. Expectations and cognitive readjustment can lead to behavioral changes, such as resuming normal daily activities, which can be observer rated. Physicians' status, whether through the general position given to them in society or through individual personality factors, may contribute to such expectations of benefit. This type of phenomenon has sometimes been termed prestige suggestion. The "Hawthorne effect" describes the phenomenon where clinical trial patients' improvements may occur because they are being observed and given special attention. A patient who is part of a study, receiving special attention, and with motivated clinicians, who are invested in the benefits of the treatment under study, is likely to have higher expectations of therapeutic benefits (Benedetti, 2021a).

Placebo-induced improvements are real and can be robust and long lasting (Benedetti, 2021b; Wager & Atlas, 2015). Individual patient factors, such as personality and genetics, may be associated with placebo responsiveness (Benedetti, 2021a). The particular illness is also relevant. For example, although placebo treatment can impact symptoms of cancer, there is no evidence that placebos can shrink tumors (Benedetti, 2021b; Kaptchuk & Miller, 2015). However, there is evidence that placebos can act as long-lasting and effective treatments for depression and various pain conditions, such as migraine and osteoarthritic knee pain (Kam-Hansen et al., 2014; Kirsch, 2019; Previtali et al., 2021). Further, some research suggests adherence to placebo medication, particularly in cardiac disease, may be associated with reduced mortality (Wager & Atlas, 2015).

The Research Setting versus the Clinical Practice Setting

Research into new medical treatments aims to control for placebo effects, and this helps ensure true evaluation of the treatment's efficacy (Enck et al., 2013). The double-blind randomized controlled trial (DBRCT), although not perfect, is the current gold standard for determining the efficacy and safety of a treatment. The DBRCT study design evolved over several centuries and became widely accepted practice in the mid-twentieth century (Lilienfeld, 1982). Of note, the term "blind" is thought to have originated in eighteenth-century France when blindfolds were used to disprove Anton Mesmer's "animal magnetism" theory and the mesmerism craze of that era (Kaptchuk, 1998). Well-designed DBRCTs minimize the impact of bias, confounding and placebo effects on findings and are the best type of study for determining whether there is a causal relationship between an intervention and an effect (Enck et al., 2013; Kabisch et al., 2011).

The reader may wonder about this requirement of differentiating placebo effects from the specific effects of an intervention and ask: If the patient improves, does it really matter why? Yes, it does, particularly for treatments that have significant risk of adverse effects. There are also broader problems raised by relying on the placebo effect. Consider prescribing antibiotics for viral infections. The patient may experience clinical benefit through a placebo effect. However, not only may some patients experience serious adverse drug reactions, but the health of the whole population is imperiled by the problem of antibiotic resistance (Llor & Bjerrum, 2014). Furthermore, informed consent is an ethical pillar of modern medicine and requires clinician honesty and transparency. Clinicians deceptively utilizing placebo treatments do not meet this requirement (Barnhill, 2012; Kaldjian & Pilkington, 2021). A medical profession that does little to distinguish placebo effects from specific treatment effects risks becoming little different from pseudoscience and the quackery that dominated medicine of past times, with likely resulting decline in public trust and deterioration in patient outcomes (Benedetti, 2021a).

Ideally, in evidence-based medicine, a new treatment undergoes rigorous research and has reasonable evidence of benefit prior to being introduced as routine treatment (although ongoing further research often continues). Clinicians can then reasonably harness and enhance the placebo effect to improve outcomes (Enck et al., 2013). A placebo effect enhancing clinical setting, in which warm and empathic clinicians provide supportive and attentive health care, creates a "therapeutic bias" in patients, giving them hope and expectation of improvement. This is "legitimate" so long as it is done without deception and in a manner consistent with informed consent, trust, and transparency (Kaptchuk & Miller, 2015).

This ideal of clinical interventions having solid evidence of efficacy before being introduced as routine practice is not always a reality. Sometimes, it is more of a situation where the "cart" of clinical practice precedes the "horse" of rigorous research evidence. Then, this catch-up research may be undertaken in a placebo effect-enhancing clinical environment, rather than a placebo effect-controlled research environment. Such situations, especially when DBRCT are not possible, present the researcher and clinician with complex research and clinical

conundrums. Some of these will now be explored using the example of the treatment of youth with gender dysphoria.

A Brief Introduction to the Gender-Affirming Treatment Model for Children and Adolescents with Gender Dysphoria

Gender dysphoria is a term used to describe the distress that is frequently felt by people whose sense of gender is incongruent with their natal sex (these people may also self-identify as transgender) and if the dysphoria is intense and persistent, alongside several other features, a DSM-5 diagnosis of gender dysphoria may be made (American Psychiatric Association, 2013). There has been a sharp rise in the numbers of children and adolescents identifying as transgender and being diagnosed with gender dysphoria (Kaltiala et al., 2020; Tollit et al., 2021; Wood et al., 2013). Many are natal sex females presenting in adolescence, and many have neurodevelopmental and psychiatric disorders (Kaltiala-Heino et al., 2018; Tollit et al., 2021; Zucker, 2019). International guidelines and child and adolescent gender clinics (CAGCs) commonly endorse a gender affirmative treatment approach (GAT) (Coleman et al., 2022; Hembree et al., 2017; Olson-Kennedy et al., 2019; Telfer et al., 2018). Key components of GAT include affirmation of a youth's stated gender identity, facilitation of early childhood social transition, provision of puberty blockers to prevent the pubertal changes consistent with natal sex, and use of cross-sex hormones (CSH) and surgical interventions to align physical characteristics with gender identity (Ehrensaft, 2017; Rosenthal, 2021). This Letter's discussion focuses primarily on the medical (puberty blockers and cross-sex hormones) and surgical elements of GAT.

GAT can achieve some of the desired masculine or feminine appearance outcomes, but the main arguments used to support the use of these treatments in GD youth are that they improve short- and long-term mental health and quality-of-life outcomes. However, this claim is only underpinned by low-quality (mostly short-term, uncontrolled, observational) studies, which provide very low certainty evidence, complemented by expert opinion (Clayton, 2022a; Hembree et al., 2017; NICE, 2020a b; Rosenthal, 2021). No randomized controlled trials (RCTs), including none using the previous treatment approach as a comparative, have been undertaken. This low-quality evidence for the efficacy of GAT is of particular concern given the potential risks associated with GAT.

Risks of Gender-Affirming Medical and Surgical Treatments

Impaired fertility is a risk of cross-sex hormones, and the extent of reversibility of this is unclear (Cheng et al., 2019; Hembree et al., 2017). If puberty blockers are commenced in early puberty and followed by cross-sex hormones, there are no proven methods of fertility preservation (Bangalore Krishna et al., 2019). Surgeries, such as gonadectomies and most genital surgeries, will result in permanent sterility. These impaired fertility and sterility outcomes are important because, firstly, as Cheng et al. (2019) reported, the widespread assumption that many transgender people do not want to have biological children is not supported by several recent studies. Secondly, children as young as ten, who do not have capacity for informed consent, are starting a treatment course that will likely render them infertile or sterile and this raises complex bioethical issues (Baron & Dierckxsens, 2021).

Other adverse effects of GAT are based on a more uncertain evidence base. I provide a brief outline of some of the areas of concern. Cross-sex hormones are associated with cardiovascular health risks, such as thromboembolic, coronary artery, and cerebrovascular diseases (Hembree et al., 2017; Irwig, 2018). Cross-sex hormones may also increase the risk of certain cancers (Hembree et al., 2017; Mueller & Gooren, 2008). Puberty blockers may have negative impact on bone mineral density, which may not be fully reversible, with an associated risk of osteoporosis and fractures (Biggs, 2021; Hembree et al., 2017). Recently, findings from animal studies have increased concerns that puberty blockers may negatively and irreversibly impact brain development due to critical time-windows of brain development. In one study on rams, long-term spatial memory deficits induced by use of puberty blockers in the peripubertal period were found to persist into adulthood (Hough et al., 2017). For those young patients who undertake surgery, there are also the risks of surgical complications (Akhavan et al., 2021). One understudied outcome of mastectomies, for those who later want to and can become pregnant, is the grief about inability to breast feed.

Puberty blockers, cross-sex hormones and genital surgery also pose risks to sexual function, particularly the physiological capacity for arousal and orgasm. It is important to be aware there is a dearth of research studying the impact of GAT on GD youth's sexual function, but I provide a brief discussion of this important topic. Estrogen use in transwomen is associated with decreased sexual desire and erectile dysfunction and testosterone for transmen may lead to vaginal atrophy and dyspareunia (Hembree et al., 2017). It seems widely assumed that testosterone simply improves transmen's sexual functioning. However, placebo-controlled studies from the non-transgender population indicate the situation is likely more complex. For example, studies indicate that testosterone may impact female sexual desire in a bell-shape curve manner, and at high levels may have no benefit or even have negative impact on sexual function (Krapf & Simon, 2017; Reed et al., 2016). Also of note, in medical conditions that are associated with high testosterone levels, such as polycystic ovarian syndrome, impaired sexual function

(e.g., arousal, lubrication, sexual satisfaction, and orgasm) has been reported (Pastoor et al., 2018).

Recently, surgeon and WPATH president-elect, Marci Bowers, raised concern that puberty blockers given at the earliest stages of puberty to birth sex males, followed by cross-sex hormones and then surgery, might adversely impact orgasm capacity because of the lack of genital tissue development (Ley, 2021). One study has reported that some young adults, who had received puberty blockers, cross-sex hormones and laparoscopic intestinal vaginoplasty, self-reported orgasmic capacity (Bouman et al., 2016). However, this finding does not negate Bower's concerns, as it did not make any assessment of the correlation between Tanner stage at initiation of puberty blockers with orgasm outcome. Of note, some of the patients in the study were over the age of 18 at start of GAT. Further, its findings do not apply to those undergoing penile skin inversion vaginoplasty. Importantly, Bouman et al. found that 32% of their participants self-reported being sexually inactive and only 52% reported having had neovaginal penetrative sex more than once. A recent literature review on sexual outcomes in adults post-vaginoplasty noted the paucity of high-quality evidence but reported that "up to 29% of patients may be diagnosed with a sexual dysfunction due to associated distress with a sexual function disturbance" (Schardein & Nikolavsky, 2022). Another recent systematic review of vaginoplasty reported an overall 24% post-surgery rate of inability to achieve orgasm (Bustos et al., 2021).

Coleman et al. (2022) claimed that "longitudinal data exists to demonstrate improvement in romantic and sexual satisfaction for adolescents receiving puberty suppression, hormone treatment and surgery." However, the supporting citation requires scrutiny. Bungener et al. (2020) was a cross-sectional study of 113 young adults, 66% of whom were transmen (most who had undergone mastectomy and gonadectomy, not genital surgery). For its claims of post-surgery increases in sexual experience, it relied on recall of pre-surgical experiences. This means it is at high risk of recall bias, especially given surgery was undertaken up to 5 years (mean 1.5 years) prior to assessment. Further, it focused on sexual experiences, which might naturally be expected to increase as adolescents enter young adulthood, and there was no evaluation of sexual function domains, such as arousal, orgasm, or pain. The study did report current sexual satisfaction but failed to compare this to pre-surgical functioning (or to the Dutch peer comparison group). Thus, it is unable to demonstrate whether sexual satisfaction improved following GAT. On the three questions about sexual satisfaction (frequency, how good sex feels, and sex life in general), 59 to 73% were reportedly moderately to very satisfied. This would appear to mean that 27 to 41% were not satisfied, which is a sizeable minority. Importantly, these sexual satisfaction questions had an approximately 45% missing data rate—an issue not discussed by the authors. This means the authors' conclusion that the majority was satisfied with their sex life is at high risk of bias. Of additional note, at the post-surgical assessment time these young transgender adults were significantly less sexually experienced than their Dutch peers. Thus, in sum, this study provides little reassurance about the sexual function outcomes of GAT in GD youth.

Lastly, in terms of risks, there are increasing reports of discontinuation of hormone treatments, regret and detransition in young people who have received GAT (Boyd et al., 2022; Hall et al., 2021; Littman, 2021; Vandenbussche, 2022). Two recent studies have relied on pharmaceutical prescription records, both using 2018 as the end date of data collection (Roberts et al., 2022; van der Loos et al., 2022). Their reported rates of discontinuation varied widely. For the US cohort, Roberts et al. (2022) reported, for those who had started CSH treatment before age 18, a 4-year CSH discontinuation rate of 25%. For the Dutch cohort, van der Loos et al. (2022) reported on CSH discontinuation rates in adolescents, evaluated according to the "meticulous" Dutch protocol, who had commenced puberty blockers before age 18. People "assigned female at birth" had a CSH discontinuation rate of 1% at a median of 2.3-years follow-up, and those "assigned male at birth" had a 4% discontinuation rate at median 3.5-years follow-up. Previous research from this Dutch group has indicated that average time to detransition was over 10 years (Wiepjes et al., 2018). Thus, given the van der Loos et al. (2022) study's short median follow-up time and young follow-up age (median 19.2 for people "assigned female at birth" and 20.2 for "assigned male at birth"), it seems likely that these discontinuation rates will increase over time. It is also concerning to note that 75% of the Dutch youth who discontinued CSH had undergone gonadectomies, but at follow-up they were receiving neither CSH nor sex hormones consistent with their birth sex.

Ongoing Research

Currently, several large long-term observational studies are underway which involve collecting and analyzing data on patients receiving routine GAT at CAGCs (Olson-Kennedy et al., 2019; Tollit et al., 2019). The aims of these studies are to provide the urgently needed rigorous empirical data to bolster the weak evidence base that currently underpins the GAT approach. However, as discussed above, it is critical to note that this type of observational research is prone to bias, confounding, and lacks ability to distinguish treatment effects from placebo effects (Fanaroff et al., 2020; Pocock & Elbourne, 2000). Thus, it is unlikely to provide the rigorous empirical data that can convincingly demonstrate a causal relationship between treatment and outcome.

Further, there seems to be a problematic tension between the research and clinical agendas of CAGCs. GAT is being

provided in a clinical environment that maximizes the placebo effect. This is the same environment in which the same clinicians are researching GAT's efficacy. As previously discussed, while a placebo effect-enhancing environment may be appropriate for a clinical environment, it is far from an ideal treatment efficacy research environment, particularly when DBRCTs are not possible and RCTs are not undertaken. In the next section, I delve more deeply into exploring this issue. First, however, I will take a brief detour with an example that illustrates the risks when expert opinion and low-quality evidence are relied on as a basis for medical interventions.

A Recent Example from Medical History of the Dangers of Medical Advice Based on Weak Evidence: The Iatrogenic Tragedy of Prone Infant Sleep Position and Sudden Infant Death Syndrome

Gender medicine clinicians and researchers have consistently stated that RCTs would be unethical (de Vries et al., 2011; Smith et al., 2001; Tollit et al., 2019). However, as Valenstein (1986) discussed in his study of the history of lobotomy, the ethics of implementing new treatments without a rigorous evidence base also need to be considered. The harm that can be done by well-intentioned, but erroneous medical advice based on prestigious physicians' clinical judgment without an adequate evidence base can be illustrated by infant sleep position and sudden infant death syndrome (SIDS). Prior to the middle of the twentieth century, it was common practice for mothers to place infants on their backs to sleep (Högberg & Bergström, 2000). The influential pediatrician, Benjamin Spock, was an early advocate of the prone position (front sleeping) for infants. He recommended it in his popular book, Baby and Childcare, from the 1956 edition through until 1985 (Gilbert et al., 2005). This recommendation, that became widespread, was mainly based on clinical wisdom that such a position reduced risk of death from aspiration of vomit and had additional benefits such as decreased crying and reduced head flattening. Early research appeared to support this clinical advice. However, by the 1980s, more rigorous research demonstrated that the prone position increased risk of SIDS. Then medical advice gradually changed to strongly recommending infant supine (back) sleeping. A marked drop in SIDS rates followed. Several biases (e.g., the healthy adopter bias and observer bias) are thought to have contributed to the erroneous clinical belief that prone sleeping position was the safest position. It has been estimated that between the 1950s and the 1990s the infant prone sleeping advice, recommended by well-meaning clinicians and prestigious medical organizations, may have contributed to the deaths of tens of thousands of infants (Gilbert et al., 2005; Sperhake et al., 2018).

Gender-Affirming Treatment for Youth with Gender Dysphoria: A Perfect Storm for Placebo Effect

The reader may ask: Why focus on GAT for GD youth? Is GAT any different from other contemporary medical treatments that also are not underpinned by rigorous evidence? I would reply—indeed, this is an issue in other areas of medicine. For example, the response rate in the placebo groups in antidepressant medication clinical trials is known to be high (Benedetti, 2021a). However, in contrast to GAT, we know this because there have been many RCTs comparing antidepressants to placebos. A recent review, that included placebo in the network meta-analysis, found that all the antidepressants under review were more efficacious than placebo in adults with major depressive disorder (Cipriani et al., 2018). This finding has been challenged by some who argue that the benefits of antidepressants beyond placebo effect seem to be minimal (Jakobsen et al., 2020). However, one of the key points to make is that placebo effect in antidepressant medication response is at least known about and discussed by many researchers, clinicians, and their patients (personal clinical experience), rather than not considered at all, as seems to be the situation to date for GAT for GD youth. Gender medicine clinicians and researchers might take note of a recent meta-analysis of antidepressants in pediatric populations, which recommended that the influence of placebo response needs to be considered in pediatric clinical trial design and implementation (Feeney et al., 2022). Furthermore, it seems particularly vital to consider the potential role of placebo effect in GAT outcomes because the stakes are high. Medical and surgical GAT, being given to vulnerable minors, lead to life-long medicalization and hold the risk of serious irreversible adverse impacts, such as sterility and impaired sexual function. Thus, we need strong evidence that they are as efficacious for critical mental health outcomes as claimed and that there are no less harmful alternatives.

In the field of GD youth medicine, there is a combination of features that seems to create a perfect storm setting for placebo effect. Thus, we have a population of vulnerable youth presenting with a condition, which has no objective diagnostic tests, and that is currently undergoing an unexplained rapid increase in prevalence and marked change in patient demographics. The treatment response is mainly based on patient-reported outcomes (yes, this can be the case for other conditions but remember we are considering the combination of features, not just a feature in isolation).

Some clinicians, who may be affiliated with prestigious institutions, enthusiastically promote GAT, including on the media, social media, and alongside celebrity patients. Some make overstated claims about the strength of evidence and the certainty of benefits of GAT, including an emphasis on their "life-saving" qualities, and underacknowledge the risks. Alternative psychosocial treatment approaches are sometimes denigrated as harmful and unethical conversion practices or as "doing nothing." This combination of features increases the likelihood that there will be a complex interplay of heightened placebo and nocebo effects in this area of medicine, with significant implications for research and clinical practice. Some examples of these types of issues are now provided.

Overstatement of the Certainty of Benefits and Under-Acknowledgment of Risks

Some professional organizations and leading GAT clinicians, in publicly available communications to GD youth, the public, and policy makers, appear to overstate the certainty of GAT's benefits and provide inadequate discussions of risks (Clayton, 2022a; Cohen, 2021a, b; Olson-Kennedy, 2015, 2019; Telfer, 2019, 2021). For example, GATs have been described in such communications as "absolutely life-saving" (Olson-Kennedy, 2015) and being underpinned by "robust scientific research" (Telfer, 2019). It is notable that these same clinicians in their peer reviewed publications acknowledge the sparse empirical evidence with critical knowledge gaps (Olson-Kennedy et al., 2019), and the urgent need for more evidence for this relatively new treatment approach (Tollit et al., 2019). Thus, there seems to be a kind of Janus-faced narrative, with a placebo effect-enhancing face of overstated certainty/strong evidence of benefit displayed to GD youth, their families, and policy makers, and the more realistic face of uncertainty/lack of evidence turned toward peer reviewers and the research community. Of note, several publications in the peer review literature that have made overstated claims about GAT have recently required correction (Bränström & Pachankis, 2020; Pang et al., 2021; Zwickl et al., 2021).

The Dangers of an Exaggerated Suicide Narrative

A specific issue that is important to discuss is the repeated claims that GD youth are at high suicide risk and that GAT reduces this risk. Parents report being told by clinicians that their child will suicide if a trans identity is not affirmed (Jude, 2021). Clinicians' public statements also indicate this message is being given, or at least implied, to parents and young people (Cohen, 2021b; Marchiano, 2017). A recent editorial in *The Lancet* stated puberty blockers reduce suicidality and to remove access to them was to "deny" life (The Lancet, 2021). However, there is no robust empirical evidence that puberty blockers reduce suicidality or suicide rates (Biggs, 2020; Clayton et al., 2021). The authors of the paper, that was the basis for *The Lancet's* claim, subsequently clarified that they were not making any causal claims that puberty blockers decreased suicidality (Rew et al., 2021). Another paper, claiming to have found that barriers to gender-affirming care was associated with suicidality, had to withdraw this claim in a significant correction to their article (Zwickl et al., 2021).

Furthermore, the suicidality of GD youth presenting at CAGCs, while markedly higher than non-referred samples, has been reported to be relatively similar to that of youth referred to generic child and adolescent mental health services (Carmichael, 2017; de Graaf et al., 2022; Levine et al., 2022). A recent study reported that 13.4% of one large gender clinic's referrals were assessed as high suicide risk (Dahlgren Allen et al., 2021). This is much less than conveyed by the often cited 50% suicide attempt figure for trans youth (Tollit et al., 2019). A recent analysis found that, although higher than population rates, transgender youth suicide (at England's CAGS) was still rare, at an estimated 0.03% (Biggs, 2022).

Of course, any elevated suicidality and suicide risk is of concern, and any at risk adolescent should be carefully assessed and managed by expert mental health professionals. However, an excessive focus on an exaggerated suicide risk narrative by clinicians and the media may create a damaging nocebo effect (e.g., a "self-fulfilling prophecy" effect) whereby suicidality in these vulnerable youths may be further exacerbated (Biggs, 2022; Carmichael, 2017). This type of risk has been discussed in other similar situations involving youth (Abrutyn et al., 2020; Canetto et al., 2021; Shain & AAP COMMITTEE ON ADOLESCENCE, 2016).

An Excessively Negative Portrayal of the Previous Standard and Current Alternative Treatment Options

Clinicians and groups advocating for GAT tend toward framing any non-affirming treatment approaches as harmful, ineffective, and unethical, and sometimes equate psychotherapeutic approaches with conversion practices (Ashley, 2022). However, others argue that there are a range of contemporary therapeutic approaches which are not "affirmative," but neither are they conversion practices (D'Angelo et al., 2021). Such approaches can include: Careful assessment and diagnostic formulation, appropriate treatment of co-existing psychological conditions, supportive and educative individual/family psychological care, group therapy, developmentally informed gender exploratory psychotherapy, trauma-informed psychotherapy, and a non-promotion of early childhood social transition (sometimes labeled under the umbrella term of "watchful-waiting," which should not be interpreted as "doing nothing") (D'Angelo et al., 2021; de

Vries & Cohen-Kettenis, 2012; Hakeem, 2012; Kozlowska et al., 2021; Lemma, 2021).

It is important to note that psychotherapeutic approaches for this group of patients are also based on limited evidence. More research into their efficacy is required. One critical consideration here seems to be that ethical psychological approaches do not hold the same adverse risk profiles as do the hormonal and surgical treatments (Baron & Dierckxsens, 2021).

Recently, two Scandinavian youth gender services have drawn similar conclusions and instigated a more cautious approach to hormonal treatments for GD minors, placing a higher emphasis on psychological care (Kaltiala-Heino, 2022; Socialstyrelsen, 2022). Furthermore, in England, the Cass Review into the country's youth gender services has released its interim report (Cass, 2022). In response, the National Health Service's "Interim Service Specification" for GD youth specialist services has specified that the primary intervention for youth will be psychosocial support and psychological interventions. A cautious approach to social transition is recommended and puberty blockers will only be available in the context of a formal research protocol (National Health Service, 2022).

Given all this, it is hard to accept the claims that GAT is prima facie the best treatment model for today's cohort of GD youth. Furthermore, the unwarranted negative portrayal of contemporary psychotherapeutic approaches likely creates nocebo effects and undermines the possibility of providing such ethical care to GD youth (Kozlowska et al., 2021).

Clinicians' Media and Social Media Promotion of Gender Affirmative Treatment

There is intense media and social media coverage of "trans youth" issues. Some surgeons are promoting their gender-affirming surgeries on social media platforms that are popular with young adolescents (Ault, 2022). Some clinicians encourage the celebratory media coverage of GAT, stating it may empower young trans people to seek GAT (Pang et al., 2020). They largely dismiss concerns that the identified association between positive media stories and increased referral rates to CAGCs may be indicative of a social contagion phenomenon. This is despite the reports of the sudden emergence of gender dysphoria, especially in adolescence, and its association with social influence (Kaltiala-Heino, 2022; Littman, 2018, 2021; Marchiano, 2017). Gender clinicians also condemn and have attempted to prevent what they consider as excessively negative media coverage of GAT (although others judge it as reasonable and balanced) (Australian Press Council, 2021; Pang et al., 2022). These clinicians are likely correct that critical media coverage of GAT could negatively impact referrals to gender clinics and might upset some patients. However, a deliberate strategy of promoting an unbalanced celebratory GAT narrative through the media and social media could contribute to social contagion and placebo effects.

What is the right balance? The Australian Press Council's judgment on a clinician's complaints about what she considered as excessively negative press coverage may, arguably, provide an example of some balance on these matters. Of note, while some of the complaints were upheld, many were not and it was judged that "even medical treatment accepted as appropriate by a specialist part of the medical profession is open to examination and criticism…needs to be debated… and was sufficiently justified in the public interest" (Australian Press Council, 2021).

The Exclusive Promotion of Gender-Affirming Treatments within Child and Adolescent Gender Clinics

There is indication of an unbalanced promotion of a celebratory GAT narrative occurring within CAGCs, where, simultaneously, there is a deep enmeshment of the clinical, advocacy, and research agendas. This has already partly been discussed in the sections above, but one detailed example is provided. The Trans20 study is a prospective cohort study based on children and adolescents seen at Melbourne's Royal Children's Hospital Gender Clinic (RCHGC) which provides a GAT model of care (Telfer et al., 2018; Tollit et al., 2019). It is important to highlight that this study's human research ethics committee (HREC) approval was not for the treatment approach, which was implemented as routine clinical care, rather it was for matters such as collection and storage of data, and longitudinal follow-up of discharged patients.

In 2019, an amended HREC approval was granted, allowing a "newsletter blog" to be sent to patients and families with the aim of improving patient engagement with the study. This change was described as raising no new ethical issues. This "first ever" newsletter asked for help with the Trans20 survey completion (Royal Children's Hospital, 2019). This research request was placed amid positive accounts of the service and its patients. For example, following attendance at the clinic's single session assessment triage (SSNac) young people were described as feeling "empowered…and more likely to start living as their preferred gender," and having improvements in mental health and quality of life. A colorful diagram showed the increased rates of social transition that followed SSNac attendance, and the section concluded "Hopefully the improvements after SSNac are a taste of things to come!" One pro-GAT parent/carer support network, that also fundraises for the RCHGC, was spotlighted. There was a "lived experience" piece in which a well-known transitioned, now young adult, patient was pictured receiving an award. This patient provided a personal testimony of the clinic's medical director: She "will always be one of my biggest heroes…an incredible person: Intelligent, compassionate and strong."

This newsletter's communication raises much to think about. The point I want to make here is that sandwiching the requests for a research survey completion between celebratory accounts of the clinic seems likely to magnify the impact of bias and placebo effect on research outcome findings. For example, consider the likely impact on patient bias (patients wanting to please the clinician by giving positive reports), response bias (patients with positive experiences of the clinic more likely to complete the surveys), social learning/contagion, prestige suggestion, and the Hawthorne phenomenon. Furthermore, consider this newsletter as part of the whole therapeutic ritual, enhancing the psychological and neurobiological placebo mechanisms. Apart from this research impact, we can also wonder whether such a newsletter is ideal clinical practice. In my opinion, there are problems. Think, for example, of the young GD patient who may be hesitant to transition. Where is the celebration of this young person's choices? Communications from the clinic, such as this newsletter, may contribute to feelings that, unless he/she transitions, he/she lacks courage (having not been "empowered") and that he/she will never be an award-winning celebrated patient. This may act as a covert form of pressure on patients to transition or, for those who do not, act as a nocebo effect negatively impacting their psychological outcomes.

Where to From Here?

There are no easy solutions to the complex research and clinical issues presented in this Letter. Here, I present a few ideas to stimulate discussion. The first step would seem to be more professional awareness and debate. Independent reviews by expert clinicians and methodologists, not currently involved in clinical practice and research in this area (thus, having some emotional distance and minimizing intellectual conflict risk), could helpfully advise further research and clinical strategies. England's Cass Review is an example of this type of approach (Cass, 2022).

Clinicians should also make measured and honest statements to patients, families, policy makers, and the public about the evidence for GAT's benefits. Placebo effects could also be noted in the limitations section of any research papers. In addition, in public discourse, the media and clinicians could present not only celebratory transition stories, but also: Realistic positive stories of those with gender dysphoria who have decided not to transition or have delayed transition until maturity; accounts of patients who have benefitted from ethical psychological approaches; and accounts of those who have had negative transition experiences. Detransition, regret, and harm from transition should be acknowledged and publicized as a significant risk. A recent paper detailing the elements of a comprehensive informed consent process is timely and important (Levine et al., 2022). However, while a comprehensive informed consent process is vital, it does not address the issue of how the whole ethos of a clinic and the media/social media milieu may act to influence young patients and their families and undermine the capacity for true informed consent.

Conclusion

In conclusion, this Letter has noted that although GAT for GD youth lacks a rigorous evidence base, it is undertaken as routine medical treatment in a strongly placebo effect enhancing environment. It is within this environment that research into its effectiveness is being undertaken. One consideration raised by this relates to clinical practice: When does such a strongly placebo effect enhancing environment meet optimal clinical practice standards? When, if at all, does it veer into the territory of unethical practice that involves deception and undue influence? This Letter has also highlighted that such a placebo effect enhancing environment presents grave problems for research (particularly non-DBRCT research). It seems unlikely that the current research being undertaken in this field will be able to untangle benefits that are due to the placebo effect from those due to the interventions' specific effectiveness. Thus, especially given the adverse risk profile of the hormonal and surgical interventions, it may be that yet again well-intentioned physicians are engaging in medical practices that cause more harm than benefit (Clayton, 2022b). The research and clinical conundrums presented in this Letter have no easy answers. However, as a first step, there is an urgent need for more awareness of the placebo effect and for rigorous and thoughtful debate over how best to proceed in research and clinical practice in this area of medicine.

Acknowledgements I thank the three peer reviewers, the Editor, and several of my colleagues for their thoughtful and helpful comments on earlier drafts of this paper.

Funding No funding was received for this paper.

Declarations

Conflict of interest The author is affiliated with the Society for Evidence-based Gender Medicine. There are no financial conflicts to interest to declare.

References

Abrutyn, S., Mueller, A. S., & Osborne, M. (2020). Rekeying cultural scripts for youth suicide: How social networks facilitate suicide diffusion and suicide clusters following exposure to suicide. *Society and Mental Health, 10*(2), 112–135. https://doi.org/10.1177/2156869319834063

Akhavan, A. A., Sandhu, S., Ndem, I., & Ogunleye, A. A. (2021). A review of gender affirmation surgery: What we know, and what we need to know. *Surgery, 170*(1), 336–340. https://doi.org/10.1016/j.surg.2021.02.013

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596.

Ashley, F. (2022). Interrogating gender-exploratory therapy. *Perspectives on Psychological Science*. https://doi.org/10.1177/17456916221102325

Ault, A. (2022, July 11). Gender surgeons on TikTok, Instagram: Appropriate or not? *Medscape*. https://medscape.com.au.

Australian Press Council. (2021). *Adjudication 1799, Complainant/Australian*. https://presscouncil.org.au.

Bangalore Krishna, K., Fuqua, J. S., Rogol, A. D., Klein, K. O., Popovic, J., Houk, C. P., Charmandari, E., Lee, P. A., Freire, A. V., Ropelato, M. G., Yazid Jalaludin, M., Mbogo, J., Kanaka-Gantenbein, C., Luo, X., Eugster, E. A., Klein, K. O., Vogiatzi, M. G., Reifschneider, K., Bamba, V., … Medina Bravo, P. G. (2019). Use of gonadotropin-releasing hormone analogs in children: Update by an international consortium. *Hormone Research in Paediatrics, 91*(6), 357–372. https://doi.org/10.1159/000501336

Barnhill, A. (2012). Clinical use of placebos: Still the physician's prerogative? *The Hastings Center Report, 42*(3), 29–37. https://doi.org/10.1002/hast.33

Baron, T., & Dierckxsens, G. (2021). Two dilemmas for medical ethics in the treatment of gender dysphoria in youth. *Journal of Medical Ethics*, *48*, 603–607. https://doi.org/10.1136/medethics-2021-107260

Benedetti, F. (2013). Responding to nocebos through observation: Social contagion of negative emotions [Commentary]. *Pain, 154*(8), 1165. https://doi.org/10.1016/j.pain.2013.05.012

Benedetti, F. (2021a). *Placebo effects: Understanding the mechanisms in health and disease* (3rd ed.). Oxford: Oxford University Press.

Benedetti, F. (2021b). Placebos and movies: What do they have in common? *Current Directions in Psychological Science, 30*(3), 274–279. https://doi.org/10.1177/09637214211003892

Biggs, M. (2020). Puberty blockers and suicidality in adolescents suffering from gender dysphoria [Letter to the Editor]. *Archives of Sexual Behavior, 49*(7), 2227–2229. https://doi.org/10.1007/s10508-020-01743-6

Biggs, M. (2021). Revisiting the effect of GnRH analogue treatment on bone mineral density in young adolescents with gender dysphoria. *Journal of Pediatric Endocrinology & Metabolism, 34*(7), 937–939. https://doi.org/10.1515/jpem-2021-0180

Biggs, M. (2022). Suicide by clinic-referred transgender adolescents in the United Kingdom [Letter to the Editor]. *Archives of Sexual Behavior, 51*(2), 685–690. https://doi.org/10.1007/s10508-022-02287-7

Bouman, M. B., van der Sluis, W. B., van Woudenberg Hamstra, L. E., Buncamper, M. E., Kreukels, B., Meijerink, W., & Mullender, M. G. (2016). Patient-reported esthetic and functional outcomes of primary total laparoscopic intestinal vaginoplasty in transgender women with penoscrotal hypoplasia. *Journal of Sexual Medicine, 13*(9), 1438–1444. https://doi.org/10.1016/j.jsxm.2016.06.009

Boyd, I., Hackett, T., & Bewley, S. (2022). Care of transgender patients: A general practice quality improvement approach. *Healthcare, 10*(1). https://doi.org/10.3390/healthcare10010121

Bränström, R., & Pachankis, J. E. (2020). Reduction in mental health treatment utilization among transgender individuals after gender-affirming surgeries: A total population study. *American Journal of Psychiatry, 177*(8), 727–734. https://doi.org/10.1176/appi.ajp.2019.19010080

Bungener, S. L., de Vries, A., Popma, A., & Steensma, T. D. (2020). Sexual experiences of young transgender persons during and after gender-affirmative treatment. *Pediatrics, 146*(6), e20191411. https://doi.org/10.1542/peds.2019-1411

Bustos, S. S., Bustos, V. P., Mascaro, A., Ciudad, P., Forte, A. J., Del Corral, G., & Manrique, O. J. (2021). Complications and patient-reported outcomes in transfemale vaginoplasty: An updated systematic review and meta-analysis. *Plastic and Reconstructive Surgery. Global Open, 9*(3), e3510. https://doi.org/10.1097/GOX.0000000000003510

Canetto, S. S., Antonelli, P., Ciccotti, A., Dettore, D., & Lamis, D. A. (2021). Suicidal as normal: A lesbian, gay, and bisexual youth script? *Crisis, 42*(4), 292–300. https://doi.org/10.1027/0227-5910/a000730

Carmichael, P. (2017*). Meeting the needs of gender diverse children and young people with mental health difficulties*. ACAMH Conference. https://soundcloud.com/user-664361280/dr-polly-carmichael-developments-and-dilemmas.

Cass, H. (2022). *The Cass Review: Independent review of gender identity services for children and young people: Interim report*. https://cass.independent-review.uk.

Cheng, P. J., Pastuszak, A. W., Myers, J. B., Goodwin, I. A., & Hotaling, J. M. (2019). Fertility concerns of the transgender patient. *Translational Andrology and Urology, 8*(3), 209–218. https://doi.org/10.21037/tau.2019.05.09

Cipriani, A., Furukawa, T. A., Salanti, G., Chaimani, A., Atkinson, L. Z., Ogawa, Y., Leucht, S., Ruhe, H. G., Turner, E. H., Higgins, J., Egger, M., Takeshima, N., Hayasaka, Y., Imai, H., Shinohara, K., Tajika, A., Ioannidis, J., & Geddes, J. R. (2018). Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: A systematic review and network meta-analysis. *Lancet, 391*(10128), 1357–1366. https://doi.org/10.1016/S0140-6736(17)32802-7

Clayton, A. (2022a). Commentary on Levine et al.: A tale of two informed consent processes. *Journal of Sex & Marital Therapy*. https://doi.org/10.1080/0092623X.2022a.2070565

Clayton, A. (2022b). The gender affirmative treatment model for youth with gender dysphoria: A medical advance or dangerous medicine? [Letter to the Editor]. *Archives of Sexual Behavior, 51*(2), 691–698. https://doi.org/10.1007/s10508-021-02232-0

Clayton, A., Malone, W. J., Clarke, P., Mason, J., & D'Angelo, R. (2021). Commentary: The signal and the noise-questioning the benefits of puberty blockers for youth with gender dysphoria-a commentary on Rew et al. (2021). *Child and Adolescent Mental Health, 27*(3), 259–262. https://doi.org/10.1111/camh.12533

Cohen, J. (Producer). (2021a). *Australian story: A balancing act/Michelle Telfer* [Television broadcast]. Australian Broadcasting Commission. http://abc.net.au.

Cohen, J. (2021b, 24 May). *Michelle Telfer has been a lifesaving advocate for hundreds of trans children. But her work doesn't come without controversy*. http://abc.net.au.

Coleman, E., Radix, A. E., Bouman, W. P., Brown, G. R., de Vries, A. L. C., Deutsch, M. B., Ettner, R., Fraser, L., Goodman, M., Green, J., Hancock, A. B., Johnson, T. W., Karasic, D. H., Knudson, G. A., Leibowitz, S. F., Meyer-Bahlburg, H. F. L., Monstrey, S. J., Motmans, J., Nahata, L., … Arcelus, J. (2022). Standards of care for the health of transgender and gender diverse people, version 8. *International Journal of Transgender Health, 23*(sup1), S1–S259. https://doi.org/10.1080/26895269.2022.2100644

Colloca, L., & Barsky, A. J. (2020). Placebo and nocebo effects. *New England Journal of Medicine, 382*(6), 554–561. https://doi.org/10.1056/NEJMra1907805

Dahlgren Allen, S., Tollit, M. A., McDougall, R., Eade, D., Hoq, M., & Pang, K. C. (2021). A waitlist intervention for transgender young people and psychosocial outcomes. *Pediatrics, 148*(2), e2020042762. https://doi.org/10.1542/peds.2020-042762

D'Angelo, R., Syrulnik, E., Ayad, S., Marchiano, L., Kenny, D. T., & Clarke, P. (2021). One size does not fit all: In support of psychotherapy for gender dysphoria [Letter to the Editor]. *Archives of Sexual Behavior, 50*(1), 7–16. https://doi.org/10.1007/s10508-020-01844-2

de Graaf, N. M., Steensma, T. D., Carmichael, P., VanderLaan, D. P., Aitken, M., Cohen-Kettenis, P. T., de Vries, A., Kreukels, B., Wasserman, L., Wood, H., & Zucker, K. J. (2022). Suicidality in clinic-referred transgender adolescents. *European Child & Adolescent Psychiatry, 31*(1), 67–83. https://doi.org/10.1007/s00787-020-01663-9

de Vries, A. L. C., & Cohen-Kettenis, P. T. (2012). Clinical management of gender dysphoria in children and adolescents: The Dutch approach. *Journal of Homosexuality, 59*(3), 301–320. https://doi.org/10.1080/00918369.2012.653300

de Vries, A. L. C., Steensma, T. D., Doreleijers, T. A., & Cohen-Kettenis, P. T. (2011). Puberty suppression in adolescents with gender identity disorder: A prospective follow-up study. *Journal of Sexual Medicine, 8*(8), 2276–2283. https://doi.org/10.1111/j.1743-6109.2010.01943.x

Ehrensaft, D. (2017). Gender nonconforming youth: Current perspectives. *Adolescent Health, Medicine and Therapeutics, 8*, 57–67. https://doi.org/10.2147/AHMT.S110859

Enck, P., Bingel, U., Schedlowski, M., & Rief, W. (2013). The placebo response in medicine: Minimize, maximize or personalize? *Nature Reviews. Drug Discovery, 12*(3), 191–204. https://doi.org/10.1038/nrd3923

Evers, A., Colloca, L., Blease, C., Annoni, M., Atlas, L. Y., Benedetti, F., Bingel, U., Büchel, C., Carvalho, C., Colagiuri, B., Crum, A. J., Enck, P., Gaab, J., Geers, A. L., Howick, J., Jensen, K. B., Kirsch, I., Meissner, K., Napadow, V., … Kelley, J. M. (2018). Implications of placebo and nocebo effects for clinical practice: Expert consensus. *Psychotherapy and Psychosomatics, 87*(4), 204–210. https://doi.org/10.1159/000490354

Fanaroff, A. C., Califf, R. M., Harrington, R. A., Granger, C. B., McMurray, J., Patel, M. R., Bhatt, D. L., Windecker, S., Hernandez, A. F., Gibson, C. M., Alexander, J. H., & Lopes, R. D. (2020). Randomized trials versus common sense and clinical observation: JACC review topic of the week. *Journal of the American College of Cardiology, 76*(5), 580–589. https://doi.org/10.1016/j.jacc.2020.05.069

Feeney, A., Hock, R. S., Fava, M., Hernández Ortiz, J. M., Iovieno, N., & Papakostas, G. I. (2022). Antidepressants in children and adolescents with major depressive disorder and the influence of placebo response: A meta-analysis. *Journal of Affective Disorders, 305*, 55–64. https://doi.org/10.1016/j.jad.2022.02.074

Gilbert, R., Salanti, G., Harden, M., & See, S. (2005). Infant sleeping position and the sudden infant death syndrome: Systematic review of observational studies and historical review of recommendations from 1940 to 2002. *International Journal of Epidemiology, 34*(4), 874–887. https://doi.org/10.1093/ije/dyi088

Hakeem, A. (2012). Psychotherapy for gender identity disorders. *Advances in Psychiatric Treatment, 18*(1), 17–24. https://doi.org/10.1192/apt.bp.111.009431

Hall, R., Mitchell, L., & Sachdeva, J. (2021). Access to care and frequency of detransition among a cohort discharged by a UK national adult gender identity clinic: Retrospective case-note review. *British Journal of Psychiatry, Open, 7*(6), e184. https://doi.org/10.1192/bjo.2021.1022

Hembree, W. C., Cohen-Kettenis, P. T., Gooren, L., Hannema, S. E., Meyer, W. J., Murad, M. H., Rosenthal, S. M., Safer, J. D., Tangpricha, V., & 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*(11), 3869–3903. https://doi.org/10.1210/jc.2017-01658

Högberg, U., & Bergström, E. (2000). Suffocated prone: The iatrogenic tragedy of SIDS. *American Journal of Public Health, 90*(4), 527–531. https://doi.org/10.2105/ajph.90.4.527

Hough, D., Bellingham, M., Haraldsen, I. R., McLaughlin, M., Robinson, J. E., Solbakk, A. K., & Evans, N. P. (2017). A reduction in long-term spatial memory persists after discontinuation of peripubertal GnRH agonist treatment in sheep. *Psychoneuroendocrinology, 77*, 1–8. https://doi.org/10.1016/j.psyneuen.2016.11.029

Irwig, M. S. (2018). Cardiovascular health in transgender people. *Reviews in Endocrine & Metabolic Disorders, 19*(3), 243–251. https://doi.org/10.1007/s11154-018-9454-3

Jakobsen, J. C., Gluud, C., & Kirsch, I. (2020). Should antidepressants be used for major depressive disorder? *BMJ Evidence-Based Medicine, 25*(4), 130. https://doi.org/10.1136/bmjebm-2019-111238

Jude. (2021). *From theory to reality: When ROGD hits your family*. Genspect Conferences. https://genspect.org.

Kabisch, M., Ruckes, C., Seibert-Grafe, M., & Blettner, M. (2011). Randomized controlled trials: Part 17 of a series on evaluation of scientific publications. *Deutsches Arzteblatt International, 108*(39), 663–668. https://doi.org/10.3238/arztebl.2011.0663

Kaldjian, L. C., & Pilkington, B. C. (2021). Why truthfulness is the first of the virtues. *American Journal of Bioethics: AJOB, 21*(5), 36–38. https://doi.org/10.1080/15265161.2021.1906991

Kaltiala, R., Bergman, H., Carmichael, P., de Graaf, N. M., Egebjerg Rischel, K., Frisén, L., Schorkopf, M., Suomalainen, L., & Waehre, A. (2020). Time trends in referrals to child and adolescent gender identity services: A study in four Nordic countries and in the UK. *Nordic Journal of Psychiatry, 74*(1), 40–44. https://doi.org/10.1080/08039488.2019.1667429

Kaltiala-Heino, R., Bergman, H., Työläjärvi, M., & Frisén, L. (2018). Gender dysphoria in adolescence: Current perspectives. *Adolescent Health, Medicine and Therapeutics, 9*, 31–41. https://doi.org/10.2147/AHMT.S135432

Kaltiala-Heino, R. (2022). Adolescent identity. *Gender: A Wider Lens Podcast Pioneer Series,* episode 62. https://podcasts.apple.com.

Kam-Hansen, S., Jakubowski, M., Kelley, J. M., Kirsch, I., Hoaglin, D. C., Kaptchuk, T. J., & Burstein, R. (2014). Altered placebo and drug labeling changes the outcome of episodic migraine attacks. *Science Translational Medicine, 6*(218), 218ra5. https://doi.org/10.1126/scitranslmed.3006175

Kaptchuk, T. (1998). Intentional ignorance: A history of blind assessment and placebo controls in medicine. *Bulletin of the History of Medicine, 72*(3), 389–433.

Kaptchuk, T. J., & Miller, F. G. (2015). Placebo effects in medicine. *New England Journal of Medicine, 373*(1), 8–9. https://doi.org/10.1056/NEJMp1504023

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

Kozlowska, K., McClure, G., Chudleigh, C., Maguire, A. M., Gessler, D., Scher, S., & Ambler, G. R. (2021). Australian children and adolescents with gender dysphoria: Clinical presentations and challenges experienced by a multidisciplinary team and gender service. *Human Systems, 1*(1), 70–95. https://doi.org/10.1177/26344041211010777

Krapf, J. M., & Simon, J. A. (2017). A sex-specific dose-response curve for testosterone: Could excessive testosterone limit sexual

interaction in women? *Menopause, 24*(4), 462–470. https://doi.org/10.1097/GME.0000000000000863

Lemma, A. (2021). *Transgender identities: A contemporary introduction*. London: Routledge.

Levine, S. B., Abbruzzese, E., & Mason, J. M. (2022). Reconsidering informed consent for trans-identified children, adolescents, and young adults. *Journal of Sex & Marital Therapy, 48*(7), 706–727. https://doi.org/10.1080/0092623X.2022.204622

Ley, D. (2021). *Does affirmative treatment impair sexual response in trans youth? Sexual health and orgasmic naivete in trans youth*. https://www.psychologytodaycom/us/blog/women-who-stray/202111/does-affirmative-treatment-impair-sexual-response-in-trans-youth.

Lilienfeld, A. M. (1982). Ceteris paribus: The evolution of the clinical trial. *Bulletin of the History of Medicine, 56*(1), 1–18.

Littman, L. (2018). Parent reports of adolescents and young adults perceived to show signs of a rapid onset of gender dysphoria. *PLoS ONE, 13*(8), e0202330. https://doi.org/10.1371/journal.pone.0202330

Littman, L. (2021). Individuals treated for gender dysphoria with medical and/or surgical transition who subsequently detransitioned: A survey of 100 detransitioners. *Archives of Sexual Behavior, 50*(8), 3353–3369. https://doi.org/10.1007/s10508-021-02163-w

Llor, C., & Bjerrum, L. (2014). Antimicrobial resistance: Risk associated with antibiotic overuse and initiatives to reduce the problem. *Therapeutic Advances in Drug Safety, 5*(6), 229–241. https://doi.org/10.1177/2042098614554919

Marchiano, L. (2017). Outbreak: On transgender teens and psychic epidemics. *Psychological Perspectives: A Quarterly Journal of Jungian Thought, 60*(3), 345–366. https://doi.org/10.1080/00332925.2017.1350804

Mueller, A., & Gooren, L. (2008). Hormone-related tumors in transsexuals receiving treatment with cross-sex hormones. *European Journal of Endocrinology, 159*(3), 197–202. https://doi.org/10.1530/EJE-08-0289

National Health Service England. (2022, 20 October). *Interim service specification: Specialist service for children and young people with gender dysphoria*. http://engage.england.nhs.uk.

National Institute for Health and Care Excellence (NICE). (2020a). *Evidence review: Gonadotrophin releasing hormone analogues for children and adolescents with gender dysphoria*. http://evidence.nhs.uk.

National Institute for Health and Care Excellence (NICE). (2020b). *Evidence review: Gender affirming hormones for children and adolescents with gender dysphoria*. http://evidence.nhs.uk.

Olson-Kennedy, J. (2015). *The future of trans care in the new millennium*. Gender Infinity Annual Conference. https://youtu.be/pO8v-tztSg.

Olson-Kennedy, J. (2019, January 17). Treating trans youth with Dr Johanna Olson-Kennedy. *Gender GP Medical Podcast*.

Olson-Kennedy, J., Chan, Y. M., Garofalo, R., Spack, N., Chen, D., Clark, L., Ehrensaft, D., Hidalgo, M., Tishelman, A., & Rosenthal, S. (2019). Impact of early medical treatment for transgender youth: Protocol for the longitudinal, observational trans youth care study. *JMIR Research Protocols, 8*(7), e14434. https://doi.org/10.2196/14434

Pang, K. C., de Graaf, N. M., Chew, D., Hoq, M., Keith, D. R., Carmichael, P., & Steensma, T. D. (2020). Association of media coverage of transgender and gender diverse issues with rates of referral of transgender children and adolescents to specialist gender clinics in the UK and Australia. *JAMA Network Open, 3*(7), e2011161. https://doi.org/10.1001/jamanetworkopen.2020.11161

Pang, K. C., Giordano, S., Sood, N., & Skinner, S. R. (2021). Regret, informed decision making, and respect for autonomy of trans young people. *Lancet. Child & Adolescent Health, 5*(9), e34–e35. https://doi.org/10.1016/S2352-4642(21)00236-4

Pang, K. C., Hoq, M., & Steensma, T. D. (2022). Negative media coverage as a barrier to accessing care for transgender children and adolescents. *JAMA Network Open, 5*(2), e2138623. https://doi.org/10.1001/jamanetworkopen.2021.38623

Pastoor, H., Timman, R., de Klerk, C., Bramer, M. W., Laan, E. T., & Laven, J. S. (2018). Sexual function in women with polycystic ovary syndrome: A systematic review and meta-analysis. *Reproductive Biomedicine Online, 37*(6), 750–760. https://doi.org/10.1016/j.rbmo.2018.09.010

Pocock, S. J., & Elbourne, D. R. (2000). Randomized trials or observational tribulations? *New England Journal of Medicine, 342*(25), 1907–1909. https://doi.org/10.1056/NEJM200006223422511

Previtali, D., Merli, G., Di Laura Frattura, G., Candrian, C., Zaffagnini, S., & Filardo, G. (2021). The long-lasting effects of "placebo injections" in knee osteoarthritis: A meta-analysis. *Cartilage, 13*(1_suppl), 185S–196S. https://doi.org/10.1177/1947603520906597.

Reed, B. G., Bou Nemer, L., & Carr, B. R. (2016). Has testosterone passed the test in premenopausal women with low libido? A systematic review. *International Journal of Women's Health, 8*, 599–607. https://doi.org/10.2147/IJWH.S116212

Rew, L., Young, C. C., Monge, M., & Bogucka, R. (2021). Response: "The signal and the noise"-a response to Clayton et al. (2021). *Child and Adolescent Mental Health*. https://doi.org/10.1111/camh.12534

Roberts, C. M., Klein, D. A., Adirim, T. A., Schvey, N. A., & Hisle-Gorman, E. (2022). Continuation of gender-affirming hormones among transgender adolescents and adults. *Journal of Clinical Endocrinology and Metabolism, 107*(9), e3937–e3943. https://doi.org/10.1210/clinem/dgac251

Rosenthal, S. M. (2021). Challenges in the care of transgender and gender-diverse youth: An endocrinologist's view. *Nature Reviews Endocrinology, 17*(10), 581–591. https://doi.org/10.1038/s41574-021-00535-9

Royal Children's Hospital. (2019). *Ethics amendment request* (Review Reference: AM/42773/RCHM-2019-183067(v1)). Obtained via freedom of information request no. FOI20200490.

Schardein, J. N., & Nikolavsky, D. (2022). Sexual functioning of transgender females post-vaginoplasty: Evaluation, outcomes and treatment strategies for sexual dysfunction. *Sexual Medicine Reviews, 10*(1), 77–90. https://doi.org/10.1016/j.sxmr.2021.04.001

Shain, B., & AAP COMMITTEE ON ADOLESCENCE. (2016). Suicide and suicide attempts in adolescents. *Pediatrics, 138*(1), e20161420. https://doi.org/10.1542/peds.2016-1420.

Smith, Y. L., van Goozen, S. H., & Cohen-Kettenis, P. T. (2001). Adolescents with gender identity disorder who were accepted or rejected for sex reassignment surgery: A prospective follow-up study. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*(4), 472–481. https://doi.org/10.1097/00004583-200104000-00017

Socialstyrelsen, The National Board of Health and Welfare. (2022). *Care of children and adolescents with gender dysphoria*. https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/artikelkatalog/kunskapss-tod/2022-3-7799.pdf.

Sperhake, J., Jorch, G., & Bajanowski, T. (2018). The prone sleeping position and SIDS. Historical aspects and possible pathomechanisms. *International Journal of Legal Medicine, 132*(1), 181–185. https://doi.org/10.1007/s00414-017-1749-5

Telfer, M. M. (2019). Witness statement of Associate Professor Michelle Telfer. *Royal Commission into Victoria's Mental Health System*. https://rcvmhs.archive.royalcommission.vic.gov.au.

Telfer, M. M. (2021). Understanding gender diversity. *RCH Kids Health Info Podcast*. rch.org.au.

Telfer, M. M., Tollit, M. A., Pace, C. C., & Pang, K. C. (2018). Australian standards of care and treatment guidelines for transgender

and gender diverse children and adolescents. *Medical Journal of Australia, 209*(3), 132–136. https://doi.org/10.5694/mja17.01044

The Lancet Child Adolescent Health. (2021). A flawed agenda for trans youth. *Lancet Child & Adolescent Health, 5*(6), 385. https://doi.org/10.1016/S2352-4642(21)00139-5

Tollit, M. A., May, T., Maloof, T., Telfer, M. M., Chew, D., Engel, M., & Pang, K. C. (2021). The clinical profile of patients attending a large, Australian pediatric gender service: A 10-year review. *International Journal of Transgender Health*. https://doi.org/10.1080/26895269.2021.1939221

Tollit, M. A., Pace, C. C., Telfer, M., Hoq, M., Bryson, J., Fulkoski, N., Cooper, C., & Pang, K. C. (2019). What are the health outcomes of trans and gender diverse young people in Australia? Study protocol for the Trans20 Longitudinal Cohort Study. *British Medical Journal Open, 9*(11), e032151. https://doi.org/10.1136/bmjopen-2019-032151

Valenstein, E. S. (1986). *Great and desperate cures: The rise and decline of psychosurgery and other radical treatments for mental illness*. Basic Books.

Vandenbussche, E. (2022). Detransition-related needs and support: A cross-sectional online survey. *Journal of Homosexuality, 69*(9), 1602–1620. https://doi.org/10.1080/00918369.2021.1919479

van der Loos, M., Hannema, S. E., Klink, D. T., den Heijer, M., & Wiepjes, C. M. (2022). Continuation of gender-affirming hormones in transgender people starting puberty suppression in adolescence: a cohort study in the Netherlands. *Lancet Child & Adolescent Health*. https://doi.org/10.1016/S2352-4642(22)00254-1

Wager, T. D., & Atlas, L. Y. (2015). The neuroscience of placebo effects: Connecting context, learning and health. *Nature Reviews. Neuroscience, 16*(7), 403–418. https://doi.org/10.1038/nrn3976

Wiepjes, C. M., Nota, N. M., de Blok, C. J. M., Klaver, M., de Vries, A. L. C., Wensing-Kruger, S. A., de Jongh, R. T., Bouman, M.-B., Steensma, T. D., Cohen-Kettenis, P., Gooren, L. J. G., Kreukels, B. P. C., & den Heijer, M. (2018). The Amsterdam Cohort of Gender Dysphoria Study (1972–2015): Trends in prevalence, treatment, and regrets. *Journal of Sexual Medicine, 15*(4), 582–590. https://doi.org/10.1016/j.jsxm.2018.01.016

Wood, H., Sasaki, S., Bradley, S. J., Singh, D., Fantus, S., Owen-Anderson, A., Di Giacomo, A., Bain, J., & Zucker, K. J. (2013). Patterns of referral to a gender identity service for children and adolescents (1976–2011): Age, sex ratio, and sexual orientation [Letter to the Editor]. *Journal of Sex & Marital Therapy, 39*(1), 1–6. https://doi.org/10.1080/0092623X.2012.675022

Zucker, K. J. (2019). Adolescents with gender dysphoria: Reflections on some contemporary clinical and research issues. *Archives of Sexual Behavior, 48*(7), 1983–1992. https://doi.org/10.1007/s10508-019-01518-8

Zwickl, S., Wong, A., Dowers, E., Leemaqz, S. Y., Bretherton, I., Cook, T., Zajac, J. D., Yip, P., & Cheung, A. S. (2021). Factors associated with suicide attempts among Australian transgender adults. *BMC Psychiatry, 21*(1), 81. https://doi.org/10.1186/s12888-021-03084-7

AAP News™

Risk of pseudotumor cerebri added to labeling for gonadotropin-releasing hormone agonists

July 1, 2022

from the Food and Drug Administration

Article type: FDA Update

Topics: Endocrinology, Hematology/Oncology, Nephrology, Therapeutics, Urology

The Food and Drug Administration (FDA) has added a warning about the risk of pseudotumor cerebri (idiopathic intracranial hypertension) to the labeling for gonadotropin-releasing hormone (GnRH) agonists that are approved for the treatment of central precocious puberty in pediatric patients. These products include Lupron Depot-Ped (leuprolide acetate), Fensolvi (leuprolide acetate), Synarel (nafarelin), Supprelin LA (histrelin) and Triptodur (triptorelin).

The new warning includes recommendations to monitor patients taking GnRH agonists for signs and symptoms of pseudotumor cerebri, including headache, papilledema, blurred or loss of vision, diplopia, pain behind the eye or pain with eye movement, tinnitus, dizziness and nausea.

The FDA assessed the potential risk of pseudotumor cerebri with use of GnRH agonists in pediatric patients by reviewing post-marketing safety data submitted by the GnRH agonist manufacturers, searching the FDA Adverse Event Reporting System and conducting a literature search.

Six cases were identified that supported a plausible association between GnRH agonist use and pseudotumor cerebri. All six cases were reported in birth-assigned females ages 5 to 12 years. Five were undergoing treatment for central precocious puberty and one for transgender care. The onset of pseudotumor cerebri symptoms ranged from three to 240 days after GnRH agonist initiation.

Symptoms included visual disturbances (n=5), headache or vomiting (n=5), papilledema (n=3), blood pressure increase (n=1) and abducens neuropathy (n=1). Treatments included lumbar puncture (n=3), acetazolamide therapy (n=5) and ventricular peritoneal shunting (n=1).

At the time of the FDA's review, symptoms had resolved in three patients, were resolving in one patient, had not resolved in one patient, and one patient's status was unknown. GnRH agonist therapy was discontinued in three patients; the status of continued therapy was unknown for the remaining three patients.

The incidence rate of pseudotumor cerebri associated with GnRH agonist use in pediatric patients could not be reliably established due to the small number of cases and data limitations.

The FDA's Office of Pediatric Therapeutics (OPT), Division of Pediatrics and Maternal Health (DPMH) and Division of General Endocrinology (DGE) contributed to this article. OPT resides in the Office of Clinical Policy and Programs in the Office of the Commissioner. DPMH resides in the Office of Rare Diseases, Pediatrics, Urologic and Reproductive Medicine. DGE resides within the Office of Cardiology, Hematology

Endocrinology and Nephrology. Both DPMH and DGE reside within the Office of New Drugs in the Center for Drug Evaluation and Research.

Resource

Information for health care professionals on reporting adverse events to the FDA

Achille et al. International Journal of Pediatric Endocrinology (2020) 2020:8
https://doi.org/10.1186/s13633-020-00078-2

International Journal of Pediatric Endocrinology

RESEARCH Open Access

Longitudinal impact of gender-affirming endocrine intervention on the mental health and well-being of transgender youths: preliminary results

Christal Achille[1], Tenille Taggart[2], Nicholas R. Eaton[2], Jennifer Osipoff[1], Kimberly Tafuri[1], Andrew Lane[1] and Thomas A. Wilson[1*]

Abstract

Background/aims: Transgender youths experience high rates of depression and suicidal ideation compared to cisgender peers. Previous studies indicate that endocrine and/or surgical interventions are associated with improvements to mental health in adult transgender individuals. We examined the associations of endocrine intervention (puberty suppression and/or cross sex hormone therapy) with depression and quality of life scores over time in transgender youths.

Methods: At approximately 6-month intervals, participants completed depression and quality of life questionnaires while participating in endocrine intervention. Multiple linear regression and residualized change scores were used to compare outcomes.

Results: Between 2013 and 2018, 50 participants (mean age 16.2 ± 2.2 yr) who were naïve to endocrine intervention completed 3 waves of questionnaires. Mean depression scores and suicidal ideation decreased over time while mean quality of life scores improved over time. When controlling for psychiatric medications and engagement in counseling, regression analysis suggested improvement with endocrine intervention. This reached significance in male-to-female participants.

Conclusion: Endocrine intervention may improve mental health in transgender youths in the US. This effect was observed in both male-to-female and female-to-male youths, but appears stronger in the former.

Keywords: Transgender, Transgender management, Transgender youth, Depression, Suicide, Suicidal ideation, Quality of life, GnRH analogue, Puberty suppression, Puberty, Testosterone, Estrogen, Cross sex hormone

Introduction

Transgender individuals have a gender identity that differs from the sex assigned at birth [1]. These individuals have a high prevalence of body image dysphoria, depression and suicidal ideation [2]. Studies in adults have shown improvement in psychological function in adulthood from endocrine and/or surgical interventions. Specifically, studies have indicated a positive impact of cross sex steroid therapy on depression scores and quality of life in the adult transgender population [3]. Guidelines for endocrine intervention in transgender youth have existed for the past decade in the United States and longer internationally. These guidelines include suppression of puberty to provide more time before cross sex steroid

* Correspondence: Thomas.A.Wilson@SUNYSB.EDU
[1]Division of Pediatric Endocrinology, Stony Brook Children's Hospital, Stony Brook University, Stony Brook, NY 11794-8111, USA
Full list of author information is available at the end of the article

therapy is introduced [4, 5]. Two studies have examined the impact of this strategy on depression and quality of life in youths. De Vries et al. demonstrated no improvement of gender dysphoria after puberty suppression alone but did report improvement only after both cross sex steroid therapy and gender confirmation surgery was complete in transgender individuals from the Netherlands [6]. These authors did not report findings after cross sex steroid therapy alone but before surgery. In the UK, Costa found that GnRH agonist suppression of puberty improved psychological functioning in transgender youth [7]. In the United States, there are few data concerning the impact of endocrine intervention on psychological function in transgender youth. Therefore, we conducted a longitudinal assessment of psychological wellbeing and quality of life in children and adolescents who have sought endocrine intervention to help with gender dysphoria. Herein, we report preliminary results of this ongoing study.

Objective

The aim of this study is to examine the impact that endocrine intervention [suppression of endogenous pubertal hormones utilizing GnRH agonists/anti-androgens/suppressors of menstruation (AKA "pubertal suppression"), or addition of cross-sex hormones] has on depression and quality of life scales of transgender youths as reported by the youths themselves over time.

Methods

Participants and procedure

This is a single center study approved by Stony Brook University IRB for children, adolescents and young adults aged 9–25 years. Subjects referred to the Pediatric Endocrine Department for gender dysphoria were approached to participate. Although we do not have exact numbers, the vast majority of eligible subjects agreed to take part in the study. Minor participants signed assent and participants over 18 years of age and parents of those less than 18 yr. of age signed consent to participate. Individuals with sex chromosome abnormalities and disorders of sexual differentiation were excluded from the study. At approximately 6-month intervals, participants completed the following validated assessments of mental health: The Center for Epidemiologic Studies Depression Scale (CESD-R) [8], The Patient Health Questionnaire Modified for Teens (PHQ-9_Modified for Teens) [9], Quality of Life Enjoyment and Satisfaction Questionnaire (QLES-Q-SF) [10]. Most subjects were followed by mental health professionals. Those that were not were encouraged to see a mental health professional.

Psychological measures

The CESD-R score is calculated as a sum of 20 questions, ranging from 0 (for those who say "not at all or less than one day" to all 20 questions) to a maximum score of 60 (for those who say "5–7 days" and/or "nearly every day for 2 weeks" for all 20 questions). A total CESD-R score less than 16 implies no clinical depression [8, 11]. The PHQ-9 consists of 9 questions describing symptoms of depression each rated 0 to 3 with the sum indicating level of depression: minimal 0–4, mild 5–9, moderate 10–14, moderately severe 15–19, severe 20–27. This questionnaire also asks the participants four additional questions relating to suicidal ideology and difficulty dealing with problems of life [9]. The QLES-Q-SF consists of 15 questions rating quality of life on a scale of 1–5 with 1 being poor and 5 being very good [10]. It was used rather than the Pediatric Quality of Life and Enjoyment Scale (PQLES-SF), which is based on QLES-Q-SF, because of the overlap in age inclusion of older adolescents and young adults and the intention of continuing the study into adulthood. Transyouths in the study were also asked if they were participating in psychological counseling and/or on psychiatric medication. ADHD medications were not included as psychiatric medication for this analysis.

Endocrine interventions

Endocrine interventions were introduced in accordance with the Endocrine Society and the WPATH guidelines [4, 5]. In our study, GnRH agonist and/or antiandrogens were used for male to female (MTF) participants, and suppression of menstruation (either GnRH agonist or Medroxyprogesterone) for female to male (FTM) participants. Collectively, these interventions were labeled "Puberty Suppression". Once eligible as determined by mental health consultants, youths, parents and according to guidelines, cross sex hormones were prescribed, either testosterone for FTM or estrogen for MTF participants.

Statistical analysis

Regression analysis was used to examine the association of various treatments with outcomes experienced by transgender youths over time. Linear multiple regression was used for continuous outcomes, and multiple logistic regression was used for dichotomous outcomes. For continuous outcomes, residualized change scores were used to compare change at outcome relative to levels at baseline. This approach thus allowed us to control for the dependent variable's level at baseline for each participant and to examine how endocrine intervention predicted change in the dependent variable over and above predicted outcome level relative to the level at baseline. Regression analyses also controlled for psychiatric medication and engagement in psychotherapy.

Results

Between December 2013 to December 2018, 116 participants entered the study. Ninety-five were naive to any endocrine intervention. Of those 95 participants, 50 completed 3 waves of questionnaires and these individuals compose the analytic sample in this report. Baseline data for this population are shown in Table 1. At wave one, none of the 50 participants were on endocrine intervention. By wave 3, 47 participants had some type of endocrine intervention (Table 2).

Mean changes over time

Mean baseline CESD-R score was 21.4 and decreased to 13.9 by wave 3 ($t(48) = 3.996$, $p < 0.001$, Fig. 1a). A score less than 16 implies no clinical depression. Mean depression scores by the PHQ-9 decreased over time as well ($t(49) = 3.753$, $p < 0.001$, Fig. 1b), while quality of life scores improved (Fig. 1c) but did not reach statistical significance ($t(48) = -1.758$, $p = .085$, Fig. 1c). Suicidal ideation decreased over time across all groups at wave 3 relative to baseline (Table 3). Thus, by all measures, depression and quality of life improved to some degree over time. Both gender subgroups demonstrated similar trends.

Regression analysis

We conducted a series of regression analyses to investigate preliminary trends in the data when controlled for reported psychiatric medications and engagement in counselling. Results are given in Table 4. Given our modest sample size, particularly when stratified by gender, most predictors did not reach statistical significance. This being said, effect sizes (R^2) values were notably large in many models. In MTF participants, only puberty suppression reached a significance level of $p < .05$ in the CESD-R. However, associations with PHQ9 and QLES-Q-SF scores approached significance. For FTM participants, only cross sex hormone therapy approached statistical significance for quality of life improvement ($p = 0.08$).

Model R^2 values ranged between small to large, even in models where the hormonal intervention's prediction of the outcome did not reach statistical significance. It is potentially noteworthy that effect sizes for endocrine interventions were notably larger for MTF than FTM participants in almost every analysis. Regression models for suicidal thoughts were not estimable due to the low frequency of endorsement and small cell sizes across gender.

Table 2 Endocrine interventions at wave 3

Type of Intervention	% of Total (n)	% of FTM (n)	% of MTF (n)
None	6% (3)	3% (1)	12% (2)
Puberty Blocker	46% (23)	24% (8)	88% (15)
Cross Sex Hormone	70% (35)	85% (28)	41% (7)
Both	22% (11)	12% (4)	41% (7)

Discussion

Cross-sex hormones and their effect on depression and quality of life has been extensively studied in adults. A meta-analysis by Costa and Colinza reported a reduction in anxiety and depression and improvement in quality of life with positive effect on personality and mood among transgender adults receiving cross-sex hormones therapy [3]. A 2006 cross-sectional study in California looked at adult FTM transgender participants on cross-sex hormone therapy and their quality of life. Participants who received testosterone therapy reported statistically significant higher quality of life than those who had not received hormonal therapy [12].

Adolescence is a particularly difficult time for transgender persons who experience the development of secondary sexual characteristics that are incongruous with their gender identity, and is associated with a high prevalence of depression and suicidal thoughts and gestures. Previous research has shown benefit to transgender youth in the Netherlands after cross sex steroid therapy AND gender confirmation surgery and in the UK after pubertal suppression alone [6, 7]. Our results extend these findings to transgender youths in the USA and apply prior to surgery.

Our results suggest that endocrine intervention is associated with improved mental health among transgender youth. This effect was observed in both MTF and FTM participants but appeared to be stronger in MTF. We speculate that this could be due to the following possibilities: 1. Testosterone has profound effects on

Table 1 Baseline characteristics at Wave 1

	Total	Female to Male	Male to Female
Number of participants	50	33	17
Age in Years (SD)	16.2 (2.2)	16.6 (2.5)	15.5 (1.6)
%Depressed in past year (n)	64% (32)	60.6% (20)	70.6% (12)
% Suicidal (n)	10% (5)	9.1% (3)	11.8% (2)
% In Counseling (n)	90% (45)	87.9% (29)	94.1% (16)
% On Psych Medication (n)	34% (17)	36.4% (12)	29.4% (5)

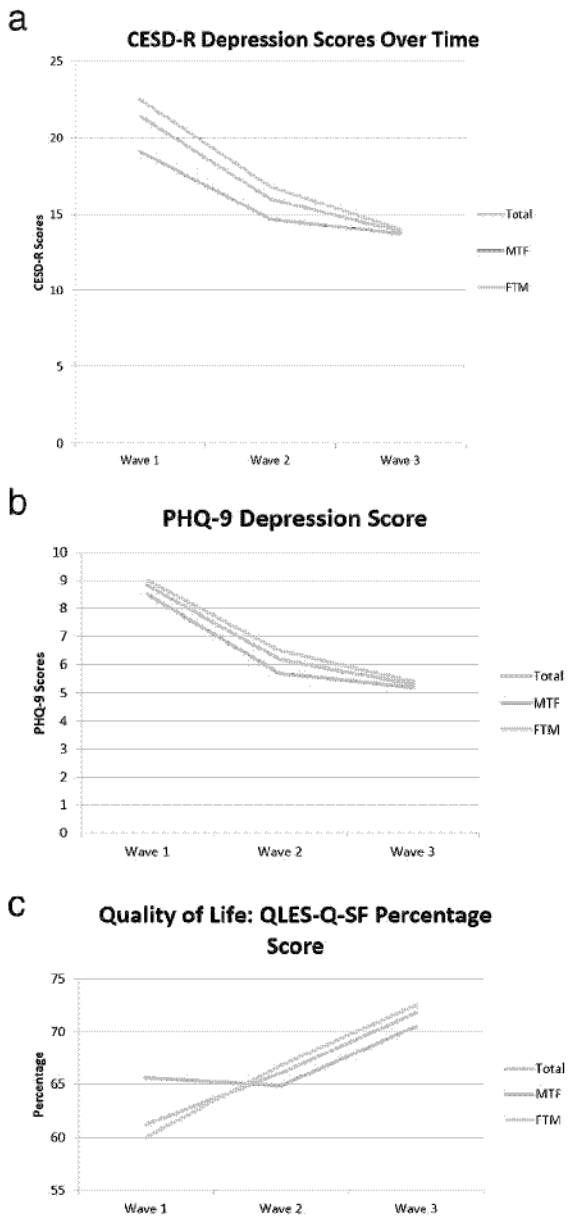


Fig. 1 Mean scores over time: **a** CESD-R. **b** PHQ-9. **c** QLES-Q-SF. Abbreviations: GnRH: Gonadotropin releasing hormone; CESD-R: Center for Epidemiologic Studies Depression Scale; PHQ-9: Patient Health Questionnaire Modifed for Teens; QLES-Q-SF: Quality of Life Enjoyment and Satisfaction Questionnaire; WPATH; World Professional Association for Transgender Health; MTF: male to female; FTM: Female to male

appearance. MTF participants may have experienced relief when serum testosterone concentrations are suppressed or antagonized; 2. The effects of testosterone in FTM transgender persons takes 6 to 12 months to become apparent and is not fully apparent until several years of exposure. Our study only extended for the first 12 months of endocrine intervention.

Table 3 Suicidal ideation

Suicidal Ideation Percentage: Wave 1 vs Wave 3		
	% at Wave 1 (n)	% at Wave 3 (n)
Total	10% (5)	6% (3)
MTF	11.8% (2)	5.9% (1)
FTM	9.1% (3)	6.1% (2)

Limitations and future directions

This is an ongoing study with preliminary results only presented herein. The numbers are too small to parse out the effects of pubertal suppression versus cross sex hormone therapy in the different genders. As our numbers continue to grow, we hope that we will be able to do so. As of now, we are only able to report trends.

Parental support has been shown to protect against mental health problems in transgender adolescents. Children who are socially transitioned at home, at school, and who use gender affirming pronouns represent those youths who are supported by their parents and caregivers. Being supported by family is associated with positive mental health outcomes [13] . Our data are somewhat limited by the fact that the majority of our participants had at least one supportive parent who was willing to facilitate medical and mental health intervention for the child and therefore may not apply to all transgender youths. In addition, regular visits with the medical team itself could influence depression and quality of life. Past studies have shown that having support from a multidisciplinary medical team – mental health provider, physician, surgeons – helped with quality of life and mental health [6].

Conclusions

Transgender children and adolescent are a high-risk population for suicide and depression. Our preliminary results show negative associations between depression scores/suicidal ideation and endocrine intervention, while quality of life scores showed positive associations with intervention, in transgender youths over time in the US. These results align with previous work in the Netherlands and the UK.

Table 4 Regression results when controlled for engagement in counselling and psychiatric medications

		MTF			FTM		
Survey	Intervention	b	*p*	R^2	b	*p*	R^2
CESD-R	Puberty Suppression	−2.41	0.008	0.52	−0.02	0.95	0.09
	Cross Sex Hormone	−0.56	0.27	0.21	−0.43	0.43	0.11
PHQ-9	Puberty Suppression	−1.89	0.07	0.28	−0.16	0.68	0.04
	Cross Sex Hormone	−0.92	0.07	0.29	−0.23	0.67	0.04
QOL	Puberty Suppression	1.26	0.21	0.13	0.71	0.86	0.01
	Cross Sex Hormone	0.87	0.06	0.08	0.93	0.08	0.11

Acknowledgements
The authors wish to thank the subjects and parents for their willingness to participate in this project and their courage.

Authors' contributions
All authors contributed to the work as outlined below: Christal Achille: Primary author, helped recruit and collect data. Tenille Taggart: Worked on data entry and data analysis. Jennifer Osipoff: Recruited subjects, collected data. Kimberly Tafuri: Recruited subjects, collected data. Andrew Lane: Recruited subjects, collected data. Nicholas Eaton: Statistical analysis and interpretation. Thomas Wilson: Senior author, initial research conceptualization, IRB approval, subject recruitment, data entry and analysis and final submission. The author(s) read and approved the final manuscript.

Funding
This study is not funded.

Availability of data and materials
Data is not available as it would compromise confidentialty of the subjects participating.

Ethics approval and consent to participate
This research effort was approved by the Stony Brook University Committee on Research Involving Human Subjects. Consent and assent to participate was obtained from subjects < 18 years of age and their respective parents and consent was obtained from those over age 18 years of age.

Consent for publication
Consent for publication was included in the consent/assent.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Division of Pediatric Endocrinology, Stony Brook Children's Hospital, Stony Brook University, Stony Brook, NY 11794-8111, USA. [2]Department of Psychology, Stony Brook University, Stony Brook, NY 11794, USA.

Received: 28 February 2020 Accepted: 3 April 2020
Published online: 30 April 2020

References

1. [APA], A.P.A., Diagnostic and Statistical Manual of Mental Disorders (5th ed.; DSM–5). 2013.
2. Aitken M, VanderLaan D, Wasserman L, Stojanovski Hon S, Zucker K. Self-harm and Suicidality in children referred for gender Dysphoria. J Am Acad Child Adolesc Psychiatry. 2016;55(6):513–9.
3. Costa R, Colizzi M. The effect of cross-sex hormonal treatment on gender dysphoria individuals' mental health: a systematic review. Neuropsychiatr Dis Treat. 2016;12:1953–66.
4. Hembree WC, Cohen-Kettenis P, Gooren L, Hannema S, Meyer W, Murad MH, et al. Endocrine treatment of gender-Dysphoric/gender-incongruent persons: an Endocrine Society* clinical practice guideline. J Clin Endocrinol Metabol. 2017;102(11):3869–903.
5. Selvaggi G, Dhejne C, Landen M, Elander A. The 2011 WPATH standards of care and penile reconstruction in female-to-male transsexual individuals. Adv Urol. 2012;2012:581712.
6. de Vries A, MCGuire J, Steensman T, Wagenaar E, Doreleijers T, Cohen-Kettenis P. Young adult psychological outcome after puberty suppression and gender reassignment. Pediatrics. 2014;134(4):696–704.
7. Costa R, Dunsford M, Skagerberg E, Holt V, Carmichael P, Colizzi M. Psychological support, puberty suppression, and psychosocial functioning in adolescents with gender Dysphoria. J Sex Med. 2015;12(11):2206–14.
8. Eaton WW, Smith C, Ybarra M, Muntaner C, Tien A. Center for Epidemiologic Studies Depression Scale: Review and Revision (CESD and CESD-R). In: The use of psychological testing for treatment planning and outcomes assessment: Instruments for adults, vol. 3. 3rd ed. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers; 2004. p. 363–77.
9. Richardson LP, McCauley E, Grossman DC, et al. Evaluation of the Patient Health Questionnaire-9 Item for detecting major depression among adolescents.Pediatrics. 2010;126(6):1117–23. https://doi.org/10.1542/peds.2010-0852.
10. Endicott J, Nee J, Harrison W, Blumenthal R. Quality of life enjoyment and satisfaction questionnaire: a new measure. Psychopharmacol Bull. 1993;29(2):321–6.
11. Heroz E, Ybarra M, Eaton W. Psychometric evaluation of a self-report scale to measure adolescent depression: the CESDR-10 in two national adolescent samples in the United States. J Affect Disord. 2014;158:154–60. https://doi.org/10.1016/j.jad.2014.02.009.
12. Newfield E, Hart S, Dibble S, Kohler L. Female-to-male transgender quality of life. Qual Life Res. 2006;15:1447–57.
13. Olson KR, Durwood L, DeMeules M, McLaughlin KA. Mental Health of Transgender Children Who are Supported in Their Identities. Pediatrics. 2016;137(3):e20153223. https://doi.org/10.1542/peds.2015-3223. Epub 2016 Feb 26.

Publisher's Note

Received: 24 June 2020 | Revised: 27 July 2020 | Accepted: 5 August 2020

DOI: 10.1111/andr.12884

ORIGINAL ARTICLE

ANDROLOGY WILEY

Long-term effect of gender-affirming hormone treatment on depression and anxiety symptoms in transgender people: A prospective cohort study

Zoë Aldridge[1,2] | Shireen Patel[1] | Boliang Guo[1] | Elena Nixon[1] | Walter Pierre Bouman[1,3] | Gemma L. Witcomb[3] | Jon Arcelus[1,3]

[1]Faculty of Medicine and Health Sciences, Institute of Mental Health, University of Nottingham, Nottingham, UK

[2]School of Sport, Exercise and Health Sciences, Loughborough University, Loughborough, UK

[3]Nottingham Centre for Transgender Health, Nottingham, UK

Correspondence
Jon Arcelus. Institute of Mental Health, Jubilee Campus, Triumph Road, University of Nottingham, Nottingham NG7 2TU, UK.
Email: Jon.Arcelus@nottingham.ac.uk

Funding information
The data analysis only was funded by the National Institute for Health Research (NIHR) Applied Research Collaboration East Midlands (ARC EM), NIHR Research for Patient Benefit (RfPB) and Nottinghamshire Healthcare NHS Foundation Trust.

Abstract

Background: Cross-sectional studies show that transgender people are more likely than cisgender people to experience depression and anxiety before gender-affirming hormone treatment (GAHT). However, the effect of GAHT on mental health in transgender people, and the role of other factors that may have a predictive effect, is poorly explored.

Objectives: Using a longitudinal methodology, this study investigated the effect of 18-month GAHT on depression and anxiety symptomatology and the predictors on mental health outcomes in a large population of transgender people.

Materials and methods: Participants (n = 178) completed a socio-demographic questionnaire, the Hospital Anxiety and Depression Scale (HADS), the Multidimensional Scale of Perceived Social Support (MSPSS) and the Autism Spectrum Quotient—Short Version (AQ-Short) at pre-assessment (T0) and at 18 months after initiation of GAHT (T1).

Results: From T0 to T1, symptomatology was significantly decreased for depression ($P < .001$) and non-significantly reduced for anxiety ($P = .37$). Scores on the MSPSS predicted reduction in depression, while scores on the AQ-Short predicted reduction in anxiety.

Discussion: GAHT reduces symptoms of depression which are predicted by having higher levels of social support. Although anxiety symptoms also reduce, the changes are not significant and high levels of anxiety still remain post-GAHT.

Conclusions: These results highlight the important mental health benefits of GAHT. Support services (professional, third sector or peer support) aiming at increasing social support for transgender individuals should be made available.

KEYWORDS

autism, hormone therapy, longitudinal, mental health, social support, transgender

1 | INTRODUCTION

Treatment-seeking transgender people who are not on hormone treatment have reported high levels of mental health problems, particularly anxiety, depression and self-harm, which are likely caused by a number of internal and external stressors.[1-12] Studies examining mental health in transgender people have primarily focused on individuals attending transgender health services and hence those who are likely to experience a higher level of distress about their assigned sex at birth. These studies have primarily looked cross-sectionally at levels of anxiety,[7,13,14] depression[7,15-17] and self-harm.[18-20]

With regard to anxiety, several studies have demonstrated high levels in transgender people before gender-affirming hormone treatment (GAHT).[21,22] For example, Bouman et al [13] found that levels of anxiety in transgender people were three times higher than those in a matched sample from the general population. This study also found that transgender males were more anxious than transgender females. Interestingly, the high scores on autistic traits found among this population have been suggested to be a product of the high levels of anxiety and low self-esteem often experienced by this group[15] and not autism per se.[23] However, a recent study has demonstrated stability in autistic traits following GAHT.[24]

Similar to anxiety, high levels of depression have also been reported in transgender individuals, prior to GAHT.[16,18,22] Witcomb et al[16] reported that transgender people prior to receiving GAHT had a fourfold increased risk of a probable depressive disorder compared to a matched control sample from the general population. Why this is the case is unclear, but social factors such as lack of general social support,[9,25-27] parental support[28] and peer support have been found to be associated with depressive symptoms among transgender people.[29,30] Experiences of transphobic discrimination are associated with increased odds of suffering from depression[31] independent of other types of discrimination, for example racism. This suggests that transgender people who are ethnic minorities are at even greater risk, because of the intersectional experience of discriminatory events. In addition, while unemployment increases the risk of depression in the general population[32] and transgender people have been found to have a higher unemployment rate than cisgender people[33]—being in employment is associated with higher levels of experienced transphobia and fear of disclosing mental health problems in the transgender population.[34,35]

Another factor that has been associated with mental health problems in treatment-seeking transgender people is age. Younger transgender people report high levels of bullying[36] and very high levels of self-harm,[19,37] which have been associated with increased anxiety as well as effects on self-esteem, family relationships and social life, which all negatively influence mental well-being.

While these studies have provided valuable insight, the use of cross-sectional methodologies to examine the impact of the above factors, particularly the role of GAHT in mental health, is limited. Therefore, it is critical to explore this on a within-subject basis using a longitudinal design. This is the most effective approach to show the effects of GAHT on mental health as it provides the opportunity to examine individuals prior to and during GAHT.

A small number of longitudinal studies that focus on the effect of GAHT on mental health do exist. Colizzi et al[38] reported significant reductions in mental health symptoms after the initiation of GAHT with anxiety reducing from 50% to 17% and depression from 24% to 11%. Heylens et al[39] also showed significant reductions in symptoms of anxiety and depression after the initiation of GAHT to the point where they resemble those of the general population. These studies are, however, not without limitations. Heylens et al's[39] study has a small sample size (n = 57), while Colizzi et al's[38] study is limited by the lack of evaluation of factors that may have impacted on the mental health of their participants, such as social support. Both studies describe the need to replicate their findings. In contrast, Bränström and Pachankis[40] using the Swedish population register showed no significant association between the likelihood of accessing mental health treatment and time since initiation of GAHT. The limitation of their study includes primarily that accessing mental health services does not necessarily reflect actual mental health and there is little additional information about the type of mental health treatment received by their participants. These limitations mean that this study cannot provide reliable evidence regarding the role of GAHT on the mental health symptoms of transgender people, and this information is vital in order to provide an evidence base of GAHT improving overall quality of life of transgender people.

While the available longitudinal studies have provided valuable evidence of the effect of GAHT on transgender people's mental health, there is a requirement to replicate these studies addressing their limitations. With this in mind, the primary aim of this study is to examine the effect of GAHT on anxiety and depression symptoms. The study will focus on those who have been on treatment for over 18 months as this allows for enough time for GAHT to produce physical, bodily changes but before surgical procedures have taken place, which could bias the results. As some physical changes can be quicker in assigned females at birth than in assigned males at birth (eg voice change with testosterone),[2] which can affect mental health outcome following GAHT, the results of GAHT in anxiety and depression for both groups will be presented separately. It is hypothesized that an improvement in mental health will take place in those assigned male and female at birth following GAHT treatment. Unfortunately, because of the long waiting list for gender-affirming surgical treatment in the United Kingdom (UK), it is unlikely that people will have undergone these interventions before this time. The secondary aim of this study is to examine pre-GAHT factors which may be predicting changes in anxiety and depression following GAHT. The predictors selected for this study are based on the literature and include ethnicity, age, assigned sex at birth, civil status, employment, social support, and autistic traits. This study hypothesized that symptoms of depression and anxiety would be significantly decreased after 18 months of GAHT.

2 | MATERIALS AND METHODS

2.1 | Participants

Participants were invited to take part through a national transgender health service in Nottingham, UK. This service is part of the National Health Service (NHS) and offers assessment for suitability of GAHT as well as chest and genital reconstructive surgery. The service also offers GAHT and speech and language therapy. The service accepts referrals from people aged 17 and over who are seeking, or considering, medical transition.

2.2 | Procedures

The sample consisted of individuals who attended an assessment at the transgender health service from November 2014 to March 2018, who agreed participation and who were not on GAHT prior to the assessment. Prior to the clinical assessment, every patient was invited to participate in the study. If agreed, they were invited to complete a baseline questionnaire pack (T0). The pack included a socio-demographics questionnaire (age, sex assigned at birth, gender identity, ethnicity, employment status, relationship status and whether participants were taking cross-sex hormones and/or blockers pre-assessment—as a significant proportion of young people are referred from the only existing child and adolescent transgender health services in the United Kingdom). Validated questionnaires regarding anxiety and depression (HADS), social support (MSPSS) and autistic traits (AQ-Short) were also included in the information pack. Data were only included if participants returned a signed consent form with the study questionnaires.

Participants who consented and returned T0 questionnaires were invited to complete a T1 questionnaire 18 months after commencing GAHT. The T1 questionnaire pack consisted of a HADS questionnaire. This allowed a comparison of changes in depression and anxiety symptoms before and after GAHT. Data were collected in October 2019. Except for the data analysis, the study was primarily unfunded and set up in a busy clinic.

2.3 | Tools

2.3.1 | The Hospital Anxiety and Depression Scale (HADS)

The HADS is a 14-item self-report screening scale originally developed to indicate the possible presence of anxiety and depression states in medical non-psychiatric outpatient clinics.[41] The HADS consists of two subscales, HADS—Depression (HADS-D) and HADS—Anxiety (HADS-A). Each subscale has seven items that are rated on a 4-point Likert scale that ranges from 0 to 4 with some items being reverse-scored. A maximum total of 21 can be obtained on each subscale. A score of 0-7 on both scales implies a non-clinical range, while a score of 8-10 suggests the possible presence of a depressive or anxiety disorder. A score of 11 or higher suggests the probable presence of a depressive or anxiety disorder. Caseness of depression and anxiety has been suggested for scores above 8.[42] The HADS has previously been used with transgender individuals.[13,43] For depression (HADS-D), this gave a specificity of 0.7 and a sensitivity of 0.9. For anxiety (HADS-A), this gave a specificity of 0.78 and a sensitivity of 0.9. In this study; Cronbach's alpha for depression was 0.76 and for anxiety 0.68.

2.3.2 | The multidimensional scale of perceived social support (MSPSS)

The MSPSS is a 12-item self-report scale to record levels of social support from family, friends and significant others.[44] The measure consists of three subscales to measure the three different types of support. Items are rated on a Likert scale that ranges from 1 ('very strongly agree') to 7 ('very strongly disagree'). To calculate subscale scores, items from each subscale are added together and divided by 4. A total score is calculated by adding together all 12 items and dividing by 12. The mean and total scores range from 1 to 7 with a higher score indicating a higher level of perceived social support. A mean total scale score ranging from 1 to 2.9 can be considered low support; a score of 3 to 5 can be considered moderate support; and a score from 5.1 to 7 can be considered high support. The MSPSS has previously been used with transgender individuals.[26,45] In this study, Cronbach's alpha was 0.89.

2.3.3 | Autism spectrum quotient—short version (AQ-Short)

The AQ-Short is a 28-item self-report questionnaire designed to measure autistic traits to give an indication of where the person lies on the continuum of the spectrum, ranging from healthy to autistic.[46,47] It is a shortened version of the validated AQ-50.[48] It consists of two higher order factors related to autistic traits, including numbers and patterns (which assess the extent to which people are fascinated by numbers, dates, patterns and categories) and social behaviours. The AQ-Short is a 4-point Likert scale ranging from 'definitely agree' to 'definitely disagree', with some items being reverse-scored. Total scores range between 28 and 112. The AQ-Short has previously been used with transgender populations.[23,24,49] Higher scores represent higher levels of autistic traits. Although not intended to be a diagnostic tool, a cut-off of ≥ 70 was found to have a sensitivity of 0.94 and specificity of 0.91 to discriminate between an autism sample and a community sample. Cronbach's alpha was 0.86.

2.3.4 | Data analysis

Data analyses were performed using the Statistical Software Package Stata 16.[50] Stata 16 was used to conduct power analysis. Only those

participants not on GAHT at assessment (T0) were included in the regression analysis. All missingness was imputed using analytical model with 20 imputed data sets generated for each model. Paired sample t tests were used to determine whether there had been a significant change in the HADS-D and HADS-A subscales from T0 to T1. Multiple regression was conducted to explore ethnicity, employment status, relationship status, age, assigned sex, MSPSS and AQ. The hypothesis regarding whether the specific factors were predictive of changes in anxiety and depression was tested via a moderator analysis, entering only the subscales found to be significant in the linear regressions and a product of their combined centred scores. This was tested via a multiple regression. Bonferroni corrections were used to correct multiplicity issue if needed. Although data are not normally distributed, Allison[51]) states that normality is the least important assumption of regression and as data met the assumptions for linearity, homoscedasticity and absence of multicollinearity or extreme outliers, a multiple regression analysis was conducted. The socio-demographic categories were split into two distinct groups for each category as seen in Table 1. Assigned sex at birth instead of gender identity was used in the socio-demographics in view of the many different gendered identities described, as an analysis based on gender identities would have made the analyses too complex to interpret; this followed previous study approaches.[23,24] The MSPSS and AQ-28 Short at T0 were significant factors predicting change in HADS-D and HADS-A at T1. To check the robustness of regression estimates sensitive to missingness, all regression models were re-run on observed data only and the results were examined against the results from imputed data set.

Ethical approval for the study was received from the NHS Ethics Committee (14/EM/0092) and the Research and Development Department at Nottinghamshire Healthcare NHS Foundation Trust in line with Health Research Authority guidance,[52] which included approval for individuals aged 17 and over to sign giving their consent without the need for additional parental consent.

3 | RESULTS

3.1 | Socio-demographic characteristics of the participants

A total of 1,271 participants were assessed between November 2014 and March 2018, completed T0 questionnaires and agreed participation in the study. Seventy-one per cent (N = 906) could be included in the analysis as they had not received hormones prior to assessment. Of these, 178 (20%) went on to complete a T1 questionnaire after 18 months of GAHT, indicating a response rate of 20%. Responders did not differ from non-responders in terms of either demographic characteristics or baseline AQ-Short scores, but they were significantly less anxious at baseline than non-responders (median 9 vs 8, $P = .001$; $z = 3.225$) (see Table 1).

The age range from the 178 participants that completed T0 and T1 questionnaires ranged from 17 to 79 years with a median age of 23 years. More than half of the participants (n = 95; 53.3%) were assigned male sex at birth, and 83 (46.7%) were assigned female sex at birth. The large majority of participants classified themselves as white (n = 167; 94%), were single (n = 120; 69%) and were in employment (n = 75; 41%) or students (n = 58; 32%). Participants who were assigned male sex at birth were more likely to be in employment compared to participants assigned female sex at birth (47% vs 33%), while more participants assigned female sex at birth were single at the time of assessment (78% vs 57%) and a higher percentage of participants assigned male sex at birth were divorced/separated (18% vs 2%) (see Table 1).

TABLE 1 Socio-demographic characteristics (n,%) and mean (SD) scores on HADS, MSPSS and AQ-Short score of all responders assigned male and female at birth

	Responders		Assigned male at birth (2)		Assigned female at birth (1)	
	n = 178		n = 95		n = 83	
	n%		n%		n%	
Age, median (range) years	23 (1774)		28 (1779)		21 (1764)	
Ethnic origin n (%)						
White	167	94	92	97	75	91
Other	11	6	3	3	8	9
Not known	-	-	-	-	-	-
Employment status n (%)						
Employed (1)	75	41	45	47	28	33
Student (2)	58	32	20	21	37	44
Housewife/ househusband (4)	-	-	-	-	-	-
Voluntary work (3)	7	4	5	5	2	2
Retired (6)	5	3	6	6	-	-
Disabled (5)	7	5	5	5	3	3
Unemployed (0)	26	14	14	14	12	14
Not known (9)	1	1			1	1
Civil status n (%)						
Single (1)	120	69	55	57	65	78
Married (2)	15	7	12	12	1	1
Civil partner (3)	5	3	-	-	5	6
In a relationship (7)	4	3	1	1	4	4
Divorced/separated (4)	20	11	18	18	2	2
Widowed (5)	2	2	2	2	1	1
Other (6)	-	-	-	-	-	-
Not known (9 + Blanks)	12	7	7	7	5	6
HADSD	7.24 (4.03)		7.03 (4.11)		7.48 (3.94)	
HADSA	8.07 (4.34)		7.54 (4.31)		8.69 (4.32)	
MSPSS	4.85 (1.29)		4.64 (1.35)		5.1 (1.16)	
AQ	64.77 (11.86)		62.83 (11.94)		66.97 (11.44)	

TABLE 2 Means (SD) of HADS-D and HADS-A scores of responders at T0 and T1

	All responders			Assigned male at birth			Assigned female at birth		
	N	Mean (SD)	Mean change from T0 to T1 (95% CI), *P*-value	n	Mean (SD)	Mean change from T0 to T1 (95% CI), *P*-value	n	Mean (SD)	Mean change from T0 to T1 (95% CI), *P*-value
HADS-D									
T0	178	7.24 (4.03)	−2.05 (−2.72 to − 1.38) *P* = .00	95	7.03 (4.11)	−1.91 (−2.80 to − 1.01) *P* = .00	83	7.48 (3.94)	−2.21 (to 3.23 to − 1.20) *P* = .00
T1	178	5.19 (3.73)		95	5.13 (3.92)		83	5.26 (3.52)	
HADS-A									
T0	178	8.07 (4.34)	−0.31 (−0.97 to 0.36), *P* = .37	95	7.54 (4.31)	−1.16 (−1.50 to 0.39) *P* = .25	83	8.69 (4.32)	−0.55 (−0.97 to 0.92) *P* = .97
T1	178	7.77 (3.90)		95	6.98 (3.96)		83	8.66 (3.65)	

3.2 | Anxiety and depression scores

The mean score for the total group for anxiety was 8.07 (SD: 4.34). It was higher in those assigned female (8.69 (SD: 4.32)) versus those assigned male at birth (7.54 (SD: 4.31). The mean score for depression was 7.24 (SD: 4.03), also higher in assigned females (7.48 (SD: 3.94)) than assigned males at birth (7.03 (SD: 4.11)).

At T0 (before hormone treatment), 51.13% of participants scored 8 or over on the HADS-A subscale, and in the case of the HADS-D subscale, 47.75% of participants scored above 8 placing these participants within the categories of possible to probable presence of an anxiety or depression disorder. At T1, 47.19% of participants scored 8 or above in the HADS-A subscale showing a reduction of 3.94%, and in the HADS-D subscale, 25.84% of people scored 8 or above showing a reduction of 21.91%.

3.3 | Change in anxiety and depression scores between T0 and T1

There was a statistically significant reduction in mean scores of HADS-D from T0 to T1 (mean change difference, −2.05; 95% CI, −2.72 to −1.38; P = .00). This indicated a reduction in depression following 18 months of GAHT. There was also a reduction in the HADS-A score from T0 to T1, but this was not statistically significant (mean change difference, −0.31; 95% CI, −0.97 to 0.36, P = .37). The same findings (a significant reduction in HADS-D and a non-significant reduction in HADS-A) were found when comparing T0 and T1 according to sex assigned at birth (see Table 2).

3.4 | Predictors of anxiety and depression change after hormone treatment

Two multiple regressions with seven predictor variables were conducted to explore the predictors of change, from T0 to T1, in scores on HADS-D and HADS-A. The predictors for each were ethnicity, employment status, relationship status, assigned sex at birth, and age. MSPSS and AQ-Short at T0 were used as independent variables for both regressions.

The results for the first regression showed that overall, the model was significant (F(7,152)=2.09, P = .04) and explained 8.8% (R^2 = 0.088) of the total variance in depression scores. The model also showed that mean MSPSS scores at T0 were the only significant predictor of HADS-D change between T0 and T1 (β = 0.81, P = .006). The second regression showed that overall, the model was significant (F(7,152)=2.09, P = .048) and explained 8.8% (R^2 = 0.088) of the total variance in anxiety scores. The model also showed mean AQ-Short scores at T0 was a significant predictor of HADS-A change between T0 and T1 (β = −0.069, P = .034). The findings suggest that only levels of social support (MSPSS scores) and autistic spectrum traits (AQ scores) were able to predict changes in anxiety and depression following 18 months of GAHT. Having higher levels of social support

(higher scores of MSPSS) predicted a reduction in depression scores following 18 months of GAHT ($P = .006$) and having lower levels of autistic spectrums traits (lower AQ scores) predicted a reduction of anxiety symptoms following 18 months of GAHT ($P = .03$), although this reduction was statistically non-significant (see Table 3).

4 | DISCUSSION

This prospective longitudinal study aimed to explore whether 18 months of GAHT reduces symptoms of anxiety and depression in transgender people, while addressing the limitations of previous studies, by recruiting a large sample of participants within the same setting. The study found a significant reduction in symptoms of depression in transgender individuals after 18 months of starting GAHT, with a more than one-fifth decrease in the number of participants who scores reflected a possible or probable depressive disorder. A statistical reduction of anxiety was not found. While reductions in depression, and to a lesser extent anxiety, were seen, a significant proportion of participants still present, post-treatment, with a possible or probable depressive disorder (25.84%) or anxiety disorder (47.19%). Data from previous studies in the field were used to compare these findings with the general population.[13,16] Acknowledging that direct comparison cannot be made, our study showed that the levels of possible and probable anxiety and depressive disorder after GAHT were still significantly higher than those reported in the general population (4.5% for possible or probable depressive disorder[16] and 34.5% for possible or probable anxiety disorder).[13] Whether these elevated levels will reduce further (when a longer use of GAHT +/− surgical interventions) needs to be explored. Thus, future longitudinal studies would benefit from following people for longer in order to track the longer-term impact of interventions.

These findings do confirm, once again, the high levels of possible anxiety and depressive disorders before GAHT and the benefit that this treatment brings. It highlights the need to facilitate the expedited use of GAHT to aid the reduction of poor mental health symptoms in the transgender population, when possible and appropriate. This conclusion supports the literature which has called for longitudinal studies such as this to replicate the findings from cross-sectional studies.[15,16,22,25,38,45] The large reduction in depression, comparing to anxiety, may indicate that GAHT targets the dysphoria that many people attending transgender health services present with, which is manifested as depression (rather than anxiety). The fact that many transgender people still feel anxious after GAHT may be because of the victimisation, discrimination and social rejection experienced by the transgender population.[3,5,10,53] Unfortunately for some, these experiences do not necessarily stop after initiation of GAHT, and in some cases, they may increase. Clinical and community services should take these findings into consideration and increase the support offered even after gender-affirming medical treatment is over.

Importantly, this study also highlights how levels of pre-treatment social support are predictive of reduced risk post-treatment, as higher levels of social support prior to receiving GAHT significantly predict a greater reduction in depression symptoms after 18 months of receiving GAHT. This indicates the importance of increasing social support in transgender people.

The study also found that those with higher levels of autistic spectrum condition (ASC) traits prior to the receipt of GAHT had lower reductions in anxiety symptoms. This could indicate that those with greater ASC traits have higher anxiety symptoms even after post-treatment or that higher ASC traits are simply reflective of difficulties in social interactions, as a result of being anxious, grounded in an individual's gender identity status. These findings corroborate cross-sectional studies which have shown that interpersonal interactions can have an impact upon transgender people's psychological well-being.[9,15,23,25,27] However, these results need to be interpreted with caution because of the lack of validity of the AQ-Short in this population.

TABLE 3 Predictive role of ethnicity, employment status, relationship status, assigned gender at birth, age, MSPSS and AQ-Short for change in HADS-D and HADS-A from T0 to T1

	HADS – D			HADS – A		
	Coef.	95% CI	*P*	Coef.	95% CI	*P*
Ethnicity (grouped as White and all other at T0)	−0.37	−3.19 to 2.45	.794	−0.67	−3.62 to 2.28	.652
Employment (grouped as unemployed and disabled and all other at T0)	−0.97	−2.75 to 0.81	.284	−0.51	−2.37 to 1.35	.591
Relationship (grouped as single, widowed, divorced/separated and other at T0 and all other)	0.31	−1.82 to 2.43	.776	−0.61	−2.83 to 1.62	.590
Assigned sex at T0	0.36	−1.21 to 1.93	.651	−0.17	−1.81 to 1.47	.841
Age at T0	0.06	−0.50 to 0.61	.843	−0.03	−0.08 to 0.03	.409
Mean MSPSS at T0	0.81	0.24 to 1.39	.006	0.56	−0.04 - 1.16	.065
Mean AQ – Short at T0	−0.04	−0.096 to 0.025	.250	−0.07	−0.13 to 0.05	.034
Constant	−1.93	−4.76 to 0.90	.179	0.41	−2.34 to 3.16	.787

With social support (from families, friends or significant others) being highlighted as a factor in depression, it is important to consider approaches to aiding in the building of social support networks for transgender people accessing transgender services. Such support does not necessarily need to be based in clinical settings, as online and offline peer support and the work of tertiary services are available, less stigmatizing and have demonstrated a positive impact.[29,30,54] Resources such as peer support workers (PSWs) and online peer-to-peer support may be a valuable tool to provide social support for transgender individuals awaiting and receiving GAHT. Additional support is recommended in particular to those with higher levels of ASC traits because of the lower reductions in anxiety symptoms found in this population. However, PSWs need to be appropriately trained in order to reduce any potential risks related to managing their own stress as well as power imbalances within peer and professional relationship.[30]

Regarding expedited access to GAHT, there are practical issues that must be considered.[55-57] Many countries either lack clinical services specializing in transgender health care or have significant waiting lists to access these services.[13,58-60] These practical issues surrounding accessing treatment can lead to self-prescribing of GAHT, with 23% of individuals referred to transgender health clinics using GAHT prior to their first appointment, 70% of which was sourced online.[61] Self-prescribing without medical oversight presents its own risks, most notably a lack of specialized knowledge required to minimize health risks.[61]

The issues surrounding access to GAHT may be compounded by how services are configured. The role of many mental health workers in transgender health services in the UK is seen as gatekeepers, focused on the assessment of transgender people with limited attention given to psychological support. A shift in roles from gatekeeping to tangible mental health support would allow for a partial addressing of the power imbalance between transgender people and mental health professionals. This would potentially allow for transgender people to feel more able to discuss issues without fear of rejection for treatment. This in turn would provide mental health professionals with the ability to focus more on supporting the mental health of those transgender people who need it. From the evidence provided by this study, it appears that this would be most important for those with low levels of social support and those with autistic traits.

The strengths of this paper are the large sample and the naturalistic prospective longitudinal design within one transgender health service, which allows for within-subject comparisons and so provides a highly valid insight into the impact of GAHT on depression and anxiety of treatment-seeking transgender people. It is important to add that this is a national NHS service, which offers gender-affirming medical interventions and assessment free at the point of access to people from different geographical regions within the UK. This study is one of the few in the literature currently addressing the role of GAHT on mental health with such a methodology.

There are, however, limitations to the study. We must acknowledge that participants may downplay their symptoms of depression and anxiety pre-assessment, for fear of not being treated. Consequently, participants may have attenuated their mental health symptoms because of concern of not being accepted for GAMT. This could indicate that the changes between T0 and T1 may be more significant than recorded. As the method of recruitment was through a transgender health clinic as part of the NHS, the findings are only generalizable to treatment-seeking transgender individuals. A limitation may also be the response rate. Although this is low, it is in line with other clinical studies and it may be a reflection of the unfunded nature of the study. Another limitation is that the sample consisted of predominantly white participants, which may explain why ethnicity was not a predictive factor and the lack of control group where the intervention (GAHT) is not being offered (eg waiting list), but this will have its limitations too. A full randomized control study within this area will not be ethically possible. Studies using data from a clinical setting must also be aware of the context in which their data are being gathered.

In conclusion, this study shows that the mental health of transgender individuals improves following GAHT, particularly for those who reported high levels of social support prior to receiving GAHT. These results highlight the important mental health benefits of GAHT and emphasises the need for interventions focused on developing social support for transgender individuals.

ACKNOWLEDGEMENTS

The authors thank all participants in the study for having given their time so generously.

CONFLICT OF INTEREST

The authors have no other conflicts of interest relevant to this article to disclose. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

AUTHORS' CONTRIBUTIONS

JA conceptualized the study and participated in the study design. ZA, SP, JA, WPB and BG handled data collection management. BG carried out the statistical analysis. ZA drafted the initial manuscript. SP drafted the initial methods and results section of the manuscript. All authors contributed to the critical review of the manuscript and approved the final version of the manuscript.

ORCID

Jon Arcelus https://orcid.org/0000-0002-3805-0180

REFERENCES

1. Birkett M, Newcomb ME, Mustanski B. Does it get better? A longitudinal analysis of psychological distress and victimization in lesbian, gay, bisexual, transgender, and questioning youth. *J Adolesc Health*. 2015;56(3):280-285.
2. Bouman WP, Arcelus J. *The Transgender Handbook: A Guide for Transgender People, Their Families, and Professionals*. New York: Nova Publishers; 2017.

3. Breslow AS, Brewster ME, Velez BL, Wong S, Geiger E, Soderstrom B. Resilience and collective action: Exploring buffers against minority stress for transgender individuals. *Psychol Sex Orientation Gender Div*. 2015;2(3):253-265.
4. Reisner SL, Vetters R, Leclerc M, et al. Mental health of transgender youth in care at an adolescent urban community health center: a matched retrospective cohort study. *J Adolesc Health*. 2015;56(3):274-279.
5. Carmel TC, Erickson-Schroth L. Mental health and the transgender population. *J Psychosoc Nurs Ment Health Serv*. 2016;54(12):44-48.
6. Claes L, Bouman WP, Witcomb G, Thurston M, Fernandez-Aranda F, Arcelus J. Non-suicidal self-injury in trans people: associations with psychological symptoms, victimization, interpersonal functioning, and perceived social support. *J Sex Med*. 2015;12(1):168-179.
7. Dhejne C, Van Vlerken R, Heylens G, Arcelus J. Mental health and gender dysphoria: A review of the literature. *Int Rev Psychiatry*. 2016;28(1):44-57.
8. Seelman KL, Colon-Diaz MJP, LeCroix RH, Xavier-Brier M, Kattari L. Transgender noninclusive healthcare and delaying care because of fear: connections to general health and mental health among transgender adults. *Transgender Health*. 2017;2(1):17-28.
9. Simons L, Schrager SM, Clark LF, Belzer M, Olson J. Parental support and mental health among transgender adolescents. *J Adolesc Health*. 2013;53(6):791-793.
10. Herman JL. Gendered Restrooms And Minority Stress: The Public Regulation of Gender and its Impact on Transgender People's Lives. *Journal of Public Management & Social Policy*. 2013;19(1):65 -80.
11. Meyer IH. Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: conceptual issues and research evidence. *Psychol Bull*. 2003;129(5):674-697.
12. Meyer IH, Schwartz S, Frost DM. Social patterning of stress and coping: does disadvantaged social statuses confer more stress and fewer coping resources? *Soc Sci Med*. 2008;67(3):368-379.
13. Bouman WP, Claes L, Brewin N, et al. Transgender and anxiety: a comparative study between transgender people and the general population. *Int J Transgenderism*. 2017;18(1):16-26.
14. Millet N, Longworth J, Arcelus J. Prevalence of anxiety symptoms and disorders in the transgender population: A systematic review of the literature. *Int J Transgenderism*. 2017;18(1):27-38.
15. Bouman WP, Davey A, Meyer C, Witcomb GL, Arcelus J. Predictors of psychological well-being among treatment seeking transgender individuals. *Sex Relats Ther*. 2016;31(3):359-375.
16. Witcomb GL, Bouman WP, Claes L, Brewin N, Crawford JR, Arcelus J. Levels of depression in transgender people and its predictors: results of a large matched control study with transgender people accessing clinical services. *J Affect Disord*. 2018;235:308-315.
17. Veale JF, Watson RJ, Peter T, Saewyc EM. Mental health disparities among Canadian transgender youth. *J Adolesc Health*. 2017;60(1):44-49.
18. Marshall E, Claes L, Bouman WP, Witcomb GL, Arcelus J. Non-suicidal self-injury and suicidality in trans people: A systematic review of the literature. *Int Rev Psychiatry*. 2016;28(1):58-69.
19. Butler C, Joiner R, Bradley R, et al. Self-harm prevalence and ideation in a community sample of cis, trans and other youth. *Int J Transgenderism*. 2019;20(4):447-458.
20. Davey A, Arcelus J, Meyer C, Bouman WP. Self-injury among trans individuals in transition and matched controls: Prevalence and associated factors. *Health Soc Care Community*. 2016;24:485-494.
21. Nobili A, Glazebrook C, Bouman WP, et al. Autistic traits in treatment-seeking transgender adults. *J Autism Dev Dis*. 2018;48(12):3984-3994.
22. Nobili A, Glazebrook C, Bouman WP, Baron-Cohen S, Allison C, Arcelus J. The stability of autistic traits in transgender adults following cross-sex hormone treatment. *Int J Transgender Health*. 2020;48(12):3984-3994.
23. Meier CSL, Fitzgerald KM, Pardo ST, Babcock J. The effects of hormonal gender affirmation treatment on mental health in female-to-male transsexuals. *J Gay Lesbian Mental Health*. 2011;15(3):281-299.
24. Budge SL, Adelson JL, Howard KAS. Anxiety and depression in transgender individuals: the roles of transition status, loss, social support, and coping. *J Consult Clin Psychol*. 2013;81(3):545-557.
25. Davey A, Bouman WP, Arcelus J, Meyer C. Social support and psychological wellbeing: a comparison of patients with gender dysphoria and matched controls. *J Sex Med*. 2014;1(12):2976-2985.
26. Hoffman B. An overview of depression among transgender women. *Depress Res Treat*. 2014;2014:394283.
27. Olson KR, Durwood L, DeMeules M, McLaughlin KA. Mental health of transgender children who are supported in their identities. *Pediatrics*. 2016;137(3):e20153223.
28. Naslund JA, Aschbrenner KA, Marsch LA, Bartels SJ. The future of mental health care: peer-to-peer support and social media. *Epidemiol Psychiatr Sci*. 2016;25(2):113-122.
29. Repper J, Carter T. A review of the literature on peer support in mental health services. *J Ment Health*. 2011;20(4):392-411.
30. Jefferson K, Neilands TB, Sevelius J. Transgender women of color:-discrimination and depression symptoms. *Ethn and Inequal Health Soc Care*. 2013;6(4):121-136.
31. Dooley D, Catalano R, Wilson G. Depression and unemployment: panel findings from the Epidemiologic Catchment Area study. *Am J Community Psychol*. 1994;22(6):745-765.
32. Leppel K. The labor force status of transgender men and women. *Int J Transgenderism*. 2016;17(3–4):155-164.
33. Mizock L, Mueser KT. Employment, mental health, internalized stigma, and coping with transphobia among transgender individuals. *Psychol Sex Orientation Gender Div*. 2014;1(2):146-158.
34. Mizock L, Woodrum TD, Riley J, Sotilleo EA, Yuen N, Ormerod AJ. Coping with transphobia in employment: strategies used by transgender and gender-diverse people in the United States. *Int J Transgenderism*. 2017;18(3):282-292.
35. Witcomb GL, Claes L, Bouman WP, Nixon E, Motmans J, Arcelus J. Experiences and psychological wellbeing outcomes associated with bullying in treatment-seeking transgender and gender-diverse youth. *LGBT Health*. 2019;6(5):216-226.
36. Arcelus J, Claes L, Witcomb GL, Marshall E, Bouman WP. Risk factors for non-suicidal self-injury among trans youth. *J Sex Med*. 2016;13(3):402-412.
37. Colizzi M, Costa R, Todarello O. Transsexual patients' psychiatric comorbidity and positive effect of cross-sex hormonal treatment on mental health: results from a longitudinal study. *Psychoneuroendocrinology*. 2014;39(1):65-73.
38. Heylens G, Verroken C, De Cock S, T'Sjoen G, De Cuypere G. Effects of different steps in gender reassignment therapy on psychopathology: A prospective study of persons with a gender identity disorder. *J Sex Med*. 2014;11(1):119-126.
39. Bränström R, Pachankis JE. Reduction in mental health treatment utilization among transgender individuals after gender-affirming surgeries: a total population study. *Am J Psychiatry*. 2019.
40. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand*. 1983;67(6):361-370.
41. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the hospital anxiety and depression scale. An updated literature review. *J Psychosom Res*. 2002;52(2):69-77.
42. Gómez-Gil E, Zubiaurre-Elorza L, Esteva I, et al. Hormone-treated transsexual report less social distress, anxiety, and depression. *Psychoneuroendocrinology*. 2012;37(5):662-670.
43. Zimet GD, Powell SS, Farley GK, Werkman S, Berkoff KA. Psychometric characteristics of the Multidimensional Scale of Perceived Social Support. *J Pers Assess*. 1990;55:610-617.

44. Bouman WP, Claes L, Marshall E, et al. Socio-demographic variables, clinical features and the role of pre-assessment cross-sex hormones in older trans people. *J Sex Med*. 2016;13(4):711-719.
45. Hoekstra RA, Vinkhuyzen AAE, Wheelwright S, et al. The construction and validation of an abridged version of the Autism-Spectrum Quotient (AQ-Short). *J Autism Dev Disord*. 2011;41(5):589-596.
46. Woodbury-Smith MR, Robinson J, Wheelwright S, Baron-Cohen S. Screening adults for Asperger Syndrome using the AQ: a preliminary study of its diagnostic validity in clinical practice. *J Autism Dev Disord*. 2005;35(3):331-335.
47. Baron-Cohen S, Wheelwright S, Skinner R, Martin J, Clubley E. The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Males and Females, Scientists and Mathematicians. *J Autism Dev Disord*. 2001;31(1):5-17.
48. Pasterski V, Gilligan L, Curtis R. Traits of autism spectrum disorders in adults with gender dysphoria. *Arch Sex Behavior*. 2014;43(2):387-393.
49. StataCorp *Stata Statistical Software: Release 16*. College Station, TX: StataCorp LLC; 2019.
50. Allison PD. Comparing logit and probit coefficients across groups. *Sociol Methods & Res*. 1999;28(2):186-208.
51. Health Research Authority (HRA). Guidance for NHS Research Studies. London: Health Research Authority; 2013.
52. Fuller KA, Riggs DW. Family support and discrimination and their relationship to psychological distress and resilience amongst transgender people. A systematic review. *Int J Transgenderism*. 2018;19(4):379-388.
53. Bockting WO, Miner MH, Swinburne Romine RE, Hamilton A, Coleman E. Stigma, mental health, and resilience in an online sample of the US transgender population. *Am J Public Health*. 2013;103(5):943-951.
54. Bouman WP, de Vries ALC, T'Sjoen G. Gender Dysphoria and Gender Incongruence: An evolving inter-disciplinary field. *Int Rev Psychiatry*. 2016;28(1):1-4.
55. Motmans J, Nieder TO, Bouman WP. Transforming the paradigm of nonbinary transgender health: A field in transition. *Int J Transgenderism*. 2019;20(2-3):118-124.
56. Richards C, Arcelus J, Barrett J, et al. Trans is not a disorder – but should still receive funding. *Sex Relats Ther*. 2015;30(3):309-313.
57. Mepham N, Bouman WP, Arcelus J, Hayter M, Wylie KR. People with gender dysphoria who self-prescribe cross-sex hormones: prevalence, sources, and side effects knowledge. *J Sex Med*. 2014;11(12):2995-3001.
58. Bergero-Miguel T, Garcia-Encinas MA, Villena-Jimena A, et al. Gender Dysphoria and social anxiety: An exploratory study in Spain. *J Sex Med*. 2016;13(8):1270-1278.
59. Bouman WP, Richards C. Diagnostic and treatment issues for people with gender dysphoria in the United Kingdom. *Sex Relats Ther*. 2013;28(3):165-171.
60. Davies A, Bouman WP, Richards C, et al. Patient satisfaction with gender identity clinic services in the United Kingdom. *Sex Relats Ther*. 2013;28(4):400-418.
61. Winter S, Diamond M, Green J, et al. Transgender people: health at the margins of society. *Lancet*. 2016;388(10042):390-400.

How to cite this article: Aldridge Z, Patel S, Guo B, et al. Long-term effect of gender-affirming hormone treatment on depression and anxiety symptoms in transgender people: A prospective cohort study. *Andrology*. 2021;9:1808–1816. https://doi.org/10.1111/andr.12884

Published in final edited form as:
J Okla State Med Assoc. 2018 October ; 111(8): 776–783.

Off-Label Medication use in Children, More Common than We Think: A Systematic Review of the Literature

H. Christine Allen, MD[a], **M. Connor Garbe, MPH**[b], **Julie Lees, MD**[b], **Naila Aziz, MD**[b], **Hala Chaaban, MD**[c], **Jamie L. Miller, PharmD, BCPS, BCPPS, FPPAG**[d], **Peter Johnson, PharmD, BCPS, BCPPS, FPPAG, FCCM**[d], **Stephanie DeLeon, MD**[b]

[a]Department of Pediatrics: Section of Pediatric Critical Care, University of Oklahoma College of Medicine, Oklahoma City, OK

[b]Department of Pediatrics: Section of General and Community Pediatrics, University of Oklahoma College of Medicine, Oklahoma City, OK

[c]Department of Pediatrics: Section of Neonatology, University of Oklahoma College of Medicine, Oklahoma City, OK

[d]University of Oklahoma College of Pharmacy, Department of Pharmacy: Clinical and Administrative Science, Oklahoma City, USA

Abstract

Content: Clinician prescribing of off-label medications is common due to a lack of pediatric-specific data regarding the dosing, efficacy and safety of medications regularly prescribed to children.

Objective: This systematic review summarizes the published incidence of off-label medication use in children from the past 10 years. We also performed a retrospective chart review to determine the incidence of off-label prescriptions for children seen in the OU Physicians clinics.

Data Sources: We conducted a literature search of PubMed and OVID Medline from 2007 to 2017. Search terms included off-label use of medications and all child. For the local review, the outpatient electronic medical record (EMR) was queried.

Study Selection: Studies were eligible for inclusion if the study included children < 18 years of age, defined off-label use in the paper, and included the incidence of off-label drug use.

Data Extraction: Each review author extracted the study data from their assigned studies. For the retrospective chart review, the EMR was queried for patients <21 years of age who had a clinic visit and received a new prescription during 2017.

Results: We identified 31 studies, with off-label prescription rates from 3.2 % to 95%. The local retrospective chart review included 1,323 prescriptions; 504 were off-label (38.1%) and 819 were

Address Correspondence to: H. Christine Allen, MD, Department of Pediatrics, Section of Pediatric Critical Care, University of Oklahoma Health Sciences Center, 1200 Everett Drive, Suite 8305, Oklahoma City, OK 73104, Phone: (405) 271-5211, Christine-Allen@ouhsc.edu, Fax: (405) 271-2945.

Conflicts of Interest and Disclosures
The authors declare no pertinent conflict of interest and did not receive funding for this project.

approved. The frequency of off-label prescriptions does not differ significantly between the meta-analysis from the systematic review and the local retrospective chart review (30.9% vs 38.1%).

Conclusions: The use of off-label medications in children remains a common practice for pediatric providers.

Introduction

In comparison to adults, there is limited data pertaining to the dosing, efficacy, and safety of medications in children. This relative lack of data can be attributed to many causes, including unfamiliarity with age-related developmental pharmacology in pediatric patients, ethical considerations with conducting pediatric research, and a lack of financial incentive for the pharmaceutical industry. As a result, over the years there have been many significant therapeutic misadventures with off-label use of medications involving children with thalidomide and chloramphenicol being prominent examples. This lack of knowledge regarding pediatric specific drug use is an on-going area of concern in need of significant research. To address these concerns, Congress approved several pieces of legislation over the last 30 years. Table 1 provides an overview of the actions created by each bill,[1] which were designed to stimulate research related to drug pharmacology in pediatric populations. Following introduction of these bills in the United States over the last 15 years, 1200 studies have been submitted to the Food and Drug Administration (FDA), resulting in changes to over 700 medication labels[2]. Other countries have also adopted similar legislation. The European Medicines Agency created the European Pediatric regulation in 2007 in an effort to facilitate the development and availability of medicines for children while Canada created the Pediatric Expert Advisory Committee in 2009 in order to promote the development and licensing of drugs for children.

Despite the passage of these bills and subsequent research, there is still a significant lack of both pediatric-specific drug information and governmental approvals for on-label pediatric prescribing. Until further research and applications occur, many clinicians are forced to prescribe medications off-label. In response to this practice gap, The American Academy of Pediatrics adopted a policy statement on the use of off-label medications in children; they defined off-label use as "use of a drug that is not included in the package insert (FDA-approved labeling) [and] does not imply improper, illegal contraindicated or investigational use"[3]. This statement also emphasizes that off-label use does not necessarily require prescribers to obtain informed consent if the decision to use the medication is supported by scientific or anecdotal evidence and is not investigational in nature. For example, enalapril has an FDA-approved indication for hypertension, heart failure, and asymptomatic left ventricular dysfunction in adults but only has a FDA-labeled indication for hypertension in children.[4] Despite this, enalapril is a commonly used to treat heart failure in pediatric patients and informed consent is not required in these situations. It is important to recognize that designation of off-label use can refer not only to the clinical indication for which a drug is prescribed but also includes administration of a medication by any route or dosing scheme that is not included in the package labeling approved by the FDA. For example, dexmedetomidine, a commonly used sedative is labeled for administration via the IV route

only in adults but is used off-label when delivered via the inhaled route in adults and children.

Despite the adoption of the previous legislation and subsequent studies, many medications continue to be used off-label in children. The Department of Pediatrics at The University of Oklahoma Health Sciences Center (OUHSC) is in the process of joining an ongoing national study, entitled "Pharmacokinetics of Understudied Drugs Administered to Children Per Standard of Care" (POPS). This trial aims to provide more data regarding drug administration in children in an effort to decrease off-label use of medications and improve FDA-labeling of medications for pediatric patients. As the study enrolls patients and gathers medication-specific data, study medications change over time; currently POPS is studying 32 drugs, which are regularly prescribed in an off-label manner. The purpose of this systematic review is to determine the incidence of off-label medication use in children across varying health care settings over the previous 10 years. We also evaluated the off-label use of medications in the outpatient pediatric clinics as part of our center's preparation for participation in the POPS study.

Methods

First, we conducted a literature review inclusive of prospective and retrospective studies on off-label use of medications in pediatric patients in any healthcare setting using PubMed (National Center for Biotechnology Information, U.S. National Library of Medicine, Bethesda MD), and OVID Medline (National Library of Medicine, Bethesda MD) databases. The initial search was performed April 6, 2018, and a second search was performed June 12, 2018. Search terms included all child (0–18 years), off-label use of medications with publication dates of 2007 to 2017. The studies included were limited to those written in English with the full article available through the OUHSC Bird Library (Appendix 1). Literature reviews, letters to the editor, or opinion papers were excluded. Studies were eligible for final inclusion if the study population included children, defined off-label use in the paper, and included the incidence of off-label drug use. As our objective was to report the overall incidence of off-label use in varying health care settings, studies were included if they focused on multiple classes of medications in any health care setting and excluded if they reported the incidence of off-label use of only a single medication or medication class. Eligibility assessment was done by two authors (CA &SD); article titles and abstracts were reviewed in an unblinded, standardized manner. Disagreements between the primary reviewers were resolved by consensus opinion of four of the authors (CA, SD, PJ, & JM). Studies from all countries were included, and we relied on the individual authors' interpretation of the local regulatory agency's ruling to determine off-label use of medications in the country the article was published.

The included studies were then divided amongst five of the authors (CA, SD, HC, NA, & JL) for review and extraction of study variables. A data extraction tool was created and pilot tested on 5 randomly selected articles; no changes were made to the data extraction tool after the pilot testing. Each review author extracted the study data from their assigned included studies. A single author (CA) checked extracted data, and disagreements were presented to a third author (SD) for resolution. Study variables included year of publication, years of data

collection or duration of data collection, study design, age range of patient population included in the study, setting of the study (i.e. inpatient unit vs. outpatient clinic), geographic location of study, total number of prescriptions, the rate of off-label medication use, and study definition of off-label use. The reason (i.e. dose, indication, frequency or route) drugs were determined to be off-label, adverse drug reactions, and handling of parental consent for off-label drug use was also collected if clearly stated in the paper.

There was no separate assessment of bias in the systematic review as all studies were either retrospective or prospective observational. There is the inherent bias due to the nature of the study design of retrospective and prospective observational studies.

Next we conducted a retrospective chart review of our electronic medical record system (Centricity EMR, GE Healthcare) to determine the prevalence of off-label mediation use in children at our institution. Data were collected for patients 0 to 20 years of age, as POPS is enrolling patients through age 20, who had at least one outpatient visit at our institution during 2017, and received a prescription with a start date in 2017 for one of the POPS current 32 drugs of interest listed in Table 2. Data collected included the generic name of the drug prescribed, date of prescription, zip code to determine rural/urban status, patient race, patient ethnicity, and determination of off-label or approved use of the drug; if the drug was used in an off-label manner, the reason was noted. IBM Micromedex (IBM Corporation) was used to determine the Food and Drug Administration approved age range, route, dose and indication for medications. Two authors (CA and SD) reviewed and discussed unclear cases until consensus reached. The local retrospective study was approved by the Institutional Review Board at OUHSC.

Descriptive statistics for the local data were calculated for demographic variables and outcome of interest, prescription off-label status (approved vs. off-label). Chi squares were calculated for categorical demographics (sex, rural/urban status, race, ethnicity), and log binomial regression was used for continuous predictors (age). A p value of < 0.05 was considered significant. Zip codes were used to determine rural/urban designation with rural urban commuting area (RUCA) codes, using the four category classification as defined by the Rural Health Research Center[5]. Comparisons of meta-analyses from the systematic review and local data were performed by observing proportions and confidence intervals of off-label prescriptions, and the frequencies of reasons for being off-label (age, dose, indication, and route), among off-label prescriptions. Meta-analyses were performed using MedCalc software's meta-analysis function for proportions for the systematic review. Analyses on local data were performed using SAS 9.4. Comparisons between Meta-analyses and local data were performed by observing estimated proportions of off-label prescriptions and their respective 95% confidence intervals.

Results

During the literature review, 285 publications were initially identified. Eight duplicates were removed, leaving 277 studies for initial screening. 54 studies met initial screening criteria and were sent on for full review. From these, 23 studies were ultimately excluded; five studies were not available in full manuscript, eleven were specific for a single medication or

medication class, and seven published no incidence of off-label prescriptions (Diagram 1). Thirty-one studies were included in final analysis, including16 retrospective studies and 15 prospective observational studies. Pediatric patients from infants to adolescents were included in 24 studies, while 7 studies included only neonates. A total of 19 of the studies were conducted in inpatient populations, 11 studies were outpatient and one study included both inpatient and outpatient locations at a single center. The majority of the studies were conducted in Europe (n=19), Asia (n=6) and Australia (n=3). The number of patients involved in each study varied between 81 and 1.9 million patients. The 31 studies in the systematic review had varied age ranges but overall included patients that were preterm infants to 19 years old

Off-label use was described in the included 31 studies. Use of the medication outside the package insert recommendations for indication, age, route, dose and frequency was the most common definition of off-label use. Four studies further defined off-label as "use of medications that had no pediatric indication or were contraindicated in children" [6–8]. The rate of off-label prescriptions varied widely in the 31 reviewed studies and was reported to be between 3.2% and 95%. Seven studies in our literature review specifically evaluated the use of off-label medications in the neonate population, with off-label drug use ranging from 26% to 95%. See Table 3 for individual article details.

Only two studies reported adverse drug reactions related to off-label use of medications. One study did not define or elaborate on the nature of the adverse drug reaction, however, the second study reported fever, diarrhea and rash as an adverse drug reaction secondary to off-label use of medications [9, 10]. No study commented on obtaining parental consent for the use of off-label medications.

In our local retrospective chart review 22 of the 32 POPS drugs of interest were prescribed as a new prescription in 2017, resulting in a total of 1,323 prescriptions for 1079 patients. Of the 1,323 prescriptions, 504 prescriptions were off-label (38.1%) and 819 prescriptions were on-label. Table 4 provides a list of the 22 included drugs and the off-label versus approved use prescription percentages. The reasons the prescriptions were classified as off-label are listed in Table 5, with patient age being the most common reason. Table 6 displays patient demographics and frequency of off-label prescriptions defined by demographic grouping. There was no significant difference in the frequency of off-label prescriptions between male and female patients (38.6%, 37.6%). Off-label prescription frequency did not significantly differ between urban and rural groups (37.7%, 39.1) or by ethnicity (Hispanic/Latino vs. Not Hispanic/Latino 41.4% v. 38.2%). Though ostensibly there is large range in off-label frequency between racial groups (27.8%–46.6%), this was not statistically significant. There is, however, a significant association between age (in years) and frequency of being prescribed an off-label drug ($p<0.001$). The probability of receiving an off-label prescription is reduced by 3% for every year increase in age, with the probability at less than a year old being 51.6% and the probability at 20 years old being 29.3%.

To estimate a combined proportion of off-label prescriptions from the systematic review for comparison with our own data, we performed a meta-analysis for proportions (Figure 1). The I^2 statistic indicated that over 99% of the variation in proportions across studies was due

to heterogeneity, and thusly the estimated proportion derived from the random effects model used was 30.9%, (95% CI: 26.0–36.0) for the systematic review. Comparing the results of local data (38.1%, 95%CI: 35.5, 40.8) with the meta-analysis from the systematic review, the frequency of off-label prescription does not differ significantly (Table 7). When comparing the reasons for off-label prescription (among those studies that listed them), OU Physicians pediatric outpatient clinics showed a significantly larger proportion of off-label prescriptions due to age (74.5%, 25.6%), and a significantly smaller proportion due to dose (10.3%, 48.3%). There was no significant difference between proportions of off-label prescriptions due to indication (21.4%, 19.5%) or route (3.2%, 3.4%).

Discussion

Our study and literature review demonstrate that off-label use of medications in pediatric patients is a common practice, with a significant number of children, inpatient and outpatient, receiving an off-label medication. With the exception of three of the reviewed studies [7, 11, 12], all studies demonstrate >12% of prescriptions are off-label. Though most of the articles reviewed were from European countries, all continents were represented. The historical practice of using off-label medications in children is a frequent worldwide occurrence and continues to be an issue despite the increased awareness and passed legislation. Eighteen of the reviewed studies were conducted in European countries and published after the European Union Pediatric Regulation was established in 2007, which proposed to stimulate pediatric specific research much like the United States regulations reviewed in the introduction. The single included United States study in children was published more than 10 years after FDAMA and Pediatric Rule, yet it continues to report a high incidence of off-label medication prescriptions (36%).[13] Our local study for new prescriptions in 2017 also demonstrated a high incidence of off-label use at 38.1%. Despite international government efforts to mandate further research for pediatric drug labeling, there continues to be a high rate of off-label use of medications.

The reason for off-label use varies among studies. Our local study had a significantly higher proportion of patients that were given a medication off-label due to age and fewer that were off-label due to dose when compared to the meta-analysis from the systematic review. As the majority of the studies included in the systematic review were outside of the United States, this may in part be due to different approved dosing regimens from the local regulatory entities in the various countries that the systematic review studies were conducted. In addition, we only included the 32 POPS drugs of interest in our local study, this may account for the disparity between age and dosage between our study and the meta-analysis.

The literature review demonstrates the incidence of off-label use is higher among younger populations, especially neonates as reflected in the neonate intensive care unit-specific articles with off-label prescription rates of at least 26%. While our local study was not focused primarily on the neonate population, we did find a significant number of off-label prescriptions were due to patient age and the risk of receiving an off-label prescription decreased as age increased.

Finally, none of the studies we reviewed addressed the issue of parental consent for off-label use despite two studies reporting an adverse drug reaction secondary to off-label medication use. This may in part be secondary to the above-referenced American Academy of Pediatrics adopted policy that specifically states providers do not need parental consent if the medication use is supported by current evidence in the literature or a practitioners experience with the medication. Additionally, there may be a lack of knowledge among prescribers about the on label indications for commonly used pediatric medications. Practitioners should be aware of the off-label medications they commonly use in children and be mindful of known potential adverse reactions and side effects but also be wary of possible new unreported drug reactions or side effects.

The real answer to this prescribing dilemma is more research in the pediatric population. Hopefully, the POPS study will generate pediatric safety data and result in FDA approval for many of the commonly prescribed medications currently used off-label, thereby eliminating the conflict providers may feel in the current prescribing environment.

Our literature review was limited to the previous ten years and excluded studies evaluating off-label use of a single specific drug or a specific class of drugs. Non-English language articles were not included in the final analysis. Exclusion of these types of studies may have limited the scope of our review. The systematic review included studies from multiple countries, a broad age range and settings such as inpatient and outpatient, which we felt would give us a current global perspective on off-label medication use but perhaps introduced bias as each setting and country defined off-label based on their local regulatory board. The studies included in the systematic review were also either retrospective or prospective observational studies, which have inherent increased risk of bias compared to randomized trials.

The local retrospective study was limited to the 32 drugs of interest in the POPs study that are frequently prescribed off-label to children. The local prevalence of off-label use could be significantly different in our patient population if inpatient prescriptions or all prescriptions were included in analysis.

Conclusions

Off-label medication use is common in the pediatric population, especially in neonates and younger age groups. More age-specific research is needed to provide adequate drug safety and effectiveness for children. Until more data is provided, clinical decision making should be guided by the best available evidence.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledegements

Research reported in this publication was supported by the Office of the Director, National Institutes of Health under Award Numbers UG1OD024950 and US4GM104938. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. [cited 2018 May 10]; Pediatric Product Development]. Available from: http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DevelopmentResources/ucm049867.htm..
2. Green DJ ZI, Burckart GJ, Pediatric Drug Development: outlook for science-based innovation. Clinical Pharmacology and Therapeutics 2018 103: p. 376–8. [PubMed: 29384202]
3. Frattarelli DA GJ, Green TP et al., Off-Label drugs in children. Pediatrics, 2014 133: p. 563–7. [PubMed: 24567009]
4. in EXPANED (package insert). Greenwood Village, CO: Silvergate Pharmaceuticals, Inc.
5. http://depts.washington.edu/uwruca/ruca-maps.php[7/6/201810:57:06AM
6. Lass J, et al., Off label use of prescription medicines in children in outpatient setting in Estonia is common. Pharmacoepidemiol Drug Saf, 2011 20(5): p. 474–81. [PubMed: 21416552]
7. Carnovale C, et al., Paediatric drug use with focus on off-label prescriptions in Lombardy and implications for therapeutic approaches. Eur J Pediatr, 2013 172(12): p. 1679–85. [PubMed: 23913312]
8. Olsson J, et al., Paediatric drug use with focus on off-label prescriptions in Swedish outpatient care--a nationwide study. Acta Paediatr, 2011 100(9): p. 1272–5. [PubMed: 21438922]
9. Ballard CD, et al., Off-label use of medicines in paediatric inpatients at an Australian teaching hospital. J Paediatr Child Health, 2013 49(1): p. 38–42. [PubMed: 23279160]
10. Palmaro A, et al., Off-label prescribing in pediatric outpatients. Pediatrics, 2015 135(1): p. 49–58. [PubMed: 25511119]
11. Langerova P, Vrtal J, and Urbanek K, Incidence of unlicensed and off-label prescription in children. Ital J Pediatr, 2014 40: p. 12. [PubMed: 24495454]
12. Muhlbauer B JK, Pichler J, Schoettler P, Off-label use of prescription drugs in childhood and adolescence: an analysis of prescription patterns in Germany. Dtsch Arztebl Int 2009 106(3): p. 25–31. [PubMed: 19564959]
13. Maltz LA, et al., Off-label drug use in a single-center pediatric cardiac intensive care unit. World J Pediatr Congenit Heart Surg, 2013 4(3): p. 262–6. [PubMed: 24327493]
14. Morales-Carpi C, et al., Drug utilization and off-label drug use among Spanish emergency room paediatric patients. Eur J Clin Pharmacol, 2010 66(3): p. 315–20. [PubMed: 19936729]
15. Ribeiro M, Jorge A, and Macedo AF, Off-label drug prescribing in a Portuguese paediatric emergency unit. Int J Clin Pharm, 2013 35(1): p. 30–6. [PubMed: 23054136]
16. Lee JL, Redzuan AM, and Shah NM, Unlicensed and off-label use of medicines in children admitted to the intensive care units of a hospital in Malaysia. Int J Clin Pharm, 2013 35(6): p. 1025–9. [PubMed: 24022725]
17. Blanco-Reina E M-CA, Vega-Jimenez MA, Ocana-Riola R, Marquez-Romero EI, Ruiz-Extremera A, Drug Utilization pattern in children and off-label use of medicines in a pediatric intensive care unit. Medicina Intensiva 2014 40(1): p. 1–8.
18. Berdkan S RL, Hajj A, Eid B, Jabbour H, Osta NE, Karam L, Khabbaz LR. , Comparative Assessment of Off-Label and Unlicensed Drug Prescriptions in Children: FDA versus ANSM Guidlines Clin Ther, 2016 38(8): p. 1833–44.
19. Lindell-Osuagwu L, et al., Prescribing for off-label use and unauthorized medicines in three paediatric wards in Finland, the status before and after the European Union Paediatric Regulation. J Clin Pharm Ther, 2014 39(2): p. 144–53. [PubMed: 24329556]
20. Aamir M, et al., Unlicensed and off-label use of drugs in pediatric surgical units at tertiary care hospitals of Pakistan. Int J Clin Pharm, 2017 39(4): p. 860–866. [PubMed: 28597174]
21. Czarniak P, et al., Clinical setting influences off-label and unlicensed prescribing in a paediatric teaching hospital. PLoS One, 2015 10(3): p. e0120630.
22. Teigen A, et al., Off-label and unlicensed medicines to hospitalised children in Norway. J Pharm Pharmacol, 2017 69(4): p. 432–438. [PubMed: 27334565]
23. Joret-Descout P, et al., Off-label and unlicensed utilisation of medicines in a French paediatric hospital. Int J Clin Pharm, 2015 37(6): p. 1222–7. [PubMed: 26394784]

24. Taylor D JP, Taylor SE., Jones A, Cheek JA, Craig SS, Graudins A, Dhir R, Krieser D, Babl FE, , Off-Label and unlicensed medicine adminsitration to pediatric emergency department patients. Emergency Medicine Australasia, 2015 27: p. 440–446.
25. Hsien L, et al., Off-label drug use among hospitalised children: identifying areas with the highest need for research. Pharm World Sci, 2008 30(5): p. 497–502. [PubMed: 18219585]
26. Abdulah R, et al., Off-label paediatric drug use in an Indonesian community setting. J Clin Pharm Ther, 2015 40(4): p. 409–12. [PubMed: 25884353]
27. Knopf H, et al., Off-label medicine use in children and adolescents: results of a population-based study in Germany. BMC Public Health, 2013 13: p. 631. [PubMed: 23822744]
28. Palcevski G, Skocibusic N, and Vlahovic-Palcevski V, Unlicensed and off-label drug use in hospitalized children in Croatia: a cross-sectional survey. Eur J Clin Pharmacol, 2012 68(7): p. 1073–7. [PubMed: 22307225]
29. Cuzzolin L AR, Off-lable and unlicensed drug treatments in Neonatal INtensive Care Units: an Italian multicentere study. Eur J Clin Pharmacol, 2016 72(1): p. 117–23. [PubMed: 26490354]
30. Kieran EA, O'Callaghan N, and O'Donnell CP, Unlicensed and off-label drug use in an Irish neonatal intensive care unit: a prospective cohort study. Acta Paediatr, 2014 103(4): p. e139–42. [PubMed: 24397831]
31. Schweigertova J DA, Doinikova D, Ondriasova E, Balazova M, Slezakova V, Kuzelova M, Off Label and unlicensed use of medicinal products in the neonatal setting in the SLovak Republic. Pediatri Int, 2016 58(2): p. 126–31.
32. Jain S, et al., Off-label use of drugs in neonatal intensive care units. Indian Pediatr, 2014 51(8): p. 644–6. [PubMed: 25128998]
33. de Souza AS Jr., et al., Off-label use and harmful potential of drugs in a NICU in Brazil: A descriptive study. BMC Pediatr, 2016 16: p. 13. [PubMed: 26795213]
34. Silva J F.-d.-L.F., Soares H, Guimaraes H, Off-Label and Unlicensed Drug Use in Neonatology: Reality in a Portugese University. Acta Med Port, 2015 28(3): p. 297–306. [PubMed: 26421781]
35. Chauthankar SA,MP, Potey AV, Nanavati RN , Drug Utilization in Neonatal Intensive Care Unit of a Tertiary-care Hospital in Mumbai India. Indian Pediatrics 2017 54: p. 931–934. [PubMed: 28849769]
36. Pereira Gomes V, et al., Off-label and unlicensed utilization of drugs in a Brazilian pediatric hospital. Farm Hosp, 2015 39(3): p. 176–80. [PubMed: 26005894]

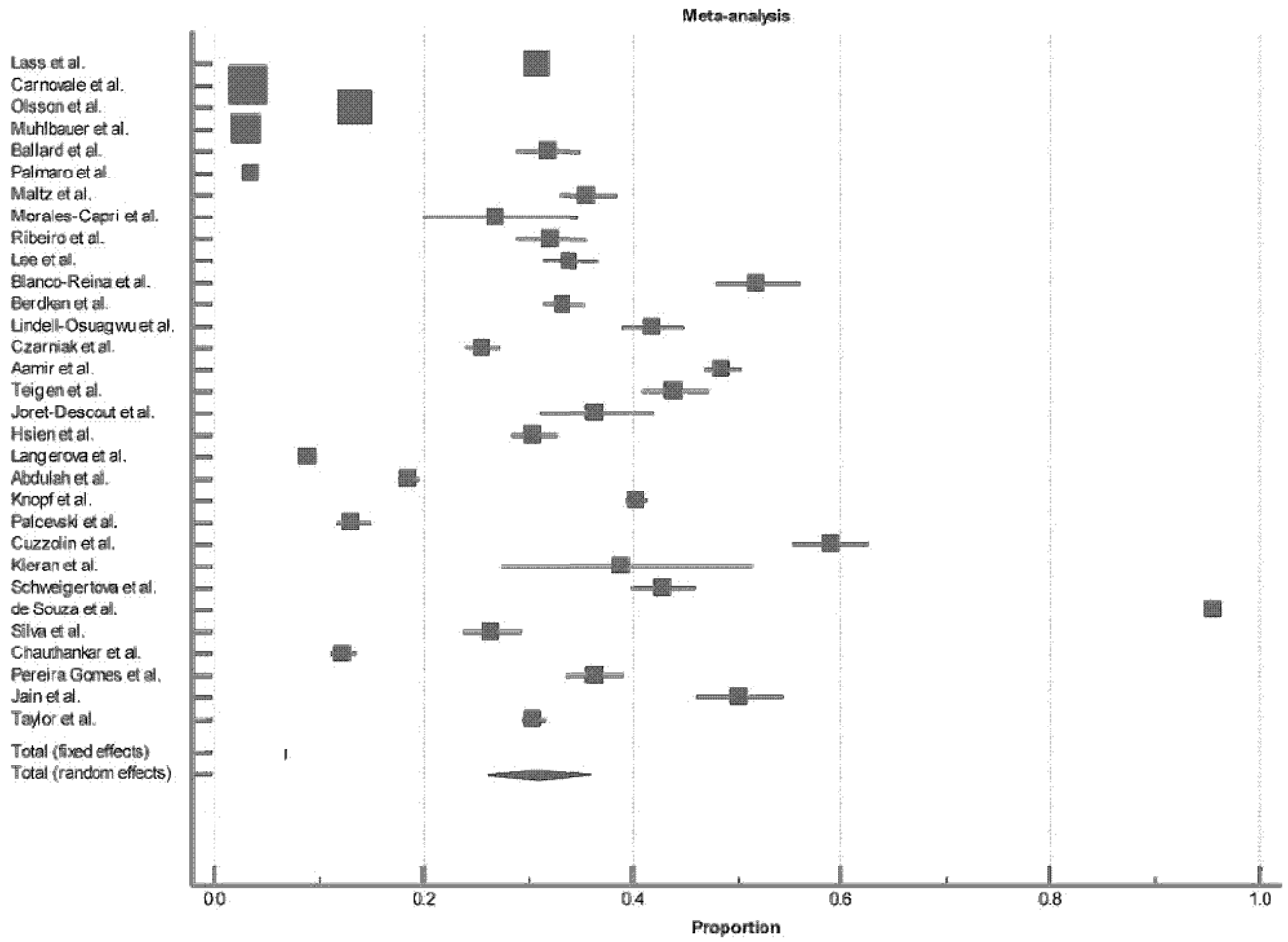


Figure 1:
Meta-analysis Forest Plot

Table 1:

Overview of Federal Legislation to Promote Pediatric Studies (Data from U.S. Drug and Food Administration[1])

Legislation [Year Enacted]	Implications of Legislation
Food and Drug Modernization Act (FDAMA) [1997]	· Encouraged pharmaceutical manufacturers to perform pediatric studies · Offered 6 months patent exclusivity as financial incentive
Pediatric Rule [1998]	· Required efficacy & safety testing for NDAs if medication could be used in children · Studies under Pediatric Rule were eligible for 6 month patent extension under FDAMA · Federal court overturned in 2002 because FDA did not have authority
Best Pharmaceuticals Act for Children (BCPA) [2002]	· Authorized FDA to request pediatric studies with NDAs (including orphan drugs) for new indications · Extended 6 month patent exclusivity through 2007 · Required NIH to publish list of needs for future study in children
Pediatric Research Equality Act (PREA) [2003]	· Required pediatric assessment & development of PSP with NDAs (expanded version of Pediatric Rule) for certain drugs (NOT orphan drugs) · Allowed for modifications to existing indications · Pediatric plan must be developed before approval in adults
Food and Drug Administration Amendments Act (FDAAA) [2007]	· Reauthorized PREA & BPCA × 5 years · Expanded the BCPA so FDA could issue request for > 1 indication (i.e., "on" & "off-label" use) · Introduced Pediatric Medical Device Safety & Improvement Act
Food and Drug Administration Safety and Administration Act (FDASIA) [2012]	· Made BPCA & PREA permanent

Table 2:

List of 32 POPS drugs of interest

1. Alfentanil
2. Amikacin
3. Atropine
4. Cefepime
5. Ceftazidime
6. Cidofovir
7. Ciprofloxacin
8. Clozapine
9. Dexmedetomidine
10. Diazepam
11. Etomidate
12. Fosphenytoin
13. Haloperidol
14. Low Molecular Weight Heparin
15. Hydromorphone
16. Lidocaine
17. Lurasidone
18. Meropenem
19. Methadone
20. Methylprednisolone
21. Midazolam
22. Molindone
23. Nafcillin
24. Pentobarbital
25. Piperacillin
26. Timolol
27. Tobramycin
28. Valproic Acid
29. Vancomycin
30. Vecuronium

31. Warfarin

32. Ziprasidone

Table 3:

List of studies included in the literature review.

Author	Year Published	Type of Report (# Patients)	Age Range	Location	Country of Origin	% Off-Label Prescriptions
Lass et al [6]	2011	Retrospective (n=151,476)	< 19 years	Multi-center outpatient clinics	Estonia	31%
Carnovale et al. [7]	2013	Retrospective database search (n=1,708,755)	0–18 year	Single-center outpatient clinics	Italy	3.3%
Olsson et al. [8]	2011	Retrospective database search (n=1,911,417)	0–18 years	Multi-center outpatient clinics	Sweden	13.5%
Muhlbauer et al [12]	2009	Retrospective database search (n=289,000)	0–16 years	Multi-center outpatient clinics	Germany	3.2%
Ballard et al. [9]	2013	Retrospective (n=300)	< 12 years	Single center general pediatric wards	Australia	32%
Palmaro et al.[10]	2015	Prospective (n=2,313)	0–16 years	Multi-center outpatient clinics	France	37.6%
Maltz et al.[13]	2013	Retrospective (n=82)	< 18 years	Single-center CVICU[3]	USA	36%
Morales-Capri et al.[14]	2010	Prospective observational (n=462)	< 14 years	Single center emergency department	Spain	27%
Ribeiro et al. [15]	2013	Retrospective (n=700)	< 18 years	Single-center emergency department	Portugal	32.2%
Lee et al. [16]	2013	Prospective observational (n=192)	Preterm to 18 years	Single-center PICUs	Malaysia	34.1%
Blanco-Reina et al. [17]	2014	Prospective observational (n=81)	Birth to 14 years	Single-center PICU and NICU[4]	Spain	52%
Berdkan et al.[18]	2016	Retrospective (n=500)	1 day to 16 years	Multi-center including multiple wards	Lebanon	30.2%
Lindell- Osuagwu et al.[19]	2014	Retrospective (n=123)	< 18 years	Single-center including multiple wards	Finland	42%
Aamir et al.[20]	2017	Prospective observational (n=895)	< 18 years	Multi-center Surgical wards	Pakistan	48.2%
Czarniak et al.[21]	2015	Retrospective (n=699)	All ages	Single Center inpatient and outpatient	Australia	25.7%
Teigen et al. [22]	2017	Prospective (n=400)	0–17 years	Multi-center including multiple wards	Norway	44%
Joret-Descout et al. [23]	2015	Retrospective (n=120)	0–18 years	Single-center including multiple wards	France	36.5%
Taylor et al. [24]	2015	Retrospective (n=3343)	0–17 years	Multi-center outpatient	Australia	30.5%
Hsien et al. [25]	2008	Prospective observational (n=417)	All ages	Single-center including multiple wards	Germany	31%
Langerova et al.[11]	2014	Prospective (n=4,282)	0–15 years	Single-center outpatient clinics	Czech Republic	9.01%
Abdulah et al.[26]	2015	Retrospective (n=4,936)	0–5 years	Multi-center outpatient clinics	Indonesia	18.6%
Knopf et al. [27]	2013	Retrospective (n=17,450)	0–17 years	Single-center outpatient clinics	Germany	40.2%
Palcevski et al. [28]	2012	Prospective (n=691)	0–19 years	Single-center including multiple wards	Croatia	13.3%
Cuzzolin et al.[29]	2016	Prospective (n=220)	Preterm/neonates	Multi-center NICUs	Italy	59%
Kieran et al.[30]	2014	Prospective (n=110)	Neonates	Single-center NICU	Ireland	39%

J Okla State Med Assoc. Author manuscript; available in PMC 2019 August 02.

Author	Year Published	Type of Report (# Patients)	Age Range	Location	Country of Origin	% Off-Label Prescriptions
Schweigertova et al.[31]	2016	Prospective (n=202)	< 29 days old	Multi-center NICUs	Slovakia	43%
Jain et al.[32]	2014	Prospective (n=156)	Neonates	Multi-center NICUs	India	26%
de Souza et al. [33]	2016	Retrospective (n=192)	< 28 days old	Single-center NICU	Brazil	95.6%
Silva et al. [34]	2015	Retrospective (n=218)	< 28 days old	Single-center NICU	Portugal	25.7%
Chauthankar et al.[35]	2017	Prospective observational (n=460)	Preterm/neonates	Single-center NICU	India	12.3%
Pereira Gomes et al. [36]	2015	Prospective observational (n=320)	2–18 years	Single- center inpatient ward	Brazil	57.2%

[1] ND = Not defined in study;

[2] PICU = Pediatric Intensive Care Unit;

[3] CVICU = Cardiac Intensive Care Unit;

[4] NICU = Neonatal Intensive Care Unit

Table 4:

Summary of Prescription Off-label Status by Drug (N=1323 prescriptions)

Drug	Off-label N (%)	Approved N (%)
amikacin	1 (50.0)	1 (50.0)
atropine	4 (100)	-
cefepime	2 (66.7)	1(33.3)
ceftazidime	-	1 (100)
ciprofloxacin	120 (66.7)	60 (33.3)
clozapine	3 (100)	-
diazepam	43 (22.3)	150 (77.7)
haloperidol	6 (54.6)	5 (45.4)
heparin	95 (100)	-
hydromorphone	6 (100)	-
lidocaine	9 (4.3)	201 (95.7)
lurasidone	18 (85.7)	3 (14.3)
methylprednisolone	-	285 (100)
midazolam	2 (66.7)	1 (33.3)
nafcillin	1 (100)	-
piperacillin	4 (100)	-
timolol	28 (100)	-
tobramycin	31 (31.3)	68 (68.7)
valproic acid	13 (61.9)	8 (38.1)
vancomycin	1 (4.3)	22 (95.7)
warfarin	89 (89.9)	10 (10.1)
ziprasidone	28 (90.3)	3 (9.7)

Table 5:

Summary of Off-label Status Frequency and Off-label Frequency by Reason

	n (%)
Off-label Status	
Approved	819 (61.9)
Off-label	504 (38.1)
Reason for Off-label Status*	
Age	338 (74.5)
Dose	52 (10.3)
Indication	108 (21.4)
Route	16 (3.2)

* Four "Off-label" missing reason. 14 prescriptions contributing to more than one reason

Table 6:

Summary of Demographic characteristics by Prescription Off-label Status (N=1,323 prescriptions)

		Off-label N (%)	Approved N (%)	
Sex				p=0.71
	Male	252 (38.6)	401 (61.4)	
	Female	252 (37.6)	418 (62.4)	
RUCA*				p=0.63
	Rural	144 (39.1)	224 (60.9)	
	Urban	360 (37.7)	595 (62.3)	
Race				Ref=White
	American Indian/ Alaska Native	27 (46.6)	31 (53.4)	p=0.21
	Asian	5 (27.8)	13 (72.2)	p=0.27
	Black/ African American	56 (36.1)	99 (63.9)	p=0.63
	Native Hawaiian/ Pacific Islander	5 (45.4)	6 (54.6)	p=0.63
	White	400 (38.1)	649 (61.9)	
Ethnicity				p=0.43
	Hispanic or Latino	72 (41.4)	102 (58.6)	
	Not Hispanic or Latino	421 (38.2)	680 (61.8)	
Age (years)				p<0.001**
	Mean (Standard Deviation)	10.2 (6.1)	11.8 (5.7)	
	Median	12	13	
	Min, Max (Interquarile Range)	<1 month, 19 (5.16)	<1 month, 20 (7.17)	

*Rural Urban Commuting Area

**Age in years: RR 0.97 (0.96–0.98)

Table 7:

Summary of Off-label Prescription Frequency and Off-label Prescription Frequency by Reason: OUHSC Children's Hospital Outpatient compared to Meta-Analysis

	OU Physicians Percent Off-label (95% CI)	Meta-Analysis Percent Off-label (95% CI)
Off-label	38.1 (35.5, 40.8)	30.9 (26.0, 36.0)
Reason for Off-label		
Age	74.5 (70.2, 78.4)	25.6 (15.4, 37.3)
Dose	10.3 (7.8, 13.3)	48.3 (35.7, 61)
Indication	21.4 (17.9, 25.3)	19.5 (12.4, 27.8)
Route	3.2 (1.8, 5.1)	3.4 (0.6, 8.3)

JAMA Surgery | Original Investigation

Association Between Gender-Affirming Surgeries and Mental Health Outcomes

Anthony N. Almazan, BA; Alex S. Keuroghlian, MD, MPH

Invited Commentary page 618

CME Quiz at jamacmelookup.com

IMPORTANCE Requests for gender-affirming surgeries are rapidly increasing among transgender and gender diverse (TGD) people. However, there is limited evidence regarding the mental health benefits of these surgeries.

OBJECTIVE To evaluate associations between gender-affirming surgeries and mental health outcomes, including psychological distress, substance use, and suicide risk.

DESIGN, SETTING, AND PARTICIPANTS In this study, we performed a secondary analysis of data from the 2015 US Transgender Survey, the largest existing data set containing comprehensive information on the surgical and mental health experiences of TGD people. The survey was conducted across 50 states, Washington, DC, US territories, and US military bases abroad. A total of 27 715 TGD adults took the US Transgender Survey, which was disseminated by community-based outreach from August 19, 2015, to September 21, 2015. Data were analyzed between November 1, 2020, and January 3, 2021.

EXPOSURES The exposure group included respondents who endorsed undergoing 1 or more types of gender-affirming surgery at least 2 years prior to submitting survey responses. The comparison group included respondents who endorsed a desire for 1 or more types of gender-affirming surgery but denied undergoing any gender-affirming surgeries.

MAIN OUTCOMES AND MEASURES Endorsement of past-month severe psychological distress (score of ≥13 on Kessler Psychological Distress Scale), past-month binge alcohol use, past-year tobacco smoking, and past-year suicidal ideation or suicide attempt.

RESULTS Of the 27 715 respondents, 3559 (12.8%) endorsed undergoing 1 or more types of gender-affirming surgery at least 2 years prior to submitting survey responses, while 16 401 (59.2%) endorsed a desire to undergo 1 or more types of gender-affirming surgery but denied undergoing any of these. Of the respondents in this study sample, 16 182 (81.1%) were between the ages of 18 and 44 years, 16 386 (82.1%) identified as White, 7751 (38.8%) identified as transgender women, 6489 (32.5%) identified as transgender men, and 5300 (26.6%) identified as nonbinary. After adjustment for sociodemographic factors and exposure to other types of gender-affirming care, undergoing 1 or more types of gender-affirming surgery was associated with lower past-month psychological distress (adjusted odds ratio [aOR], 0.58; 95% CI, 0.50-0.67; $P < .001$), past-year smoking (aOR, 0.65; 95% CI, 0.57-0.75; $P < .001$), and past-year suicidal ideation (aOR, 0.56; 95% CI, 0.50-0.64; $P < .001$).

CONCLUSIONS AND RELEVANCE This study demonstrates an association between gender-affirming surgery and improved mental health outcomes. These results contribute new evidence to support the provision of gender-affirming surgical care for TGD people.

Author Affiliations: Harvard Medical School, Boston, Massachusetts (Almazan, Keuroghlian); Harvard T. H. Chan School of Public Health, Boston, Massachusetts (Almazan); The Fenway Institute, Fenway Health, Boston, Massachusetts (Keuroghlian); Department of Psychiatry, Massachusetts General Hospital, Boston (Keuroghlian).

Corresponding Author: Anthony N. Almazan, BA, Harvard Medical School, 25 Shattuck St, Boston, MA 02215 (anthony_almazan@hms.harvard.

JAMA Surg. 2021;156(7):611-618. doi:10.1001/jamasurg.2021.0952
Published online April 28, 2021.

Transgender and gender diverse (TGD) people experience a disproportionate burden of mental health problems compared with the general population.[1,2] Prior studies of mental health among TGD people have demonstrated a 41% lifetime prevalence of suicide attempts,[2] 7% to 61% lifetime prevalence of binge drinking,[3] and a 33% prevalence of tobacco use.[4] Increased adverse mental health outcomes among TGD people are likely attributable to stigma, discrimination, pathologization, economic marginalization, violence, and dysphoria associated with an incongruence between gender identity and societal expectations based on one's sex assigned at birth.[5]

According to *Standards of Care* published by the World Professional Association for Transgender Health, gender-affirming surgery is a medically necessary treatment to alleviate psychological distress for many TGD people.[6] The term *gender-affirming surgery* refers to any surgical procedures offered to affirm the gender identities of TGD people. The process of surgical gender affirmation is individually tailored because not all TGD people desire or access these procedures.[7] In the largest survey of the TGD community to our knowledge to date, 25% of respondents reported undergoing some type of gender-affirming surgery.[8]

As a result of professional recommendations, insurance nondiscrimination laws, and expansion of dedicated transgender health practices, demand for gender-affirming surgery is steadily rising.[9] In the United States, incidence of gender-affirming surgeries has increased annually since 2000.[10] Despite growing demand for and access to gender-affirming surgery, there is a paucity of high-quality evidence regarding its effects on mental health outcomes among TGD people.

Existing evidence on the association between gender-affirming surgeries and mental health outcomes is largely derived from small-sample, cross-sectional, and uncontrolled studies.[1,11,12] A seminal 1998 review of the experiences of more than 2000 TGD people from 79 predominantly uncontrolled follow-up studies demonstrated qualitative improvement in psychosocial outcomes following gender-affirming surgery.[11] Attempts since then to empirically demonstrate mental health benefits from gender-affirming surgery have generated mixed results. A meta-analysis of 1833 TGD people across 28 studies concluded that studies offered "low-quality evidence" for positive mental health benefits from surgical gender affirmation.[12] The largest existing study on this subject to our knowledge,[13] a total population study including 2679 people diagnosed as having gender incongruence in Sweden, demonstrated a longitudinal association between gender-affirming surgery and reduced mental health treatment utilization.[13] However, a 2020 published correction of this study[14] demonstrated no mental health benefit from gender-affirming surgery after comparison with a control group of TGD people who had not yet undergone surgery. Mental health effects of gender-affirming surgery thus remain controversial.

Given the increasing incidence of surgical gender affirmation among TGD people, there is a significant need for clarification of the mental health benefits of gender-affirming surgery. In this article, we present the largest study to our knowledge to date on the association between gender-affirming surgeries and mental health outcomes. Using the 2015 US Transgender Survey, the largest existing data set on surgical and mental health experiences of TGD people, we investigate the hypothesis that gender-affirming surgeries are associated with improved mental health outcomes, including psychological distress, substance use, and suicidality.

Key Points

Question Are gender-affirming surgeries associated with better mental health outcomes among transgender and gender diverse (TGD) people?

Findings In this secondary analysis of the 2015 US Transgender Survey (n = 27 715), TGD people with a history of gender-affirming surgery had significantly lower odds of past-month psychological distress, past-year tobacco smoking, and past-year suicidal ideation compared with TGD people with no history of gender-affirming surgery.

Meaning These findings support the provision of gender-affirming surgeries for TGD people who seek them.

Methods

Study Design

In this study, we performed a secondary analysis of the 2015 US Transgender Survey (USTS).[8] This investigation is reported using Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guidelines.

Study Population and Data Source

The 2015 USTS was a cross-sectional, nonprobability sample of responses from 27 715 TGD adults from 50 US states, Washington, DC, US territories, and US military bases abroad. The survey was developed by researchers, advocates, people with lived experience, and subject experts over the course of a year. The final survey contained 324 possible questions with 32 domains addressing subjects including health and health care access. It was disseminated by community-based outreach and administered online from August 19, 2015, to September 21, 2015. The USTS protocol was approved by the University of California, Los Angeles institutional review board.[8] The protocol for the present study was reviewed by the Fenway Institute institutional review board and did not meet criteria for human subjects research. For this reason, consent was not obtained.

Outcomes

Five binary mental health outcomes were examined, including endorsement or denial of the following: (1) past-month severe psychological distress (score on the Kessler Psychological Distress Scale meeting the previously validated threshold of ≥13),[15] (2) past-month binge alcohol use (≥5 alcoholic drinks on one occasion), (3) past-year tobacco smoking, (4) past-year suicidal ideation, and (5) past-year suicide attempt.

Exposure Group

The exposure group included respondents who endorsed a history of gender-affirming surgery, defined as undergoing 1 or

more types of gender-affirming surgery at least 2 years prior to submitting responses to the USTS. Respondents were asked about their experiences with gender-affirming surgeries through the question, "Have you had or do you want any of the health care listed below for gender transition?" Respondents were presented with 1 of 2 lists of gender-affirming surgeries based on their self-reported sex assigned at birth. For each surgery, respondents were able to indicate one of the following answers: "Have had it," "Want it some day," "Not sure if I want this," or "Do not want this." Respondents were included in the exposure group if they answered "Have had it" to 1 or more of the following types of gender-affirming procedures: breast augmentation, orchiectomy, vaginoplasty/labiaplasty, trachea shave, facial feminization surgery, or voice surgery. Respondents were also included in the exposure group if they answered "Have had it" to one or more of the following types of gender-affirming procedures: chest surgery, hysterectomy, clitoral release/metoidioplasty/centurion procedure, or phalloplasty.

In this study, outcomes of interest included mental health symptoms in the year prior to taking the USTS. To ensure that exposure to gender-affirming surgeries temporally preceded all outcomes of interest, respondents were included in the exposure group if they had received their first gender-affirming surgery at least 2 years prior to submitting responses to the USTS. For each respondent with a history of gender-affirming surgery, the number of years since their first surgery was calculated by subtracting age at first surgery from current age.

Control Group

The control group included respondents who desired gender-affirming surgeries but had not yet received any. Respondents were included in this group if they answered "Want it some day" for at least 1 of the aforementioned gender-affirming procedures but did not answer "Have had it" for any of them. We excluded participants who did not report desire for any gender-affirming surgeries.

Covariates

The following sociodemographic covariates were examined: age (18-44 years, 45-64 years, and ≥65 years), education level (less than high school or high school graduate up to associate degree, bachelor degree, or higher), employment status (employed, unemployed, or out of labor force), gender identity (transgender woman, transgender man, nonbinary, or cross-dresser), health insurance status (uninsured or insured), household income (<$25 000, $25 000-$99 999, or ≥$100 000), race (Alaska Native/American Indian, Asian/Pacific Islander, Black/African American, Latinx/Hispanic, other/biracial/multiracial, or White), sex assigned at birth (female or male), and sexual orientation (asexual, lesbian/gay/bisexual, or heterosexual).

Family rejection was included as a covariate and was defined by the USTS as history of any of the following experiences with a family member owing to the respondent's gender identity: ending the relationship, physical violence, being forced out of their home, being prevented from wearing desired gender-concordant clothing, and exposure to gender identity conversion efforts. Lifetime exposures to other types of gender-affirming care were also examined, including gender-affirming counseling, pubertal suppression, and hormone therapy. Given the possibility that any of these covariates could confound the relationship between gender-affirming surgeries and mental health outcomes, all covariates were included in the final multivariable models.

Statistical Analysis

All analyses were conducted using Stata, version 16.1 (StataCorp). Unweighted descriptive statistics for exposure and control groups were calculated and are presented as frequencies and percentages.

Multivariable logistic regression models adjusted for all covariates were generated to examine whether undergoing gender-affirming surgery is associated with each of the examined mental health outcomes.[16,17] To account for the survey's nonprobability sampling, all models incorporated survey weights to correct sampling biases related to age and race/ethnicity. Adjusted odds ratios (aORs), 95% CIs, and 2-sided *P* values are reported.

We performed a post hoc analysis to determine whether associations between gender-affirming surgeries and mental health outcomes differ based on the degree of surgical affirmation. The exposure variable was recoded as 3 categories: those who received all desired surgeries, some desired surgeries, and no desired surgeries. Because the USTS did not collect information on timing of each respondent's last surgery, respondents for this post hoc analysis could not be excluded to ensure that all exposures temporally preceded mental health outcomes. The recoded 3-category exposure variable was substituted into 5 additional multivariable logistic regression models, adjusted for all aforementioned covariates.

Owing to concerns that baseline mental health status may confound associations between gender-affirming surgery and mental health outcomes, we conducted an additional post hoc analysis to determine whether lifetime mental health measures were associated with exposure to gender-affirming surgeries. We did not incorporate these measures into the primary models due to collinearity. Four separate post hoc models, adjusted for all aforementioned covariates, regressed exposure to gender-affirming surgeries against lifetime suicidal ideation, lifetime suicide attempts, lifetime alcohol use, and lifetime smoking.

To account for multiple hypothesis testing, a Bonferroni correction was applied to adjust for 19 total tests. A *P* value of less than .002 was used as the corrected threshold for statistical significance.

Less than 2% of the study sample had missing data for exposure and outcome variables, and less than 9% of the study sample had missing data for any covariates. Given that these are acceptably low levels of missingness,[18] respondents with missing data were excluded without compensatory methods.

Results

Of the 27 715 respondents, 3559 (12.8%) endorsed undergoing 1 or more types of gender-affirming surgery at least 2 years

Table 1. Sample Sociodemographics[a]

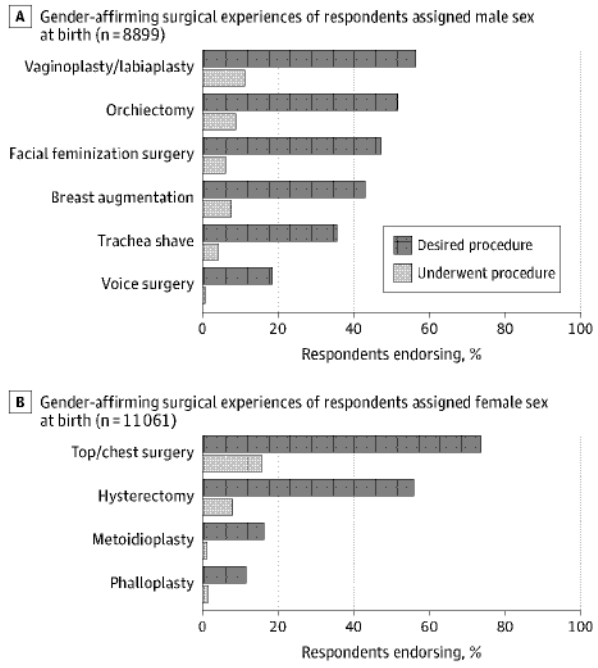
Characteristic	No. (%)		Difference, % (95% CI)
	No history of surgery (n = 16 401)	History of surgery (n = 3559)	
Age, y			
18-44	14 170 (86.4)	2012 (56.5)	29.9 (28.2 to 31.6)
45-64	1922 (11.7)	1261 (35.4)	−23.7 (−25.4 to −22.1)
≥65	309 (1.9)	285 (8.0)	−6.1 (−7.0 to −5.2)
Education			
Less than high school	682 (4.2)	37 (1.0)	3.1 (2.7 to 3.6)
High school graduate up to associate degree	10 918 (66.6)	1243 (34.9)	31.6 (29.9 to 33.3)
Bachelor degree or higher	4801 (29.3)	2279 (64.0)	−34.8 (−36.5 to −33.0)
Employment			
Employed	10 306 (62.8)	2585 (72.6)	−9.8 (−11.4 to −8.2)
Unemployed	2474 (15.1)	202 (5.7)	9.4 (8.5 to 10.3)
Out of labor force	3537 (21.6)	755 (21.2)	0.4 (−1.1 to 1.8)
Family rejection			
Yes	7466 (45.5)	2328 (65.4)	−19.9 (−21.6 to −18.2)
No	7360 (44.9)	1173 (33.0)	11.9 (10.2 to 13.6)
Gender identity			
Transgender woman	6277 (38.3)	1474 (41.4)	−3.1 (−4.9 to −1.4)
Transgender man	4764 (29.1)	1725 (48.5)	−19.4 (−21.2 to −17.6)
Nonbinary	4958 (30.2)	342 (9.6)	20.6 (19.4 to 21.8)
Cross-dresser	402 (2.5)	18 (0.5)	2.0 (1.6 to 2.3)
Health insurance			
Uninsured	2397 (14.6)	304 (8.5)	6.1 (5.0 to 7.1)
Insured	13 959 (85.1)	3253 (91.4)	−6.3 (−7.4 to −5.2)
Household income			
<$25 000	5960 (36.3)	768 (21.6)	14.7 (13.2 to 16.3)
$25 000-$99 999	6829 (41.6)	1804 (50.7)	−9.1 (−10.9 to −7.2)
≥$100 000	2073 (12.6)	840 (23.6)	−11.0 (−12.4 to −9.5)
Race/ethnicity			
Alaska Native/American Indian	206 (1.3)	39 (1.1)	0.2 (−0.2 to 0.5)
Asian/Pacific Islander	436 (2.7)	64 (1.8)	0.9 (0.4 to 1.4)
Black/African American	459 (2.8)	124 (3.5)	−0.7 (−1.3 to −0.03)
Latinx/Hispanic	929 (5.7)	154 (4.3)	1.3 (0.6 to 2.1)
Other/biracial/multiracial	963 (5.9)	200 (5.6)	0.3 (−0.6 to 1.1)
White	13 408 (81.8)	2978 (83.7)	−1.9 (−3.3 to −0.6)
Sex assigned at birth			
Female	9032 (55.1)	2029 (57.0)	−1.9 (−3.7 to −0.1)
Male	7369 (44.9)	1530 (43.0)	1.9 (0.1 to 3.7)
Sexual orientation			
Asexual	2002 (12.2)	228 (6.4)	5.8 (4.9 to 6.7)
Lesbian, gay, bisexual	11 433 (69.7)	2393 (67.2)	2.5 (0.8 to 4.2)
Heterosexual	1729 (10.5)	782 (22.0)	−11.4 (−12.9 to −10.0)
Other gender-affirming care			
Counseling	9016 (55.0)	3099 (87.1)	−32.1 (−33.4 to −30.8)
Pubertal suppression	197 (1.2)	94 (2.6)	−1.4 (−2.0 to −0.9)
Hormone therapy	7104 (43.3)	3213 (90.3)	−47.0 (−48.2 to −45.7)

[a] Column percentages may not add up to 100% because missing data are not displayed.

prior to submitting survey responses, while 16 401 respondents (59.2%) endorsed a desire to undergo 1 or more types of gender-affirming surgery but denied undergoing any of these.

Compared with the control group, the exposure group had higher percentages of respondents who were older, employed, more educated, endorsed family rejection, reported having health insurance, and reported higher household income. Respondents in the exposure group were more likely to endorse a history of gender-affirming counseling, pubertal suppression, and hormone therapy (Table 1).

Figure 1. Desire for and History of Gender-Affirming Surgical Procedures in Study Sample



Includes 2015 US Transgender Survey respondents who indicated they desired and either had or had not undergone at least 1 type of gender-affirming surgery. Respondents were presented with 1 of 2 lists of gender-affirming surgeries based on their self-reported sex assigned at birth.

Figure 2. Comparison of Mental Health Outcomes Among Respondents Who Did and Did Not Undergo Gender-Affirming Surgery

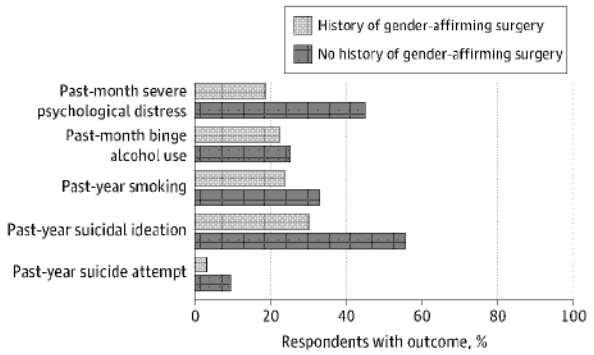


Table 2. Association Between History of Gender-Affirming Surgery and Mental Health Outcomes[a]

Variable	aOR (95% CI)[b]	*P* value
Severe psychological distress (past month)[c]	0.58 (0.50-0.67)	<.001
Substance use		
Binge alcohol use (past month)[d]	0.83 (0.72-0.96)	.01
Smoking (past year)	0.65 (0.57-0.75)	<.001
Suicidality (past year)		
Ideation	0.56 (0.50-0.64)	<.001
Attempt	0.65 (0.47-0.90)	.009

Abbreviation: aOR, adjusted odds ratio.

[a] Adjusted for age, education, employment status, family rejection, gender identity, health insurance, household income, race/ethnicity, sex assigned at birth, sexual orientation, history of gender-affirming counseling, pubertal suppression, and history of gender-affirming hormone therapy.

[b] Reference/control group (n = 16 401) is composed of individuals who desired at least 1 type of gender-affirming surgery but had not received any surgeries. Exposure group (n = 3559) is limited to respondents who had their first surgery at least 2 years prior to submitting survey responses.

[c] Defined as a score of at least 13 on the Kessler Psychological Distress Scale.

[d] Defined as consuming at least 5 alcoholic drinks on the same occasion.

For each surgical procedure, the percentage of people who desired it was higher than the percentage of people who endorsed undergoing it (**Figure 1**). For every adverse mental health outcome, the percentage of respondents who endorsed it was lower in the exposure group than in the control group (**Figure 2**).

After adjustment for sociodemographic factors and exposure to other types of gender-affirming care, undergoing 1 or more types of gender-affirming surgery was associated with lower past-month psychological distress (aOR, 0.58; 95% CI, 0.50-0.67; $P < .001$), past-year smoking (aOR, 0.65; 95% CI, 0.57-0.75; $P < .001$), and past-year suicidal ideation (aOR, 0.56; 95% CI, 0.50-0.64; $P < .001$). After Bonferroni correction, there was no statistically significant association between gender-affirming surgeries and past-month binge alcohol use or past-year suicide attempts (**Table 2**).

In the post hoc analysis stratifying by degree of surgical affirmation, 16 401 respondents were in the reference group who received no desired surgeries. Respondents who had undergone all desired surgeries (n = 2448) had significant reductions in the odds of each adverse mental health outcome, and these reductions were more profound than those among respondents who had received only some desired surgeries (n = 3311) (**Table 3**).

Measures of lifetime mental health were not associated with exposure to gender-affirming surgeries. After adjustment for all aforementioned covariates, undergoing gender-affirming surgery was not associated with lifetime suicidal ideation (aOR, 1.00; 95% CI, 0.85-1.20; $P = .92$), lifetime suicide attempts (aOR, 1.16; 95% CI, 1.01-1.34; $P = .04$), lifetime alcohol use (aOR, 1.00; 95% CI, 0.99-1.01; $P = .96$), or lifetime smoking (aOR, 1.00; 95% CI, 1.00-1.01; $P = .34$).

Discussion

To our knowledge, this is the first large-scale, controlled study to demonstrate an association between gender-affirming surgery and improved mental health outcomes. In this study, we demonstrate that undergoing gender-affirming surgery is associated with decreased odds of past-month severe psychological distress, past-year smoking, and past-year suicidal ideation. The post hoc analysis stratifying by degree of surgical affirmation demonstrates that TGD people who underwent all desired surgeries had significantly lower odds of all adverse mental health outcomes, and these benefits were stronger than

Table 3. Association Between Degree of Surgical Gender Affirmation and Mental Health Outcomes[a]

Variable	Received some desired surgeries (n = 3311)[b]		Received all desired surgeries (n = 2448)[b]	
	aOR (95% CI)	*P* value	aOR (95% CI)	*P* value
Severe psychological distress (past month)[c]	0.70 (0.60-0.81)	<.001	0.47 (0.39-0.56)	<.001
Substance use				
Binge alcohol use (past month)[d]	0.97 (0.84-1.11)	.63	0.75 (0.64-0.87)	<.001
Smoking (past year)	0.75 (0.66-0.86)	<.001	0.58 (0.49-0.68)	<.001
Suicidality (past year)				
Ideation	0.72 (0.63-0.81)	<.001	0.44 (0.38-0.51)	<.001
Attempt	0.70 (0.53-0.93)	.01	0.44 (0.28-0.70)	<.001

Abbreviation: aOR, adjusted odds ratio.

[a] Adjusted for age, education, employment status, family rejection, gender identity, health insurance, household income, race/ethnicity, sex assigned at birth, sexual orientation, history of gender-affirming counseling, pubertal suppression, and history of gender-affirming hormone therapy.

[b] Reference group is individuals who received none of their desired surgeries (n = 16 401).

[c] Defined as a score of at least 13 on the Kessler Psychological Distress Scale.

[d] Defined as consuming at least 5 alcoholic drinks on the same occasion.

among TGD people who only received some desired surgeries.

The observed associations between gender-affirming surgery, psychological distress, and suicide risk reinforce previous small-sample studies suggesting that gender-affirming surgery improves mental health and quality of life among TGD people.[1,12] Our findings also reflect evidence from qualitative studies indicating perceived mental health benefits of gender-affirming surgeries among TGD people.[19-21] In our primary analysis, although gender-affirming surgery was associated with lower odds of past-year suicidal ideation, there was no statistically significant association between gender-affirming surgeries and past-year suicide attempts. However, in a post hoc analysis respondents who underwent all desired gender-affirming surgeries had significantly lower odds of past-year suicide attempts.

The association observed between gender-affirming surgeries and reduction in substance use behaviors is consistent with previous studies involving small community samples that demonstrated associations between gender-affirming medical care and lower odds of high-risk substance use.[22,23] In the primary analysis, undergoing gender-affirming surgery was not significantly associated with past-month binge alcohol use. This may be consistent with evidence that after adjustment for sociodemographic factors, gender minority identity itself does not predict high-risk alcohol use.[24] However, in a post hoc analysis, respondents who underwent all desired gender-affirming surgeries had significantly lower odds of past-month binge alcohol use.

This investigation offers evidence to support the clinical practice of gender-affirming surgery. Guidelines for provision of gender-affirming medical and surgical care have historically been challenged based on a limited evidence base. The American Psychiatric Association has previously concluded that the quality of evidence for treatment of gender dysphoria is low, and consequently, recommendations regarding gender-affirming care have been driven by clinical consensus where empirical evidence is lacking.[25] This study offers new data that substantiate the current clinical consensus by expanding the evidence base in support of gender-affirming surgical care.

The observed mental health benefits of gender-affirming surgeries in this study highlight the importance of policies that facilitate access to surgical gender affirmation. In the present study, the percentages of people who had undergone each gender-affirming surgical procedure were substantially lower than the percentages of people who desired them, suggesting significant barriers to accessing gender-affirming surgeries. State-level prohibitions against insurance exclusions for gender-affirming care have been associated with more extensive coverage of gender-affirming surgical procedures.[26] In light of this study's results, such policies may be of even greater public health interest. US federal policies related to gender-affirming care have included a recent reversal of Affordable Care Act insurance protections for gender affirmation and the continued prohibition of Veterans Affairs funding allocation for gender-affirming surgeries.[27,28] Formulation of evidence-based policies for the financing of gender-affirming surgery will be crucial for advancing the health and well-being of TGD communities.

Strengths and Limitations

This study's strengths include aspects of its design that address prior limitations in the existing literature on this subject. Multiple meta-analyses of studies examining the association between gender-affirming surgeries and mental health outcomes have demonstrated that much of the existing literature consists of evidence derived with small sample sizes, lack of control groups, and lack of adjustment for other kinds of gender-affirming care.[12,29] Our study is responsive to these methodologic concerns.

First, we used the largest existing data set containing information on the surgical and mental health experiences of TGD people. Second, this is, to our knowledge, the first large-scale study on this subject to use the ideal control group to examine associations between gender-affirming surgeries and mental health outcomes: individuals who desire gender-affirming surgery but have not yet received it. Experts have cautioned against using comparison groups that conflate TGD people who did not undergo gender-affirming surgery because they were waiting for it with TGD people not seeking it in the first place. Inability to differentiate these 2 groups likely

contributed to the lack of significant mental health benefit observed in the 2019 large-scale study on this subject.[13,30]

Third, although this survey-based investigation uses a cross-sectional study design, we constructed an exposure group that includes only individuals exposed to their first gender-affirming surgery prior to the window of assessment for any adverse mental health outcomes. Thus, we ensured that our exposure temporally preceded our outcomes, allowing us to better understand the direction of observed associations. These exclusions could not be performed in our post hoc analysis stratifying by degree of surgical affirmation, and that analysis should therefore be interpreted with caution.

Fourth, our data set allowed us to control for previous experiences of gender-affirming counseling, pubertal suppression, and hormone therapy. Consequently, this study is, to our knowledge, the first large-scale investigation to ascertain the mental health benefits of gender-affirming surgeries independent of other common forms of gender-affirming health care.

Our study has several limitations. The nonprobability sampling of the USTS may limit generalizability. All measures are self-reported and may be subject to response bias. Furthermore, the USTS only offers data on experiences with 10 specific types of gender-affirming surgeries and does not capture the full range of procedures that constitute gender-affirming surgery. Lastly, because this is an observational study, it may be subject to unmeasured confounding. Much of the literature on mental health benefits of gender-affirming surgery has been complicated by inability to adjust for a key confounder: baseline mental health status. Our post hoc analysis demonstrates that lifetime suicidality and substance use behaviors are not associated with the exposure variable in this sample. Therefore, prior mental health factors do not appear to confound associations between gender-affirming surgery and subsequent mental health outcomes in our study. There may nevertheless be other types of mental health problems not captured in the USTS that confound these associations. These limitations highlight the need for larger probability-based surveys with TGD communities, more consistent gender identity data collection across health care systems, and more comprehensive baseline health data collection with TGD populations.

Conclusions

In this article, we present the largest study to our knowledge to date on associations between gender-affirming surgeries and mental health outcomes. Our results demonstrate that undergoing gender-affirming surgery is associated with improved past-month severe psychological distress, past-year smoking, and past-year suicidal ideation. Our findings offer empirical evidence to support provision of gender-affirming surgical care for TGD people who seek it. Furthermore, this study provides evidence to support policies that expand and protect access to gender-affirming surgical care for TGD communities.

ARTICLE INFORMATION

Accepted for Publication: February 5, 2021.

Published Online: April 28, 2021. doi:10.1001/jamasurg.2021.0952

Author Contributions: Mr Almazan had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Concept and design: All authors.
Acquisition, analysis, or interpretation of data: All authors.
Drafting of the manuscript: All authors.
Critical revision of the manuscript for important intellectual content: All authors.
Statistical analysis: Almazan.
Obtained funding: Keuroghlian.
Administrative, technical, or material support: Keuroghlian.
Supervision: Keuroghlian.

Conflict of Interest Disclosures: Dr Keuroghlian reported grants from Patient-Centered Outcomes Research Institute Contract AD-2017C1-6569 (PI: Sari L. Reisner) during the conduct of the study; in addition, Dr Keuroghlian stands to receive future royalties as editor of a forthcoming McGraw-Hill Education textbook on transgender and gender diverse care. No other disclosures were reported.

Funding/Support: This work was supported by contract AD-2017C1-6569 from the Patient-Centered Outcomes Research Institute (PI: Dr Sari L. Reisner).

Role of the Funder/Sponsor: The funding source had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Additional Contributions: We thank the National Center for Transgender Ecuality for granting us access to the data from the 2015 US Transgender Survey.

REFERENCES

1. Dhejne C, Van Vlerken R, Heylens G, Arcelus J. Mental health and gender dysphoria: a review of the literature. *Int Rev Psychiatry*. 2016;28(1):44-57. doi:10.3109/09540261.2015.1115753

2. Grant J, Motter L, Tanis J. Injustice at every turn: a report of the National Transgender Discrimination Survey. 2011. Accessed October 11, 2020. https://www.transequality.org/sites/default/files/docs/resources/NTDS_Report.pdf

3. Gilbert PA, Pass LE, Keuroghlian AS, Greenfield TK, Reisner SL. Alcohol research with transgender populations: a systematic review and recommendations to strengthen future studies. *Drug Alcohol Depend*. 2018;186:138-146. doi:10.1016/j.drugalcdep.2018.01.016

4. Wheldon CW, Wiseman KP. Tobacco use among transgender and gender non-conforming adults in the United States. *Tob Use Insights*. 2019;12:X19849419. doi:10.1177/1179173X19849419

5. Reisner SL, Poteat T, Keatley J, et al. Global health burden and needs of transgender populations: a review. *Lancet*. 2016;388(10042):412-436. doi:10.1016/S0140-6736(16)00684-X

6. Coleman E, Bockting W, Botzer M, et al Standards of care for the health of transsexual, transgender, and gender-nonconforming people, version 7. *Int J Transgenderism*. 2012;13(4):165-232. doi:10.1080/15532739.2011.700873

7. Berli JU, Knudson G, Fraser L, et al. What surgeons need to know about gender confirmation surgery when providing care for transgender individuals: a review. *JAMA Surg*. 2017;152(4):394-400. doi:10.1001/jamasurg.2016.5549

8. James S, Herman J, Rankin S, Keisling M, Mottet L, Anafi M. The Report of the 2015 U.S. Transgender Survey. National Center for Transgender Equality; 2016. Accessed April 2, 2020. https://ncvc.dspacedirect.org/handle/20.500.11990/1299

9. Weissler JM, Chang BL, Carney MJ, et al. Gender-affirming surgery in persons with gender dysphoria. *Plast Reconstr Surg*. 2018;141(3):388e-396e. doi:10.1097/PRS.0000000000004123

10. Canner JK, Harfouch O, Kodadek LM, et al. Temporal trends in gender-affirming surgery among transgender patients in the United States. *JAMA Surg*. 2018;153(7):609-616. doi:10.1001/jamasurg.2017.6231

11. Pfafflin F, Junge A. Sex reassignment thirty years of international follow-up studies SRS: a comprehensive review, 1961-1991. *International Journal of Transgenderism*. Published online 1998.

12. Murad MH, Elamin MB, Garcia MZ, et al. Hormonal therapy and sex reassignment: a systematic review and meta-analysis of quality of life and psychosocial outcomes. *Clin Endocrinol (Oxf)*. 2010;72(2):214-231. doi:10.1111/j.1365-2265.2009.03625.x

13. Bränström R, Pachankis JE. Reduction in mental health treatment utilization among transgender individuals after gender-affirming surgeries: a total population study. *Am J Psychiatry*. doi:10.1176/appi.ajp.2019.19010080

14. Correction to Bränström and Pachankis. Correction to Bränström and Pachankis. *Am J Psychiatry*. 2020;177(8):734-734. doi:10.1176/appi.ajp.2020.1778correction

15. Kessler RC, Green JG, Gruber MJ, et al. Screening for serious mental illness in the general population with the K6 screening scale: results from the WHO World Mental Health (WMH) survey initiative. *Int J Methods Psychiatr Res*. 2010;19(S1)(suppl 1):4-22. doi:10.1002/mpr.310

16. Tolles J, Meurer WJ. Logistic regression: relating patient characteristics to outcomes. *JAMA*. 2016;316(5):533-534. doi:10.1001/jama.2016.7653

17. Stoltzfus JC. Logistic regression: a brief primer. *Acad Emerg Med*. 2011;18(10):1099-1104. doi:10.1111/j.1553-2712.2011.01185.x

18. Bennett DA. How can I deal with missing data in my study? *Aust N Z J Public Health*. 2001;25(5):464-469. doi:10.1111/j.1467-842X.2001.tb00294.x

19. Schneiders M. *Values and Preferences of Transgender People: A Qualitative Study*. World Health Organization; 2014:65.

20. Jokić-Begić N, Lauri Korajlija A, Jurin T. Psychosocial adjustment to sex reassignment surgery: a qualitative examination and personal experiences of six transsexual persons in croatia. *ScientificWorldJournal*. 2014;2014:960745. doi:10.1155/2014/960745

21. McGuire JK, Doty JL, Catalpa JM, Ola C. Body image in transgender young people: Findings from a qualitative, community based study. *Body Image*. 2016;18:96-107. doi:10.1016/j.bodyim.2016.06.004

22. Keuroghlian AS, Reisner SL, White JM, Weiss RD. Substance use and treatment of substance use disorders in a community sample of transgender adults. *Drug Alcohol Depend*. 2015;152:139-146. doi:10.1016/j.drugalcdep.2015.04.008

23. Wilson EC, Chen Y-H, Arayasirikul S, Wenzel C, Raymond HF. Connecting the dots: examining transgender women's utilization of transition-related medical care and associations with mental health, substance use, and HIV. *J Urban Health*. 2015;92(1):182-192. doi:10.1007/s11524-014-9921-4

24. Blosnich JR, Lehavot K, Glass JE, Williams EC. Differences in alcohol use and alcohol-related health care among transgender and nontransgender adults: findings from the 2014 behavioral risk factor surveillance system. *J Stud Alcohol Drugs*. 2017;78(6):861-866. doi:10.15288/jsad.2017.78.861

25. Byne W, Bradley SJ, Coleman E, et al; American Psychiatric Association Task Force on Treatment of Gender Identity Disorder. Report of the American Psychiatric Association task force on treatment of gender identity disorder. *Arch Sex Behav*. 2012;41(4):759-796. doi:10.1007/s10508-012-9975-x

26. Almazan AN, Benson TA, Boskey ER, Ganor O. Associations between transgender exclusion prohibitions and insurance coverage of gender-affirming surgery. *LGBT Health*. 2020;7(5):254-263. doi:10.1089/lgbt.2019.0212

27. Arnold JD, Nelson AE, Loubier EM. Trends in insurance coverage for gender-affirming surgeries. *JAMA Surg*. 2018;153(10):972. doi:10.1001/jamasurg.2018.2120

28. Kuzon WM Jr, Sluiter E, Gast KM. Exclusion of medically necessary gender-affirming surgery for America's armed services veterans. *AMA J Ethics*. 2018;20(4):403-413. doi:10.1001/journalofethics.2018.20.4.sect1-1804

29. Nobili A, Glazebrook C, Arcelus J. Quality of life of treatment-seeking transgender adults: a systematic review and meta-analysis. *Rev Endocr Metab Disord*. 2018;19(3):199-220. doi:10.1007/s11154-018-9459-y

30. Bränström R, Pachankis JE. Toward rigorous methodologies for strengthening causal inference in the association between gender-affirming care and transgender individuals' mental health: response to letters. *Am J Psychiatry*. 2020;177(8):769-772. doi:10.1176/appi.ajp.2020.20050599

Invited Commentary

Gender-Affirming Surgeries and Improved Psychosocial Health Outcomes

Andrew A. Marano, MD; Matthew R. Louis, MD; Devin Coon, MD, MSE

There is a growing body of literature supporting the positive outcomes of gender-affirming surgery (GAS) on transgender and gender diverse individuals. Mental health outcomes are among the most vital end points to study, given the fundamental intent of GAS to provide patients with relief from gender dysphoria and improvement of psychosocial distress. Much of the data on this topic come from observational studies that lack either control groups or adequate sample size.[1,2] In this issue of *JAMA Surgery*, Almazan and Keuroghlian[3] contribute an analysis of the US Transgender Survey (USTS), examining the topic of mental health outcomes following GAS.

Related article page 611

This study[3] compared individuals who desired but had not undergone GAS with those who had, finding significantly lower rates of psychosocial distress, smoking, and suicidal ideation in the surgery group. When the analysis was broadened to include lifetime rather than recent symptoms (ie, the temporal association between surgery and symptoms was removed), the association became insignificant. The authors[3] concluded the significant associations were not because of prior mental health status but rather a result of surgical intervention.

We commend the authors[3] on their thorough exploration of the USTS, the largest collection of data on the experience of transgender and gender diverse individuals to our knowledge to date. They provide a controlled, well-powered study, and their findings align with prior studies demonstrating the efficacy of GAS. However, the largest challenge in interpreting this association lies in the mental health screening typically necessary to be a candidate for GAS, which may convolute the specific connection between these 2 variables. The authors have fashioned a surrogate temporal association from cross-sectional data, but it is one that inevitably depends on certain key assumptions to hold true.

The second challenge is the use of USTS survey questions to quantify psychosocial distress, rather than a validated outcome instrument targeted toward psychosocial assessment in the transgender and gender diverse population. This is not as much a critique of the method as an acknowledgment of the scarcity of prospective longitudinal data sets measuring robust outcomes. Prospective cohort-level analyses (rather than population-level analyses) with well-validated outcome instruments are widely recognized as the area requiring greater progress. In the interim, though, this report[3] contributes additional evidence to support the efficacy of GAS in alleviating dysphoria.

The availability of data on this community is a major impediment to addressing its needs and 1 reason the USTS was conducted in the first place, since nearly all governmental surveys continue to omit gender identity as a survey item. This issue has been recognized by numerous key public health

JAMA Pediatrics | Original Investigation

Top Surgery and Chest Dysphoria Among Transmasculine and Nonbinary Adolescents and Young Adults

Mona Ascha, MD; Daniel C. Sasson, BA; Rachita Sood, MD; Jeremy W. Cornelius, MA; Jacob M. Schauer, PhD; Adariane Runge, MS; Abigail L. Muldoon, MA; Noopur Gangopadhyay, MD; Lisa Simons, MD; Diane Chen, PhD; Julia F. Corcoran, MD, MHPE; Sumanas W. Jordan, MD, PhD

IMPORTANCE Transgender and nonbinary (TGNB) adolescents and young adults (AYA) designated female at birth (DFAB) experience chest dysphoria, which is associated with depression and anxiety. Top surgery may be performed to treat chest dysphoria.

OBJECTIVE To determine whether top surgery improves chest dysphoria, gender congruence, and body image in TGNB DFAB AYA.

DESIGN, SETTING, AND PARTICIPANTS This is a nonrandomized prospective cohort study of patients who underwent top surgery between December 2019 and April 2021 and a matched control group who did not receive surgery. Patients completed outcomes measures preoperatively and 3 months postoperatively. This study took place across 3 institutions in a single, large metropolitan city. Patients aged 13 to 24 years who presented for gender-affirming top surgery were recruited into the treatment arm. Patients in the treatment arm were matched with individuals in the control arm based on age and duration of testosterone therapy.

EXPOSURES Patients in the surgical cohort underwent gender-affirming mastectomy; surgical technique was at the discretion of the surgeon.

MAIN OUTCOMES AND MEASURES Patient-reported outcomes were collected at enrollment and 3 months postoperatively or 3 months postbaseline for the control cohort. The primary outcome was the Chest Dysphoria Measure (CDM). Secondary outcomes included the Transgender Congruence Scale (TCS) and Body Image Scale (BIS). Baseline demographic and surgical variables were collected, and descriptive statistics were calculated. Inverse probability of treatment weighting (IPTW) was used to estimate the association of top surgery with outcomes. Probability of treatment was estimated using gradient-boosted machines with the following covariates: baseline outcome score, age, gender identity, race, ethnicity, insurance type, body mass index, testosterone use duration, chest binding, and parental support.

RESULTS Overall, 81 patients were enrolled (mean [SD] age, 18.6 [2.7] years); 11 were lost to follow-up. Thirty-six surgical patients and 34 matched control patients completed the outcomes measures. Weighted absolute standardized mean differences were acceptable between groups with respect to body mass index, but were not comparable with respect to the remaining demographic variables baseline outcome measures. Surgical complications were minimal. IPTW analyses suggest an association between surgery and substantial improvements in CDM (−25.58 points; 95% CI, −29.18 to −21.98), TCS (7.78 points; 95% CI, 6.06-9.50), and BIS (−7.20 points; 95% CI, −11.68 to −2.72) scores.

CONCLUSIONS AND RELEVANCE Top surgery in TGNB DFAB AYA is associated with low complication rates. Top surgery is associated with improved chest dysphoria, gender congruence, and body image satisfaction in this age group.

Editorial page 1071

Supplemental content

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Sumanas W. Jordan, MD, PhD, Division of Plastic and Reconstructive Surgery, Northwestern University Feinberg School of Medicine, 675 North St Clair St, Galter Sta 19-250, Chicago, IL 60611 (sum

JAMA Pediatr. 2022;176(11):1115-1122. doi:10.1001/jamapediatrics.2022.3424
Published online September 26, 2022.

Recent studies estimate that up to 9% of adolescents and young adults (AYA) identify as transgender or nonbinary (TGNB).[1] Many TGNB youth designated female at birth (DFAB) experience chest dysphoria, defined as distress related to the development of breasts.[2] Chest dysphoria can lead to negative physical and emotional consequences, such as lack of participation in exercise and sports, chest-binding practices that can be physically harmful when not performed safely, psychosocial distress, functional limitations, and suicidal ideation.[2-4] A retrospective review by our group of 156 TGNB DFAB AYA aged 12 to 18 years found that chest dysphoria was associated with greater anxiety and depression independent of gender dysphoria, degree of appearance congruence, and social transition status.[5] TGNB youth may perform chest binding to relieve gender dysphoria, improve social acceptance, and reduce misgendering.[3] However, chest binding may cause skin irritation, musculoskeletal pain, and shortness of breath.[6] Prior research has shown that most TGNB youth who perform chest binding desire definitive top surgery.[3]

Top surgery (ie, mastectomy or reduction) is a common gender-affirming surgery sought by TGNB DFAB individuals and has been shown to improve chest dysphoria and quality of life in adults.[7-9] Top surgery among TGNB DFAB patients increased 15% in 2020 compared with 2019.[10] While outcomes of top surgery in adults are well reported, outcomes in youth are not well described.[11-13] A qualitative study of 30 TGNB DFAB AYA reported that patients who underwent top surgery experienced near or total resolution of chest dysphoria, lack of surgical regret, and improved quality of life and functioning.[4] The Chest Dysphoria Measure (CDM) was recently introduced to assess chest dysphoria in TGNB patients, and youth who had not received top surgery had CDM scores nearly 10 times higher than those who did receive top surgery.[2] Despite these promising findings, these studies are limited by their retrospective design. There is a need for robust prospective studies in TGNB youth to guide evidence-based practices, especially with increasing proposed legislation to criminalize gender-affirming health care for minors.

The present study evaluates the association of top surgery with chest dysphoria, gender congruence, and body image in TGNB DFAB AYA. We hypothesize that top surgery will significantly improve chest dysphoria compared with patients who do not undergo top surgery. To our knowledge, this is the first and only prospective matched cohort study evaluating top surgery in youth to date.

Key Points

Question Does gender-affirming top surgery improve chest dysphoria, gender congruence, and body image in transmasculine and nonbinary adolescents and young adults?

Findings This nonrandomized, multicenter, prospective, control-matched study showed that top surgery was associated with statistically significant improvement in chest dysphoria, gender congruence, and body image at 3 months postsurgery. Surgical complications were minimal.

Meaning This study suggests that gender-affirming top surgery is associated with improved chest dysphoria, gender congruence, and body image in this age group.

Methods

Study Sample and Design

This is a multicenter, prospective, matched cohort study, in which TGNB DFAB AYA were recruited between December 2019 and April 2021 to 1 of 2 study arms: patients undergoing surgical mastectomy (treatment group) or patients not undergoing surgical mastectomy (control group). There was no attempt to randomize patients. Patients in the treatment group were recruited from Northwestern Memorial Hospital, The University of Illinois at Chicago, or Ann & Robert H. Lurie Children's Hospital of Chicago at their preoperative consults for top surgery. Inclusion criteria for treatment patients were 13 to 24 years of age at the time of surgery, transmasculine or nonbinary gender identity, DFAB, and English speaking. Gender-affirming hormone therapy was not a requirement in order to recruit patients who may not be interested in hormones but were interested in top surgery.[14] Control patients were recruited from Ann & Robert H. Lurie Children's Hospital of Chicago's Gender and Sex Development Program. Inclusion criteria for control patients were 13 to 24 years of age, presenting for gender-affirming care, transmasculine or nonbinary gender identity, DFAB, and English speaking.

Multicenter institutional review board approval was obtained. Oral and written informed consent were obtained from patients older than 18 years, and from parents when patients were younger than 18 years. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guidelines were followed.

Intervention and Matching

All patients in the treatment arm were assessed for surgical readiness according to World Professional Association for Transgender Health guidelines Standards of Care version 7.[15] Health care professionals tailored surgical techniques to each individual's presentation and preferences. Preoperative and postoperative care protocols varied by institution, and no attempt to standardize surgical technique or protocol was made. Patients in the control group did not receive surgical intervention for chest dysphoria during the study period. Control patients who underwent top surgery during the study period were removed from the study. Two control patients were matched to each treatment patient; the 2:1 matching rate was chosen to account for expected control patient attrition, with a goal of a final 1:1 matching rate for analysis. Patients were matched by age (±1 year) and duration of testosterone therapy (±1 year) at the time of surgery.

Study Variables and Baseline Measures

Demographic variables, date of testosterone therapy initiation, and surgical details were collected via patient medical record review. Race and ethnicity information was obtained from the patient medical record; categories for these variables are provided by the electronic medical record. Age, gender

identity, sex designated at birth, and chest binding practices at initial presentation were collected with a survey distributed through Research Electronic Data Capture hosted at Northwestern University.

Parental Support of Gender Variance—Child/Youth

The Parental Support of Gender Variance (PSGV-CY) is a 21-item questionnaire assessing the respondent's perception of their relationship with their parents and perceived support for their gender identity.[16] Items are scored on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree), with higher scores indicating greater parental support. The PSGV-CY score was only measured at baseline assessment.

Patient-Reported Outcomes

Patient-reported outcomes were obtained via Research Electronic Data Capture preoperatively as a baseline and 3 months postoperatively for surgery patients or postbaseline for control patients. The 3-month time frame was chosen as a period beyond immediate postsurgical pain and euphoria, while remaining in proximity to the study intervention.

Chest Dysphoria

The primary outcome measure of this study is the CDM. The CDM is a 17-item questionnaire that measures aspects of chest dysphoria.[2] It is the only validated instrument published to date that measures chest dysphoria in TGNB AYA patients. Questions assess physical functioning, hygiene, exercise, intimate partnerships, and disruption of plans. Items are rated on a 4-point scale (0 = never, 3 = all the time) and summed, with higher scores indicating greater chest dysphoria (range, 0-51). This measure has high internal consistency (Cronbach α = 0.79 for postsurgical patients, and α = 0.89 for nonsurgical patients).[2]

Gender Congruence

The Transgender Congruence Scale (TCS) is a 12-item questionnaire that evaluates congruence between gender identity and life experience. The TCS comprises 2 subscales, appearance congruence (TCS-AC) and gender identity acceptance (TCS-GIA). The 9-item TCS-AC measures the degree to which respondents feel their external appearance represents their gender identity. The 3-item TCS-GIA measures the degree to which patients accept their self-identified gender identity rather than the gender identity designated by society. Items are rated on a 5-point scale (1 = strongly disagree, 5 = strongly agree) and summed, with higher scores representing favorable gender congruence. The TCS has strong internal consistency (Cronbach α = 0.94 for TCS-AC, α = 0.77 for TCS-GIA, and α = 0.92 for TCS).[17]

Body Image

The Body Image Scale (BIS) is a 30-item questionnaire that assesses satisfaction with various body parts, rated on a 5-point scale (1 = very satisfied, 5 = very dissatisfied) and summed.[18,19] The BIS includes primary sex characteristics, secondary sex characteristics, and neutral (nonsex-related) body parts.[18,20] The BIS includes subscales based on 6 primary, 14 secondary, and 10 neutral characteristics. The BIS has moderate internal consistency (Cronbach α = 0.65 for primary characteristics, α = 0.84 for secondary characteristics, and α = 0.81 for neutral characteristics).[20]

Data Collection and Analysis

Primary (CDM) and secondary (TCS, BIS) outcomes were analyzed with inverse probability treatment weighting (IPTW) propensity score analysis with linear regression adjustments. As robustness checks, we also estimated association of top surgery with outcomes via nearest-neighbor and poststratification matching, linear regression analysis with covariate adjustment, sensitivity analyses, and analysis of gain scores (secondary analyses).

The IPTW propensity score analysis was performed to improve balance of covariates between surgery groups and control groups on baseline measures.[21] The propensity score analysis first involved estimating the probability of receiving surgery using a gradient boosted machine as a function of the following covariates: race, ethnicity, baseline outcome scores for outcomes of interest, age, insurance type, body mass index, testosterone use duration, chest binding, and parental support of gender variance score. Probability of treatment was computed based on propensity scores, and IPTW analyses used weights inversely proportional to the probability of treatment. Balance was checked among covariates with unweighted absolute standardized mean differences (SMDs) and weighted absolute SMDs after matching. To minimize the association of remaining imbalance between the surgery and control groups after weighting, our final IPTW analysis used (weighted) linear adjustments for covariates that did not exhibit satisfactory balance after weighting. Prior to IPTW analysis, we excluded control patients with low chest dysphoria (CDM score <19), high transgender congruence (TCS score >50), and low body image dissatisfaction (BIS score <50) to better ensure that the surgery and control groups were comparable. The final cohort for IPTW analysis included 25 control patients and 35 surgical patients. Methods describing the secondary analyses are available in the online supplement (eMethods in the Supplement).

Subgroup analysis was conducted for patients younger than 18 years. This was done because the age of 18 is often used as an arbitrary cutoff for insurance companies for reimbursing care. The same covariates were used in these models, except for age. IPTW propensity score analysis was recalculated for the subgroup and outliers were removed in a similar fashion. The final cohort for IPTW subgroup analysis included 9 control patients and 15 surgical patients. All analyses were performed using R (The R Project) version 4.0.1 and RStudio (The R Project) version 1.3.959.[22]

Results

Table 1 compares baseline characteristics between surgery and control groups (mean [SD] age, 18.6 [2.7] years), with unweighted absolute SMDs and weighted absolute SMDs after propensity score matching. There were 36 patients in

Table 1. Demographic Characteristics

Variable	No. (%)			Absolute standardized mean difference	
	Overall (n = 70)	Control (n = 34)	Surgical (n = 36)	Unweighted[a]	Weighted[b]
Age, mean (SD), y	18.6 (2.7)	18.3 (2.3)	18.8 (3.1)	0.092	0.232
Gender identity[c]					
Transgender man	59 (84)	27 (79)	32 (89)	0.134	0.133
Nonbinary/genderqueer	9 (13)	6 (18)	3 (8)	0.134	0.133
Other	2 (3)	1 (3)	1 (3)	NA	NA
BMI, mean (SD)	26.0 (6.7)	26.8 (6.8)	25.3 (6.7)	0.307	0.089
Race[d]					
African American	4 (6)	0 (0)	4 (11)	0.343	0.423
Asian	2 (3)	1 (3)	1 (3)	NA	NA
Unknown	2 (3)	1 (3)	1 (3)	NA	NA
White	62 (89)	32 (94)	30 (83)	0.343	0.423
Ethnicity[d]					
Hispanic or Latinx	13 (19)	8 (24)	5 (14)	0.343	0.175
Not Hispanic or Latinx	56 (80)	25 (74)	31 (86)	0.343	0.175
Unknown	1 (1)	1 (3)	0 (0)		
Smoking status[c]					
Current smoker	1 (1)	1 (3)	0 (0)	0.269	0.278
Former smoker	7 (10)	1 (3)	6 (17)	NA	NA
Never smoker	60 (86)	31 (91)	29 (81)	0.269	0.278
Unknown	2 (3)	1 (3)	1 (3)	NA	NA
Performs chest binding	26 (37)	9 (27)	17 (47)	0.278	0.227
Comorbidities					
ADHD	10 (14)	8 (24)	2 (6)	0.445	0.436
Anxiety	21 (30)	10 (30)	11 (31)	0.165	0.166
Asthma	7 (10)	2 (6)	5 (14)	0.114	0.116
Depression	27 (39)	15 (44)	12 (33)	0.200	0.200
Diabetes	2 (3)	1 (3)	1 (3)	0.064	0.063
Using testosterone[c]	67 (96)	32 (94)	35 (97)	0.236	0.228
Testosterone use in months, mean (SD)	20.1 (9.3)	22.4 (7.7)	18.1 (10.3)	0.371	0.236
PSGV total score, mean (SD)	78.9 (17.0)	76.3 (18.3)	81.23 (17.1)	0.405	0.251
Insurance type					
Commercial/managed care	51 (73)	23 (68)	28 (78)	0.444	0.318
Medicaid	16 (23)	11 (32)	5 (14)	0.444	0.318
Self-pay	3 (4)	0 (0)	3 (8)		
Days between surveys, mean (SD)[c]	94.7 (40.1)	97.9 (45.8)	91.7 (34.2)	0.226	0.218
Baseline, mean (SD)					
CDM	28.8 (9.4)	25.7 (10.6)	31.8 (6.9)	0.202	0.314
TCS	37.9 (8.8)	40.7 (9.4)	35.3 (7.5)	0.284	0.257
BIS	87.2 (17.0)	88.3 (19.4)	86.2 (14.6)	0.459	0.387

Abbreviations: ADHD, attention-deficit/hyperactivity disorder; BMI, body mass index; BIS, Body Image Scale; CDM, Chest Dysphoria Measure; NA, not applicable; PSGV, Parental Support of Gender Variance; TCS, Transgender Congruence Scale.

[a] Unweighted standardized mean differences were calculated for the entire cohort.

[b] Weighted standardized mean differences were calculated after excluding patients with CDM scores less than 19, TCS scores more than 50, and BIS scores less than 50.

[c] Indicates covariate was not used in propensity score matching.

[d] Race and ethnicity were self-identified.

the surgery group and 34 patients in the control group. The planned 2:1 matching rate resulted in the anticipated 1:1 final matched rate; recruitment of control patients was difficult given the limited number of available patients and lack of testosterone-matched and age-matched controls for some surgical patients. Eleven patients were lost to follow-up (6 surgical patients and 5 control patients). Drop-out patients had significantly greater median (IQR) baseline CDM (52.0 [46.0-53.0] vs 29.0 [23.0-35.8]) and BIS (103.0 [91.5-109] vs 87.0 [75.3-96.8]) compared with retained patients. Reasons for dropping out of the study are unknown. After propensity score matching, most absolute SMDs improved with weighting, but did not achieve desired balance of less than 0.1 for all covariates.

Average age was comparable in the treatment and control groups. Most patients in both groups identified as transmasculine, were White, were not Hispanic/Latinx, and were never smokers (Table 1). Testosterone use in months was comparable in both groups; eFigure 1 in the Supplement illustrates testosterone use vs age for the surgical and control groups. Two

Table 2. Inverse Probability Treatment Weighting Propensity Score Model Results for Entire Cohort[a]

Outcome	Unweighted mean (SD)		Weighted PS model (surgery estimate [95% CI])
	Control (n = 25)	Surgery (n = 35)	
Primary outcome			
3-mo CDM score	30.50 (8.37)	3.80 (3.47)	-25.58 (-29.18 to -21.98)
Secondary outcomes			
3 month			
TCS	36.90 (4.25)	44.40 (3.59)	7.78 (6.06-9.50)
TCS AC	26.80 (4.64)	33.50 (3.23)	7.11 (5.44-8.79)
BIS	91.90 (15.80)	77.00 (13.00)	-7.20 (-11.68 to -2.72)
BIS secondary	38.50 (8.08)	34.10 (7.37)	-0.11 (-2.27 to 2.05)

Abbreviations: AC, appearance congruence; BIS, Body Image Scale; CDM, Chest Dysphoria Measure; PS, Propensity Score; TCS, Transgender Congruence Scale.

[a] Summary statistics and models were calculated after excluding patients with CDM scores less than 19, TCS scores more than 50, and BIS scores less than 50.

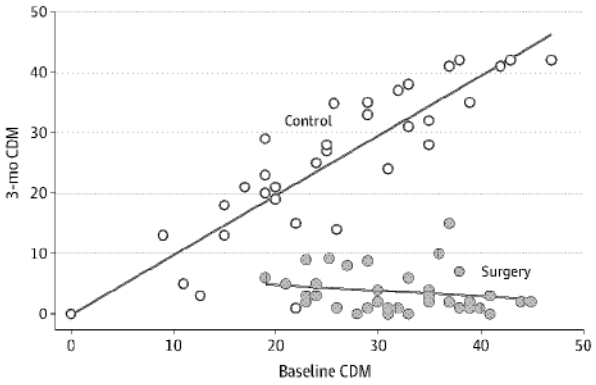
patients in the control group and 1 patient in the surgical group were not taking testosterone. Seventeen surgical patients reported performing chest binding, while 9 control patients reported binding. Complete case analysis was performed comparing demographic variables between complete and incomplete patient medical records. There were no statistically significant findings between complete and incomplete cases.

The most common surgical technique was double incision mastectomy with nipple areolar complex grafting (n = 24 [67%]), followed by nipple-sparing mastectomy (n = 10 [28%]), and periareolar mastectomy (n = 2 [6%]). The resection weight was calculated as the average value of both breasts. Median (IQR) resection weight in grams was 329.5 (111.9-552.8). Seven patients (19%) received concomitant liposuction and 29 patients (81%) received drains. Median (IQR) drain duration was 8.3 (7.0-9.0) days. Complications were minimal, with 1 hematoma (3%), 2 seromas (6%), and 1 instance of nipple loss (3%). There were no incidences of infection or delayed wound healing.

Chest Dysphoria

An exploratory data analysis was performed. Covariates including age, body mass index, and testosterone use in months appeared normally distributed. **Table 2** outlines IPTW analysis results for the primary and secondary outcomes of interest. Summary statistics for 3-month outcomes scores in each group are also reported. eFigure 2 in the Supplement illustrates boxplot distribution of propensity scores for each group. eFigure 3 in the Supplement illustrates mean and standard deviation of propensity scores across quintiles for each group. The IPTW model estimated a 25.58 (95% CI, -29.18 to -21.98) point decrease in CDM for the surgery group relative to the control group. The **Figure** illustrates 3-month CDM as a function of baseline CDM. Three-month CDM scores for all surgery patients were low, regardless of baseline CDM score. The mean (SD) difference between baseline and 3-month scores in the treatment group was -28.12 (8.32). In contrast, among control patients, CDM scores did not appear to change significantly at 3 months. The mean (SD) difference between baseline and 3-month scores in the control group was -0.52 (6.29). **Table 3** outlines subgroup analysis of outcomes among patients under 18 years. The IPTW model estimated a 25.48 (95% CI, -32.85 to -18.11) point decrease in CDM for the surgery group relative to the control group.

Figure. Baseline vs 3-Month Chest Dysphoria Measure (CDM) Scores Among Surgery and Control Patients



The x-axis represents the baseline or preoperative CDM score and the y-axis represents the 3-month CDM score. A higher CDM score indicates worsening chest dysphoria. Three-month CDM scores for all surgery patients were low, regardless of baseline CDM score. Surgery patients with high baseline CDM scores experienced greater decreases in CDM scores after top surgery, as indicated by the cluster of points at the bottom right of the graph. Among control patients, there appears to be a linear trend in baseline CDM scores vs 3-month CDM scores, indicating that CDM scores remained unchanged or increased at 3 months.

Gender Congruence

The IPTW model estimated a 7.78 (95% CI, 6.06-9.50) point increase in TCS for the surgery group relative to the control group (Table 2). Results for the TCS-AC subscore were similar (Table 2).

Table 3 reports results for the cohort of individuals younger than 18 years. The IPTW model estimated a 5.36 (95% CI, 0.47-10.24) point increase in TCS for the surgery group relative to the control group. Results for the TCS-AC subscore were similar (Table 3).

Body Image Dissatisfaction

The IPTW model estimated a 7.20 (95% CI, -11.68 to -2.72) point decrease in BIS for the surgery group relative to the control

Table 3. Inverse Probability Treatment Weighting Propensity Score Model Results for the Younger Than 18 Years Cohort

Outcome	Mean (SD)		Weighted PS model (surgery estimate [95% CI])
	Control (n = 9)	Surgery (n = 15)	
Primary outcome			
3-mo CDM score	32.80 (8.89)	3.47 (3.14)	-25.48 (-32.85 to -18.11)
Secondary outcomes			
3 month			
TCS	38.10 (3.06)	45.30 (3.92)	5.36 (0.47-10.24)
TCS AC	27.30 (2.4)	34.40 (3.83)	5.41 (0.83-10.00)
BIS	97.30 (14.20)	79.90 (12.40)	-7.83 (-13.98 to -1.67)
BIS secondary	40.00 (7.30)	35.20 (6.58)	0.99 (-2.52 to 4.51)

Abbreviations: AC, appearance congruence; BIS, Body Image Scale; CDM, Chest Dysphoria Measure; PS, propensity score; TCS, Transgender Congruence Scale.

group (Table 2), indicating reduced body dissatisfaction. Results for the BIS secondary subscore did not demonstrate a significant association between surgery and score improvement (Table 2).

Table 3 reports results for the younger than 18 years cohort. The IPTW model estimated a 7.83 (95% CI, -13.98 to -1.67) point decrease in BIS for the surgery group relative to the control group. Results for the BIS secondary subscore did not demonstrate a significant association between surgery and score improvement (Table 3).

Secondary Analyses

Results for the secondary analyses are available in the eResults in the Supplement. eTables 1 through 3 in the Supplement present these results, which are similar to those of the IPTW analysis.

Discussion

To our knowledge, we present the first and largest prospective matched study evaluating quantitative outcomes of top surgery among TGNB DFAB AYA. We found that top surgery was associated with significant improvement in chest dysphoria (measured by the CDM) 3 months postoperatively; patients receiving surgery exhibited substantial decreases in CDM from presurgery to postsurgery and their reductions in CDM were significantly greater than those for patients who did not undergo top surgery. Top surgery also led to significant improvements in gender congruence (TCS) and body image satisfaction (BIS) at 3 months postoperatively. Surgical complications were minimal and comparable with those in adult patients.[23-28]

Chest dysphoria can be treated with top surgery, and patients with minimal or no chest dysphoria may not require or desire top surgery. We removed participants with baseline CDM scores of less than 19 from the control group prior to conducting our primary analysis; this was the lowest CDM score in the surgery group. We acknowledge that several control patients with low CDM scores have not sought top surgery because of minimal chest dysphoria. Surgeons should consider treating patients with a diagnosis of gender dysphoria and who experience the negative mental and physical health effects of chest dysphoria.

In our practice, there is no predetermined timeline for gender-affirming medical or surgical treatment; patients are assessed individually by a multidisciplinary team for readiness. There is no evidence to support delaying surgery for eligible patients based on age. With all other covariates remaining equal, patients' baseline outcome score was significantly associated with improvement in the 3-month outcome score in the IPTW models. For all outcome measures, the effect size of improvement was less than 1 point, and the clinical significance of this finding is minimal. This does not support the notion that dysphoria will improve over time without intervention. Treatment patients demonstrated significant improvement and a greater effect size in all measures after surgery compared to control patients.

TGNB youth frequently encounter systemic barriers to care. Obtaining insurance coverage for top surgery is a common problem; while World Professional Association for Transgender Health Standards of Care version 7 do not recommend a specific age for top surgery, many insurance plans deny coverage for patients younger than 18 years.[29] Medicolegal barriers exist as many states have ratified legislation prohibiting, and even criminalizing, medical and surgical therapy in TGNB minors.[30,31] This can lead to delays in care and health care avoidance, which result in worsening mental health burden.[32] The American Academy of Pediatrics and the Pediatric Endocrine Society have issued statements in support of providing gender-affirming medical and surgical care to TGNB minors, contending that it is effective, safe, the standard of care, and should be covered by insurers.[33-35] In this study, patients younger than 18 years demonstrated parallel trends of improved CDM, TCS, and BIS compared with the entire cohort, demonstrating the robustness of our findings in this age subgroup. Findings from this study can help dispel misconceptions that gender-affirming treatment is experimental and support evidence-based practices of top surgery. Our retrospective review demonstrated associations between chest dysphoria and anxiety and depression; top surgery can improve chest dysphoria, thereby leading to improvements in quality of life.[5] Our results also corroborate studies that gender-affirming therapy improves mental health and quality of life among TGNB youth.

Limitations

Our study is not without limitations. Analyses omit 11 patients whose outcomes were not measured due to attrition. Furthermore, we acknowledge that patients in the treatment group who were able to access surgery may have greater socioeconomic status and parental support, possibly introducing sampling bias. Patients in the control group may have desired top surgery but were unable obtain it due to both medical and nonmedical barriers; this could not be controlled during analysis. We were unable to achieve a high degree of balance on baseline measures between surgery and control groups, even after propensity score adjustments. Weighted SMDs exceeded 0.1 for many covariates, rendering analyses susceptible to confounding. As a sensitivity analysis, we computed E-values in the Supplement to quantify the potential effect of confounding; reported E-values suggest an extremely high degree of confounding would need to be present to negate estimated treatment effects. We acknowledge that there may be differences between groups and intrinsic differences in the gender experience that are unmeasurable. In the future, control patients should be better matched to treatment patients, with matching using several demographic variables and baseline outcome measures. This will permit estimation of a causal effect on subsequent analyses. Some insurance plans require testosterone therapy for 1 year prior to surgery; this may negatively affect nonbinary patients who do not undergo testosterone therapy and may have limited inclusion of this patient population in our cohort.[2] Nonbinary patients continue to be underrepresented in the literature; subgroup analysis of nonbinary patients was not possible due to small sample size and future studies should focus on this population. One-year follow up data are currently being collected to determine the long-term effect of top surgery on chest dysphoria, gender congruence, and body image satisfaction.

Conclusions

To our knowledge, we present the first prospective study evaluating the association of top surgery with chest dysphoria and gender congruence in TGNB DFAB AYA. Top surgery in this age group is associated with improved chest dysphoria, gender congruence, and body image satisfaction.

ARTICLE INFORMATION

Accepted for Publication: June 8, 2022.

Published Online: September 26, 2022. doi:10.1001/jamapediatrics.2022.3424

Author Affiliations: Division of Plastic and Reconstructive Surgery, Northwestern University Feinberg School of Medicine, Chicago, Illinois (Ascha, Sasson, Sood, Jordan); Department of Obstetrics & Gynecology, Northwestern University Feinberg School of Medicine, Chicago, Illinois (Cornelius); Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois (Schauer); Potocsnak Family Division of Adolescent and Young Adult Medicine, Ann and Robert H. Lurie Children's Hospital, Chicago, Illinois (Runge, Muldoon, Simons, Chen); Pritzker Department of Psychiatry and Behavioral Health, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, Illinois (Gangopadhyay, Chen); Department of Pediatrics, Northwestern University Feinberg School of Medicine, Chicago, Illinois (Simons, Chen); Department of Psychiatry and Behavioral Sciences, Northwestern University Feinberg School of Medicine, Chicago, Illinois (Chen); Division of Plastic, Reconstructive and Cosmetic Surgery, University of Illinois at Chicago, Chicago (Corcoran).

Author Contributions: Drs Ascha and Jordan had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.
Concept and design: Ascha, Sasson, Sood, Runge, Simons, Chen, Corcoran, Jordan.
Acquisition, analysis, or interpretation of data: Ascha, Sasson, Sood, Cornelius, Schauer, Runge, Muldoon, Gangopadhyay, Chen, Corcoran, Jordan.
Drafting of the manuscript: Ascha, Sasson, Cornelius, Schauer, Runge, Muldoon, Simons, Jordan.
Critical revision of the manuscript for important intellectual content: Ascha, Sasson, Sood, Muldoon, Gangopadhyay, Simons, Chen, Corcoran, Jordan.
Statistical analysis: Ascha, Sasson, Schauer, Jordan.
Obtained funding: Sood, Jordan.
Administrative, technical, or material support: Ascha, Sasson, Sood, Cornelius, Runge, Muldoon, Gangopadhyay, Chen, Jordan.
Supervision: Ascha, Gangopadhyay, Chen, Jordan.
Other - Data collection: Runge.

Conflict of Interest Disclosures: Dr Simons reported consultant fees from CVS Caremark for reviewing medication protocols for gender dysphoria, outside the submitted work. Dr Chen reported grants from the National Institutes of Health (R01 HD097122, R01 MH123746, and R01 HD082554), outside the submitted work. Dr Jordan reported grants from the Plastic Surgery Foundation during the conduct of the study. No other disclosures were reported.

Funding/Support: Northwestern University Clinical and Translational Sciences Institute is funded in part by a Clinical and Translational Science Award grant from the National Institutes of Health (UL1TR001422) that contributed to collection, management, analysis, and interpretation of the data. This study was also funded by the Plastic Surgery Foundation and American Association of Pediatric Plastic Surgery Combined Research grant that contributed to collection, management, analysis, and interpretation of the data.

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

REFERENCES

1. Kidd KM, Sequeira GM, Douglas C, et al. Prevalence of gender-diverse youth in an urban school district. *Pediatrics*. 2021;147(6):e2020049823. doi:10.1542/peds.2020-049823

2. Olson-Kennedy J, Warus J, Okonta V, Belzer M, Clark LF. Chest reconstruction and chest dysphoria in transmasculine minors and young adults: comparisons of nonsurgical and postsurgical cohorts. *JAMA Pediatr*. 2018;172(5):431-436. doi:10.1001/jamapediatrics.2017.5440

3. Julian JM, Salvetti B, Held JI, Murray PM, Lara-Rojas L, Olson-Kennedy J. The impact of chest binding in transgender and gender diverse youth and young adults. *J Adolesc Health*. 2021;68(6):1129-1134. doi:10.1016/j.jadohealth.2020.09.029

4. Mehringer JE, Harrison JB, Quain KM, Shea JA, Hawkins LA, Dowshen NL. Experience of chest dysphoria and masculinizing chest surgery in transmasculine youth. *Pediatrics*. 2021;147(3):e2020013300. doi:10.1542/peds.2020-013300

5. Sood R, Chen D, Muldoon AL, et al. Association of chest dysphoria with anxiety and depression in transmasculine and nonbinary adolescents seeking gender-affirming care. *J Adolesc Health*. 2021;68(6):1135-1141. doi:10.1016/j.jadohealth.2021.02.024

6. Peitzmeier SM, Silberholz J, Gardner IH, Weinand J, Acevedo K. Time to first onset of chest binding-related symptoms in transgender youth. *Pediatrics*. 2021;147(3):e20200728. doi:10.1542/peds.2020-0728

7. Rachlin K, Green J, Lombardi E. Utilization of health care among female-to-male transgender individuals in the United States. *J Homosex*. 2008;54(3):243-258. doi:10.1080/00918360801982124

8. van de Grift TC, Elaut E, Cerwenka SC, Cohen-Kettenis PT, Kreukels BPC. Surgical satisfaction, quality of life, and their association after gender-affirming surgery: a follow-up study. *J Sex Marital Ther*. 2018;44(2):138-148. doi:10.1080/0092623X.2017.1326190

9. Poudrier G, Nolan IT, Cook TE, et al. Assessing quality of life and patient-reported satisfaction with masculinizing top surgery: a mixed-methods descriptive survey study. *Plast Reconstr Surg*. 2019;

143(1):272-279. doi:10.1097/PRS.0000000000005113

10. American Society of Plastic Surgeons. 2020 Plastic Surgery Statistics Report. Accessed August 12, 2021. https://www.plasticsurgery.org/documents/News/Statistics/2020/plastic-surgery-statistics-full-report-2020.pdf

11. Skordis N, Butler G, de Vries MC, Main K, Hannema SE. ESPE and PES international survey of centers and clinicians delivering specialist care for children and adolescents with gender dysphoria. *Horm Res Paediatr*. 2018;90(5):326-331. doi:10.1159/000496115

12. Lopez CM, Solomon D, Boulware SD, Christison-Lagay ER. Trends in the use of puberty blockers among transgender children in the United States. *J Pediatr Endocrinol Metab*. 2018;31(6):665-670. doi:10.1515/jpem-2018-0048

13. Delahunt JW, Denison HJ, Sim DA, Bullock JJ, Krebs JD. Increasing rates of people identifying as transgender presenting to Endocrine Services in the Wellington region. *N Z Med J*. 2018;131(1468):33-42.

14. Cocchetti C, Ristori J, Romani A, Maggi M, Fisher AD. Hormonal Treatment Strategies Tailored to Non-Binary Transgender Individuals. *J Clin Med*. 2020;9(6):E1609. doi:10.3390/jcm9061609

15. Coleman E, Bockting W, Botzer M, et al. Standards of care for the health of transsexual, transgender, and gender-nonconforming people, version 7. *Int J Transgenderism*. 2012;13(4):165-232. doi:10.1080/15532739.2011.700873

16. Forbes C, Banisaba B. Trans Youth Parental Support Measure. [Unpublished]. Published online 2010.

17. Kozee HB, Tylka TL, Bauerband LA. Measuring transgender individuals' comfort with gender identity and appearance: development and validation of the transgender congruence scale. *Psychol Women Q*. 2012;36(2):179-196. doi:10.1177/0361684312442161

18. Lindgren TW, Pauly IB. A body image scale for evaluating transsexuals. *Arch Sex Behav*. 1975;4(6):639-656. doi:10.1007/BF01544272

19. Kuper LE, Stewart S, Preston S, Lau M, Lopez X. Body dissatisfaction and mental health outcomes of youth on gender-affirming hormone therapy. *Pediatrics*. 2020;145(4):e20193006. doi:10.1542/peds.2019-3006

20. van de Grift TC, Cohen-Kettenis PT, Steensma TD, et al. Body satisfaction and physical appearance in gender dysphoria. *Arch Sex Behav*. 2016;45(3):575-585. doi:10.1007/s10508-015-0614-1

21. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med*. 2015;34(28):3661-3679. doi:10.1002/sim.6607

22. R Core Team. A language and environment for statistical computing. Accessed August 24, 2022. http://www.R-project.org/

23. McEvenue G, Xu FZ, Cai R, McLean H. Female-to-male gender affirming top surgery: a single surgeon's 15-year retrospective review and treatment algorithm. *Aesthet Surg J*. 2017;38(1):49-57. doi:10.1093/asj/sjx116

24. Cuccolo NG, Kang CO, Boskey ER, et al. Masculinizing chest reconstruction in transgender and nonbinary individuals: an analysis of epidemiology, surgical technique, and postoperative outcomes. *Aesthetic Plast Surg*. 2019;43(6):1575-1585. doi:10.1007/s00266-019-01479-2

25. Agarwal CA, Scheefer MF, Wright LN, Walzer NK, Rivera A. Quality of life improvement after chest wall masculinization in female-to-male transgender patients: a prospective study using the BREAST-Q and Body Uneasiness Test. *J Plast Reconstr Aesthet Surg*. 2018;71(5):651-657. doi:10.1016/j.bjps.2018.01.003

26. van de Grift TC, Elfering L, Bouman M-B, Buncamper ME, Mullender MG. Surgical indications and outcomes of mastectomy in transmen: a prospective study of technical and self-reported measures. *Plast Reconstr Surg*. 2017;140(3):415e-424e. doi:10.1097/PRS.0000000000003607

27. van de Grift TC, Kreukels BPC, Elfering L, et al. Body image in transmen: multidimensional measurement and the effects of mastectomy. *J Sex Med*. 2016;13(11):1778-1786. doi:10.1016/j.jsxm.2016.09.003

28. Wilson SC, Morrison SD, Anzai L, et al. Masculinizing top surgery: a systematic review of techniques and outcomes. *Ann Plast Surg*. 2018;80(6):679-683. doi:10.1097/SAP.0000000000001354

29. Kuper LE, Rider GN, St Amand CM. Recognizing the importance of chest surgery for transmasculine youth. *Pediatrics*. 2021;147(3):e2020029710. doi:10.1542/peds.2020-029710

30. Lyman B; Montgomery Advertiser. Alabama Senate approves bill banning transgender youth treatments amid protests at state Capitol. Accessed August 24, 2022. https://www.montgomeryadvertiser.com/story/news/2021/03/02/alabama-senate-approves-bill-banning-transgender-youth-treatments/6867380002/

31. Freedom for All Americans. Legislative tracker: anti-transgender Medical Care bans. Accessed August 24, 2022. https://freedomforallamericans.org/legislative-tracker/medical-care-bans/

32. Park BC, Das RK, Drolet BC. Increasing criminalization of gender-affirming care for transgender youths—a politically motivated crisis. *JAMA Pediatr*. 2021;175(12):1205-1206. doi:10.1001/jamapediatrics.2021.2969

33. Walch A, Davidge-Pitts C, Safer JD, Lopez X, Tangpricha V, Iwamoto SJ. Proper care of transgender and gender diverse persons in the setting of proposed discrimination: A policy perspective. *J Clin Endocrinol Metab*. 2021;106(2):305-308. doi:10.1210/clinem/dgaa816

34. Rafferty J; Committee on Psychosocial Aspects of Child And Family Health; Committee on Adolescence; Section on Lesbian, Gay, Bisexual, and Transgender Health and Wellness. Ensuring comprehensive care and support for transgender and gender-diverse children and adolescents. *Pediatrics*. 2018;142(4):e20182162. doi:10.1542/peds.2018-2162

35. Turban JL, Kraschel KL, Cohen IG. Legislation to criminalize gender-affirming medical care for transgender youth. *JAMA*. 2021;325(22):2251-2252. doi:10.1001/jama.2021.7764

Journal of the Endocrine Society, 2021, Vol. 5, No. 4, 1–16
doi:10.1210/jendso/bvab011
Meta-Analysis

Meta-Analysis

Hormone Therapy, Mental Health, and Quality of Life Among Transgender People: A Systematic Review

Kellan E. Baker,[1,3] Lisa M. Wilson,[1,3] Ritu Sharma,[1,3] Vadim Dukhanin,[1,3] Kristen McArthur,[1,3] and Karen A. Robinson[2,3]

[1]Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA; [2]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA; and [3]Johns Hopkins Evidence-Based Practice Center, 615 North Wolfe Street, Baltimore, MD 21205, USA

ORCiD number: 0000-0002-9716-7936 (K. E. Baker).

Abbreviations: BDI, Beck Depression Inventory; ENIGI, European Network for the Investigation of Gender Incongruence; GnRH, gonadotropin-releasing hormone; HADS, Hospital Anxiety and Depression Scale; QOL, quality of life; RCT, randomized controlled trial; SF-36, Short Form-36 Health Survey; WPATH, World Professional Association for Transgender Health.

Received: 5 October 2020; Editorial Decision: 25 January 2021; First Published Online: 2 February 2021; Corrected and Typeset: 19 February 2021.

Abstract

We sought to systematically review the effect of gender-affirming hormone therapy on psychological outcomes among transgender people. We searched PubMed, Embase, and PsycINFO through June 10, 2020 for studies evaluating quality of life (QOL), depression, anxiety, and death by suicide in the context of gender-affirming hormone therapy among transgender people of any age. We excluded case studies and studies reporting on less than 3 months of follow-up. We included 20 studies reported in 22 publications. Fifteen were trials or prospective cohorts, one was a retrospective cohort, and 4 were cross-sectional. Seven assessed QOL, 12 assessed depression, 8 assessed anxiety, and 1 assessed death by suicide. Three studies included trans-feminine people only; 7 included trans-masculine people only, and 10 included both. Three studies focused on adolescents. Hormone therapy was associated with increased QOL, decreased depression, and decreased anxiety. Associations were similar across gender identity and age. Certainty in this conclusion is limited by high risk of bias in study designs, small sample sizes, and confounding with other interventions. We could not draw any conclusions about death by suicide. Future studies should investigate the psychological benefits of hormone therapy among larger and more diverse groups of transgender people using study designs that more effectively isolate the effects of hormone treatment.

Key Words: Transgender, hormone therapy, sex hormones, mental health, systematic review

ISSN 2472-1972

Transgender people are those whose gender identity is different from the sex they were assigned at birth. Estimates of the size of the transgender population vary depending on how the data are collected [1]. In studies that rely on clinical records, estimates range between 1 and 30 people per 100 000 (0.001% to 0.03%) [2]. Studies that focus instead on self-report among nonclinical populations find estimates that range between 0.1% and 2% [2].

Many transgender people seek medical services to affirm their gender identity. According to the *Standards of Care for Transsexual, Transgender, and Gender Non-Conforming People* maintained by the World Professional Association for Transgender Health (WPATH), gender-affirming medical care is different for each individual and may include a variety of services and procedures, such as psychological support, hormone therapy, and surgeries [3]. Hormone therapy, which typically involves estrogens and anti-androgens for transgender women and other trans-feminine people and testosterone for transgender men and other trans-masculine people, is a common component of medical gender affirmation [4]. Because hormone treatment can have a powerful effect on physical appearance, it is often a priority for transgender people seeking medical gender affirmation [5]. Gender-affirming hormone therapy can be managed for most patients by primary care providers, as it typically involves long-term maintenance on doses similar to those used for cisgender patients with conditions such as hypogonadism [6, 7]. Some clinicians require a minimum period of psychological counseling before hormone therapy can be initiated, while others provide hormone therapy on the basis of informed consent [8].

The need for gender-affirming care is often characterized using psychiatric diagnoses such as gender dysphoria, which replaced gender identity disorder in the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) [9]. The 11th International Classification of Diseases (ICD-11) replaces these terms with a diagnosis called gender incongruence (codes: HA60, HA61, HA6Z), which is located in a new chapter on sexual health. These changes clarify that the target of gender-affirming medical interventions is not the person's gender identity itself but rather the clinically significant distress that can accompany a misalignment between gender identity and sex assigned at birth [10]. Some countries have further underscored that transgender identity is not a pathology by recognizing gender affirmation as fundamental to the human right to self-definition and removing requirements that transgender people seeking gender-affirming medical care present with a diagnosis such as gender dysphoria [11].

Several previous reviews have indicated that gender-affirming hormone therapy is associated with psychological benefits that include reductions in depression and anxiety and improvements in quality of life (QOL) among transgender people [12-17]. Most of these reviews did not require a minimum duration of hormone therapy [14-17]. One review that did impose a minimum follow-up requirement is 10 years old [12]. The other that required a minimum of 3 months of therapy included only uncontrolled prospective cohorts, which resulted in a sample of only 3 studies [13]. A comprehensive review without a minimum follow-up period assessed gender-affirming hormone therapy and surgeries only in adolescents [17]. By requiring a minimum duration of hormone treatment but considering all ages and a variety of study designs, we sought to update and more completely summarize the growing evidence base regarding the relationship between gender-affirming hormone therapy and psychological outcomes in transgender people.

Search Strategy and Selection Criteria

This review is one of a series of systematic reviews on gender-affirming care conducted for WPATH to inform the eighth revision of the *Standards of Care*. The protocol is registered on PROSPERO (CRD42018115379) [18], and we followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines in reporting our findings [19].

We searched PubMed, Embase, and PyscINFO from inception to October 2018 and updated the search through June 10, 2020, for studies assessing QOL, depression, anxiety, and death by suicide among transgender participants of any age in the context of gender-affirming hormone therapy [20]. We also reviewed the reference lists of previous reviews and hand-searched the *International Journal of Transgenderism*. Using DistillerSR [21], 2 reviewers independently screened titles, abstracts, and full-text articles. Differences were resolved through consensus adjudication.

We included studies that evaluated the psychological effects of any testosterone, estrogen, or anti-androgen formulation used for gender affirmation. We also considered gonadotropin-releasing hormone (GnRH) analogues used as anti-androgens or for puberty delay. Study participants must have been on hormone therapy for at least 3 months in order to reflect a minimum time for expected onset of effects [3]. Health care provider supervision was not required. We excluded studies that did not state therapy type and duration, including the range for cross-sectional studies. We included studies regardless of language (the search terms were in English) and country of origin, and we accepted any study design except case reports.

We created standardized forms for data extraction using the Systematic Review Data Repository system. The data extracted included participant demographics; study design

and methods; hormone therapy type, dose, and duration; potential confounders such as gender-affirming surgery status; outcome scales [20]; and psychological outcomes. From studies that used the Short Form-36 Health Survey (SF-36) to measure QOL, we extracted scores in all domains [22]. For studies that used measures with depression or anxiety subscales, we extracted only the subscale scores corresponding to the psychological outcomes of interest (eg, the depression subscale of the Minnesota Multiphasic Personality Inventory [MMPI]). We extracted comparisons with cisgender controls or general population norms only when longitudinal findings in a transgender population or comparisons with an untreated transgender control group were not reported. We used WebPlotDigitizer to extract data reported only in figures [23].

Two reviewers independently assessed risk of bias [20]. For randomized controlled trials (RCTs), we used the revised Cochrane tool [24]. For non-randomized studies, we used the Cochrane Risk of Bias Assessment Tool for Non-Randomized Studies of Interventions (ROBINS-I) [25]. One reviewer graded strength of evidence for each outcome using the Agency for Healthcare Research and Quality Methods Guide for Conducting Comparative Effectiveness Reviews [26]. We considered the directionality and magnitude of effects reported in cross-sectional studies as additional context for our evaluation of evidence from trials and prospective and retrospective cohorts. Each strength of evidence assessment was confirmed by a second reviewer.

WPATH provided the research question and reviewed the protocol, evidence tables, and report. WPATH had no role in study design, data collection, analysis, interpretation, or drafting. The corresponding author had full access to all the data and had final responsibility for the decision to submit for publication. The authors are responsible for all content, and statements in this report do not necessarily reflect the official views of or imply endorsement by WPATH.

Results

We retrieved 1753 nonduplicate studies for the broader systematic review project of which this review was a part (Fig. 1). After screening and full-text review for the specific research question on the psychological effects of gender-affirming hormone therapy, 20 studies reported in 22 publications were included (Table 1): 1 RCT [27], 2 before-after trials [28, 29], 12 prospective cohorts reported in 13 publications [30-42], 1 retrospective cohort reported in 2 publications [43, 44], and 4 cross-sectional studies [45-48]. De Vries (2014) [35] reported on a subset of the participants in de Vries (2011) [34] who continued in care. We counted these publications as a single study but extracted and reported data separately because the characteristics of the study's adolescent population changed substantially in the period between the 2 publications. Similarly, Asscheman (2011) [44] reported on an extension of Asscheman (1989) [43]; we counted these as a single study but extracted data separately. In Table 1 and in the subsequent tables for each outcome, studies are ordered first by study design (RCTs, before-after trials, prospective cohorts, retrospective cohorts, and cross-sectional studies); within these categories, studies are presented in the following order according to how the study results were reported: adult transgender women only, adult transgender men only, adult transgender women and transgender men together, and transgender adolescents (no study reported separate results by gender identity for transgender youth). Where multiple studies shared the same study design and population, they are additionally ordered chronologically.

The time frame covered in the included studies began in 1972 [43], but most studies dated from post-2000. Eight studies were conducted in Italy [27-29, 31, 32, 36, 39, 41]; 2 each in Belgium [37, 48], the Netherlands [34, 35, 43, 44], the United States [30, 47], and Spain [38, 45]; and 1 in the United Kingdom [33], Turkey [42], and France [46]. One study recruited participants from Switzerland and Germany [40]. One study was part of the European Network for the Investigation of Gender Incongruence (ENIGI), which is a research collaborative between clinics providing gender-affirming care to transgender people in Ghent (Belgium), Amsterdam (Netherlands), Oslo (Norway), and Hamburg (Germany). The ENIGI study included in this review drew participants only from the Ghent clinic [37].

The study sizes ranged from 20 to 1331, although most had fewer than 60 participants. Fourteen studies reported on testosterone formulations in adult transgender men [27, 29, 31-33, 36, 39-46, 48]. These formulations were typically injectable testosterone cypionate or enanthate, although some studies used long-acting injectable testosterone undecanoate or daily transdermal gels. Ten studies reported on estrogen formulations in adult transgender women, usually in conjunction with an anti-androgen such as cyproterone acetate or spironolactone [28, 31, 33, 36, 37, 39, 43-47]. Estrogen formulations included transdermal, oral, or injectable estradiol (commonly estradiol valerate) or conjugated estrogens. Three studies reported on the psychological effects of GnRH therapy for puberty delay among mixed-gender groups of transgender adolescents [30, 34, 35, 38]. No study reported on hormone therapy among nonbinary people.

All studies that reported information about recruitment drew their participants largely or exclusively from specialized clinics dedicated to providing gender-affirming care for transgender people. These clinics were typically part of larger systems such as university hospitals. Clinic-based

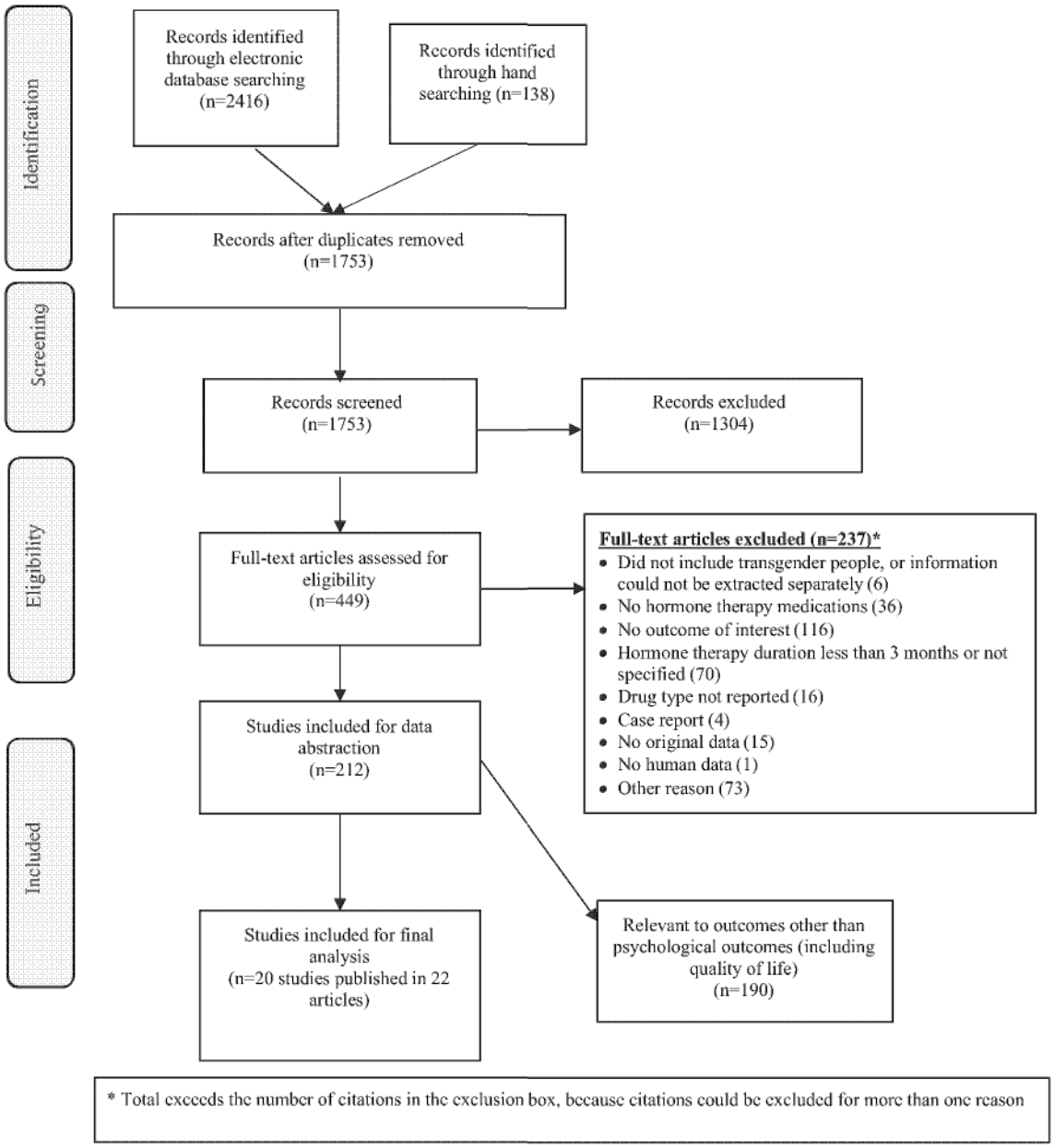


Figure 1. PRISMA flow diagram.

studies often applied strict eligibility criteria that included a period of psychiatric evaluation and a formal diagnosis of gender dysphoria before hormone therapy was initiated. Some studies also reported that psychological counseling was either available or required during the course of hormone therapy. In many cases, hormone therapy was considered a prerequisite for gender-affirming surgeries. The type and timing of gender-affirming surgeries and the proportion of participants for whom hormone therapy and surgeries were assessed simultaneously varied widely: some studies assessed only participants who had not had any type of gender-affirming surgery [27, 28, 30-32, 34, 36, 38-40, 42, 46, 47], while in others some or all participants underwent gender-affirming surgeries during the study period [29, 33, 35, 43-45, 48].

Quality of Life

Seven studies, including 1 RCT [27], 2 before-after trials [28, 29], 2 prospective cohorts [30, 39], and 2 cross-sectional studies [46, 48], assessed QOL (Table 2). An RCT found an improvement of approximately 5.5 points on a 10-point measure of life satisfaction across 3 groups of transgender men (n = 15 each) after 1 year of testosterone treatment ($P < 0.05$) [27]. A before-after trial similarly reported that life satisfaction scores almost

Table 1. Studies Reporting Effects of Gender-Affirming Hormone Therapy on Psychological Outcomes Among Transgender People

Author, year Location Study name	Study design	Start year	Transgender population	Overall N	Age in years	Baseline HT status	Outcomes	GAS status	Risk of bias
Pelusi, 2014 [27] Italy	Randomized controlled trial[a]	NR	Men	45	Mean: 29.5	No previous HT	QOL	No GAS before or during study	High
Gava, 2016 [28] Italy	Before-after trial	NR	Women	40	Mean: 3.2 (range, 19–55)	No previous HT	QOL, Depression	No GAS before or during study	Low
Gava, 2018 [29] Italy	Before-after trial[a]	NR	Men	50	Mean: 30.1 (range, 21–42)	No previous HT	QOL	72% (n = 36) had gonadectomy during study	Serious
Fuss, 2015 [37] Belgium ENIGI (NCT01072825)	Prospective cohort	2010	Women	20	Mean: 33.9 (range, 17–48)	No previous HT	Anxiety	NR	Serious
Costantino, 2013 [32] Italy	Prospective cohort	2001	Men	50	Mean: 29.8	No previous HT	Depression	No GAS before or during study	Serious
Motta, 2018 [41] Italy	Prospective cohort	2013	Men	52	Mean: 28.3	No previous HT	Anxiety	NR	Moderate
Turan, 2018 [42] Turkey	Prospective cohort[b]	NR	Men	37	Mean: 24.6	No previous HT	Depression, Anxiety	No GAS before or during study	Moderate
Metzger, 2019 [40] Switzerland, Germany	Prospective cohort[b]	2013	Men	23	Mean: 27.2 (range, 18–51)	No previous HT	Depression	No GAS before or during study	Moderate
Colizzi, 2014 [31] Italy	Prospective cohort	2008	Women and men	107	Mean: 29.2	No previous HT	Depression, Anxiety	No GAS before or during study	Low
Manieri, 2014 [39] Italy	Prospective cohort	NR	Women and men	83	Mean: 32.7 (women), 30.2 (men)	No previous HT	QOL	No GAS before or during study	Moderate
Fisher, 2016 [36] Italy	Prospective cohort	2012	Women and men	54	Mean: 32.5 (women), 26.3 (men)	No previous HT	Depression	No GAS before or during study	Low
Defreyne, 2018 [33] UK	Prospective cohort	2012	Women and men	155	Median: 27 (range, 18–52)	No previous HT	Depression, Anxiety	Some had GAS during study; % and type NR	Serious
Asscheman, 1989 [43] Netherlands	Retrospective cohort[b,d]	1972	Women and men	425	Median: 32 (women, range, 16–67); 25.4 (men, range, 16–54)	Previous HT for at least 6 months	Death by suicide	78% (n = 235) of transgender women had GAS during study; data NR for transgender men	Serious

Table 1. Continued

Author, year Location Study name	Study design	Start year	Transgender population	Overall N	Age in years	Baseline HT status	Outcomes	GAS status	Risk of bias
Asscheman, 2011 [44] Netherlands	Retrospective cohort[b,d]	1975	Women and men	1331	Mean: 31.4 (women, range, 16–76); 26.1 (men, range, 16–57)	Previous HT for at least 1 year	Death by suicide	87% (n = 834) of transgender women and 94% (n = 343) of transgender men had GAS during study	Serious
Leavitt, 1980 [47] US	Cross-sectional	1976	Women	41	Range, 18–35	54% (n = 22) on HT	Depression	No previous GAS	Serious
Wierckx, 2011 [48] Belgium	Cross-sectional[b]	2009	Men	47	Mean: 37 (range, 22–54)	100% on HT	QOL	100% had GAS, but not within previous year	Serious
Gómez-Gil, 2012 [45] Spain	Cross-sectional	NR	Women and men	187	Mean: 29.9 (range, 15–61)	64% (n = 120) on HT	Depression, Anxiety	42% (n = 79) of all participants and 64% (n = 77) of participants on HT had previous GAS	Serious
Gorin-Lazard, 2012 [46] France	Cross-sectional[b]	NR	Women and men	61	Mean: 34.7	72% (n = 44) on HT	QOL	No previous GAS	Serious
de Vries, 2011 [34] Netherlands	Prospective cohort	2000	Girls and boys	70	Mean: 14.8 (range, 11.3–18.6)	No previous HT	Depression, Anxiety	No GAS before or during study	Moderate
de Vries, 2014 [35] Netherlands	Prospective cohort[b,c]	2000	Girls and boys	55	Mean: 14.8 (range, 11.5–18.5)	No previous HT	Depression, Anxiety	100% had GAS during study	Serious
Achille, 2020 [30] US	Prospective cohort	2013	Girls and boys	50	Mean: 16.2	No previous HT	QOL, Depression	No GAS before or during study	Moderate
López de Lara, 2020 [38] Spain	Prospective cohort[b]	2018	Girls and boys	23	Mean: 16 (range, 14-18)	No previous HT	Depression, Anxiety	No GAS before or during study	Moderate

Abbreviations: ENIGI, European Network for the Investigation of Gender Incongruence; GAS, gender-affirming surgery; HT, hormone therapy; NR, not reported; QOL, quality of life.

[a]25 participants were included in both Pelusi [27] and Gava (2018) [29]

[b]Included a cisgender control group or a comparison to general population norms

[c]All participants were also included in de Vries (2011) [34]

[d]An unknown number of participants were included in both Asscheman (1989) [43] and Asscheman (2011) [44]

Table 2. Effects of Gender-Affirming Hormone Therapy on Quality of Life Among Transgender People

Author, year Study design	Transgender population	Treatment / comparison (n)	QOL measures	Length of treatment	Findings
Pelusi, 2014 [27] RCT[a]	Men	Testoviron depot (15) vs testosterone gel (15) vs testosterone undecanoate (15)	VAS (general life satisfaction)	54 weeks	Mean QOL scores increased from 2.8 to 8.5 ($P < 0.05$) in the testoviron depot arm, from 3.2 to 8.9 ($P < 0.05$) in the testosterone gel arm, and from 2.6 to 8.0 ($P < 0.05$) in the testosterone undecanoate arm.[d] There was no difference across arms.
Gava, 2016 [28] Before-after trial	Women	Cyproterone acetate + estradiol (20) vs leuprolide acetate + estradiol (20)	VAS (general life satisfaction) SF-36	12 months	Mean QOL scores did not change in either arm. No comparisons across arms were reported.
Gava, 2018 [29] Before-after trial[a]	Men	Testosterone undecanoate (25)[c] vs testosterone enanthate (25)[c]	VAS (general satisfaction)	5 years	Mean QOL scores increased from 4.3 ± 3.1 to 8.1 ± 1.8 ($P < 0.001$) in the testosterone undecanoate arm and from 4.3 ± 3.8 to 8.3 ± 1.7 ($P < 0.001$) in the testosterone enanthate arm. No comparisons across arms were reported.
Manieri, 2014 [39] Prospective cohort	Women	HT (56)	WHOQOL	12 months	Mean QOL scores increased from 62.5 to 72.2 ($P < 0.05$).[d]
Manieri, 2014 [39] Prospective cohort	Men	HT (27)	WHOQOL	12 months	Mean QOL scores did not change.
Wierckx, 2011 [48] Cross-sectional[b]	Men	HT (47)[c]	SF-36	At least 3 years	Mean QOL scores on the VT and MH subscales were lower for transgender men than cisgender men (VT subscale: 62.1 ± 20.7 vs 71.9 ± 18.3, $P = 0.002$; MH subscale: 72.6 ± 19.2 vs 79.3 ± 16.4, $P = 0.020$). There were no other differences between transgender men and either cisgender men or cisgender women.
Gorin-Lazard, 2012 [46] Cross-sectional[b]	Women and men	HT (44) vs no HT (17)	SF-36	Median: 20 months (range, 12–42 months)	Mean QOL scores were generally higher in the group receiving HT vs the group not receiving HT (MCS: 51.0 ± 7.7 vs 39.8 ± 12.7, $P = 0.003$; MH subscale: 76.4 ± 14.1 vs 59.1 ± 19.6, $P = 0.004$; RE subscale: 88.6 ± 22.7 vs 54.9 ± 40.7, $P = 0.001$; SF subscale: 83.2 ± 23.3 vs 69.9 ± 24.2, $P = 0.026$). There were no differences in the other subscales.
Achille, 2020 [30] Prospective cohort	Girls and boys	GnRH treatment + HT (47)	Q-LES-Q-SF	12 months	Mean QOL scores did not change.

Abbreviations: GnRH, gonadotropin-releasing hormone; HT, hormone therapy; MCS, Mental Component Summary; MH, mental health; QOL, quality of life; RCT, randomized controlled trial; RE, role functioning/emotional; SF, social functioning; SF-36, Short Form-36 Health Survey; VAS, visual analog scale; VT, vitality; WHOQOL, World Health Organization Quality of Life measure.

[a]10 participants on testosterone enanthate and 15 participants on testosterone undecanoate were included in both Pelusi [27] and Gava (2018) [29]

[b]Included a cisgender control group or a comparison to general population norms

[c]Included participants who had undergone gender-affirming surgery/surgeries, or surgery status not reported

[d]No standard deviations reported

doubled among transgender men (n = 50) over 5 years [29]. A prospective study found a 16% improvement in QOL scores among transgender women (n = 56) after 1 year of treatment ($P < 0.05$) but no change among transgender men (n = 27) [39]. Another before-after trial reported no difference in SF-36 scores among 2 groups of transgender women (n = 20 each) after 1 year [28]. Among adolescents, a mixed-gender prospective cohort (n = 50) showed no difference in QOL scores after a year of endocrine interventions, which included combinations of GnRH analogues and estrogen or testosterone formulations [30]. No study found that hormone therapy decreased QOL scores. We conclude that hormone therapy may improve QOL among transgender people. The strength of evidence for this conclusion is low due to concerns about bias in study designs, imprecision in measurement because of small sample sizes, and confounding by factors such as gender-affirming surgery status.

Depression

Twelve studies, including 1 before-after trial [28], 9 prospective cohorts [30-36, 38, 40, 42], and 2 cross-sectional studies [45, 47], assessed depression (Table 3). A prospective study found that the proportion of transgender men and transgender women (n = 107) showing symptoms of depression decreased from 42% to 22% over 12 months of treatment ($P < 0.001$) [31]. In 2 other prospective cohorts, Beck Depression Inventory (BDI-II) scores improved by more than half among both transgender men (n = 26) and transgender women (n = 28) after 24 months of therapy ($P < 0.001$) [36] and improved from 15.7 ± 12.3 to 8.1 ± 6.2 among transgender men (n = 23) after 6 months ($P < 0.001$) [40]. A fourth prospective study reported improvements of 1.05 points (95% CI: −1.87, −0.22) and 1.42 points (95% CI: −2.61, −0.24) on the 21-point Hospital Anxiety and Depression Scale (HADS) among 91 transgender women and 64 transgender men after 12 months ($P = 0.013$ and $P = 0.019$, respectively) [33]. A before-after trial, however, found no change in BDI-II scores among 2 groups of transgender women (n = 20 each) after 1 year [28]. Two prospective studies reported no difference among transgender men (n = 37) after 24 weeks [42] or among transgender men (n = 50) after 12 months [32], although in the latter study this outcome did not change from a baseline median of 0.0 ("not at all depressed") on an unvalidated 4-point scale. Among adolescents, 2 mixed-gender prospective cohorts (n = 50 and n = 23, respectively) showed improvements in depression scores after 1 year of treatment with GnRH analogues and estrogen or testosterone formulations (both $P < 0.001$) [30, 38]. Another prospective study reported that BDI scores improved almost by half among adolescents (n = 41) after a mean of 1.88 years of treatment with GnRH analogues to delay puberty ($P = 0.004$) [34]. The overall improvement after several subsequent years of testosterone or estrogen therapy in this cohort (n = 32) was smaller, however, resulting in no significant change from baseline [35]. No study found that hormone therapy increased depression. We conclude that hormone therapy may decrease depression among transgender people. The strength of evidence for this conclusion is low due to concerns about study designs, small sample sizes, and confounding.

Anxiety

Eight studies, including 7 prospective cohorts [31, 33-35, 37, 38, 41, 42] and 1 cross-sectional study [45], assessed anxiety (Table 4). One prospective study found that Symptom Checklist 90-Revised scores indicating a probable anxiety disorder among a mixed-gender group of adults (n = 107) improved from borderline to normal over 12 months ($P < 0.001$) [31]. Another prospective study, however, did not find a difference in HADS anxiety scores among either transgender men (n = 64) or transgender women (n = 91) after 1 year [33], and a third study reported no change in the number of transgender men (6/52, 12%) with a diagnosed anxiety disorder after 7 months [41]. Likewise, 2 other prospective studies found no difference in anxiety scores among transgender men (n = 37) after 24 weeks of treatment [42] or transgender women (n = 20) after 12 months [37], although this latter finding represented no change from a baseline median score of 0 (answering "no" to the question, "do you feel anxious?") on an unvalidated 3-point scale. Among adolescents, 1 prospective study saw mean anxiety scores in a mixed-gender group (n = 23) improve from 33.0 ± 7.2 to 18.5 ± 8.4 after 1 year ($P < 0.001$) [38], but another reported no changes in anxiety after approximately 2 years of puberty delay treatment with GnRH analogues and 4 years of hormone therapy (n = 32) [35]. No study found that hormone therapy increased anxiety. We conclude that hormone therapy may decrease anxiety among transgender people. The strength of evidence for this conclusion is low due to concerns about study designs, small sample sizes, and confounding.

Death by Suicide

One retrospective study reported in 2 publications assessed death by suicide (Table 5) [43, 44]. The first publication reported that 3 transgender women in the Amsterdam gender dysphoria study cohort (n = 303) died by suicide between 1972 and 1986 [43]. The authors calculated the number of suicide deaths expected in an age-matched stratum of

Table 3. Effects of Gender-Affirming Hormone Therapy on Depression Among Transgender People

Author, year Study design	Transgender population	Treatment / comparison (n)	Depression measures	Length of treatment	Findings
Gava, 2016 [28] Before-after trial	Women	Cyproterone acetate + estradiol (20) vs Leuprolide acetate + estradiol (20)	BDI-II	12 months	Mean depression scores did not change in either arm. No comparisons across arms were reported.
Fisher, 2016 [37] Prospective cohort	Women	HT (28)	BDI-II	24 months	Mean depression score decreased from 10.12 to 4.58 ($P < 0.001$).[d,e]
Defreyne, 2018 [33] Prospective cohort	Women	HT (91)[c]	HADS (depression subscale)	1 year	Median depression score decreased by 1.05 (95% CI: −1.87, −0.22) on a 21-point scale ($P = 0.013$).
Costantino, 2013 [32] Prospective cohort	Men	HT (50)	Ad hoc questionnaire	12 months	Depression score did not change from a median of 0.0 at baseline (IQR: 0.0, 1.0).
Fisher, 2016 [36] Prospective cohort	Men	HT (26)	BDI-II	24 months	Mean depression score decreased from 9.31 to 4.25 ($P < 0.001$).[d,e]
Defreyne, 2018 [33] Prospective cohort	Men	HT (64)[c]	HADS (depression subscale)	1 year	Median depression score decreased by 1.42 (95% CI: −2.61, −0.24) on a 21-point scale ($P = 0.019$).
Turan, 2018 [42] Prospective cohort[b]	Men	HT (37)	SCL-90-R (depression subscale)	24 weeks	Mean depression score did not change.
Metzger, 2019 [40] Prospective cohort[b]	Men	HT (23)	BDI-II	6 months	Mean depression score decreased from 15.7 ± 12.3 to 8.1 ± 6.2 ($P < 0.001$).
Colizzi, 2014 [31] Prospective cohort	Women and men	HT (107)	Zung SDS SCL-90-R (depression subscale)	12 months	Mean Zung SDS score improved from 48.40 ± 10.5 to 39.98 ± 10.79 ($P < 0.001$), and the proportion with Zung SDS scores indicating mild, moderate, or severe depression (vs no depression) decreased from 42% to 22% ($\chi^2 = 19.05$, $P < 0.001$). Mean SCL-90-R score decreased from 0.83 ± 0.74 to 0.51 ± 0.49 ($P < 0.001$), which represents an improvement from possible borderline depression to no depression.
Leavitt, 1980 [47] Cross-sectional	Women	HT (22) vs No HT (19)	MMPI (depression subscale)	At least 12 months	Mean depression score was lower in the group receiving HT vs the group not receiving HT (53.1 ± 14.7 vs 65.7 ± 11.2, $P = 0.004$).

Table 3. Continued

Author, year Study design	Transgender population	Treatment / comparison (n)	Depression measures	Length of treatment	Findings
Gómez-Gil, 2012 [45] Cross-sectional	Women and men	HT (120)[c] vs No HT (67)[c]	HADS (depression subscale)	Mean: 11.0 years (women, range, 1–46 years); 4.7 years (men, range, 1–22 years)	Mean depression score was lower in the group receiving HT vs the group not receiving HT (3.3 ± 3.2 vs 5.2 ± 4.2, $P = 0.002$).[f] The proportion with scores indicating depression (vs no depression) was larger in the group not receiving HT (31% vs 8%, $\chi^2 = 16.46$, $P = 0.001$).[f]
de Vries, 2011 [34] Prospective cohort	Girls and boys	GnRH treatment (41)	BDI	1.88 years	Mean depression score decreased from 8.31 ± 7.12 to 4.95 ± 6.72 ($P = 0.004$).
de Vries, 2014 [35] Prospective cohort[a,b]	Girls and boys	GnRH treatment + HT (32)[c]	BDI	5.9 years	Mean depression score did not change.
Achille, 2020 [30] Prospective cohort	Girls and boys	GnRH treatment + HT (47)	CESD-R, PHQ-9 (modified for adolescents)	12 months	Mean CESD-R score decreased from 21.4 to 13.9 ($P < 0.001$);[d] a score of <16 indicates no clinical depression. Mean PHQ-9 score decreased from 9.0 to 5.4 ($P < 0.001$).[d]
López de Lara, 2020 [38] Prospective cohort[b]	Girls and boys	GnRH treatment + HT (23)	BDI-II	1 year	Mean depression score decreased from 19.3 ± 5.5 to 9.7 ± 3.9 ($P < 0.001$).

Abbreviations: BDI/BDI-II, Beck Depression Inventory; GAS, gender-affirming surgery; GnRH, gonadotropin-releasing hormone; HADS, Hospital Anxiety and Depression Scale; HT, hormone therapy; IQR, interquartile range; MMPI, Minnesota Multiphasic Personality Inventory; NA, not applicable; SCL-90-R, Symptom Checklist 90-Revised; Zung SDS, Zung Self-Rating Depression Scale.

[a]All participants were also included in de Vries (2011) [34]

[b]Included a cisgender control group or a comparison to general population norms

[c]Included participants who had undergone gender-affirming surgery/surgeries, or surgery status not reported

[d]No standard deviations reported

[e]Adjusted for age, gender role, and surgery status

[f]Adjusted for age, gender, and education level

Table 4. Effects of Gender-Affirming Hormone Therapy on Anxiety Among Transgender People

Author, year	Transgender population	Treatment / comparison (n)	Anxiety measures	Length of treatment	Findings
Fuss, 2015 [37] Prospective cohort	Women	HT (20)[c]	Ad hoc questionnaire	12 months	Anxiety score did not change from a median of 0.0 at baseline.
Defreyne, 2018 [33] Prospective cohort	Women	HT (91)[c]	HADS (anxiety subscale)	1 year	Median anxiety score did not change.
Defreyne, 2018 [33] Prospective cohort	Men	HT (64)[c]	HADS (anxiety subscale)	1 year	Median anxiety score did not change.
Motta, 2018 [41] Prospective cohort	Men	HT (46)[c]	DSM	7 months	Proportion diagnosed with an anxiety disorder (6/46, 12%) did not change.
Turan, 2018 [42] Prospective cohort[b]	Men	HT (37)	SCL-90-R (anxiety subscale)	24 weeks	Mean anxiety score did not change.
Colizzi, 2014 [31] Prospective cohort	Women and men	HT (107)	SCL-90-R (anxiety subscale) Zung SAS	12 months	Mean SCL-90-R score decreased from 1.05 ± 0.95 to 0.54 ± 0.56 ($P < 0.001$), which represents an improvement from borderline anxiety disorder to no anxiety disorder. Mean Zung SAS score improved from 44.91 ± 9.59 to 37.90 ± 8.97 ($P < 0.001$), and the proportion with Zung SAS scores indicating mild, moderate, or severe anxiety (vs no anxiety) decreased from 50% to 17% ($\chi^2 = 33.03$, $P < 0.001$).
Gómez-Gil, 2012 [45] Cross-sectional	Women and men	HT (120)[c] vs No HT (67)[c]	HADS (anxiety subscale) SADS	Mean: 11.0 years (women, range, 1-46 years); 4.7 years (men, range, 1-22 years)	Mean HADS and SADS scores were lower in the group receiving HT vs the group not receiving HT (6.4 ± 3.7 vs 9.0 ± 4.0, $P = 0.001$; 8.5 ± 7.8 vs 11.0 ± 7.3, $P = 0.038$, respectively).[d] The proportion with scores indicating anxiety (vs no anxiety) was higher in the group not receiving HT ($\chi^2 = 14.46$, $P < 0.001$).[d]
de Vries, 2011 [34] Prospective cohort	Girls and boys	GnRH treatment (41)	STAI (trait subscale)	1.88 years	Mean anxiety score did not change.
de Vries, 2014 [35] Prospective cohort[a,b]	Girls and boys	GnRH treatment + HT (32)[c]	STAI (trait subscale)	5.9 years	Mean anxiety score did not change.
López de Lara, 2020 [38] Prospective cohort[b]	Girls and boys	GnRH treatment + HT (23)	STAI (trait subscale)	1 year	Mean anxiety score decreased from 33.0 ± 7.2 to 18.5 ± 8.4 ($P < 0.001$).

Abbreviations: BAI, Beck Anxiety Inventory; DSM, Diagnostic and Statistical Manual of Mental Disorders; GAS, gender-affirming surgery; GnRH, gonadotropin-releasing hormone; HADS, Hospital Anxiety and Depression Scale; HT, hormone therapy; IQR, interquartile range; SADS, Social Avoidance and Distress Scale; SCL-90-R, Symptom Checklist 90-Revised; STAI, State-Trait Anxiety Inventory; Zung SAS, Zung Self-Rating Anxiety Scale.

[a]All participants were also included in de Vries (2011) [34]

[b]Included a cisgender control group or a comparison to general population norms

[c]Included participants who have undergone gender-affirming surgery/surgeries, or surgery status not reported

[d]Adjusted for age, gender, and education level

the general male Dutch population over this period to be 0.208. No data were reported for transgender men (n = 122). An update to this study reported 17 deaths by suicide among transgender women (n = 966) and 1 among transgender men (n = 365) between 1975 and 2007 [44]. The age- and sex-stratified standardized mortality ratios were 5.70 (95% CI: 4.93, 6.54) and 2.22 (95% CI: 0.53, 6.18), respectively. The risk of bias for this study was serious due to the difficulty of identifying appropriate comparison groups and uncontrolled confounding by surgery

Table 5. Effects of Gender-Affirming Hormone Therapy on Death by Suicide Among Transgender People

Author, year	Transgender population	Treatment / comparison (n)	Measures	Length of treatment	Findings
Asscheman, 1989 [43] Retrospective cohort[a,b]	Women	HT (303)[c]	Death by suicide (confirmed by autopsy report)	Median: 4.4 years (range, 6 months to 13 years)	3 transgender women (1%) died by suicide between 1972 and 1986. The adjusted number of suicide deaths expected among the general Dutch male population was 0.208.
Asscheman, 2011 [44] Retrospective cohort[a,b]	Women	HT (966)[c]	Death by suicide (confirmed by medical report or physician information)	Median: 18.6 years (range, 0.7–44.5 years)	17 transgender women (2%) died by suicide between 1975 and 2007. The age-stratified SMR compared to the general Dutch male population was 5.70 (95% CI: 4.93, 6.54).
Asscheman, 1989 [43] Retrospective cohort[a,b]	Men	HT (122)[c]	Death by suicide (confirmation procedure NR)	Median: 3.6 years (range, 6 months to 13 years)	No deaths by suicide among transgender men were reported during the study period.
Asscheman, 2011 [44] Retrospective cohort[a,b]	Men	HT (365)[c]	Death by suicide (confirmed by medical report or physician information)	Median: 18.4 years (range, 4.7–42.6 years)	1 transgender man (0.3%) died by suicide between 1975 and 2007. The age-stratified SMR compared to the general Dutch female population was 2.22 (95% CI: 0.53, 6.18).

Abbreviations: HT, hormone therapy; NR, not reported; SMR, standardized mortality ratio.
[a]An unknown number of participants were included in both Asscheman (1989) [43] and Asscheman (2011) [44]
[b]Included a cisgender control group or a comparison to general population norms
[c]Includes participants who had undergone gender-affirming surgery/surgeries, or surgery status not reported

status and socioeconomic variables such as unemployment. We cannot draw any conclusions on the basis of this single study about whether hormone therapy affects death by suicide among transgender people.

Discussion

This systematic review of 20 studies found evidence that gender-affirming hormone therapy may be associated with improvements in QOL scores and decreases in depression and anxiety symptoms among transgender people. Associations were similar across gender identity and age. The strength of evidence for these conclusions is low due to methodological limitations (Table 6). It was impossible to draw conclusions about the effects of hormone therapy on death by suicide.

Uncontrolled confounding was a major limitation in this literature. Many studies simultaneously assessed different types of gender-affirming care and did not control for gender-affirming surgery status, making it difficult to isolate the effects of hormone therapy. Others failed to report complete information about surgery status. Additional factors that may influence both access to care and psychological outcomes, including extent of social or legal gender affirmation and exposure to determinants of health such as discrimination, were typically not considered. In addition, some evidence indicates that cyproterone acetate, a common anti-androgen assessed in many studies alongside estrogen therapy, may increase depression, which may be a source of confounding [49].

Another source of potential bias was recruitment of participants from specialized clinics that impose strict diagnostic criteria as a prerequisite for gender-affirming care. The dual role of clinicians and researchers as both gatekeepers and investigators may force transgender study participants to over- or understate aspects of their mental health in order to access gender-affirming care [8]. Similarly, transgender clinic patients may feel that they cannot opt out of research-related activities, which is a serious concern for the validity of psychological outcome measurements.

Clinic-based recruitment also overlooks transgender people who cannot access these clinics for financial or other reasons and misses those whose need for gender affirmation does not fit into current medical models. This is a particular concern for nonbinary and other gender-diverse people, for whom a model of gender affirmation as a linear transition from one binary gender to another is inaccurate [50].

Most studies used well-known scales for measuring psychological outcomes. None of these scales, however, have been specifically validated for use in transgender populations [51]. Furthermore, many scales are normed

Table 6. Strength of Evidence of Studies that Evaluate the Psychological Effects of Hormone Therapy Among Transgender People

Outcome	Number of studies (n)	Strength of evidence	Summary[a]
Quality of life	1 randomized controlled trial [27] (45)[b] 2 before-after trials [28, 29] (65)[b] 2 prospective cohorts [30, 39] (133) 2 cross-sectional studies [46, 48] (108)	Low[e]	Hormone therapy may improve quality of life among transgender people.[g]
Depression	1 before-after trial [28] (40) 9 prospective cohorts [30-36, 38, 40, 42] (569)[c] 2 cross-sectional [45, 47] (228)	Low[e]	Hormone therapy may alleviate depression among transgender people.[g]
Anxiety	7 prospective cohorts [31, 33-35, 37, 38, 41, 42] (464)[c] 1 cross-sectional [45] (187)	Low[e]	Hormone therapy may alleviate anxiety among transgender people.[g]
Death by suicide	1 retrospective cohort [43, 44] (1756)[d]	Insufficient[f]	There is insufficient evidence to draw a conclusion about the effect of hormone therapy on death by suicide among transgender people.

[a]Due to similarity of findings, the summary is the same for transgender men and transgender women and for adolescents and adults
[b]25 participants are included in both Pelusi [27] and Gava (2018) [29] and are counted once
[c]All 55 participants in de Vries (2014) [35] were also included among the 70 participants in de Vries (2011) [34] and are counted once
[d]An unknown number of participants were included in both Asscheman (1989) [43] and Asscheman (2011), [44] so the unique sample size is smaller than indicated here
[e]Evidence downgraded due to study limitations, including uncontrolled confounding, and imprecision because of small sample sizes
[f]Evidence downgraded due to study limitations, including confounding and a lack of meaningful comparison groups, and imprecision in measurement of a rare event
[g]The body of cross-sectional evidence tended to align with the conclusion

separately for (presumed cisgender) men and women [52]. Inconsistency in identification of appropriate general population norms hinders comparisons between transgender and cisgender groups, which is a major related research question that requires further investigation.

Beyond methodological concerns in the studies we assessed, our review has other limitations. First, it is likely subject to publication bias, as we may have missed studies not published in the peer-reviewed literature. Second, a number of potentially relevant studies could not be included because the authors did not report on a minimum of 3 months of treatment or did not clearly state the type and/or duration of therapy, including the range for cross-sectional studies [53-65]. Finally, even where outcome measurements were similar across studies, heterogeneity in study designs, study populations, intervention characteristics, and reporting of results (ie, some studies reported results separately by gender identity, while others did not), prevented us from quantitatively pooling results.

More research is needed to further explore the relationship between gender-affirming hormone therapy and QOL, death by suicide, and other psychological outcomes, especially among adolescents. Future studies should investigate these outcomes in larger groups of diverse participants recruited outside clinical settings. Studies assessing the relationship between gender-affirming hormone therapy and mental health outcomes in transgender populations should be prospective or use strong quasi-experimental designs; consistently report type, dose, and duration of hormone therapy; adjust for possible confounding by gender-affirming surgery status; control for other variables that may independently influence psychological outcomes; and report results separately by gender identity. Despite the limitations of the available evidence, however, our review indicates that gender-affirming hormone therapy is likely associated with improvements in QOL, depression, and anxiety. No studies showed that hormone therapy harms mental health or quality of life among transgender people. These benefits make hormone therapy an essential component of care that promotes the health and well-being of transgender people.

Acknowledgments

Financial Support: This review was partly funded by the World Professional Association for Transgender Health.

Author Contributions: R.S. developed and implemented the search strategy with input from K.B., L.W., and K.R. K.B., L.W., R.S., V.D., K.M., and K.R. screened and assessed studies, extracted data, and graded strength of evidence. K.B. wrote the report, which was reviewed by all co-authors.

Additional Information

Correspondence: Kellan E. Baker, MPH, MA, Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, 624 North Broadway, Baltimore, MD 21205, USA. Email: kbaker39@jhu.edu.

Disclosure Summary: This review was partially funded by the World Professional Association for Transgender Health (WPATH). The authors of this manuscript are responsible for its content. Statements in the manuscript do not necessarily reflect the official views of or imply endorsement by WPATH.

Data Availability: Some or all data generated or analyzed during this study are included in this published article or in the data repository listed in the References.

References

1. Collin L, Reisner SL, Tangpricha V, Goodman M. Prevalence of transgender depends on the "case" definition: a systematic review. *J Sex Med*. 2016;**13**(4):613-626.
2. Goodman M, Adams N, Corneil T, Kreukels B, Motmans J, Coleman E. Size and distribution of transgender and gender nonconforming populations: a narrative review. *Endocrinol Metab Clin North Am*. 2019;**48**(2):303-321.
3. Coleman E, Bockting W, Botzer M, et al. Standards of care for the health of transsexual, transgender, and gender-nonconforming people, version 7. *Int J Transgenderism*. 2012;**13**(4):165-232.
4. Hembree WC, Cohen-Kettenis PT, Gooren L, et al. Endocrine treatment of gender-dysphoric/gender-incongruent persons: an Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab*. 2017;**102**(11):3869-3903.
5. James SE, Herman JL, Rankin S, Keisling M, Mottet L, Anafi M. *The Report of the 2015 U.S. Transgender Survey*. National Center for Transgender Equality; 2016.
6. Deutsch MB, ed. Guidelines for the primary and gender-affirming care of transgender and gender nonbinary people. 2016. Accessed December 19, 2020. https://transcare.ucsf.edu/guidelines
7. Wylie K, Knudson G, Khan SI, Bonierbale M, Watanyusakul S, Baral S. Serving transgender people: clinical care considerations and service delivery models in transgender health. *Lancet*. 2016;**388**(10042):401-411.
8. Schulz SL. The informed consent model of transgender care: an alternative to the diagnosis of gender dysphoria. *J Humanist Psychol*. 2018;**58**(1):72-92.
9. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. American Psychiatric Association; 2013.
10. Robles R, Fresán A, Vega-Ramírez H, et al. Removing transgender identity from the classification of mental disorders: a Mexican field study for ICD-11. *Lancet Psychiatry*. 2016;**3**(9):850-859.
11. Arístegui I, Radusky PD, Zalazar V, Romero M, Schwartz J, Sued O. Impact of the Gender Identity Law in Argentinean transgender women. *Int J Transgenderism*. 2017;**18**(4):446-456.
12. Murad MH, Elamin MB, Garcia MZ, et al. Hormonal therapy and sex reassignment: a systematic review and meta-analysis of quality of life and psychosocial outcomes. *Clin Endocrinol (Oxf)*. 2010;**72**(2):214-231.
13. White Hughto JM, Reisner SL. A systematic review of the effects of hormone therapy on psychological functioning and quality of life in transgender individuals. *Transgend Health*. 2016;**1**(1):21-31.
14. Costa R, Colizzi M. The effect of cross-sex hormonal treatment on gender dysphoria individuals' mental health: a systematic review. *Neuropsychiatr Dis Treat*. 2016;**12**:1953-1966.
15. Nobili A, Glazebrook C, Arcelus J. Quality of life of treatment-seeking transgender adults: a systematic review and meta-analysis. *Rev Endocr Metab Disord*. 2018;**19**(3):199-220.
16. Rowniak S, Bolt L, Sharifi C. Effect of cross-sex hormones on the quality of life, depression and anxiety of transgender individuals: a quantitative systematic review. *JBI Database System Rev Implement Rep*. 2019;**17**(9):1826-1854.
17. Mahfouda S, Moore JK, Siafarikas A, et al. Gender-affirming hormones and surgery in transgender children and adolescents. *Lancet Diabetes Endocrinol*. 2019;**7**(6):484-498.
18. Sharma R, Robinson K, Wilson L, Baker KE. Effects of hormone therapy in transgender people. Accessed December 19, 2020. https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=115379
19. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*. 2009;**339**:b2535.
20. Baker KE, Wilson LM, Sharma R, Dukhanin V, McArthur K, Robinson KA. Data associated with the publication: Hormone therapy, mental health, and quality of life among transgender people: a systematic review. *Johns Hopkins Univ Data Arch*. **V1**. doi: 10.7281/T1/E70MXR.
21. *Evidence Partners*. DistillerSR [software]; 2020.
22. Ware JE Jr, Kosinski M, Bayliss MS, McHorney CA, Rogers WH, Raczek A. Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: summary of results from the Medical Outcomes Study. *Med Care*. 1995;**33**(4 Suppl):AS264-AS279.
23. Rohatgi A. WebPlotDigitizer: an HTML5-based online tool for to extract numerical data from plot images. 2020. https://automeris.io/WebPlotDigitizer/index.html
24. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*. 2019;**366**:l4898.
25. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. 2016;**355**:i4919.
26. Berkman ND, Lohr KN, Ansari MT, et al. Grading the strength of a body of evidence when assessing health care interventions: an EPC update. *J Clin Epidemiol*. 2015;**68**(11):1312-1324.
27. Pelusi C, Costantino A, Martelli V, et al. Effects of three different testosterone formulations in female-to-male transsexual persons. *J Sex Med*. 2014;**11**(12):3002-3011.
28. Gava G, Cerpolini S, Martelli V, Battista G, Seracchioli R, Meriggiola MC. Cyproterone acetate vs leuprolide acetate in combination with transdermal oestradiol in transwomen: a comparison of safety and effectiveness. *Clin Endocrinol (Oxf)*. 2016;**85**(2):239-246.
29. Gava G, Mancini I, Cerpolini S, Baldassarre M, Seracchioli R, Meriggiola MC. Testosterone undecanoate and testosterone

enanthate injections are both effective and safe in transmen over 5 years of administration. *Clin Endocrinol (Oxf)*. 2018;**89**(6):878-886.
30. Achille C, Taggart T, Eaton NR, et al. Longitudinal impact of gender-affirming endocrine intervention on the mental health and well-being of transgender youths: preliminary results. *Int J Pediatr Endocrinol*. 2020;**2020**:8.
31. Colizzi M, Costa R, Todarello O. Transsexual patients' psychiatric comorbidity and positive effect of cross-sex hormonal treatment on mental health: results from a longitudinal study. *Psychoneuroendocrinology*. 2014;**39**:65-73.
32. Costantino A, Cerpolini S, Alvisi S, Morselli PG, Venturoli S, Meriggiola MC. A prospective study on sexual function and mood in female-to-male transsexuals during testosterone administration and after sex reassignment surgery. *J Sex Marital Ther*. 2013;**39**(4):321-335.
33. Defreyne J, T'Sjoen G, Bouman WP, Brewin N, Arcelus J. Prospective evaluation of self-reported aggression in transgender persons. *J Sex Med*. 2018;**15**(5):768-776.
34. de Vries AL, Steensma TD, Doreleijers TA, Cohen-Kettenis PT. Puberty suppression in adolescents with gender identity disorder: a prospective follow-up study. *J Sex Med*. 2011;**8**(8):2276-2283.
35. de Vries AL, McGuire JK, Steensma TD, Wagenaar EC, Doreleijers TA, Cohen-Kettenis PT. Young adult psychological outcome after puberty suppression and gender reassignment. *Pediatrics*. 2014;**134**(4):696-704.
36. Fisher AD, Castellini G, Ristori J, et al. Cross-sex hormone treatment and psychobiological changes in transsexual persons: two-year follow-up data. *J Clin Endocrinol Metab*. 2016;**101**(11):4260-4269.
37. Fuss J, Hellweg R, Van Caenegem E, et al. Cross-sex hormone treatment in male-to-female transsexual persons reduces serum brain-derived neurotrophic factor (BDNF). *Eur Neuropsychopharmacol*. 2015;**25**(1):95-99.
38. López de Lara D, Pérez Rodríguez O, Cuellar Flores I, et al. Evaluación psicosocial en adolescentes transgénero. *An Pediatría*. 2020;**93**(1):41-48.
39. Manieri C, Castellano E, Crespi C, et al. Medical treatment of subjects with gender identity disorder: the experience in an Italian Public Health Center. *Int J Transgenderism*. 2014;**15**(2):53-65.
40. Metzger NY, Boettger S. The effect of testosterone therapy on personality traits of trans men: a controlled prospective study in Germany and Switzerland. *Psychiatry Res*. 2019;**276**:31-38.
41. Motta G, Crespi C, Mineccia V, Brustio PR, Manieri C, Lanfranco F. Does testosterone treatment increase anger expression in a population of transgender men? *J Sex Med*. 2018;**15**(1):94-101.
42. Turan Ş, Aksoy Poyraz C, Usta Sağlam NG, et al. Alterations in body uneasiness, eating attitudes, and psychopathology before and after cross-sex hormonal treatment in patients with female-to-male gender dysphoria. *Arch Sex Behav*. 2018;**47**(8):2349-2361.
43. Asscheman H, Gooren LJ, Eklund PL. Mortality and morbidity in transsexual patients with cross-gender hormone treatment. *Metabolism*. 1989;**38**(9):869-873.
44. Asscheman H, Giltay EJ, Megens JA, de Ronde WP, van Trotsenburg MA, Gooren LJ. A long-term follow-up study of mortality in transsexuals receiving treatment with cross-sex hormones. *Eur J Endocrinol*. 2011;**164**(4):635-642.
45. Gómez-Gil E, Zubiaurre-Elorza L, Esteva I, et al. Hormone-treated transsexuals report less social distress, anxiety and depression. *Psychoneuroendocrinology*. 2012;**37**(5):662-670.
46. Gorin-Lazard A, Baumstarck K, Boyer L, et al. Is hormonal therapy associated with better quality of life in transsexuals? A cross-sectional study. *J Sex Med*. 2012;**9**(2):531-541.
47. Leavitt F, Berger JC, Hoeppner JA, Northrop G. Presurgical adjustment in male transsexuals with and without hormonal treatment. *J Nerv Ment Dis*. 1980;**168**(11):693-697.
48. Wierckx K, Van Caenegem E, Elaut E, et al. Quality of life and sexual health after sex reassignment surgery in transsexual men. *J Sex Med*. 2011;**8**(12):3379-3388.
49. Heinemann LA, Will-Shahab L, van Kesteren P, Gooren LJ; Collaborating Centers. Safety of cyproterone acetate: report of active surveillance. *Pharmacoepidemiol Drug Saf*. 1997;**6**(3):169-178.
50. Reisner SL, Hughto JMW. Comparing the health of non-binary and binary transgender adults in a statewide non-probability sample. *Plos One*. 2019;**14**(8):e0221583.
51. Thompson HM, Reisner SL, VanKim N, Raymond HF. Quality-of-life measurement: assessing the WHOQOL-BREF scale in a sample of high-HIV-risk transgender women in San Francisco, California. *Int J Transgend*. 2015;**16**(1):36-48.
52. Webb A, Heyne G, Holmes J, Peta J. Assessment norms for gender and implications for transgender, nonbinary populations. *Division 44 Newsletter*. 2016. Accessed June 9, 2020. https://www.apadivisions.org/division-44/publications/newsletters/division/2016/04/nonbinary-populations
53. Heylens G, Verroken C, De Cock S, T'Sjoen G, De Cuypere G. Effects of different steps in gender reassignment therapy on psychopathology: a prospective study of persons with a gender identity disorder. *J Sex Med*. 2014;**11**(1):119-126.
54. Gorin-Lazard A, Baumstarck K, Boyer L, et al. Hormonal therapy is associated with better self-esteem, mood, and quality of life in transsexuals. *J Nerv Ment Dis*. 2013;**201**(11):996-1000.
55. Gómez-Gil E, Vidal-Hagemeijer A, Salamero M. MMPI-2 characteristics of transsexuals requesting sex reassignment: comparison of patients in prehormonal and presurgical phases. *J Pers Assess*. 2008;**90**(4):368-374.
56. Oda H, Kinoshita T. Efficacy of hormonal and mental treatments with MMPI in FtM individuals: cross-sectional and longitudinal studies. *BMC Psychiatry*. 2017;**17**(1):256.
57. Elaut E, De Cuypere G, De Sutter P, et al. Hypoactive sexual desire in transsexual women: prevalence and association with testosterone levels. *Eur J Endocrinol*. 2008;**158**(3):393-399.
58. Warmuz-Stangierska I, Stangierski A, Ziemnicka K, et al. Emotional functions in transsexuals after the first step in physical transformation. *Endokrynol Pol*. 2015;**66**(1):47-52.
59. Colton Meier SL, Fitzgerald KM, Pardo ST, Babcock J. The effects of hormonal gender affirmation treatment on mental health in female-to-male transsexuals. *J Gay Lesbian Ment Health*. 2011;**15**(3):281-299.

60. Davis SA, Colton Meier S. Effects of testosterone treatment and chest reconstruction surgery on mental health and sexuality in female-to-male transgender people. *Int J Sex Health*. 2014;**26**(2):113-128.
61. Keo-Meier CL, Herman LI, Reisner SL, Pardo ST, Sharp C, Babcock JC. Testosterone treatment and MMPI-2 improvement in transgender men: a prospective controlled study. *J Consult Clin Psychol*. 2015;**83**(1):143-156.
62. Newfield E, Hart S, Dibble S, Kohler L. Female-to-male transgender quality of life. *Qual Life Res*. 2006;**15**(9):1447-1457.
63. Gooren LJ, Sungkaew T, Giltay EJ, Guadamuz TE. Cross-sex hormone use, functional health and mental well-being among transgender men (Toms) and Transgender Women (Kathoeys) in Thailand. *Cult Health Sex*. 2015;**17**(1):92-103.
64. van Kesteren PJ, Asscheman H, Megens JA, Gooren LJ. Mortality and morbidity in transsexual subjects treated with cross-sex hormones. *Clin Endocrinol (Oxf)*. 1997;**47**(3):337-342.
65. Wiepjes CM, den Heijer M, Bremmer MA, et al. Trends in suicide death risk in transgender people: results from the Amsterdam Cohort of Gender Dysphoria study (1972-2017). *Acta Psychiatr Scand*. 2020;**141**(6):486-491.

ELSEVIER

Journal of Clinical Epidemiology 64 (2011) 401–406

Journal of Clinical Epidemiology

GRADE guidelines: 3. Rating the quality of evidence

Howard Balshem[a,*], Mark Helfand[a,b], Holger J. Schünemann[c], Andrew D. Oxman[d], Regina Kunz[e], Jan Brozek[c], Gunn E. Vist[d], Yngve Falck-Ytter[f], Joerg Meerpohl[g,h], Susan Norris[i], Gordon H. Guyatt[c]

[a]*Oregon Evidence-based Practice Center, Oregon Health and Science University, 3181 SW Sam Jackson Park Rd., Portland, OR 97239, USA*
[b]*Portland VA Medical Center, Portland, OR, USA*
[c]*Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario L8N 3Z5, Canada*
[d]*Norwegian Knowledge Centre for the Health Services, PO Box 7004, St Olavs plass, 0130 Oslo, Norway*
[e]*Basel Institute of Clinical Epidemiology, University Hospital Basel, Hebelstrasse 10, 4031 Basel, Switzerland*
[f]*Division of Gastroenterology, Case Medical Center and VA, Case Western Reserve University, Cleveland, OH 44106, USA*
[g]*German Cochrane Center, Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, 79104 Freiburg, Germany*
[h]*Division of Pediatric Hematology and Oncology, Department of Pediatric and Adolescent Medicine, University Medical Center Freiburg, 79106 Freiburg, Germany*
[i]*Department of Medical Informatics and Clinical Epidemiology, Oregon Health and Science University, Portland, OR 97239-3098, USA*

Accepted 30 July 2010

Abstract

This article introduces the approach of GRADE to rating quality of evidence. GRADE specifies four categories—high, moderate, low, and very low—that are applied to a body of evidence, not to individual studies. In the context of a systematic review, quality reflects our confidence that the estimates of the effect are correct. In the context of recommendations, quality reflects our confidence that the effect estimates are adequate to support a particular recommendation. Randomized trials begin as high-quality evidence, observational studies as low quality. "Quality" as used in GRADE means more than risk of bias and so may also be compromised by imprecision, inconsistency, indirectness of study results, and publication bias. In addition, several factors can increase our confidence in an estimate of effect. GRADE provides a systematic approach for considering and reporting each of these factors. GRADE separates the process of assessing quality of evidence from the process of making recommendations. Judgments about the strength of a recommendation depend on more than just the quality of evidence.

Keywords: Quality assessment; Body of evidence; Imprecision; Indirectness; Inconsistency; Publication bias

1. Introduction

In the two previous articles in this series, we introduced GRADE; provided an overview of the GRADE process for developing recommendations and the final outputs of that process, the evidence profile, and Summary of Findings table; and described the process for framing questions and identifying outcomes [1,2]. In this third article, we will introduce GRADE's approach to rating the quality of evidence. The goal is to provide a conceptual overview of the approach. A more detailed description, accompanied by examples, will follow in articles dealing with factors that may lead to rating down or rating up the quality of evidence [3–7].

2. What we do not mean by quality of evidence

In discussions of quality of evidence, confusion often arises between evidence and opinion and between quality of evidence and strength of recommendations. We, therefore, begin by explaining what we do not mean by quality of evidence.

3. Opinion is not evidence

In the absence of high-quality evidence, clinicians must look to lower quality evidence to guide their decisions.

The GRADE system has been developed by the GRADE Working Group. The named authors drafted and revised this article. A complete list of contributors to this series can be found on the *Journal of Clinical Epidemiology* Web site.

* Corresponding author. Oregon Evidence-based Practice Center, OHSU-BICC, 3181 SW Sam Jackson Park Rd., Portland, OR 97239, USA. Tel.: 503-220-8262 x54487; fax: 503-418-3332.
E-mail address: balshemh@ohsu.edu (H. Balshem).

0895-4356/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi: 10.1016/j.jclinepi.2010.07.015

Key Points

- GRADE provides a framework for assessing quality that encourages transparency and an explicit accounting of the judgments made.
- GRADE distinguishes between quality assessment conducted as part of a systematic review and that undertaken as part of guideline development.
- The optimal application of GRADE requires systematic review of the impact of alternative management strategies on all patient-important outcomes.
- Information about study limitations, imprecision, inconsistency, indirectness, and publication bias is necessary for decision makers, clinicians, and patients to understand and have confidence in the assessment of quality and estimate of effect size.

Confusion arises when, in such situations, guideline developers classify "expert opinion" as a type of evidence. Developing recommendations always requires the opinion of experts, the basis of which includes experience with patients, an understanding of biology and mechanism, and knowledge and understanding of preclinical and early clinical research as well as of the results of randomized clinical trials and observational studies. Guideline developers should always engage experts to help understand the evidence; they must also uncover and make clear the evidence that underlies the experts' opinions and rate the quality of that evidence, not the opinions that follow from the evidence and its interpretation.

An example illustrates the difference between evidence and expert opinion. Suppose that during attending rounds with medical students and residents, an endocrinologist explains the rationale for tight glycemic control in diabetes. Table 1 shows the two assertions he makes and the evidence he cites to support them. The evidence he cites for opinion 1 is exclusively his personal clinical experience. For opinion 2, he cites his own experience and refers (with no more than a general statement) to evidence from clinical research.

It seems highly plausible that opinion 1 might reasonably be based on careful observation. If patients who complain of fatigue, polyuria, or other symptoms return in a few days saying they are better, initiation of treatment is the likeliest explanation. The phenomenon of a patient who had no complaints returning, a few days later, to say how much better she is would be particularly memorable. Unfortunately, there are many other potential explanations of these observations. The endocrinologist's impression of the extent of patients' reports of benefit may be inaccurate, he may be forgetting many patients who failed to improve, or the apparent improvement in some patients may be because of natural history, placebo effects, leading questions on the part of the clinician, or the patient's desire to please. Without, at the very least, a rigorous and structured approach to data collection, we could consider the endocrinologist's report of his clinical experience (but not the opinion that he arrived at from his interpretation of that experience) as evidence from an uncontrolled case series and classify it as very low quality.

Whereas the implicit study design underlying the evidence for opinion 1 is a before–after study, opinion 2 suggests a parallel group comparison, which in this case has serious problems. If indeed his memory is accurate (patients with tighter control in his practice do achieve better outcomes), the reason may be that their success in controlling their glucose reflects differences in their underlying disease strongly associated with their likelihood of suffering complications. This risk of bias from unrecognized prognostic imbalance, as well as from the uncertainty and imprecision associated with the endocrinologist's memory of the events, would lead us again to classify his observations as very low quality evidence.

4. A particular quality of evidence does not necessarily imply a particular strength of recommendation

A second area of confusion relates to the distinction between assessing the quality of evidence and making a recommendation. Later articles in this series will provide a detailed discussion of GRADE's approach to deciding on the direction and strength of recommendations. We note here the importance of GRADE's explicit separation of the process for assessing the quality of a body of evidence from the process for making recommendations based in part on those assessments. Although higher quality evidence is more likely to be associated with strong recommendations than lower quality evidence, a particular level of quality does not imply a particular strength of recommendation. Sometimes, low or very low quality evidence can lead to a strong recommendation.

For instance, consider the decision to administer aspirin or acetaminophen to children with chicken pox. Observational studies have observed an association between aspirin administration and Reye's syndrome [8–11]. Because aspirin and acetaminophen are similar in their analgesic and antipyretic effects, the low-quality evidence regarding the potential harms of aspirin does not preclude a strong recommendation for acetaminophen.

Similarly, high-quality evidence does not necessarily imply strong recommendations. For example, faced with a first deep venous thrombosis (DVT) with no obvious provoking factor patients must, after the first months of anticoagulation, decide whether to continue taking warfarin long term. High-quality randomized controlled trials show that continuous warfarin will decrease the risk of recurrent thrombosis but at the cost of increased risk of bleeding and inconvenience [12–15]. Because patients with varying values and

Table 1
Expert opinion vs. evidence

Expert opinion	Evidence
Tight control will make a patient feel better	"In my 20 years in practice I have started treatment for newly diagnosed diabetes many times. I almost always see these patients back a week or so after starting treatment, and the great majority say they feel much better than they did before. Even a patient who denied having any complaints or symptoms will come back and say she has more energy, particularly in the afternoons, and will marvel at how much better she feels in general."
Tight control will reduce the long-term risk of developing kidney disease, neuropathy, and blindness	"I institute tight control on every patient—I believe they all deserve the best possible treatment—so I have a lot of experience with this. I have many patients who have been with me for a decade, or even several decades, and who take their medicine faithfully and have great blood sugars. These patients also have very few complications. On the other hand, I have a lot of patients who have terrible control and develop complications early on. Also, there are a lot of studies showing that tight control reduces the risk of complications."

preferences are likely to make different choices, guideline panels addressing whether patients should continue or terminate warfarin may—despite the high-quality evidence—offer a weak recommendation.

5. So what do we mean by "quality of evidence"?

GRADE distinguishes between quality assessment conducted as part of a systematic review and that undertaken in the process of guideline development. We, therefore, provide two definitions of "quality of evidence."

The optimal application of GRADE requires systematic reviews of the impact of alternative management approaches on all patient-important outcomes [1]. In the context of a systematic review, the ratings of the quality of evidence reflect the extent of our confidence that the estimates of the effect are correct. In the context of making recommendations, the quality ratings reflect the extent of our confidence that the estimates of an effect are adequate to support a particular decision or recommendation.

The reason for the different definitions is that the conduct of systematic reviews does not include processes required for making rigorous recommendations. In particular, unless the systematic review team includes members who will use the review as part of guideline development, authors of systematic reviews are, generally, not in a position to weigh the trade-offs between the desirable and undesirable consequences of adhering to a recommendation. Relevant stakeholders are in a better position to make these judgments. For example, in the DVT case described earlier, a systematic review might provide reliable estimates of the magnitude of effect and associated confidence intervals (CIs) for symptomatic thromboembolism and bleeding and the mortality associated with both of these events, but the reviewers who wrote it would not be able to provide reliable judgments about whether the benefit of warfarin treatment is worth the risk. Such judgments must also include considerations of values, cost, and pertinent stakeholder input.

On the other hand, a guideline (or a clinician applying the evidence from a systematic review) must assess the quality of the evidence in the context of the decision regarding anticoagulation. In considering this trade-off, a guideline panel must decide whether or not to recommend anticoagulation (and the strength of that recommendation) in light of the effect on the risk of symptomatic thromboembolism, their confidence in the effect estimates, and the corresponding risks and confidence in estimates of serious bleeding. Although the processes for assessing quality are the same, authors of systematic reviews and authors of guidelines will apply the criteria differently. We will highlight this different application of criteria in the fifth article in this series, which addresses the assessment of precision in rating the quality of the evidence [5].

6. Quality in GRADE means more than risk of bias

In the clinical epidemiological literature, when used at all, "quality" commonly refers to a judgment on the internal validity (i.e., risk of bias) of an individual study. To arrive at a rating, reviewers consider features in controlled trials such as randomization, allocation concealment, blinding, and use of intention to treat analysis. In observational studies, they consider appropriate measurement of exposure and outcome as well as appropriate control of confounding; in both controlled trials and observational studies, they consider loss to follow-up and may consider other aspects of design, conduct, and analysis that influence the risk of bias.

GRADE judgments refer not to individual studies but to a body of evidence, and quality, as used in GRADE, means more than risk of bias. A body of evidence (for instance, a number of well-designed and executed trials) may be associated with a low risk of bias, but our confidence in effect estimates may be compromised by a number of other factors (imprecision, inconsistency, indirectness, and publication bias). There are also factors, particularly relevant to observational studies, that may lead to rating up quality, including the magnitude of treatment effect and the presence of a dose–response gradient.

GRADE's specific uses of the terms "quality" and "risk of bias" (labeled "study limitations" in previous GRADE publications) require authors to take care in using these terms when they describe their findings and reasoning in

Table 2
Significance of the four levels of evidence

Quality level	Current definition	Previous definition
High	We are very confident that the true effect lies close to that of the estimate of the effect	Further research is very unlikely to change our confidence in the estimate of effect
Moderate	We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different	Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate
Low	Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect	Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate
Very low	We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect	Any estimate of effect is very uncertain

the context of a systematic review or guideline. Well-conducted studies may be part of a body of evidence rated low quality because they only provide indirect or imprecise evidence for the question of interest. Although clinical epidemiologists and others have attributed other meanings to the word "quality" (typically risk of bias), we believe the meaning described here corresponds more closely to the common and nontechnical understanding of "quality."

7. GRADE specifies four categories for the quality of a body of evidence

Although the quality of evidence represents a continuum, the GRADE approach results in an assessment of the quality of a body of evidence as high, moderate, low, or very low. Table 2 presents what GRADE means by each of these four categories and contrasts their current definition with the previous definition [16], which focused on the implications of the levels of evidence for future research (the lower the quality, the more likely further research would change our confidence in the estimates, and the estimates themselves). The earlier characterization has been criticized—we believe legitimately—because there are many situations in which we cannot expect higher quality evidence to be forthcoming. We, nevertheless, consider the prior characterization of quality to provide an alternative under circumstances when obtaining new compelling evidence is plausible.

8. Arriving at a quality rating

When we speak of evaluating quality, we are referring to an overall rating for each important outcome across studies. As discussed in the previous article in this series that addressed the framing of the question [2], before assessing the quality of the evidence, systematic reviewers and guideline developers should identify all potential patient-important outcomes, including benefits, harms, and costs. Reviewers will then assess the quality of evidence for each important outcome.

Table 3 summarizes GRADE's approach to rating the quality of evidence, which begins with the study design (trials or observational studies) and then addresses five reasons to possibly rate down the quality of evidence and three to possibly rate up the quality. Subsequent articles in this series will address, in detail, the meaning and use of each of these criteria. Here, we discuss why these criteria, in particular, have been identified as important in assessing the quality of a body of evidence.

Table 3
A summary of GRADE's approach to rating quality of evidence

Study design	Initial quality of a body of evidence	Lower if	Higher if	Quality of a body of evidence
Randomized trials	High ⇒	Risk of Bias −1 Serious −2 Very serious Inconsistency −1 Serious −2 Very serious Indirectness −1 Serious −2 Very serious Imprecision −1 Serious −2 Very serious Publication bias −1 Likely −2 Very likely	Large effect +1 Large +2 Very large Dose response +1 Evidence of a gradient All plausible residual confounding +1 Would reduce a demonstrated effect +1 Would suggest a spurious effect if no effect was observed	High (four plus: ⊕⊕⊕⊕) Moderate (three plus: ⊕⊕⊕◯) Low (two plus: ⊕⊕◯◯) Very low (one plus: ⊕◯◯◯)
Observational studies	Low ⇒			

9. Rationale for using GRADE's definition of quality

To be useful to decision makers, clinicians, and patients, systematic reviews must provide not only an estimate of effect for each outcome but also the information needed to judge whether these estimates are likely to be correct. What information about the studies in a review affects our confidence that the estimate of an effect is correct?

To answer this question, consider an example. Suppose you are told that a recent Cochrane review reported that, in patients with chronic pain, the number needed to treat (NNT) for clinical success with topical salicylates was 6 (95% CI = 4–13) compared with placebo. What additional information would you seek to help you decide whether to believe this estimate and how to apply it?

The most obvious questions might be the following: how many studies were pooled to get this estimate; how many patients did they include; and how wide were the CIs around the effect estimate? Were they randomized controlled trials? Did the studies have important limitations, such as lack of blinding or large or differential loss to follow-up in the compared groups? The questions thus far relate to GRADE categories of imprecision and risk of bias.

But there are also other important questions. Is there evidence that more studies of this treatment were conducted, but some were inaccessible to the reviewers? If so, how likely is it that the results of the review reflect the overall experience with this treatment? Did the trials have similar or widely varying results? Was the outcome measured at an appropriate time, or were the studies too short in duration to have much relevance? What part of the body was involved in the interventions (and thus, to what part of the body can we confidently apply these results)? These latter questions refer to the GRADE categories of publication bias, inconsistency, and indirectness. Without answers to (or at least information about) these questions, it is not possible to determine how much confidence to attach to the reported NNT and CIs.

GRADE identified its five categories—risk of bias, imprecision, inconsistency, indirectness, and publication bias—because they address nearly all issues that bear on the quality of evidence. For any given question, moreover, information about each of these categories is likely to be essential to judge whether the estimate is likely to be correct. These categories were arrived at through a case-based process by members of GRADE, who identified a broad range of issues and factors related to the assessment of the quality of studies. All potential factors were considered, and through an iterative process of discussion and review, concerns were scrutinized and solutions narrowed by consensus to these five categories.

GRADE's approach to quality implies that every systematic review should provide information about each of the categories (and any other pertinent issues in a particular case). Decision makers, whether they are guideline developers or clinicians, find it difficult to use a systematic review that does not provide this information. Good systematic reviews and clinical practice guidelines have commonly emphasized appraisal of the risk of bias (study limitations) using explicit criteria. Often, however, the focus has been on assessments across outcomes for each study rather than on each important outcome across studies. Assessment of other factors that determine how much confidence can be placed in estimates of effect has often been lacking. Before the adoption of GRADE, standards for reporting systematic reviews have not made clear how this information should be presented. GRADE provides a structure for systematic reviews and clinical practice guidelines to ensure they address the key questions that are pertinent to rating the quality of the evidence for all outcomes relevant to a particular question in a consistent systematic manner.

10. Conclusion

In closing, we caution against a mechanistic approach toward the application of the criteria for rating the quality of the evidence up or down. Although GRADE suggests the initial separate consideration of five categories of reasons for rating down the quality of evidence, and three categories for rating up, with a yes/no decision regarding rating up or down in each case, the final rating of overall evidence quality occurs in a continuum of confidence in the validity, precision, consistency, and applicability of the estimates. Fundamentally, the assessment of evidence quality is a subjective process, and GRADE should not be seen as obviating the need for or minimizing the importance of judgment or as suggesting that quality can be objectively determined.

As we repeatedly stress throughout this series, use of GRADE will not guarantee consistency in assessment, whether of the quality of evidence or of the strength of recommendations. There will be cases in which competent reviewers will have honest and legitimate disagreement about the interpretation of evidence. In such cases, the merit of GRADE is that it provides a framework that guides one through the critical components of this assessment and an approach to analysis and communication that encourages transparency and an explicit accounting of the judgments involved.

References

[1] Guyatt GH, Oxman AD, Kunz R, Vist GE, Brozek J, Norris S, et al. GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2011;64:383–94 [in this issue].

[2] Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist GE, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. J Clin Epidemiol 2011;64:395–400 [in this issue].

[3] Guyatt GH, Oxman AD, Vist GE, Kunz R, Brozek J, Alonso-Coello, et al. GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). J Clin Epidemiol 2011;64:407–15.

[4] Guyatt GH, Oxman AD, Montori V, Vist GE, Kunz R, Brozek J, et al. GRADE guidelines: 5. Rating the quality of evidence - publication bias. J Clin Epidemiol. In press.

[5] Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines: 6. Rating the quality of evidence—imprecision (random error). J Clin Epidemiol. In press.

[6] Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 7. Rating the quality of evidence—inconsistency. J Clin Epidemiol. In press.

[7] Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 8. Rating the quality of evidence—indirectness. J Clin Epidemiol. In press.

[8] Waldman RJ, Hall WN, McGee H, Van Amburg G. Aspirin as a risk factor in Reye's syndrome. JAMA 1982;247:3089–94.

[9] Starko KM, Ray CG, Dominguez LB, Stromberg WL, Woodall DF. Reye's syndrome and salicylate use. Pediatrics 1980;66:859–64.

[10] Halpin TJ, Holtzhauer FJ, Campbell RJ, Hall LJ, Correa-Villasenor A, Lanese R, et al. Reye's syndrome and medication use. JAMA 1982;248:687–91.

[11] Hurwitz ES, Barrett MJ, Bregman D, Gunn WJ, Pinsky P, Schonberger LB, et al. Public health service study of Reye's syndrome and medications: report of the main study. JAMA 1987;257:1905–11.

[12] Kearon C, Gent M, Hirsh J, Weitz J, Kovacs MJ, Anderson DR, et al. A comparison of three months of anticoagulation with extended anticoagulation for a first episode of idiopathic venous thromboembolism. N Engl J Med 1999;340:901.

[13] Campbell IA, Bentley DP, Prescott RJ, Routledge PA, Shetty HGM, Williamson IJ. Anticoagulation for three versus six months in patients with deep vein thrombosis or pulmonary embolism, or both: randomised trial. BMJ 2007;334:674.

[14] Kearon C, Ginsberg JS, Anderson DR, Kovacs MJ, Wells P, Julian JA, et al. Comparison of 1 month with 3 months of anticoagulation for a first episode of venous thromboembolism associated with a transient risk factor. J Thromb Haemost 2004;2:743–9.

[15] Agnelli G, Prandoni P, Santamaria MG, Bagatella P, Iorio A, Bazzan M, et al. Three months versus one year of oral anticoagulant therapy for idiopathic deep venous thrombosis. N Engl J Med 2001;345:165.

[16] Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336:924–6.