Volume 2
Issue 7
July 2019

BARBELL MEDICINE

WITH YOU FROM BENCH TO BEDSIDE

Monthly Research Review

Table of Contents

Author Information

Jordan Feigenbaum, MD

@jordan_barbellmedicine

info@barbellmedicine.com

Transgender Athletes in Sport: Solving A Wicked Problem With Science

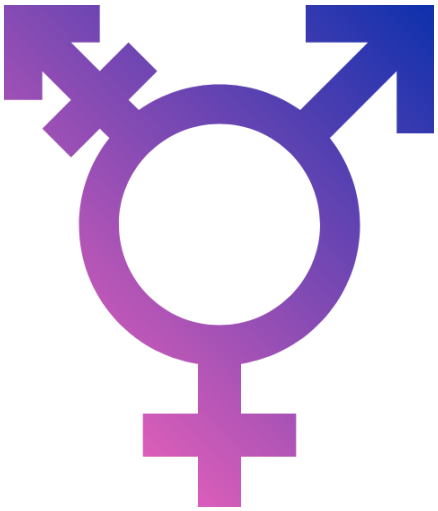
Transsexuals and competitive sports by Gooren et al. 2004.

Key Points:

1. Gender and sex are not the same thing. Biologically, sex is difficult to dichotomize into two binary divisions, as there are a number of different chromosome compositions consistent with either male or female sex. Effectively regulating sports divisions based on sex is extremly unlikely, as there are no singular or combinations of objective criteria that define male and female.

2. On average, the so-called “gender gap” between men and women’s performance in sport is estimated at 10% to 12%. This number increases to about 15% for sports requiring a high degree of upper body strength. However, the gap remains at 10-12% for sprint events in running, cycling, and swimming. Additionally, the gap is much narrower at approximately 5% for ultra-endurance sports. Sports like Olympic weightlifting and powerlifting report a much larger gap of ~40%, but these numbers appear to be confounded by other variables such as equity in participation and other sociocultural influences.

3. The present data on transgender individuals suggest that any competitive advantage maintained or acquired through the process of transitioning likely falls within an acceptable range given the precedent set in the 2015 Dutee Chand case. The Olympic Charter also encourages inclusion of individuals irrespective without discrimination based on race, color, sex, sexual orientation, language, religion, political or other opinion, national or social origin, property, birth, or other status. While rules of sport to ensure “fair-play” are arbitrary, depending on the interpretation of what this actually means, it is important for those in a position of power to continually review information and policies to ensure equity and equality in sports participation, to the extent that’s possible.



Introduction

The role of gender in sports dates back to 1890's and the birth of the modern Olympic movement. When asked about the inclusion of women in the Olympics, the founder of the International Olympic Committee (IOC) and father of the modern Olympic Games, Pierre de Coubertin, said, "*Women have but one task: that of crowning the winner with garlands.*"[Warner 2006] The Olympics of 1900 held in Paris, France allowed women to participate for the first time in just two sports: tennis and golf. Unsurprisingly, women's participation in the Summer Olympics was low, at about 2% of all participants until 1924, the final year Coubertin was president of the IOC.[IOC Factsheet]

Over the years, the IOC has gradually made strides towards including more competitive opportunities for women as well as leadership roles in sports administration. This has been a slow process, as women were only recently allowed to participate in Olympic Weightlifting at the 2000 Olympic games in Sydney, Australia.[IOC Factsheet] Additionally, a 2013 survey of over 1500 individuals who held leadership positions in international sports bodies revealed that only 13% were women, and 14 of the 56 executive committees included no women at all.[SportAccord Factsheet] Despite this, women's participation in sport has been rising, as demonstrated by increased presence at the highest level of sport. At the London Games of 2012 and Rio Games of 2016, women made up about 44% and 45% of all athletes, respectively.[IOC Factsheet]

As women's participation rates have steadily increased, a different question has emerged with respect to gender in sports: **How do we determine who is eligible for the women's division in sport?** Put another way, how do we define "woman" and "man" for the purposes of athletic competition?

This month, I'd like to take a deep dive into the issue of gender in sports. Specifically, we'll review the history of gender eligibility testing in sport, gender differences in athletic performance, transgender-specific considerations, and a smattering of related biological, social, and legal issues. Overall, the issue of gender in sports is a sort of wicked problem - meaning it is difficult to solve due to incomplete knowledge, contradictory information, and changing requirements or inputs that prevent a neat and tidy solution. In any case, I'll try my best to provide a representative view of the issue to help better inform you, the reader, to draw your own conclusions. Let's begin!

Gender Eligibility Testing

Humans are typically born with 46 chromosomes, which are DNA (Deoxyribonucleic Acid) molecules carrying genetic information known as their "*genotype*". The expression of this genetic information leads to the physical appearance and manifestations of a person known as their "*phenotype*". Individuals receive 23 chromosomes from each parent during the fertilization process. Additionally, there is one specialized set of chromosomes that determine sex, e.g. the "X" and the "Y" chromosomes.

While most women are genotypically 46 XX and most men are genotypically 46 XY, some individuals will be born with a single sex chromosome (45 X or 45 Y), known as *sex monosomies*, or even multiple sex chromosomes (47 XXX, 47 XYY, etc.), which are known as *sex polysomies*. Additionally, some phenotypic males are born with 46 XX via the inclusion of the small Sex-determining Region of the Y chromosome (SRY).[WHO Gender and Genetics] Similarly, some phenotypic females may be born 46 XY due to mutations in the Y chromosome that compromise its function. **Interestingly, the prevalence of 46 XY in the female athletic population is about 7 per 1000, which is 140 times greater than what is seen in the general population.**[Bermon 2014]

Is is clear that science does not view sex in a binary way based on chromosomes, as there are many different chromosomal makeups that can be consistent with either the "male" or "female" sex. At present the best definition of sex is put forth by the Universisty of California San Francisco (UCSF) Transgender Care and Treatment Guidelines:

*"**Sex**: Historically has referred to the sex assigned at birth, based on assessment of external genitalia, as well as chromosomes and gonads. In everyday language is often used interchangeably with gender, however there are differences, which become important in the context of transgender people."*[UCSF Transgender Care and Treatment Guidelines]

While the current IOC and International Association of Athletics Federation (IAAF) both use the term "sex testing" in technical policies for eligibility testing, we will see through the lens of history that these organizations often conflate gender and sex. For a definition of gender, let's again turn to the UCSF Transgender Care and Treatment Guidelines:

"Gender identity: A person's internal sense of self and how they fit into the world, from the perspective of gender.

Gender expression: The outward manner in which an individual expresses or displays their gender. This may include choices in clothing and hairstyle, or speech and mannerisms. Gender identity and gender expression may differ; for example a woman (transgender or non-transgender) may have an androgynous appearance, or a man

(transgender or non-transgender) may have a feminine form of self-expression."[UCSF Transgender Care and Treatment Guidelines]

In other words, sex and gender are **not** the same thing. Gender identity also has multiple lines of evidence suggesting it has a biological basis:

- A recent systematic review looked at studies involving individuals with Disorders (or Differences) of Sex Development (DSDs), which refers to anomalies of the sex chromosomes, the gonads, the reproductive ducts, and the genitalia. Two important studies looked at individuals who were 46 XY and had severe non-hormonal anatomic abnormalities of genital development including "penile agenesis, cloacal exstrophy, and penile ablation." For many years, the standard treatment for these patients included female gender assignment and surgical feminization, yet some would forgo these treatments and be raised as males. While none of the patients raised as males initiated a gender change to female, two studies reported that 22% and 50% of those raised as females ended up changing their gender identity to male. The authors conclude, "*Although the cohort sizes in these studies were small, the data provide the strongest evidence for the biologic underpinnings of gender identity.*"Saraswat 2015

- Twin studies also show a greater transgender concordance (40% compared to 0%), i.e. the presence of the same trait in both twins, among identical twins compared to fraternal twins.Heylens 2012 Identical twins occur when a single fertilized egg splits into two, whereas fraternal twins result when two different eggs are fertilized by two different sperm.

- Neuroimaging studies show different brain structures and neurophysiology in those with "gender-incongruence", i.e. those who have a gender identity that is different from their sex chromosomes.Luders 2012

- Prenatal androgen exposure has been shown to play a role in gender identity formation, though there is dispute over the magnitude of this effect.Wilson 1999

Taken together, the evidence suggests that there are biological inputs to both sex and gender, yet "biological sex" is a misnomer. Science has repeatedly shown us sex cannot be determined by any single characteristic or combination of characteristics. However, lines in the sand must somehow be drawn for sports that have gender divisions. Let's take a trip down memory lane to see how it's been done over time:

- **Pre-1966**: Reports and data on early gender eligibility testing are murky, however it appears that for registration at international events, e.g. the World Championships or Olympic Games, national federations relied on a medical certificate attached to the application form to validate the gender of a female athlete. In many cases, suspicions were raised as to the legitimacy of these documents and subsequently physical examinations were performed on women

athletes.[Ferguson 1991]

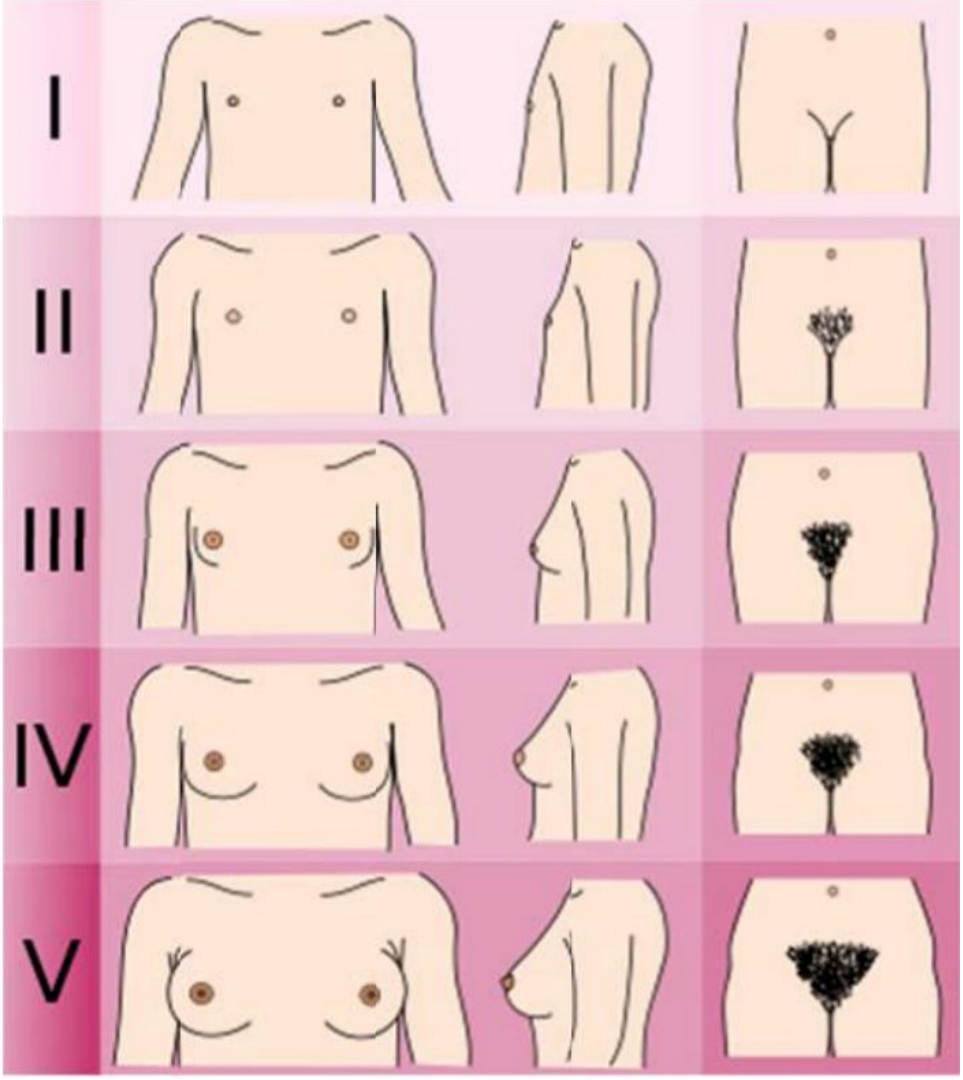
- **1966-1967**: Universal testing of all women was introduced at the 1966 European Athletics Championships in Budapest. According to multiple sources, female athletes were required to walk naked and undergo visual genital inspection by a panel of doctors to obtain eligibility to participate in competitive sport, which were termed “naked parades.”[Ferguson 1991, Heggie 2010]

- **1967-1996:** The IOC introduced the “buccal smear” test, which is comprised of a swab taken from inside the cheek that allowed doctors to identify the presence of multiple X chromosomes and thus, in theory, provide an objective basis for distinguishing between women and men. The test was found to be unreliable, as it falsely identified women with various genetic make-ups as men.[Genel 2005]

- **1996-2000:**, The IAAF and IOC decided to use a different genetic test that identified the sex-determining region of the Y chromosome to again, try to objectively distinguish female from male. The new tests were used in the 1996 Summer Olympics in Atlanta, Georgia where 8 of 3,387 female athletes failed them, although all were allowed to compete as women.[Heggie 2010]

- **2000-2011**: The IOC decided to abandon universal testing for female athletes in favor of a “suspicion-based” approach, where athletes suspected of being men competing in the women’s divisions would be tested. While, this move appeared to be successful initially, as no screenings were performed at the Winter Olympic Games of Salt Lake City in 2002, the Summer Olympic Games in Athens 2004, and Winter Olympic Games in Torino 2006, the suspicion approach ultimately led to officials subjecting athletes like India’s Dutee Chand and South Africa’s Caster Semenya to testing and public scrutiny. We’ll cover this in some detail below.

- **2011-2018:** In 2011, the IAAF put forth a new set of guidelines to help determine, once and for all, who was eligible for the female division in track and field sports. To be eligible, individuals had to be recognized by law as female or intersex, have testosterone levels less than 10 nmol/L for 6 months, which had to be maintained during periods where individuals wished to remain eligible. In cases where there was still uncertainty, the IAAF recommended a physical examination to look for “feminine” characteristics, such as lack of body hair, shape of breasts, and the presence of typical genital hair. To accomplish this, the IAAF used The Tanner-Whitehouse Scale seen below.[IAAF 2011 Hyperandrogenism Guideline]



- **2018-Present:** In April of 2018 the IAAF introduced new rules concerning eligibility for the women's divisions in track and field sports. Briefly, individuals have to be recognized by law as female or intersex, have a testosterone level of less than 5 nmol/L for 6 months prior to eligibility, and maintain this level for the duration of their eligibility period. All physical exam criteria were removed from the 2018 regulations.

Consider the case of Indian sprinter Dutee Chand, winner of two gold medals at the 2014 Asian Junior Athletics Championships in the 100m sprint and 4x400m relays. Due to her impressive performance and "masculine build", Chand had her testosterone levels tested by the Sports Authority of India.[Mohanty 2014] She failed that test and was no longer eligible to compete, as she had a testosterone level in excess of what was

allowed by the 2011 IAAF Hyperandrogenism Eligibility Regulations. Chand appealed to the Court of Arbitration for Sport (CAS), the Swiss-based organization that settles disputes within international sports recognized by the IOC. In July of 2015, the CAS ruled in her favor stating:

> *"According to the evidence reported by the IAAF, the competitive advantage that men have over women is approximately in the range of 10 to 12% while that the one enjoyed by hyperandrogenic athletes over other women would be between 1 and 3%. This advantage is not sufficient to justify a separation in the category of female athletes since many other factors such as nutrition, coaching, other genetic and biological variations have an impact on athletic performance."*[CAS 2015]

Chand was ultimately reinstated and ended up qualifying for the 2016 Summer Olympic Games in Rio in the 100m sprint event, though she did not advance past the qualifying heats. While the CAS found that Chand *maybe* possessed a 1-3% performance advantage due to her testosterone levels, *other* factors, such as nutrition, coaching, genetics, etc. also have the potential to impact performance - yet these are not formally regulated in sport.

Consider the case of Eero Antero Mäntyranta, the Finnish distance skier who won five Olympic medals (two gold, one silver, and two bronze) in the 1960s. In endurance events, the oxygen carrying capacity of the blood is an important biological factor in success. Hemoglobin, the protein inside red blood cells that binds oxygen for transport to active tissues, is produced by the bone marrow in response to erythropoietin or *EPO*, a hormone produced by the kidneys. Athletes who use EPO do so in order to increase their hemoglobin levels so that they have more capacity to carry oxygen to active tissues. With more oxygen carrying capacity, the athletes can work at higher intensities for longer periods of time.[Momaya 2015]

EPO works by binding to the EPO receptor on cells in the bone marrow. When EPO binds to the receptor, the cell starts making hemoglobin. There is also a portion of the EPO receptor that modulates this response, so that the cell doesn't make *too much* hemoglobin. With that in mind, what would happen if an individual had a mutation in the EPO receptor that resulted in a broken "stop" signal, and the cell never got the message that there was already enough hemoglobin being produced? You'd get **A LOT** more hemoglobin production, almost as if someone was using exogenous EPO as a performance enhancing drug.

Well, Mäntyranta (and much of his family) happens to have a mutation just like this. His EPO receptor gene has a small, single mutation that results in him having a hemoglobin level that's about 65% higher than everyone else *not* taking EPO.[Epstein 2013]

Interestingly enough, there has never been any official regulation on hemoglobin levels for the purposes of athletic eligibility. Similarly, there's never been any type of eligibility

criteria with respect to any other genetic condition, such as gigantism (e.g. Romania's basketball player Gheorghe Mureșan) or Marfan's syndrome (e.g. US volleyball player Flo Hyman) in professional sports. Nor do we use testosterone levels to determine male eligibility in sport, rather it is used solely for the purpose of determining eligibility in women's sport.

All told, I do not envy those who are in positions of power with respect to rule-making in sport. It seems incredibly difficult, if not impossible, to determine who is eligible for the women's division in sport in both an equitable and equal manner. This is a wicked problem indeed! Let's see if the science on the gender gap in performance can lend a hand in approaching this problem.

The Gender Gap in Performance

One way to objectively measure the difference between men and women's performances at the elite level would be to compare the top 10 performers's results from each gender across a wide variety of different events. A 2010 paper by Thibault *et al.* did just that, with the authors reporting the following findings:

> "*A stabilization of the gender gap in world records is observed after 1983, at a mean difference of 10.0% ± 2.94 between men and women for all events. The gender gap ranges from 5.5% (800-m freestyle, swimming) to 18.8% (long jump). The mean gap is 10.7% for running performances, 17.5% for jumps, 8.9% for swimming races, 7.0% for speed skating and 8.7% in cycling. The top ten performers' analysis reveals a similar gender gap trend with a stabilization in 1982 at 11.7%.*"Thibault 2010

The 10-12% difference in performance between different genders has been corroborated in a number of other reviews, with exceptions in events where upper-body power is a major contributor and ultra-endurance swimming. In the former, the gender gap is larger, at more than 15%, and in the latter, the gap is now less than 5%.Sandbakk 2017

However, there's some nuance here, as different results emerge when we consider individual sports. Consider that the gender gap in Olympic weightlifting is at 36.8%, which is *far* greater than the average difference of 10-12% cited above.Thibault 2010 A similar gap is seen when comparing international-level powerlifters, as a recent review of IPF competitors by the USAPL showed approximately a 40% difference in strength performance.Personal Communication Hunt 2019

One way to look at this is to suggest that some biological factor such as testosterone, presence of a Y-chromosome, growth hormone levels, or some combination of biological factors have a greater opportunity to widen the gap in barbell sports compared to other sports where the onus on absolute strength is not as significant. There is evidence for all of these biological factors, which is reviewed in a 2018 paper by Handselman *et al.*Handsleman 2018

On the other hand, one might observe the larger gender gap in barbell sports and wonder if there may be other, non-biological factors involved. For example, women weren't allowed to compete in Olympic weightlifting until 1998 at any level and the 2000 Summer Games in Sydney was the first Olympics where that women participated in weightlifting.[Thibault 2010] Also, about 50 world records in women's weightlifting are set by Chinese athletes, which speaks to both the success of the Chinese recruitment and development systems as well as the relative competitiveness compared to the men's divisions.[Guillaume 2009] Competitiveness is driven by a number of factors, one of which is participation rates. Larger pools of competitors tend to increase the probability that there are multiple standouts, which drives competitiveness (and records) up. If we consider that it wasn't until the 2012 Summer Olympic Games in London that every country's delegation included a female competitor combined with the short history of women's participation in barbell sports, we might suggest that psychosocial and cultural factors are involved in the larger gender gap seen in barbell sports.[Capranica 2013]

Based on present evidence, I think it's reasonable to suggest that all of these factors contribute to the ranges of gender gap observed in sport. Sports with a long history of women's inclusion and large athlete pools, such as running, cycling, and swimming demonstrate the 10-12% performance gap, even for sprint events like the 100 m race and all 100 m swim events.[Sandbakk 2017] Conversely, sports that require more upper-body muscle strength, e.g. canoeing, kayaking, and skiing, tend to have a gender gap of around 15%.[Sandbakk 2017] Finally, women actually tend to outperform *men* in open-water ultra-endurance swimming as the records for both the 32-km 'Catalina Channel Swim' and 46-km 'Manhattan Island Marathon Swim' are held by women.[Knechtle 2014]

Despite this, the current data on powerlifting and Olympic weightlifting suggest a nearly 3-fold increase in the performance gap between genders, which I think may be inflated by non-biological factors such as participation and coaching access, sociocultural factors, inequality at every level of sport, and more. Nevertheless, there are certainly real biological differences between men and women that we should address to lay the groundwork for this month's review:

- **Height**: On average, men tend to be 10-15 cm taller than women.[Janssen 2000] One way to predict this in pediatric patients is to subtract 13 cm from the father's height and average that value with the mother's height for girls. For boys, 13 cm is added from the mother's height and averaged with the father's height. The 13 cm represents the average difference in height of men and women.[Tanner 1970] It is thought that testosterone release during puberty is responsible for the average greater height in men.[Courant 2010]

- **Muscle Mass:** On average, men tend to carry 30-40% more lean body mass than women.[Janssen 2000] However, height and weight influence muscle carrying capacity for an individual regardless of sex or gender.[Heymsfield 2015,Gallagher 1997] Interestingly, when correcting for height and total body weight, the gap narrows

to about a 10% difference in lean body mass between men and women across the entire lifespan.[Kirchengast 2010] Additionally, a number of other factors contribute to the regulation of adult muscle mass, including genetics, race, adiposity, training, diet, and birth order.[Heymsfield 2015] Importantly, data from twin studies suggests about ~50% to 60% of the difference in muscle mass and strength is due to genetic factors.[Silventoinen 2003, Beunen 2004, Silventoinen 2008]

- **Body Fat:** Whereas men tend to have 10% greater lean body mass than women of the same height and weight, women have about 10% greater body fat than men. This gap also tends to be preserved over the entire lifespan.[Kirchengast 2010]

- **Strength/Power:** As described above, men tend to be taller, heavier, and carry more muscle and less fat than women. However, when we normalize existing strength data for fat-free mass, which is an approximation of skeletal muscle mass, there are no sex differences in strength and power.[van den Tillaar 2004] For example, Slawinski *et al*. looked at world-class sprinters' peak acceleration after 1 second. After normalizing for body mass, men and women had approximately the same peak acceleration power.[Slawinski 2017] Additionally, power output performances for the Wingate test, which is used to express the anaerobic power during cycling, showed no differences between men and women when normalized for fat-free mass.[Perez-Gomez 2008] Finally, men and women appear to respond similarly to resistance training, with no gender-specific responses noted after 10-weeks of resistance training.[Gentil 2016]

- **Hemoglobin:** As described above in the Eero Mäntyranta story, hemoglobin is responsible for carrying oxygen around to active tissues. On average, hemoglobin levels are 12% higher in males than females, which is thought to be due to the effect of testosterone.[Shahani 2009]

- **Testosterone:** In healthy males, testosterone has a wide range of normal, e.g. 250 -1000 ng/dL, with some slight variation on the upper and lower limits depending on the laboratory and age of the individual.[Morales 2015] Women also have a range of normal, e.g. 20-60 ng/dL, with additional variation based on the lab and phase of the menstrual cycle.[Longcope 1986] Experimental evidence reflects a nearly 5-fold difference in average testosterone levels between men and women.[Clark 2018] With respect to testosterone, the CAS said the following in the Chand case,"*The Panel has accepted that testosterone is the best indicator of performance difference between male and female athletes. However, the evidence does not equal the level of testosterone in females with a percentage increase in competitive advantage."*[CAS 2015]

Taken together, we now have a good idea of the size and components of the performance gap between the genders. What we *don't* know, however, is what sort of performance gap there would be, if any, when an individual changes gender. In other words, does a trans-woman maintain the performance advantages after transitioning?

One way to assess this is to look at women with *hyperandrogenism*, i.e. women with testosterone levels that approach (and in some cases exceed) the lower end of the normal range for men. Hyperandrogenism can occur a number of different ways including Disorders (or Differences) of Sexual Development (DSD), Polycystic Ovarian Syndrome (PCOS), Congenital Adrenal Hyperplasia (CAH), and more.

To reiterate, in the case of Dutee Chand discussed earlier, the CAS weighed in on this issue using a panel of medical experts:

> *"According to the evidence reported by the IAAF, the competitive advantage that men have over women is approximately in the range of 10 to 12% while that the one enjoyed by hyperandrogenic athletes over other women would be between 1 and 3%. This advantage is not sufficient to justify a separation in the category of female athletes since many other factors such as nutrition, coaching, other genetic and biological variations have an impact on athletic performance."* CAS 2015

This has interesting implications for transgender women interested in participating in sports, as it would seem that a competitive advantage of less than 3% would grant them eligibility to participate in the women's division. Conversely, a competitive advantage of 10-12% or greater would require a different solution for inclusion into sport. Let's now turn our focus towards transgender-specific considerations.

Transgender-Specific Considerations

Let's start out by defining what it means to be transgender. Again, the UCSF Transgender Care and Treatment Guidelines provide insight:

- **Transgender***: A person those gender identity differs from the sex that was assigned at birth. May be abbreviated to trans. A transgender man is someone with a male gender identity and a female birth assigned sex; a transgender woman is someone with a female gender identity and a male birth assigned sex. A non-transgender person may be referred to as cisgender (cis=same side in Latin).* UCSF Transgender Care and Treatment Guidelines

At present, there are a number of different policies put forth by international organizations. Most of them have similar requirements to the IOC, which is listed in full below:

- **International Olympic Committee (IOC)**:
  - Those who transition from female to male are eligible to compete in the male category without restriction.
  - Those who transition from male to female are eligible to compete in the female category under the following conditions:

    - The athlete has declared that her gender identity is female. The declaration cannot be changed, for sporting purposes, for a minimum of four years.
    - The athlete must demonstrate that her total testosterone level in serum has been below 10 nmol/L for at least 12 months prior to her first competition (with the requirement for any longer period to be based on a confidential case-by-case evaluation, considering whether or not 12 months is a sufficient length of time to minimize any advantage in women's competition).
    - The athlete's total testosterone level in serum must remain below 10 nmol/L throughout the period of desired eligibility to compete in the female category.[IOC Consensus 2015]

The other organizations listed below tend to have similar, if not identical procedures to the IOC, however there are some notable differences that I'll briefly list:

- **National Collegiate Athletic Association (NCAA):** The NCAA differs from the IOC in that it requires Female-to-Male (FTM) trans athletes to obtain a medical exception for treatment with testosterone for diagnosed Gender Identity Disorder or gender dysphoria and/or Transsexualism. These individuals are no longer eligible to compete on a women's team without changing that team status to a mixed team. While there are no set cut-offs for testosterone levels published for either FTM or Male-to-Female (MTF) trans athletes, MTF athletes must have documentation of testosterone suppression for one year prior to being eligible to compete on a women's team. That being said, MTF athletes can compete prior to the full one year of suppression, but the women's team designation will be changed to "mixed team" status until completion of one year of suppression.[NCAA Inclusion of Transgender Student-Athletes 2011]

- **CrossFit:** CrossFit follows the IOC's lead in all policies outside of requiring legal documentation of their gender, e.g. state ID or driver's license, and having a provision for athlete's whose total testosterone levels are greater than the IOC's cut-off of 10 nmol/L. These athletes must demonstrate that they have a genetic condition that affects the bioavailability of their serum testosterone, such as hyperandrogenism (e.g., PCOS in rare cases), androgen insensitivity syndrome or 5-alpha reductase deficiency to remain eligible for participation in the women's division.[CrossFit Rulebook 2019]

- **United States Weightlifting (USAW):** The USAW uses the IOC's policy to the letter, though they do require more extensive documentation and hormonal suppression. Athletes wishing to change their gender identify need to have a written letter requesting eligibility to participate as a different gender, a confirmation of gender identity from a medical professional, and legal documentation of their identified gender. With respect to hormonal suppression in MTF athletes, the USAW requires two years of hormone therapy, instead of just

one as listed by the IOC to minimize gender related competitive advantages. For FTM athletes desiring hormone therapy, documentation must be provided certifying the athlete's therapy is being monitored by a medical doctor and is not being used as a way to enhance athletic advantage.[USAW Transgender Policy]

- **United States of America Powerlifting (USAPL) and International Powerlifting Federation (IPF)**: While both organizations recognize the IOC's policy on transgender participation in sports, there is only limited opportunity for trans athletes to participate. Specifically, transgender women are not allowed to compete in the women's division, but are allowed to compete in the men's division. If they are taking spironolactone, a common anti-androgen used in combination with synthetic estrogen for transitioning individuals, they must obtain a therapeutic use exemption (TUE) in order to participate and comply with existing WADA regulations. Transgender men, or any athlete taking testosterone, may not participate in USAPL or IPF competitions, with or without a Therapeutic Use Exemption form (no TUEs are available in USAPL or IPF for testosterone therapy). However, transgender men may participate in the men's division if they are not taking testosterone or any other prohibited substances.[USAPL Transgender Policy]

At present, it appears that registering for an event sanctioned by CrossFit, USAW, or IAAF requires transgender women to maintain testosterone levels of less than 10 nmol/L for a period of either 1 year (CrossFit and IAAF) or 2 years (USAW) to be eligible for the women's division. Conversely, no objective limits have been placed on transgender men for testosterone levels, despite receiving exogenous testosterone support.

Interestingly, none of the organizations listed here define sex or gender for the purposes of participation eligibility. Said differently, nobody has definitively said who *is* and who *isn't* a female (or male). What's more, is that none of the organizations list testing policies and procedures for sex (or gender) verification, which follows the IOC's decision from 2000 to eliminate routine sex testing in individuals. When discussing this issue with the USAPL's TUE Chair, Dr. Kris Hunt, he lamented that there isn't any way to do this objectively and even if there was, the costs would likely be prohibitive.[Personal Communication Hunt 2019]

It appears there is a concern for "gender fraud", specifically the fear that men will attempt to impersonate women athletes in pursuit of athletic success and accompanying accolades. However, there are precisely zero examples of a man posing as a woman in elite athletics in at least the past 50 years.[Pielke 2016] Conversely, there are examples of women posing as men in order to participate in sport, e.g. Kathrine Switzer using only her initials to run the 1967 Boston Marathon that was then closed to women.[Switzer 2007] It would seem that gender fraud in sport is an example of an "academic urban legend" —a falsehood repeated so often by academics (and others) that it comes to be accepted as true.[Rekdal 2014]

If gender fraud isn't really "a thing", then what's with all the regulation? It would seem as though ensuring a level playing field while fostering participation in sport are the major goals. However, these policies arguably limit eligibility for participation under the guise of fairness, i.e. inclusion in the absence of advantage. Yet some may interpret them as discriminatory, especially when considering the fourth and sixth Fundamental Principles of Olympism listed in the Olympic Charter:

Fundamental Principle Number Four
The practice of sport is a human right. Every individual must have the possibility of practising sport, without discrimination of any kind and in the Olympic spirit, which requires mutual understanding with a spirit of friendship, solidarity and fair play.

Fundamental Principle Number Six
The enjoyment of the rights and freedoms set forth in this Olympic Charter shall be secured without discrimination of any kind, such as race, colour, sex, sexual orientation, language, religion, political or other opinion, national or social origin, property, birth or other status. Olympic Charter 2018

For transgender athletes, the question of eligibility is complex and raises scientific, social, and legal issues. While it would be inappropriate to look only to science for a clear-cut answer, gathering more information might allow us to have a more informed opinion and offer better solutions to this wicked problem. In this month's research review, let's take a look at one of the seminal studies on changes observed in transgender individuals before and after their transitions.

Purpose

The main purpose of the article being reviewed and this month's BMR is to answer the question, "*What, if any, objective performance differences persist in transgender athletes after the process of transitioning?*"

Additionally, my BMR contribution this month aims to provide much-needed background information in understanding the topic of transgender athletes participation in sport, i.e. terminology, current rules and guidelines, and related scientific data.

Finally, while it is inappropriate for me, an individual with no rule-making power in the world of sport, to extrapolate these findings to determine what constitutes "fair competition", as there will always be an element of arbitrariness when it comes to making the rules of sport, I do feel compelled to offer some possible solutions to the problem. Nevertheless, I expect these recommendations to be interpreted differently, accepted, or criticized depending on the levels of arbitrariness an athletic organization (or individual) is willing to accept.

Subjects

This retrospective study looked at data from 17 female-to-male (FTM) and 19 male-to-female (MTF) transgender individuals.

The FTM individuals had an average age of 24.6 years (range 16-33) and were healthy, as assessed by their medical histories, physical examinations, and laboratory measurements. These individuals received 250 mg of testosterone every two weeks delivered via injection. Data was collected prior to and after 1 and 3 years of testosterone administration. Prior to hormonal administration, the individuals had the following characteristics:

- Average weight 60.7 kg
- Average height 167.1 cm
- Average BMI 21.7 kg/m^2
- Average Muscle Area 238.8 cm^2
- Serum Testosterone 1.6 nmol/L

The MTF individuals had an average age of 26.6 years (range 18-37) and were healthy, as assessed by their medical histories, physical examinations, and laboratory measurements. Data was collected prior to and after 1 and 3 years of hormonal suppression using the anti-androgen cyproterone acetate, which is stopped after the first 2 years, and the synthetic estrogen ethinyl estradiol, taken indefinitely. Both medications were administered orally. Prior to hormonal suppression, the individuals had the following characteristics:

- Average weight 66.1 kg
- Average height 177.8 cm
- Average BMI 20.8 kg/m^2
- Average Muscle Area 306.9 cm^2
- Serum Testosterone 21.5 nmol/L

Of note, cyproterone acetate is a medication brought to market as an oral contraceptive for women (in combination with estrogen) in Europe in the 1970's and later Canada in the 1980s. It was never approved for use in the United States. With respect to its anti-androgenic effects, its mechanism of action is interesting, as it inhibits Luteinizing Hormone (LH) release from the pituitary gland. In response to this reduction in pituitary LH secretion, the testes produce less testosterone. Cyproterone also blocks the binding of dihydrotestosterone (DHT), which is a potent metabolite of testosterone that has a greater affinity for the androgen receptor when compared to testosterone.[Neumann 1999]

In the United States, spironolactone (trade name Aldactone) is the most widely used drug for transgender women (e.g. MTF individuals). Mechanistically, spironolactone inhibits binding of testosterone to the androgen receptor (e.g. competitive inhibition) and prevents the testes from making testosterone.

Spironolactone can be prescribed off label to treat hormonal acne (e.g. acne vulgaris) in adult women.[Zaenglein 2016] Additionally, spironolactone is commonly prescribed as an adjunctive treatment for *hirsutism* (facial hair growth) associated with polycystic ovarian syndrome (PCOS) if initial treatment with oral contraceptives produce suboptimal results.[Martin 2018] With that being said, spironolactone is currently on the World Anti-Doping Agency's prohibited substance list, as it can be used as both a masking agent and a diuretic.[WADA 2019] However, athletes can have their physician submit a request for a therapeutic use exemption or "TUE" in order to use spironolactone outside of competition, e.g. the athletes must stop using the medication 5 days prior to competition.

Interestingly, the USAPL has performed a retrospective internal data review on female powerlifters who received a TUE for spironolactone to characterize the effect of the medication on strength performance. Based on the data of ~25 female competitors, it appears that there is a slight reduction in strength performance (e.g. powerlifting total) in the short-term and no long-term performance effect.[Personal Communication Hunt 2019]

Methods

Anthropometric data was measured as follows:

- Body weight was measured to the nearest 0.1 kg using a digital scale
- Height was measured to the nearest 0.1 cm with subjects wearing only underwear
- Magnetic resonance imaging (MRI) was used to determine muscle cross sectional area at the level of the thigh using the same anatomical markers for all subjects.
- Body fat measurements were assessed via both skin caliper testing and bioelectrical impedance (BIA)

Laboratory data was measured as follows:

- All blood samples were obtained after an overnight fast.
- Standardized radioimmunoassays were used to measure levels of testosterone and estrogen levels.
- Standardized immunoradiometric assays were used to measure sex hormone-binding globulin and IGF-1 levels.
- Hemoglobin levels were measured from blood samples using an automated cell counter

Food intake was assessed with the use of the Dutch EPIC food frequency questionnaire.

All anthropometric and laboratory data were measured prior to any intervention and then remeasured at 12 months. Some data was also remeasured at 36 months post intervention, e.g. testosterone levels and muscle cross sectional area.

Findings

Data after 1 year of testosterone suppression in MTF are reflected in Figure 1.

Male to Female (MTF) Individuals (n=19)		
Metrics	**Before Testosterone Suppression**	**After Testosterone Suppression**
Height (cm)	177.8± 7.9	177.8± 7.9
Body weight (kg)	66.1± 11.7	69.9± 11.3
BMI (kg/m^2)	20.8± 2.6	22.0± 2.7
Muscle area (cm^2)	306.9± 46.5	277.8± 37.0
Serum Testosterone (nmol/l)	21.5± 5.8	1.0± 0.0
Hemoglobin (mmol/l)	9.3± 0.7	8.0± 0.7
IGF-1	38.0± 10.0	14.0± 8.0

Figure 1: Male to Female transgender individuals prior to testosterone suppression and 1 year after testosterone suppression.

We can see that body weight increased by 3.8 kg (8.3 lbs) while muscle cross-sectional area dropped by nearly 40 cm^2, which represents an increase in body fat and loss of muscle mass. Hemoglobin and IGF-1 also decreased, which likely corresponds to the reduction of testosterone levels to castration levels. Unpublished data from these same authors show a further reduction in muscle cross sectional area to 271cm^2 after three years of hormone suppression, however no weight or BMI data is available from the three year follow up.

Data after 1 year of testosterone administration in FTM are reflected in Figure 2.

Female to Male (FTM) Individuals (n=17)		
Metrics	**Before Testosterone Administration**	**After Testosterone Administration**
Height (cm)	167.1± 7.8	167.1± 7.8
Body weight (kg)	60.7± 11.8	63.4± 11.4
BMI (kg/m^2)	21.7± 3.5	22.6± 3.0
Muscle area (cm^2)	238.8± 31.1	285.3± 35.6
Serum Testosterone (nmol/l)	1.6± 0.6	30.8± 11.4
Hemoglobin (mmol/l)	8.2± 0.7	9.4± 0.8
IGF-1	26.0± 12.0	36.0± 14.0

Figure 2: Female to male transgender individuals prior to testosterone administration and 1 year after testosterone administration.

We can see that body weight increased by 3.4kg (7.5lbs) while muscle cross-sectional area increased by nearly 47cm^2, which represents a decrease in body fat and increase in muscle mass - the exact opposite of the data seen in MTF individuals. Unpublished data from these same authors show a slight reduction in muscle cross sectional area to 280cm^2 after three years of hormone suppression, suggesting a likely plateau effect of bringing testosterone levels up to that of a "eugonadal" male.

Additionally, quantitative comparisons for muscle cross sectional area at the thigh were also made between the subjects prior to transitioning and after hormonal suppression or administration had taken place for one year, as shown in Figure 3.

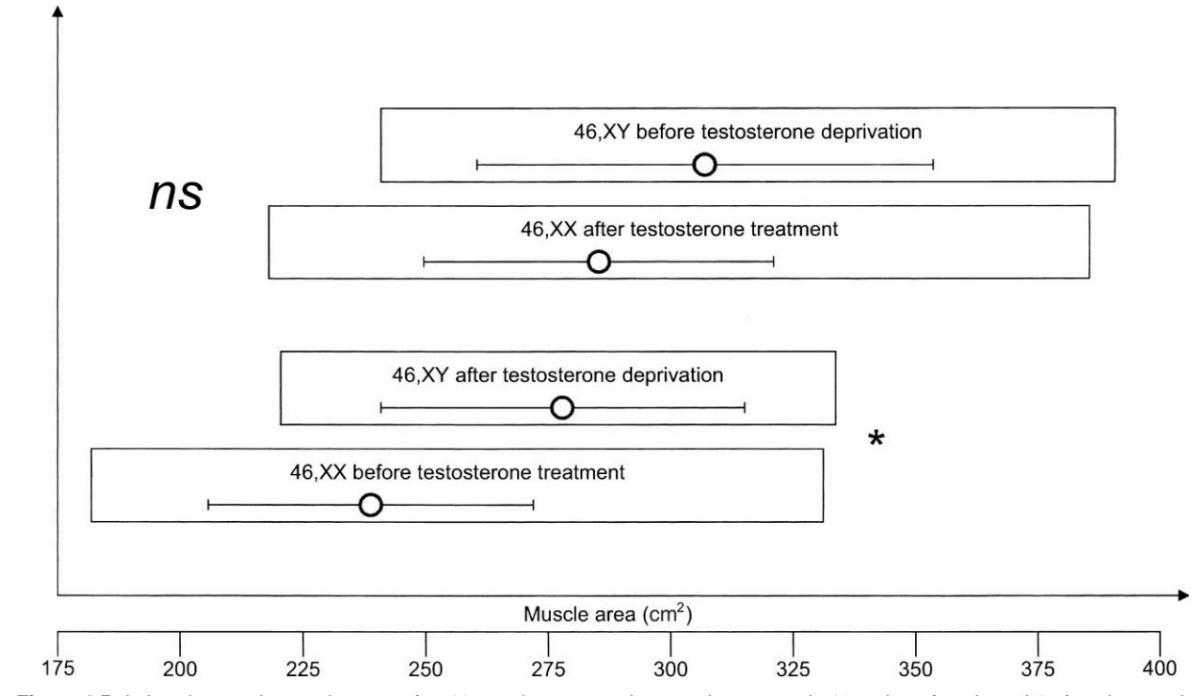


Figure 3: Muscle cross sectional area comparisons between subjects pre and post hormonal administration (FTM) or suppression (MTF), respectively.

The casual reader will notice the significant overlap in muscle cross sectional areas between all of the subjects. The mean cross sectional areas, which are labeled with an empty circle, suggest that pre-transition men (46, XY) tend to have greater muscle cross sectional area than pre-transition women (46, XX). However, if we take a closer look we can also see the considerable overlap between trans women (46, XY after testosterone deprivation) and cis women (46 XX before testosterone treatment).

While the average amount of muscle cross sectional area in trans women is about 39 cm^2 higher than cis women, the overlap in range of muscle cross sectional areas is remarkable. The authors stumbled upon this as well, remarking:

> ***"After 1 year of androgen deprivation, mean muscle area in MTF had decreased significantly but remained significantly greater than in FTM before testosterone treatment, though with an overlap that was almost complete when androgen-deprived MTF were again compared to untreated FTM."***

The authors also found that height strongly predicted muscle cross-sectional area. Subjects with an XY genotype were on average 10.7 cm taller, which strongly correlated to muscle cross sectional area after correction for the effect of genotype, e.g. XY or XX.

Why does this article matter?

This article, published by a leading expert in endocrinology and transgender medicine, is one of the only published studies on the musculoskeletal effects of gender transition. It provides useful insight into the expected outcomes from the standard treatments in both MTF and FTM transgender individuals. When combined with other existing literature on similar topics, we can start to get a clearer picture of this issue and become more informed.

For example, Joanna Harper is a medical physicist, distance runner, and advisor to the IOC on matters of gender and sport. She was intstrumental in developing the 2015 IOC transgender policy and advises on matters of gender and sport. She is also a transgender woman. In 2015, she published a study looking at race times of 8 transgender women before and after transition. She reported:

> *"Collectively, the eight runners were much slower in the female gender; slow enough, in fact, that their age-graded performances were almost identical to their male AGs. Two of the runners had higher average AGs in male gender than in female gender, while one runner had higher female AGs than male ones. The changes in the age grades of these runners mirrored changes in their training habits."*[Harper 2015]

*AG= age grading, a method of comparing the performance of athletes for all ages and both sexes. For running events, the athlete's race time (RT) is compared to the fastest time ever run by a person of that age and sex, or the age standard (AS)

A 2009 cross-sectional study by T'Sjoen *et al.* looked at 50 individuals who transitioned from male to female with anti-androgen medication and sex reassigment surgery. After 12 months of HRT, they reported a "strikingly lower muscle mass" and additionally, increased fat mass compared to pretreatment values and reference controls. However, no absolute values for change in muscle or fat mass were reported.[T'Sjoen 2009] Taken together with the Gooren *et al.* study discussed in this month's research review, it's clear that muscle mass decreases markedly in MTF transgender individuals.

Despite the overlap in muscle cross-sectional area between cis-women (pre-transition) and trans-women, it also appears that trans-women *do* carry more muscle mass than cis-women at 1 year post-transition. While Gooren *et al.* concluded that transgender male individuals are likely to be able to compete without an athletic advantage 1-year post-cross-sex hormone treatment, they did not test this objectively. In other words, we can't say if trans-women are stronger after transition than cis-women, on average.

Unfortunately, that's all we have for quantitative data on transgender athletes. All of the other literature discusses policy, fairness, and other qualitative data. In the 2011 NCAA Policy on Transgender Student-Athletes, Eric Vilain, a physician, professor, and Director of the Center for Gender-Based Biology and Chief Medical Genetics Department of Pediatrics at UCLA, says:

> *"Androgen deprivation and cross sex hormone treatment in male-to-female transsexuals reduces muscle mass; accordingly, one year of hormone therapy is an appropriate transitional time before a male-to-female student-athlete competes on a women's team."* NCAA Inclusion of Transgender Student-Athletes

And yet, there is still uncertainty regarding the "fairness" aspect. Reflecting back to the Dutee Chand case that was settled by the CAS, we have precedent that competitive advantage of 3% or less due secondary to testosterone exposure is considered acceptable. *CAS 2015* Perhaps if we had robust retrospective data looking at transgender athletes across a wide variety of sports we could characterize what advantage, if any, transgender individuals have. At present however, we focus on testosterone, which I think is a bit reductionist. You see, sports are divided into women's and men's categories, and not "low testosterone" and "high testosterone" divisions - or any other biological variable for that matter. The CAS added the following in the case of Dutee Chand:

> *"The Panel found that the Athlete has established that it is prima facie discriminatory to require female athletes to undergo testing for levels of endogenous testosterone when male athletes do not. In addition, it is not in dispute that the Hyperandrogenism Regulations place restrictions on the eligibility of certain female athletes to compete on the basis of a natural physical characteristic (namely the amount of testosterone that their bodies produce naturally) and are therefore prima facie discriminatory on that basis too."* *CAS 2015*

In 2012, Dr. Katrina Karkazis and Dr. Rebecca Jordan-Young, two influential bioethicists published a landmark paper, *Out of Bounds, A Critique of the New Policies on Hyperandrogenism in Elite Female Athletes* in the American Journal of Bioethics. In this article, the authors review the controversies over eligibility policies and conclude with respect to testosterone:

> *"Despite the many assumptions about the relationship between testosterone and athletic advantage, there is no evidence showing that successful athletes have higher testosterone levels than less successful athletes."* Karkazis 2012

Overall, my view is that testosterone levels don't matter nearly as much as people think in the context of athletic performance when it comes to cis-gendered athletic competition, and therefore *should not be used to determine eligibility for participation in women's sports*. With that said, I am less confident about the importance of *previous* exposure to endogenous testosterone (in the case of MTF individuals) or current exposure to exogenous testosterone (in the case of FTM individuals) among transgender athletes. For the sake of argument let's assume that in the case of transgender women, previous exposure to testosterone does confer an advantage and is associated with greater athletic prowess. My response -- which may be different than yours -- is "So what?"

In a thought experiment, let's suppose the previous paragraph was rewritten with "height" or "body weight", as there are numerous examples of cis-women with exceptional height and/or weight, typically reserved for cis-men, who are very successful athletically. Should they be ineligible to participate in the women's division because of their more typically "male" height or weight? While endocrinology in sports is fascinating, my view is that it's no more relevant to gender eligibility policy than any other naturally occurring human characteristic, regardless of gender. Drawing arbitrary lines in the sand have the potential to hurt people, however, as demonstrated by Dutee Chand and transgender individuals who are ineligible to compete in their sport of choice based on current policies.

As it stands however, I have no rule-making power or authority within the athletic community to effect change outside of conveying complex topics through the lens of science. However, I do think it would be unfair of me to point out all these issues without suggesting potential solutions. In my view, there are three viable solutions to this problem:

1. **Assess Gender like Nationality**: Given that science is having an incredibly difficult time trying to figure out what makes a woman a woman and a man a man, perhaps we should stop trying to use science as the end-all, be-all. Instead, we could use a process similar to how international sports organizations regulate nationality. Nationality is also a touchy subject, as an "anything goes" policy might result in a situation where athletes are traded between competing nations. Additionally, a person's nationality is not binary and, like gender and sex, cannot be defined satisfactorily by science. Can you imagine trying to describe someone's "American-ness"?

   A parent organization like the IOC could require individuals to register as a particular gender at a qualifying event, e.g. a national or

internationally-sanctioned competition, which would take place across all age divisions as well. Then, once an athlete becomes an adult, he or she would sign an affidavit testifying to his or her gender. Policies would also have to be in place to cover if an adult wishes to change their gender, similar to the current IOC rules about declaring one's gender no more frequently than once every four years. This policy has the benefit of being consistent with how science currently views gender and sex without limiting the inclusion of individuals based on arbitrary consensus decisions.[Harper 2018]

2. **Open an Additional Division:** Another idea -- though one that does not address the inherent problems of regulating gender and sex in sport -- could involve the addition of a third division in sports for transgender individuals to participate. The LGBT International Powerlifting Championships introduced the "MX" division for inclusion of transgender and intersex athletes in 2018.[Gay Games Website] The benefits of this idea are two-fold: 1) Transgender, intersex, and non-binary individuals get to compete in their sport of choice immediately; and 2) Objective data can be collected about performances relative to cis-gendered athletes in order to further the discussion about inclusion of transgendered and intersex individuals in conventional sports' two gender divisions, if desired.

   Joanna Harper, who was discussed earlier, has lent some support to this idea.[Harper 2018] Additionally, the TUE Chair for the USAPL, Dr. Kris Hunt, as well as long-time USAPL president, Larry Maile, both agreed they were in favor of this idea being proposed at the next committee meeting.[Personal Communication Hunt 2019] On the other hand, implementation of this idea poses two substantial drawbacks: 1) It fails to address the inherent problems with regulating gender and sex based on science alone; and 2) It limits participation of trans and intersex individuals in the gender division they identify with.

3. **Keep Things the Same:** Practically speaking, yes, we could just keep things the same, and I do not say this tongue-in-cheek. Rather, sports - and the rules that govern them - are just collections of arbitrary consensus statements that have been collated into a rulebook. With respect to powerlifting, there are over 50 separate federations within the United States, 14 of which have international affiliates, and each having their own collection of arbitrary consensus statements called "rules". My personal view is that sports organizations should continue to review their rules and rule-making policies, in order to fulfill their own mission statements and advance the interests of their stakeholders. I can envision a number of situations where revising rules and rule-making policies based on the information presented here may not support those aims.

   Conversely, it is also clear to me that sports "punch above their weight class"

with respect to their impacts on society. Consider that the NFL has the largest revenue of any sporting organization in the United States, earning $12 billion in 2015.[Kaplan 2015] To put this into perspective, Apple earned about $12 billion every two weeks in the first quarter of 2016.[Apple 2016] The entire 2013 revenue of all major US sports (including NASCAR and the NCAA) was about $23 billion, or 0.15% of the total US gross domestic product. Despite this, we give sports a ton of attention, namely via television and social media. In 2015, ESPN was the most-watched channel on television and had 18 of the top 27 most-watched shows.[ESPN 2015] On instagram, Cristiano Ronaldo has the highest number of followers at 173.5 million.[Fuentes 2019]That's close to 30 million more than Dwayne "The Rock" Johnson! Thus, I think sports - and by extension, sporting organizations - have a responsibility to uphold when it comes to furthering the narrative of equity and equality in both sports and society.[Schultz 2012] While we *can* sit back and do nothing, I hope we don't.

Thank you for reading this month's research review. I know this was a bit different than my normal contribution, as I used a study that was 15 years-old, reviewed a ton of other data, and even cited an article from *Seventeen* magazine. With that said, I am hopeful that this article helped shed some light on the topic of transgender participation in sport. Who knows? Maybe in 5 years there will be a ton of new data and I can write an update article. In closing, I'll leave you with a quote from Arne Ljungqvist, former high jumper, vice chairman of WADA, and chairman of the IOC's Medical Commission:

> **"***Ultimately, the number of transsexual athletes who can successfully compete in open international events is likely to be small, in accord with the estimated incidence of gender dysphoria of one in about every 12,000 men and one in about every 30 000 women. Furthermore, the recommended process for gender reassignment as described is rather arduous.*
>
> *Finally, individuals who fulfil these criteria will likely be at a relatively advanced age athletically, at least in many sports, though there are notable exceptions—eg, in golf, such as Mianne Bagger who recently qualified and has been competing on the Ladies European Tour after competing in the Swedish Telia Tour in 2004. Inevitably there will be transgendered athletes, such as Renee Richards, who will be competitive at a high level, but most will probably wish to compete only at a masters level or at local and regional events.*
>
> *The recommendations of the International Olympic Committee are being adopted by various sports governing bodies, such as the US Golf Association and Great Britain's Ladies Golf Union. We believe that they provide a fair and equitable standard."*[Ljungqvist 2005]

Thanks for reading.

References

1. IOC Consensus Meeting on Sex Reassignment and Hyperandrogenism November 2015. https://bit.ly/29IcVCv
2. Fuentes, Tamara. The 10 Most Followed Celebrities on Instagram. *Seventeen.* April. 2019. https://bit.ly/2XkZumU
3. Shahani S, Braga-Basaria M, Maggio M, Basaria S. Androgens and erythropoiesis: past and present. J Endocrinol Invest. 2009;32(8):704–716
4. Van den Tillaar R, Ettema G. Effect of body size and gender in overarm throwing performance. Eur J Appl Physiol 2004;91(4):413-8

5. Gender and Genetics. *World Health Organization.* February 23, 2012. https://www.who.int/genomics/gender/en/index1.html
6. NCAA Inclusion of Transgender Student-Athletes 2011. http://www.ncaapublications.com/p-4335-ncaa-inclusion-of-transgender-student-athletes.aspx
7. Zaenglein AL, Pathy AL, Schlosser BJ, et al. Guidelines of care for the management of acne vulgaris. J Am Acad Dermatol. 2016; 74(5):945-973.e33.
8. Neumann, Friedmund (1996). "Pharmacology of Cyproterone Acetate — A Short Review". *Antiandrogens in Prostate Cancer*. pp. 31–44.
9. Bermon S, Garnier PY, Hirschberg AL, et al. Serum androgen levels in elite female athletes. J Clin Endocrinol Metab. 2014;99(11):4328-4335
10. Factsheet: Women in Leadership Positions within SportAccord Members. Sport-Accord.com. https://bit.ly/30bJcK3
11. Schultz, Jaime. New Standards, Same Refrain: The IAAF's Regulations on Hyperandrogenism, The American Journal of Bioethics, 2012.12:7, 32-33.
12. Court of Arbitration for Sport (July 2015). *CAS 2014/A/3759 Dutee Chand v. Athletics Federation of India (AFI) & The International Association of Athletics Federations (IAAF)* (PDF). Court of Arbitration for Sport.
13. Momaya, A., Fawal, M., & Estes, R. (2015). *Performance-Enhancing Substances in Sports: A Review of the Literature. Sports Medicine, 45(4), 517–531.*
14. Heymsfield SB, Gonzalez MC, Lu J, Jia G, Zheng J. Skeletal muscle mass and quality: evolution of modern measurement concepts in the context of sarcopenia. *Proc Nutr Soc. 2015 Nov; 74(4):355-66.*
15. Silventoinen K, Sammalisto S, Perola M, Boomsma DI, Cornes BK, Davis C, Dunkel L, De Lange M, Harris JR, Hjelmborg JV, Luciano M, Martin NG, Mortensen J, Nistico L, Pedersen NL, Skytthe A, Spector TD, Stazi MA, Willemsen G, Kaprio J. Heritability of adult body height: a comparative study of twin cohorts in eight countries. Twin Res. 2003;6(5):399–408.
16. Beunen G, Thomis M. Gene powered? Where to go from heritability (h2) in muscle strength and power? Exerc Sport Sci Rev. 2004;32(4):148–154.
17. Silventoinen K, Magnusson PK, Tynelius P, Kaprio J, Rasmussen F. Heritability of body size and muscle strength in young adulthood: a study of one million Swedish men. Genet Epidemiol. 2008;32(4):341–349.

18. Harper, Joanna. “Race Times for Transgender Athletes.” (2015). https://bit.ly/2LyNQh0
19. Courant F, Aksglaede L, Antignac JP, Monteau F, Sorensen K, Andersson AM, Skakkebaek NE, Juul A, Bizec BL. Assessment of circulating sex steroid levels in prepubertal and pubertal boys and girls by a novel ultrasensitive gas chromatography-tandem mass spectrometry method. J Clin Endocrinol Metab. 2010;95(1):82–92.
20. Janssen, Ian, Heymsfield, Steven B, Wang, ZiMian, & Ross, Robert. Skeletal muscle mass and distribution in 468 men and women aged 18–88 yr. Journal of Applied Physiology 2000 89:1, 81-88
21. Kaplan, Daniel. NFL projecting revenue increase of $1B over 2014. Sports Business Journal, March. 2015. https://bit.ly/2Xj1gox
22. Olympic Charter 2018. https://bit.ly/2Yslngv
23. Slawinski J, Termoz N, Rabita G, Guilhem G, Dorel S, Morin JB, Samozino P. How 100- m event analyses improve our understanding of world-class men's and women's sprint performance. Scand J Med Sci Sports 2017;27(1):45-54.
24. Clark, R. V., Wald, J. A., Swerdloff, R. S., Wang, C., Wu, F. C. W., Bowers, L. D., & Matsumoto, A. M. Large Divergence in Testosterone Concentrations between Men and Women: Frame of Reference for Elite Athletes in Sex-Specific Competition in Sports, a Narrative Review. Clinical Endocrinology. 2018.
25. Ferguson-Smith MA, Ferris EA. Gender verification in sport: The need for change? Br J Sports Med 1991; 25:17-20.
26. IAAF, Regulations Governing Eligibility of Females with Hyperandrogenism. 2018. https://bit.ly/2vOgXGH
27. IAAF, Regulations Governing Eligibility of Females with Hyperandrogenism. 2011.https://bit.ly/2J7BUkJ
28. Heggie V. Testing sex and gender in sports; reinventing, reimagining and reconstructing histories. Endeavour. 2010;34(4):157–163.
29. Kirchengast, Sylvia.Gender Differences in Body Composition from Childhood to Old Age: An Evolutionary Point of View, Journal of Life Sciences, 2010. 2:1, 1-10.
30. Harper, J., Martinez-Patino, M.-J., Pigozzi, F., & Pitsiladis, Y. (2018). Implications of a Third Gender for Elite Sports. Current Sports Medicine Reports, 17(2).
31. T’Sjoen, G., Weyers, S., Taes, Y., Lapauw, B., Toye, K., Goemaere, S., & Kaufman, J.-M. (2009). Prevalence of Low Bone Mass in Relation to Estrogen Treatment and Body Composition in Male-to-Female Transsexual Persons. Journal of Clinical Densitometry, 12(3), 306–313.
32. Karkazis K, Jordan-Young R, Davis G, Camporesi S. Out of bounds? A critique of the new policies on hyperandrogenism in elite female athletes [published correction appears in Am J Bioeth. 2012;12(10):56]. Am J Bioeth. 2012;12(7):3–16.
33. 2019 CrossFit Rulebook https://games.crossfit.com/rules/open
34. 2019 USAPL Transgender Participation Policy https://www.usapowerlifting.com/transgender-participation-policy/

35. Genel, Myron, and Arne Ljungqvist. “Essay: Gender Verification of Female Athletes.” *The Lancet*, vol. 366, Dec. 2005.
36. Kathrine Switzer, “The Girl Who Started It All,” Runner’s World, May 2007, http://kathrineswitzer.com/site/wp-content/uploads/SwitzerStory_RunnersWorld.pdf.
37. Gallagher, D., Visser, M., De Meersman, R. E., Sepúlveda, D., Baumgartner, R. N., Pierson, R. N.,Heymsfield, S. B. (1997). Appendicular skeletal muscle mass: effects of age, gender, and ethnicity. Journal of Applied Physiology, 83(1), 229–239.
38. Factsheet: Women in Leadership Positions within SportAccord Members,” Sport-Accord.com. https://bit.ly/2YyF3iY
39. UCSF Transgender Care, Department of Family and Community Medicine, University of California San Francisco. Guidelines for the Primary and Gender-Affirming Care of Transgender and Gender Nonbinary People; 2nd edition. Deutsch MB, ed. June 2016.
40. Perez-Gomez J, Rodriguez GV, Ara I, Olmedillas H, Chavarren J, Gonzalez-Henriquez JJ, Dorado C, Calbet JA. Role of muscle mass on sprint performance: gender differences? Eur J Appl Physiol 2008;102(6):685-94.
41. LGBT INTERNATIONAL POWERLIFTING ANNOUNCES NEW “MX” CATEGORY, Jan 2018.https://gaygames.org/latest-news/5656204
42. Mohanty, D.I am who I am: Dutee Chand. The Indian Express, September, 2014. https://indianexpress.com/article/sports/sport-others/big-picture-i-am-who-i-am/
43. Saraswat A, Weinand JD, Safer JD. Evidence supporting the biologic nature of gender identity. Endocr Pract 2015; 21:199.
44. Heylens G, De Cuypere G, Zucker KJ, et al. Gender identity disorder in twins: a review of the case report literature. J Sex Med 2012; 9:751.
45. Luders E, Sánchez FJ, Tosun D, et al. Increased Cortical Thickness in Male-to-Female Transsexualism. J Behav Brain Sci 2012; 2:357.
46. Martin KA, Anderson RR, Chang RJ, et al. Evaluation and Treatment of Hirsutism in Premenopausal Women: An Endocrine Society Clinical Practice Guideline. J Clin Endocrinol Metab 2018; 103:1233.
47. Tanner JM, Goldstein H, Whitehouse RH. Standards for children's height at ages 2-9 years allowing for heights of parents. Arch Dis Child 1970; 45:755.
48. Sandbakk, Ø., Solli, G. S., & Holmberg, H.-C. (2018). Sex Differences in World-Record Performance: The Influence of Sport Discipline and Competition Duration. International Journal of Sports Physiology and Performance, 13(1), 2–8.
49. Guillaume M., El Helou N., Nassif H., Berthelot G., Len S., et al. Success in Developing Regions: World Records Evolution through a Geopolitical Prism. PLoS ONE, 2009.4(10)
50. Handelsman DJ, Hirschberg AL, Bermon S. Circulating Testosterone as the Hormonal Basis of Sex Differences in Athletic Performance. *Endocr Rev*. ;39(5):803–829.
51. Capranica, Laura & Piacentini, Maria Francesca & Halson, Shona & Myburgh, Kathryn & Ogasawara, Etsuko & Millard-Stafford, Mindy. (2013). The Gender

Gap in Sport Performance: Equity Influences Equality. International journal of sports physiology and performance. 8. 99-103.
52. Morales A, Bebb RA, Manjoo P, et al. Diagnosis and management of testosterone deficiency syndrome in men: clinical practice guideline. CMAJ 2015; 187:1369.
53. Pielke, Roger. The Edge: The War against Cheating and Corruption in the Cutthroat World of Elite Sports (p. 173). Roaring Forties Press. 2016.Kindle Edition.
54. WADA Prohibited List .2019. https://bit.ly/2RDGM3s
55. Hunt, K. (2019, June). Phone interview with USAPL Medical Director, Kris Hunt, MD.
56. Gentil P, Steele J, Pereira MC, Castanheira RP, Paoli A, Bottaro M. Comparison of upper body strength gains between men and women after 10 weeks of resistance training.PeerJ. 2016;4:e1627, Feb, 2016.
57. Longcope C. Adrenal and gonadal androgen secretion in normal females. Clin Endocrinol Metab 1986; 15:213.
58. Epstein, David. The Sports Gene: Inside the Science of Extraordinary Athletic Performance (p. 267). Penguin Publishing Group. Kindle Edition.
59. Knechtle B, Rosemann T, Lepers R, Rust CA. Women outperform men in ultradistance swimming: the Manhattan Island Marathon Swim from 1983 to 2013. Int J Sports Physiol Perform 2014;9(6):913-24.
60. IOC, “Factsheet: Women in the Olympic Movement,” Olympic.org, updated January 2016, https://bit.ly/2J6MBUM
61. Redkal,Ole Bjorn, “Academic Urban Legends,” *Social Studies of Science 44, no.4(2014):638-654.*
62. Thibault V, Guillaume M, Berthelot G, et al. Women and Men in Sport Performance: The Gender Gap has not Evolved since 1983. *J Sports Sci Med*. 2010;9(2):214–223, June, 2010.

Author Information

Austin Baraki, MD

@austin_barbellmedicine

draustin@barbellmedicine.com

Is there a Ceiling Effect for Training Volume in Muscle Hypertrophy and Strength?

Evidence of a Ceiling Effect for Training Volume in Muscle Hypertrophy and Strength in Trained Men – Less is More? by Barbalho *et al.* 2019

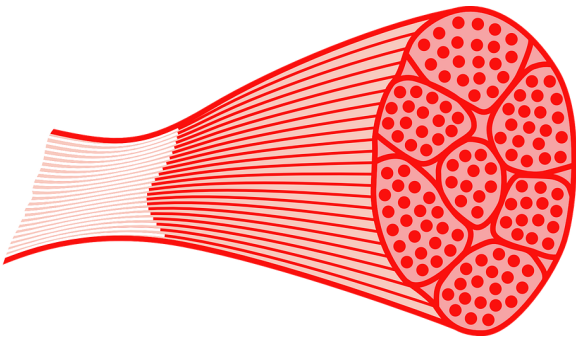
Key Points:

1. This study recruited forty young, healthy men into a 24-week study where they trained three days per week, randomized to different training volumes (5, 10, 15, or 20 total sets to failure per session). Day 1 involved three pressing exercises, day 2 involved three "pulling"/"rowing" exercises, and day 3 involved three lower body exercises.

2. All groups had significant improvements in all variables. However, the 5- and 10-set groups showed greater improvements in 10RM strength compared to 15- and 20-set groups. With respect to muscle thickness, all groups appeared to plateau at around 12 weeks, with little improvement -- and some regression -- observed beyond that, without significant differences detected between groups.

3. In contrast to the authors' conclusions regarding an "optimal weekly set volume", this design of this study instead suggests a maximum *per-session* volume of 5-10 sets per muscle group, on average, when sets are taken to failure. The overall weight of evidence still points towards a dose-response relationship between weekly training volume and strength/hypertrophy outcomes into the range of 10-20 sets per muscle group per week, particularly when distributed across a higher weekly frequency. Data on the effects of *very* high weekly set volumes are mixed. We also have little evidence on how these relationships and thresholds might change with submaximal training (i.e., *not* to failure).

Introduction



There has been extensive ongoing debate in the resistance training literature in recent years regarding the determinants of strength and hypertrophy outcomes and the extent of the relationship (if any) between the two. Taber 2019 Loenneke 2019

Numerous recent resistance training studies have examined the role of intensity Nóbrega 2018 Lasevicius 2018, frequency Schoenfeld 2019, rest periods McKendry 2016 Schoenfeld 2016, and training volume Schoenfeld 2017 Ralston 2017 on strength and hypertrophy.

One area of particular interest is *how much* training volume is necessary to achieve optimal outcomes. While some researchers have found evidence of a substantial dose-response effect up to very high training volumes, others have found conflicting results with stagnation or even regression at higher doses (specific references will be discussed in the discussion section of this piece). The authors of the current study have suggested evidence of a "ceiling effect" for training volume, referring to a maximum threshold above which no further gains result from further increases in the dose of training volume applied.

As with any field of study, there will be conflicting data and controversy, and we must be cautious regarding the strength of our conclusions and base them on the overall body of literature rather than any single study. But given this conflict, we will take a closer look at this paper to see how it fits in to our understanding.

Purpose

The authors' stated purpose was to "*compare the effects of different resistance training volumes on muscle performance and hypertrophy in trained men.*"

Subjects

Subjects included forty young men who were at least 18 years old (average age approximately 24.5 years) and cleared to train by a physician. This number of subjects was calculated based on the number needed to show an effect size (ES) of 0.5 with a statistical power of 80%. Effect sizes are a method of reporting the *magnitude* of difference, either between groups or within a group across the study. This helps to give us a more useful idea of "significance" than the standard *p*-value calculation for statistical significance (to learn more about effect sizes, see here).

Subjects were required to have participated in un-interrupted resistance training at least three times per week for the prior 3 years. Minimum strength standards were also required for recruitment: subjects needed to be able to perform at least a bodyweight x 10RM bench press, and at least a 1.5 times bodyweight x 10RM leg press. For example, a potential male subject weighing 185 lbs would need to be able to complete 10 reps on the bench press at 185 lbs and 10 reps on the leg press at 277.5 lbs to qualify for inclusion.

Methods

Participants were divided into groups of 10 and randomized to four different training protocols. Each muscle group was trained once per week at different volumes: 5 sets per week (G5), 10 sets per week (G10), 15 sets per week (G15), or 20 sets per week (G20) (see Fig. 1). However, on examination of the weekly training schedule, note that all sets for a particular muscle group were performed on the same day (i.e., training frequency per group was once per week).

Figure 1 – Weekly Training Schedule & Volume Distribution By Training Group

Monday	Wednesday	Friday	G5	G10	G15	G20
Barbell Bench Press	Lat Pulldown	45° Leg Press	2	4	5	7
Barbell Incline Bench press	Cable Row	Barbell Squat	2	4	5	7
Military Press	Upright Barbell Row	Stiff-Leg Deadlift	1	2	5	6

Training duration was 24 weeks in length using a non-linear periodization approach. The 24 weeks were divided into six 4-week blocks, as shown in Fig 2.

Figure 2 – Periodization Scheme

Week	Rep Range	Rest Interval
1, 5, 9, 13, 17, 21	12-15RM	30-60 seconds
2, 6, 10, 14, 18, 22	4-6RM	3-4 minutes
3, 7, 11, 15, 19, 23	10-12RM	1-2 minutes
4, 8, 12, 16, 20, 24	6-8RM	2-3 minutes

Training was performed in an in-person supervised setting. Lifts were instructed to be performed with a 2-second concentric and a 2-second eccentric, and all sets were performed to *momentary failure*. Momentary failure is defined as "*the point where, despite the greatest effort, a person is unable to meet and overcome the demands of the exercise causing an involuntary set end point*". Steele 2017 This standardization was instituted in order to control for as many variables as possible other than total volume.

Participants were also instructed to maintain their usual diets, and no further control was implemented over nutrition habits.

Strength testing was performed at baseline, mid-way through, and after the end of the study. This involved working up to a 10RM on the bench press, lat pulldown, 45 degree leg press, and stiff-leg deadlift, distributed across three consecutive days.

Muscle thickness measurements were obtained at baseline, mid-way through, and after the end of the study for the right biceps brachii, triceps brachii, pectoralis major, quadriceps femoris, and gluteus maximus. The measurements were obtained using ultrasound at standardized anatomic landmarks, using standardized technique, at the same time of day, and 3-5 days after the last training session. The ultrasound was operated by a single experienced technician who was blinded to study group allocation.

Given that there were multiple study groups with outcome measures repeated across three time points (0, 12, and 24 weeks), repeated measures Analysis of Variance (ANOVA) was performed to determine the effects of any time x group interactions.

Findings

Absolute changes in muscle thickness per muscle group (in mm) are presented from pre- to mid-intervention (Fig. 3) and pre- to post-intervention (Fig. 4).

Fig. 3 - Changes in muscle thickness pre- to mid-intervention (in mm, mean & 95% CI range)

	G5	G10	G15	G20
Biceps	3.7 +/- 0.6	3.8 +/- 0.6	2.7 +/- 0.7	2.3 +/- 0.7
Triceps	4.1 +/- 0.7	4.1 +/- 0.7	3.4 +/- 0.7	2.8 +/- 0.8
Pec major	4.9 +/- 0.8	4.8 +/- 0.7	3.7 +/- 0.8	2.9 +/- 0.9
Quadriceps	6.2 +/- 1.0	6.0 +/- 1.0	4.4 +/- 1.0	4.0 +/- 1.1
Gluteus	4.2 +/- 0.7	4.2 +/- 0.7	3.2 +/- 0.7	2.6 +/- 0.7

Fig. 4 - Changes in muscle thickness pre- to post-intervention (in mm, mean & 95% CI range)				
	G5	G10	G15	G20
Biceps	4.5 +/- 0.7	3.9 +/- 0.7	2.1 +/- 0.7	0.9 +/- 0.8
Triceps	5.1 +/- 0.8	4.5 +/- 0.8	2.9 +/- 0.9	1.3 +/- 1.0
Pec major	6.0 +/- 0.9	5.3 +/- 0.9	3.0 +/- 1.0	1.0 +/- 1.1
Quadriceps	8.0 +/- 1.1	6.6 +/- 1.1	3.7 +/- 1.2	1.8 +/- 1.3
Gluteus	5.2 +/- 0.8	4.5 +/- 0.8	2.8 +/- 0.8	1.0 +/- 0.9

Absolute changes in strength (in kg) from pre- to mid-intervention (Fig. 5) and pre- to post-intervention (Fig. 6).

Fig. 5 - Changes in strength pre- to mid-intervention (in kg, mean & 95% CI range)				
	G5	G10	G15	G20
Bench	16.3 +/- 2.0	15.9 +/- 2.0	7.7 +/- 2.2	4.4 +/- 2.3
Pulldown	12.8 +/- 1.7	12.8 +/- 1.7	6.3 +/- 1.8	3.9 +/- 2.0
Leg Press	29.2 +/- 3.8	29.7 +/- 3.8	14.0 +/- 4.0	8.3 +/- 3.7
SLDL	14.8 +/- 1.9	14.9 +/- 1.9	7.1 +/- 2.0	

Fig. 6 - Changes in strength pre- to post-intervention (in kg, mean & 95% CI range)				
	G5	G10	G15	G20
Bench	24.1 +/- 3.1	23.6 +/- 3.1	11.3 +/- 3.2	6.5 +/- 3.5
Pulldown	19.0 +/- 2.7	18.9 +/- 2.7	9.3 +/- 2.8	5.6 +/- 3.0
Leg Press	43.6 +/- 5.8	44.3 +/- 5.8	20.7 +/- 6.1	12.1 +/- 6.4
SLDL	22.0 +/- 2.9	22.2 +/- 2.9	10.4 +/- 3.0	

The authors found that all groups had significant improvements in all variables. However, the 5- and 10-set groups showed greater improvements in 10RM strength compared to 15- and 20-set groups.

With respect to muscle thickness, all groups appeared to plateau at around 12 weeks, with little improvement -- and some regression -- observed beyond that. The observed muscle thickness differences between groups across time did not reach statistical significance (i.e., no significant time x group interactions were noted), although the post hoc comparisons suggested the lower volume groups may have had better outcomes.

Percent changes are presented in graphical form in Fig 7 and 8.

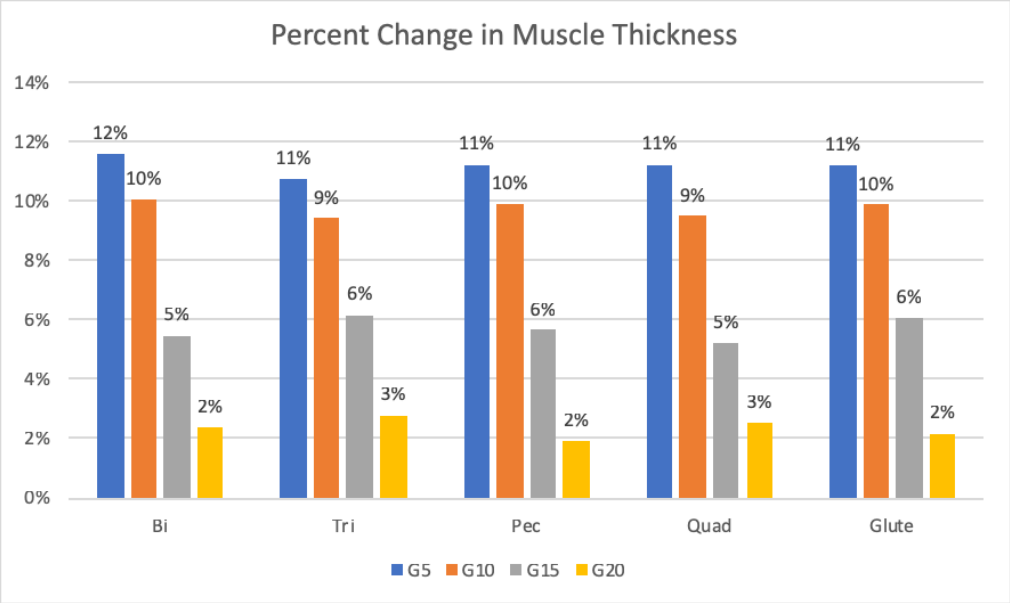


Fig. 7 - Percent change in muscle thickness by subgroup *note that these differences between groups were not statistically significant

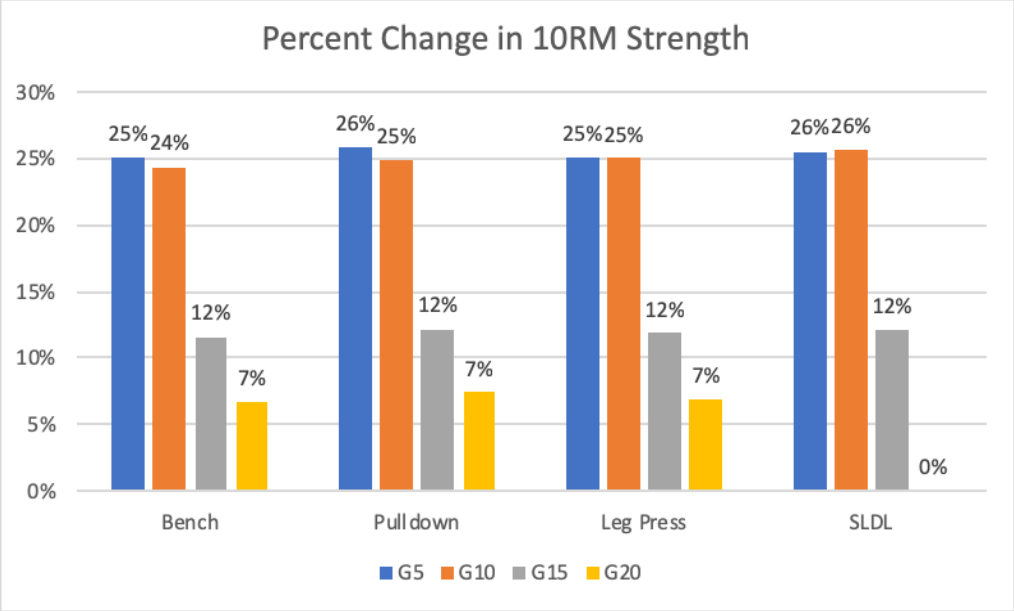


Fig. 8 - Percent change in 10-rep max strength by subgroup

Why does this article matter?

This paper is a particularly interesting one that replicates another study by the same group using the same training protocol among young women, where these authors originally proposed such a “ceiling” effect of training volume. Barbalho 2019

This describes a situation where increasing training volume fails to induce further gains, a finding that seems to conflict with other data suggesting a *dose-response relationship* of weekly training volume on hypertrophy outcomes. In contrast, a *dose-response relationship* describes a phenomenon by which increasing the dose of stimulus generates a larger effect, though it is important to note that this relationship does not necessarily need to be *linear* in nature (i.e., "diminishing returns" still represents a dose-response relationship as long as *some* continued positive effect is observed). This conflict makes Barbalho *et al*'s data interesting and worthy of examination to see how it might fit into our overall understanding.

The authors recruited forty young, healthy men into a 24-week study where they trained three days per week, randomized to different "doses" of training volume as described above. Perhaps the most significant aspect of the program to point out is that each muscle group was trained once per week.

In other words: day 1 involved pressing exercises performed for a total of 5, 10, 15, or 20 sets to failure, day 2 involved "pulling"/"rowing" exercises performed for 5, 10, 15, or 20 total sets to failure, and day 3 involved lower body exercises performed for 5, 10, 15, or 20 total sets to failure. All sets were performed under supervision of trainers to ensure that they were performed to momentary failure.

While the general study methodology was strong, the specific programming approach used in this paper adds some nuance to the interpretation of these results in the context of the broader literature, as well as its generalizability to the average gym population.

Rep ranges and rest periods were periodized as shown in Fig. 2 above. Of note, most of the available evidence on rest periods shows that longer rest periods (about 2 minutes or greater) tends to generate better strength and hypertrophy outcomes than shorter rest periods. McKendry 2016 Schoenfeld 2016

Furthermore, we also have evidence showing that training to failure does ***not*** produce more hypertrophy than stopping a few reps short of failure, but ***does*** induce substantially greater fatigue (which can limit subsequent motor unit recruitment) and prolongs neuromuscular recovery. Nóbrega 2018 Martorelli 2017 Pareja-Blanco 2018 Moran-Navarro 2017 The differences in accumulated fatigue between groups (e.g., between performing 5 sets to failure per session vs. 20 sets to failure per session) could plausibly have affected the outcomes on mid- and post-intervention 10RM strength measurements, as well as impaired hypertrophy outcomes over the course of the study.

Of note, there was no control exerted over subjects' dietary habits in this study; they were simply instructed to continue their usual nutritional intake, and were asked throughout the study period about whether they made any substantial changes to their intake. Post-intervention body mass measurements were not provided, and strength or hypertrophy outcomes may have been different if the training protocol were paired with a nutrition intervention as well.

This brings us to the primary question at hand in this study: the authors claim that these data support an *optimal weekly volume of 5-10 sets per muscle group* when taken to failure. However, we feel that this is an inappropriate conclusion to draw from these data given the study design. Considering the once-weekly frequency of exposure, we view the results as providing greater support for a maximum ***per-session*** volume of 5-10 sets per muscle group, on average, when sets are taken to failure.

From a practical standpoint, most experienced lifters would probably not be surprised to learn that a program of 20 sets to failure on a single exercise or muscle group per session might not produce optimal outcomes. With that said, let's take a look at how this fits in to the overall body of evidence on the matter.

The Context

We have evidence from animal models suggesting that the muscle protein synthetic (MPS) response to training maximizes and plateaus beyond a certain acute dose of training volume. Ogasawara 2017 We also have evidence that MPS measurements predict longer-term hypertrophy outcomes, assuming measurement duration is sufficient (~6 hours for protein ingestion, >24 hours for exercise) and individuals are sufficiently trained that exercise-induced muscle damage is attenuated. Damas 2016 Brook 2015

Furthermore, we know that among trained individuals the time course of the *total* training-induced muscle protein synthetic response is shorter, and that the *myofibrillar* protein synthetic response is blunted. Damas 2015 We view this as a sort of training-induced "anabolic resistance" that arises as an adaptive response to being more trained.

In the context of dietary protein as the anabolic stimulus driving MPS, we have evidence suggesting a dose-response effect up to a maximum threshold (with inter-individual variation depending upon the degree of anabolic sensitivity/resistance present), beyond which we see no further increases in MPS rates and progressive increases in amino acid oxidation. Moore 2009 Witard 2014 Moore 2015

In the context of physical training as the anabolic stimulus, it is plausible that there might be a similar maximum *single-session* dose of stimulus to which an individual can mount an adaptive response (with a degree of inter-individual variation). Beyond this threshold, less adaptation or even regression might occur. A potential reason for this might be the effects of muscle damage and the associated acute inflammation limiting the adaptive response, although we don't know for sure.

It is also plausible that there might be a maximum total *weekly* dose of stimulus to which an individual can mount an adaptive response (with a degree of inter-individual variation), although given a session-maximum limit this may be influenced by the frequency across which the training volume is distributed. This concept has been put forth by several authors including James Krieger as an "inverted U" hypothesis.

The Studies

For example, in a 6-month training study by Radaelli *et al*. subjects were randomized to perform 1, 3, or 5 sets per exercise using 9 exercises per session that trained varied muscle groups, all distributed across three sessions per week. Radaelli 2015 A dose-response effect was seen with respect to strength and hypertrophy outcomes up to *very high* training volumes. This finding was replicated in a shorter 8-week training study by Schoenfeld et al. using 1, 3, or 5 sets per exercise in 7 exercises per session, distributed across 3 sessions per week. Schoenfeld 2019 It should be noted that both of these studies have come with their own share of controversy in the interpretation and significance of their results, and a detailed analysis of each of these papers could match the length of the present article.

While we do tend to see a dose-response effect from low to moderate weekly training volumes, several other papers have shown that weekly set ranges of 9-18 tend to match (and sometimes *outperform*) weekly set ranges of 20-40. Paulsen 2003 Baker 2013 Rønnestad 2007 Ostrowski 1997 Heaselgrave 2019 This suggests that beyond a certain point, this dose-response effect for hypertrophy tends to "fall off" and may reverse altogether if the training becomes more than the athlete can tolerate and adapt to.

It should also be noted that even when the total number of sets are matched at three sets per exercise, moderate-load training for sets of 8-12 reps appears to produce greater hypertrophy than heavy-load training for sets of 2-4 reps. Schoenfeld 2016 This finding has added some nuance to discussions about how best to "measure" training volume. One method is to count the number of "hard" sets (in which full motor unit recruitment is achieved), although in light of these findings it has been suggested that this method should be limited to the number of sets of at least 6 repetitions that are taken to within at least 3 reps of failure (i.e., RPE 7). Baz-Valle J 2018

To summarize, meta-analytic data on total weekly training volume with respect to hypertrophy also finds evidence of a graded dose-response for <5, 5-9, and 10+ weekly sets with respect to muscle hypertrophy. Schoenfeld 2017 Similar findings have been demonstrated for strength outcomes as well. Ralston 2017

In contrast, meta-analytic data examining the role of training *frequency* has generally shown little to no effect on hypertrophy under volume-matched scenarios (though the majority of such studies do not involve very high weekly training volumes), while the overall body of evidence points towards a benefit with respect to strength. Schoenfeld 2019 Nuckols

There remains a plausible argument for frequency as a useful variable under very high volume situations, and as a simple practical tool for adherence purposes. In very high volume situations, a low-frequency training arrangement may exceed the single-session threshold for productive training, and fatigue generated may become disproportionate to

the stimulus delivered, leading to stagnation or regression compared to a broader weekly distribution of stimulus. Non-volume equated studies on training frequency have shown modest to moderate effects of higher training volumes on hypertrophic outcomes.

Caveats and Conclusions

When assessing the body of literature as a whole, it is important to note a few (among *many* other) caveats:

> 1) Many studies have low statistical power due to relatively small sample sizes, leaving us with a situation where many individual studies (particularly short-term studies) are set up to fail to find significant effects. This can be compounded by the choice of measurement methods for hypertrophy, which is a notoriously tricky endeavor. Vigotsky 2018 Haun 2019 In contrast, pooled data in meta-analytic studies can tell a different story, though the quality of the "input" data always needs to be considered as well.
>
> 2) There is an *enormous* inter-individual variability in training response (a concept previously discussed in my April 2019 BMR article discussing Ahtiainen 2016). Some individuals demonstrate better training responses to lower volume training, others to higher volume training, and we find similar variations in response to the other training variables as well. Damas 2019 It is thus ideal when resistance training studies report *individual subject-level data* (which were not provided in the present study), so we can more easily observe the degree of variation in responsiveness, outliers, and other trends that may be more difficult to detect based on summary statistics alone.
>
> 3) The current review is limited to the evidence with respect to strength and hypertrophy outcomes. The discussion regarding exercise volume, frequency, and intensity for *health outcomes* is separate and outside the scope of this piece.

This overall pattern suggests that there is no one-size-fits-all "optimal" training approach, but rather that we seem to have many potential options for training approaches that may work to produce strength & hypertrophy for an individual. Indeed, this is exactly what we observe in the real world -- lots of people achieving impressive outcomes using a variety of training methods.

Based on the current body of evidence, we might conclude that there is an interaction between training volume and frequency such that 5-10 sets for a particular muscle group is a reasonable *average single-session dose* in a trained individual, but that by distributing the stimulus across the week we might observe a continued dose-response effect into the 10-20 set per week range - though this needs to be studied more directly.

While some studies have found continued dose-response effects to much higher volumes, there is an increasing amount of conflicting data in this range, as well as an increased likelihood of inter-individual variation in training tolerance/responsiveness influencing outcomes in practice. Given what we observe in practice, combined with evidence regarding anabolic responses in trained individuals, it seems likely that the "inverted U" curve might *shift* over the course of an individual's training career, from a relatively *low* tolerance for training volume in untrained individuals to a substantially higher tolerance in more advanced stages.

Finally, the majority of data for these thresholds involves studies where subjects are training to failure; how these thresholds may vary with habitual submaximal training is less clear.

Ultimately, it is impossible to predict who will respond best to a particular approach up front, but we can use the overall body of evidence to apply broad principles and individualize things beyond that based on demonstrated training responses and individual goals and preferences.

References

1. Taber et al. "Exercise-Induced Myofibrillar Hypertrophy is a Contributory Cause of Gains in Muscle Strength" Sports Med (2019) 49: 993.
2. Loenneke et al. "Exercise-Induced Changes in Muscle Size do not Contribute to Exercise-Induced Changes in Muscle Strength" Sports Med (2019) 49: 987.
3. Nóbrega et al. "Effect of Resistance Training to Muscle Failure vs. Volitional Interruption at High- and Low-Intensities on Muscle Mass and Strength." J Strength Cond Res. 2018 Jan;32(1):162-169.
4. Lasevicius et al. "Effects of different intensities of resistance training with equated volume load on muscle strength and hypertrophy." Eur J Sport Sci. 2018 Jul;18(6):772-780.
5. Schoenfeld et al. "How many times per week should a muscle be trained to maximize muscle hypertrophy? A systematic review and meta-analysis of studies examining the effects of resistance training frequency." J Sports Sci. 2019 Jun;37(11):1286-1295.
6. McKendry et al. "Short inter-set rest blunts resistance exercise-induced increases in myofibrillar protein synthesis and intracellular signalling in young males." Exp Physiol. 2016 Jul 1;101(7):866-82.
7. Schoenfeld et al. "Longer Interset Rest Periods Enhance Muscle Strength and Hypertrophy in Resistance-Trained Men." J Strength Cond Res. 2016 Jul;30(7):1805-12.
8. Schoenfeld et al. "Dose-response relationship between weekly resistance training volume and increases in muscle mass: A systematic review and meta-analysis." J Sports Sci. 2017 Jun;35(11):1073-1082.
9. Ralston et al. "The Effect of Weekly Set Volume on Strength Gain: A Meta-Analysis." Sports medicine (Auckland, N.Z.) vol. 47,12 (2017): 2585-2601.

10. Steele et al. “Clarity in reporting terminology and definitions of set endpoints in resistance training” Muscle Nerve. 2017 Sep;56(3):368-374.
11. Barbalho et al. “Evidence for an Upper Threshold for Resistance Training Volume in Trained Women.” Med Sci Sports Exerc. 2019 Mar;51(3):515-522.
12. Martorelli et al. “Strength Training with Repetitions to Failure does not Provide Additional Strength and Muscle Hypertrophy Gains in Young Women.” Eur J Transl Myol. 2017 Jun 27;27(2):6339.
13. Pareja-Blanco et al. “Time Course of Recovery From Resistance Exercise With Different Set Configurations.” J Strength Cond Res. 2018 Jul 20.
14. Morán-Navarro et al. “Time course of recovery following resistance training leading or not to failure.” Eur J Appl Physiol. 2017 Dec;117(12):2387-2399.
15. Ogasawara et al. “Relationship between exercise volume and muscle protein synthesis in a rat model of resistance exercise.” J Appl Physiol (1985). 2017 Oct 1;123(4):710-716.
16. Damas et al. “Resistance training-induced changes in integrated myofibrillar protein synthesis are related to hypertrophy only after attenuation of muscle damage.” J Physiol. 2016 Sep 15;594(18):5209-22.
17. Brook et al. “Skeletal muscle hypertrophy adaptations predominate in the early stages of resistance exercise training, matching deuterium oxide-derived measures of muscle protein synthesis and mechanistic target of rapamycin complex 1 signaling.” FASEB J. 2015 Nov;29(11):4485-96.
18. Damas et al. “A review of resistance training-induced changes in skeletal muscle protein synthesis and their contribution to hypertrophy.” Sports Med. 2015 Jun;45(6):801-7.
19. Moore et al. “Ingested protein dose response of muscle and albumin protein synthesis after resistance exercise in young men.” Am J Clin Nutr. 2009 Jan;89(1):161-8.
20. Witard et al. “Myofibrillar muscle protein synthesis rates subsequent to a meal in response to increasing doses of whey protein at rest and after resistance exercise.” Am J Clin Nutr. 2014 Jan;99(1):86-95.
21. Moore et al. “Protein ingestion to stimulate myofibrillar protein synthesis requires greater relative protein intakes in healthy older versus younger men.” J Gerontol A Biol Sci Med Sci. 2015 Jan;70(1):57-62.
22. Radaelli et al. “Dose-response of 1, 3, and 5 sets of resistance exercise on strength, local muscular endurance, and hypertrophy.” J Strength Cond Res. 2015 May;29(5):1349-58.
23. Schoenfeld et al. “Resistance Training Volume Enhances Muscle Hypertrophy but Not Strength in Trained Men.” Med Sci Sports Exerc. vol. 51,1 (2019): 94-103.
24. Paulsen et al. “The influence of volume of exercise on early adaptations to strength training.” J Strength Cond Res. 2003 Feb;17(1):115-20.
25. Baker et al. “Strength and Body Composition Changes in Recreationally Strength-Trained Individuals: Comparison of One versus Three Sets

Resistance-Training Programmes" BioMed research international vol. 2013 (2013): 615901.
26. Rønnestad et al. "Dissimilar effects of one- and three-set strength training on strength and muscle mass gains in upper and lower body in untrained subjects." J Strength Cond Res. 2007 Feb;21(1):157-63.
27. Ostrowski et al. "The effect of weight training volume on hormonal output and muscular size and function". J Strength Cond Res. 1997;11(1):148-154.
28. Heaselgrave et al. "Dose-Response Relationship of Weekly Resistance-Training Volume and Frequency on Muscular Adaptations in Trained Men." Int J Sports Physiol Perform. 2019 Mar 1;14(3):360-368.
29. Schoenfeld et al. "Differential Effects of Heavy Versus Moderate Loads on Measures of Strength and Hypertrophy in Resistance-Trained Men" J Sports Sci Med. 2016 Dec 1;15(4):715-722.
30. Baz-Valle. "Total Number of Sets as a Training Volume Quantification Method for Muscle Hypertrophy: A Systematic Review." J Strength Cond Res. 2018 Jul 30.
31. Nuckols. "Training Frequency for Strength Development: What the Data Say." Retrieved from: https://www.strongerbyscience.com/training-frequency/
32. Vigotsky et al. "Methods matter: the relationship between strength and hypertrophy depends on methods of measurement and analysis" PeerJ. 2018 Jun 27;6:e5071.
33. Haun et al. "A Critical Evaluation of the Biological Construct Skeletal Muscle Hypertrophy: Size Matters but So Does the Measurement" Front Physiol. 2019 Mar 12;10:247.
34. Damas et al. "Individual Muscle Hypertrophy and Strength Responses to High vs. Low Resistance Training Frequencies." J Strength Cond Res. 2019 Apr;33(4):897-901.

Author Information

Michael Ray, MS, DC

@michael_barbellmedicine

mike@barbellmedicine.com

Leg Length Discrepancy: Much ado about nothing?

Leg Length Discrepancy: The Natural History (And What Do We Really Know) by Gordon et al. 2019.

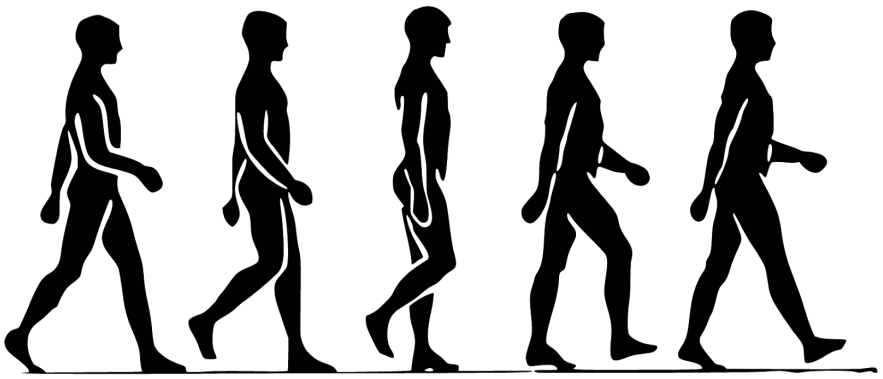
Key Points:

1. Leg Length Discrepancy (LLD) has a high prevalence rate, with ~90% of the population displaying a <1.0 cm difference.

2. There is very weak evidence of poor quality demonstrating correlation between a particular leg length discrepancy and symptoms; broad generalizations should not be drawn from the available data at this time.

3. It is important to recognize that we adapt to such deviations from textbook “norm”.

Introduction



A guiding principle to follow when discussing potential correlates to pain is “does it matter?” And if “it” matters, we need to figure out 1) when it matters, and 2) what can / should we actually do about it?

Leg Length Discrepancy (LLD, a.k.a. anisomelia) is defined as a measurable difference between the length of the lower extremities, and is often blamed for numerous issues related to pain or dysfunction due to the asymmetry. It’s important to point out the underlying flawed premise here: that a structural “problem” necessarily results in symptoms. We’ve discussed at length the flaws of such a biomedical approach in the past, and will continue to do so in the future. Pain is a complex experience that cannot be reduced to simple anatomical observations that might superficially seem problematic (see Lederman 2011).

Two variations of LLD have been described: “Anatomical” and “Functional”.

An “anatomical” LLD describes structural differences in femur and/or tibia length, either naturally occurring in the course of development or *acquired* at some point during life (e.g., due to fracture, bony disease, or joint replacement).

“Functional” LLD isn’t as straightforward -- and although frequently discussed, lacks supporting research evidence. Typically the discussion of functional LLD is centered around pseudoscientific, poorly defined ideas such as “pelvic torsion”, “tight” vs “loose” muscles, or “subluxations”.

Measuring Leg Length Discrepancies

Many perform assessments of LLD in clinics/gyms via palpation, tape measures, or blocks (to level the pelvis). Overall, the available research shows that such approaches are ineffective. A more accurate assessment can be obtained via radiological imaging, such as X-ray or CT scanogram. Sabharwal 2008, Gibbons 2002, Cooperstein 2017

But even though we *can* measure LLDs, before we pathologize this supposed “issue” the more important question is: how readily identifiable are leg length discrepancies in the general population?

Purpose

In 1978, Richard Gross was the first to try and quantify the amount of LLD necessitating intervention. At the time, he concluded any LLD less than 2 cm likely wasn’t an issue. The purpose of the article in review this month by Gordon *et al* was to assess the current state and quality of evidence regarding LLDs since Gross’s initial publication.

Methods

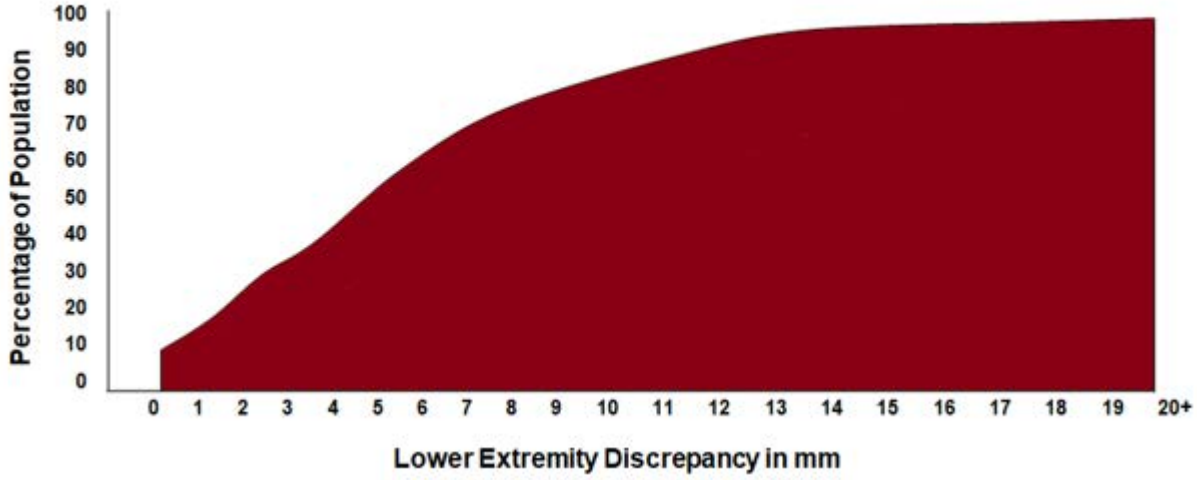
The authors performed a search for medical literature assessing the effects and treatment of leg length discrepancies. The authors excluded any level 5 evidence (case-reports, expert opinion, and personal observations), articles not written in English, and any article assessing the risk(s) of procedures resulting in LLD. The authors categorized their findings into 2 broad topics: natural history of LLD and gait analyses of patients with LLD.

Recall - “natural history” is typically defined as the usual trajectory of a supposed issue without intervention. In essence, these studies are examining a group of patients with LLD to assess for any associated problems.

Results

The authors found that approximately 90% of the population has an LLD (see figure 1). On average, a 5.2 mm difference has been observed.[Knutson 2005]

Figure1. Leg Length Discrepancy



According to Knutson et al, 10% of the population has equal lower limb lengths, ~50% has a 4 mm LLD or less, and ~90% has a 10 mm LLD or less.[Knutson 2005]

Given the high prevalence of LLD, we must ask whether an LLD of a particular measurement matters (i.e., is it correlated to pain and/or dysfunction)?

Not necessarily. Knutson's review concludes:

> *"In summary,* ***childhood-onset anatomic leg-length inequality appears to have little clinical significance up to 20 mm****. Several authors agree, most recently with Kakushima et al who stated: 'Therefore, although conflicts in the literature exist,* ***3 cm of LLD [leg length discrepancy] can be characterized as a minimum LLD, which should be treated in the clinical practice'****. This estimation of clinical significance dovetails nicely with the findings on the effects of LLI, particularly pelvic torsion. Passive structural changes – pelvic torsion, mild lumbar scoliosis, facet angulation, changes in muscle length – seem capable of compensating for anatomic LLI of up to 20 mm. Past the ~ 20 mm point, passive structural changes give way to active muscular compensatory measures." [emphasis ours]*

Based on this information, 2-3 cm seems to be the ***minimum threshold*** for warranting investigation and possible intervention, and this is likely correlated more to *quality of life* than to pain.

Taking a closer look at the articles included in Gordon *et al*, we find inconclusive evidence on LLD and future issues such as joint replacement, osteoarthritis, pain, etc. The available data aren't even capable of showing whether the long leg or the short leg are at greater risk for negative effects.

Joint replacement

Tallroth *et al* 2005 examined 100 participants undergoing hip-replacement surgery. 39 displayed a longer right leg and 42 with left. Nineteen participants did **NOT** have an LLD. Sixty-eight participants had the hip replacement performed on the longer leg (mean difference in leg length was 7.5 mm, standard deviation 4.7 mm), and 13 on the shorter leg (mean difference in leg length was 4.4 mm, standard deviation 3.2 mm).

Tallroth *et al* 2017 - 193 participants initially underwent X-ray and then followed-up 29 years later to see who received joint replacements of the hip or knee. The breakdown of participants and LLDs was as follows:

- 24 (12%) **no** LLD
- 62 (32%) 1 - 4 mm LLD
- 74 (38%) 5 - 8 mm LLD
- 21 (11%) 9 - 12 mm LLD
- 12 (6%) over 12 mm LLD

Of the 193 participants, **only 16 (8%) went on to receive a joint replacement for the diagnosis of primary osteoarthritis**. The authors go on to report 10 participants had replacement completed on their longer leg and 3 on their shorter leg. This is a low number of participants receiving joint replacement from this cohort and it would be very difficult to conclude receiving a joint replacement was primarily due to an LLD. Three participants **WITHOUT** an LLD went on to receive a joint replacement as well, leading us to further question LLD as a major correlate, if one at all, to the need for a future joint replacement.

Overall, this evidence isn't sufficient to support LLD as a major correlate leading to joint replacement.

"Degenerative" Findings

Before diving into these articles, let's recall we have a plethora of data showing asymptomatic degenerative findings (deviation from textbook norm) in various body regions (see table 1). Many of the studies included in Gordon *et al* are examining the relationship between LLD and degenerative findings such as osteoarthritis or lumbar disc herniations. Even if we see a correlation between the presence of LLD and the presence of degenerative findings, this does not mean the person has symptoms or will later develop symptoms purely because of the LLD.

****Continued on next page****

Table 1 - Asymptomatic Imaging Findings	
Disc "issues"	Brinjikji 2015
Facet Osteoarthritis	Kalichman 2008
Sacroiliac Osteoarthritis	Eno 2015
Hip Osteoarthritis	Kim 2014
Knee Osteoarthritis	Culvenor 2018

Furthermore, such a biomedical premise further reduces the person experiencing symptoms to solely a biological tissue issue where we have mounting evidence has a large prevalence rate in the general population (as discussed above).

With that said, let's take a closer look at the included studies examining degenerative findings and LLD.

Murray et al 2017- lumbo-pelvic x-ray imaging reviewed from 255 adults to assess LLDs and degenerative findings in the hips and lumbar spine. The authors noted an increased risk for degenerative joint disease of the hip and lower lumbar spine in patients with LLD > 5 mm. A potential issue with this study is the lack of full-length x-rays from pelvis and lower extremities.

Ten Brinke *et al* 1999 - 132 participants being seen for neurological symptoms supposedly attributed to lumbar disc herniation were examined for an LLD. LLDs were measured indirectly (non-radiographically), making the results questionable. The authors report a mean LLD of 5.4 ± 5.2 mm (range of 0 - 26 mm). 104 participants demonstrated an LLD > 1 mm. 28 showed no LLD. 64 (62%) of the 104 participants with LLD displayed radiating symptoms into shorter leg.

Harvey et al 2010 - 2964 patients with full-length x-rays were followed for 30 months and examined for increase in osteoarthritis. Overall, 14.5% of participants (n=429) displayed an LLD ≥ 1 cm and 0.9% (n=27) had an LLD ≥ 2 cm. Participants with an LLD > 1 cm were further examined. Of those participants, 53% demonstrated increased OA on the short-side, while 36% showed increased OA on the long-side. Interestingly, the incident rate (new cases) of osteoarthritis were not influenced by an LLD ≥ 1 cm or ≥ 2 cm over the 30 month follow-up period. Participants demonstrating an LLD ≥ 1 cm, compared to those with < 1 cm LLD, did have an increased odds of developing knee symptoms over the 30-month follow-up period (shorter limb = 15% vs 9%, OR 1.7, 95%CI, 1.2-2.4 and longer limb = 13% vs 9%, OR 1.5, 95%CI, 1.0-2.1). An LLD ≥ 1 cm, compared to those with an LLD < 1cm, did demonstrate a 1.3 times greater odds of having progressive knee osteoarthritis in the shorter limb over the follow-up period (95% CI, 1.0-1.7, 29% vs 24%). Significance wasn't reached for progressive knee

osteoarthritis symptoms in the longer limb over the follow-up period. Oddly, an LLD ≥ 2 cm did not demonstrate a significant increase in the odds of developing progressive knee osteoarthritis in the shorter limb. Add more importantly, only 6 of 26 participants with an LLD ≥ 2 cm demonstrated progress of knee osteoarthritis - this is an interesting finding given the authors' premise that LLD increases the risk of developing knee osteoarthritis and claim, *"...leg length inequality as small as 0.5 to 1cm increased the risk of prevalent knee osteoarthritis, primarily in the shorter limb."* One would think this would be observed in what are being considered "large" LLDs more so than smaller LLDs, but this doesn't appear to be the case.

Low Back Pain

Defrin *et al* 2005 - 33 participants with persistent low back pain were examined for an LLD and given a shoe insert to "correct" the discrepancy. The fact that the cohort was labeled with persistent pain should immediately raise questions regarding the premise of this study, since the chronicity of pain symptoms drastically diminishes the correlation between identifiable tissue pathology and symptoms (See Durmez 2017).

The authors examined LLDs with ultrasonography and then divided participants into study vs control groups. The study group (22 participants) received a shoe insert to correct the LLD and the control group (11 participants) didn't. Baseline measurements of pain intensity and function were performed. The authors reported a significant baseline difference between groups for pain intensity, and this matters when trying to assess the efficacy of an intervention. Ideally there are no significant differences in baseline characteristics between groups that could otherwise confound results. After 12 weeks of intervention, the participants were re-examined for pain intensity and functional status. The authors conclude:

> *"This study suggests that the correction of an LLD of 10mm or less can significantly reduce CLBP. Shoe inserts are simple, inexpensive, and noninvasive means for alleviating CLBP and are therefore recommended to be included in the treatment of patients with LBP who have mild LLD."*

Examining the data more closely, these are rather bold conclusions. The study uses a small sample size, has baseline differences in pain intensity, and shows large standard deviations in treatment outcomes - all calling into question such claims. Even if we wanted to follow this reductionist line of thought, we have higher quality evidence demonstrating that shoe inserts lack efficacy and have low-level evidence for effectiveness. Recall, efficacy and effectiveness are not the same thing.

Efficacy, in this context, is examining how well a treatment performs under ideal or perfect circumstances. Typically, interventions are examined for efficacy in well conducted randomized controlled trials. These studies help answer the question - "Does 'X' intervention actually work?".

Effectiveness is assessing a treatments generalizability to real-world situations such as dealing with humans in clinical practice. Many treatments appear effective (reduce pain and improve function) for a myriad of reasons, often involving placebo-like contextual effects. Treatments appearing effective may lack efficacy, thus questioning their necessity in clinical practice for long-term positive outcomes.

In regards to shoe lifts - Campbell *et al* completed a systematic review of the literature in 2018 on adults experiencing musculoskeletal symptoms. The authors concluded:

> *"We sought evidence to answer fundamental questions for guiding clinical treatment of LLD for common painful musculoskeletal conditions. In the setting of mechanical LBP, hip, and knee OA, correction of LLD using a shoe lift may reduce pain, improve function and increase ROM; however, these benefits remain uncertain due to very low-quality evidence. We were unable to make evidence-based conclusions regarding the magnitude or proportion of LLD that should be corrected. More rigorous, high-quality studies evaluating which LLD-associated conditions benefit from shoe lift correction, shoe lift correction strategy, and relevant patient outcomes are required to guide clinical treatment. An appropriate comparison group would be helpful in this regard."*

The absence of a clear level of LLD necessitating intervention, and the lack of appropriate RCTs on the usage of shoe inserts negates our ability to draw any pragmatic conclusions.

The good news: it appears we adapt

Khamis and Carmeli *et al* 2018 - 7 healthy participants were recruited to go through a simulated LLD gait analysis. Participants were equipped with shoes simulating LLDs of 5, 10, 15, 20, 30, and 40 mm. The authors found at 5 mm LLD compensatory strategies were employed. At 10 mm the authors noted significant differences in compensatory gait strategies involving shortening of the longer limb, lengthening of shorter limb, or both.

Song *et al* 1997 - 35 children recruited with LLD. The reasons for the LLD ranged from idiopathic (unknown), congenital, fracture, and dislocation. The included children displayed an LLD ranging from 0.6 to 11.1 cm. The authors found children with small mean discrepancies of 1.6 cm, no compensatory gait strategies were observed. However, children with larger mean LLD of 6.5 cm displayed a compensatory gait strategy of toe walking. Eight children displaying LLDs between 2 and 15.8 cm demonstrated pelvic obliquity (drop of pelvis on short side). Finally, 9 children displayed increased range of pelvic obliquity as a compensatory gait strategy.

Aiona *et al* 2014 demonstrated compensated gait strategies for LLD. Forty-five children recruited with an LLD > 2 cm. The average LLD was 4.6 cm (range 2 - 12.2 cm) - this is considered a large discrepancy compared to what we've been discussing. The LLDs

were due to a variety of issues: Legg-Calve-Perthes disease, hip dysplasia, growth plate abnormalities due to trauma or infection, and congenital shortening of the femur or tibia to name a few. The authors also included 20 children in a control group for comparative gait analysis. Various biomechanical compensations were observed in the study group, from a single kinematic deviation to several adaptations (see table 2).

Table 2. Gait Strategy by Location of Short Bone

		Short Side – Ankle Strategies [n (%)]		Long Limb	Both
Short Segment	Pelvic Obliquity Only [n (%)]	Equinis Only	Vaulting Only	Flexed Knee Only [n (%)]	Multiple Strategies [n (%)]
Femur (n = 18)	1 (5.6)	2 (11.1)	5 (27.8)	5 (27.8)	5 (27.8)
Tibia (n = 9)	5 (55.6)				4 (44.4)
Both (n = 18)	5 (27.8)	2 (11.1)	3 (16.7)		8 (44.4)

Interestingly, the authors found very similar gait velocities between groups. The control group had an average walking velocity of 1.3 ± 0.2 m/s (range 0.9 to 1.6 m/s) vs. the study group with an average walking velocity of 1.2 ± 0.2 m/s (range 0.8 to 1.5 m/s). The authors state:

> *"Our study demonstrated a variety of gait compensations for LLD. The magnitude of the difference and the location of the difference appear to be important factors in determining the compensation strategy. If the discrepancy was >7 cm, all patients used a combination strategy. As the differences became less, a greater variety of pattern choice was noted. For example, if >4 cm but <7 cm, only 3/11 chose multiple strategies, with the most frequent isolated pattern being pelvic tilt."*

The primary takeaway from the above articles is that compensation is evident in gait based on the size of the LLD. However, whether these compensations lead to any negative outcomes over the longer term is not known. We also don't know if a certain size LLD warrants intervention given these compensatory strategies.

Why does this article matter?

Leg length discrepancies are readily identifiable in the general population, with some sources reporting an average difference of approximately 5 mm. Furthermore, screening for LLDs would not be advisable given the large prevalence rate in the asymptomatic general population and lack of strong correlation to symptomatic development or future negative effects. There is substantial controversy over a

threshold LLD necessitating intervention, and after review of the available evidence, this remains unknown.

Many clinicians and coaches anecdotally report LLDs being causative for pain or performance issues, however, there is a paucity of evidence linking a particular LLD to either. Given what we know about the complexities of pain, this would be an extremely reductionist approach that at this time is indefensible based on current evidence. This means that if a patient reports in clinic with low back pain, it would not be recommended to check for an LLD. Low back pain has a high prevalence in the 4th decade of life and yet many clinicians check for LLDs to see whether it is related to the patient's symptoms. However, if the patient is in their 40s, a finding of an LLD would certainly have been adapted to at this point. Even if we wanted to follow this line of clinical reasoning, the typical recommendation is then a shoe insert - an intervention greatly lacking in high quality evidential support.

Regarding LLDs and performance, at this time it would appear we adapt to a range of LLDs. Is there potentially a point where an LLD becomes problematic for a patient, sacrificing quality of life? Potentially, but we've yet to find a generalizable, predictive rule. Such scenarios are likely much more related to non-idiopathic situations involving congenital deficits, fracture, post-joint replacement, tumor, infection, etc. We even have examples of high-level athletes with LLDs. Gordon *et al.* discuss Usain Bolt having a reported 1.3 cm LLD and video analysis demonstrating "*...right leg striking the ground with 13% more force and his left leg spending 14% more time on the ground.*" This makes it difficult to single out such an isolated biomechanical finding and claim it as a "problem" necessitating fixing. Overall, Gordon *et al* conclude:

> *"In conclusion, the evidence for the effect of leg length discrepancy and the amount of leg length discrepancy that we should be treating is quite poor and probably has advanced little since Gross's initial survey of pediatric orthopaedic surgeons."*

What does this mean for us as clinicians and coaches?

We should ensure we aren't making problems out of normative findings in the general population, and we should understand that either way, *we are adaptable*. We certainly should NOT be screening for LLDs in routine practice. Finally, attempting to make claims about an LLD's relationship with pain or performance issues would not only be overly reductionist but stands in opposition to what evidence is currently showing us. There may be cases where a non-idiopathic LLD is an issue, but for the majority of clinicians and coaches, these cases are likely an exceedingly small fraction of real-world encounters.

References:

1. Gordon JE, Davis LE. Leg Length Discrepancy: The Natural History (And What Do We Really Know). Journal of pediatric orthopedics. 2019; 39(Issue 6, Supplement 1 Suppl 1):S10-S13.
2. Lederman E. The fall of the postural-structural-biomechanical model in manual and physical therapies: exemplified by lower back pain. Journal of bodywork and movement therapies. 2011; 15(2):131-8.
3. Sabharwal S, Kumar A. Methods for Assessing Leg Length Discrepancy Clin Orthop Relat Res. 2008; 466(12):2910-2922.
4. Gibbons P, Dumper C, Gosling C. Inter-examiner and intra-examiner agreement for assessing simulated leg length inequality using palpation and observation during a standing assessment Journal of Osteopathic Medicine. 2002; 5(2):53-58.
5. Cooperstein R, Lucente M. Comparison of Supine and Prone Methods of Leg Length Inequality Assessment Journal of Chiropractic Medicine. 2017; 16(2):103-110.
6. Knutson GA. Anatomic and functional leg-length inequality: a review and recommendation for clinical decision-making. Part I, anatomic leg-length inequality: prevalence, magnitude, effects and clinical significance. *Chiropr Osteopat*. 2005;13:11. Published 2005 Jul 20. doi:10.1186/1746-1340-13-11.
7. Tallroth K, Ylikoski M, Lamminen H, Ruohonen K. Preoperative leg-length inequality and hip osteoarthrosis: a radiographic study of 100 consecutive arthroplasty patients. Skeletal radiology. 2005; 34(3):136-9.
8. Tallroth K, Ristolainen L, Manninen M. Is a long leg a risk for hip or knee osteoarthritis?. *Acta Orthop*. 2017;88(5):512–515. doi:10.1080/17453674.2017.1348066
9. Murray KJ, Molyneux T, Le Grande MR, Castro Mendez A, Fuss FK, Azari MF. Association of Mild Leg Length Discrepancy and Degenerative Changes in the Hip Joint and Lumbar Spine Journal of Manipulative and Physiological Therapeutics. 2017; 40(5):320-329.
10. ten Brinke A, van der Aa HE, van der Palen J, Oosterveld F. Is leg length discrepancy associated with the side of radiating pain in patients with a lumbar herniated disc? Spine. 1999; 24(7):684-6.
11. Harvey WF, Yang M, Cooke TD, et al. Association of leg-length inequality with knee osteoarthritis: a cohort study. Annals of internal medicine. 2010; 152(5):287-95.
12. Defrin R, Ben Benyamin S, Aldubi RD, Pick CG. Conservative correction of leg-length discrepancies of 10mm or less for the relief of chronic low back pain. Archives of physical medicine and rehabilitation. 2005; 86(11):2075-80.
13. Durnez W, Van Damme S. Let it be? Pain control attempts critically amplify attention to somatosensory input. Psychological research. 2017; 81(1):309-320.
14. Campbell TM, Ghaedi BB, Tanjong Ghogomu E, Welch V. Shoe Lifts for Leg Length Discrepancy in Adults With Common Painful Musculoskeletal Conditions:

A Systematic Review of the Literature. Archives of physical medicine and rehabilitation. 2018; 99(5):981-993.e2.
15. Khamis S, Carmeli E. The effect of simulated leg length discrepancy on lower limb biomechanics during gait. Gait & posture. 2018; 61:73-80.
16. Song KM, Halliday SE, Little DG. The effect of limb-length discrepancy on gait. The Journal of bone and joint surgery. American volume. 1997; 79(11):1690-8.
17. Aiona M, Do KP, Emara K, Dorociak R, Pierce R. Gait patterns in children with limb length discrepancy. Journal of pediatric orthopedics. ; 35(3):280-4.

Author Information

Derek Miles, PT, DPT

@derek_barbellmedicine

derek@barbellmedicine.com

BARBELL MEDICINE

Are We Certain Heavy Eccentrics Work for Mid-Portion Achilles Tendinopathy?

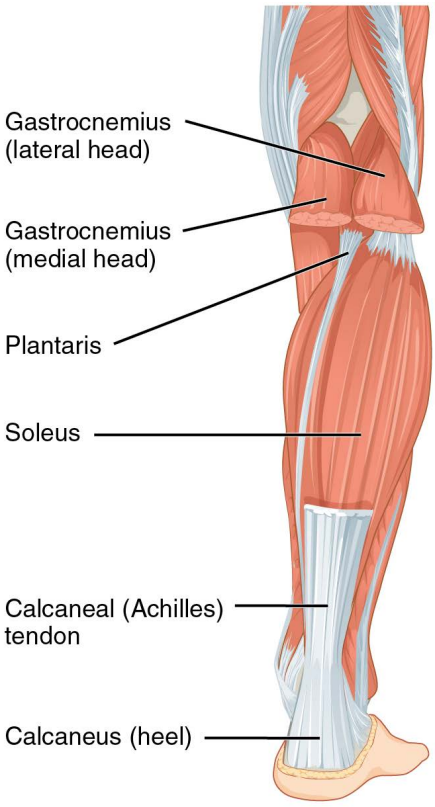
Efficacy of heavy eccentric calf training for treating mid-portion Achilles tendinopathy: a systematic review and meta-analysis by Calder Murphy et al 2019

Key Points:

1. Eccentric calf training may be better than natural history or the use of passive modalities in the treatment of mid-portion Achilles tendinopathy.

2. Heavy slow resistance training may be slightly better than eccentric calf training for the treatment of mid-portion Achilles tendinopathy.

3. The certainty with which we can claim these interventions work for the treatment of Achilles tendinopathy should be questioned and challenge current clinical practice.



Introduction

There are very few things we can claim with certainty in the rehabilitation world. Working with humans is a multifactorial experience where unique individual factors make it difficult to predict individual responses to our interventions. With that said, if a word association game were played for the treatment of Achilles tendinopathy, the immediate answer would most likely include eccentric training.

This dates back to the original Alfredson study from 1998 on recreational distance runners with mid-portion Achilles tendinopathy.[Alfredson 1998] The original treatment paradigm was 3 sets of 15 calf raises with both a gastroc and soleus bias, done 2 times/day, 7 days/week, for 12 weeks. This loading protocol would involve a significant time demand on athletes but has been accepted as the gold standard for treatment of mid-portion Achilles tendinopathy.

The subjects of this study (*n*=15) had an average age of 44.3 years and were compared against a surgical group (yes, back in the 90's surgery was a treatment for tendinopathy, and some still advocate for this approach today). While the subjects in this study did achieve good results, it is difficult to extrapolate results for middle aged, recreational runners to high level competitive athletes across other sports.

The past twenty years has seen eccentrics accepted as the go-to treatment and, anecdotally, I have taught courses highly endorsing their utility. Recommendations for eccentrics evolved with the introduction of Kongsgaard papers to include *heavy slow resistance (HSR) training,* and the conversation turned to the possibility that the *magnitude* of the eccentric load was as important as the dosage. Kongsgaard 2009, Kongsgaard 2010 The 2010 Kongsgaard paper was also unique in that it identified fibril morphology changes with a 12 week loading program. This would seem to run contrary to the "treat the donut, not the hole" mantra proposed by some groups and adopted by more.Docking 2014 In this paradigm it is generally accepted that once tendon tissue enters the degenerative phase it is unlikely to adapt to further stimulus. Hence, an intervention is not going to *change* the degenerative tissue but rather possibly lay down tissue in the periphery.

However, both of those studies were performed on patellar tendons. It took until the Beyer paper included in the current meta-analysis to examine specific heavy slow resistance training for Achilles tendinopathy.Beyer 2015 Around the same time is when the *tendon-pathology continuum* became popularized, categorizing tendons into "normal", "reactive", "degenerative", or "reactive-on-degenerative". Cook 2016 A continuum had also emerged for treatment, advocating for an initial phase of isometrics for analgesia, followed by eccentrics/heavy slow resistance, followed by energy storage and release (e.g. plyometric or power movements). The original paper utilizing isometrics involved a small cohort (*n*=6) in an uncontrolled randomized cross-over design for patellar tendinopathy. They ran a protocol of 5 sets of 45 second isometrics at 70% 1RM load.Rio 2015 A larger cohort study (*n*=29) was performed by the same group in 2016 that did not show any difference between isometric and isotonic exercises for reducing pain in patellar tendinopathy, with both demonstrating equivalent effectiveness. Once again, these findings have been extrapolated to treatment of Achilles tendinopathy as well.

The current systematic review by Calder Murphy *et al* sought to determine the magnitude of effect of eccentric exercises compared to other interventions for the treatment of Achilles tendinopathy. The authors do refer to the Alfredson protocol as Heavy Eccentric Calf Training (HECT) which is an interesting nomenclature given the *lack of resistance* generally accepted with the protocol. Without meta-analyses such as this, it is difficult to say with confidence which treatments should be the first-line approach for a particular pathology. There is often a conflation of "*may* be used in

treatment" with "*should* be used in treatment" in clinical paradigms. Without controlled evidence, it is difficult to endorse one training modality being superior to another.

Methods

This study is a pre-registered systematic review of randomized and quasi-randomized trials with one arm of the study including HECT to treat mid-portion Achilles tendinopathy and the other using natural history, sham treatment, "traditional" physical therapy, or another intervention.

Included subjects

Studies with subjects older than 18, active or sedentary, with a greater than three month history of mid-portion Achilles tendinopathy were eligible for inclusion.

Interventions

Interventions were categorized as *sham* if patients underwent exercise determined by the authors unlikely to result in a physiological response i.e. exercises that did not overload contractile tissue to induce a strength response. *Traditional Physical Therapy* interventions were classified as those *without* exercise intervention that included deep friction massage to the tendon, other forms of manual therapy to local tissue, ultrasound, or taping. None of those interventions have shown efficacy in the treatment of mid-portion Achilles tendinopathy but the authors elected to classify the interventions as traditional physical therapy instead of sham.[Sussmilch-Leitch 2012] The authors elected to compare HECT to natural history, sham exercise, traditional physical therapy, and different exercise interventions.

Primary Outcome Measure

Heterogeneity of outcomes measures has made conducting meta-analyses difficult.[Habets 2015] The authors of this study elected to only include studies with a validated and reliable outcome measure for mid-portion Achilles tendinopathy. For this, studies that utilized the VISA-A outcome measure were included.[Robinson 2001] The VISA-A is the only valid, reliable outcome measure specific to mid-portion Achilles tendinopathy.[Murphy 2018] The mean difference (MD) and 95% confidence intervals (95% CI) were included for all trials. The final time point at which the original study was conducted was used for analysis. The authors elected to not include studies only utilizing the visual analog scale (VAS) or numeric rating scale (NRS) as they have been shown to possess poor test-retest reliability for Achilles tendinopathy.[Silbernagel 2001]

Data Collection

Two authors conducted the search according to PRISMA guidelines.[Liberati 2009] The included studies were assessed for risk of bias using the Cochrane Risk of Bias Tool.[Higgins 2011] This allows for the included studies to be classified as low risk of bias, some concerns of bias, or high risk of bias. If data was missing the authors contacted the original study authors for the data and if it could not be provided the study was excluded from the meta-analysis.

Assessments

The authors conducted an assessment of heterogeneity, but due to the stringent inclusion criteria of this meta-analysis using a specific diagnosis and outcome measure, the authors assumed it would be low. Where heterogeneity was identified, a subgroup analysis was performed to determine the impact on the overall findings via sensitivity analysis. In registering their systematic review the authors also planned to conduct a sensitivity analysis in instances where:

- The standard deviation had to be manually input
- Studies in which adherence was not reported
- Different exercise interventions versus HECT
- Studies in which HECT and other exercise interventions were compared against placebo
- Studies in which there was a high risk of bias as determined by the high risk of bias tool

If a study contained less than 50 participants it was determined to possess a high risk of small sample bias. Assessment of the quality of evidence was performed using GRADE criteria.[Atkins 2004] Data were then synthesized using an inverse variance and random effects model.

Results

Seven studies met criteria for inclusion in the meta-analysis (Table 1). One study compared HECT to natural history, no studies compared HECT to placebo/sham, two studies compared HECT to traditional physical therapy, and four studies compared HECT to another exercise intervention.

The seven included studies contained a total of 241 participants with a mean (SD) age range of 36.6 (7.2) to 49.2 (11.3) years. Five of the trials reported a gender distribution with participants 45% male and 55% female. Four of the trials reported body mass index (BMI) with it ranging from 25.0 (5.0) to 31.6 (6.1) kg/m^2. Five of the trials reported the mean duration of symptoms ranging from 6.2 to 27.6 months. All studies

reported a baseline VISA-A ranging from 36 (23.4) to 62 (18). The duration of follow up ranged from 6 weeks to 16 weeks.

There was no heterogeneity between the studies that compared HECT to traditional physical therapy (I^2=0%) but there was significant heterogeneity between studies comparing HECT to different exercise interventions (I^2=68%). One study comparing HECT to traditional physical therapy had a high risk of bias.[Rompe 2007] One study comparing HECT to traditional physical therapy had some concern for risk of bias.[Wiedmann 2017]

Study Name	Study Design	Sample Size	Mean Age	Symptom Duration (mo)	Intervention	Follow-up (wks)
Beyer *et al*	RCT	47	48	17-19	HECT vs HSR	12
Herrington and McCullough	RCT	25	37	21.3-27.6	HECT vs Traditional Physical Therapy	12
Rompe *et al*	RCT	50	47	9.2-10.9	HECT vs Natural History	16
Stasinopoulos and Manias	RCT	41	48	6.9-7.1	HECT vs Stanish Protocol	12
Stevens and Tan	RCT	26	49	6.2-8.9	HECT vs mHECT	6
Tumilty *et al*	RCT	32	47	N/A	HECT vs mHECT	12
Wiedmann *et al*	RCT	20	43	N/A	HECT vs Traditional Physical Therapy	12

Table 1- Included studies in the meta-analysis. mHECT=modified eccentric protocol, HECT= heavy eccentric calf training protocol.

Effects of Interventions

Only one study compared HECT to natural history with a statistically significant mean difference (MD) of 20.6 (11.69 to 29.51) in favor of HECT. Unfortunately, the study was rated as a high risk of bias and prone to small sample size bias. Two studies compared HECT to traditional physical therapy (Figure 1) with results favoring the use of HECT. The MD (+/-95% CI) was 17.70 (3.75 to 31.66) with both of the trials at risk for small sample size bias and one raising some concerns for risk of bias.

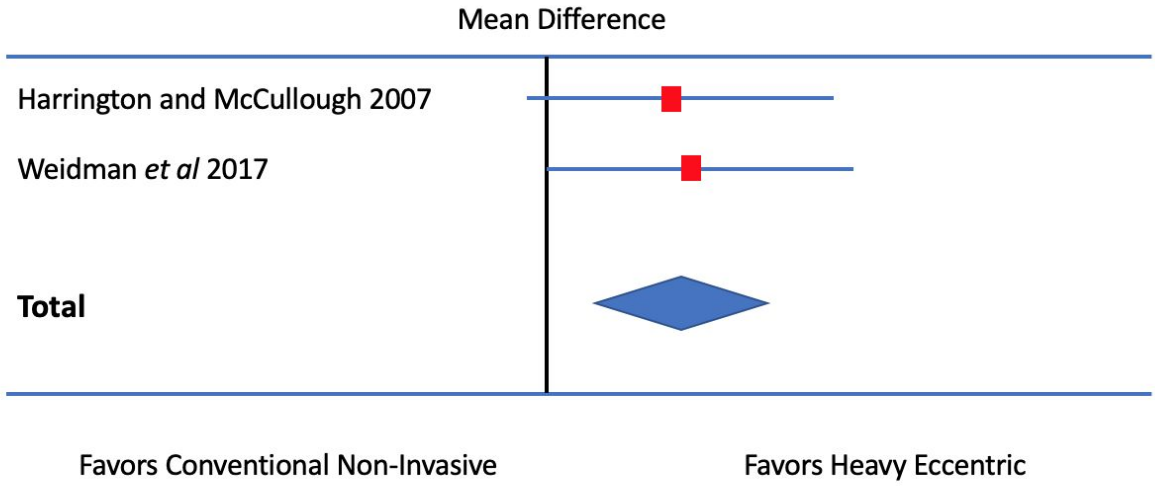


Figure 1- Forest Plot of HECT versus traditional physical therapy.

Four studies were included in the meta-analysis of HECT versus other exercise interventions. Two of these compared HECT to a modified heavy eccentric protocol, one compared to heavy slow resistance training, and one compared to the Stanish Protocol. The Stanish protocol is a 12 week program that changes exercises weekly with a focus on speed then load.[Stasinopoulos 2013] Here, there was a non-significant pooled MD (95% CI) -1.19 (-9.40 to 7.01) (Figure 2). All studies were at a risk for small sample bias and the Stanish protocol study was at a high risk of bias by Grade scoring.

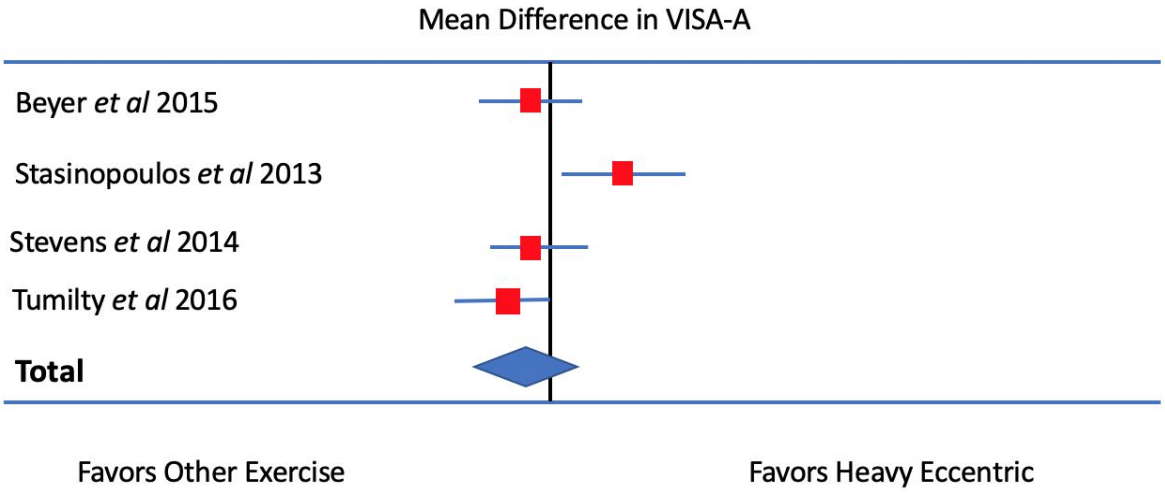


Figure 2-HECT versus other exercise interventions

The preplanned sensitivity analysis by the authors called for the removal of the Stasinopoulos study due to a lack of reporting of adherence and high risk of bias. The MD of this study (13.00) was also discongruent with the other three (MD -3.8 to -7.20). This study was subsequently removed with the updated forest plot seen in figure 3.

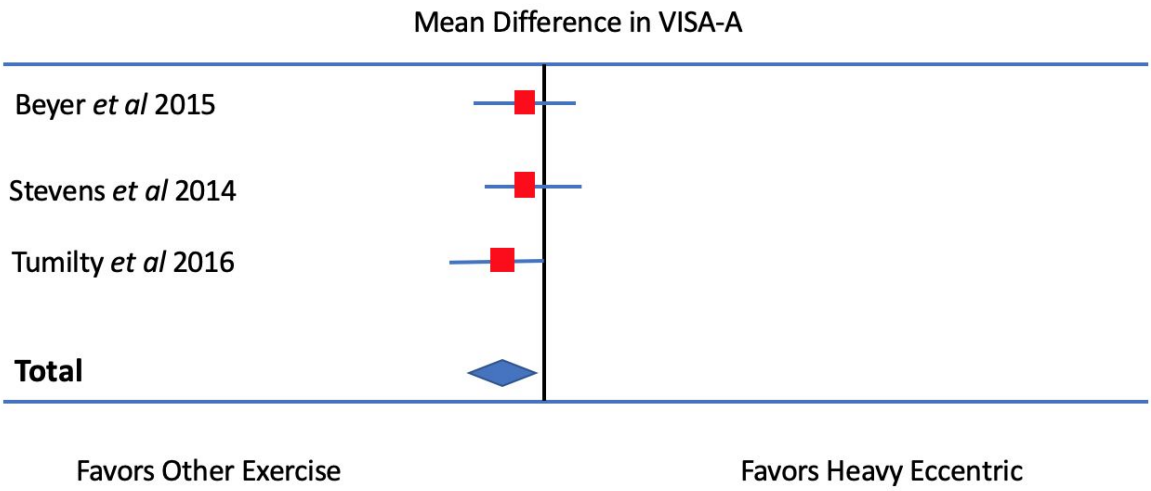


Figure 3- Sensitivity Analysis Adjusted Forest Plot of HECT versus other exercise.

Why This Study Matters

If the reader is experiencing a slight existential crisis after seeing these data, that is okay. Meta-analyses like this are excellent for showing the limitations of what our treatments actually can change. We advocate for a biopsychosocial approach to treatment, but sometimes we are also guilty of applying a treatment with the expectation that changing tissue alone is enough to improve function.

Unsurprisingly, there is evidence that psychosocial factors play a role in the manifestation and treatment of tendinopathy.[Mallows 2017] What an exercise scheme does to a tendon at a local level does not necessarily correlate to functional outcomes, even under the best-prescribed loading schemes. This meta-analysis is unique in that it did focus on the VISA-A as an outcome measure which is a *functional* outcome measure. While pain and function are inherently linked in the treatment of any condition, only looking at pain misses an individual's ability to participate in activity.

While the evidence here would state with some degree of certainty that the Alfredson protocol is better than passive modalities, it is still not the panacea for

treatment of Achilles tendinopathy. It is also interesting that the authors of this study refer to the Alfredson protocol as *heavy* eccentric calf training since most of the exercises are not performed with a substantial external load. Six sets of fifteen, two times a day is also a substantial time commitment on the part of the athlete. In the Beyer article comparing heavy slow resistance to the Alfredson protocol, both groups did improve, but the time constraint and patient satisfaction were initially higher in the heavy slow group. The heavy slow athletes were also committing 107 min/week to training versus 308 min/week in the HECT group. That is an additional ***3+ hours*** that an individual has to train other components or do other things.

We must also consider the subjects involved in these studies. The average age of the overall cohort was well into their 40's and were not competitive athletes. Translating loading protocols from recreational athletes or inactive individuals to those competing at a higher levels of sport may not be valid. The magnitude of loading seems to play an integral role in tendon adaptation, with tendons needing higher loads with which to adapt.[Magnusson 2010] In the younger population, Mersmann *et al* advocate for low repetition, very high intensity (>85% 1RM) in order to elicit tendon adaptation.[Mersmann 2017] This is well above anything advocated for in HECT and even the Beyer paper on heavy slow resistance. It could be that if we are wanting true structural adaptation, *all protocols advocated for in this meta-analysis may be under-dosed*. The Beyer protocol advocates for the following set-up:

- Week 1: 3 sets of 15 to failure (15 RM)
- Weeks 2-3: 3 sets of 12 to failure (12RM)
- Weeks 4-5: 4 sets of 10 to failure (10RM)
- Weeks 6-8: 4 sets of 8 to failure (8RM)
- Weeks 9-12: 4 sets of 6 to failure (6RM)

It would not be until week 9 that this protocol would approach the recommended dosage for tendon adaptation by Mersmann. The Beyer article also advocates for only performing the exercises 3x/week and taking a 2-3 minute rest in between sets. This is one of the only protocols that discusses intensity, rest, and frequency of dosage. While we may be prescribing the right *type* of exercises for our athletes, if they continue to be *under-dosed* we still may not achieve the outcomes we seek.

All of this still pertains to the local tendon response and adaptation. Communication regarding expectations on a human level of patient goals, barriers, and expectations is needed to maximize outcomes. If these are not accounted for, the effect size for any treatment approaches will likely be un-impressive.

We advocate against viewing a patient as a “herniated disc,” “degenerated rotator cuff,” or “worn-out knee”, and we similarly cannot view them as an “Achilles tendinopathy.” Developing a treatment plan from an entirely structural approach will always fall short compared to a plan that addresses the *person*. This review shows that while exercise certainly has a role in the treatment of tendinopathy, the role of eccentrics is likely less than what is currently accepted. As with almost all conditions, an individualized approach is necessary to account for psychosocial factors, as well as to titrate loading to the current goals and needs of each patient. Haphazardly dosing 6 sets of 15, 2 times a day was likely insufficient from the start. Currently, there are four main loading protocols within the literature for the treatment of tendinopathy (Table 2), while they account for symptoms and most have a form of progression, intensity is lacking on all but the heavy slow resistance protocol.

Program	Type of Exercise	Set/Reps	Frequency	Progression	Pain
Alfredson	Eccentric	3/15	Twice Daily	Load	Enough load to achieve moderate Pain
Stanish	Eccentric-Concentric, then power	3/10-20	Daily	Speed, then Load	Enough load to be painful on the third set
Silbernagel	Eccentric-concentric, eccentric, faster, balance, plyometric	Various	Daily	Volume, then type of exercise	Acceptable within defined limits
HSR	Eccentric-Concentric	4/15-6	3x/week	15-6RM	Acceptable if not worse after

Table 2- Current loading protocols for the treatment of tendinopathy

It is generally accepted that a good training program accounts for sets, reps, intensity, and rest between sets. Failing to account for any of these variables in a rehabilitation program likely makes it subpar as well. While there is evidence that eccentric loading protocols create tendon adaptation, the dose makes the cure, the poison, or the treatment inert. If we apply dosing protocols for recreational, middle aged athletes to athletes who perform at a high level and consistently it is hard to surmise why outcomes are less than ideal.

As with all evidence, more is needed and welcomed. This meta-analysis does make an excellent case that, with what we currently have, we should still possess some hesitancy before seeing eccentric training as the ultimate fix for Achilles tendinopathy.

References

1. Murphy MC, Travers MJ, Chivers P, *et al*. Efficacy of heavy eccentric calf training for treating mid-portion Achilles tendinopathy: a systematic review and meta-analysis. Br J Sports Med. 2019 Jan 13. pii: bjsports-2018-099934.
2. Alfredson H, Pietila T, Jonsson P, *et al*. Heavy-load eccentric calf muscle training for the treatment of chronic Achilles tendinosis. Am J Sports Med. 1998 May-Jun;26(3):360-6.
3. Kongsgaard M, Kovanen V, Aargaard P, *et al*. Corticosteroid injections, eccentric decline squat training and heavy slow resistance training in patellar tendinopathy. Scand J Med Sci Sports. 2009 Dec;19(6):790-802.
4. Kongsgaard M, Qvortrup K, Larsen J, *et al*. Fibril morphology and tendon mechanical properties in patellar tendinopathy: effects of heavy slow resistance training. Am J Sports Med. 2010 Apr;38(4):749-56.
5. Docking S, Rosengarten S, Daffy J, *et al*. Treat the donut, not the hole: The pathological Achilles and patellar tendon has sufficient amounts normal tendon structure. JSAMS. 2014 Dec; 14(18): e2.
6. Beyer R, Kongsgaard M, Kjaer B, *et al*. Heavy Slow Resistance Versus Eccentric Training as Treatment for Achilles Tendinopathy: A Randomized Controlled Trial. Am J Sports Med. 2015 Jul;43(7):1704-11.
7. Cook J, Rio E, Purdam CR, *et al*. Revisiting the continuum model of tendon pathology: what is its merit in clinical practice and research? Br J Sports Med. 2016 Oct;50(19):1187-91.
8. Rio E, Kidgell D, Purdam CR, *et al*. Isometric exercise induces analgesia and reduces inhibition in patellar tendinopathy. Br J Sports Med. 2015 Oct;49(19):1277-83.
9. Sussmilch-Leitch SP, Collins NJ, Bialocerkowski A, *et al*. Physical therapies for Achilles tendinopathy: systematic review and meta-analysis. Foot Ankle Res. 2012 Jul 2;5(1):15.
10. Habets B, van Cingel R. Eccentric exercise training in chronic mid-portion Achilles tendinopathy: a systematic review on different protocols. Scand J Med Sci Sports. 2015 Feb;25(1):3-15.
11. Robinson J, Cook J, Purdam C, *et al*. The VISA-A questionnaire: a valid and reliable index of the clinical severity of Achilles tendinopathy. Br J Sports Med. 2001 Oct;35(5):335-41.
12. Murphy M, Rio E, Debenham J, *et al*. Evaluating the progress of mid-portion Achilles tendinopathy during rehabilitation: a review of outcome measures for self-reported pain and function. Int J Sports Phys Ther. 2018 Apr;13(2):283-292.

13. Silbernagel K, Thomee R, Thomee P, *et al*. Eccentric overload training for patients with chronic Achilles tendon pain--a randomised controlled study with reliability testing of the evaluation methods. Scand J Med Sci Sports. 2001 Aug;11(4):197-206.
14. Liberati A, Altman D, Tetzlaff J, *et al*. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med. 2009 Jul 21;6(7):e1000100.
15. Higgins C. Cochrane Handbook for Systematic Reviews of Interventions.
16. Atkins D, Best D, Briss P, *et al*. Grading quality of evidence and strength of recommendations. BMJ. 2004 Jun 19;328(7454):1490.
17. Rompe J, Nafe B, Furia J, *et al*. Eccentric loading, shock-wave treatment, or a wait-and-see policy for tendinopathy of the main body of tendo Achillis: a randomized controlled trial.
18. Wiedmann M, Mauch F, Huth J, et al. Treatment of mid-portion Achilles tendinopathy with eccentric training and its effect on neovascularization. Sports Orthopaed Trauma 2017;33
19. Stasinopoulos D, Manias P. Comparing two eccentric exercise programmes for the management of Achilles tendinopathy. A pilot trial. J Bodyw Mov Ther. 2013 Jul;17(3):309-15.
20. Mallows A, Debenham J, Walker T, *et al*. Association of psychological variables and outcome in tendinopathy: a systematic review. Br J Sports Med. 2017 May;51(9):743-748.
21. Magnusson SP, Langberg H, Kjaer M. The pathogenesis of tendinopathy: balancing the response to loading. Nat Rev Rheumatol. 2010 May;6(5):262-8.
22. Mersmann F, Bohm S, Arampatzis A. Imbalances in the Development of Muscle and Tendon as Risk Factor for Tendinopathies in Youth Athletes: A Review of Current Evidence and Concepts of Prevention. Front Physiol. 2017 Dec 1;8:987.