Technische Universität München

The Impact of GnRHa on the Bone Health in Transgender and Gender Diverse Youth

Current State of Research

Wissenschaftliche Arbeit zur Erlangung des Grades

B.Ed. Berufliche Bildung Gesundheits- und Pflegewissenschaften

an TUM School of Social Sciences and Technology der Technischen Universität München.

Betreuer/-in Jennifer Queisser, M.Ed.
Lehrstuhl für Präventive Pädiatrie

Aufgabensteller/-in Jennifer Queisser, M.Ed.
Lehrstuhl für Präventive Pädiatrie

Eingereicht von Petra Stiefenhofer

[redacted]

Eingereicht am München, den 31.05.2023

Table of Contents

Abstract

Background: Transgender and Gender Diverse (TGD) youth often experience gender dysphoria which can be intensified by the appearance of secondary sex characteristics during puberty. This development of secondary sex characteristics and thus the worsening of gender dysphoria can be alleviated or even prevented by treatment with Gonadotropin-releasing hormone agonists (GnRHa). GnRHa prevent the progression of puberty as sex hormone production is suppressed. These sex hormones, however, are a key factor of growth during adolescence – including bone growth. Therefore, it is possible that GnRHa impact bone health. Thus, for an empirically based use of GnRHa, it is necessary to consider possible impacts and effects of GnRHa on bone health in TGD youth.

Aim: Within this work, the current state of research on the impact of GnRHa monotherapy on the bone health in TGD youth is to be presented.

Methods: To answer the research question, a systematic literature search was carried out from December 2022 to February 2023. For this purpose, literature was sought out in the following databases: Google Scholar, PubMed, Medline, and Scopus. The search was guided by a research strategy plan as well as inclusion and exclusion criteria. In the end, 7 studies were included, analyzed, and discussed in this paper.

Results: The literature reviewed suggests a negative impact of GnRHa on bone health in TGD youth. Bone health parameters (bone mineral density, bone mineral apparent density, bone mineral content, and bone turnover markers) do not develop according to age with GnRHa monotherapy. Z-scores of bone mineral density and bone mineral apparent density decrease compared to peers. In addition, TGD adolescents appear to have insufficient or deficient levels of vitamin D and calcium.

Discussion and conclusion: Current research shows that bone health (particularly bone mass) is negatively impacted by GnRHa monotherapy in TGD adolescents. Among other consequences, this can result in an increased risk of bone fractures. However, further studies on this topic are necessary to solidify this assumption. Overall, more research needs to be done on the impact of GnRHa monotherapy on the bone health in TGD youth – especially in transgender girls.

Hintergrund: Transgender- und genderdiverse (TGD)-Jugendliche leiden häufig unter Geschlechtsdysphorie. Diese kann sich mit dem Auftreten sekundärer Geschlechtsmerkmale während der Pubertät weiter verstärken. Die Entwicklung der sekundären Geschlechtsmerkmale und somit eine Verschlimmerung der Geschlechtsdysphorie lässt sich durch eine Behandlung mit Gonadotropin-Releasing-Hormon-Agonisten (GnRHa) mildern oder sogar verhindern. GnRHa verhindern das körperliche Fortschreiten der Pubertät, indem die Produktion von Sexualhormonen unterbunden wird. Diese Sexualhormone sind ihrerseits jedoch ein wichtiger Faktor für das Wachstum während der Pubertät – einschließlich des Knochenwachstums. Somit ist es möglich, dass GnRHa einen Einfluss auf die Knochengesundheit haben. Für einen empirisch fundierten Einsatz von GnRHa ist es somit notwendig, mögliche Einflüsse und Auswirkungen von GnRHa auf die Knochengesundheit von TGD-Jugendlichen zu berücksichtigen.

Ziel: In dieser Arbeit soll der aktuelle Forschungsstand zum Einfluss von GnRHa Monotherapie auf die Knochengesundheit von TGD-Jugendlichen dargestellt werden.

Methoden: Zur Beantwortung der Forschungsfrage wurde von Dezember 2022 bis Februar 2023 eine systematische Literaturrecherche durchgeführt. Zu diesem Zweck wurde Literatur in den folgenden Datenbanken durchsucht: Google Scholar, PubMed, Medline und Scopus. Die Suche wurde anhand eines Forschungsstrategieplans sowie Ein- und Ausschlusskriterien durchgeführt. Am Ende der Literaturrecherche wurden 7 Studien in die vorliegende Arbeit aufgenommen und anschließend analysiert sowie diskutiert.

Ergebnisse: Die untersuchte Literatur deutet auf einen negativen Einfluss von GnRHa auf die Knochengesundheit bei TGD-Jugendlichen hin. Parameter der Knochengesundheit (u.a. Knochenmineraldichte) entwickeln sich während GnRHa-Monotherapie nicht altersgemäß, und Z-Scores dieser Parameter nehmen im Vergleich zu Gleichaltrigen ab. Darüber hinaus scheinen TGD-Jugendliche unzureichende oder mangelhafte Vitamin-D- und Kalziumspiegel zu haben.

Diskussion und Schlussfolgerung: Aktuelle Forschungsergebnisse zeigen, dass die Knochengesundheit (insbesondere die Knochenmasse) während der GnRHa-Monotherapie bei TGD-Jugendlichen negativ beeinflusst wird. Dies kann unter anderem das Risiko für Knochenbrüche erhöhen. Um diese Annahme zu bestätigen sind weitere Studien zu diesem Thema erforderlich. Es bedarf gezielterer und speziell fokussierter Forschung zur Auswirkung von GnRHa Monotherapie auf die Knochengesundheit bei TGD-Jugendlichen – insbesondere transgender Mädchen betreffend.

List of Abbreviations

1CTP
Carboxy Terminal Cross Linked Telopeptide of Type 1 Collagen 9

25OHD
Serum 25-Hydroxyvitamin D..... 9

aBMD
Areal Bone Mineral Density..... 8

AFAB
Assigned Female at Birth 4

AMAB
Assigned Male at Birth 4

BMAD
Bone Mineral Apparent Density..... 8

BMC
Bone Mineral Content..... 8

BMD
Bone Mineral Density 8

BMFSFJ
Bundesministerium für Familie, Senioren, Frauen und Jugend..... 2

BTM
Bone Turnover Markers..... 9

CSH
Cross-Sex Hormone..... 5

DXA
Dual Enery X-Ray Absorptiometry..... 8

FN
Femoral Neck..... 8

GnRHa
Gonadotropin-Releasing Hormone analogues / agonists..... 5

IOM
US Institute of Medicine 9

LS
Lumbar Spine..... 8

LTH
Left Total Hip..... 36

NB
Nonbinary..... 4

NCI
National Cancer Institute..... 6

List of Tables

List of Figures

1. Introduction

“What the board has sought to do is to protect our children from therapies that have been shown to create irreversible harm.” – Dr. Hector Vila, 2023

Transgender health is a timely topic that has become increasingly more significant in recent years. This becomes clear when one looks at the current political debates on this topic in various countries around the world. In the United States of America (short: USA), for example, the political situation regarding transgender health and gender-affirming care is currently very charged. To illustrate the relevance and timeliness of the issue of Gonadotropin-releasing hormone agonist treatment in transgender and gender diverse adolescents suffering from gender dysphoria, the current political situation in the United States (short: U.S.) is presented in brief:

Since the start of 2023, multiple states in the USA have restricted gender-affirming care for minors who wish to transition (The General Assembly of the State of Iowa, 2023; The General Assembly of the State of Arkansas, 2023; The Legislature of the State of Montana, 2023; The Legislature of the State of South Dakota, 2023). The therein newly introduced bills are called “anti-trans bills”. Medical procedures affected by these bills include – but are not limited to – the suppression of puberty using puberty blockers such as Gonadotropin-releasing-hormone agonists (= GnRHa). Cross-sex hormone therapies and gender reassignment surgeries are also affected. The wording within these “anti-trans bills” varies, but it is striking that the omission of such medical interventions applies to minors only when they are administered “for the purpose of attempting to alter the appearance of, or to validate a minor's perception of, the minor’s sex, if that appearance or perception is inconsistent with the minor's sex” (The Legislature of the State of South Dakota, 2023). Under different circumstances, such as a “medically verifiable disorder of sex development [… or a] disorder of sexual development” (The Legislature of the State of South Dakota, 2023), the procedures may still be performed.

Furthermore, the Florida Department of Health announced a new rule in March 2023 stating that certain “therapies and procedures performed for the treatment of gender dysphoria in minors are prohibited” (Florida Department of Health - Board of Medicine, 2023). Again, these procedures include inter alia the suppression of puberty. When asked about it in an interview, Dr. Hector Vila – member of the Florida Board of Medicine – explained, the board seeks to “protect [the] children from therapies that have been shown to create irreversible harm” (NBC News, 2023). This, however, bases the “protective changes” for gender transitioning in minors on the assumption that puberty blockers can cause irreversible damage.

In Germany, there are different opinions on this topic. The use of puberty blockers in minors is currently not regulated by law in Germany. Following an online debate on this topic, the German Federal Ministry for Family Affairs, Senior Citizens, Women and Youth (*Bundesministerium für Familie, Senioren, Frauen und Jugend*, short: BMFSFJ) issued a statement regarding the handling of puberty blockers:

> Puberty blockers can only be prescribed:
>
> - after careful medical [consideration and] indication,
> - on the basis of scientific guidelines,
> - by specialists.
>
> The Federal Government does not recommend the use of puberty blockers. The decision to prescribe puberty blockers is **at the sole discretion of the treating specialists**.
>
> (BMFSFJ, 2022)

Leaving it to medical professionals to decide whether gender dysphoria in minors should be treated with puberty blockers, the German Federal Government still stresses that such treatment must not endanger the mental and physical health of children (BMFSFJ, 2022). Nevertheless, there are members of the federal parliament in Germany (= "Bundestag") that demand a change of the legislation. In 2022, the right-wing populist party *Alternative für Deutschland* (AfD) called for the Federal Government to submit a bill which should prevent children (who are not yet able to consent but still suffer from gender dysphoria) from being treated with puberty blockers, cross-sex hormones, and similar drugs. These members – similar to the legislatures in certain U.S. states – reference the (current) state of research and its alleged lack of knowledge regarding the physical and psychological side effects and consequences of puberty blockers. (von Storch, et al., 2022)

But does GnRHa treatment lead to irreversible changes? Researchers agree that puberty blockers, like most medical interventions, have side effects. Some authors found that while the scale of negative effects of puberty blockers lacked clarification, GnRH analogues do delay the increase of bone density during puberty, a time that is critical for the development of peak bone mass (Ahrbeck & Felder, 2022).

While the topic of transgender health is of growing political interest, the present work aims to determine whether the use of puberty blockers in form of GnRHa monotherapy leads to harmful side effects regarding bone health in transgender and gender diverse youth. This manuscript focuses solely on the scientific and evidence-based medical discussion of this topic. For this purpose, a systematic review was conducted to present and discuss the current state of research on this subject in a structured manner.

2. Theoretical Background

2.1. Important Terminology and Definitions

“Where, after all, do universal human rights begin? In small places, close to home – so close and so small that they cannot be seen on any maps of the world. Yet they are the world of the individual person” - Eleanor Roosevelt, 1958

These words, spoken by Eleanor Roosevelt, first Chairperson of the Commission on Human Rights, mark the importance of treating everyone with respect and dignity. Article I of the Declaration of Human Rights states that “all human beings are born free and equal in dignity and rights” (UDHR, 1948). To maintain this dignity and to avoid misunderstandings, the correct linguistic handling of the terms of different gender identities and the aspects contained therein is essential. The terms relevant to this paper as well as medical terms and parameters to determine bone health are therefore defined below.

2.1.1. Gender Identity

Sex corresponds to the biological trait assigned to an individual at birth (binary, i.e., female, or male). Gender, on the other hand, is a person's gender identity. (Martin & Hadwin, 2022) Therefore, gender identity is not about the sex one is born with but rather the gender one feels like and identifies with (Coleman, et al., 2022). There is a variety of genders one can identify as, one of them being *transgender*. The use of ***transgender and gender diverse*** (short: TGD) in the title and all other parts of this work that do not explicitly refer to individual studies is based on the terminology used in the Standards of Care Version 8 (short: SOC-8). SOC-8 is an evidence-based guideline for medical professionals treating transgender and gender diverse people, published by the World Professional Association for Transgender Health (short: WPATH), which is an “international, multidisciplinary, professional association whose mission is to promote evidence-based care, education, research, public policy, and respect in transgender health” (Coleman, et al., 2022, p. S5). Within this guideline the purpose of using the term *TGD* is “to be as broad and comprehensive as possible in describing members of the many varied communities globally of people with gender identities or expressions that differ from the gender socially attributed to the sex assigned to them at birth” (Coleman, et al., 2022, p. S11). Therefore, *TGD* youth was determined to be the adequate term to be used in this paper. Referring to WPATH, the terms *transgender* and *gender diverse* can be defined further:

Gender Diverse applies to people “with gender identities and/or expressions that are different from social and cultural expectations attributed to their sex assigned at birth” (Coleman, et al., 2022, p. S252). In the context of this paper *gender diverse* discusses youth identifying as both “transgender” and “nonbinary” (compare the following two paragraphs).

Transgender as an umbrella term refers to people “whose gender identities and/or gender expressions are not what is typically expected for the sex to which they were assigned at birth” (Coleman, et al., 2022, p. S252). WPATH emphasizes that "transgender" should never be applied as a noun or verb, but only as an adjective, which is why this paper continuously speaks of TGD youth or TGD people. Since adolescents are of interest in this paper, the term **transgender boys** will be used instead of the term transgender men used in SOC-8. However, the definition of transgender men as “people who have gender identities as men and who were assigned female at birth [and] may or may not have undergone any transition” (Coleman, et al., 2022, p. S253) can also be applied to transgender boys. For the same reason, **transgender girls** will be used instead of transgender women. Again, the definition of transgender women as “people who have gender identities as women and who were assigned male at birth [and] may or may not have undergone any transition” (Coleman, et al., 2022, p. S253) is applicable to transgender girls. WPATH determines terms such as Female-to-Male and Male-to-Female to be “falling out of use” (Coleman, et al., 2022, p. S253); therefore, those terms will not be used in this work.

Nonbinary (short: NB), in turn, is defined in SOC-8 as people identifying “as partially a man and partially a woman or […] as sometimes a man and sometimes a woman, or […] as a gender other than a man or a woman, or as not having a gender at all” wherein some nonbinary people can “consider themselves to be transgender [and] some do not because they consider transgender to be part of the gender binary” (Coleman, et al., 2022, p. S252).

Given that **sex assigned at birth** (short: SAB) is relevant when using the adjective *transgender*, it is also important to define this term. The SOC-8 defines SAB as follows:

> Sex assigned at birth refers to a person’s status as male, female, or intersex based on physical characteristics. Sex is usually assigned at birth based on appearance of the external genitalia. AFAB is an abbreviation for “assigned female at birth.” AMAB is an abbreviation for “assigned male at birth.”
>
> (Coleman, et al., 2022, p. S252)

To limit the scope of this paper, the use of these two abbreviations (AFAB and AMAB) is resorted to in some places, for example in Table 5.

In addition to the terms contained in the title of this paper, other technical terms, used in this work, need to be defined:

Gender Dysphoria (short: GD) as described in the SCO-8 is a “state of distress or discomfort that may be experienced because a person’s gender identity differs from that which is physically and/or socially attributed to their sex assigned at birth” (Coleman, et al., 2022, p. S252). The American Psychology Association’s Diagnostic and Statistical Manual of Mental Disorders defines GD as a “clinically significant distress or impairment related to gender incongruence, which may include desire to change primary and/or secondary sex characteristics” (APA, n.d.). The term listed in the manual mentioned above serves the purpose of diagnostic classification.

The process of changing “from the gender expression associated with their assigned sex at birth to another gender expression that better matches their gender identity“ (Coleman, et al., 2022, p. S253) is called **transition**. One can transition in different ways, e.g., socially by disclosing their preferred pronouns or physically using inter alia GnRHa (Coleman, et al., 2022, p. S253).

2.1.2. Medical Terms

Apart from terms relating to TGD people, several medical terms must be introduced to this paper.

In most countries (e.g., the United Kingdom and Germany), gender-affirming treatment measures leading to irreversible, phenotypic changes, such as Cross Sex Hormone (short: CSH) therapy may only be used from the age of 16 (NHS, 2020; Meyer, 2021). Therefore, suppressing puberty in TGD youth with GD is a good measure which can be applied even at a younger age, to reduce the experienced pain and emotional suffering due to GD. The following briefly explains the working mechanisms of so-called *puberty blockers*. Furthermore, *youth* is defined, as are puberty assessment criteria relevant to this paper.

Sex hormones cause bodily changes and the development of secondary sex characteristics during puberty. AFABs usually start puberty at 8 – 13 years old (short: y/o) while puberty in AMABs usually occurs a little later at an age of 9 – 14 years. (Breehl & Caban, 2022). To achieve a momentary arrest of puberty in TGD youth, the release of the sex hormone Gonadotropin-releasing Hormone, which causes the bodily changes during puberty and the development of secondary sex characteristics (Breehl & Caban, 2022), needs to be suppressed. This suppression can be achieved using *puberty blockers*. The SOC-8 explicitly recommends the use of **Gonadotropin-Releasing Hormone Analogues** (also: Gonadotropin-releasing-hormone agonists; short: GnRHa) to suppress puberty in TGD youth (Coleman, et al., 2022, p. 111 and 113). GnRHa in

the simplest of terms, as described by the National Cancer Institute (short: NCI), is “a substance that keeps the testicles and ovaries from making sex hormones by blocking other hormones that are needed to make them. In men, GnRHas cause the testicles to stop making testosterone. In women, they cause the ovaries to stop making estrogen and progesterone” (NIH (National Cancer Institute), n.d.). Importantly, the use of GnRHa leads to reversable changes and only results in a delay of the "development of irreversible pubertal changes" (Guss & Gordon, 2022). Consequently, when GnRHa treatment is stopped, puberty progresses again.

As this paper discusses the use of GnRHa to suppress puberty a definition of youth, young people and adolescents is needed. **Youth** and **young people** as defined by the United Nations (short: UN) are interchangeable terms and describe “persons between the ages of 15 and 24 years” (UN, n.d.). The World Health Organization (short: WHO) sets a different age range of 10 to 19 for **adolescents** in the “phase of life between childhood and adulthood” where they – among other things – experience “rapid physical […] growth” (WHO, n.d.). In the context of this work, an age between approximately 12 and 16 years is assumed and referred to as *youth* or *adolescents*.

The development of the body during puberty can be classified by **Tanner Stages**. These are subdivided into five stages, ranging from pre-pubertal (Stage 1) to adult form (Stage 5). For both, AMAB as well as AFAB one aspect of classification is the Pubic Hair Scale. In addition, the Male External Genitalia Scale assesses AMABs, and the Female Breast Development Scale is applied to AFABs. (Emmanuel & Bokor, 2022) The characteristics of each stage are portrayed in the following table:

	Pubic Hair Scale	**Male External Genitalia Scale**	**Female Breast Development Scale**
Stage 1	No hair	*Testicular volume* < 4 ml or *long axis* < 2.5 cm	No glandular breast tissue palpable
Stage 2	Downy hair	4 ml - 8 ml or 2.5 to 3.3 cm long **= 1st pubertal sign in AMAB**	Breast bud palpable under the areola **= 1st pubertal sign in AFAB**
Stage 3	Scant terminal hair	9 ml - 12 ml (or 3.4 to 4.0 cm long)	Breast tissue palpable outside areola; no areolar development

	Pubic Hair Scale	Male External Genitalia Scale	Female Breast Development Scale
Stage 4	Terminal hair that fills the entire triangle overlying the pubic region	15 - 20 ml (or 4.1 to 4.5 cm long)	Areola elevated above the contour of the breast, forming a “double scoop” appearance
Stage 5	Terminal hair that extends beyond the inguinal crease onto the thigh	> 20 ml (or > 4.5 cm long)	Areolar mound recedes into single breast contour with areolar hyperpigmentation, papillae development, and nipple protrusion

Table 1: *Tanner Stages* based on Emmanuel & Bokor, 2022

Tanner Stages are often described according to the scales above, though the full name of the scale is rarely written out. They are often abbreviated to:

- “Tanner P” for the Pubic Hair Scale,
- "Tanner G" for the Male External Genitalia Scale, and
- "Tanner B" for the Female Breast Development Scale

A "Tanner G2", for example, corresponds to the classification of Stage 2 on the Male External Genitalia and suggests the person under consideration has a testicular volume of 4 - 8 ml or a long axis of 2.5 to 3.3 cm.

Given a reliable diagnosis of GD, it is recommended to start puberty suppression at a Tanner Stage of 2 (Coleman, et al., 2022, p. S48 and Statement 6.12f), which defines the first pubertal sign in AMABs and AFABs, respectively. After the start of GnRHa treatment, “there is typically a regression of one Tanner Stage” (Coleman, et al., 2022, p. S112 ff.), so the sex characteristics seen in Tanner Stage 2 regress back to Tanner Stage 1 (Coleman, et al., 2022, p. S113). However, Tanner Stage 2 defines only the lower limit for initiation of GnRHa therapy; “GnRH agonists may be appropriate in late stages or in the post-pubertal period (e.g., Tanner Stage 4 or 5)” (Carmichael, et al., 2021, p. S64).

2.1.3. Bone Health

Key part of this work is the impact of GnRHa on *bone health* in TGD youth. To objectively assess bone health, several factors need to be analyzed. Indices of bone health are inter alia bone mass measurements and bone turnover markers. In the following, bone mass measurements, associated measuring techniques, evaluation instruments, and bone turnover markers will be defined and explained. Furthermore, the importance

of calcium and vitamin D will be explained. How insufficient development of these parameters can affect bone health is described in 5.1.

Bone mineral content (short: BMC) and **bone mineral density** (short: BMD), also called bone density, bone mass, areal BMD or aBMD, are used to identify bone mass.

BMC provides data on bone minerals regardless of the area or volume examined, though it does depend on bone length, width, and density (Mølgaard, Thomsen, Prentice, Cole, & Fleischer Michaelsen, 1997). It is of importance as a determinant of bone strength (Almeida, et al., 2017) and reported in grams (g).

BMD (also: aBMD) is the quantity of minerals (e.g., calcium) "contained in a certain volume of bone" (NIH (National Cancer Institute), n.d.). It is typically reported as absolute areal values in grams per square centimeter (g/cm^2) (= BMC per bone area) and measured using e.g., DXA scans. Measuring sites include:

- the lumbar spine (short: LS; often at lumbar vertebrae L1 – L4),
- the region of the hip, specifically the femoral neck (short: FN), often of the non-dominant or left hip and
- sometimes the whole body (less head)

In children and growing youth, specifically, **bone mineral apparent density** (short: BMAD) is used to calculate volumetric BMD (short: vBMD) (Crabtree, et al., 2014, p. 47). BMAD is corrected for the growing size of the body and bone (as most youth still increase in height). It is most commonly reported as grams per cubic centimeter (g/cm^3) and usually measured at the LS and / or FN.

The accumulation and level of bone mass for BMC or BMD can be determined based on **dual energy X-ray absorptiometry** (short: DXA) scans (Institute of Medicine (US), 2011). According to the WHO, DXA scans are the gold standard for determining bone mineral density (Varacallo, Seaman, Jandu, & Pizzutillo, 2022). **DXA scans** are typically performed to assess "bone metabolic status (BMD, T-Score and Z-Score for lumbar, femur, forearm and whole-body scan)" (Bazzocchi, Ponti, Albisinni, Battista, & Guglielmi, 2016, p. 1483). Advantages of this method compared to others are low cost, easy handling, and its accuracy as well as a "very low radiation dose to the patient" (Bazzocchi, Ponti, Albisinni, Battista, & Guglielmi, 2016, p. 1482). Additionally, "DXA is precise and permits monitoring of patients over time" (U.S. Department of Health and Human Services (Pt. 4), 2004). DXA scans of the LS to determine BMD are "best for monitoring treatment effect" (Sheu & Diamond, 2016).

To interpret results of DXA scans of a person in comparison to "age-, gender-, and ethnicity-matched norms" (U.S. Department of Health and Human Services (Pt. 4), 2004), **z-scores** are used for BMD and BMAD. The "z-score is compared to a healthy, age-

matched control group" (Varacallo, Seaman, Jandu, & Pizzutillo, 2022). Since it is unclear how much of the world's population consists of TGD people and data – especially in young TGD people – is scarce, the studies at hand compare the data of TGD youth with their peers according to their sex assigned at birth, respectively. Most studies reviewed in this paper report BMD z-scores adjusted for birth sex and age as well as occasionally height or ethnicity. For BMAD z scores, adjustments vary across studies depending on different reference groups.

Bone turnover markers (short: BTM; also: serum bone markers) can help predict fracture risk and provide complementary information to radiological measurements of bone mass (Greenblatt, Tsai, & Wein, 2017). They also "respond rapidly to changes in bone physiology" (Greenblatt, Tsai, & Wein, 2017). Multiple BTM, such as bone formation markers procollagen type 1 N-terminal propeptide (short: P1NP) and osteocalcin, as well as the bone resorption marker carboxy terminal cross linked telopeptide of type 1 collagen (short: 1CTP) (Garnero, 2009), are relevant to this work.

Besides the parameters of determining bone mass listed above, it is necessary to identify factors for maintaining and promoting bone health such as **calcium**. To reach peak bone mass, an adequate intake of calcium inter alia during adolescence is necessary – mostly for bone mineral deposition (Greer, Krebs, & Committee on Nutrition, 2006). Inadequate calcium intake can lead to reduced bone mass and osteoporosis and thus a higher fracture risk (Institute of Medicine (US), 1997). AFABs reach peak calcium-accretion rate at 12.5 years on average, earlier than AMABs at 14.0 years of age (Greer, Krebs, & Committee on Nutrition, 2006). According to the U.S. National Institutes of Health (short: NIH), children aged 9 to 13 as well as teens aged 14 to 18 should have a daily intake of 1.300 mg calcium (NIH (Office of Dietary Supplements), n.d.).

Vitamin D promotes the absorption of calcium and is another vital component of bone health. More accurately, vitamin D is needed for the body to produce the hormone calcitriol which "stimulates the intestines to absorb enough calcium […] and also affects bone directly" (U.S. Department of Health and Human Services (Pt. 1), 2004). Its status can be measured using serum 25-hydroxyvitamin D (short: 25OHD) (Holick, 2009), which corresponds to calcitriol and can be measured in nmol/L. According to the Robert Koch Institute (short: RKI), which in turn references data provided by the US Institute of Medicine (short: IOM), a 25OHD level between 30 and 50 nmol/L indicates a vitamin D insufficiency and a level below 30 nmol/L indicates a vitamin D deficiency (RKI, 2019).

2.2. Status Quo and Current Problem

Currently, a new law called the “Selbstbestimmungsgesetz” (in English: right of self-determination law) is being planned by the German *Federal Ministry for Family Affairs* and the *Federal Ministry of Justice* to abolish the outdated and partly unconstitutional "Transsexuellengesetz" (in English: German Transsexuals Act (short: TSG)) (BMFSFJ, 2023). In this context, it is important to keep up with time and ensure the integrity, safety, health as well as access to health care of the TGD population – not only, but also in Germany.

TGD youth should be protected and cared for – especially when it comes to the gender-affirming interventions they are taking to transition to the desired gender. The first step in this journey is often puberty suppression. GnRHa treatment delays the development of secondary sex characteristics and thus spares TGD adolescents from suffering from gender dysphoria whilst giving them time to solidify their decision to transition. Indicated use of puberty suppression as part of the gender-affirming interventions is deemed as appropriate care in the SOC-8 for TGD youth (Coleman, et al., 2022, p. S6). Additionally, discussing bone health with TGD people is recommended by the SOC-8 (Coleman, et al., 2022, p. S153).

Adolescence is a vulnerable time regarding bone accumulation and the development of bone health. Rapid bone growth is triggered by puberty, “stimulating a rise in bone mineral density” (Joseph, Ting, & Butler, 2019, p. 1). So, while the treatment with GnRHa in TGD youth with gender dysphoria can relieve their emotional suffering, it also impacts bone health, among other things, as it prevents the release of sex hormones.

Currently, researchers are not completely sure how exactly GnRHa affects bone health. In 2017, the Endocrine Society, whose cosponsoring associations are:

- the American Association of Clinical Endocrinologists,
- the American Society of Andrology,
- the European Society for Pediatric Endocrinology,
- the European Society of Endocrinology, Pediatric Endocrine Society,
- the World Professional Association for Transgender Health.

… published a Clinical Practice Guideline for the *Endocrine Treatment of Gender-Dysphoric/Gender-Incongruent Persons* (Hembree, et al., 2017). It states, among other things, that “few data are available on the effect of GnRH analogs on BMD in adolescents with GD/gender incongruence” (Hembree, et al., 2017, p. 3882). However, new studies have been published since 2017.

To be able to protect the well-being and health of TGD young people in the future, it is important to further investigate the effects of puberty blockers (especially GnRHa) on bone health – which are taken during the vulnerable phase of adolescents. Therefore, this paper aims to present the current state of research (between 2013 and 2023) on the impact of GnRHa on bone health in TGD youth. This will be done in the form of a systematic review. The resulting, summarized knowledge could help shape the application of GnRHa in the future.

3. Methodology

3.1. Topic of Research and Research Design

The focus of this paper is *The Impact of GnRHa on the Bone Health in Transgender and Gender Diverse Youth*. This research question as well as the methodology on which this work is based, is to be specified in the following section.

The research question can be structured by the PICO framework. PICO is short for:

- P = Population
- I = Intervention
- C = Comparison
- O = Outcome

(Methley, Campbell, Chew-Graham, McNally, & Cheraghi-Sohi, 2014)

Population	Transgender and gender diverse youth, diagnosed with GD
Intervention	Suppression of Puberty, using GnRHa monotherapy
Comparison	Normative data of the parameters regarding bone health
Outcome	Parameters regarding bone health • BMD, • BMAD, • BMC, • BTM (serum bone markers), • calcium, • Vitamin D

Table 2: *PICO-Framework*

A more detailed description of the contents of the table follows in the next paragraphs.

This paper focuses on *transgender and gender diverse youth, diagnosed with gender dysphoria*. A more narrowly age group than “youth” is not set, though application period of GnRHa must be at approximately 12 to 16 y/o. The studies discussed in this thesis include mostly, but are not limited to, data about transgender youth. To also respect youth who identify as e.g., nonbinary, not only transgender but also gender diverse youth is explicitly mentioned in the research question. In short, the **population** of this paper is TGD adolescents, diagnosed with gender dysphoria and subsequently receiving GnRHa therapy.

The **intervention** of interest is the *suppression of puberty* using *GnRHa monotherapy*. Focus is the impact of this procedure on the *bone health* of TGD youth. The magnitude of this impact is measured using different parameters. The main criterion for this is BMD. For the purpose of this work, BMAD is seen as a deepening of this criterion and is therefore a subordinate main criterion. Secondary criteria are BMC, BTM (= serum bone markers) as well as calcium and vitamin D levels.

Derived from population and intervention, the **comparison** is made up. Instead of an actual control group, normative data of the parameters serves as comparison. This is achieved using BMD and BMAD z-scores.

Any changes in the previously mentioned parameters (BMD, BMAD, BMC, serum bone markers, calcium, and vitamin D levels) mark the **outcome**. It is important to note that the impact of GnRHa on the bone health of the previously defined population will only be reviewed regarding *GnRHa monotherapy*; any changes in bone health which happen in combination of GnRHa with CSH therapy are not part of this work.

Aside from the sole impact of GnRHa on bone health, there are other related questions that should be answered. Without claiming to be exhaustive, potential further questions are:

- Is the impact of GnRHa on the bone health of TGD youth positive or negative?
- If GnRHa affects bone health negatively / positively, how does it show, respectively?
- What does an effect of GnRHa on the bone health imply for future treatment of TGD youth?

Another aspect could be the differences of the impact of GnRHa on bone health depending on the sex assigned at birth and its consequent differences in physiological development. This gives rise to considerations such as:

- Does GnRHa have a bigger impact on one sex?
- How significant are possible differences?

The goal of this research paper is to answer the questions described above and to identify possible existing gaps in the current state of research.

As this bachelor thesis should be conducted within a few weeks and the population of interest is extremely specific, answering the research question within the framework of a literature-based, systematic review is most coherent. Given the actuality and the specificity of the topic, there are very few studies on this area of research. The following paragraphs show how the selected studies on the topic at hand were distinguished.

3.2. Literature Selection

3.2.1. Inclusion and Exclusion Criteria

Before starting the literature search, inclusion and exclusion criteria were defined to structure the search for fitting studies and provide a framework for this paper. During the research process it turned out that the search parameters had to be very specific and targeted to find suitable literature. Readjustments were made, in particular, with regard to the selected publication period of the literature, to achieve a minimum number of 5 studies. The final inclusion and exclusion criteria are listed below.

The following inclusion criteria were established:

To determine the current state of research, only studies published in the past 10 years (2013 – 2023) were included. Furthermore, studies had to be published in English or German (as these languages are understood by the author on a native speaker level). The vast majority of studies on this subject aren't solely focused on GnRHa monotherapy and monotherapy in combination with the needed study population (TGD youth) proofed to be too narrow of a parameter. Therefore, studies which collected data during GnRHa monotherapy prior to combination with CSH therapy were included in addition to studies investigating GnRHa monotherapy itself. The study participants also had to have been diagnosed with gender dysphoria - or diagnoses that correspond to it, such as gender incongruence or gender identity disorder. The changes in bone health during GnRHa monotherapy had to be presented regarding "young adults" or "adolescents". Accordingly, the criterion set was that the data collected during GnRHa monotherapy had to relate to the age between 12 and 16 or until the start of cross-sex hormone therapy. Tanner Stages (or alternatively menarche) had to be given as an indicator of puberty stage to ensure the comparison of the studies. Regarding bone health, measurements of the BMD or BMAD of the LS, the FN or in the hip area had to be stated explicitly, though it was more important that BMD was investigated. Z-scores were to be used to describe changes in comparison to the peers. Another criterion to ensure comparability was the use of DXA scans as method of determining bone density. The last criterion was the inclusion of studies according to the Open Access principle; only studies of which the full version was accessible without payment were included.

Besides inclusion criteria, exclusion criteria were also determined:

Studies which determine the effects of GnRHa only in combination with CSH therapy as well as studies discussing the impact of GnRHa in populations which are not TGD youth were excluded. Studies examining solely the psychological effects or any other aspects regarding GnRHa use in TGD youth were also disregarded. Besides that, reviews on the topic were excluded.

3.2.2. Literature Search

The literature search was conducted online using various databases, namely Google Scholar, PubMed, MedLine, and Scopus. Prior to the literature search, a research strategy was established to then draw up different search strings. Umbrella terms were connected using Boolean operators (AND, OR and NOT) as well as truncations, according to the database, respectively. The research strategy plan, which includes the umbrella terms, is shown below:

	The Impact of GnRHa on the Bone Health in Transgender and Gender Diverse Youth			
Synonyms	GnRHa	Bone Health	Transgender	Youth
	GnRH analogue	Bone	Transgender Youth	Adolescent
	Gonadotropin-Releasing Hormone		Transgender Person	Adolescence
	Gonadotropin-releasing hormone agonist		Transsexualism	
	GnRH agonist			
	Puberty Blocker			
	Puberty Suppressant			

Table 3: *Research Strategy Plan*

Besides the terms used in the table above, it became apparent that, in order to get fitting results, it was useful to include terms which were not originally part of the research strategy plan, such as *gender dysphoria*.

During the research process it turned out that the same search strings did not always achieve the desired results across all databanks. Therefore, the multitude of search strings used in different databanks and their outcomes were documented individually. Besides this adjustment, the publication period was also gradually extended during the research progress, after it became clear that too short of a period does not lead to enough citable literature. Therefore, no fixed publication frame was used to narrow down the results, if the search didn't put out more than 1.500 results – which was considered a checkable number of results. In the end, a time frame of 10 years was set – meaning only studies published between January 2013 and February 2023 were included. However, this does not mean that the published data had to have been collected during that time. After performing a rough overview on the topic in December 2022, the research took place from January 2023 to the beginning of March 2023.

To keep track of the research, the literature search was conducted sequentially in the databases listed hereafter: firstly Google Scholar, followed by PubMed and Medline, and finally Scopus. Search strings were gradually adjusted due to an initially rather substantial number of results. If a search string resulted in studies that matched the title of this paper, that search term was documented so that in the future it would be possible to reconstruct how the literature included in this paper was originally retrieved. Duplicates within the title screening were marked as such, so that only literature that had not yet been recorded progressed to abstract screening. Studies included by title were screened by abstract for the beforementioned inclusion criteria. If they met the inclusion criteria to the greatest extent possible and were provided via open access, the full text was read and checked for the inclusion and exclusion criteria (see: 3.2.1).

Using this study selection progress, 5 studies were found. Another 2 studies were found by screening primary literature of reviews regarding this topic (see: p.17). Each of the studies was found in at least two of the databases. Since Google Scholar was the first database searched for matching literature and search results in Google Scholar are often cross-linked to other databases, one search string returned all 5 of the studies found via this literature search. This particular search string, the found studies, and the databases on which these studies were also found, are shown in the following table:

Search string used on Google Scholar: *transgender youth bone health GnRH analogue and gender affirming hormones*	
Usable Study found on Google Scholar	**Databases the studies were also found on**
Bone Development in Transgender Adolescents Treated With GnRH Analogues and Subsequent Gender-Affirming Hormones (Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020)	• PubMed
Effect of pubertal suppression and cross-sex hormone therapy on bone turnover markers and bone mineral apparent density (BMAD) in transgender adolescents (Vlot, et al., 2017)	• PubMed • Scopus
Pubertal Suppression, Bone Mass, and Body Composition in Youth With Gender Dysphoria (Navabi, Tang, Khatchadourian, & Lawson, 2021)	• PubMed • Medline

Search string used on Google Scholar: *transgender youth bone health GnRH analogue and gender affirming hormones*	
Usable Study found on Google Scholar	**Databases the studies were also found on**
The effect of GnRH analogue treatment on bone mineral density in young adolescents with gender dysphoria: findings from a large national cohort (Joseph, Ting, & Butler, 2019)	• PubMed • Medline
Short-term outcomes of pubertal suppression in a selected cohort of 12 to 15 year old young people with persistent gender dysphoria in the UK (Carmichael, et al., 2021)	• PubMed • Medline

Table 4: *Dublicates and Databases*

As mentioned in 3.2.1, reviews are not included in this paper. But if by their title and abstract reviews appeared to contain relevant primary literature, a search of the reference list (i.e., screening of primary literature) was conducted. This resulted in further, possibly suitable studies, which also went through the process described above (checked for title, abstract, Open Access and full text). In most cases, the reviews could be found on multiple of the databases used to conduct the literature research for this paper. In addition, a considerable number of these reviews often use the same primary sources. Therefore, the studies generated from reviews were not assigned to a specific search string. The screening of primary literature ultimately led to another 2 studies included in this work:

- *Bone Mass in Young Adulthood Following Gonadotropin-Releasing Hormone Analog Treatment and Cross-Sex Hormone Treatment in Adolescents With Gender Dysphoria* (Klink, Caris, Heijboer, van Trotsenburg, & Rotteveel, 2015)
- *Physical changes, laboratory parameters, and bone mineral density during testosterone treatment in adolescents with gender dysphoria* (Stoffers, de Vries, & Hannema, 2019)

A total of 7 studies were found through the previously described literature search and screening of the primary literature of reviews and included in the present work. The flowchart on the next page illustrates the search in brief:

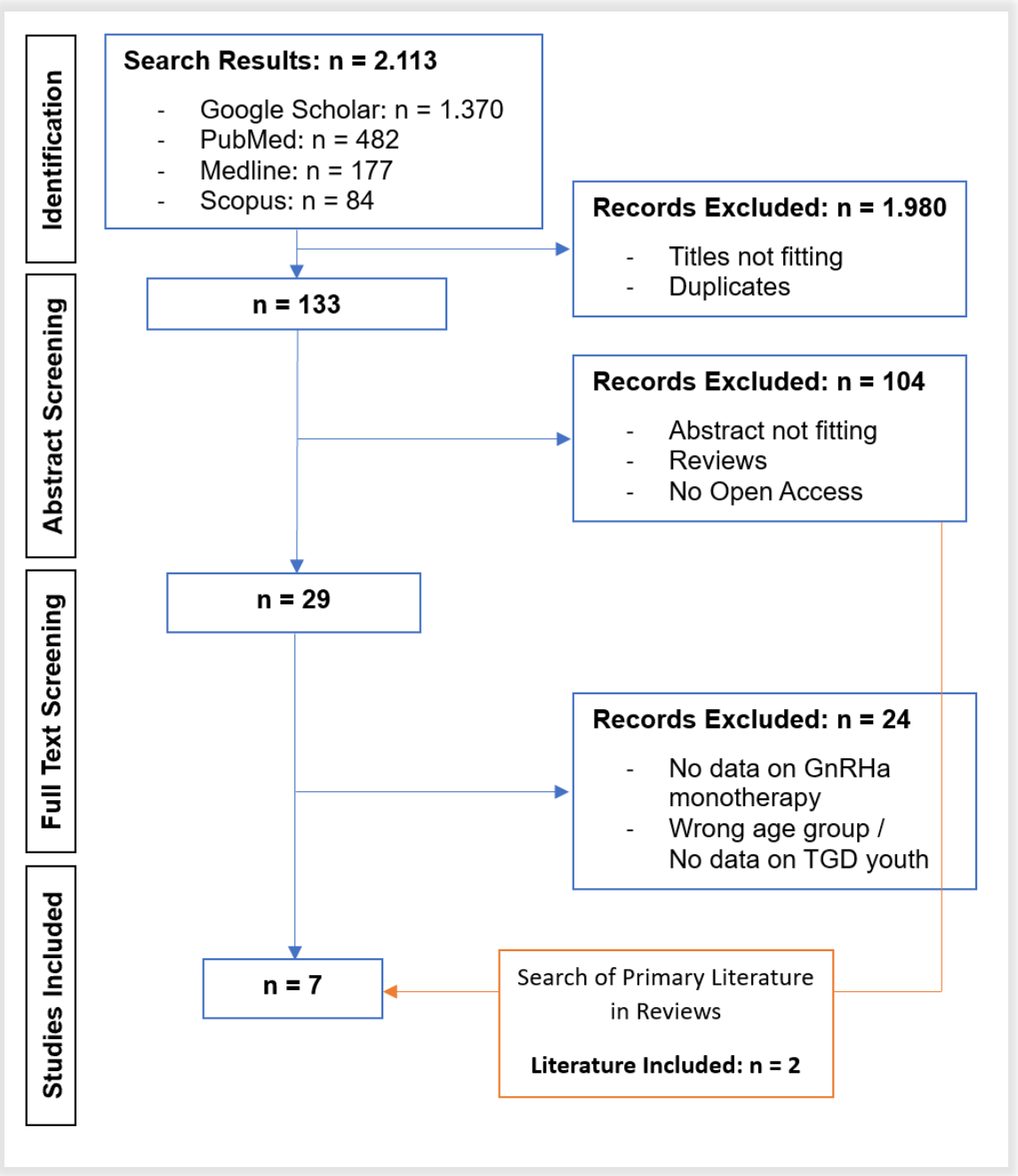

Figure 1: *Flowchart of the Systematic Research*

3.2.3. Quality Assessment

The quality of the studies was assessed according to the *Quality Assessment Tool for Before-After (Pre-Post) Studies With No Control Group* (short: QAT) provided by the NIH (NIH (National Heart, Lung and Blood Institute), 2021). After reviewing the various study quality assessment tools provided by the NIH, it was found that the assessment of all studies with the selected QAT was feasible. The key factor in this decision was that the studies, while considering z-scores, did not include control groups of TGD adolescents with GD who were not treated with GnRHa.

The selected QAT is presented in the form of a checklist with twelve questions (NIH (National Heart, Lung and Blood Institute), 2021). These questions can be answered with "Yes", "No" or "Cannot Determine / Not Applicable / Not Reported". The checklist was adopted by the NIH, and a grading system based on a point system was established to assess the quality of the studies:

- Questions answered with "Yes" get +1 point.
- Questions answered with “No” get -1 point.
- Questions answered with "Cannot Determine / Not Applicable / Not Reported" get 0 points.

In total, a study can receive a maximum of 12 and a minimum of 0 points. If a study got a score below zero in the presented scoring system, it was automatically considered as a 0-point study. The grades 1 - 6 were awarded in 2-point increments, with 1 being considered "very good" and 6 being "very poor".

The completed checklists of the quality assessments of the individual studies can be found in the Appendix (Tables 7 - 13). For the blank versions and instructions for completing the checklist, please refer to the NIH website (NIH (National Heart, Lung and Blood Institute), 2021). The results of the quality assessments are included in the tabular summary of the studies (see: 4.1).

4. Presentation of Studies and Findings

In this chapter, the studies included in this paper are analyzed. First, an overview of the studies is given in tabular form. The findings of all 7 studies are then narratively explained. Finally, the most important findings are summarized.

4.1. Tabular Overview of the Studies

The following table shows the most important aspects of the reviewed studies as well as the quality rating they got according to the QAT. It is important to note that the rating applies to the entire study. Therefore, it may be influenced by aspects of the studies which are not included in the table nor the narrative explanation, seeing as they were not relevant to the research question of this paper. This is especially true for studies primarily related to CSH therapy, as they often address changes in bone parameters during GnRHa monotherapy only in single sentences or short sections.

Studies are listed chronologically, by year of publication. To make similarities and differences apparent at first glance, the tabular presentation does not use the terminology for TGD youth as used in the studies; instead, the terms *AFAB* for transgender boys and *AMAB* for transgender girls – which are generally defined in 2.1.1 – are used.

Title, Author, Publishing Year and Country	Structure of the Study	Study Population (Data at Start of GnRHa)	Findings Relevant to this Paper #	Quality of the Study*
Bone Mass in Young Adulthood Following Gonadotropin-Releasing Hormone Analog Treatment and Cross-Sex Hormone Treatment in Adolescents with Gender Dysphoria Authors: Klink, D., Caris, M., Heijboer, A., van Trotsenburg, M., Rotteveel, J. Publishing Year: 2015 Publishing Country: The Netherlands	Study Design: Longitudinal observational Survey Instrument: DXA scan - at LS and FN - using Hologic QDR 4500 Bone Health Parameters: - aBMD - aBMD z-scores - BMAD - BMAD z-scores Median GnRHa Treatment Duration[!]: - In AFAB: 1.3 - In AMAB: 1.5 Type of GnRHa used: Triptorelin, s.c.	**Smaller Cohort** AFAB: - n = 19 - median age[!]: 15.0 - mean bone age!: 15.0 - median Tanner P: 5 - median Tanner B: 4 AMAB: - n = 15 - median age[!]: 14.9 - mean bone age[!]: 15.5 - median Tanner P: 5 - median Tanner B: 5	**Smaller Cohort** aBMD: - decrease at LS and FN in AFAB aBMD z-scores: - decrease at LS and FN in AFAB - n.s. decrease at LS in AMAB → scores below population mean at start of GnRHa BMAD: - decrease at LS in AFAB BMAD z-scores: - decrease at LS and FN in AFAB	3

Title, Author, Publishing Year and Country	Structure of the Study	Study Population (Data at Start of GnRHa)	Findings Relevant to this Paper #	Quality of the Study*
	Additional Data on Larger Cohort: Median GnRHa Treatment Duration[!]: - In AFAB: 1.6 - In AMAB: 1.8	**Additional Data on Larger Cohort**: AFAB - n = 48 - median age[!]: 15.1 - mean bone age[!]: 15.0 - median Tanner P: 5 - median Tanner B: 4 AMAB - n = 30 - median age[!]: 14.1 - mean bone age!: 14 - median Tanner P: 3.5 - median Tanner B: 4.5 **This cohort includes the smaller cohort shown above!**	**Additional Data on Larger Cohort**: aBMD: - decreased at LS and FN in AFAB aBMD z-scores: - decreased at LS in AMAB - decreased at LS and FN in AFAB BMAD: - decreased at FN in AMAB BMAD z-scores: - decreased at LS in AMAB - decreased at LS and FN in AFAB	

Title, Author, Publishing Year and Country	Structure of the Study	Study Population (Data at Start of GnRHa)	Findings Relevant to this Paper #	Quality of the Study*
Effect of pubertal suppression and cross-sex hormone therapy on bone turnover markers and bone mineral apparent density (BMAD) in transgender adolescents Authors: Vlot, M., Klink, D., den Heijer, M., Blankenstein, M., Rotteveel, J., Heijboer, A. Publishing Year: 2016 Publishing Country: The Netherlands	Study Design: Retrospective Survey Instrument: DXA scan - at LS and FN of the non-dominant hip - using Hologic QDR 4500 Bone Health Parameters: - BMAD - BMAD z-scores - BTM (P1NP, osteocalcin, 1CTP) Median GnRHa Treatment Duration: --- Type of GnRHa used: Triptorelin, s.c.	AFAB: - n = 42 - median age[!]: 15.1 - mean bone age!: 15 - median Tanner P: 5 - median Tanner B: 5 AMAB: - n = 28 - median age[!]: 13.5 - mean bone age[!]: 13.5 - median Tanner P: 3 - median Tanner G: 3	BMAD: - decrease at LS and FN in old AFAB BMAD z-scores: - decrease at LS and FN in all AFAB - decrease at LS and FN in young AMAB → scores in old and young AMAB below zero at start of GnRHa P1NP: - decrease in young AFAB and all AMAB 1CTP: - decrease in young AFAB and young AMAB	3

Title, Author, Publishing Year and Country	Structure of the Study	Study Population (Data at Start of GnRHa)	Findings Relevant to this Paper #	Quality of the Study*
Physical changes, laboratory parameters, and bone mineral density during testosterone treatment in adolescents with gender dysphoria Authors: Stoffers, I., de Vries, M., Hannema, S. Publishing Year: 2019 (June) Publishing Country: The Netherlands	Study Design: Retrospective Survey Instrument: DXA scan - at LS, left and right hip (neck area) - using Hologic Discovery A Bone Health Parameters: - BMD - BMD z-scores - BMAD - BMAD z-scores - Vitamin D Median GnRHa Treatment Duration: 8 months Type of GnRHa used: triptorelin, s.c.	**Study just looked at AFAB (= transgender males); data on GnRHa monotherapy on only 18 out of 62** AFAB: - n = 18 - median age[!]: 16.5 - mean bone age[!]: --- - most subjects were in Tanner B5 - most subjects were post menarche	BMD: - significant decrease at LS, left and right hip BMD z-scores: - significant decrease at LS, left and right hip Vitamin D: - at start of GnRHa: 74% of subjects had Vitamin D insufficiency. - even with supplementation, 37% of subjects still had Vitamin D insufficiency at end of GnRHa monotherapy	2

Title, Authoring Year and Country	Structure of the Study	Study Population (Data at Start of GnRHa)	Findings Relevant to this Paper #	Quality of the Study*
The effect of GnRH analogue on bone mineral density in young adolescents with gender dysphoria: findings from a large national cohort Authors: Joseph, T., Ting, J., Butler, G. Publishing Year: 2019 (July) Publishing Country: United Kingdom	Study Design: Retrospective Review + Complete Longitudinal Analysis Survey Instrument: DXA scan - at LS and FN - using Hologic Discovery QDR series model 010-1549 Bone Health Parameters: - BMD (aBMD) - BMD z-scores - BMAD - BMAD z-scores GnRHa Treatment Duration: At least one year Type of GnRHa used: ---	**Study just looked at subjects of white Caucasian origin.** **Complete Longitudinal Analysis of n = 31 subjects:** AFAB: - n = 21 - median age[!]: 12.9 - mean bone age: --- - majority was post menarche AMAB: - n = 10 - median age[!]: 13.0 - mean bone age: --- - 57%: Tanner G2 – 3 - 43%: Tanner G4 – 5	aBMD z-scores: - lower in AFAB at LS and FN - most significant drop after 1 year in AFAB and AMAB BMAD z-scores: - lower in AFAB at LS and FN - most significant drop after 1 year in AFAB and AMAB → **the bone mass accrual does not happen according to age**	2

Title, Author, Publishing Year and Country	Structure of the Study	Study Population (Data at Start of GnRHa)	Findings Relevant to this Paper #	Quality of the Study*
		Larger Cohort, looking at just the first year of GnRHa treatment: AFAB: - n = 39 - median age[!]: 12.6 AMAB: - n = 31 - median age[!]: 13.2 **This cohort includes the smaller cohort shown above!**	**Larger Cohort, looking at just the first year of GnRHa treatment**: aBMD z-scores: - decrease at LS and FN in AMAB and AFAB BMAD z-scores: - decrease in AMAB and AFAB	

Title, Author, Publishing Year and Country	Structure of the Study	Study Population (Data at Start of GnRHa)	Findings Relevant to this Paper #	Quality of the Study*
Bone Development in Transgender Adolescents Treated With GnRH Analogues and Subsequent Gender-Affirming Hormones Authors: Schagen, S. E. E., Wouters, F. E., Cohen-Kettenis, P. T., Gooren, L. J., Hannema, S. E. Publishing Year: 2020 Publishing Country: The Netherlands	Study Design: Observational Prospective Survey Instrument: DXA scan - at LS, FN, and whole body - using Hologic QDR 4500 Bone Health Parameters: - aBMD (BMD) - aBMD z-scores - BMAD - BMAD z-scores - P1NP / P3NP - Osteocalcin - 1CTP GnRHa Treatment Duration: About 2 years, some individuals up to 4 years Type of GnRHa used: Triptorelin, i.m.	AFAB: - n = 70 - median age[!]: 14.5 - mean bone age: --- - mostly late puberty (Tanner 4-5) AMAB: - n = 51 - median age[!]: 14.1 - mean bone age: --- - mostly late puberty (Tanner 4-5)	aBMD z-scores: - decrease at LS, FN, and whole boy in earls- and late-pubertal AMAB - decrease at LS, FN, and whole boy in earls- and late-pubertal AFAB BMAD: - decrease at LS and FN in late-pubertal AFAB - decrease at FN in early-pubertal AFAB - decrease at FN in late-pubertal AMAB BMAD z-scores: - lower at LS and FN in AMAB than AFAB - decreased at LS in early- and late-pubertal AMAB and AFAB - decreased at FN in early- and late-pubertal AFAB	2

Title, Author, Publishing Year and Country	Structure of the Study	Study Population (Data at Start of GnRHa)	Findings Relevant to this Paper #	Quality of the Study*
			<u>BMAD z-scores</u>: - decreased at FN in late-pubertal AMAB - some individuals showed BMAD z-scores below -2 <u>P1NP, P3NP, Osteocalcin, 1CTP</u>: - decrease in early- and late-pubertal AMAB - decrease in early-pubertal AFAB - decrease mostly in the 1st year of treatment - only P3NP and 1CTP: decrease in late-pubertal AFAB	

Title, Author, Publishing Year and Country	Structure of the Study	Study Population (Data at Start of GnRHa)	Findings Relevant to this Paper #	Quality of the Study*
Short-term outcomes of pubertal suppression in a selected cohort of 12 to 15 year old young people with persistent gender dysphoria in the UK Authors: Carmichael, P., Butler, G., Masic, U. Cole, T. J., De Stavola, B. L., Davidson, S., Skageberg, E. M., Khadr, S., Viner, R. M. Publishing Year: 2021 (February) Publishing Country: United Kingdom	Study Design: Uncontrolled, Observational Prospective Survey Instrument: DXA scan - at LS and hip - using Hologic Discovery QDR series model 010-1549 Bone Health Parameters: - aBMD (BMD) - aBMD z-scores - BMC Median GnRHa Treatment Duration: 31 months Type of GnRHa used: Triptorelin, i.m.	- Median age at consent[!]: 13.6 - 89% of white ethnicity AFAB: - n = 19 - Puberty Stage 4 or 5 - 79% post menarche AMAB: - n = 25 - Puberty Stage 3 **Age at consent does not necessarily mark the start of GnRHa!** **Puberty Stage** **= Tanner P + Tanner B / Tanner P + Tanner G**	aBMD: - increase at LS after 2 years aBMD z-scores: - decrease at LS and FN after 1 and 2 years, but not at 3-year mark BMC: - increase at LS after 2 and 3 years	2

Title, Author, Publishing Year and Country	Structure of the Study	Study Population (Data at Start of GnRHa)	Findings Relevant to this Paper #	Quality of the Study*
Pubertal Suppression, Bone Mass, and Body Composition in Youth With Gender Dysphoria Authors: Navabi, B., Tang, K., Khatchadourian, K., Lawson, M. L. Publishing Year: 2021 (October) Publishing Country: Canada	Study Design: Retrospective Survey Instrument: DXA scan - at LS, left total hip and total body less head - using Lunar Prodigy system Bone Health Parameters: - aBMD (BMD) - aBMD z-scores - BMC - BMAD - BMAD z-scores - vitamin D - calcium GnRHa Treatment Duration: Follow-up time of min. 18 months Type of GnRHa used: Leuprolide acetate, i.m.	AFAB: - n = 80 - median age[!]: 15.2 - mean bone age: --- - mostly Tanner 4-5 AMAB: - n = 36 - median age[!]: 15.4 - mean bone age: --- - mostly Tanner 4-5	aBMD z-scores: - decrease at LS, left total hip and total body less head in AFAB and AMAB - less pronounced changes in younger AFAB - some individuals showed a drop > 2 in LS aBMD z-score BMAD z-scores: - decrease at LS in AFAB BMC z-scores: - decrease in both AFAB and AMAB Vitamin D: - deficiency in 17.6% at baseline - insufficiency in 37.6% at baseline - sufficiency in 87.3% after 2 years of supplementation	1

Title, Author, Publishing Year and Country	Structure of the Study	Study Population (Data at Start of GnRHa)	Findings Relevant to this Paper #	Quality of the Study*
			Calcium: low at baseline	

Table 5: *Overview of the Studies*

Legend:
- # referring to changes in the studies' bone health parameters during monotherapy, unless explicitly stated otherwise, these are mostly changes that the authors described as "significant"
- * quality according to the QAT (NIH (National Heart, Lung and Blood Institute), 2021)
- ! in years
- --- no data found in the study
- LS = Lumbar Spine
- FN = Femoral Neck
- n.s. = not significant
- s.c. = subcutaneously
- i.m. = intramuscular
- Bone Health Parameters: Please refer to 2.1.3 for explanations on the parameters

4.2. Narrative Compilation of the Studies´ Findings

The study results on bone health during GnRHa monotherapy are presented in more detail below. For a short overview on the studies, please refer to the table under 4.1. To structure the systematic presentation and summary of the studies, multiple key points will be addressed. They can be organized as follows:

- The impact of GnRHa on BMD
- The impact of GnRHa on BMAD
- The impact of GnRHa on BMC
- The impact of GnRHa on BTM
- Calcium and Vitamin D in TGD youth

Given the complexity of the topic and the study situation on the individual parameters, it is important to maintain an overview so that the relationships between GnRHa and the bone health parameters studied become as clear as possible. To make the main section easier to read and compare, it is structured as uniformly as possible. To achieve this, the studies are not listed and summarized individually; instead, general information on the studies is summarized and presented in chronological order. Then, the results on the two main parameters (BMD and BMAD) are listed. To do so, a description of the measurement tools used in each study is provided first. Thereafter, the studies' results are presented. This presentation is sorted by study and in the order in which they are mentioned in the description of the measurement instruments. Finally, the studies' results on the other parameters (BMC, BTM, vitamin D and calcium) are portrayed. A discussion of the significance of the findings can be found in 5.1.

4.2.1. Important Insertion

To achieve full transparency, it needs to be highlighted that all terms of TGD youth discussed in 2.1.1 are only defined as such when they are not applied to any of the included studies. Since most referenced studies lack a precise definition of “TGD people” the formerly specified definitions cannot necessarily be transferred to the wording used in the studies. Therefore, the wording of the currently cited study will be employed. The same is true for the use of “birth-registered” and “natal” sex. Admittedly, this leads to the main section sounding somewhat obscure at times, since a multitude of terms were used. Nevertheless, this approach was considered useful to avoid assigning definitions to foreign studies that were not intended as such by the authors. To check which study uses which terms, please refer to Table 6 in the Appendix.

4.2.2. General Information

In this chapter, the study populations of each study at the start of GnRHa monotherapy will be described. In addition – if the information is given – the GnRHa used and information such as the duration of use are given.

Most studies only looked at one main study population. However, two studies provide further data on slightly bigger study populations, in which each main population is included, respectively (Klink, Caris, Heijboer, van Trotsenburg, & Rotteveel, 2015; Joseph, Ting, & Butler, 2019). For more information on how the main and the expanded study populations of these two studies are made up, please refer to the table in 4.1. Another study collected data on n = 172 participants (including n = 2 nonbinary youth), though they only analyzed a subgroup of n = 116 of this study population for changes due to GnRHa (Navabi, Tang, Khatchadourian, & Lawson, 2021).

Three of the studies included in this paper applied the GnRHa triptorelin at a dosage of 3.75 mg subcutaneously, every four weeks (Klink, Caris, Heijboer, van Trotsenburg, & Rotteveel, 2015; Vlot, et al., 2017; Stoffers, de Vries, & Hannema, 2019). Two other studies used the same GnRHa at the same dosage, though they applied it intramuscularly every four weeks (Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020; Carmichael, et al., 2021). In one of those studies, the first doses were "administered with a 3-week interval followed by injections every 4 weeks to suppress endogenous sex steroid production" (Carmichael, et al., 2021, p. e4253). For one study, no information could be found on which exact GnRHa was used to treat GD (Joseph, Ting, & Butler, 2019). Only the latest study names a different kind of GnRHa: leuprolide acetate, which was administered intramuscularly, "starting with 3 doses of 7.5 mg […] every 4 weeks, followed by 11.25 mg […] every 12 weeks" (Navabi, Tang, Khatchadourian, & Lawson, 2021, p. 2).

Besides one study, which only looked at transgender males (Stoffers, de Vries, & Hannema, 2019), all other studies conducted data on both AFAB and AMAB. More data exists on transgender boys (= AFABs). Including all the supplemental data, information for a total of n = 316 transgender boys and n = 201 transgender girls treated with GnRHa monotherapy is given across all herein included studies. There is only one study, which conducted more data on birth-registered males (= transgender girls) than birth-registered females (Carmichael, et al., 2021).

The patients were of different ages and Tanner Stages across all studies. 5 out of 7 studies give information on the Tanner Stages participants were in, though some only specify the required minimum of Tanner Stage 2 for start of GnRHa treatment (Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020). Others primarily give information on whether AFABs had had their menarche before GnRHa treatment

(Stoffers, de Vries, & Hannema, 2019; Joseph, Ting, & Butler, 2019; Carmichael, et al., 2021). The median ages and the puberty stages (Tanner Stages, pre-/post-menarche) most participants were in in each study, are described below. First, information is given on AMABs and then AFABs.

At the start of GnRHa treatment, the youngest AMABs (= transgender girls) considered were at a median age of 13.0 years with 57% being in Tanner G2-3 and 43% in Tanner G4-5 (Joseph, Ting, & Butler, 2019), and the oldest were a median of 15.4 y/o with 80% in Tanner Stages 4-5 (Navabi, Tang, Khatchadourian, & Lawson, 2021). The range in between reached from to 13.5 y/o and mostly Tanner G3 (Vlot, et al., 2017), to 14.1 y/o and most being Tanner 4-5 (Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020), to 14.9 y/o with most subjects being Tanner Stages P and G 5 – plus supplemental data on transwomen of 14.1 years and Tanner Stages P3.5 and G4.4 (Klink, Caris, Heijboer, van Trotsenburg, & Rotteveel, 2015). In Carmichael et al. (2021), the age of consent to treatment is 13.4 y/o in birth-registered males; however, age of consent does not necessarily mean start of GnRHa treatment. In this study, Tanner Stages of both the P-Scale and the G-Scale are combined to create “Puberty Stages”. 68% of birth-registered males were in Puberty Stage 3. (Carmichael, et al., 2021)

Most data on AFABs within the studies is collected in older subjects than in AMABs. At the start of treatment, in the study with data on the youngest subjects out of all, trans-boys had a median age of 12.9 years, though the majority was post-menarche and considered to be in later stages of puberty (Joseph, Ting, & Butler, 2019). The oldest subjects were studied by Stoffers et al. (2019) who only looked at transgender males, which were a median age of 16.5 y/o at start of GnRHa and most were in Tanner B5 and post-menarche (Stoffers, de Vries, & Hannema, 2019). The range between oldest and youngest study participants stretches from a median of 14.5 y/o with most patients being in late puberty at Tanner 4-5 (Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020), to 15.0 y/o with most being in Tanner P5 and B4 (Klink, Caris, Heijboer, van Trotsenburg, & Rotteveel, 2015) and 15.1 y/o with the majority being in Tanner B5 (Vlot, et al., 2017), to 15.2 y/o with 90% in Tanner Stages 4-5 (Navabi, Tang, Khatchadourian, & Lawson, 2021). Again, Carmichael et al. (2021) specified the age of consent, which in this case was 13.9 y/o. Most of these birth-registered females were in later Puberty Stages (4-5) and post-menarche. (Carmichael, et al., 2021)

Though not all studies clarify the ethnicity of their study subjects, Joseph et al. (2019) only collected data on “white Caucasian” patients (Joseph, Ting, & Butler, 2019) and in Carmichael et al. (2021), eighty-nine percent of participants were also of white ethnicity (Carmichael, et al., 2021).

The median duration of the treatment differs; from 8 months (Stoffers, de Vries, & Hannema, 2019), to 31 months (Carmichael, et al., 2021), to 1.6 - 1.8 years (in transmen and transwomen, respectively) (Klink, Caris, Heijboer, van Trotsenburg, & Rotteveel, 2015), to 2 years (Joseph, Ting, & Butler, 2019), to up to 4 years (Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020). Two studies did not state the median duration patients were treated with GnRHa, though monotherapy was usually stopped at approximately 16 y/o (Vlot, et al., 2017; Navabi, Tang, Khatchadourian, & Lawson, 2021).

4.2.3. BMD Measurement Techniques

All studies considered for this work compiled data on the impact of GnRHa on bone mineral density (= BMD; also: aBMD) using DXA scans. The exact kind of densitometer, however, varies. The resulting measurements of the DXA scans are reported as g/cm^2 across all studies. The following paragraphs describe the studies' measuring instruments and measuring sites for determining BMD. Furthermore, the studies' reference populations for determining BMD z-scores are indicated.

Klink, et al. (2015), **Schagen, et al.** (2020) as well as **Vlot, et al.** (2017) use the *Hologic QDR 4500* as bone densitometer. All three studies take BMD measurements at the LS as well as the FN of the non-dominant hip. Schagen et al. (2020) also include whole body measurements. While Klink et al. (2015) indicate the exact examined area of the lumbar spine with L1 - L4, no further specification can be found in Schagen et al. (2020) and Vlot et al. (2019).

Klink et al. (2015) referenced their aBMD z-scores to the *National Health and Nutrition Examination Survey* (short: NHANES) (Ward, Ashby, Roberts, Adams, & Zulf Mughal , 2007). They calculated the z-scores according to natal sex, age, and ethnicity.

Schagen et al. (2020) also used the NHANES to calculate BMD z-scores, though they did so based only on age and sex.

Although Vlot et al. (2017) obtained BMD using DXA scans, these values are not discussed further in the study results, instead detailing BMAD to correct for height and height gain. Information on this can be found in the section on BMAD.

A similar densitometer, the *Hologic Discovery QDR series model 010-1549*, is used by **Joseph, et al.** (2019) and **Carmichael, et al.** (2021). Both take measurements at the LS (L1 - L4), but whereas Joseph et al. (2019) took scans of the FN, Carmichael et al. (2021) – unlike the previous studies – assessed the total hip.

As mentioned above (see: 4.1 and 4.2.2), Joseph et al. (2019) conducted two study populations. In the smaller study (n = 31 subjects) all participants received DXA scans at LS and FN, initially at baseline and then every 12 months with three scans total; in

the larger study population, data from the baseline scan were compared to those after just one year (Joseph, Ting, & Butler, 2019). BMD z-scores were “produced by the internal software associated with this commercially available equipment” (Joseph, Ting, & Butler, 2019, p. 2) and adjusted for birth sex and age.

In Carmichael et al. (2021), BMD z-scores were adjusted for age, birth-registered sex and adjusted for height. It is important to note that no data was reported for sample sizes below n = 8 (Carmichael, et al., 2021, p. 10).

Stoffers et al. (2019) use the *Hologic Discovery A scanner* to determine BMD of the LS as well as the “neck area of the left and right hip” (Stoffers, de Vries, & Hannema, 2019, p. 1461) at two times relevant to this paper: Once before GnRHa treatment started and once before the start of CSH therapy. To calculate BMD z-scores in participants under 16 y/o, the authors referenced data on the birth-assigned sex (= female) from the *Bone Mineral Density in Childhood Study* (Kalkwarf, et al., 2007).

Navabi et al. (2021) are the only ones which did not use a Hologic densitometer for DXA scans; instead, they used the *Lunar Prodigy system*. They conducted aBMD for the lumbar spine (L2 - L4), the left total hip (short: LTH) and additionally the total body less head (short: TBLH). Based on birth-assigned sex, age and ethnicity, the authors calculated aBMD z-scores initially at baseline and then annually. Only in exceptional cases (= a significant drop of more than 1 standard derivation in LS aBMD z-scores or LS aBMD z-scores below 2 standard derivations), BMD was recorded more frequently. (Navabi, Tang, Khatchadourian, & Lawson, 2021, p. 2)

4.2.4. Changes in BMD and BMD Z-Scores

The findings of the studies regarding changes in BMD and BMD z-scores as well as the data collected within are presented in the following. To maintain an overview on the findings of the individual studies, the study results on the impact of GnRHa on BMD and BMD z-scores are listed in the same order as they are mentioned in the previous chapter.

Klink et al. (2015) compare data on aBMD and aBMD z-scores at the start of GnRHa treatment to data at the start of CSH therapy, which marks the end of GnRHa monotherapy. They set their level of significance at $p < 0.017$ (Klink, Caris, Heijboer, van Trotsenburg, & Rotteveel, 2015, p. E272 (Table 2)). They give information about transwomen and transmen – in the mentioned order the results are explained hereafter. In the case of transwomen, they present data on aBMD of n = 12 at the LS and n = 14 at the FN. Information on the aBMD z-scores of the same group is available for n = 11 at the LS and n = 6 at the FN. The hereby recorded data shows that aBMD in transwomen “did not change during GnRHa monotherapy” (Klink, Caris, Heijboer, van

Trotsenburg, & Rotteveel, 2015, p. E273) on either measuring site. Transwomen's "LS aBMD z-score was below the population mean at the start of treatment" (Klink, Caris, Heijboer, van Trotsenburg, & Rotteveel, 2015, p. E273). Although there were changes in transwomen's aBMD z-scores, these changes were deemed "not significant" by the authors (Klink, Caris, Heijboer, van Trotsenburg, & Rotteveel, 2015, p. E273). The supplemental data included in this study (Klink, Caris, Heijboer, van Trotsenburg, & Rotteveel, 2015, p. E271 "Supplemental Data") gives information, inter alia, on LS aBMD for n = 25 and on FN aBMD for n = 27 transwomen – the subjects previously discussed are included herein. For LS aBMD z-scores, there are data for a same sample size of n = 25, but for the FN aBMD z-scores, there are data for only n = 17 transwomen. This larger cohort shows a slight increase of LS aBMD in transwomen. However, aBMD z-scores of both the LS and the FN showed a significant decrease during GnRHa monotherapy ($p < 0.001$ and $p = 0.001$, respectively). (Klink, Caris, Heijboer, van Trotsenburg, & Rotteveel, 2015, pp. "Supplemental Data", p. 3 and Table 4) In the smaller study population, aBMD data of the LS and the FN are available for n = 18 transmen, each. The sample size for LS aBMD z-scores was n = 18 and for FN aBMD z-scores it was n = 13. At both sites, aBMD decreased significantly (aBMD at LS: $p = .006$; aBMD at FN: $p = .005$) in transmen (Klink, Caris, Heijboer, van Trotsenburg, & Rotteveel, 2015, p. E272 (Table 2)). Consequently, aBMD z-scores at both sites also showed a significant decrease (aBMD z-score at LS: $p < .001$; aBMD z-score at FN: $p = .001$) in this group (Klink, Caris, Heijboer, van Trotsenburg, & Rotteveel, 2015, p. E272 (Table 2)). While the supplemental data shows a different development of aBMD and aBMD z-scores in transwomen, for transmen the numbers of the larger cohort are in line with the ones from the cohort just described. (Klink, Caris, Heijboer, van Trotsenburg, & Rotteveel, 2015, pp. "Supplemental Data", p. 3)

Schagen et al. (2020) divided their study samples of transgirls and transboys into "early" and "late" pubertal. *Early pubertal* classifies adolescents in Tanner Stages 2-3, whereas *late pubertal* applies to study participants in Tanner Stages 4-5 (Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020, p. e4254). The level of significance was set at $p < 0.05$ (Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020, p. e4246 (Table 2)). Most changes during a two-year GnRHa treatment were deemed significant ($p < 0.05$) with no further specification of p-values. Over the course of two years of GnRHa monotherapy, the aBMD of the LS and FN and BMD of the whole body showed a slight increase in early-pubertal transgirls and -boys. While changes in whole body BMD and FN aBMD in late-pubertal transgirls were deemed non-significant, they too showed a slight increase in LS aBMD. In late-pubertal transboys, on the other hand, aBMD decreased at all three measuring sites. Z-scores

showed a significant fall in all four groups (early- and late-pubertal transgirls and -boys) at all three sites. (Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020, p. e4256 (Table 2)) While both transboys and transgirls had aBMD values “within normal range” at start of GnRHa, “transgirls had z-scores well below zero” (Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020, p. e4259). Some of the younger (early-pubertal) participants “were treated for up to 4 years” (Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020, p. e4256), but in this case only aBMD and aBMD z-scores of the LS and FN were recorded. During three years of GnRHa treatment, n = 4 transgirls showed a slight increase of LS aBMD over the first two years with a following stagnation, while FN aBMD showed a slow but steady increase over the course of three years. While transboys (n = 11) also showed increasing and then stagnant LS aBMD values, FN aBMD did not change during the three years. In both, transgirls and transboys, these changes are not significant. While aBMD seemed to remain mostly stable, aBMD z-scores decreased at both sites and in both groups. However, only the FN aBMD z-score in transgirls and the LS aBMD z-score in transboys showed significant changes ($p = 0.007$ and $p = 0.008$, respectively). (Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020, p. e4258 (Table 3)) It is important to note that the referenced sample sizes regarding the data on transgirls and -boys treated for more than two years in-text and in the cited table are reversed. As the table provides more detailed information, data from the table was used for the analysis in the paper at hand. In addition, the results in-text ("z-scores on the other hand decline" (Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020, p. p. e4256)) and the table (see previous explanations) are not consistent with the discussion, which states: “In most individuals with prolonged (3-4 years) GnRHa treatment, no further decrease in aBMD z-scores was obtained in the last year, suggesting that z-scores might stabilize.” (Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020, p. e4260)

Joseph et al. (2019) analyzed data of n = 31 subjects with two years of GnRHa treatment and DXA scans at baseline and then annually, and of a larger cohort of n = 70. The larger cohort includes the subjects with three DXA scans, but also an additional n = 39 subjects with just two DXA scans (at baseline and after one year of GnRHa). Same as in Schagen et al. (2020), $p < 0.05$ is considered “statistically significant” (Joseph, Ting, & Butler, 2019, p. 2). BMD of the LS and the FN show small fluctuations during the two-year GnRHa treatment. They are, however, insignificant enough for "no change in BMD for the lumbar spine or hip […] over the course of 3 years" (Joseph, Ting, & Butler, 2019, p. 2) to be recorded in the written summary of the results. BMD z-scores decrease significantly at both measuring sites during GnRHa treatment, with p-values between $p = 0.002$ (FN BMD z-score in transgirls), $p = 0.001$ (FN BMD z-score

in transboys) and p = 0.000 (LS BMD z-scores in both groups) (Joseph, Ting, & Butler, 2019, p. 2 (Table 1)). Regarding absolute BMD, even in the larger sample (n = 70), "no significant change […] in the spine or hip values" (Joseph, Ting, & Butler, 2019, p. 2) could be determined after one year of GnRHa treatment. Again, the z-scores decrease significantly (LS BMD z-score in transgirls: p = 0.003; FN BMD z-score in transgirls: p = 0.002; LS BMD z-score in transboys: p = 0.000; FN BMD z-score in transboys: p = 0.000) (Joseph, Ting, & Butler, 2019, p. 4 (Table 2)). The authors conclude that bone mass accrual does not happen according to the usual pattern (Joseph, Ting, & Butler, 2019, p. 2), as “GnRHa treatment is interrupting the rapidity of bone size increase” (Joseph, Ting, & Butler, 2019, p. 3).

Carmichael et al. (2021) looked at short- and medium-term outcomes of GnRHa treatment in birth-assigned males and females who – at the start of GnRHa monotherapy – were mostly in Tanner Stages 3 (birth-registered males) and 4 (birth-registered females) (Carmichael, et al., 2021, p. 11 / 26 (Table 1)). The authors set the level of significance at p = 0.003 (Carmichael, et al., 2021, p. 9 / 26). DXA scans of the LS were conducted in all n = 44 participants at baseline. Over the course of the treatment, the number of existing BMD values progressively decreases (data on n = 43 participants after 12 months; data on n = 24 participants after 24 months; data on n = 12 participants after 36 months). Less data is available for the hip than for the LS (data on n = 43 at baseline; data on n = 39 after 12 months; data on n = 22 after 24 months; data on n = 12 after 36 months). (Carmichael, et al., 2021, p. 13 / 26 (Table 3)) At baseline, age-adjusted BMD z-scores of the LS as well as the hip were within normal range. Although no changes in BMD of the LS were documented after the first 12 months of treatment, LS BMD did show an increase at the 24-month testing compared to baseline. BMD of the hip, however, did not show any changes at any time of testing. (Carmichael, et al., 2021, p. 12 / 26) Age- as well as height-adjusted BMD z-score “point-estimates fell at 12 and 14 months but not at 36 months” (Carmichael, et al., 2021, p. 12 / 26). Even though the authors demonstrate BMD of the LS increases and BMD of the hip stays stable (Carmichael, et al., 2021, p. 21 / 26) during up to four years of GnRHa treatment, they also acknowledge that this growth is slower than in peers not treated with GnRHa (thus the fall in BMD z-scores) (Carmichael, et al., 2021, p. 18 / 26).

Stoffers et al. (2019) only included transgender males and had the oldest study population with the shortest treatment period (see: 4.2.2). Although their focus was not on GnRHa monotherapy, they did report data at both the start of GnRHa therapy and the start of CSH therapy (= end of GnRHa monotherapy). The level of significance in this study was set at p < 0.05 (Stoffers, de Vries, & Hannema, 2019, p. 1462). After GnRHa

therapy, they report a significant fall in BMD and BMD z-scores at the LS and the left and right hip in n = 18 ($p < 0.001$) (Stoffers, de Vries, & Hannema, 2019, p. 1463). Again, the table does not quite match the in-text descriptions. The table shows data for n = 62 transgender males, whereas in the in-text results, changes for n = 18 transgender males are seen as significant. As the significance of the results does matter, the paper at hand references the text. Overall, it is assumed that the n = 18 transgender males are relevant to answering the research question in this paper. Therefore, only n = 18 are mentioned in the tabular summary of the studies.

Navabi et al. (2021) set their level of significance at $p < 0.0011$, "to correct for multiple testing" (Navabi, Tang, Khatchadourian, & Lawson, 2021, p. 5 (Table 3)). Raw data for BMD of the LS and the LTH is provided at baseline, otherwise, they only provide data on aBMD z-scores. Overall, baseline values for aBMD of the LS and LTH are lower in transgender females compared to the transgender males of the study (Navabi, Tang, Khatchadourian, & Lawson, 2021, p. 3). With $p < 0.001$, aBMD z-scores at all three measuring sites of this study (LS, LTH, TBLH) showed a significant decrease in both transgender males and transgender females (Navabi, Tang, Khatchadourian, & Lawson, 2021, p. 5 (Table 3)). Since their study primarily looked at transgender adolescents in Tanner Stages 4 and 5, "the degree of BMD changes after puberty suppression in early puberty is not known" (Navabi, Tang, Khatchadourian, & Lawson, 2021, p. 5). However, if transgender males started GnRHa treatment at an earlier stage of puberty, changes in LS aBMD z-score seemed "less pronounced" (Navabi, Tang, Khatchadourian, & Lawson, 2021, p. 4). Subjects got further testing (via lateral spine radiograph) if they showed certain changes in values in their scans. This was the case for n = 20 transgender males, who showed a LS aBMD z-score drop > 1 in their follow-up. Furthermore, low LS aBMD z-scores (< 2) at baseline were found in n = 3 birth-assigned males and n = 1 birth-assigned female. A vertebral fracture was not found in any of these imaging. (Navabi, Tang, Khatchadourian, & Lawson, 2021, p. 3)

4.2.5. BMAD Measurement Techniques

BMAD is calculated to adjust for height and height gain. The impact of GnRHa monotherapy on the bone mineral apparent density (= BMAD) is conducted by 6 of the studies reviewed in this paper. The resulting BMAD measurements are reported as g/cm^3 across all studies. Again, different reference populations are used to calculate BMAD z-scores. Which studies investigated BMAD, and which reference populations were used to calculate BMAD z-scores is listed below.

Klink et al. (2015), **Schagen et al.** (2020) and **Vlot et al.** (2017) all measured BMAD at the LS and FN, and calculated BMAD z-scores according to a *UK reference population* described by Ward et al. in 2007. Schagen et al. (2020) also used the same source

to calculate BMAD according to “LMS data from an English reference population” (Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020, p. e4254). Instead of using bone age – as they used before to classify transmen and -women as “young” and “old” – Vlot et al. (2017) used “the chronological calendar age of the transgender adolescents” (Vlot, et al., 2017, p. 12), as well as the just mentioned LMS data, to calculate BMAD.

While **Stoffers et al.** (2019) do calculate BMAD and compare LS and left FN BMAD z-scores to Ward et al. (2007), they only describe change during CSH therapy compared to before GnRHa. As this is not relevant to this paper, it will not be discussed here.

Both **Joseph et al.** (2019) and **Navabi et al.** (2021) only measure BMAD at the LS. However, they used different reference populations to estimate LS BMAD z-scores.

Referring to a more recent study (Crabtree, et al., 2017), Joseph et al. (2019) calculated BMAD z-scores according to the *ALPHABET study*, “using UK norms for Caucasian subjects” (Joseph, Ting, & Butler, 2019, p. 2) as a reference.

Navabi et al. (2021) calculated BMAD z-scores for “age-matched birth assigned gender BMAD mean and SD” (Navabi, Tang, Khatchadourian, & Lawson, 2021, p. 2) according to an older study (Kröger, Kotaniemi, Vainio, & Alhava, 1992). It is striking that – although this study was only published in 2021 – comparative data from 1992 was used to calculate the BMAD z-scores.

4.2.6. Changes in BMAD and BMAD Z-Scores

The following shows the studies’ documented changes in BMAD and BMAD z-scores during GnRHa monotherapy. The results are presented in the same order as they are mentioned in 4.2.5.

Klink et al. (2015) recorded changes in transwomen’s BMAD for n = 11 at LS and n = 12 at FN at the start of GnRHa. The number of study subjects for BMAD z-scores at those two sites is reversed. For transmen, data regarding the changes of BMAD and BMAD z-scores at start of GnRHa monotherapy was conducted for n = 18 across all examination sites. The changes raised in BMAD and BMAD z-scores of the LS and FN during GnRHa monotherapy in transwomen were deemed not significant. (Klink, Caris, Heijboer, van Trotsenburg, & Rotteveel, 2015, p. E272 (Table 2)) The supplemental data on n = 78 subjects during GnRHa monotherapy provided within this study showed significant changes in a larger group of transwomen (n = 20 for FN BMAD z-score to n = 24 for LS BMAD z-score). While LS BMAD showed no changes, FN BMAD decreased significantly ($p < 0.002$). The z-scores at both the LS and the FN also decreased significantly ($p < 0.001$ and $p = 0.004$, respectively). (Klink, Caris, Heijboer, van Trotsenburg, & Rotteveel, 2015, pp. “Supplemental Data”, p. 3 and Table 4) Whilst

BMAD at LS and FN as well as BMAD z-scores at FN in transmen were also seen as insignificant, the changes in BMAD z-scores at LS in transmen were significant ($p = 0.004$ at a level of significance at $p < 0.017$). (Klink, Caris, Heijboer, van Trotsenburg, & Rotteveel, 2015, p. E272 (Table 2)) In the larger cohort ($n = 42$) of the supplemental data, both LS BMAD and LS BMAD z-score showed a significant decrease ($p < 0.001$) in transmen. At $p < 0.001$ and $p = 0.001$, FN BMAD and FN BMAD z-score also decreased. (Klink, Caris, Heijboer, van Trotsenburg, & Rotteveel, 2015, pp. "Supplemental Data", p. 3)

Schagen et al. (2020), on one hand, report no changes in LS BMAD in early-pubertal transgirls and -boys as well as late-pubertal transgirls during two years of GnRHa treatment. Late-pubertal transboys show a "small but significant decrease" (Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020, p. e4255) of LS BMAD. On the other hand, FN BMAD does show a significant decrease in early- and late-pubertal transboys and late-pubertal transgirls, but not in early-pubertal transgirls. (Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020, p. e4256 (Table 2)) BMAD z-scores of the LS and the FN significantly decreased. While this was observed in all groups (early- and late-pubertal transgirls and transboys) for LS BMAD z-scores ($p \leq 0.001$ at a level of significance at $p < 0.05$), this change occurs only in late-pubertal transgirls and both groups of transboys for FN BMAD z-scores ($p = 0.006$, $p = 0.002$, $p < 0.001$) (Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020, pp. e4255, e4256 (Table 2)). Furthermore, three transgirls "had a z-score of the lumbar spine below -2 [and four] had a z-score of the hip below -2" (Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020, p. e4255). In transboys, none had a LS BMAD z-score below -2, but two did have a "z-score of the hip below -2" (Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020, p. e4255).

Vlot et al. (2017) report no differences in BMAD of the FN between young (bone age < 14 years in transmen; bone age < 15 years in transwomen) and old subjects of both transmen and transwomen, and the two groups themselves, at start of GnRHa treatment. Over the course of the treatment, FN BMAD decreased only in old transmen. At the start of GnRHa, FN BMAD z-scores in both groups of transwomen were below zero. (Vlot, et al., 2017, p. 13 (3.3.1)) In the study's summary of the results in tabular form (Vlot, et al., 2017, p. 14 (Table 2)), a decrease in the FN BMAD z-score can be seen in young and old transmen as well as in young transwomen, whereby only the decrease (of $p = 0.002$) in the group of men is explicitly mentioned in the text (Vlot, et al., 2017, p. 13 (3.3.1)). At baseline (= start of GnRHa treatment) the only difference in BMAD of the LS is between young transwomen and young transmen, with young transwomen having lower values than young transmen. On average, LS BMAD z-scores were lower

in transwomen than in transmen. Looking at only the transmen, “young transmen showed a lower Z-score compared to old transmen ($p = 0.02$) at baseline” (Vlot, et al., 2017, p. 13 (3.3.2)). During GnRHa treatment, LS BMAD z-scores decreased in young and old transmen (young transmen: $p = 0.003$; old transmen: $p \leq 0.0001$) as well as in young transwomen ($p = 0.001$). (Vlot, et al., 2017, p. 13 (3.3.2))

Joseph et al. (2019) showed a fluctuation of LS BMAD in both transgirls and transboys (total of n = 31), though no changes were observed during a three-year GnRHa treatment. However, as it did not increase either, BMAD z-scores display a significant fall ($p = 0.000$ in transgirls; $p = 0.001$ in transboys) (Joseph, Ting, & Butler, 2019, p. 3 (Table 1)). When looking at a larger study cohort (n = 70, including the n = 31 subjects), absolute BMAD did not change significantly (Joseph, Ting, & Butler, 2019, p. 3). But again, BMAD z-scores of both transgirls and transboys showed a decrease after the first year of GnRHa treatment (Joseph, Ting, & Butler, 2019, p. 4 (Table 2)). This indicates a slower increase in bone size, as the rapid increase typical at this age is prevented (Joseph, Ting, & Butler, 2019, p. 3).

Navabi et al. (2021) did not report significant changes regarding BMAD z-scores of the LS during GnRHa treatment in n = 30 transgender females ($p = 0.003$ with $p = 0.0011$ being considered statistically significant). However, LS BMAD z-scores of the n = 80 transgender males showed a significant ($p < 0.001$) decrease. (Navabi, Tang, Khatchadourian, & Lawson, 2021, p. 5 (Table 3)) Especially in late-pubertal transgender males, BMAD z-scores of the LS were “negatively affected” (Navabi, Tang, Khatchadourian, & Lawson, 2021, p. 4) by GnRHa monotherapy.

4.2.7. GnRHa and Other Parameters of Bone Health

The focus of this work is on the changes in bone health as detected by BMD and – if conducted – BMAD. Therefore, not all the studies included in this work take up BMC, BTM or calcium and vitamin D. The ones that do and their findings can be found below.

4.2.7.1. GnRHa and BMC

Only the two most recent studies included in this paper present data on BMC.

Though they did not collect data on BMAD, **Carmichael et al.** (2021) did measure BMC in addition to BMD. The same densitometer as for BMD (see: 4.2.3) was used and measurements were taken at the same sites (LS and hip). Data on the BMC of the LS was conducted for n = 44 subjects at baseline, n = 42 after 12 months of treatment, n = 24 after 24 months and n = 12 after 36 months. At the FN, BMC measurements were taken in n = 43 subjects at baseline, n = 39 after the first 12 months of treatment, n = 22 after 24 months and, again, n = 12 after 36 months. While BMC of neither the

LS nor the hip showed changes at the 12-month mark of GnRHa treatment, BMC of the LS did increase after 24 and 36 months, each ($p < 0.0001$ and $p = 0.0007$ at a level of significance of $p = 0.003$). (Carmichael, et al., 2021, p. 13 / 26 (Table 3)) This increase happened more slowly than in peers (Carmichael, et al., 2021, p. 18 / 26). The BMC of the FN did increase, although not significantly (Carmichael, et al., 2021, p. 13 / 26 (Table 3)).

Navabi et al. (2021) showed that, "relative to the birth-assigned sex population […] transgender females had a lower baseline BMC z-score" (Navabi, Tang, Khatchadourian, & Lawson, 2021, p. 4). BMC showed an increase of $p < 0.001$ (at a significance level of $p = 0.0011$) in transgender females. While BMC z-scores showed significant decrease ($p < 0.001$) in transgender males after GnRHa treatment, transgender females did not show such change. (Navabi, Tang, Khatchadourian, & Lawson, 2021, p. 5 (Table 3))

4.2.7.2.GnRHa and BTM

Same as BMC, BTM (= serum markers) is only discussed in two studies.

They are mentioned very briefly in **Schagen et al.** (2020), who studied P1NP, P3NP (= amino terminal peptide of type 3 procollagen), osteocalcin and 1CTP. To do so, fasting blood samples were "drawn before noon on the same days as the DXA scans" (Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020, p. E4254). Besides osteocalcin, which was measured using an immunometric assay, all serum bone markers were measured via radioimmunoassay (Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020, p. E4254). Before the start of GnRHa treatment, serum levels in both groups of transgirls (early- and late-pubertal) were similar, while in transboys, the early-pubertal group showed significantly higher levels than late-pubertal transboys in all examined serum markers. Over the course of the two years in which they were treated with GnRHa, all beforementioned serum bone markers decreased significantly in early- and late-pubertal transgirls as well as early-pubertal transboys. The strongest drop occurred during the first twelve months of GnRHa treatment. In late-pubertal transboys, P3NP and 1CTP decreased slightly but significantly over the whole course of the treatment. (Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020, p. p. e4256 and e4258 (Figure 2))

Vlot et al. (2017) also collected data related to the bone turnover markers P1NP, osteocalcin, and 1CTP. Unlike Schagen et al. (2020), they measured these serum bone markers in a non-fasting state; however, they did use similar measuring instruments (Vlot, et al., 2017, p. 12 (2.2.2)). The authors found that the younger groups of transmen and transwomen each had higher P1NP and osteocalcin serum levels than their

associated older group at the start of GnRHa treatment. In young transmen as well as young and old transwomen, P1NP levels decreased over the course of the treatment. Osteocalcin on the other hand did not show changes in these groups, though it did increase in old transmen. 1CTP was higher in young transmen than in old ones at start of GnRHa treatment but showed no differences in the two groups of transwomen. Only the young groups of transwomen and transmen showed a decrease of 1CTP during GnRHa treatment. (Vlot, et al., 2017, p. 13 (3.2))

4.2.7.3. Calcium and Vitamin D Status in TGD Youth

While several studies address vitamin D levels in study participants, calcium levels are mentioned explicitly only by **Navabi et al.** (2021). This study suggests that calcium intake is low in transgender youth, therefore the authors recommend calcium supplementation "for all youth with GD, particularly those who seek pubertal suppression" (Navabi, Tang, Khatchadourian, & Lawson, 2021, p. 6).

Three of the more recent studies provide information regarding the status of vitamin D and how they dealt with low values. For reference values on vitamin D insufficiency and deficiency, please read "Calcium and Vitamin D" in Chapter 2.1.3.

Carmichael et al. (2021) report that 18% of their participants (n = 8) "had vitamin D insufficiency at baseline and were given vitamin D supplements" (Carmichael, et al., 2021, p. 11).

The number of participants affected by vitamin D insufficiency in **Stoffers et al.** (2019) is significantly higher, with 74% of the study's participants (= transgender males) having a vitamin D value of < 50 nmol/L at the start of GnRHa treatment. On average, participants had a Vitamin D level of 36 nmol/L at baseline, though the range of these levels (24 - 54 nmol/L) spans from vitamin D deficiency (24 nmol/l < 30 nmol/l) to sufficient vitamin D level (54 nmol/L > 50 nmol/L). (Stoffers, de Vries, & Hannema, 2019, p. 1464 (Table 3)) Individuals with vitamin D insufficiency (and deficiency) were prescribed vitamin D (Stoffers, de Vries, & Hannema, 2019, p. 1466). Even so, at the start of CSH therapy (= end of GnRHa monotherapy), 37% of participants still had a vitamin D insufficiency (with < 50 nmol/L) (Table 3, p. 1464).

At baseline, 17.6 % (n = 30) of the n = 170 transgender youth studied by **Navabi et al.** (2021) had vitamin D deficiency (25OHD < 30 nmol/L) and another 37.6% (n = 64) had vitamin D insufficiency (25OHD: 30 – 50 nmol/L) (Navabi, Tang, Khatchadourian, & Lawson, 2021, p. 4 (Table 2)). These individuals were given vitamin D supplementation (Navabi, Tang, Khatchadourian, & Lawson, 2021, p. 3) which showed improvement of the vitamin D status. Despite the smaller number of study participants during follow-ups (third follow-up: n = 63), the vitamin D levels of the transgender adolescents improve

over time – probably as a result of supplementation. At the third follow-up, 87.3% (n = 55) participants had sufficient (25OHD > 50 nmol/L) vitamin D levels (Navabi, Tang, Khatchadourian, & Lawson, 2021, p. 4 (Table 2)).

4.3. Summary of the Findings

In this chapter the most important findings on each of the parameters, will be briefly presented. For more detail, please refer to back the previous chapter (4.2). The findings are sorted by authors, to ensure clarity, and presented in chronological order. However, if there are results that can be arranged according to different findings (4.3.2) or proportion of those affected (4.3.5), this order is preferred and applied.

4.3.1. BMD

"Absolute bone mass of the LS and FN decreases during GnRHa monotherapy" (Klink, Caris, Heijboer, van Trotsenburg, & Rotteveel, 2015, p. E273). The oldest included study (**Klink et al.**, 2015) shows that LS and FN aBMD and aBMD z-scores decrease significantly in transmen, who were primarily in Tanner Stages 4-5 at start of GnRHa treatment. aBMD z-scores also decrease slightly in transwomen in Tanner Stage 5, just not as rapidly as in transmen. (Klink, Caris, Heijboer, van Trotsenburg, & Rotteveel, 2015) In a larger cohort with transwomen being in earlier Tanner Stages (P3.5 and G4.4), however, aBMD z-scores in transwomen do show a significant decrease (Klink, Caris, Heijboer, van Trotsenburg, & Rotteveel, 2015, p. "Supplemental Data").

The subsequently following study (**Stoffers et al.**, 2019) analyzed the oldest study population – consisting of just transgender males (Stoffers, de Vries, & Hannema, 2019). The data of this study shows a "significant decrease in BMD and BMD z-scores" (Stoffers, de Vries, & Hannema, 2019, p. 1466) at LS, left and right hip, in n = 18 transgender males. Therefore, the transgender males studied had less bone mineral accrual than their peers.

The study conducted by **Joseph et al.** (2019) contained the youngest individuals in both AMABs and AFABs. In contrast to other studies, which present data on BMD z-scores at baseline (e.g., Klink, et al., 2015; Schagen, et al., 2020), in this study, z-scores at baseline were lower in transboys than in transgirls (Joseph, Ting, & Butler, 2019, p. 2). The authors of this study proved that "the usual pattern of accruing bone mass according to age does not happen" (Joseph, Ting, & Butler, 2019, p. 2) during GnRHa application. And although absolute BMD shows "very little actual change" and "does not change substantially over a 2-year period […] on GnRHa" (Joseph, Ting, & Butler, 2019, p. 4), there is an immediate drop in BMD z-scores after the first year of treatment (Joseph, Ting, & Butler, 2019, p. 3). Still, pubertal status at start of GnRHa

was not associated with the previously described later changes (Joseph, Ting, & Butler, 2019, p. 3).

Schagen et al. (2020) differentiated between early- and late-pubertal transgirls and transboys. In all four groups, aBMD z-scores of the LS, FN and whole body showed a significant decrease over the course of two years of GnRHa monotherapy. Still, the aBMD of the LS and FN and BMD of the whole body showed a slight increase in early-pubertal transgirls and -boys of this study. (Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020, p. e4256 (Table 2)) Furthermore, individuals who received GnRHa for more than three years showed almost no changes in aBMD at the LS nor the FN – which suggests a stabilization of z-scores in prolonged treatment (Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020, p. e4260).

In short, **Carmichael et al.** (2021) found that in transgender youth, BMD "increased in the lumbar spine indicating greater bone strength, but more slowly than in peers so [the] BMD z-score fell" (Carmichael, et al., 2021, p. 18 / 26). Same as Joseph et al. (2019), they too state that pubertal stages are not associated with later changes in BMD (Carmichael, et al., 2021, p. 18 / 26).

The last and most recent study (**Navabi et al.**, 2021) that discusses BMD and BMD z-scores reinforces the previous results. It found that "GnRHa monotherapy negatively affected transgender youth aBMD *z* scores at LS, LTH, and TBLH levels" (Navabi, Tang, Khatchadourian, & Lawson, 2021, p. 4); especially of the LS and mostly in late pubertal transgender men. However, since this study primarily looked at transgender adolescents in Tanner Stages 4 and 5, "the degree of BMD changes after puberty suppression in early puberty is not known" (Navabi, Tang, Khatchadourian, & Lawson, 2021, p. 5). The authors note that if transgender males started GnRHa at an earlier stage of puberty, changes in LS aBMD z-score seemed "less pronounced" (Navabi, Tang, Khatchadourian, & Lawson, 2021, p. 4).

4.3.2. BMAD

Looking at the smaller study population of the study conducted by **Klink et al.** (2015), "BMAD [does] not change during GnRHa monotherapy" (Klink, Caris, Heijboer, van Trotsenburg, & Rotteveel, 2015, p. E273) and BMAD z-scores do decrease, but not significantly. However, supplemental data of a larger cohort shows that during GnRHa monotherapy, FN BMAD, as well as FN and LS BMAD z-scores decrease in transwomen, and both LS BMAD and LS BMAD z-score and FN BMAD and FN BMAD z-score decrease in transmen (Klink, Caris, Heijboer, van Trotsenburg, & Rotteveel, 2015, p. "Supplemental Data").

No changes of BMAD at the LS were found in transgirls and transboys in **Joseph et al.** (2019). BMAD z-scores of both transgirls and transboys show a significant decrease after longer GnRHa treatment, though the most significant decrease can be seen after the first year of GnRHa treatment in both small (n = 31) and large (n = 70) sample sizes. This suggests a slower accrual of calcium in transgirls and -boys of which the long-term effects are not known, yet. (Joseph, Ting, & Butler, 2019, p. 3)

In contrast to those findings, **Vlot et al.** (2017) provided data, showing BMAD LS and FN decrease in old transmen with a bone age greater than 14. BMAD z-scores, predominantly at the LS, decreased in young transwomen as well as young and old transmen. (Vlot, et al., 2017, p. 13) This suggests that "sex steroid deprivation due to GnRHa treatment results in stable bone mass, which implies a loss of Z-scores in transgender adolescents compared to their peers" (Vlot, et al., 2017, p. 16).

Schagen et al. (2020) present similar findings to Vlot et al. (2017). In their study, BMAD z-scores at LS significantly decreased in all study groups (early- and late-pubertal transgirls and -boys) (Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020, p. e4255). Furthermore, three transgirls "had a z-score of the lumbar spine below -2" and four "had a z-score of the hip below -2" and two transboys had a "z-score of the hip below -2" (Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020, p. e4255).

Navabi et al. (2021) could show that BMAD z-scores of the LS were "negatively affected" (Navabi, Tang, Khatchadourian, & Lawson, 2021, p. 4) by GnRHa monotherapy – especially in late pubertal transgender males. But while transgender females did not show significant changes in LS BMAD z-scores, transgender males' LS BMAD z-scores showed a significant decrease during GnRHa treatment. (Navabi, Tang, Khatchadourian, & Lawson, 2021, p. 5 (Table 3)). Again, this decrease was most present in those in late puberty (Tanner Stages 4-5).

4.3.3. BMC

Data on BMC during GnRHa monotherapy is provided by the two most recent studies.

Without giving specific data sorted for transgirls and transboys, **Carmichael et al.** (2021) showed a significant increase of BMC at the LS after two to three years of GnRHa monotherapy (Carmichael, et al., 2021, p. 13 / 26 (Table 3)). However, this increase happened "more slowly than in peers" (Carmichael, et al., 2021, p. 18 / 26).

The findings of **Navabi et al.** (2021) are consistent with this. They, too, showed an increase of BMC in transgender females with no changes in BMC z-score. In transgender males, however, BMC z-scores decreased significantly. (Navabi, Tang, Khatchadourian, & Lawson, 2021, p. 5 (Table 3))

4.3.4. BTM

Both studies that included data on BTM subdivided their study populations into “young” (= early-pubertal) and “old” (= late-pubertal).

Schagen et al. (2020) examined P1NP, P3NP, osteocalcin and 1CTP. These BTM decreased significantly in early- and late-pubertal transgirls and early-pubertal transboys over the course of a two-year GnRHa treatment, with the strongest drop occurring in the first year (Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020, p. e4256). P3NP and 1CTP showed a slight (but significant) decrease in late-pubertal transboys, over the whole course of the (2-year) treatment, while “P1NP and osteocalcin did not change” (Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020, p. e4256).

Similarly, **Vlot et al.** (2021) showed that both groups of younger individuals had higher P1NP and osteocalcin serum levels at start of GnRHa juxtaposed to their older groups (Vlot, et al., 2017, p. 13 (3.2)). Again, in young and old transwomen as well as young transmen, P1NP decreased during the treatment period while 1CTP decreased only in the younger groups (Vlot, et al., 2017, p. 13 (3.2.1 and 3.2.3)). Osteocalcin did not change in any of these three groups, but it did increase in old transmen (Vlot, et al., 2017, p. 13 (3.2.2)).

4.3.5. Calcium and Vitamin D

According to **Navabi et al.** (2021), calcium levels in transgender youth are low, which is why they suggest calcium supplementation – especially for those transgender youth, who are treated with GnRHa (Navabi, Tang, Khatchadourian, & Lawson, 2021, p. 6).

Vitamin D levels were found to be insufficient in many TGD youth before GnRHa treatment:

Carmichael et al. (2021) report a vitamin D insufficiency at baseline in 18% of their subjects (Carmichael, et al., 2021, p. 11 / 26 (Table 1)).

At baseline, in 37.6% of the subjects in **Navabi et al.**’s (2021) study, vitamin D levels were insufficient, with another 17.6% of individuals having a vitamin D deficiency (Navabi, Tang, Khatchadourian, & Lawson, 2021, p. 4 (Table 2)).

In **Stoffers et al.** (2019), the number of individuals (transgender males) which showed a vitamin D deficiency was highest at baseline at 74%. Even with supplementation (during GnRHa monotherapy at a median duration of eight months), 37% had insufficient vitamin D levels at end of GnRHa monotherapy. (Stoffers, de Vries, & Hannema, 2019, p. 1464 (Table 3))

5. Discussion

5.1. Discussion of Study Results

The aim of this work is to determine the impact of GnRHa on the bone health of TGD youth using primarily BMD and BMAD. Other useful parameters for determining bone health are included but seen as subsidiary. As depicted in previous chapters (see: 4.2 and 4.3), the studies included in this work show a negative association between GnRHa monotherapy and bone health in TGD youth. The most important aspects of the studies' findings are discussed hereafter. It is important to know that terms BMD, aBMD, bone mass and bone density are interchangeable.

Besides decrease, stagnation (= "no change") in BMD or BMAD can also lead to decreased BMD and BMAD z-scores, since both BMD and BMAD increase during adolescence in peers (without GnRHa treatment) (Crabtree, et al., 2017, p. "Supplementary Material").

Several studies found that **BMD** z-scores at baseline are lower in transgirls (=AMABs) than in transboys (=AFABs) (Klink, Caris, Heijboer, van Trotsenburg, & Rotteveel, 2015; Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020; Navabi, Tang, Khatchadourian, & Lawson, 2021). Contrary to these results, Joseph et al. (2019) showed that transboys had lower BMD z-scores at baseline than transgirls (Joseph, Ting, & Butler, 2019). This suggests that even before GnRHa, bone density is not ideal in TGD youth.

The studies' findings agree that BMD z-scores decrease during GnRHa treatment (Klink, Caris, Heijboer, van Trotsenburg, & Rotteveel, 2015; Stoffers, de Vries, & Hannema, 2019; Joseph, Ting, & Butler, 2019; Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020; Carmichael, et al., 2021; Navabi, Tang, Khatchadourian, & Lawson, 2021), whether due to stagnation or actual decrease of BMD. This change is particularly strong in late pubertal transgender boys (Klink, Caris, Heijboer, van Trotsenburg, & Rotteveel, 2015; Navabi, Tang, Khatchadourian, & Lawson, 2021). Low BMD z-scores indicate unusually low BMD (McKiernan, Berg, & Linneman, 2011). Unusually low BMD, in turn, increases the risk for fractures (NHS, 2022). A decrease of z-values of -2.0 – as shown in Navabi, et al. (2021) – indicates a loss of bone mass compared to an individual's peer group (U.S. Department of Health and Human Services (Pt. 4), 2004) and thus a higher facture risk due to osteopenia (Varacallo, Seaman, Jandu, & Pizzutillo, 2022). Even with that higher fracture risk, Navabi et al. (2021) could not find vertebral fractures in further imaging of individuals with z-scores below 2 (Navabi, Tang, Khatchadourian, & Lawson, 2021, p. 3).

Although fracture risk increases with decreased BMD (Varacallo, Seaman, Jandu, & Pizzutillo, 2022), lower BMD and BMD z-scores do not necessarily result in a fracture (U.S. Department of Health and Human Services (Pt. 4), 2004). As BMD decreases with ageing, one's z-score may remain stable throughout their entire life – therefore, BMD z-scores might not be suitable for diagnostic use. The comparison of an individual with their peers, however, is still possible. (U.S. Department of Health and Human Services (Pt. 4), 2004) The impact of GnRHa on BMD appears to be stronger as the study participants get older, and Tanner Stages are higher. Although BMD is lower in TGD youth treated with GnRHa compared to peers, it seems to stabilize with prolonged GnRHa use.

When considering the ongoing bone growth in TGD adolescents by calculating **BMAD** values, the data situation of the cited studies is not as conclusive as data regarding BMD. Two of four studies with data regarding BMAD did not find any changes in BMAD values in smaller populations (Klink, Caris, Heijboer, van Trotsenburg, & Rotteveel, 2015; Joseph, Ting, & Butler, 2019). In larger cohorts, however, BMAD did show a significant decrease (Klink, Caris, Heijboer, van Trotsenburg, & Rotteveel, 2015; Vlot, et al., 2017; Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020; Navabi, Tang, Khatchadourian, & Lawson, 2021). It seems as if BMAD z-scores decrease predominantly at the LS (Klink, et al., 2015; Vlot, et al., 2017; Schagen, et al., 2020; Navabi, et al., 2021). The decrease in BMAD z-scores seems to be especially prominent in older transgender boys (Vlot, et al., 2017; Navabi, Tang, Khatchadourian, & Lawson, 2021). Again, one study even showed a drop greater than -2 in BMAD z-scores recorded for individual study participants (Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020). In growing children and adolescents, low LS BMAD (and BMAD z-scores) can indicate an increased fracture risk (Jones, Ma, & Cameron, 2006).

BMC, BTM and **calcium and vitamin D** are considered subsidiary criteria of bone health in this work. Therefore, they were not part of the inclusion criteria. Consequently, less data is provided by the cited studies regarding these aspects.

BMC does increase in TGD youth treated with GnRHa (Carmichael, et al., 2021; Navabi, et al., 2021). However, in line with the development of BMD, this increase does not happen as fast as it does in their peers (Carmichael, et al., 2021), and thus BMC z-scores show a significant decrease (Navabi, Tang, Khatchadourian, & Lawson, 2021). In this case, the evolution of BMC is consistent with the demonstrated changes in BMD and BMAD and supports the finding that bone mineral density is not evolving as expected during GnRHa monotherapy. Thus, this parameter also indicates that TGD youth might have an increased risk of fractures.

BTM (serum bone markers or biochemical markers of bone turnover), append to the information obtained by determining BMD. While not useful for estimating bone mass, they can be used to estimate the rate at which bone loss is occurring. (U.S. Department of Health and Human Services (Pt. 1), 2004) Both studies looking at BTM agree that P1NP, osteocalcin and 1CTP decrease during the use of GnRHa (Schagen, et al., 2020; Vlot, et al., 2017). This indicates that bone loss might be happening faster than in peers. Again, this shows an increased fracture risk (NHS, 2022).

Calcium and vitamin D are already lower in TGD adolescents before the administration of GnRHa than in peers, with very few having even adequate levels of vitamin D (Stoffers, de Vries, & Hannema, 2019; Carmichael, et al., 2021; Navabi, Tang, Khatchadourian, & Lawson, 2021). Especially in older transgender males who only received vitamin D supplements for a median of 8 months of GnRHa monotherapy, the documented deficiency and insufficiency had not been fully resolved at the end of GnRHa monotherapy (Stoffers, de Vries, & Hannema, 2019). Seeing as both calcium and Vitamin D serve as protective factors of bone health (see: 2.1.3), this is less than ideal for people whose fracture risk seems to be increased due to the treatment they receive. Therefore, TGD youth should receive supplementation of both calcium, but especially vitamin D.

Based on the parameters analyzed, it can be said that the studies reviewed in this work concluded that the bone health of TGD adolescents – especially in older transgender boys – is affected by GnRHa monotherapy and the fracture risk increases. However, further research is needed, as shown in 6.

5.2. Methodological Critique

5.2.1. Methodological Critique of the Cited Studies

The need for further research on this topic is evident from various circumstances.

First, the sample sizes of the studies cited are rather small. Of course, it should be considered that these are samples from an already smaller population (TGD youth). Still, smaller sample sizes may result in changes in bone health parameters during GnRHa monotherapy going undetected.

Furthermore, in some cases, there was missing data (Carmichael, et al., 2021; Joseph, Ting, & Butler, 2019). In Joseph et al. (2019) for example, this was due to the retrospective design, as sometimes data was not recorded annually (Joseph, Ting, & Butler, 2019, p. 4). This also influences the evaluation of the observed changes.

In other studies (Stoffers, de Vries, & Hannema, 2019; Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020), the findings recorded in-text did not correspond

to the results presented in tables. This is unmeriting when it comes to assessing the results of studies.

In addition, the comparability of the studies is questionable since the conception and study designs of the included studies differ. This is noticeable, among other things, when considering the given levels of significance. For example, Joseph et al. deem $p < 0.05$ as “statistically significant” (Joseph, Ting, & Butler, 2019, p. 2), whereas Carmichael et al. deem $p = 0.003$ as the “appropriate threshold for statistical significance” (Carmichael, et al., 2021, p. 9).

Another aspect is the number of people of whom the data is interpreted. It is questionable to what extent findings for a small n can be applied to larger populations at all. Carmichael et al. (2021), for example, omit data for n = 8 or less (Carmichael, et al., 2021, p. 10). Other studies, however, collect very small study populations from the outset, e.g., data on n = 10 transgender girls (Joseph, Ting, & Butler, 2019), or report data on sample sizes as small as n = 5 (Vlot, et al., 2017, p. 14). As a result, findings that are considered significant in one study would not be considered as such in another study. This in turn makes comparability and thus the assessment of the results more difficult.

It is also unclear to what extent the use of DXA scans leads to comparable results. DXA scans are considered the gold standard. Therefore, it makes sense to only include studies that meet this gold standard. This also creates at least a basis for comparability. However, according to earlier studies, “BMD measured on one type of machine cannot be accurately compared to BMD measured on a different type of machine, nor can BMD [be] performed on the same type of machine at two different locations” (U.S. Department of Health and Human Services (Pt. 4), 2004, p. 208).

Some of the studies reviewed in this paper refer to only one or two parameters of bone health (Klink, Caris, Heijboer, van Trotsenburg, & Rotteveel, 2015; Vlot, et al., 2017). This means that other factors influencing the development of bone mass – such as physical activity, weight, smoking, alcohol consumption, socioeconomic status, diseases, and medications (Zhu & Zheng, 2021), and in most studies even vitamin D and calcium – were not taken into account. How different constellations between protective and risk factors and GnRHa monotherapy affect bone health in TGD adolescents, cannot be said exactly.

Furthermore, there is mostly data on the influence of GnRHa on the bone health of transgender boys. If one adds up all the study populations of the studies considered, it results in available data for n = 316 transgender boys, and for n = 201 transgender girls. This leads to the conclusion that transgender girls seem to be underrepresented in current research.

In addition, in the included studies there is more data to be found on transgender youth (n = 316 + 201 = 517) than on nonbinary youth (just n = 2) regarding the GnRHa monotherapy for treatment of GD. This group of TGD youth should in no way be neglected, as they can also benefit from treating their suffering of GD with GnRHa. On top of that, the studies mostly consider short- to mid-term treatment durations. However, since nonbinary adolescents are unlikely to continue with CSH therapy after GnRHa – as they do not feel like they belong to their sex assigned at birth, nor to the other sex – it is important to investigate the effects of longer application durations.

Also, none of the studies had a control group consisting of TGD adolescents without GnRHa treatment. Any similarities and differences between TGD adolescents with and without GnRHa monotherapy are therefore not recognizable. Instead, the z-values of the parameters collected are often compared with the values of peers of the birth sex. The reasonableness of such comparisons is doubtful – as mentioned in the studies themselves (Klink, Caris, Heijboer, van Trotsenburg, & Rotteveel, 2015, p. E273; Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020, p. e4260) – since the release of sex hormones by peers of the congenital sex naturally has a different effect on the bones than the lack of these sex hormones in TGD adolescents under GnRHa treatment.

Still, the overall quality of the considered studies can be rated as “good”. It is noticeable that the quality has increased in recent years.

Most of the studies reviewed here were conducted in the Netherlands or the UK. Future international research should be conducted.

5.2.2. Methodological Critique of this Paper

This paper, too, has its limitations. For one, only a limited number of databases (Google Scholar, PubMed and Medline, Scopus) were used to conduct the literature search. The research was also carried out by only one person. Moreover, only studies in English (or German) and ones which were published on the Open Access principle were screened for their Abstract and Full Text – and, if fitting the research question, included in this work. It is therefore quite possible that relevant literature was not found.

However, the inclusion of literature found by checking the primary sources of reviews is a strength of this work. For example, the study by Stoffers et al. (2019) was initially excluded after the abstract screening and was listed as relevant and thus included in this work, after it seemed to provide useful information by sifting through reviews and their primary literature. The detailed presentation of the studies’ contents and findings is also a strength, which above all compensates for the fact that literature for this paper was recorded, and the paper itself was written, by only one person.

5.3. Contribution to Current Research

The aim of this work is to present the current state of research and evidence-based medical knowledge on the impact of GnRHa on the bone health of TGD youth. It is important to note that the political discussions taken up in the introduction only serve to illustrate the topicality and urgency of this topic. This work does not serve as a basis for political discussions and does not intend to solve or fuel the current political discussions.

Since "transgender people" is a very current topic, this work is not the first to address it. How this work differs from others and why it therefore contributes to the current state of research should therefore be recorded.

Against the same background of the planned change in the TSG law in Germany, an article on *Genderinkongruenz: zur Problematik einer geschlechtsangleichenden medikamentösen Intervention bei Kindern und Jugendlichen* (in English: Gender Incongruence: on the problem of a gender-adapting drug intervention in children and adolescents) was published in the “Arzneimittelbrief” in March 2023. However, this article primarily deals with CSH therapy and gender-reassignment surgeries, as well as what this change in law means for health care and medical professionals. (Der Arzneimittelbrief, 2023)

Furthermore, some previously conducted reviews reference the same studies as this paper, but most of those focus on bone health as only one aspect of the possible effects of GnRHa therapy on TGD adolescents – most focus primarily on the psychological aspects or include a multitude of aspects (e.g., body mass index). Others reference the same findings with the added aspect of cross-sex hormone treatment, which, when looking at the cumulated effects of GnRHa and CSH, yield different results.

For example, in May last year, Giacomelli et al. (2022) published a review on the *Bone Health in Transgender People* which does include GnRHa use in transgender youth. However, sole focus is not on TGD youth, as data on older TGD people are also included. Furthermore, the studies presented herein are only very briefly summarized (Giacomelli & Meriggiola, 2022).

A study by Ciancia et al. was found only after the start of this work, which bears a very similar title: *Impact of gender-affirming treatment on bone health in transgender and gender diverse youth*. The goal of this paper is to describe “differences in age and pubertal stage at the start of puberty suppression, chosen strategy to block puberty progression, duration of puberty suppression, and the timing of re-evaluation after estradiol or testosterone administration” (abstract). In addition to GnRHa, this study also looks at the use of anti-androgens such as cyproterone acetate and progestins such as lyn-

estrenol (Ciancia, Dubois, & Cools, 2022, p. 5/6) . The development of bone health during CSH therapy is also discussed (Ciancia, Dubois, & Cools, 2022, p. 6 ff.). Despite the similar title, the work at hand is in no way intended to duplicate the review published by Ciancia et al.

The focus of this paper is to give detailed information on the studies considered, to coordinate the findings of the individual studies and to come to a general conclusion regarding the bone health of TGD adolescents under the influence of GnRHa. Such an in-depth study on the effects of GnRHa on the bone health of just TGD youth (diagnosed with gender dysphoria), which summarizes and presents the findings of individual studies in a thorough and structured manner, has not been found in the search for literature within this paper. Therefore, this work can add to the current study situation and especially helps both those who want to get an initial overview of, as well as those who are interested in an in-depth presentation of multiple studies on the impact of GnRHa on bone health in TGD adolescents.

6. Conclusion

Preliminary study results suggest that the use of GnRHa monotherapy to suppress puberty in TGD youth with GD affects bone health. The following briefly summarizes the tendency of the effects, whether stronger effects can be detected in AMABs or AFABs (e.g., dependent on different ages and Tanner stages), and what this means for future research.

GnRHa negatively impacts bone health in TGD youth. This impact can manifest itself through stagnation, but also a decrease of bone mass. The developments of BMD, BMAD and BMC during GnRHa suggest an increased fracture risk in TGD youth. Seeing as "bone mass is a key determinant of osteoporosis and fragility fractures" (Zhu & Zheng, 2021), it is important to monitor bone health of TGD adolescents during GnRHa. The studies cited here examined mostly smaller sample sizes. For a better picture of the (negative) effects and a better generalizability of study results, further studies of a larger scope are necessary. Studies which include follow-ups at later ages and look at a multitude of bone parameters could also be helpful to determine whether the increased fracture risk is reflected in more frequent fractures in TGD people.

The impact of GnRHa on BMD appears to be stronger as the study participants get older, and Tanner Stages are higher. Especially in transgender boys in later stages of puberty, the negative impact of GnRHa on parameters of bone health seems to be more pronounced. However, the studies considered here could present the actual conditions in a distorted manner, seeing as they, for the most part, examine transgender boys – who are also predominantly in later Tanner Stages. Future studies should therefore collect and evaluate data of younger individuals, more transgender girls, and non-binary youth.

But, although bone mass is lower in TGD treated with GnRHa compared to peers, it seems to stabilize with prolonged GnRHa use. Since most of the study participants in the studies considered were older (over 14 y/o) at the start of the GnRHa monotherapy and in later puberty, and GnRHa is in most cases only used as a monotherapy up to the age of 16, the study results only briefly represent to medium-term effects. In most countries, however, GnRHa is permitted from the age of 12, or (according to SOC-8) from Tanner Stage 2. The long-term effects and how GnRHa affects the bone health of younger TGD adolescents cannot be illustrated by the studies presented here. It is advisable to conduct further research in this regard.

In short, GnRHa tends to lead to a decrease in bone mass in the TGD adolescents and an increased fracture risk – especially in older TGD youth. Based on this finding, further research regarding bone health and treatment optimization in TGD is warranted.

References

Ahrbeck, B., & Felder, M. (2022). Nutzen-Risiko-Abwägung einer pubertätssupprimierenden Behandlung mit GnRH-Analoga. In B. Ahrbeck, & M. Felder, *Geboren im falschen Körper: Genderdysphorie bei Kindern und Jugendlichen.* Kohlhammer.

Almeida, M., Laurent, M. R., Dubois, V., Claessens, F., O´Brien, C. A., Bouillon, R., . . . Manolagas, S. C. (2017). Estrogens and Androgens in Skeletal Physiology and Pathophysiology. *Physiological reviews, 97*(1), 135 - 187. doi:https://doi.org/10.1152/physrev.00033.2015

APA. (n.d.). *What is Gender Dysphoria?* Retrieved May 2023, from American Psychiatric Association: https://www.psychiatry.org/patients-families/gender-dysphoria/what-is-gender-dysphoria#section_3

Bazzocchi, A., Ponti, F., Albisinni, U., Battista, G., & Guglielmi, G. (2016). DXA: Technical aspects and application. *European Journal of Radiology, 85*(8), 1481 - 1492. doi:https://doi.org/10.1016/j.ejrad.2016.04.004

BMFSFJ. (2022, October). *Bundesregierung empfiehlt nicht die Einnahme von Pubertätsblockern*. Retrieved May 2023, from Bundesministerium für Familie, Senioren, Frauen und Jugend: https://www.bmfsfj.de/bmfsfj/aktuelles/alle-meldungen/bundesregierung-empfiehlt-nicht-die-einnahme-von-pubertaetsblockern-202506

BMFSFJ. (2023, May). *Queerpolitik und geschlechtliche Vielfalt: Gesetz über die Selbstbestimmung in Bezug auf den Geschlechtseintrag (SBGG)*. Retrieved May 2023, from Bundesministerium für Familie, Senioren, Frauen und Jugend: https://www.bmfsfj.de/bmfsfj/themen/gleichstellung/queerpolitik-und-geschlechtliche-vielfalt/gesetz-ueber-die-selbstbestimmung-in-bezug-auf-den-geschlechtseintrag-sbgg--199332

Breehl, L., & Caban, O. (2022). Physiology, Puberty. *StatPearls [Internet].* Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK534827/

Carmichael, P., Butler, G., Masic, U., Cole, T. J., De Stavola, B. L., Davidson, S., . . . Viner, R. M. (2021, February). Short-term outcomes of pubertal suppression in a selected cohort of 12 to 15 year old young people with persistent gender dysphoria in the UK. *PLoS ONE, 16*(2). doi:https://doi.org/10.1371/journal.pone.0243894

Ciancia, S., Dubois, V., & Cools, M. (2022). Impact of gender-affirming treatment on bone health in transgender and gender diverse youth. *Endocrine Connections, 11*(11). doi:https://doi.org/10.1530/EC-22-0280

Coleman, E., Radix, A. E., Bouman, G. R., de Vries, A. L., Deutsch, M. B., Ettner, R., . . . Arcelus, J. (2022). Standards of Care for the Health of Transgender and Gender Diverse People, Version 8. *International Journal of Transgender Health, 23*(S1), S1-S259. doi:https://doi.org/10.1080/26895269.2022.2100644

Crabtree, N. J., Arabi, A., Bachrach, L. K., Fewtrell, M., El-Hajj Fuleihan, G., Kecskemethy, H. H., . . . International Society for Clinical Densitometry. (2014). Dual-energy X-ray absorptiometry interpretation and reporting in children and adolescents: the revised 2013 ISCD Pediatric Official Positions. *Journal of clinical densitometry : the official journal of the International Society for Clinical Densitometry, 17*(2), 225 - 242. doi:https://doi.org/10.1016/j.jocd.2014.01.003

Crabtree, N. J., Shaw, N. J., Bishop, N. J., Adams, J. E., Mughal, M. Z., Arundel, P., . . . ALPHABET Study Team. (2017). Amalgamated Reference Data for Size-Adjusted Bone Densitometry Measurements in 3598 Children and Young Adults-the ALPHABET Study. *Journal of bone and mineral research: the official journal of the American Society for Bone and Mineral Research, 31*(1), 172-180. doi:https://doi.org/10.1002/jbmr.2935

Der Arzneimittelbrief. (2023, 03). Genderinkongruenz: zur Problematik einer geschlechtsangleichenden medikamentösen Intervention bei Kindern und Jugendlichen. *Der Arzneimittelbrief (Unabhängiges Informationsblatt der ISDB), 57*(3), 22 - 23.

Emmanuel, M., & Bokor, B. R. (2022). Tanner Stages. *StatsPearls [Internet].* Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK470280/

Florida Department of Health - Board of Medicine. (2023, March). *Standards of Practice for the Treatment of Gender Dysphoria in Minor.* Retrieved May 2023, from Florida Administrative Rules, Law, Code, Register: https://www.flrules.org/gateway/RuleNo.asp?id=64B8-9.019

Garnero, P. (2009). Bone markers in osteoporosis. *Current Osteoporosis Reports, 7*, 84 - 90. doi:https://doi.org/10.1007/s11914-009-0014-3

Giacomelli, G., & Meriggiola, M. C. (2022). Bone Health in Transgender People: a narrative review. *Therapeutic advances in endocrinology and metabolism, 13*, 1 - 15. doi:https://doi.org/10.1177/20420188221099346

Greenblatt, M. B., Tsai, J. N., & Wein, M. N. (2017). Bone Turnover Markers in the Diagnosis and Monitoring of Metabolic Bone Disease. *Clinical Chemistry, 2*, 464–474. doi:https://doi.org/10.1373/clinchem.2016.259085

Greer, F. R., Krebs, N. F., & Committee on Nutrition. (2006). Optimizing Bone Health and Calcium Intakes of Infants, Children, and Adolescents. *Pediatrics, 117*(2), 578 - 585. doi:https://doi.org/10.1542/peds.2005-2822

Guss, C., & Gordon, C. M. (2022). Pubertal Blockade and Subsequent Gender Affirming Therapy: True, True and Unrelated? *JAMA Netw Open, 5*(11). doi:10.1001/jamanetworkopen.2022.39763

Hembree, W. C., Cohen-Kettenis, P. T., Gooren, L., Hannema, S. E., Meyer, W. J., Murad, M., . . . T´Sjoen, G. G. (2017). Endocrine Treatment of Gender-Dysphoric/Gender-Incongruent Persons: An Endocrine Society* Clinical Practice Guideline. *The Journal of Clinical Endocrinology & Metabolism, 102*(11), 3869 - 3903. doi:https://doi.org/10.1210/jc.2017-01658

Holick, M. F. (2009). VITAMIN D STATUS: MEASUREMENT, INTERPRETATION AND CLINICAL APPLICATION. *Annals of epidemiology, 19*(2), 73 - 78. doi:https://doi.org/10.1016/j.annepidem.2007.12.001

Institute of Medicine (US). (1997). 4 Calcium. In S. C. Intakes, *Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D, and Fluoride.* Washington (DC): National Academies Press (US). Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK109827/#_ch4_s1_

Institute of Medicine (US). (2011). Bone Mass Measurements Associated With Calcium. In I. o. Calcium, A. C. Ross, C. L. Taylor, A. L. Yaktine, & H. B. Del Valle (Eds.), *Dietary Reference Intakes for Calcium and Vitamin D.* Washington DC: National Academies Press (US). doi:10.17226/13050

Jones, G., Ma, D., & Cameron, F. (2006). Bone density interpretation and relevance in Caucasian children aged 9-17 years of age: insights from a population-based fracture study. *Journal of clinical densitometry: the official journal of the International Society for Clinical Densitometry, 9*(2), 202 - 209. doi:https://doi.org/10.1016/j.jocd.2006.02.004

Joseph, T., Ting, J., & Butler, G. (2019, October). The effect of GnRH analogue treatment on bone mineral density in young adolescents with gender dysphoria: findings from a large national cohort. *Journal of pediatric endocrinology & metabolism: JPEM, 32*(10), 1077–1081. doi:https://doi.org/10.1515/jpem-2019-0046

Kalkwarf, H. J., Zemel, B. S., Gilsanz, V., Lappe, J. M., Horlick, M., Oberfield, S., . . . Shepherd, J. A. (2007). The bone mineral density in childhood study: bone mineral content and density according to age, sex, and race. *The Journal of clinical endocrinology and metabolism, 92*(6), 2087-2099. doi:https://doi.org/10.1210/jc.2006-2553

Klink, D., Caris, M., Heijboer, A., van Trotsenburg, M., & Rotteveel, J. (2015). Bone mass in young adulthood following gonadotropin-releasing hormone analog treatment and cross-sex hormone treatment in adolescents with gender dysphoria. *The Journal of clinical endocrinology and metabolism, 100*(2), E270 - E275. doi:https://doi.org/10.1210/jc.2014-2439

Kröger, H., Kotaniemi, A., Vainio, P., & Alhava, E. (1992). Bone densitometry of the spine and femur in children by dual-energy-x-ray absorptiometry. *Bone Miner, 17*(1), 75-85.

Martin, J., & Hadwin, J. A. (2022). The roles of sex and gender in child and adolescent mental health. *JCPP Advances, 2*(1). doi:https://doi.org/10.1002/jcv2.12059

McKiernan, F. E., Berg, R. L., & Linneman, J. G. (2011). The utility of BMD Z-score diagnostic thresholds for secondary causes of osteoporosis. *Osteoporosis International, 22*, 1069 – 1077. doi:https://doi.org/10.1007/s00198-010-1307-1

Methley, A. M., Campbell, S., Chew-Graham, C., McNally, R., & Cheraghi-Sohi, S. (2014). PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. *BMC health services research, 14*(579). doi:https://doi.org/10.1186/s12913-014-0579-0

Meyer, G. (2021). Geschlechtsangleichende Hormontherapie bei Transidentität: Voraussetzungen und Therapiemanagement. *Journal for Gynäkologische Endokrinologie / Schweiz, 24*, 180 - 189. doi:https://doi.org/10.1007/s41975-021-00215-x

Mølgaard, C., Thomsen, B. L., Prentice, A., Cole, T. J., & Fleischer Michaelsen, K. (1997). Whole body bone mineral content in healthy children and adolescents. *Archives of Disease in Childhood, 76*, 9 - 15. Retrieved from https://adc.bmj.com/content/76/1/9.short

Navabi, B., Tang, K., Khatchadourian, K., & Lawson, M. L. (2021, October). Pubertal Suppression, Bone Mass, and Body Composition in Youth With Gender Dysphoria. *PEDIATRICS, 148*(4). doi:https://doi.org/10.1542/peds.2020-039339

NBC News. (2023, March 16). *Florida becomes eighth state to restrict transgender care for minors*. Retrieved May 2023, from NBC News, Out Politics and Policy: https://www.nbcnews.com/nbc-out/out-politics-and-policy/florida-becomes-eighth-state-restrict-transgender-care-minors-rcna75337

NHS. (2020, May 28). *Treatment - Gender dysphoria*. Retrieved May 2023, from National Health Service, UK: https://www.nhs.uk/conditions/gender-dysphoria/treatment/

NHS. (2022, October 13). *Overview - Osteoporosis*. Retrieved May 2023, from National Health Service (UK): https://www.nhs.uk/conditions/osteoporosis/

NIH (National Cancer Institute). (n.d.). *NCI Dictionary of Cancer Terms: bone mineral density*. Retrieved May 2023, from National Institutes of Health: https://www.cancer.gov/publications/dictionaries/cancer-terms/def/bone-mineral-density

NIH (National Cancer Institute). (n.d.). *NCI Dictionary of Cancer Terms: GnRHa*. Retrieved May 2023, from National Institutes of Health: https://www.cancer.gov/publications/dictionaries/cancer-terms/def/gnrha

NIH (National Heart, Lung and Blood Institute). (2021, 06). *Study Quality Assessment Tools (Quality Assessment Tool for Before-After (Pre-Post) Studies With No Control Group)*. Retrieved May 2023, from National Institutes of Health: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools

NIH (Office of Dietary Supplements). (n.d.). *Calcium: Fact Sheet for Consumers*. Retrieved May 2023, from National Institutes of Health: https://ods.od.nih.gov/factsheets/Calcium-Consumer/

RKI. (2019). *Wie wird der Vitamin-D-Status bestimmt und beurteilt?* Retrieved May 2023, from Robert Koch Institut: https://www.rki.de/SharedDocs/FAQ/Vitamin_D/FAQ07.html#:~:text=25(OH)D%20ist%20ein,Wert%20durch%202%2C5).&text=Mangelhafte%20Versorgung%20mit%20einem%20erh%C3%B6hten,wie%20Rachitis%2C%20Osteomalazie%20und%20Osteoporose,

Schagen, S. E., Wouters, F. M., Cohen-Kettenis, P. T., Gooren, L. J., & Hannema, S. E. (2020). Bone Development in Transgender Adolescents Treated With GnRH Analogues and Subsequent Gender-Affirming Hormones. *The Journal of clinical endocrinology and metabolism, 105*(12), e4252 – e4263. doi:https://doi.org/10.1210/clinem/dgaa604

Sheu, A., & Diamond, T. (2016). Bone mineral density: testing for osteoporosis. *Australian prescriber, 39*(2), 35 - 39. doi:10.18773/austprescr.2016.020

Stoffers, I. E., de Vries, M. C., & Hannema, S. E. (2019). Physical Changes, Laboratory Parameters, and Bone Mineral Density During Testosterone Treatment in Adolescents with Gender Dysphoria. *The Journal of Sexual Medicine, 16*(9), 1459 - 1468. doi:https://doi.org/10.1016/j.jsxm.2019.06.014

The General Assembly of the State of Arkansas. (2023, March). *An Act Concerning Medical Malpractice and Gender Transition in Minors; To Create the Protectiong Minors from Medical Malpractice Act of 2023; And for Other Purposes.* Retrieved May 2023, from Arkansas State Legislature: https://www.arkleg.state.ar.us/Bills/Detail?id=sb199&ddBienniumSession=2023%2F2023R

The General Assembly of the State of Iowa. (2023, March). *An Act Relating to Prohibited Activities Regarding Gender Transition Procedures Relative to Minors, and Including Effective Date and Applicability Provisions.* Retrieved May 2023, from The Iowa Legislature: https://www.legis.iowa.gov/legislation/BillBook?ga=90&ba=SF538

The Legislature of the State of Montana. (2023, April). *Youth Health Protection Act.* Retrieved May 2023, from Montana Legislature: https://laws.leg.mt.gov/legprd/LAW0210W$BSIV.ActionQuery?P_BILL_NO1=99&P_BLTP_BILL_TYP_CD=SB&Z_ACTION=Find&P_SESS=20231

The Legislature of the State of South Dakota. (2023, February). *An Act to Prohibit Certain Medical and Surgical Interventions on Minor Patients.* Retrieved May 2023, from South Dakota Legislature Legislative - Research Council: https://sdlegislature.gov/Session/Bill/24100

U.S. Department of Health and Human Services (Pt. 1). (2004). PART ONE: WHAT IS BONE HEALTH?, Chapter 2: The Basics of Bone Health and Disease. In United States Public Health Service, & Surgeon General of the United States, *Bone Health and Osteoporosis: A Report of the Surgeon General* (pp. 16 - 36). Rockville, MD, USA: Department of Health and Human Services, Office of the Surgeon General. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK45504/#_ch2_s4_

U.S. Department of Health and Human Services (Pt. 4). (2004). PART FOUR: WHAT CAN HEALTH CARE PROFESSIONALS DO TO PROMOTE BONE HEALTH?, Chapter 8: Assessing the Risk of Bone Disease and Fracture, Step 2: Measuring Bone Mineral Density (BMD) for Individuals at Risk of Osteoporosis. In United States Public Health Service, & Surgeon General of the United States, *Bone Health and Osteoporosis: A Report of the Surgeon General* (pp. 185 - 212). Rockville, MD, USA: Department of Health and Human Services, Office of the Surgeon General. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK45525/

UDHR. (1948, December 10). *Universal Declaration of Human Rights.* (U. N. Assembly, Ed.) Retrieved May 2023, from United Nations: https://www.un.org/en/about-us/universal-declaration-of-human-rights

UN. (n.d.). *Who Are the Youth?* Retrieved May 2023, from United Nations, Global Issues, Youth: https://www.un.org/en/global-issues/youth

Varacallo, M., Seaman, T. J., Jandu, J. P., & Pizzutillo, P. (2022). Osteopenia. *StatPearls [Internet]*. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK499878/#_article-26405_s1_

Vlot, M. C., Klink, D. T., den Heijer, M., Blankenstein, M. A., Rotteveel, J., & Heijboer, A. C. (2017). Effect of pubertal suppression and cross-sex hormone therapy on bone turnover markers and bone mineral apparent density (BMAD) in transgender adolescents. *Bone, 95*, 11 - 19. doi:https://doi.org/10.1016/j.bone.2016.11.008

von Storch, B., Reichardt, M., Harder-Kühnel, M. I., Ehrhorn, T., Bollman, G., Wenzel Schmidt, J., . . . Peterka, T. M. (2022, October 31). *20/4213 - Antrag: Schutz von Kindern und Jugendlichen mit Geschlechtsdysphorie vor.* Retrieved May 2023, from Deutscher Bundestag: https://dserver.bundestag.de/btd/20/042/2004213.pdf

Ward, K., Ashby, R., Roberts, S., Adams, J., & Zulf Mughal , M. (2007). UK reference data for the Hologic QDR Discovery dual-energy x ray absorptiometry scanner in healthy children and young adults aged 6–17 years. *Archives of Disease in Childhood, 92*(1), 53-59.

WHO. (n.d.). *Adolescent Health - Overview*. Retrieved May 2023, from World Health Organization: https://www.who.int/health-topics/adolescent-health#tab=tab_1

Zhu, X., & Zheng, H. (2021). Factors influencing peak bone mass gain. *Frontiers of Medicine, 15*, 53 - 69. doi:https://doi.org/10.1007/s11684-020-0748-y

Appendix

Study	TGD Terms sorted according to: • AMAB (= transgender girls) • AFAB (= transgender boys)
Klink et al. (2015)	• Transwomen • Transmen
Vlot et al. (2017)	Old and young: • Transwomen • Transmen
Stoffers et al. (2019)	• --- • Transgender males
Joseph et al. (2019)	• Transgirls • Transboys
Schagen et al. (2020)	Early- and late-pubertal: • Transgirls • Transboys
Carmichael et al. (2021)	• Birth-registered males • Birth-registered females
Navabi et al. (2021)	• Transgender females • Transgender males

Table 6: *TGD Terms Used in the Included Studies*

Checklists of the Quality of the Studies According to the QAT (NIH (National Heart, Lung and Blood Institute), 2021)

Study: Bone Mass in Young Adulthood Following Gonadotropin-Releasing Hormone Analog Treatment and Cross-Sex Hormone Treatment in Adolescents With Gender Dysphoria (Klink, Caris, Heijboer, van Trotsenburg, & Rotteveel, 2015)

Criteria	Yes	No	Cannot Determine; Not Applicable; Not Reported
(1) Was the study question or objective clearly stated?	☒	☐	☐
(2) Were eligibility/selection criteria for the study population prespecified and clearly described?	☒	☐	☐
(3) Were the participants in the study representative of those who would be eligible for the test/service/intervention in the general or clinical population of interest?	☒	☐	☐
(4) Were all eligible participants that met the prespecified entry criteria enrolled?	☐	☐	☒
(5) Was the sample size sufficiently large to provide confidence in the findings? [$]	☐	☒	☐
(6) Was the test/service/intervention clearly described and delivered consistently across the study population?	☒	☐	☐
(7) Were the outcome measures prespecified, clearly defined, valid, reliable, and assessed consistently across all study participants?	☒	☐	☐
(8) Were the people assessing the outcomes blinded to the participants' exposures/interventions?	☐	☐	☒

Criteria	Yes	No	Cannot Determine; Not Applicable; Not Reported
(9) Was the loss to follow-up after baseline 20% or less? Were those lost to follow-up accounted for in the analysis? !	☒	☐	☐
(10) Did the statistical methods examine changes in outcome measures from before to after the intervention? Were statistical tests done that provided p values for the pre-to-post changes?	☒	☐	☐
(11) Were outcome measures of interest taken multiple times before the intervention and multiple times after the intervention (i.e., did they use an interrupted time-series design)?	☒	☐	☐
(12) If the intervention was conducted at a group level (e.g., a whole hospital, a community, etc.) did the statistical analysis take into account the use of individual-level data to determine effects at the group level?	☒	☐	☐

Table 7: *Quality Assessment Klink et al. (2015)*

Legend: $ The study's discussion was used to answer the question;
! If either question can be answered with "yes", this criterion gets +1 point, even if loss to follow-up was > 20%

Study: Effect of pubertal suppression and cross-sex hormone therapy on bone turnover markers and bone mineral apparent density (BMAD) in transgender adolescents (Vlot, et al., 2017)

Criteria	Yes	No	Cannot Determine; Not Applicable; Not Reported
(1) Was the study question or objective clearly stated?	☒	☐	☐
(2) Were eligibility/selection criteria for the study population prespecified and clearly described?	☒	☐	☐
(3) Were the participants in the study representative of those who would be eligible for the test/service/intervention in the general or clinical population of interest?	☒	☐	☐
(4) Were all eligible participants that met the prespecified entry criteria enrolled?	☐	☐	☒
(5) Was the sample size sufficiently large to provide confidence in the findings? $	☐	☒	☐
(6) Was the test/service/intervention clearly described and delivered consistently across the study population?	☒	☐	☐
(7) Were the outcome measures prespecified, clearly defined, valid, reliable, and assessed consistently across all study participants?	☒	☐	☐
(8) Were the people assessing the outcomes blinded to the participants' exposures/interventions?	☐	☐	☒
(9) Was the loss to follow-up after baseline 20% or less? Were those lost to follow-up accounted for in the analysis? !	☒	☐	☐

Criteria	Yes	No	Cannot Determine; Not Applicable; Not Reported
(10) Did the statistical methods examine changes in outcome measures from before to after the intervention? Were statistical tests done that provided p values for the pre-to-post changes?	☒	☐	☐
(11) Were outcome measures of interest taken multiple times before the intervention and multiple times after the intervention (i.e., did they use an interrupted time-series design)?	☒	☐	☐
(12) If the intervention was conducted at a group level (e.g., a whole hospital, a community, etc.) did the statistical analysis take into account the use of individual-level data to determine effects at the group level?	☒	☐	☐

Table 8: *Quality Assessment Vlot et al. (2017)*

Legend: $ The study's discussion was used to answer the question;
! If either question can be answered with "yes", this criterion gets +1 point, even if loss to follow-up was > 20%

Study: Physical changes, laboratory parameters, and bone mineral density during testosterone treatment in adolescents with gender dysphoria (Stoffers, de Vries, & Hannema, 2019)

Criteria	Yes	No	Cannot Determine; Not Applicable; Not Reported
1. Was the study question or objective clearly stated?	☒	☐	☐
2. Were eligibility/selection criteria for the study population prespecified and clearly described?	☒	☐	☐
3. Were the participants in the study representative of those who would be eligible for the test/service/intervention in the general or clinical population of interest?	☒	☐	☐
4. Were all eligible participants that met the prespecified entry criteria enrolled?	☐	☒	☐
5. Was the sample size sufficiently large to provide confidence in the findings? $	☒	☐	☐
6. Was the test/service/intervention clearly described and delivered consistently across the study population?	☒	☐	☐
7. Were the outcome measures prespecified, clearly defined, valid, reliable, and assessed consistently across all study participants?	☒	☐	☐
8. Were the people assessing the outcomes blinded to the participants' exposures/interventions?	☐	☐	☒
9. Was the loss to follow-up after baseline 20% or less? Were those lost to follow-up accounted for in the analysis? !	☒	☐	☐

Criteria	Yes	No	Cannot Determine; Not Applicable; Not Reported
10. Did the statistical methods examine changes in outcome measures from before to after the intervention? Were statistical tests done that provided p values for the pre-to-post changes?	☒	☐	☐
11. Were outcome measures of interest taken multiple times before the intervention and multiple times after the intervention (i.e., did they use an interrupted time-series design)?	☒	☐	☐
12. If the intervention was conducted at a group level (e.g., a whole hospital, a community, etc.) did the statistical analysis take into account the use of individual-level data to determine effects at the group level?	☒	☐	☐

Table 9: *Quality Assessment Stoffers et al. (2019)*

Legend: $ The study's discussion was used to answer the question;
! If either question can be answered with "yes", this criterion gets +1 point, even if loss to follow-up was > 20%

Study: The effect of GnRH analogue treatment on bone mineral density in young adolescents with gender dysphoria: findings from a large national cohort (Joseph, Ting, & Butler, 2019)

Criteria	Yes	No	Cannot Determine; Not Applicable; Not Reported
1. Was the study question or objective clearly stated?	☒	☐	☐
2. Were eligibility/selection criteria for the study population prespecified and clearly described?	☒	☐	☐
3. Were the participants in the study representative of those who would be eligible for the test/service/intervention in the general or clinical population of interest?	☒	☐	☐
4. Were all eligible participants that met the prespecified entry criteria enrolled?	☐	☐	☒
5. Was the sample size sufficiently large to provide confidence in the findings? $	☒	☐	☐
6. Was the test/service/intervention clearly described and delivered consistently across the study population?	☒	☐	☐
7. Were the outcome measures prespecified, clearly defined, valid, reliable, and assessed consistently across all study participants?	☒	☐	☐
8. Were the people assessing the outcomes blinded to the participants' exposures/interventions?	☐	☐	☒
9. Was the loss to follow-up after baseline 20% or less? Were those lost to follow-up accounted for in the analysis? !	☒	☐	☐

Criteria	Yes	No	Cannot Determine; Not Applicable; Not Reported
10. Did the statistical methods examine changes in outcome measures from before to after the intervention? Were statistical tests done that provided p values for the pre-to-post changes?	☒	☐	☐
11. Were outcome measures of interest taken multiple times before the intervention and multiple times after the intervention (i.e., did they use an interrupted time-series design)?	☒	☐	☐
12. If the intervention was conducted at a group level (e.g., a whole hospital, a community, etc.) did the statistical analysis take into account the use of individual-level data to determine effects at the group level?	☒	☐	☐

Table 10: *Quality Assessment Joseph et al. (2019)*

Legend: $ The study's discussion was used to answer the question;
! If either question can be answered with "yes", this criterion gets +1 point, even if loss to follow-up was > 20%

Study: Bone Development in Transgender Adolescents Treated With GnRHa Analogues and Subsequent Gender-Affirming Hormones (Schagen, Wouters, Cohen-Kettenis, Gooren, & Hannema, 2020)

Criteria	Yes	No	Cannot Determine; Not Applicable; Not Reported
1. Was the study question or objective clearly stated?	☒	☐	☐
2. Were eligibility/selection criteria for the study population prespecified and clearly described?	☒	☐	☐
3. Were the participants in the study representative of those who would be eligible for the test/service/intervention in the general or clinical population of interest?	☒	☐	☐
4. Were all eligible participants that met the prespecified entry criteria enrolled?	☐	☐	☒
5. Was the sample size sufficiently large to provide confidence in the findings? $	☐	☐	☒
6. Was the test/service/intervention clearly described and delivered consistently across the study population?	☒	☐	☐
7. Were the outcome measures prespecified, clearly defined, valid, reliable, and assessed consistently across all study participants?	☒	☐	☐
8. Were the people assessing the outcomes blinded to the participants' exposures/interventions?	☐	☐	☒
9. Was the loss to follow-up after baseline 20% or less? Were those lost to follow-up accounted for in the analysis? !	☒	☐	☐

Criteria	Yes	No	Cannot Determine; Not Applicable; Not Reported
10. Did the statistical methods examine changes in outcome measures from before to after the intervention? Were statistical tests done that provided p values for the pre-to-post changes?	☒	☐	☐
11. Were outcome measures of interest taken multiple times before the intervention and multiple times after the intervention (i.e., did they use an interrupted time-series design)?	☒	☐	☐
12. If the intervention was conducted at a group level (e.g., a whole hospital, a community, etc.) did the statistical analysis take into account the use of individual-level data to determine effects at the group level?	☒	☐	☐

Table 11: *Quality Assessment Schagen et al. (2020)*

Legend: $ The study's discussion was used to answer the question;
! If either question can be answered with "yes", this criterion gets +1 point, even if loss to follow-up was > 20%

Study: Short-term outcomes of pubertal suppression in a selected cohort of 12 to 15 year old young people with persistent gender dysphoria in the UK (Carmichael, et al., 2021)

Criteria	Yes	No	Cannot Determine; Not Applicable; Not Reported
1. Was the study question or objective clearly stated?	☒	☐	☐
2. Were eligibility/selection criteria for the study population prespecified and clearly described?	☒	☐	☐
3. Were the participants in the study representative of those who would be eligible for the test/service/intervention in the general or clinical population of interest?	☒	☐	☐
4. Were all eligible participants that met the prespecified entry criteria enrolled?	☒	☐	☐
5. Was the sample size sufficiently large to provide confidence in the findings? [$]	☐	☒	☐
6. Was the test/service/intervention clearly described and delivered consistently across the study population?	☒	☐	☐
7. Were the outcome measures prespecified, clearly defined, valid, reliable, and assessed consistently across all study participants?	☒	☐	☐
8. Were the people assessing the outcomes blinded to the participants' exposures/interventions?	☐	☐	☒
9. Was the loss to follow-up after baseline 20% or less? Were those lost to follow-up accounted for in the analysis? [!]	☒	☐	☐

Criteria	Yes	No	Cannot Determine; Not Applicable; Not Reported
10. Did the statistical methods examine changes in outcome measures from before to after the intervention? Were statistical tests done that provided p values for the pre-to-post changes?	☒	☐	☐
11. Were outcome measures of interest taken multiple times before the intervention and multiple times after the intervention (i.e., did they use an interrupted time-series design)?	☒	☐	☐
12. If the intervention was conducted at a group level (e.g., a whole hospital, a community, etc.) did the statistical analysis take into account the use of individual-level data to determine effects at the group level?	☒	☐	☐

Table 12: *Quality Assessment Carmichael et al. (2021)*

Legend: $ The study's discussion was used to answer the question;
! If either question can be answered with "yes", this criterion gets +1 point, even if loss to follow-up was > 20%

Study: Pubertal Suppression, Bone Mass, and Body Composition in Youth With Gender Dysphoria
(Navabi, Tang, Khatchadourian, & Lawson, 2021)

Criteria	Yes	No	Cannot Determine; Not Applicable; Not Reported
1. Was the study question or objective clearly stated?	☒	☐	☐
2. Were eligibility/selection criteria for the study population prespecified and clearly described?	☒	☐	☐
3. Were the participants in the study representative of those who would be eligible for the test/service/intervention in the general or clinical population of interest?	☒	☐	☐
4. Were all eligible participants that met the prespecified entry criteria enrolled?	☒	☐	☐
5. Was the sample size sufficiently large to provide confidence in the findings? [$]	☒	☐	☐
6. Was the test/service/intervention clearly described and delivered consistently across the study population?	☒	☐	☐
7. Were the outcome measures prespecified, clearly defined, valid, reliable, and assessed consistently across all study participants?	☒	☐	☐
8. Were the people assessing the outcomes blinded to the participants' exposures/interventions?	☐	☐	☒
9. Was the loss to follow-up after baseline 20% or less? Were those lost to follow-up accounted for in the analysis? [!]	☒	☐	☐

Criteria	Yes	No	Cannot Determine; Not Applicable; Not Reported
10. Did the statistical methods examine changes in outcome measures from before to after the intervention? Were statistical tests done that provided p values for the pre-to-post changes?	☒	☐	☐
11. Were outcome measures of interest taken multiple times before the intervention and multiple times after the intervention (i.e., did they use an interrupted time-series design)?	☒	☐	☐
12. If the intervention was conducted at a group level (e.g., a whole hospital, a community, etc.) did the statistical analysis take into account the use of individual-level data to determine effects at the group level?	☒	☐	☐

Table 13: *Quality Assessment Navabi et al. (2021)*

Legend: $ The study's discussion was used to answer the question;
! If either question can be answered with "yes", this criterion gets +1 point, even if loss to follow-up was > 20%

Declaration of Authorship

I hereby declare that I have written this scientific paper independently, without the unauthorized help of third parties, and without using sources other than those indicated. The data, concepts and other contents taken directly or indirectly from other sources are marked as such by references. This work has not been submitted in the same or similar form to any other examination authority in Germany or abroad, has not been part of a course of study or examination and has not yet been published.

Ich erkläre hiermit an Eides statt, dass ich die vorliegende wissenschaftliche Ausarbeitung eigenständig ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten, Konzepte und anderen Inhalte sind unter Angabe des Literaturzitats gekennzeichnet. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt, war bisher nicht Bestandteil einer Studien- oder Prüfungsleistung und ist noch nicht veröffentlicht.

Ort, Datum, Unterschrift