APPLICATION OF STANDARD, INTERMEDIATE, AND ADVANCED EVIDENCE SYNTHESES METHODS TO INFORM DECISION MAKING ABOUT OPIOID USE AND GENDER-AFFIRMING CARE

By Anna Miroshnychenko, M.Sc.

A Thesis submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

McMaster University DOCTOR OF PHYLOSOPHY (2025)

Hamilton, Ontario (Health Research Methodology)

Title: Application of standard, intermediate, and advanced evidence syntheses methods to inform decision making about opioid use and gender-affirming care

AUTHOR: Anna Miroshnychenko, M.Sc.

SUPERVISOR: Dr. Romina Brignardello-Petersen, Associate Professor

NUMBER OF PAGES: xii, 200

ABSTRACT

Opioid crisis in North America called for an evidence synthesis to compare the effects of analgesics for the management of acute dental pain. Uncertainty about the effects of gender-affirming interventions required a series of systematic reviews and meta-analyses. We used standard, intermediate, and advanced methods to create these evidence syntheses. This thesis presents four systematic reviews that address a total of 44 comparisons, 54 outcomes, 185 included studies. In terms of advanced methods, the best available evidence assessing the comparative effectiveness of acetaminophen, NSAIDs and opioids, ranging from moderate to high certainty, was derived from numerous RCTs, and we performed a systematic review and network meta-analysis. We used Grading of Recommendations Assessment, Development and Evaluation (GRADE) guidance for network meta-analyses and used an automated tool to rate the certainty of the evidence for direct, indirect, and network estimates of effect. For data interpretation and clarity of presentation, we classified the interventions from the most to the least effective by considering the estimate of effect and the certainty of the evidence and organized these data according to a colour coding system. Based on moderate and high certainty evidence, our systematic review and network meta-analysis demonstrated that NSAIDs with or without acetaminophen result in better pain-related outcomes than opioids with or without acetaminophen. As numerous outdated systematic reviews and meta-analyses about the effects of corticosteroids have been published, for our systematic review, we searched the Epistemonikos database and the Living Overview of Evidence (LOVE) platform that utilizes artificial

intelligence. With low and very low certainty evidence, our systematic review and meta-analysis suggested that there is a trivial (unimportant) difference in postoperative pain intensity and postoperative infection after administration of corticosteroids orally, submucosally, or intra-muscularly compared to placebo in patients undergoing third molar extractions. Research about gender dysphoria has been a subject of contentious discussion. Therefore, when conducting systematic reviews and meta-analyses about gender-affirming hormone therapy and gender-affirming mastectomy for individuals experiencing gender dysphoria, we devised a plan for minimization and management of conflicts of interest to demonstrate the integrity of our work. The systematic reviews and meta-analyses about the interventions to manage gender dysphoria in children and young adults showed that the current best available evidence about the effects of gender-affirming hormone therapy and mastectomy comes mostly from the methodologically limited before-after and case series studies, and ranges from high to very low certainty. As the fields of dentistry and gender medicine are advancing rapidly, researchers are challenged with creating and appropriately using methods for synthesizing evidence into systematic reviews and (network) meta-analyses to produce authentic results.

ACKNOWLEDGMENTS

The publications included in this thesis are a culmination of great effort from numerous talented individuals, and I am tremendously grateful to them all.

Thank you to my supervisor, Dr. Romina Brignardello-Petersen, for her invaluable expertise, counsel, and mentorship throughout my PhD degree. Thank you, Romina, for your support, guidance, and for providing the freedom to explore relevant professional initiatives throughout my degree to enrich my experience.

Thank you to my thesis committee members, Dr. Gordon Guyatt and Dr. Mo Bhandari, for providing their invaluable expertise and guidance with the projects included in this thesis.

Thank you to my colleagues who made enormous contributions to these publications, including Dr. Yetiani Roldan, Sara Ibrahim, Maria Azab, and many others.

Thank you to the team at the American Dental Association, University of Pennsylvania, and the University of Pittsburgh for their leadership on the projects related to acute dental pain. Thank you to the team at the Society for Evidence-based Gender Medicine for their contributions to the projects related to gender-affirming treatments.

Thank you to the staff at the Department of Health Research Methods, Evidence, and Impact for their guidance, mentorship, and assistance throughout both of my degrees.

And most importantly, an enormous thank you to my family for all of their unwavering support throughout all of my undergraduate and graduate studies.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

LIST OF TABLES

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

Table Number and Title	Page

CHAPTER 6

Table Number and Title	Page

LIST OF APPENDIXES

CHAPTER 6

Table Number and Title	Page

LIST OF ABBREVIATIONS

Abbreviation	Definition
ADA	American Dental Association
ADASRI	American Dental Association Science & Research Institute
AI	Artificial Intelligence
BMD	Bone Mineral Density
CI	Confidence Interval
COX	Cyclooxygenase
CPG	Clinical Practice Guideline
CSH	Cross-Sex Hormones
DSM	Diagnostic and Statistical Manual of Mental Disorders
EBM	Evidence-Based Medicine
FDA	United States Food and Drug Administration
FST	Follicle Stimulating Hormone
GAHT	Gender-Affirming Hormone Therapy
GD	Gender Dysphoria
GI	Gastrointestinal
GnRHa	Gonadotropin Releasing Hormone Analogues
GRADE	Grading of Recommendations, Assessment, Development, and Evaluation
ICD-11	International Classification of Diseases, Eleventh Edition
ICEMAN	Instrument for Credibility of Effect Modification Analyses
LH	Luteinizing Hormone
LOVE	Living Overview of Evidence Platform
MA	Meta-Analysis
MC	Mean Change
MD	Mean Difference
NF	Natal Females
NM	Natal Males
NMA	Network Meta-Analysis
NNT	Needed to Treat
NR	Not Reported
NSAIDs	Non-Steroidal Anti-Inflammatory Drugs
OIS	Optimal Information Size
OR	Odds Ratio
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT	Randomized Controlled Trial
RD	Risk Difference
ROBINS-I	Risk of Bias Tool for Non-Randomized Studies of Interventions
SD	Standard Deviation
SEGM	Society for Evidence-based Gender Medicine
SMD	Standardized Mean Difference
SPID	Summed Pain Intensity Difference
SR	Systematic Review
TOTPAR	Total Pain Relief
USA	United States of America

DECLARATION OF ACADEMIC ACHIEVEMENT

Chapter 1: This chapter is unpublished. AM is the sole author.

Chapter 2: This chapter is published in the *Journal of Dental Research*. AM contributed to conception and design, data acquisition, analysis and interpretation, and drafted and critically revised the manuscript. SI and MA contributed to conception and design, data acquisition, analysis and interpretation, and critically revised the manuscript. YR, DT, LH, and JWL contributed to data acquisition. JPDM contributed to data analysis. OU, MP, DEP, PAM, EVH, BC, ACL contributed to conception and design, data interpretation, and critically revised the manuscript. RBP contributed to conception and design, data analysis and interpretation, and critically revised the manuscript.

Chapter 3: This chapter is published in the *Journal of the American Dental Association*. AM contributed to conception and design, data acquisition, analysis and interpretation, and drafted and critically revised the manuscript. SI and MA contributed to conception and design, data acquisition, analysis and interpretation, and critically revised the manuscript. YR, DT, and LH contributed to data acquisition. JPDM contributed to data analysis. OU, FVP, MT contributed to conception and design, data acquisition, and critically revised the manuscript. DEP, PAM, EVH, RBP and ACL contributed to conception and design, data analysis and interpretation, and critically revised the manuscript.

Chapter 4: This chapter is published in the *Archives of Disease in Childhood by the BMJ Group*. AM contributed to the conception and design, data collection, analysis and interpretation, and drafted and critically revised the manuscript. YR and SI contributed to data collection, analysis and interpretation, and critically revised the manuscript. CKM contributed to the conception and design and critically revised the manuscript. SM contributed to the conception and design, data interpretation, and critically revised the manuscript. RC contributed to data collection. GG critically revised the manuscript. RBP contributed to conception and design, data interpretation, and critically revised the manuscript. RBP is the guarantor of this work.

Chapter 5: This chapter is published in the *Journal of Plastic and Reconstructive Surgery*. AM contributed to the conception and design, data acquisition, analysis and interpretation, and drafted and critically revised the manuscript. YR and SI contributed to data acquisition, analysis and interpretation, and critically revised the manuscript. CHM and KD contributed to the conception and design and critically revised the manuscript. SM contributed to the conception and design, data interpretation, and critically revised the manuscript. RC contributed to data acquisition. GG critically revised the manuscript. RBP contributed to conception and design, data interpretation, and critically revised the manuscript.

Chapter 6: This chapter is unpublished. AM is the sole author.

CHAPTER 1: INTRODUCTION OF THE THESIS

INTRODUCTION

Evidence Syntheses and Clinical Practice Guidelines for Decision-Making

A clinical practice guideline (CPG) is a series of statements that include recommendations for the best management of a particular clinical condition or diagnosis [1]. The development of a CPG is informed by rigorous evidence syntheses and an assessment of the benefits and harms of existing management options. A CPG is developed by a team of methodologists, an evidence synthesis team, and a panel of clinical experts in the field. The team of methodologists assists the panel of experts throughout the entire process of developing a CPG. The organization spearheading the development of the CPG, along with the panel of experts and methodologists, determines the scope, purpose, target audience, and clinical questions for the guideline. The evidence synthesis team conducts the evidence syntheses that aim to address the clinical questions. Subsequently, the team of methodologists, along with the panel of experts, uses the results of the evidence syntheses to assess the benefits and harms of the management options, patients' values and preferences related to the management options, and resources, acceptability, feasibility, and equity associated with the management options in order to formulate recommendations for the CPG.

Evidence syntheses can include a series of systematic reviews and meta-analyses or network meta-analyses that synthesize the best available evidence to address the clinical questions posed by the CPG. In addition to developing CPGs, evidence syntheses can be used independently by clinicians, patients, and policy makers to make

decisions about the management of a particular clinical condition or diagnosis. Evidence syntheses can answer questions related to the benefits and harms associated with choosing a particular treatment option in comparison to placebo or other treatment options. Evidence syntheses can also be used when conducting economic analyses and health technology assessments by providing data about the clinical effectiveness of a treatment option.

There are two topics for which I led an extensive evidence synthesis. The first evidence synthesis, consisting of four publications, was used to inform the development of two clinical practice guidelines for the management of acute dental pain in adults and children. The second evidence synthesis, consisting of five publications, was conducted to inform patients, clinicians, and policy makers about the current best available evidence for the management of gender dysphoria in individuals below 26 years of age when making decisions about existing treatment options. In this thesis, I describe and discuss the standard, intermediate and advanced syntheses methods to inform decision making about the use of opioids and gender-affirming treatments.

Acute Dental Pain

In North America, dentistry as a field has existed for several centuries. One of the first dentists in America was an English surgeon and dentist John Baker, who settled in Boston in 1763 [2]. Various analgesics used to manage the acute dental pain today, such as acetaminophen, non-steroidal anti-inflammatory drugs (NSAIDs) and opioids, have existed for many years and originally had a different purpose.

Acetaminophen was first synthesized in 1878 and was first used to treat any pain and fever in 1893 [3]. An existing explanation for its mechanism is that it blocks one of the cyclo-oxygenase (COX) enzymes at the pain site [3, 4], which is also a known mechanism of action of NSAIDs [3, 5]. However, the effect of acetaminophen at the pain site is too weak to be responsible for pain relief, therefore it is hypothesized that its effect may be exerted by blocking the enzyme production in the brain, thus blocking further transmission of the pain nerve impulses [3, 6]. If this is the case, it could be hypothesized that prolonged use of acetaminophen may undesirably affect the development of the brain, especially in children [7].

Ibuprofen, a NSAID, was derived from propionic acid (a food preservative) during the 1960s in efforts to find a safer alternative to aspirin [8]. In 1969, ibuprofen was launched in the United Kingdom to manage rhematic diseases and marketed as prescription medication, Brufen [8]. In 1983, ibuprofen was approved as an over-the-counter medicine in the United Kingdom, followed by the United States of America (USA) in 1984 [8]. Naproxen, another NSAID, was developed by Atnahs Pharma [9]. In 1976, the United States Food and Drug Administration approved Naproxen for the management of autoimmune disorders such as rheumatoid arthritis and juvenile rheumatoid arthritis, under the brand name Naprosyn [9]. Neither ibuprofen nor naproxen treat the root cause of the rhematic diseases.

The use of opioids to treat pain became prevalent in the USA in the early 1860s as a way to treat wounded soldiers during the civil war [10]. These soldiers were treated with morphine, and many developed dependencies and addictions to the drug in the

years following the war [10]. Ironically, in 1898, the Bayer Company introduced heroin as a safe and effective pain reliever and cough suppressant, with the assertion that it was less habit-forming than morphine [11]. Throughout the 1910s-1920s, the USA placed restrictions on opioids and narcotics, requiring a formal prescription. In 1995, Purdue Pharma developed OxyContin, a version of oxycodone, which was presented as a "gentler and less-addictive opioid" [12]. Over the next two decades, clinicians increased the prescription of this and other opioids to treat pain and, therefore, increased the number of individuals who developed an addiction [12]. Despite lawsuits taken against Purdue Pharma, opioids continued to be heavily prescribed, resulting in large numbers of individuals presenting with addictions and dying by overdose. In October 2017, the USA was officially declared to be in a public health crisis and measures keep being taken to reduce it.

Dentists prescribe acetaminophen, NSAIDs, opioids and other analgesics such as corticosteroids and local anesthetics to alleviate acute pain following various dental procedures and conditions, including tooth extraction and symptomatic irreversible pulpitis. Over the last few years in North America, prescription of opioids by dental clinicians appears to account for 5% to 10% of all opioid prescriptions [13]. Dentists prescribe opioids to manage the acute dental pain and have been recorded to prescribe large amounts of analgesics for extended periods of time [14]. Among individuals with at least one dental visit, it is estimated that 28.6% of adults and 2.7% of children received an opioid in 2012 [15]. Although these numbers have presumably declined to

12.2% of adults and 1.6% of children in 2019, opioid use and misuse are still a leading cause of morbidity and mortality in North America [15].

Management of acute dental pain with opioids is associated with a series of adverse effects [16-19]. Seventy five percent of individuals who develop an opioid use disorder start with a prescription for opioid analgesics [20]. Fifty four percent of patients fill but do not fully consume their prescribed opioids, thus potentially increasing the risk of misuse [21]. Adolescents are at an increased risk of developing an opioid use disorder [22, 23]. In 2020, 19% of adolescents aged 12-18 years had used an opioid prescription in the past 12 months; 15.7% had used medically as prescribed and 3.2% had misused [24]. Further, dental prescriptions at the recommended morphine milligram equivalents have been associated with an increased risk of adverse outcomes, including major depressive and anxiety disorders [25, 26]. Since the announcement of an opioid crisis in 2017, various academic and government bodies in the USA and Canada have taken action to provide guidance for the use of opioids for dental procedures.

In Canada, the Royal College of Dental Surgeons of Ontario released opioid prescribing guidelines for dental clinicians in November of 2015 [27]. The guideline recommended prescription of opioids as a third-line therapy for acute dental pain after attempting acetaminophen and NSAIDs. After publication of the guideline, dentists in the Canadian province of Ontario issued 1,571,897 opioid prescriptions to 1,157,102 patients over a period of five years (i.e., between 2012 and 2017). The guideline was not associated with a change in opioid dispensing rates, however it was associated with a statistically significant reduction in the volume of opioids dispensed (28.1%

reduction in opioid units such as milligram morphine equivalents per 100 population between 2015 and 2017) [27]. The guideline suggested that there remained to be uncertainty about the comparative effectiveness of opioids, NSAIDs, and acetaminophen [27].

In 2020, the American Dental Association Science & Research Institute (ADASRI), the University of Pittsburgh School of Dental Medicine, and the Center for Integrative Global Oral Health at the University of Pennsylvania aimed to produce the evidence-based clinical practice guidelines (CPGs) for the pharmacological management of acute dental pain consecutive to simple and surgical tooth extractions and pain associated with pulpitis and its complications. In order to provide the best available evidence to inform the CPGs [28, 29], we conducted a systematic review and network meta-analysis to assess the comparative effectiveness of opioids, NSAIDs, and acetaminophen in adults [30]. Further, we conducted three systematic reviews and meta-analyses to assess the comparative effectiveness of analgesics to manage the acute dental pain in children [31], corticosteroids to manage the acute dental pain in adults [32], and local anesthetics to treat acute dental pain in adults [30].

In this evidence synthesis, we included randomized controlled trials (RCTs) only, as we identified a relatively large amount of RCTs meeting the inclusion criteria. We assessed a set of specific outcomes at specific time points determined by the guideline panel with the guidance from peer-reviewed evidence. For the systematic review and network meta-analysis assessing the comparative effectiveness of analgesics to manage the acute dental pain, each treatment option was carefully selected by the guideline

panel and could include more than one dose for a monotherapy or a combination [30]. In terms of analyses, we performed frequentists network meta-analyses (NMAs) using a random effects model in RStudio. We used an automated tool to rate the certainty of the evidence for direct, indirect, and network estimates of effect, and used a single threshold approach for the assessment of imprecision using the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidance. Further, we used a classification system to rate the interventions from the most to the least effective in the management of acute dental pain by taking into consideration the magnitude of effect and the certainty of the evidence, and we organized this information according to an easily interpretable colour coding system. For the systematic review and meta-analysis assessing the effectiveness of corticosteroids for the management of acute pain following dental extractions, we used a similar methodology as well as a unique approach to retrieving relevant evidence via a collaboration with the Epistemonikos database and the Living Overview of Evidence (LOVE) platform that utilizes artificial intelligence in addition to input from methodologists [32, 33].

Gender Dysphoria

In North America, the field of gender medicine seems to have emerged with the John Hopkins Gender Clinic, which opened in 1965. It was the first gender clinic in the USA to provide comprehensive care for transgender individuals including counseling, hormone therapy, and gender-affirming surgeries [34].

Later in North America, Harry Benjamin, an endocrinologist studying gender diversity, published a book, *The Transsexual Phenomenon* [35]. Before publishing this book, Harry Benjamin studied with Magnus Hirschfeld at the institute for Sexual Science in Berlin. Magnus Hirschfeld attempted to explain the experience of gender diversity by creating an initial framework [35]. While Hirschfeld's colleagues aimed to cure gender diverse patients, he developed and implemented "adaptation therapy" at his institute in Berlin to help patients live "according to their nature" [35]. Harry Benjamin adopted many of Hirschfeld's beliefs while studying with him. *The Transsexual Phenomenon* provided the foundation for modern transgender and gender diverse health care by advising that hormonal and surgical treatments are therapeutic and life-saving for this patient population [35].

In 1980, the American Psychiatric Association added the diagnosis of "gender identity disorder" to the third *Diagnostic and Statistical Manual of Mental Disorders (DSM-3)* [35]. In 2013, this diagnosis was reconceptualized as "gender dysphoria" in the *DSM-5* [35]. According to the *DSM-5*, gender dysphoria refers to psychological distress resulting from a "marked incongruence between one's experienced or expressed gender and assigned gender" [36]. According to the World Health Organization's *International Classification of Diseases, Eleventh Edition (ICD-11)*, a similar condition of "Gender Incongruence of Adolescence or Adulthood" refers to "a marked and persistent incongruence between an individual's experienced gender and the assigned sex, which often leads to a desire to transition in order to live and be accepted as a person of the experienced gender, through hormonal treatment, surgery or other

health care services to make the individual's body align, as much as desired and to the extent possible, with the experienced gender". Further, the *ICD-11* defines "Gender Incongruence of Childhood" as a "marked incongruence between an individual's experienced/expressed gender and the assigned sex in pre-pubertal children".

Today, individuals experiencing gender dysphoria or gender incongruence may choose to pursue gender-affirming treatments [37]. These gender-affirming treatments may include social gender transition (SGT), medical affirmation such as puberty blockers and gender-affirming hormones, and surgical affirmation such as mastectomy or masculinizing chest surgery, masculinizing phalloplasty, feminizing vaginoplasty, facial feminization procedures, reduction thyrochondroplasty, and voice surgery [37].

Currently, there are two approaches to gender-affirmation treatment. The first approach is the gender-affirmation approach. This approach prioritizes assisting youth in social and medical changes to affirm their experienced gender [38]. The second approach is the watchful waiting approach. The aim of this approach is to allow time for youth to explore and mature [39-41]. This aim is based on the evidence that for a proportion of individuals who experience childhood gender dysphoria, the dysphoria resolves by late adolescence or young adulthood [39, 42, 43].

There is currently uncertainty with respect to the effects of gender-affirming treatments. There were no systematic reviews and meta-analyses assessing the effects of gender affirming treatments, i.e., social gender transition, chest binding and genital tucking, gender-affirming mastectomy, puberty blockers and gender-affirming

hormones, that examined the risk of bias in each individual study as well as assessed the certainty of the evidence of each outcome of interest. In order to investigate the effects of gender-affirming treatments, we performed an evidence synthesis consisting of five systematic reviews and meta-analyses to assess the effects of social gender transition (submitted as a report), chest binding and genital tucking (submitted for publication), mastectomy [44], puberty blockers [45], and gender-affirming hormones [46].

The evidence synthesis to address interventions for gender dysphoria included observational studies only, ranging from high to very low certainty, and we did not identify any RCTs meeting our inclusion criteria. In this evidence synthesis, due to the sensitive nature of the subject matter, we aimed to provide an impartial introduction in every publication, describing the existing definition of gender dysphoria and highlighting all existing management approaches for gender dysphoria. We aimed to transparently disclose any conflicts of interest as well as devised a plan for the minimization and management of conflicts of interest. In our efforts to communicate the results in accordance with the latest methodological guidance, we were diligent to not make recommendations within the evidence syntheses, as only CPGs are appropriately positioned to do this.

The objective of this thesis is to illustrate and discuss the standard, intermediate, and advanced syntheses methods used to inform decision making about the use of opioids and gender-affirming treatments. This thesis is based on four peer-reviewed publications that illustrate the application of standard, intermediate and advanced

methods, two publications included in the first evidence synthesis and two publications included in the second evidence synthesis. The last chapter summarizes the standard, intermediate, and advanced methods in all included publications. The last chapter also highlights important findings of all included publications and discusses possible future research for both topics.

Chapter 2 focuses on the systematic review and network meta-analysis to compare the effectiveness of opioids, NSAIDs, and acetaminophen for the management of acute dental pain. This chapter presents the application of the standard, intermediate, and advanced methods used in this systematic review and network meta-analysis. This chapter also presents important findings of this review that were used to inform a CPG for the management of acute dental pain in adults by the American Dental Association.

Chapter 3 focuses on the systematic review and meta-analysis to assess the comparative effectiveness of corticosteroids for the management of pain subsequent to surgical tooth extraction in adults. This chapter presents the application of the standard, intermediate, and advanced methods used in this systematic review and meta-analysis. This chapter also presents important findings of this review that were used to inform a CPGs for the management of acute dental pain in adults by the American Dental Association.

Chapter 4 focuses on the assessment of the effects of gender-affirming hormone therapy for youth experiencing gender dysphoria. This chapter presents the application of the standard, intermediate, and advanced methods used in this systematic review and

meta-analysis. This chapter also highlights important findings of this review that are being used by patients, clinicians, policy makers and other stakeholders regarding the management of gender dysphoria in youth.

Chapter 5 focuses on the assessment of the effects of mastectomy for individuals experiencing gender dysphoria. This chapter presents application of the standard, intermediate, and advanced methods used in this systematic review and meta-analysis. This chapter also highlights important findings of this review that are being used by patients, clinicians, policy makers and other stakeholders regarding the management of gender dysphoria in youth and young adults.

This thesis ends with **Chapter 6**, which is a discussion of the standard, intermediate, and advanced methods used to generate evidence syntheses to inform decision-making for the management of acute dental pain and gender dysphoria. This chapter also highlights important findings of all included publications, while exploring future direction for the use of analgesics to manage acute dental pain and gender-affirming interventions to manage gender dysphoria.

REFERENCES

1. Brignardello-Petersen, R., A. Carrasco-Labra, and G.H. Guyatt, *How to Interpret and Use a Clinical Practice Guideline or Recommendation: Users' Guides to the Medical Literature.* JAMA, 2021. **326**(15): p. 1516-1523.
2. Kamen, S. and J. Dorfman. *Dentistry in 18th- and 19th-century America*. 2024; Available from: https://www.britannica.com/science/dentistry/Dentistry-in-18th-and-19th-century-America.
3. Shaughnessy, A., *How Does Acetaminophen Work?* 2022.
4. Botting, R.M., *Mechanism of Action of Acetaminophen: Is There a Cyclooxygenase 3?* Clinical Infectious Diseases, 2000. **31**(Supplement_5): p. S202-S210.
5. Ghlichloo I. and Gerriets V., *Nonsteroidal Anti-Inflammatory Drugs (NSAIDs).* StatPearls [Internet], 2023.
6. Ohashi, N. and T. Kohno, *Analgesic Effect of Acetaminophen: A Review of Known and Novel Mechanisms of Action.* Front Pharmacol, 2020. **11**: p. 580289.
7. Parker, W., et al., *The Dangers of Acetaminophen for Neurodevelopment Outweigh Scant Evidence for Long-Term Benefits.* Children (Basel), 2023. **11**(1).
8. Connelly, D. *A brief history of ibuprofen*. 2017; Available from: https://pharmaceutical-journal.com/article/infographics/a-brief-history-of-ibuprofen.
9. Zhao, J.-H., et al., *Clinically approved small-molecule drugs for the treatment of rheumatoid arthritis.* European Journal of Medicinal Chemistry, 2023. **256**: p. 115434.
10. *Opiate Addiction in the Civil War's Aftermath*. Available from: https://virginiahistory.org/learn/opiate-addiction-civil-wars-aftermath#:~:text=During%20the%20Civil%20War%2C%20surgeons,%E2%80%94too%20well%2C%20in%20fact.
11. Hosztafi, S., *[The history of heroin].* Acta Pharm Hung, 2001. **71**(2): p. 233-42.
12. Van Zee, A., *The promotion and marketing of oxycontin: commercial triumph, public health tragedy.* Am J Public Health, 2009. **99**(2): p. 221-7.
13. Richard, P., et al., *Opioid prescribing for surgical dental procedures in dental clinics of military treatment facilities.* J Am Dent Assoc, 2021. **152**(2): p. 94-104.e18.
14. Koppen, L., et al., *Dentists' prescribing of antibiotics and opioids to Medicare Part D beneficiaries: Medications of high impact to public health.* J Am Dent Assoc, 2018. **149**(8): p. 721-730.
15. Okunev, I., J. Frantsve-Hawley, and E. Tranby, *Trends in national opioid prescribing for dental procedures among patients enrolled in Medicaid.* The Journal of the American Dental Association, 2021. **152**(8): p. 622-630.e3.

16. Brummett, C.M., et al., *New Persistent Opioid Use After Minor and Major Surgical Procedures in US Adults.* JAMA Surg, 2017. **152**(6): p. e170504.
17. Donohue, J.M., et al., *Patterns of Opioid Administration Among Opioid-Naive Inpatients and Associations With Postdischarge Opioid Use: A Cohort Study.* Ann Intern Med, 2019. **171**(2): p. 81-90.
18. Goesling, J., et al., *Trends and predictors of opioid use after total knee and total hip arthroplasty.* Pain, 2016. **157**(6): p. 1259-1265.
19. Inacio, M.C., et al., *Risk factors for persistent and new chronic opioid use in patients undergoing total hip arthroplasty: a retrospective cohort study.* BMJ Open, 2016. **6**(4): p. e010664.
20. Cicero, T.J., et al., *The changing face of heroin use in the United States: a retrospective analysis of the past 50 years.* JAMA Psychiatry, 2014. **71**(7): p. 821-6.
21. Maughan, B.C., et al., *Unused opioid analgesics and drug disposal following outpatient dental surgery: A randomized controlled trial.* Drug Alcohol Depend, 2016. **168**: p. 328-334.
22. Quinn, P.D., et al., *Association of Opioid Prescription Initiation During Adolescence and Young Adulthood With Subsequent Substance-Related Morbidity.* JAMA Pediatr, 2020. **174**(11): p. 1048-1055.
23. Schroeder, A.R., et al., *Association of Opioid Prescriptions From Dental Clinicians for US Adolescents and Young Adults With Subsequent Opioid Use and Abuse.* JAMA Intern Med, 2019. **179**(2): p. 145-152.
24. Griesler, P.C., et al., *Assessment of Prescription Opioid Medical Use and Misuse Among Parents and Their Adolescent Offspring in the US.* JAMA Network Open, 2021. **4**(1): p. e2031073-e2031073.
25. Khouja, T., et al., *Serious opioid-related adverse outcomes associated with opioids prescribed by dentists.* Pain, 2022. **163**(8): p. 1571-1580.
26. Rosoff, D.B., G.D. Smith, and F.W. Lohoff, *Prescription Opioid Use and Risk for Major Depressive Disorder and Anxiety and Stress-Related Disorders: A Multivariable Mendelian Randomization Analysis.* JAMA Psychiatry, 2021. **78**(2): p. 151-160.
27. Guan, Q., et al., *Assessing the impact of an opioid prescribing guideline for dentists in Ontario, Canada.* J Am Dent Assoc, 2020. **151**(1): p. 43-50.
28. Carrasco-Labra, A., et al., *Evidence-based clinical practice guideline for the pharmacologic management of acute dental pain in children: A report from the American Dental Association Science and Research Institute, the University of Pittsburgh School of Dental Medicine, and the Center for Integrative Global Oral Health at the University of Pennsylvania.* J Am Dent Assoc, 2023. **154**(9): p. 814-825.e2.
29. Carrasco-Labra, A., et al., *Evidence-based clinical practice guideline for the pharmacologic management of acute dental pain in adolescents, adults, and older adults: A report from the American Dental Association Science and Research Institute, the University of Pittsburgh, and the University of Pennsylvania.* The Journal of the American Dental Association, 2024. **155**(2): p. 102-117.e9.

30. Miroshnychenko, A., et al., *Acute Postoperative Pain Due to Dental Extraction in the Adult Population: A Systematic Review and Network Meta-analysis.* J Dent Res, 2023. **102**(4): p. 391-401.
31. Miroshnychenko, A., et al., *Analgesics for the management of acute dental pain in the pediatric population: A systematic review and meta-analysis.* J Am Dent Assoc, 2023. **154**(5): p. 403-416.e14.
32. Miroshnychenko, A., et al., *Corticosteroids for managing acute pain subsequent to surgical extraction of mandibular third molars: A systematic review and meta-analysis.* J Am Dent Assoc, 2023. **154**(8): p. 727-741.e10.
33. *Epistemonikos*. 2024; Available from: https://www.epistemonikos.org/.
34. *LGBTQ History at JHU.*
35. Naz Khan, F., *A History of Transgender and Gender Diverse Health Care: From Medical Mistreatment to Gender-Affirmative Health Care*, in *Transgender and Gender Diverse Health Care: The Fenway Guide*, A.S. Keuroghlian, J. Potter, and S.L. Reisner, Editors. 2022, McGraw Hill: New York, NY.
36. American Psychiatric Association, *The Diagnostic and Statistical Manual of Mental Disorders (5th ed DSM-5)*, ed. A. American Psychiatric and D.S.M.T.F. American Psychiatric Association. 2013, Arlington, VA: American Psychiatric Association.
37. Vandermorris, A. and D.L. Metzger, *An affirming approach to caring for transgender and gender-diverse youth.* Paediatr Child Health, 2023. **28**(7): p. 437-448.
38. Byne, W., et al., *Report of the American Psychiatric Association Task Force on Treatment of Gender Identity Disorder.* Arch Sex Behav, 2012. **41**(4): p. 759-96.
39. de Vries, A.L. and P.T. Cohen-Kettenis, *Clinical management of gender dysphoria in children and adolescents: the Dutch approach.* J Homosex, 2012. **59**(3): p. 301-20.
40. Drescher, J., P. Cohen-Kettenis, and S. Winter, *Minding the body: situating gender identity diagnoses in the ICD-11.* Int Rev Psychiatry, 2012. **24**(6): p. 568-77.
41. Zucker, K.J., *Adolescents with Gender Dysphoria: Reflections on Some Contemporary Clinical and Research Issues.* Arch Sex Behav, 2019. **48**(7): p. 1983-1992.
42. Cantor, J.M., *Transgender and Gender Diverse Children and Adolescents: Fact-Checking of AAP Policy.* J Sex Marital Ther, 2020. **46**(4): p. 307-313.
43. Levine, S.B., E. Abbruzzese, and J.W. Mason, *Reconsidering Informed Consent for Trans-Identified Children, Adolescents, and Young Adults.* J Sex Marital Ther, 2022. **48**(7): p. 706-727.
44. Miroshnychenko, A., et al., *"Mastectomy for individuals with gender dysphoria below 26 years of age: A systematic review and meta-analysis".* Plast Reconstr Surg, 2024.
45. Miroshnychenko, A., et al., *Puberty blockers for gender dysphoria in youth: A systematic review and meta-analysis.* Arch Dis Child, 2025.

46. Miroshnychenko, A., et al., *Gender affirming hormone therapy for individuals with gender dysphoria aged <26 years: a systematic review and meta-analysis.* Arch Dis Child, 2025.

CHAPTER 2: ACUTE POSTOPERATIVE PAIN DUE TO DENTAL EXTRACTION IN THE ADULT POPULATION: A SYSTEMATIC REVIEW AND NETWORK META-ANALYSIS

Authors: Anna Miroshnychenko, Sara Ibrahim, Maria Azab, Yetiani Roldan, Juan Pablo Diaz Martinez, Divya Tamilselvan, Leon He, Joshua Little, Olivia Urquhart, Malavika Tampi, Deborah Polk, Paul Moore, Elliot Hersh, Bill Claytor, Alonso Carrasco-Labra, Romina Brignardello-Petersen

Published: Journal of Dental Research [January 11, 2023]

ABSTRACT

Objective: This study compares the effectiveness of pharmacological treatments to develop guidelines for the management of acute pain after tooth extraction.

Methods: We searched Medline, EMBASE, CENTRAL, and US Clinical Trials registry on November 21, 2020. We included randomized clinical trials (RCTs) of participants undergoing dental extractions comparing 10 interventions, including acetaminophen, nonsteroidal anti-inflammatory drugs (NSAIDs), opioids, and combinations to placebo. After duplicate screening and data abstraction, we conducted a frequentist network meta-analysis for each outcome at 6 h (i.e., pain relief, total pain relief [TOTPAR], summed pain intensity difference [SPID], global efficacy rating, rescue analgesia, and adverse effects). We assessed the risk of bias using a modified Cochrane RoB 2.0 tool and the certainty of evidence using the Grading of Recommendations, Assessment, Development, and Evaluation approach. We implemented the analyses in RStudio version 3.5.3 and classified interventions from most to least beneficial or harmful.

Results: We included 82 RCTs. Fifty-six RCTs enrolling 9,095 participants found moderate- and high-certainty evidence that ibuprofen 200 to 400 mg plus acetaminophen 500 to 1,000 mg (mean difference compared to placebo [MDp], 1.68; 95% confidence interval [CI], 1.06-2.31), acetaminophen 650 mg plus oxycodone 10 mg (MDp, 1.19; 95% CI, 0.85-1.54), ibuprofen 400 mg (MDp, 1.31; 95% CI, 1.17-

1.45), and naproxen 400-440 mg (MDp, 1.44; 95% CI, 1.07-1.80) were most effective for pain relief on a 0 to 4 scale. Oxycodone 5 mg, codeine 60 mg, and tramadol 37.5 mg plus acetaminophen 325 mg were no better than placebo. The results for TOTPAR, SPID, global efficacy rating, and rescue analgesia were similar.

Conclusion: *Based on moderate- and high-certainty evidence, NSAIDs with or without acetaminophen result in better pain-related outcomes than opioids with or without acetaminophen (except acetaminophen 650 mg plus oxycodone 10 mg) or placebo. Based on low- and very low-certainty evidence, most interventions were classified as no more harmful than placebo for most selected adverse effects.*

INTRODUCTION

North America is amid an opioid crisis, which is a leading public health and safety concern. In dentistry, many patients are prescribed opioids for the first time to manage acute postoperative pain after dental impaction surgery. This often results in prescription of an excess number of opioid pills, thereby increasing the risk of misuse, abuse, and addiction [1-3].

Acute dental pain includes pain from both surgical and nonsurgical dental conditions. Of over half a million dental patient visits between 2011 and 2015, 29% of prescribed opioids exceeded the recommended morphine equivalent for appropriate management of acute pain, and over half exceeded the recommended days of supply [4].

The current clinical practice guidelines lack evidence-based guidance on effective management of acute dental pain [5]. The National Academies' report, titled *Framing Opioid Prescribing Guidelines for Acute Pain*, highlighted the need to formalize evidence-based alternatives to opioid analgesics in a clinical practice guideline [6].

The objective of this systematic review (SR) and network meta-analysis (NMA) was to assess the comparative effectiveness of pharmacological treatments for the management of pain subsequent to simple and surgical tooth extraction, as well as pain associated with pulpitis or its complications. This SR was conducted to inform the 2022 evidence-based clinical practice guidelines produced by the American Dental

Association (ADA) Council on Scientific Affairs, the ADA Science & Research Institute (ADASRI), and the University of Pittsburgh's and the University of Pennsylvania's Schools of Dental Medicine in partnership with the US Food and Drug Administration (FDA) for the management of acute dental pain.

METHODS

We report this SR following the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist for SR and NMAs (see Appendix Table 1) [7]. We did not register this SR but followed preestablished methodology outlined in the plan for guideline development and used eligibility criteria determined by the recommendation questions addressed by the guideline panel.

Eligibility criteria

We included randomized clinical trials (RCTs) including individuals ages 12 and above undergoing simple or surgical tooth extraction or affected by symptomatic pulpitis or its complications that compared 10 interventions, including acetaminophen, nonsteroidal anti-inflammatory drugs (NSAIDs), opioids, and combinations against one another or placebo. We limited the literature to peer-reviewed articles and English language.

The interventions and dosages were selected by the guideline panel as the most commonly prescribed and relevant to acute dental pain management after a prioritization exercise that considered more than 30 different drugs as monotherapies or combinations. Similar to Cochrane Library overviews of multiple analgesics, a follow-up time point of 6 h was selected to eliminate distortions of findings (i.e., summed pain intensity difference [SPID], total pain relief [TOTPAR], proportion of rescue medications) when comparing short-duration (4–6 h) and long-duration (8–12 h) agents. The outcomes included pain relief at 6 h, TOTPAR at 6 h, SPID at 6 h, global efficacy rating at 6 h, proportion of participants receiving rescue analgesia at 6 h, and adverse effects (central nervous system and gastrointestinal) at the longest reported follow-up time point. See Appendix Table 2 for a list of included interventions and outcomes.

To establish the aforementioned eligibility criteria, the guideline panel relied on peer-reviewed literature regarding dental pharmacology, professional experience, a national survey of US oral surgeons of drug selection and prescribing behaviors, and selective randomized clinical trials of analgesic efficacy following third molar extractions. All decisions regarding the eligibility criteria for drug/dose selection and time points were made with the consensus of the panel.

Information sources

We performed searches in Medline, EMBASE, CENTRAL, and US Clinical Trials registry from inception through November 21, 2020. See Appendix Table 3 for a sample search strategy.

Study selection

We performed searches in Medline, EMBASE, CENTRAL, and US Clinical Trials registry from inception through November 21, 2020. See Appendix Table 3 for a sample search strategy.

Data collection

For each eligible trial, pairs of reviewers, following training and calibration exercises, extracted data independently using a standardized, pilot-tested data extraction form. Reviewers collected information on trial characteristics (i.e., design), patient characteristics (i.e., age, gender, country), and outcomes of interest. Reviewers resolved discrepancies by discussion and, when necessary, with adjudication by a third party.

Risk of bias within individual studies

For each eligible trial and outcome, pairs of reviewers, following training and calibration exercises, independently used a modification of the Cochrane tool to assess

risk of bias in randomized trials (RoB 2.0). Reviewers resolved discrepancies by discussion and, when necessary, with adjudication by a third party.

Data synthesis

Treatments were grouped into nodes that could include more than 1 dose for a monotherapy or a combination, according to the eligibility criteria listed above. The comparator/reference for all networks was selected to be the intervention/placebo reported in the highest number of studies.

We summarized the effect of interventions on dichotomous outcomes (i.e., proportion of participants receiving rescue analgesia and adverse effects) using odds ratios (ORs) and risk differences (RDs), as well as their corresponding 95% confidence interval (CI). For continuous outcomes (i.e., pain relief, TOTPAR, SPID, global efficacy rating), we used the mean difference (MD) (between pre- and postoperative scores) and corresponding 95% CI. When studies reported the same outcome using a scale with a different range, we converted data to the scale range most commonly reported before conducting analyses [8]. When standard deviation (SD) was not reported, we calculated SD using standard error, confidence intervals, means, and sample sizes. In rare instances, when neither of the beforementioned statistics were reported, we imputed SD by choosing a median SD of 3 studies with similar means [9].

We performed frequentist NMAs for outcomes with sufficient data. If data were insufficient for an NMA but adequate for a pairwise meta-analysis (i.e., at least 2 studies), we conducted pairwise meta-analyses comparing specific interventions. All analyses were completed using a random-effects model and weighting studies according to the inverse of their variance.

We implemented the analyses in RStudio version 3.5.3 (R Studio). We used the package *netmeta* [10] to conduct NMAs and the packages *meta* [11] and *metafor* [12] to conduct pairwise meta-analyses.

Certainty of the evidence

We assessed the certainty of the evidence using the Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) approach for NMAs [13-15]. The GRADE domains included risk of bias, inconsistency, indirectness, publication bias, imprecision, transitivity, and incoherence (i.e., agreement between direct and indirect evidence). Two methodologists rated each domain for each outcome and comparison independently, resolving discrepancies by discussion. We used a minimally contextualized approach to rate our certainty that there was an important effect [16], using a threshold of 10% of the length of the scale as the minimally important difference for continuous outcomes and baseline risk for dichotomous outcomes.

Presentation of results

To facilitate interpretation of results of dichotomous outcomes, we calculated absolute effects (95% CI) per 100 participants. To draw conclusions, we classified interventions in groups from the most to the least effective by considering the estimates of effect and the certainty of the evidence [17]

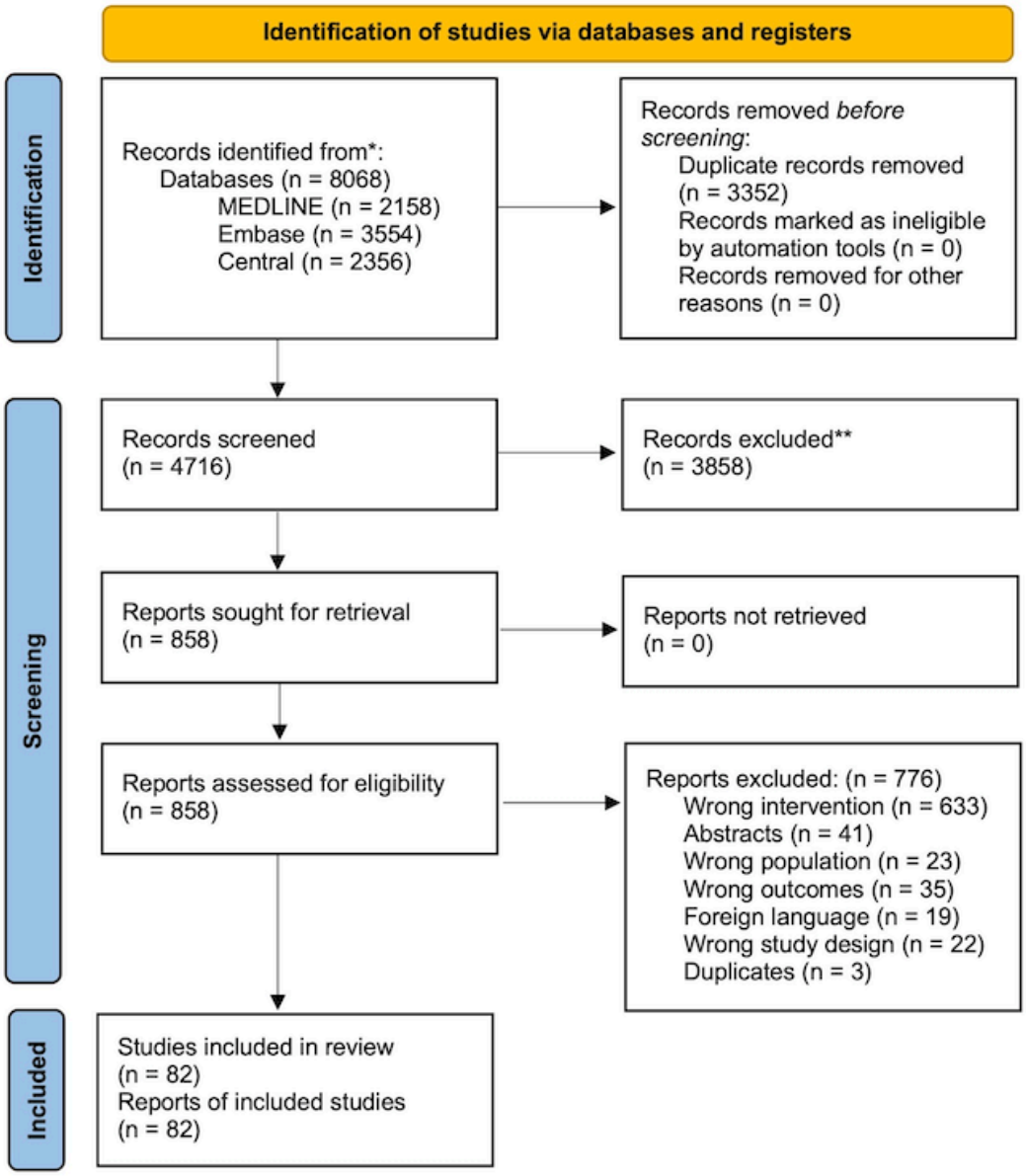
Subgroup and sensitivity analyses

Subgroup and sensitivity analyses were not planned.

RESULTS

After screening 4,716 titles and abstracts, we included 85 unique RCTs reported in 82 publications (Figure 1). Reasons for exclusion at the full-text screening stage (n = 776) are presented the Figure 1.

Figure 1. Study identification and selection flowchart.



Characterstics of included studies

Characteristics of the included studies are reported in Table 1. All studies were parallel group RCTs. Most studies were conducted in the United States (75%). Number of participants ranged from 31 to 540. Mean age of participants across studies varied from 13.5 (SD = 2.64) to 40.7 (SD not reported). Interventions were administered orally. Surgical tooth extraction was the type of extraction performed in all included studies. Studies assessing the interventions of interest in patients with symptomatic pulpitis or its complications were not found.

Table 1. The characteristics of included studies.

Study ID	Year	Study design	Country	Number of participants randomized	Age (years) Overall	Gender	Type of extraction	Type of tooth (report all information available in study)	Interventions
					Mean/Range (SD/SE)	Female (%)			
Cooper	1988	Parallel group	United States of America	80	Mean: 22.60 (SD=4.31)	65	Surgical tooth extraction	Not specified	Ibuprofen 400 mg (fast acting or acid), Placebo/ no treatment
Cooper	1981	Parallel group	United States of America	116	Mean: 23.19 (SD=NR)	59.46	Surgical tooth extraction	Impacted third molars	Acetaminophen 650 mg, Acetaminophen 650 mg/codeine 60 mg, Placebo/ no treatment
Seymour	1996	Parallel group	United Kingdom	123	Mean: 25.37 (SD=6.10)	68.29	Surgical tooth extraction	Impacted third molars	Acetaminophen 500 mg, Acetaminophen 1000 mg, Placebo/ no treatment
Gay	1996	Parallel group	Spain	80	Mean: 23.70 (SD=4.59)	53.66	Surgical tooth extraction	Impacted third molars of the lower jaw	Ibuprofen 400 mg, Placebo/ no treatment

Kiersch	1994	Parallel group	United States of America	134	Mean: 23.64 (SD=5.06)	58	Surgical tooth extraction	Partially or completely impacted third molars of the lower jaw	Acetaminophen 1000 mg, Naproxen sodium 440 mg, Placebo/ no treatment
Dolci	1994	Parallel group	Italy	148	Mean: 27.54 (SD=NR)	61	Surgical tooth extraction	Impacted third molars	Acetaminophen 500 mg, Placebo/ no treatment
Seymour	2003	Parallel group	United Kingdom	94	Mean: 25.03 (SD=5.05)	69.35	Surgical tooth extraction	Impacted third molars of the lower jaw	Solid Acetaminophen 1000 mg, Placebo/ no treatment
Mehlisch	2010	Parallel group	United States of America	440	Mean: 20.38 (SD=3.61)	62.42	Surgical tooth extraction	Partially or completely impacted third molars of the lower jaw	Ibuprofen 400 mg/Acetaminophen 1000 mg, Ibuprofen 200 mg/Acetaminophen 500 mg, Ibuprofen 400 mg, Acetaminophen 1000 mg, Acetaminophen 500 mg, Placebo/ no treatment
Cooper	2019	Parallel group	United States of America	385	Mean: 19.0 (SD=2.8)	53.7	Surgical tooth extraction	Partially or completely impacted third molars of the lower jaw	Ibuprofen 400mg, Naproxen sodium 440mg, Placebo/ no treatment
Mehlisch	2002	Parallel group	United States of America	300	Mean: 26.4 (SD=NR)	NR	Surgical tooth extraction	Impacted third molars	Conventional Ibuprofen 400mg, Ibuprofen Arginate 400mg, Placebo/ no treatment
Seymour	1991	Parallel group	England	95	Mean: 25.01 (SD=6.84)	59.38	Surgical tooth extraction	Impacted third molars	Soft Gelatin Ibuprofen 400mg, Ibuprofen Tablet 400mg, Placebo/ no treatment
Seymour	1991	Parallel group	England	92	Mean: 25.67 (SD=4.64)	78.13	Surgical tooth extraction	Impacted third molars	Soluble Ibuprofen 400mg, Ibuprofen Tablet 400mg, Placebo/ no treatment

Zelenakas	2004	Parallel group	United States of America	101	Mean: 22.4 (SD=4.51)	64.7	Surgical tooth extraction	Impacted third molars of the lower jaw	Ibuprofen 400mg, Placebo/ no treatment
Weiser1	2018	Parallel group	United States of America	279	Mean: 19.52 (SD=1.91)	65.7	Surgical tooth extraction	Impacted third molars of the lower jaw	Ibuprofen (acid) 400mg, Placebo/ no treatment
Al-Sukhun	2012	Parallel group	Finland	98	Mean: 30.29 (SD=7.40)	47.17	Surgical tooth extraction	Impacted third molars of the lower jaw	Ibuprofen 400mg, Placebo/ no treatment
Malmstrom	2002	Parallel group	United States of America	90	Mean: 22.50 (SD=6.72)	66.7	Surgical tooth extraction	Impacted third molars of the lower jaw	Ibuprofen 400mg, Placebo/ no treatment
Matthews	1984	Parallel group	United Kingdom	36	NR	NR	Surgical tooth extraction	Impacted third molars of the lower jaw	Acetaminophen 500mg, Placebo/ no treatment
Gazal	2017	Parallel group	Saudi Arabia	80	Mean: 40.7 (SD=NR)	NR	Surgical tooth extraction	Impacted third molars	Ibuprofen 400mg, Acetaminophen 1000mg
Kubitzek	2003	Parallel group	Germany	162	Mean: 26.0 (SD=NR)	NR	Surgical tooth extraction	Impacted third molars	Acetaminophen 1000mg, Placebo/ no treatment
Malmstrom	2004	Parallel group	United States of America	100	Mean: 23.30 (SD=5.0)	56	Surgical tooth extraction	Partially or completely impacted third molars of the lower jaw	Acetaminophen 600mg/Codeine 60mg, Placebo/ no treatment
Qi	2012	Parallel group	United States of America	540	Mean: 18.41 (SD=2.09)	54.8	Surgical tooth extraction	Partially or completely impacted third molars of the upper or lower jaw	Acetaminophen 1000mg, Acetaminophen 650mg, Placebo/ no treatment
Forbes	1986	Parallel group	United States of America	86	Mean: 21.61 (SD=NR)	48	Surgical tooth extraction	Impacted third molars	Codeine 60mg, Placebo/ no treatment

Fricke	2002	Parallel group	United States of America	100	Mean: 21.20 (SD=4.24)	50	Surgical tooth extraction	Impacted third molars	Tramadol 37.5 mg plus Acetaminophen 325 mg, Placebo/ no treatment
Cooper	1989	Parallel group	United States of America	184	Mean: 22.85 (SD=4.72)	66	Surgical tooth extraction	Impacted third molars	Ibuprofen 400 mg, Acetaminophen 1000 mg, Placebo/ no treatment
Cooper	1996	Parallel group	United States of America	97	Mean: 25.01 (SD=NR)	53	Surgical tooth extraction	Impacted third molar of the lower jaw	Ibuprofen 400 mg, Placebo/ no treatment
Melzack	1985	Parallel group	Canada	60	Mean: 34.0 (SD=NR)	NR	Surgical tooth extraction	Impacted third molars	Acetaminophen 500 mg, Placebo/ no treatment
Tong	2012	Parallel group	United States of America	104	Mean: 22.49 (SD=3.62)	59	Surgical tooth extraction	Partially or completely impacted third molars of the lower jaw	Ibuprofen 400mg, Placebo/ no treatment
Daniels	2011	Parallel group	United States of America	108	Mean: 21.5 (SD=NR)	59.7	Surgical tooth extraction	Partially or completely impacted third molars of the lower jaw	Acetaminophen 600 mg/Codeine 60 mg, Placebo/ no treatment
Chang	2004	Parallel group	United States of America	150	Mean: 22.0 (SD=NR)	60	Surgical tooth extraction	Partially or completely impacted third molars of the upper or lower jaw	Acetaminophen 650 mg plus Oxycodone 10 mg, Placebo/ no treatment
Christensen	2018	Parallel group	United States of America	80	Mean: 19.43 (SD=2.15)	72	Surgical tooth extraction	Partially or completely impacted third molars of the lower jaw	Ibuprofen 400 mg, Placebo/ no treatment
Desjardins	2007	Parallel group	United States of America	152	Mean: 21.98 (SD=4.06)	65.6	Surgical tooth extraction	Partially or completely impacted third molars of the lower jaw	Acetaminophen 650 mg plus Oxycodone 10 mg, Placebo/ no treatment

Seymour	1999	Parallel group	United Kingdom	80	Mean: 25.43 (SD=5.15)	68.29	Surgical tooth extraction	Impacted third molars	Ibuprofen 400 mg, Placebo/ no treatment
Morrison	1999	Parallel group	United States of America	101	Mean: 18.25 (SD=1.96)	52.9	Surgical tooth extraction	Partially or completely impacted third molars of the lower jaw	Ibuprofen 400 mg, Placebo/ no treatment
Ziccardi	2000	Parallel group	United States of America	76	Mean: 24.91 (SD=5.56)	67.3	Surgical tooth extraction	Partially or completely impacted third molars of the lower jaw	Acetaminophen 600 mg/Codeine 60 mg, Placebo/ no treatment
Moore	1998	Parallel group	United States of America	57	Range: 18-70	NR	Surgical tooth extraction	Impacted third molars	Codeine 60mg, Placebo/ no treatment
Cooper	1989	Parallel group	United States of America	184	Mean: 23.13 (SE=0.60)	66	Surgical tooth extraction		Ibuprofen 400, Acetaminophen 1000, Placebo/ no treatment
Yue_1	2013	Parallel group	United States of America	300	Mean: 23.08 (SD=3.79)	56	Surgical tooth extraction	Partially or completely impacted third molars of the lower jaw	Fast-Dissolving Acetaminophen 1000 mg, Fast-Dissolving Acetaminophen 500 mg, Placebo/ no treatment
Yue_2	2013	Parallel group	United States of America	401	Mean: 20.4 (SD=2.80)	63	Surgical tooth extraction	Partially or completely impacted third molars of the lower jaw	Fast-Dissolving Acetaminophen 1000 mg, Acetaminophen 650 mg, Placebo/ no treatment
Chang	2001	Parallel group	United States of America	211	Mean: 20.7 (SD=4.67)	70.6	Surgical tooth extraction	Partially or completely impacted third molars of the upper or lower jaw	Acetaminophen 600 mg/Codeine 60mg, Placebo/ no treatment
Bakshi	1994	Parallel group	Germany	162	Mean: 27.41 (SD=NR)	42.5	Surgical tooth extraction	Impacted third molars of the lower jaw	Ibuprofen 400 mg, Placebo/ no treatment

Chang	2004	Parallel group	United States of America	125	Mean: 22.0 (SD=4.3)	54	Surgical tooth extraction	Partially or completely impacted third molars of the lower jaw	Acetaminophen 650 mg plus Oxycodone 10 mg, Placebo/ no treatment
Daniels	2009	Parallel group	United Kingdom	321	Mean: 21.3 (SD=3.99)	61.3	Surgical tooth extraction	Partially or completely impacted third molars of the lower jaw	Ibuprofen/poloxamer 400 mg, Acetaminophen 1000 mg, Sodium ibuprofen 400 mg, Placebo/ no treatment
Hersh	1993	Parallel group	United States of America	100	NR	NR	Surgical tooth extraction	Impacted third molars	Ibuprofen 400 mg, Placebo/ no treatment
Mehlisch	2010	Parallel group	United States of America	234	Mean: 20.8 (SD=3.1)	73.13	Surgical tooth extraction	Partially or completely impacted third molars of the lower jaw	Ibuprofen 400mg/Acetaminophen 1000 mg, Ibuprofen 200mg/Acetaminophen 500 mg, Ibuprofen 400mg, Acetaminophen 1000 mg, Placebo/ no treatment
Schou	1998	Parallel group	Denmark	105	Mean: 26.23 (SD=6.04)	46.9	Surgical tooth extraction	Impacted third molars of the lower jaw	Ibuprofen 400 mg, Placebo/ no treatment
Malmstrom	2006	Parallel group	United States of America	40	Mean: 23.0 (SD=4.42)	85	Surgical tooth extraction	Partially or completely impacted third molars of the lower jaw	Acetaminophen 650 mg plus Oxycodone 10 mg, Placebo , Active Placebo
Mehlisch	1995	Parallel group	United States of America	240	Mean: 24.99 (SD=7.33)	64	Surgical tooth extraction	Partially or completely impacted third molars	Ibuprofen 400 mg, Acetaminophen 1000 mg, Placebo/ no treatment
Kyselovic	2020	Parallel group	Czech Republic	351	Mean: 28.0 (SD=7.83)	58.9	Surgical tooth extraction	Partially or completely impacted third molars	Ibuprofen lysinate 400 mg, Ibuprofen acid 400 mg, Placebo/ no treatment

Cheung	2007	Parallel group	United States of America	114	Mean: 21.6 (SD=4.6)	52.6	Surgical tooth extraction	Partially or completely impacted third molars of the lower jaw	Ibuprofen 400 mg, Placebo/ no treatment
Kellstein	2020	Parallel group	United States of America	394	Mean: 18.1 (SD=1.95)	51.1	Surgical tooth extraction	Partially or completely impacted third molars of the lower jaw	Ibuprofen 200mg/Acetaminophen 500 mg, Ibuprofen 250mg/Acetaminophen 500 mg, Ibuprofen 300mg/Acetaminophen 500 mg, Ibuprofen 400 mg, Placebo/ no treatment
Schwartz	2007	Parallel group	United States of America	31	Mean: 23.0 (SD=5.98)	20	Surgical tooth extraction	Impacted third molars	Ibuprofen 400 mg, Placebo/ no treatment
Ahlstrom	1993	Parallel group	Sweden	62	Mean: 25.33 (SD=NR)	53.13	Surgical tooth extraction	Impacted third molars of the lower jaw	Ibuprofen 400 mg, Placebo/ no treatment
Fricke	1993	Parallel group	United States of America	201	Mean: 23.28 (SD=4.91)	65	Surgical tooth extraction	Partially or completely impacted third molars of the lower jaw	Ibuprofen 400 mg, Naproxen Sodium 440mg, Placebo/ no treatment
Forbes	1991	Parallel group	United States of America	76	Mean: 23.37 (SD=NR)	65	Surgical tooth extraction	Impacted third molars	Ibuprofen 400 mg, Placebo/ no treatment
Cooper	1991	Parallel group	United States of America	120	Mean: 23.06 (SD=NR)	69.23	Surgical tooth extraction	Impacted third molars	Acetaminophen 600 mg/Codeine 60mg, Acetaminophen 650 mg, Placebo/ no treatment
VanDyke	2004	Parallel group	United States of America	311	Mean: 24.44 (SD=5.22)	54.3	Surgical tooth extraction	Partially or completely impacted third molars	Oxycodone 5 mg, Ibuprofen 400 mg, Placebo/ no treatment

Cooper	1988	Parallel group	United States of America	107	Mean: 25.1 (SD=NR)	80.56	Surgical tooth extraction	Impacted third molars	Acetaminophen 600 mg/Codeine 60mg, Acetaminophen 650 mg, Placebo/ no treatment
Giglio	1990	Parallel group	United States of America	79	Mean: 22.7 (SD=NR)	84.61	Surgical tooth extraction	Impacted third molars	Codeine 60mg, Placebo/ no treatment
Forbes	1990	Parallel group	United States of America	140	Mean: 22.88 (SD=NR)	59	Surgical tooth extraction	Impacted third molars	Ibuprofen 400 mg, Acetaminophen 600 mg, Acetaminophen 600 mg/Codeine 60 mg, Placebo/ no treatment
Chang	2005	Parallel group	United States of America	210	Mean: 18.87 (SD=3.89)	55	Surgical tooth extraction	Partially or completely impacted third molars of the lower jaw	Acetaminophen 600 mg/Codeine 60 mg, Placebo/ no treatment
Moller	2000	Parallel group	Denmark	242	Mean: 25.07 (SD=3.8)	55	Surgical tooth extraction	Impacted third molars of the lower jaw	Acetaminophen Effervescent 1000mg, Acetaminophen Tablet 1000mg, Effervescent Placebo, Tablet Placebo
Black	2002	Parallel group	United States of America	298	Mean: 21.43 (SD=NR)	55.56	Surgical tooth extraction	Impacted third molars of the upper and lower jaw	Ibuprofen Arginate 400mg, Ibuprofen 400mg, Placebo/ no treatment
McQuay	1996	Parallel group	United Kingdom	41	Mean: 13.50 (SD=2.64)	54.55	Surgical tooth extraction	Impacted third molars of the lower jaw	Ibuprofen 400mg, Placebo/ no treatment
Bentley	1987	Parallel group	Canada	79	Mean: 24.61 (SD=9.55)	56.09	Surgical tooth extraction	Impacted third molars	Acetaminophen 1000 mg, Codeine 60 mg, Placebo/ no treatment
Skoglund	1991	Parallel group	Norway	65	Mean: 24.55 (SD=6.79)	50	Surgical tooth extraction	Impacted third molars	Acetaminophen 1000mg, Placebo/ no treatment

Malmstrom	2005	Parallel group	United States of America	202	Mean: 23.08 (SD=5.71)	64	Surgical tooth extraction	Partially or completely impacted third molars of the lower jaw	Acetaminophen 650 mg plus Oxycodone 10 mg, Acetaminophen 600 mg/Codeine 60mg, Placebo/ no treatment
Searle_1	2020	Parallel group	United States of America	393	Mean: 19.38 (SD=2.12)	59.3	Surgical tooth extraction	Partially or completely impacted third molars of the lower jaw	Fixed-dose Ibuprofen 250 mg/Acetaminophen 500 mg, Acetaminophen 650 mg, Placebo/ no treatment
Searle_2	2020	Parallel group	United States of America	123	Mean: 21.8 (SD=3.82)	54.9	Surgical tooth extraction	Partially or completely impacted third molars of the lower jaw	Fixed-dose Ibuprofen 250 mg/Acetaminophen 500 mg, Placebo/ no treatment
Malmstrom	1999	Parallel group	United States of America	91	Mean: 23.0 (SD=4.2)	70	Surgical tooth extraction	Partially or completely impacted third molars of the lower jaw	Ibuprofen 400 mg, Placebo/ no treatment
Mehlisch	1984	Parallel group	United States of America	113	Mean: 29.77 (SD=NR)	55.17	Surgical tooth extraction	Partially or completely impacted third molars	Acetaminophen 1000 mg, Placebo/ no treatment
Cooper	1982	Parallel group	United States of America	125	Mean: 23.55 (SD=NR)	63.41	Surgical tooth extraction	Impacted third molars	Codeine 60 mg, Ibuprofen 400 mg, Placebo/ no treatment
Jain	1986	Parallel group	United States of America	96	Mean: 23 (SD=4.95)	53.06	Surgical tooth extraction	Impacted third molars	Ibuprofen 400 mg, Placebo/ no treatment
Dionne	1994	Parallel group	United States of America	76	Mean: 29.08 (SD=NR)	48.15	Surgical tooth extraction	Impacted third molars	Acetaminophen 650 mg, Acetaminophen 650 mg/Codeine 60 mg, Placebo/ no treatment

Hersh	1993	Parallel group	United States of America	44	Mean: 29.46 (SD=NR)	18.75	Surgical tooth extraction	Partially or completely impacted third molars	Codeine 60 mg, Ibuprofen 400 mg, Placebo
Olson	2001	Parallel group	Puerto Rico	172	Mean: 22.66 (SD=NR)	68.7	Surgical tooth extraction	Impacted third molars	Ibuprofen Liquigel 400 mg, Acetaminophen 1000 mg, Placebo/ no treatment
Seymour	1998	Parallel group	United Kingdom	146	Mean: 25.0 (SD=2.84)	47.37	Surgical tooth extraction	Impacted third molars	Ibuprofen 400 mg, Placebo/ no treatment
Sunshine	1986	Parallel group	Puerto Rico	91	Mean: 22.36 (SD=NR)	80	Surgical tooth extraction	Impacted third molars	Acetaminophen 650 mg, Acetaminophen 650 mm/Codeine 60 mg, Placebo/ no treatment
Hersh	2000	Parallel group	United States of America	149	Mean: 23.21 (SD=4.55)	69.5	Surgical tooth extraction	Partially or completely impacted third molars of the lower jaw	Ibuprofen Liquigel 400 mg, Acetaminophen 1000 mg, Placebo/ no treatment
Forbes	1990	Parallel group	United States of America	59	Mean: 22.00 (SD=NR)	74	Surgical tooth extraction	Impacted third molars	Acetaminophen 600 mg/Codeine 60 mg, Placebo/ no treatment
Akural	2009	Parallel group	Finland	38	Mean: 24.0 (SD=2.43)	61	Surgical tooth extraction	Impacted third molars	Acetaminophen 1000 mg, Placebo/ no treatment
Desjardins	1984	Parallel group	United States of America	80	Range: 18-68	NR	Surgical tooth extraction	Impacted third molars	Codeine 60 mg, Placebo/ no treatment
Moore	1987	Parallel group	United States of America	63	Mean: 24.12 (SD=NR)	48.48	Surgical tooth extraction	Partially or completely impacted third molars of the lower jaw	Codeine 60 mg, Placebo/ no treatment
Quiding	1984	Parallel group	Finland	92	Mean: 26.32 (SD=5.26)	NR	Surgical tooth extraction	Partially or completely impacted third molars	Codeine 60 mg, Acetaminophen 500 mg, Acetaminophen 1000 mg

Squires	1981	Parallel group	Canada	61	Mean: 28.14 (SD=10.12)	NR	Surgical tooth extraction	Impacted third molars	Ibuprofen 400 mg, Placebo/ no treatment
Daniels	2018	Parallel group	New Zealand & United States of America	296	Mean: 24.97 (SD=NR)	62.73	Surgical tooth extraction	Impacted third molars	Ibuprofen 292.5mg/Acetaminophen 975mg, Acetaminophen 975mg, Placebo/ no treatment

Risk of bias in included studies

Appendix Table 4 to 9 present the assessment of risk of bias of the included RCTs for each outcome. The domains in which most of the RCTs were judged at high risk of bias were missing outcome data and selection of reported results. Overall, 9 of 85 RCTs were judged at low or probably low risk of bias.

Effects of the interventions

Table 2 presents a summary of the effects of interventions on pain outcomes. Appendix Tables 10 and 11 summarize the effect of the interventions on adverse effects. Appendix Figures 1 to 21.6 and Appendix Tables 12 to 28 include network plots and forest plots of pairwise meta-analyses for all outcomes, as well as detailed relative and absolute effect estimates from the NMAs and the certainty of evidence for all comparisons and outcomes. All 85 RCTs were included in at least 1 of the 5 NMAs.

Table 2. Summary of benefit outcomes compared with placebo (no treatment).

	Pain relief	TOTPAR (Total Pain Relief)	SPID (Summed Pain Intensity Difference)	Global Efficacy Rating	Rescue Analgesia
Time point	6 hours	6 hours	6 hours	6 hours	6 hours
Scale	0 (none) – 4 (complete) [c]	(0–24)- higher better [d]	18 points- higher better [e]	0 (poor) – 4 (excellent) [c]	
Thresholds	-0.4, 0.4	-2.4, 2.4	-1.8, 1.8	-0.4, 0.4	-8, 8
Placebo[a]	0.62	4.1	0.345	0.69	80 per 100
Ibuprofen 200-400 mg plus Acetaminophen 500-1,000 mg	1.68 (1.06 to 2.31)	11.07 (8.23 to 13.91)	4.41 (5.78 to 3.04)	-	-55.60 (-70.27 to -31.22)
Oxycodone 5 mg or Codeine 60 mg	0.10 (-0.06 to 0.25)[b]	1.13 (0.17 to 2.09)[b]	0.78 (0.02 to 1.55)	0.23 (-0.14 to 0.61)	-3.64 (-20.49 to 7.57)
Acetaminophen 650 mg plus Oxycodone 10 mg	1.19 (0.85 to 1.54)	7.91 (6.49 to 9.32)	5.54 (5.26 to 6.02)	1.76 (1.35 to 2.18)	-45.18 (-62.93 to -22.10)
Ibuprofen 400 mg (fast acting or acid)	1.31 (1.17 to 1.45)	8.65 (7.82 to 9.48)	5.58 (4.85 to 6.31)	1.47 (1.27 to 1.68)	-43.01 (-49.50 to -36.02)
Tramadol 37.5 mg plus Acetaminophen 325 mg	0.01 (-0.34 to 0.36)[b]	-	-	-	-
Acetaminophen 500-1,000 mg	0.42 (0.23 to 0.62)	4.20 (3.30 to 5.09)	2.95 (2.31 to 3.60)	0.85 (0.65 to 1.06)	-24.00 (-32.02 to -16.30)
Acetaminophen 600-650 mg plus Codeine 60 mg	0.49 (0.27 to 0.71)	5.03 (4.04 to 6.03)	2.92 (2.32 to 3.53)[b]	0.98 (0.72 to 1.25)	-21.20 (-32.13 to -11.10)[b]
Naproxen 400-440 mg	1.44 (1.07 to 1.80)	8.47 (6.15 to 10.79)	5.27 (3.50 to 7.03)[b]	-	-51.49 (-64.71 to -33.31)
Ibuprofen 200 mg plus Hydrocodone 5 mg	-	-	-	-	-
Hydrocodone 5 mg plus Acetaminophen 300-325 mg	-	-	-	-	-

a The expected risk of each outcome with placebo is reported in the grey row. Numbers in the coloured cells are the estimated mean differences (95% CI) or risk differences (95% CI) per 100 patients when compared to placebo.
b The best estimate of effect was obtained from direct evidence.
Empty cells: there was no evidence for the specific intervention.

TOTPAR: total pain relief, SPID: sum of pain intensity differences
c We used this scale range as it was the most reported scale for this outcome among the included studies.
d The range of possible scores ranged from 0 to 24.
e The range of possible scores ranged from -6 to 12, a total length of 18 points.

Legend		
	BENEFIT OUTCOMES	
	High/Moderate certainty evidence	Low/Very low certainty evidence
AMONG THE BEST	Better than placebo and some alternatives	May be better than placebo and some alternatives
INTERMEDIATE	Better than placebo, but no better than any alternatives	May be better than placebo, but no better than any alternatives
AMONG THE WORST	No better than placebo	May be no better than placebo

Pain relief

Fifty-six studies including 9,095 participants were included in the NMA for pain relief. Using a scale from 0 to 4, where higher scores represent more pain relief, the interventions classified as among the most effective for this outcome were (effect estimates in reference to placebo) as follows: ibuprofen 200 to 400 mg plus acetaminophen 500 to 1,000 mg (MD, 1.68; 95% 95% CI, 1.06−2.31; moderate certainty), acetaminophen 650 mg plus oxycodone 10 mg (MD, 1.19; 95% CI, 0.85−1.54; moderate certainty), and ibuprofen 400 mg (fast acting or acid) (MD, 1.31; 95% CI, 1.17−1.45; moderate certainty). There was no convincing evidence that any of the other interventions were different from placebo for this outcome.

TOTPAR

Forty-four studies including 7,282 participants were included in the NMA for TOTPAR. Using a scale from 0 to 24, where higher scores represent more total pain relief, the interventions classified as among the most effective for this outcome were (effect estimates in reference to placebo) as follows: ibuprofen 200 to 400 mg plus acetaminophen 500 to 1,000 mg (MD, 11.07; 95% CI, 8.23−13.9; moderate certainty), acetaminophen 650 mg plus oxycodone 10 mg (MD, 7.91; 95% CI, 6.49−9.32; moderate certainty), ibuprofen 400 mg (fast acting or acid) (MD, 8.65; 95% CI, 7.82−9.48; moderate certainty), and naproxen 400 to 440 mg (MD, 8.47; 95% CI, 6.15−10.79; moderate certainty). Interventions that were more effective than placebo but less effective than the interventions above were acetaminophen 500 to 1,000 mg (MD, 4.20; 95% CI, 3.30−5.09; moderate certainty) and acetaminophen 600 to 650 mg plus codeine 60 mg (MD, 5.03; 95% CI, 4.04−6.03; moderate certainty). Oxycodone 5 mg and codeine 60 mg were not more effective than placebo.

SPID

Thirty-one studies including 6,721 participants reported on SPID. Using an 18-point scale, where higher scores represent better outcomes, the interventions classified as the most effective were (effect estimates in reference to placebo) as follows: acetaminophen 650 mg plus oxycodone 10 mg (MD, 5.54; 95% CI, 5.26−6.02;

moderate certainty) and ibuprofen 400 mg (fast acting or acid) (MD, 5.58; 95% CI, 4.85−6.31; moderate certainty). Acetaminophen 500 to 1,000 mg (MD, 2.95; 95% CI, 2.31−3.60; moderate certainty), acetaminophen 600 to 650 mg plus codeine 60 mg (MD, 2.92; 95% CI, 2.32−3.53; moderate certainty), and naproxen 400 to 440 mg (MD, 5.27; 95% CI, 3.50−7.03; moderate certainty) were better than placebo but less effective than the interventions above. Ibuprofen 200 to 400 mg plus acetaminophen 500 to 1,000 mg (MD, 4.41; 95% CI, 5.78−3.04; low certainty) were classified as possibly better than placebo. Oxycodone 5 mg and codeine 60 mg were not more effective than placebo.

Global efficacy rating, rescue analgesia, and adverse effects are reported in the Appendix.

DISCUSSION

This SR and NMA summarizes the comparative effects of the analgesic drugs considered by the guideline panel when making recommendations for treating dental acute pain. Based on moderate- and high-certainty evidence, in individuals undergoing surgical tooth extractions, the interventions classified as the most effective for pain relief were ibuprofen 200 to 400 mg plus acetaminophen 500 to 1,000 mg, acetaminophen 650 mg plus oxycodone 10 mg, ibuprofen 400 mg, and naproxen 400 to 440 mg. Oxycodone 5 mg or codeine 60 mg and tramadol 37.5 mg plus

acetaminophen 325 mg were no better than placebo. The results for TOTPAR, SPID, global efficacy rating, and rescue analgesia were similar to pain relief. Based on low- and very low-certainty evidence, most interventions were classified as no more harmful than placebo for most adverse effects.

Evidence for ibuprofen 200 mg plus hydrocodone 5 mg and hydrocodone 5 mg plus acetaminophen 300 to 325 mg (i.e., Vicodin) was not available. A comprehensive national survey conducted in the United States showed that hydrocodone and acetaminophen combinations (i.e., Vicodin, Vicodin ES, Vicodin HP) are the most preferred combination of analgesics prescribed by oral and maxillofacial surgeons for the management of pain following third molar extractions [18]. Future research should focus on exploring the comparative effect of hydrocodone and acetaminophen formulations. Furthermore, evidence regarding the effect of the 10 selected analgesics on the temporary management of symptomatic pulpitis or its complications prior to dental treatment was not found.

A recent overview of SRs summarizing benefits and harms of analgesic agents for the management of acute dental pain concluded that "relief of postoperative pain in dental practice with the use of nonsteroidal anti-inflammatory drugs, with or without acetaminophen, is equal or superior to that provided by opioid-containing medications" [19]. This finding is similar to the finding of this SR and NMA that, except for acetaminophen 650 mg plus oxycodone 10 mg, NSAIDs with or without

acetaminophen were superior to placebo as opposed to opioid-containing interventions that were not. Furthermore, similar to this NMA, the overview of SRs suggested that opioid analgesics and their combinations are associated with higher rates of acute adverse events [19].

Another overview of SRs aiming to summarize the efficacy of analgesics for acute dental pain in adults investigated 41 single-dose analgesics or analgesic combinations for acute postoperative pain and showed that the best interventions in terms of number needed to treat (NNT) for at least 50% maximum pain relief over 4 to 6 h compared with placebo were ibuprofen 200 mg plus acetaminophen 500 mg, ibuprofen 200 mg, ibuprofen 200 mg plus caffeine 100 mg, diclofenac potassium 50 mg, and etoricoxib 120 mg [20]. The worst intervention was codeine 60 mg [20]. These results were comparable to the results of this SR and NMA, specifically the superiority of ibuprofen plus acetaminophen and ibuprofen alone and inferiority of codeine 60 mg compared to placebo.

An overview of SRs summarizing adverse event rates associated with 41 single-dose oral analgesics or analgesic combinations compared with placebo for acute postoperative pain in adults showed that there were few instances of participants experiencing significantly more or fewer adverse events than with placebo for most NSAIDs, acetaminophen, and combinations not containing opioids [21]. However, for aspirin 1,000 mg, diflunisal 1,000 mg, opioids, or fixed-dose combination drugs

containing opioids, participants often experienced significantly more adverse effects than with placebo. In this SR and NMA, most included opioid-containing interventions (oxycodone 5 mg and codeine 60 mg, acetaminophen 650 mg plus oxycodone 10 mg, and acetaminophen 600–650 mg plus codeine 60 mg) were worse than placebo in terms of at least 1 adverse effect. Ibuprofen 400 and acetaminophen 500 to 1,000 mg were worse than placebo in terms of drowsiness.

Several differences between this SR and NMA and review of reviews by P.A. Moore et al. [19] and R.A. Moore et al. [20-22] may have contributed to the differences in conclusions between these reviews. In this SR and NMA, data from individual RCTs were included in a single NMA that comprised the results, whereas in the overviews of SRs, results from multiple analyses were aggregated to constitute the results. In terms of pain relief, P.A. Moore et al. and R.A. Moore et al. [20-22] reported NNT for at least 50% pain relief, whereas this SR and NMA reported mean pain relief measured on a scale from 0 (none) to 4 (complete) at 6 h. The variability in inclusion of interventions may have contributed to the differences in the conclusions between these publications.

Furthermore, in this SR and NMA, a modification of the Cochrane tool for assessing risk of bias (RoB 2.0) was used to examine risk of bias in individual RCTs, whereas the overviews of SRs by P.A. Moore et al. [19] and R.A. Moore et al. [20-22] did not mention risk of bias assessment. The quality of the evidence using the GRADE

approach was assessed in this SR, but it was not addressed in the overviews by P.A. Moore et al. [19]and R.A. Moore et al. [20-22].

Strengths of the review process

This review is the first to synthesize NMAs for comparison of analgesics for treatment of acute dental pain associated with postsurgical extractions, allowing for streamlined clinical decision-making. Each of the review process stages was conducted in duplicate, with adjudication of conflicts by a third reviewer. The risk of bias of each individual study, as well as the certainty of the evidence for each outcome of interest, was assessed. We performed analyses and interpreted the results using the latest methodological guidance from the GRADE Working Group. In order to make the results easier to interpret, instead of using standardized mean difference, we reported continuous outcomes using mean difference by converting all scale scores to the most reported scale [8].

Limitations of the review process

This SR and NMA was limited to inclusion of only 10 interventions, which may limit its applicability. These 10 interventions, however, were chosen by a panel of clinical experts who determined that they were the most relevant to practice. Furthermore, this SR and NMA included only peer-reviewed studies published in English, but this likely did not affect the study conclusions.

CONCLUSION

Based on moderate- and high-certainty evidence, in individuals undergoing surgical tooth extractions, the interventions classified as the most effective for pain relief were ibuprofen 200 to 400 mg plus acetaminophen 500 to 1,000 mg, acetaminophen 650 mg plus oxycodone 10 mg, ibuprofen 400 mg, and naproxen 400–440 mg. Oxycodone 5 mg or codeine 60 mg and tramadol 37.5 mg plus acetaminophen 325 mg were no better than placebo. The results for TOTPAR, SPID, global efficacy rating, and rescue analgesia were similar to pain relief. Based on low- and very low-certainty evidence, most interventions were classified as no more harmful than placebo for most selected adverse effects. Future research should focus on the assessment of ibuprofen 200 mg plus hydrocodone 5 mg, hydrocodone 5 mg plus acetaminophen 300 to 325 mg, and tramadol 37.5 mg plus acetaminophen 325 mg through RCTs.

Supplemental material

A supplemental appendix to this article is available online.

Conflict of interest statement

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

Funding statement

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: E.V. Hersh and the Trustees of the University of Pennsylvania have received grant funding from Pfizer Consumer Healthcare and Bayer Healthcare. In addition, within the past 5 y, E.V. Hersh received consulting funds for reviewing data and sharing his expertise with the National Advertising Division of the Better Business Bureau concerning pain studies that compared Aleve and Extra Strength Tylenol.

This project was financially supported by the FDA of the US Department of Health and Human Services (HHS). The contents are those of the author(s) and do not necessarily represent the official views of, nor an endorsement by, FDA/HHS or the US government.

The findings and conclusions in this presentation are those of the author(s) and do not necessarily reflect the official views, position, or policy of the American Dental Association, nor does this presentation constitute an endorsement by the American Dental Association.

Ethics approval statement

The ethics approval was not required.

REFERENCES

1. Denisco, R.C., et al., *Prevention of prescription opioid abuse: the role of the dentist.* J Am Dent Assoc, 2011. **142**(7): p. 800-10.
2. Maughan, B.C., et al., *Unused opioid analgesics and drug disposal following outpatient dental surgery: A randomized controlled trial.* Drug Alcohol Depend, 2016. **168**: p. 328-334.
3. Moore, P.A., et al., *Why do we prescribe Vicodin?* J Am Dent Assoc, 2016. **147**(7): p. 530-3.
4. Suda, K.J., et al., *Overprescribing of Opioids to Adults by Dentists in the U.S., 2011-2015.* Am J Prev Med, 2020. **58**(4): p. 473-486.
5. Herzig, S.J., et al., *Safe Opioid Prescribing for Acute Noncancer Pain in Hospitalized Adults: A Systematic Review of Existing Guidelines.* J Hosp Med, 2018. **13**(4): p. 256-262.
6. National Academies of Sciences, E., et al., in *Framing Opioid Prescribing Guidelines for Acute Pain: Developing the Evidence*. 2019, National Academies Press (US)

Copyright 2020 by the National Academy of Sciences. All rights reserved.: Washington (DC).

7. Hutton, B., et al., *The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations.* Ann Intern Med, 2015. **162**(11): p. 777-84.
8. Thorlund, K., et al., *Pooling health-related quality of life outcomes in meta-analysis-a tutorial and review of methods for enhancing interpretability.* Res Synth Methods, 2011. **2**(3): p. 188-203.
9. Higgins, J.G., S., ed. *Cochrane Handbook for Systematic Reviews of Interventions*. Vol. Version 5.1.0 [updated March 2011]. 2011, The Cochrane Collaboration.
10. Rucker, G., Krahn, U., Konig, J., Efthimiou, O., Davies, A., Papakonstantinou, T., Schwarzer, G., *Network Meta-Analysis using Frequentist Methods*. 2021.
11. Schwarzer, G., *meta: General Package for Meta-Analysis.* 2015(R package meta version 5.1-0).
12. Viechtbauer, W., *metafor: Meta-Analysis Package for R.* 2021(R package metafor version 3.0-2).
13. Brignardello-Petersen, R., et al., *Advances in the GRADE approach to rate the certainty in estimates from a network meta-analysis.* J Clin Epidemiol, 2018. **93**: p. 36-44.
14. Brignardello-Petersen, R., et al., *GRADE approach to rate the certainty from a network meta-analysis: addressing incoherence.* J Clin Epidemiol, 2019. **108**: p. 77-85.

15. Puhan, M.A., et al., *A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis.* Bmj, 2014. **349**: p. g5630.
16. Zeng, L., et al., *GRADE guidelines 32: GRADE offers guidance on choosing targets of GRADE certainty of evidence ratings.* J Clin Epidemiol, 2021. **137**: p. 163-175.
17. Brignardello-Petersen, R., et al., *GRADE approach to drawing conclusions from a network meta-analysis using a partially contextualised framework.* Bmj, 2020. **371**: p. m3907.
18. Levy, B., et al., *Trends in Opioid Analgesic-Prescribing Rates by Specialty, U.S., 2007-2012.* Am J Prev Med, 2015. **49**(3): p. 409-13.
19. Moore, P.A., et al., *Benefits and harms associated with analgesic medications used in the management of acute dental pain: An overview of systematic reviews.* J Am Dent Assoc, 2018. **149**(4): p. 256-265.e3.
20. Moore, R.A., et al., *Single dose oral analgesics for acute postoperative pain in adults - an overview of Cochrane reviews.* Cochrane Database Syst Rev, 2015. **2015**(9): p. Cd008659.
21. Moore, R.A., et al., *Adverse events associated with single dose oral analgesics for acute postoperative pain in adults - an overview of Cochrane reviews.* Cochrane Database Syst Rev, 2015. **2015**(10): p. Cd011407.
22. Moore, R.A., et al., *Non-prescription (OTC) oral analgesics for acute pain - an overview of Cochrane reviews.* Cochrane Database Syst Rev, 2015. **2015**(11): p. Cd010794.

CHAPTER 3: CORTICOSTEROIDS FOR MANAGING ACUTE PAIN SUBSEQUENT TO SURGICAL EXTRACTION OF MANDIBULAR THIRD MOLARS: A SYSTEMATIC REVIEW AND META-ANALYSIS

Authors: Anna Miroshnychenko, Maria Azab, Sara Ibrahim, Yetiani Roldan, Juan Pablo Diaz Martinez, Divyalakshmi Tamilselvan, Leon He, Olivia Urquhart, Francisca Verdugo-Paiva, Malavika Tampi, Deborah E. Polk, Paul A. Moore, Elliot V. Hersh, Romina Brignardello-Petersen, and Alonso Carrasco-Labra

Published: Journal of American Dental Association [August 2023]

ABSTRACT

Objective: Corticosteroids are used to manage pain after surgical tooth extractions. The authors assessed the effect of corticosteroids on acute postoperative pain in patients undergoing surgical tooth extractions of mandibular third molars.

Methods: The authors conducted a systematic review and meta-analysis. The authors searched the Epistemonikos database, including MEDLINE, Embase, Cochrane Central Register of Controlled Trials, and the US clinical trials registry (ClinicalTrials.gov) from inception until April 2023. Pairs of reviewers independently screened titles and abstracts, then full texts of trials were identified as potentially eligible. After duplicate data abstraction, the authors conducted random-effects meta-analyses. Risk of bias was assessed using Version 2 of the Cochrane Risk of Bias tool and certainty of the evidence was determined using the Grading of Recommendations Assessment, Development and Evaluation approach.

Results: Forty randomized controlled trials proved eligible. The evidence suggested that corticosteroids compared with a placebo provided a trivial reduction in pain intensity measured 6 hours (mean difference, 8.79 points lower; 95% CI, 14.8 to 2.77 points lower; low certainty) and 24 hours after surgical tooth extraction (mean difference, 8.89 points lower; 95% CI, 10.71 to 7.06 points lower; very low certainty). The authors found no important difference between corticosteroids and a placebo with regard to incidence of postoperative infection (risk difference, 0%; 95% CI, −1% to

1%; low certainty) and alveolar osteitis (risk difference, 0%; 95% CI, −3% to 4%; very low certainty).

Conclusion: *Low and very low certainty evidence suggests that there is a trivial difference in postoperative pain intensity and postoperative infection when corticosteroids administered orally, submucosally, or intramuscularly are compared with placebo in patients undergoing third-molar extractions.*

INTRODUCTION

Surgical removal of impacted mandibular third molars is 1 of the most frequently performed surgical interventions in dental surgery, with more than 10 million teeth extracted per year [1]. The most common complications, including pain, swelling, and trismus, can severely affect a patient's quality of life during the immediate postoperative period. Analgesics and anti-inflammatory drugs prescribed postoperatively should relieve pain, reduce swelling and trismus, and improve healing without undesirable adverse effects. Therefore, medications that exert both analgesic and anti-inflammatory effects, such as corticosteroids, could be used for the management of postoperative discomfort.

Corticosteroids can be divided into 2 major groups, that is, glucocorticoids and mineralocorticoids. Glucocorticoids are used for the management of postoperative complications after surgical tooth extraction because of their substantial anti-inflammatory effects [2]. The term corticosteroids will be used to represent glucocorticoids in our review.

Corticosteroids are classified according to their duration of action and relative anti-inflammatory potency compared with hydrocortisone, a reference standard with a potency of 1. The higher the relative potency anti-inflammatory score, the higher the corticosteroid's anti-inflammatory potency. Short-acting glucocorticoids include cortisol and cortisone, with a duration of action of fewer than 12 hours and anti-

inflammatory potency of 1 [3]. Intermediate-acting corticosteroids include prednisone and prednisolone, with an anti-inflammatory potency of 4, and 6-methylprednisolone and triamcinolone, which both have an anti-inflammatory potency of 5. Intermediate corticosteroids have a duration of action of 12 through 36 hours. The long-acting glucocorticoids include dexamethasone, with a duration of action of more than 36 hours and anti-inflammatory potency of 25 [3]. The administration of corticosteroids in dentistry typically varies among oral, intramuscular, and submucosal routes.

Available systematic reviews (SRs) to inform the effect of corticosteroids for managing postoperative acute pain consecutive to surgical tooth extractions have several limitations. Almost all were published before 2018 [4-7], with only 1 published in 2020 [8]. They did not use the Grading of Recommendations Assessment, Development and Evaluation (GRADE) [9] approach to assess the certainty of the evidence—an essential component of SRs and a key step to inform the formulation of guideline recommendations—and they applied suboptimal methods to synthesize pain-related outcomes.

The purpose of our SR was to determine the effect of corticosteroids administered orally, submucosally, or intramuscularly on the management of pain subsequent to surgical tooth extraction, including impacted mandibular third-molar extractions. Our review informed the clinical questions posed by the forthcoming evidence-based clinical practice guideline for the pharmacologic management of acute dental pain consecutive to tooth extractions (A. Carrasco-Labra D.E. Polk, O. Urquhart, and

colleagues, unpublished data, 2023). This review and associated clinical practice guidelines were led by the American Dental Association Science and Research Institute, the School of Dental Medicine at the University of Pittsburgh, the Center for Integrative Global Oral Health at the University of Pennsylvania, and a guideline panel including primary care dentists, oral and maxillofacial surgeons, public health practitioners, pharmacoepidemiologists, biostatisticians, and health research methodologists, among others.

METHODS

This article follows the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement [10] (Appendix Table 1). We also followed established methodological considerations defined for evidence synthesis and guideline development, used eligibility criteria determined per the recommendation questions that the guideline panel proposed, and used information outlined in the National Academy of Medicine's *Framing Opioid Prescribing Guidelines for Acute Pain: Developing the Evidence [11]*.

Eligibility criteria

We included randomized controlled trials (RCTs) that compared the effect of corticosteroids administered orally, submucosally, or intramuscularly at any dose with that of a placebo in adolescent, adult, or older adult participants undergoing surgical (that is, extraction of a tooth with the need of a flap and osteotomy) third-molar

extraction, regardless of the language of publication. We included the following continuous outcomes: pain intensity at 6 hours, pain at 24 hours, total pain relief at 6 hours, and global efficacy rating at 6 hours. We included the following dichotomous outcome: adverse effects (for example, postoperative surgical site infection, alveolar osteitis, mood alteration, and gastrointestinal [GI] adverse effect at any time). We excluded studies that administered corticosteroids intravenously and studies that only reported outcomes associated with the management of inflammatory complications (that is, trismus, facial swelling, or infection).

Search methods to identify and select studies

We performed the evidence search in 2 steps. First, we conducted searches in the Epistemonikos database, a comprehensive mate-search engine and updated source of relevant SRs and primary studies to inform health decision making. Using artificial intelligence technology, Epistemonikos periodically screens across the following databases: Cochrane Database of Systematic Reviews, PubMed, MEDLINE, Embase, Cumulative Index to Nursing and Allied Health Literature, PsycINFO, Latin American and Caribbean Health Sciences Literature, Database of Abstracts of Reviews of Effects, Health Technology Assessment Database, Campbell Collection online library, Joanna Briggs Institute Database of Systematic Reviews and Implementation Reports, Evidence for Policy and Practice Centre Evidence Library, and the US clinical trials registry (ClinicalTrials.gov) [12]. Our search strategy included a combination of free

and controlled key words. These terms include specific key words to represent the concepts of surgical extraction, third molars, and corticosteroids (Appendix Table 2).

In addition, we used the Living Overview of Evidence (L·OVE) platform (Epistemonikos Foundation), which maps the question of interest to a repository maintained through additional searches on PubMed, Embase, and Cochrane Central Register of Controlled Trials. We searched L·OVE for third-molar–related literature without restriction according to study design, language, or publication status. The search covered the period from the inception date of each database through April 2023 and had no language restrictions. The results of the searches in each source were deduplicated by means of an algorithm that compares unique identifiers (that is, database identification, digital object identifier, and trial registry identification) and citation details (that is, author names, journal, year of publication, volume, number, pages, article title, and article abstract).

Pairs of reviewers (A.M., M.A., Y.R., J.P.D.M., D.T., L.H., F.V.-P.) independently evaluated the titles, abstracts, and full text of potentially eligible studies across all databases. When eligibility consensus was elusive, a third reviewer served as arbiter (R.B.-P., A.C.-L.).

Data collection

After training and calibration exercises, pairs of reviewers independently extracted data for each eligible trial using a standardized, pilot-tested data extraction form. We

collected information on trial characteristics (for example, intervention, comparison, and co-interventions), patient characteristics (for example, age, sex, country, and type of extraction), and outcomes of interest. Reviewers resolved discrepancies by means of discussion, and, when necessary, a third reviewer served as arbiter.

Risk of bias in studies

After training and calibration exercises, pairs of reviewers used Version 2 of the Cochrane Risk of Bias tool for each eligible trial and outcome to assess the risk of bias in RCTs, rating trials as being at low risk of bias, probably at low risk of bias, probably at high risk of bias, or at high risk of bias across the domains of bias arising from the randomization process, bias due to deviations from the intended intervention, bias due to missing data, bias due to measurement of the outcome, and bias in selection of the reported results [13]. We rated trials as high risk of bias overall if 1 or more domains were rated as probably high risk of bias or at high risk of bias, and we rated trials as low risk of bias overall if all domains were rated as probably at low risk of bias or at low risk of bias. Reviewers resolved discrepancies by means of discussion and, when necessary, a third reviewer served as arbiter.

Data synthesis

For dichotomous outcomes, we summarized the effect of interventions using odds ratios. When the incidence of the outcome was low across studies (that is, there were no events in several study groups), we used the risk difference. For continuous

outcomes, we used the mean difference. When studies reported the same outcome using a scale with a different range, we converted the data to the most reported scale before conducting analyses. In addition, we calculated 95% CIs around all of these estimates and created forest plots (in which the black diamond represents the pooled estimate across studies). In instances when the SD was not reported, we calculated SDs using SE, CIs, means, and sample sizes. In rare cases when none of these statistics were reported, we imputed the SD by means of averaging the SDs of 3 studies with similar means. We conducted random-effects meta-analyses using Review Manager (Cochrane) software.

Certainty of the evidence

We assessed the certainty of the evidence using the GRADE approach [9]. Two methodologists (F.V.-P., A.C.-L.) formally trained in using GRADE rated each domain for each comparison and outcome independently, resolving discrepancies by means of discussion. We rated the certainty as high, moderate, low, or very low, taking into consideration risk of bias, inconsistency, indirectness, publication bias, and imprecision. We used a minimally contextualized approach with a null effect threshold to rate the certainty that there is a benefit or harm [14]. When the point estimate was close to the null effect, we rated our certainty as having a trivial effect (that is, no important difference) using a threshold of 10% of the baseline risk of dichotomous outcomes and 10% of the scale range for continuous outcomes [15]. For dichotomous outcomes, we calculated absolute estimates of effect using the mean baseline risk

across trials. We created GRADE summary of findings tables using GRADEpro software (McMaster University and Evidence Prime).

Subgroup analyses

We performed 2 subgroup analyses to determine the extent to which treatment effects vary according to the type of corticosteroid (for example, dexamethasone, methylprednisolone, prednisolone, and triamcinolone acetonide) and routes of administration (oral, intramuscular, submucosal) compared with a placebo.

RESULTS

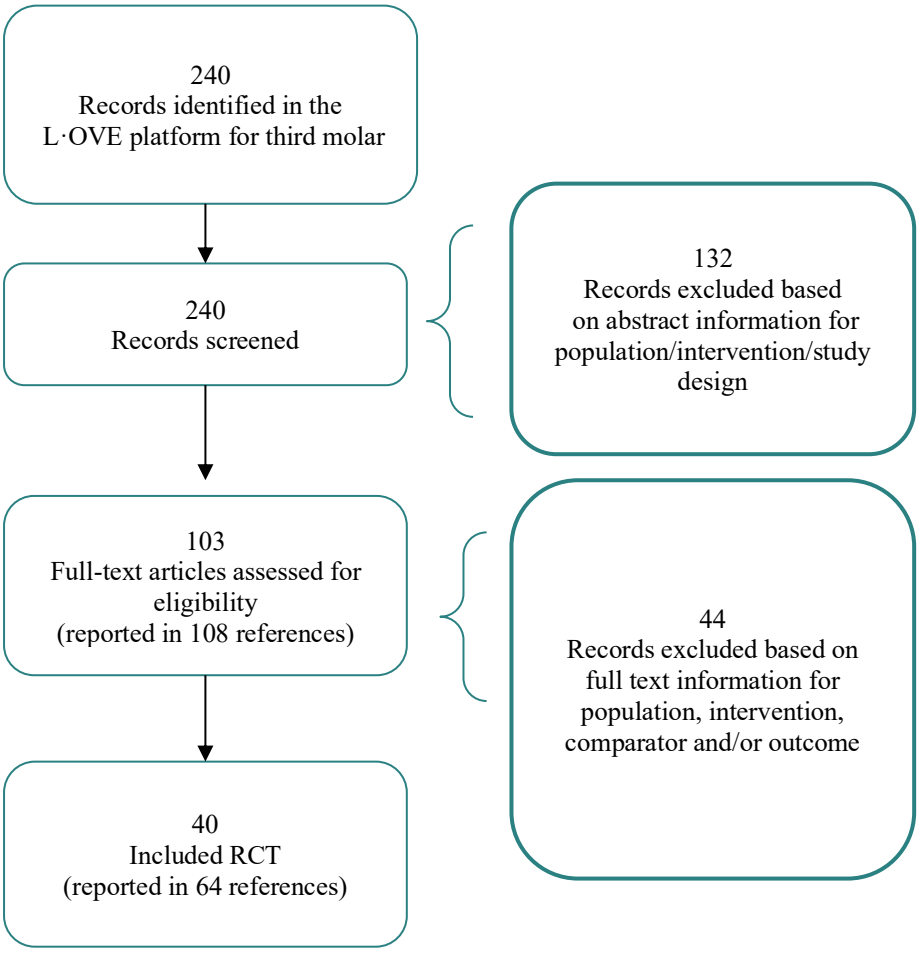
The search of the Epistemonikos database initially identified a total of 44 search results. Titles and abstracts of all SRs were screened for our inclusion criteria, and 19 SRs proved relevant. These SRs included a total of 79 RCTs (reported in 81 references) comparing the use of corticosteroids with a placebo for patients undergoing surgical third-molar extractions. All identified RCTs were entered into our database. The following link provides access to the interactive version of the matrix of evidence that we built, as described above (corticosteroids vs a placebo for patients undergoing third-molar

extractions: http://www.epistemonikos.org/matrixes/60cfb16e7aaac86eee79456c).

After removal of duplicates, the search in the L·OVE platform for “corticosteroids” and “third molar” yielded 240 records to screen at the title and abstract stage. Then,

103 of those records (108 references) were potentially eligible, and their full texts were evaluated. Finally, 40 RCTs (64 references) were included [16-55]. The complete study selection process, including the reasons for excluding studies at the time of the full-text review, is summarized in a Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart (Figure 1).

Figure 1. Study identification and selection flow chart.



Most RCTs were parallel group in design (88%), with the number of participants ranging from 30 through 450. Mean (SD) age of participants across studies ranged from 18 (not reported) to 31.42 (11.76) years. All populations across included studies underwent surgical extraction of the impacted third molars (Table 1).

Table 1. Characteristics of the included studies examining corticosteroids.

Study ID	Study design	Country	Number of participants randomized	Age (years)	Gender	Type of extraction	Interventions
				Mean/Range (SD/SE)	Female (%)		
Selimovic_2017	Parallel group	Bosnia and Herzegovina	40	Range: 18-45	NR	Surgical tooth extraction (impacted third molar)	Methylprednisolone 32 mg + Meloxicam 15 mg (oral), Meloxicam 15 mg
Gopinath_2017	Parallel group	India	80	NR	NR	Surgical tooth extraction (impacted third molar)	Dexamethasone 4mg (submucosal), No treatment (placebo)
Ghensi_2017A	Parallel group	Italy	80	Mean: 27 (SD=7.1)	46.25	Surgical tooth extraction (impacted third molar)	Dexamethasone 4 mg (submucosal), No treatment (placebo)
Ghensi_2017B							Dexamethasone 4 mg + Bromelain 40 mg (submucosal), Bromelain 40 mg
Gozali_2017	Split mouth	Thailand	96	Range: 18-30	60.42	Surgical tooth extraction (impacted third molar)	Dexamethasone 8 mg (submucosal), No treatment (placebo)
Prashar_2016	Parallel group	India	30	Mean: 26.5 (NR)	NR	Surgical tooth extraction (impacted third molar)	Methylprednisolone 8 mg + Diclofenac 50 mg (oral), Diclofenac 50 mg
Zerener_2015A	Parallel group	Turkey	78	Mean: 22.6 (SD=6.3)	48.72	Surgical tooth extraction	Dexamethasone 4 mg (submucosal), No treatment (placebo)

Zerener_2015B						(impacted third molar)	Triamcinolone acetonide 4 mg (submucosal), No treatment (placebo)
Deo_2016_1	Parallel group	Nepal	40	Mean: 24.93 (SD=NR)	36.67	Surgical tooth extraction (impacted third molar)	Dexamethasone 2 mL of 4 mg/mL (submucosal), No treatment (placebo)
Deo_2016_2	Parallel group	Nepal	40	Range: 20-41	NR	Surgical tooth extraction (impacted third molar)	Dexamethasone 2 mL of 4 mg/mL (submucosal), No treatment (placebo)
Ibikunle_2016A	Parallel group	Nigeria	191	Mean:28.1 (SD=7.4)	62.9	Surgical tooth extraction (impacted third molar)	Prednisolone 40 mg (oral), No treatment (placebo)
Ibikunle_2016B							Prednisolone 40 mg (submucosal), No treatment (placebo)
Larsen_ 2021A	Split mouth	Denmark	104	Mean: 25.92 (SD=5.99)	69.23	Surgical tooth extraction (impacted third molar)	Methylprednisolone 20 mg (intramuscular), No treatment (placebo)
Larsen_ 2021B							Methylprednisolone 30 mg (Intramuscular), No treatment (placebo)
Larsen_ 2021C							Methylprednisolone 40 mg (intramuscular), No treatment (placebo)
Gholami_ 2021A	Parallel group	Iran	60	Mean: 26.83 (SD=4.19)	51.67	Surgical tooth extraction (impacted third molar)	Methylprednisolone 40 mg (intramuscular - masseter), No treatment (placebo)
Gholami_ 2021B							Methylprednisolone 40 mg (intramuscular - gluteal), No treatment (placebo)

Imon_2021	Parallel group	Bangladesh	294	Mean: 25 (SD=NR)	44.9	Surgical tooth extraction (impacted third molar)	Dexamethasone tapering dose (oral), No treatment (placebo)
Larsen_2020A	Split mouth	Denmark	104	Mean: 25.92 (SD=5.99)	69.23	Surgical tooth extraction (impacted third molar)	Methylprednisolone 20 mg (intramuscular), No treatment (placebo)
Larsen_2020B							Methylprednisolone 30 mg (intramuscular), No treatment (placebo)
Larsen_2020C							Methylprednisolone 40 mg (intramuscular), No treatment (placebo)
Pansard_2020	Parallel group	Brazil	114	Mean: 31.43 (SD=11.76)	65.59	Surgical tooth extraction (impacted third molar)	Dexamethasone 8 mg (oral), No treatment (placebo)
Chugh_2018A	Parallel group	India	60	Mean: 29.79 (SD=8.37)	36.67	Surgical tooth extraction (impacted third molar)	Dexamethasone 8 mg (submucosal), No treatment (placebo)
Chugh_2018B							Methylprednisolone 40 mg (submucosal), No treatment (placebo)
Mojsa_2017A	Parallel group	Poland	60	Range: 18-42	64.44	Surgical tooth extraction (impacted third molar)	Dexamethasone 4 mg/mL preoperatively (submucosal), No treatment (placebo)
Mojsa_2017B							Dexamethasone 4 mg/mL postoperatively (submucosal), No treatment (placebo)
Bhargava_2014A	Parallel group	India	40	NR	NR	Surgical tooth extraction (impacted third molar)	Dexamethasone 4 mg/mL (submucosal), No treatment (placebo)

Bhargava_2014B							Dexamethasone 4 mg/mL (intramuscular), No treatment (placebo)
Bhargava_2014C							Dexamethasone 4 mg (oral), No treatment (placebo)
Afkan_2018A	Parallel group	Iran	75	Mean: 28 (SD=NR)	NR	Surgical tooth extraction (impacted third molar)	Dexamethasone 2 mg preoperatively and 2 mg postoperatively (oral), No treatment (placebo)
Afkan_2018B							Dexamethasone 2 mg (oral), No treatment (placebo)
Atalay_2020A	Parallel group	Turkey	60	Mean: 25.18 (NR=5.26)	46.67	Surgical tooth extraction (impacted third molar)	Dexamethasone 1mL of 4 mg/mL solution (submucosal), No treatment (placebo)
Atalay_2020B							Dexamethasone 2mL of 8 mg/mL solution (submucosal), No treatment (placebo)
Sahu_2020	Parallel group	India	40	Range: 18-60	NR	Surgical tooth extraction (impacted third molar)	Dexamethasone 4 mg/mL (submucosal), No treatment (placebo)
Lim_2017A	Parallel group	Malaysia	65	Mean: 25 (SD=4)	81.67	Surgical tooth extraction (impacted third molar)	Dexamethasone 4 mg/mL (submucosal), No treatment (placebo)
Lim_2017B							Methylprednisolone 40 mg (submucosal), No treatment (placebo)
Khalida_2017	Parallel group	Pakistan	60	Mean: 28.77 (SD=7.04)	36.67	Surgical tooth extraction (impacted third molar)	Dexamethasone 4 mg/m L (submucosal), No treatment (placebo)

Bortoluzzi_2013A	Parallel group	Brazil	50	Mean: 22.5 (SD=NR)	NR	Surgical tooth extraction (impacted third molar)	Dexamethasone 8 mg + Amoxicillin 2 g, Amoxicillin 2 g (oral)
Bortoluzzi_2013B							Dexamethasone 8 mg (oral), No treatment (placebo)
Antunes_2011A	Parallel group	Brazil	60	Mean: 21 (SD=5.4)	38.4	Surgical tooth extraction (impacted third molar)	Dexamethasone 8 mg (intramuscular), No treatment (placebo)
Antunes_2011B							Dexamethasone 8 mg (oral), No treatment (placebo)
Deo_2011	Parallel group	India	30	Mean: 24.93 (SD=NR)	40	Surgical tooth extraction (impacted third molar)	Dexamethasone 8 mg (submucosal), No treatment (placebo)
Majid_2011_1A	Parallel group	Iraq	33	Mean: 26.9 (SD=6.1)	51.51	Surgical tooth extraction (impacted third molar)	Dexamethasone 4 mg (intramuscular), No treatment (placebo)
Majid_2011_1B							Dexamethasone 4 mg (submucosal), No treatment (placebo)
Majid_2011_2A	Parallel group	Iraq	30	Mean: 26.7 (SD=6.3)	46.66	Surgical tooth extraction (impacted third molar)	Dexamethasone 4 mg (intramuscular), No treatment (placebo)
Majid_2011_2B							Dexamethasone 4 mg (submucosal), No treatment (placebo)
Kang_2010A	Parallel group	Korea	450	Range: 20-30	NR	Surgical tooth extraction (impacted third molar)	Prednisolone 10 mg (oral), No treatment (placebo)
Kang_2010B							Prednisolone 20 mg (oral), No treatment (placebo)

Tiigimae_2010	Parallel group	Estonia	78	Mean: 30.57 (SD=13.73)	73	Surgical tooth extraction (impacted third molar)	Etorikoxib 120 mg preoperatively + Prednisolone 30 mg immediately postoperatively (oral), Etorikoxib 120 mg preoperatively
Vegas-Bustamante_2008	Parallel study	Spain	80	Mean: 25 (SD=5)	46	Surgical tooth extraction (impacted third molar)	Methylprednisolone 40 mg (intramuscular), No treatment
Grossi_2007A	Parallel group	Italy	61	Mean: 27.7 (SD=6.5)	45.9	Surgical tooth extraction (impacted third molar)	Dexamethasone 4 mg (submucosal), No treatment (placebo)
Grossi_2007B							Dexamethasone 8 mg (submucosal), No treatment (placebo)
Buyukkurt_2006	Parallel group	Turkey	30	Range: 18-36	44.44	Surgical tooth extraction (impacted third molar)	Prednisolone 25 mg postoperatively (intramuscular), No treatment (placebo)
Schmelzeisen_1993	Split mouth	Germany	80	Mean: 18 (SD=NR)	45	Surgical tooth extraction (impacted third molar)	Dexamethasone 6 mg 12h preoperatively and 12h postoperatively (oral), No treatment (placebo)
Pedersen_1995	Split mouth	Denmark	66	Mean: 22 (SD=NR)	56.66	Surgical tooth extraction (impacted third molar)	Dexamethasone 4 mg (intramuscular), No treatment (placebo)
El Hag_1985	Parallel group	England	70	Mean: 23.5 (SD=NR)	51.42	Surgical tooth extraction (impacted third molar)	Dexamethasone 10 mg 1h preoperatively and 10-18h postoperatively (intramuscular), No corticosteroid treatment (placebo)
Nair_2013	Parallel group	India	100	NR	NR	Surgical tooth extraction (impacted third molar)	Dexamethasone 4 mg (submucosal), No treatment (placebo)

Bauer_2012	Parallel group	Brazil	54	Mean: 22 (SD=3.6)	68.09	Surgical tooth extraction (impacted third molar)	Ibuprofen 600 mg, Dexamethasone 8 mg (oral), No corticosteroid treatment (placebo)
Simone_2013	Parallel group	Brazil	34	NR	70.59	Surgical tooth extraction (impacted third molar)	Dexamethasone 8 mg (oral), No treatment (placebo)
Majid_2013A	Parallel group	Iraq	47	Mean: 25.69 (SD=5.53)	57.45	Surgical tooth extraction (impacted third molar)	Dexamethasone 4 mg (intramuscular), No treatment (placebo)
Majid_2013B							Dexamethasone 4 mg (submucosal), No treatment (placebo)
Majid_2013C							Dexamethasone 1 mg (oral), No treatment (placebo)
Saravanan_2016 A	Parallel groups	India	60	NR	NR	Surgical tooth extraction (impacted third molar)	Dexamethasone 4 mg (intramuscular), No treatment (placebo)
Saravanan_2016 B							Dexamethasone 4 mg (submucosal), No treatment (placebo)

Risk of bias in included studies

Randomization, deviations from the intended intervention, and measurement of outcomes were the risk of bias domains judged as high or probably high risk of bias most frequently across the included studies (Appendix Table 3).

Subgroup analyses

The various types of corticosteroids identified in this review included dexamethasone, methylprednisolone, prednisolone, and triamcinolone acetonide. The route of administration varied among oral, intramuscular, and submucosal. We did not find evidence of a subgroup effect for the comparison of each specific corticosteroid with a placebo (Appendix Figures 1-5). Similarly, we did not find compelling evidence of a subgroup effect for the comparison of corticosteroids with a placebo based on administration modes (Appendix Figures 6-10). Any minor quantitative differences observed across analyses were not clinically significant.

Figure 2. Corticosteroids versus placebo for the outcome of pain at 6 hours.

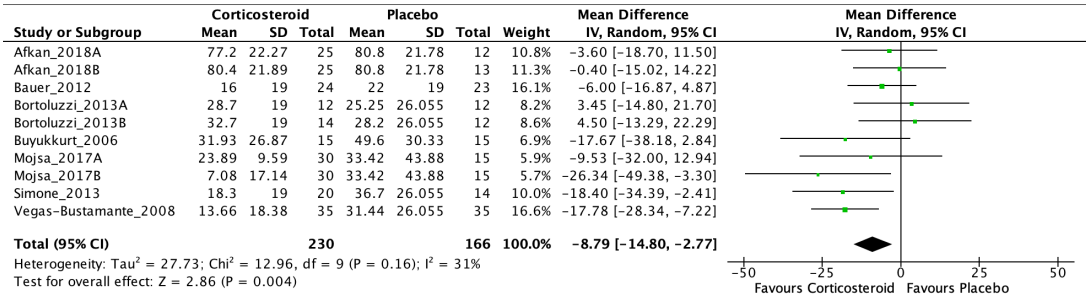


Figure 3. Corticosteroids versus placebo for the outcome of pain at 24 hours.

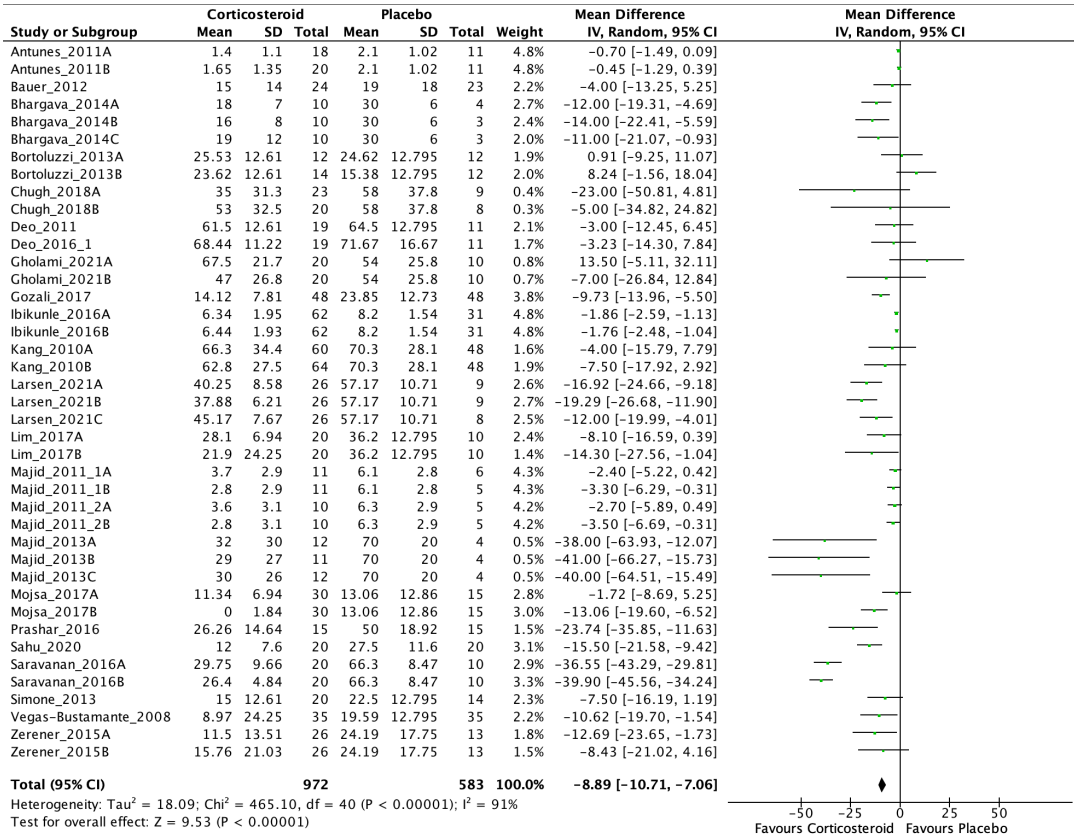


Figure 4. Corticosteroids versus placebo for the outcome of postoperative infection adverse effect.

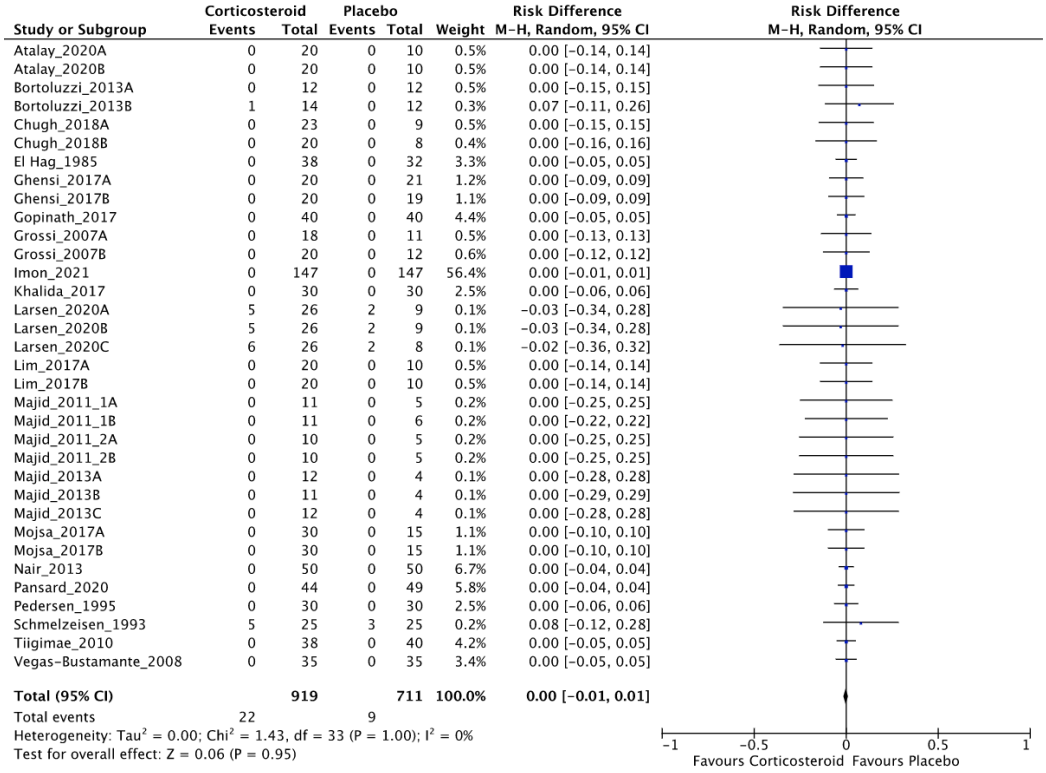


Figure 5. Corticosteroids versus placebo for the outcome of alveolar osteitis adverse effect.

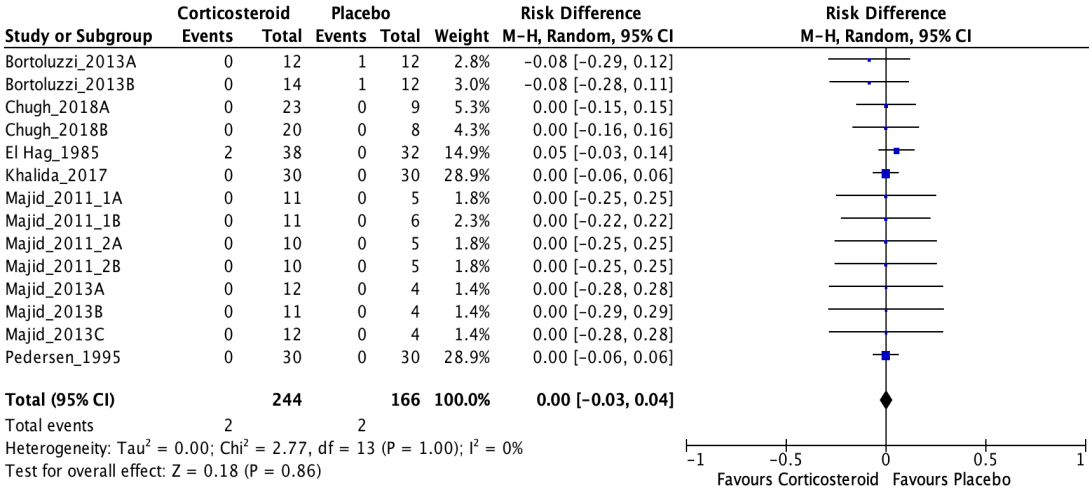
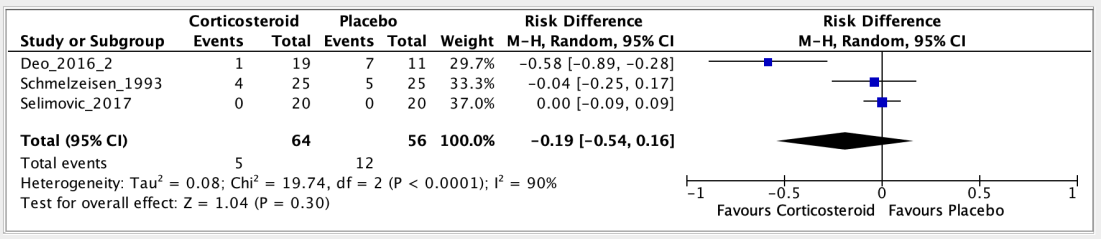


Figure 6. Corticosteroids versus placebo for the outcome of nausea/vomiting adverse effect.



Outcome Measures

Corticosteroids compared with a placebo (i.e., no treatment with corticosteroids)

1. ***Postoperative pain***

Seven RCTs, including 396 participants, assessed pain using a visual analog scale ranging from 0 (no pain) through 100 (worst pain imaginable) at 6 hours [16, 19, 21, 22, 43, 52, 54]. The results suggested that there may be a trivial benefit of corticosteroids compared with a placebo in terms of pain intensity measured 6 hours postoperatively (mean difference, 8.79 points lower; 95% CI, 14.8 to 2.77 points lower; low certainty) (Table 2, Figure 2). Twenty-three RCTs, including 1,555 participants, assessed pain using a visual analog scale ranging from 0 (no pain) through 100 (worst pain imaginable) at 24 hours [17, 19-21, 23-25, 29, 31, 33, 35, 38-43, 47-49, 52, 54-56]. Very low certainty evidence suggested that there may be a trivial difference between corticosteroids and a placebo in terms of pain intensity measured 24 hours postoperatively (mean difference, 8.89 points lower; 95% CI 10.71 to 7.06 points lower; very low certainty) (Table 2, Figure 3).

Table 2. Corticosteroids vs placebo (no treatment with corticosteroids) for acute dental pain.

Outcome № of participants (studies)	**Relative effect (95% CI)**	**Anticipated absolute effects (95% CI)**			**Certainty**	**Comments**
		With No Treatment (placebo)	**With Corticosteroids**	**Difference**		
Pain assessed with: Visual Analogue Scale from 0 (no pain) to 100 (worst pain) follow up: 6 hours № of participants: 396 (7 RCTs)	-	The median pain was **32.43** points	-	MD **8.79 points lower** (14.8 lower to 2.77 lower)	⨁⨁◯◯ LOW [a,b]	There may be a trivial benefit of corticosteroids compared to no treatment (placebo) in terms of pain measured 6 hours postoperatively.
Pain assessed with: Visual Analogue Scale from 0 (no pain) to 100 (worst pain) follow up: 24 hours № of participants: 1555 (23 RCTs)	-	The median pain was **26.06** points	-	MD **8.89 points lower** (10.71 lower to 7.06 lower)	⨁◯◯◯ VERY LOW[c,d,e]	There is very low certainty evidence regarding the difference between corticosteroids and no treatment (placebo) in terms of pain measured 24 hours postoperatively.
Adverse Effect - Postoperative Infection assessed with: Proportion of patients experiencing postoperative infection follow up: any time № of participants: 1630 (21 RCTs)	not estimable	1.3%	**2.4%**	**0% fewer** (1 fewer to 1 more)	⨁⨁◯◯ LOW [f,g]	There may be no difference between corticosteroids and no treatment (placebo) with regards to incidence of postoperative infection

Adverse Effect - Alveolar Osteitis assessed with: Proportion of patients experiencing alveolar osteitis follow up: any time № of participants: 410 (8 RCTs)	not estimable	1.2%	**0.8%**	**0% fewer** (3 fewer to 4 more)	⊕◯◯◯ VERY LOW[f,h]	There is very low certainty evidence regarding the difference between corticosteroids and no treatment (placebo) in terms of incidence of alveolar osteitis.
Adverse Effect - Gastrointestinal assessed with: Proportion of patients experiencing nausea/vomiting follow up: any time № of participants: 120 (3 RCTs)	not estimable	21.4%	**7.8%**	**19% fewer** (54 fewer to 16 more)	⊕◯◯◯ VERY LOW[i,j,k]	There is very low certainty evidence regarding the difference between corticosteroids and no treatment (placebo) in terms of incidence of nausea/vomiting .
Total Pain Relief (TOTPAR) at 6 Hours - not measured	-	-	-	-	-	
Global Efficacy Rating at 6 Hours - not measured	-	-	-	-	-	
Adverse Effect - Mood Alteration - not measured	-	-	-	-	-	

*__The risk in the intervention group__ (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; **MD:** Mean difference

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanation
a. Four studies were at a high risk of reporting bias as they did not provide measures of variability. Three studies were at either a high or probably high risk of attrition bias due to missing outcome data. Therefore, we rated down one level due to risk of bias.

b. Using a threshold of 10 points (based on 10% of the scale range), the lower bound of the confidence interval suggests an important difference favouring corticosteroids while the upper bound suggests a trivial difference favouring corticosteroids. Therefore, we rated down one level due to imprecision.
c. Several studies were probably at a high risk of bias arising from the randomization process because there was no mention of allocation concealment and the healthcare providers were not blinded or it was unclear whether they were blinded. Several studies were also at a high risk of performance and detection bias due to a lack of blinding of participants (outcome assessors). Additionally, some studies were at a high risk of reporting bias as they did not report measures of variability. Therefore, we rated down one level due to risk of bias.
d. There is high statistical heterogeneity ($I2 = 91\%$, $p < 0.00001$) and the effect estimates of some studies are importantly different from each other. Therefore, we rated down one level due to inconsistency.
e. Using a threshold of 10 points (based on 10% of the scale range), the lower bound of the confidence interval suggests an important difference favouring corticosteroids while the upper bound suggests a trivial difference favouring corticosteroids. Therefore, we rated down one level due to imprecision.
f. Several studies were probably at a high risk of bias arising from the randomization process because there was no mention of allocation concealment, and the healthcare providers were not blinded or it was unclear whether they were blinded. Several studies were also at a high risk of bias due to a lack of blinding of participants. Therefore, we rated down one level due to risk of bias.
g. Given the very low event rate, the optimal information size for this outcome was not met. Therefore, we rated down one level due to imprecision.
h. Using a threshold of 0.12% (based on 10% of the baseline risk, i.e., the risk with placebo), the lower bound of the confidence interval suggests an important difference favouring corticosteroids while the upper bound suggests an important difference favouring no treatment (placebo). Therefore, we rated down two levels due to imprecision.
i. Two studies were at a high risk of attrition bias due to missing outcome data. Two studies were at a high or probably high risk of detection bias as participants were not blinded. One of these studies also had concerns regarding reporting bias as well as the randomization process as it did not mention allocation concealment and it was unclear whether healthcare providers were blinded. Therefore, we rated down one level due to risk of bias.
j. There is high statistical heterogeneity ($I2 = 90\%$, $p < 0.0001$) and the confidence interval of one study that contributes 29.7% to the pooled estimate does not overlap with the others. Therefore, we rated down one level due to inconsistency.
k. Using a threshold of 2.14%, (based on 10% of the baseline risk, i.e., the risk with placebo) the lower bound of the confidence interval suggests an important difference favouring corticosteroids while the upper bound suggests an important difference favouring no treatment (placebo). Therefore, we rated down one level due to imprecision.

2. Reported adverse effects

1. Infection

Twenty-one RCTs, including 1,630 participants, assessed the incidence of postoperative infection as an adverse effect at different follow-up times [18, 21, 23, 27, 28, 30, 32, 34, 36, 37, 39-46]. The studies suggested that there may be no difference between corticosteroids and a placebo with regard to incidence of postoperative infection (risk difference, 0%; 95% CI, −1% to 1%; low certainty) (Table 2, Figure 4).

2. Alveolar osteitis

Eight RCTs, including 410 participants, examined the occurrence of alveolar osteitis as the proportion of participants experiencing this outcome at any time [21, 23, 27, 36, 40-42, 46]. Very low certainty evidence suggested that there may be no difference between corticosteroids and a placebo in terms of incidence of alveolar osteitis (risk difference, 0%; 95% CI, −3% to 4%; very low certainty) (Table 2, Figure 5).

3. GI adverse effects

Three RCTs, including 120 participants, provided evidence on the incidence of GI adverse effects expressed as the proportion of participants experiencing nausea or vomiting at any time [30, 54, 55] (Table 2, Figure 6).

Corticosteroids may reduce the incidence of GI adverse effects compared with a placebo; however, the evidence is uncertain (risk difference, −19%; 95% CI, −54% to 16%; very low certainty) (Table 2, Figure 6). In addition, although we identified a statistically significant difference in GI adverse effects depending on the route of administration (oral multidose vs submucosal single dose), the small number of events and studies did not provide compelling evidence to claim a subgroup effect (Table 2, Appendix Figure 10).

4. Total pain relief, global efficacy rating, and mood alteration

None of the included studies provided evidence related to the effect of corticosteroids compared with a placebo on total pain relief, global efficacy rating, or mood alteration outcomes.

DISCUSSION

Low certainty evidence suggested that corticosteroids administered orally, submucosally, or intramuscularly in adolescent, adult, or older adult participants may decrease pain intensity a trivial amount at 6 hours compared with a placebo (no treatment with corticosteroids). This difference remained trivial at 24 hours (very low certainty). We found no difference between corticosteroids compared with a placebo with regard to the incidence of postoperative infection (low certainty) and alveolar osteitis (very low certainty). With regard to adverse events, in particular GI events (for example, nausea and vomiting), the available evidence suggested that corticosteroids could result in a small benefit (that is, a reduction in GI adverse events) compared with a placebo (very low certainty); however, these results are not trustworthy due to the small number of studies and participants. Most of the included RCTs had serious issues related to the randomization process, deviations from the intended intervention, and measurement of outcomes. In addition, none of the primary studies were conducted in the United States, where intravenous corticosteroids (not oral, submucosal, intramuscular) are often administered, patients regularly receive intravenous sedation,

and more than 1 mandibular third molar may be extracted at a single appointment. To improve the certainty of the evidence, future trials should focus on conducting methodologically rigorous RCTs and increasing the sample size.

In an SR including 12 RCTs, researchers examined the impact of corticosteroids on postoperative morbidity after third-molar extraction [6]. The researchers examined the effects of perioperative corticosteroid administration (that is, betamethasone, prednisolone, methylprednisolone, methylprednisolone acetate, methylprednisolone sodium succinate, dexamethasone) on pain intensity, trismus, and swelling. Their finding suggested that corticosteroids reduced swelling and trismus in the early (1–3 days) and late (> 3 days) postoperative phases. The researchers could not determine the effect of corticosteroids on pain intensity. This differed from our findings, which suggested a negligible difference between corticosteroids and a placebo at 6 and 24 hours postoperatively. In another SR, researchers also examined the effect of corticosteroids and included 28 RCTs [4]. The researchers concluded that corticosteroids (that is, dexamethasone, prednisolone, methylprednisolone, betamethasone) improved patients' postoperative experiences and had a significant benefit on trismus and inflammation. As we focused our review on pain intensity and adverse effects, we did not examine the evidence regarding trismus, swelling, or inflammation; therefore, we cannot compare the results of the reviews cited above with the results of our review.

In a third SR and meta-analysis, including 8 RCTs, researchers assessed the effect of submucosal injection of dexamethasone after third-molar extraction [7]. The results of the review suggested that the submucosal injection of dexamethasone resulted in a reduction in swelling and pain from impacted third-molar surgery. In that review, the researchers also found no difference between dexamethasone and a placebo in relation to trismus. These results are similar to those from our review in terms of reduction of pain with the use of corticosteroids.

In a fourth SR, including 7 RCTs, researchers examined the impact of different dosages of corticosteroids (that is, cortisol, prednisone, prednisolone, methylprednisolone, dexamethasone, and betamethasone) and administration routes on facial swelling, pain, and trismus [5]. The researchers concluded that preoperative submucosal injection of corticosteroids significantly diminished facial swelling, postoperative pain, and trismus compared with a placebo. Although we did not examine facial swelling and trismus, our findings are aligned regarding pain reduction, although the magnitude of effect that we found was lower than our clinically significant threshold (10%). The researchers concluded that the optimal dosage of corticosteroids and administration route for decreasing postsurgical morbidity and improving quality of life after surgical removal of mandibular third molars was unknown [5].

In a fifth SR, researchers analyzed the efficacy of corticosteroids for pain management after mandibular third-molar extraction [8]. The review included 27 RCTs and the researchers concluded that corticosteroids (that is, dexamethasone,

methylprednisolone, and betamethasone) could be used as an adjuvant for pain reduction after an impacted third-molar extraction. The researchers also suggested that methylprednisolone and dexamethasone delivered via an intramuscular route were the best interventions for effective pain reduction. The ideal time for administration of corticosteroids was the preoperative period. These results were similar to the results of our review in terms of reduction of pain. However, our review included dexamethasone, methylprednisolone, prednisolone, and triamcinolone acetonide and did not find a difference in pain intensity among the different corticosteroids.

The researchers in the SRs mentioned did not assess the certainty of the evidence. The results of our SR established that the best evidence supporting the outcomes of choice ranged from low through very low certainty, highlighting the need for RCTs of higher methodological and reporting quality, as well as statistical precision.

Strengths and limitations

The strengths of our review relied on each stage of the process being conducted independently and in duplicate. We assessed the risk of bias for each RCT and the certainty of the evidence for each outcome of interest. We performed the analyses and interpreted the results using the latest methodological guidance from the GRADE working group. A limitation is that we restricted our eligibility criteria to peer-reviewed research articles published in English. However, we believe it is unlikely that our conclusions would have been different had we included studies in a different language.

CONCLUSION

The results of our SR and meta-analysis indicated that there was low and very low certainty evidence informing the effect of corticosteroids administered orally, submucosally, or intramuscularly in adolescent, adult, or older adult participants compared with a placebo for the management of acute pain after surgical tooth extraction. Patients receiving corticosteroids in the modalities above may experience a trivial reduction in pain intensity compared with those receiving a placebo at 6- and 24-hour follow-ups; however, the evidence is uncertain. Future research should focus on examining the effect of corticosteroids on various patient-important outcomes in patients undergoing surgical tooth extractions.

Supplemental material

A supplemental appendix to this article is available online https://doi.org/10.1016/j.adaj.2023.04.018.

Conflicts of interest statement

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

Funding statement

This project was financially supported by the FDA of the US Department of Health and Human Services (HHS). The contents are those of the author(s) and do not necessarily represent the official views of, nor an endorsement by, FDA/HHS or the US government.

The findings and conclusions in this project are those of the authors and do not necessarily reflect the official views, position, or policy of, nor an endorsement by, the US Food and Drug Administration, US Department of Health and Human Services, the US government, and the American Dental Association. Presented at the International Association for Dental Research General Session and Exhibition, June 25, 2022. https://www.iadr.org/2022iags

Ethics approval statement

The ethics approval was not required.

REFERENCES

1. Friedman, J.W., *The prophylactic extraction of third molars: a public health hazard.* Am J Public Health, 2007. **97**(9): p. 1554-9.
2. Barnes, P.J., *How corticosteroids control inflammation: Quintiles Prize Lecture 2005.* Br J Pharmacol, 2006. **148**(3): p. 245-54.
3. Ngeow, W.C. and D. Lim, *Do Corticosteroids Still Have a Role in the Management of Third Molar Surgery?* Adv Ther, 2016. **33**(7): p. 1105-39.
4. Herrera-Briones, F.J., et al., *Update on the use of corticosteroids in third molar surgery: systematic review of the literature.* Oral Surg Oral Med Oral Pathol Oral Radiol, 2013. **116**(5): p. e342-51.

5. Larsen, M.K., et al., *Different Dosages of Corticosteroid and Routes of Administration in Mandibular Third Molar Surgery: a Systematic Review.* J Oral Maxillofac Res, 2018. **9**(2): p. e1.
6. Markiewicz, M.R., et al., *Corticosteroids reduce postoperative morbidity after third molar surgery: a systematic review and meta-analysis.* J Oral Maxillofac Surg, 2008. **66**(9): p. 1881-94.
7. Moraschini, V., R. Hidalgo, and E. Porto Barboza, *Effect of submucosal injection of dexamethasone after third molar surgery: a meta-analysis of randomized controlled trials.* Int J Oral Maxillofac Surg, 2016. **45**(2): p. 232-40.
8. Sugragan, C., et al., *Do corticosteroids reduce postoperative pain following third molar intervention?* J Dent Anesth Pain Med, 2020. **20**(5): p. 281-291.
9. Guyatt, G.H., et al., *What is "quality of evidence" and why is it important to clinicians?* Bmj, 2008. **336**(7651): p. 995-8.
10. Page, M.J., et al., *The PRISMA 2020 statement: an updated guideline for reporting systematic reviews.* Bmj, 2021. **372**: p. n71.
11. National Academies of Sciences Engineering Medicine Health and Medicine Division Board, *National Academies of Sciences Engineering Medicine Health and Medicine Division Board on Health Care Services Committee on Evidence-Based Clinical Practice Guidelines for Prescribing Opioids for Acute Pain,*, in *Framing Opioid Prescribing Guidelines for Acute Pain: Developing the Evidence*. 2019, National Academies Press (US)
Copyright 2020 by the National Academy of Sciences. All rights reserved.: Washington (DC).
12. Rada, G., et al., *Epistemonikos: a comprehensive database of systematic reviews for health decision-making.* BMC Med Res Methodol, 2020. **20**(1): p. 286.
13. Sterne, J.A.C., et al., *RoB 2: a revised tool for assessing risk of bias in randomised trials.* BMJ, 2019. **366**: p. l4898.
14. Hultcrantz, M., et al., *The GRADE Working Group clarifies the construct of certainty of evidence.* J Clin Epidemiol, 2017. **87**: p. 4-13.
15. Zeng, L., et al., *GRADE guidelines 32: GRADE offers guidance on choosing targets of GRADE certainty of evidence ratings.* J Clin Epidemiol, 2021. **137**: p. 163-175.
16. Afkan SS, M.M., Razavi M, Kazemi P., *Clinical Evaluation of Dexamethasone Pills on Side Effects of Mandibular Third Molars Surgery.* J Res Med Dent Sci, 2018. **6**(1): p. 98-106.
17. Antunes, A.A., et al., *Effect of two routes of administration of dexamethasone on pain, edema, and trismus in impacted lower third molar surgery.* Oral Maxillofac Surg, 2011. **15**(4): p. 217-23.
18. Atalay, B., A.T. Şitilci, and D. Onur Ö, *Analgesic and Anti-Inflammatory Effects of Articaine and Perineural Dexamethasone for Mandibular Third Molar Surgery: A Randomized, Double-Blind Study.* J Oral Maxillofac Surg, 2020. **78**(4): p. 507-514.

19. Bauer, H.C., et al., *Assessment of preemptive analgesia with ibuprofen coadministered or not with dexamethasone in third molar surgery: a randomized double-blind controlled clinical trial.* Oral Maxillofac Surg, 2013. **17**(3): p. 165-71.
20. Bhargava, D., K. Sreekumar, and A. Deshpande, *Effects of intra-space injection of Twin mix versus intraoral-submucosal, intramuscular, intravenous and per-oral administration of dexamethasone on post-operative sequelae after mandibular impacted third molar surgery: a preliminary clinical comparative study.* Oral Maxillofac Surg, 2014. **18**(3): p. 293-6.
21. Bortoluzzi, M.C., et al., *A single dose of amoxicillin and dexamethasone for prevention of postoperative complications in third molar surgery: a randomized, double-blind, placebo controlled clinical trial.* J Clin Med Res, 2013. **5**(1): p. 26-33.
22. Buyukkurt, M.C., M. Gungormus, and O. Kaya, *The effect of a single dose prednisolone with and without diclofenac on pain, trismus, and swelling after removal of mandibular third molars.* J Oral Maxillofac Surg, 2006. **64**(12): p. 1761-6.
23. Chugh, A., et al., *Submucosal injection of dexamethasone and methylprednisolone for the control of postoperative sequelae after third molar surgery: randomized controlled trial.* Int J Oral Maxillofac Surg, 2018. **47**(2): p. 228-233.
24. Deo, S.P., *Effect of submucosal injection of dexamethasone on post-operative sequelae of third molar surgery.* JNMA J Nepal Med Assoc, 2011. **51**(182): p. 72-8.
25. Deo, S.P., *Role of addition of dexamethasone to lignocaine 2% with adrenaline in dental nerve blocks for third molar surgery: A prospective randomized control trial.* Ann Maxillofac Surg, 2016. **6**(2): p. 260-266.
26. Deo, S.P., *Single-Dose of Submucosal Injection of Dexamethasone Affects the Post Operative Quality of Life After Third Molar Surgery.* J Maxillofac Oral Surg, 2016. **15**(3): p. 367-375.
27. ElHag, M., et al., *The anti-inflammatory effects of dexamethasone and therapeutic ultrasound in oral surgery.* Br J Oral Maxillofac Surg, 1985. **23**(1): p. 17-23.
28. Ghensi, P., et al., *Effect of Oral Administration of Bromelain on Postoperative Discomfort After Third Molar Surgery.* J Craniofac Surg, 2017. **28**(2): p. e191-e197.
29. Gholami, M., et al., *What Are the Effects of Methylprednisolone Injection Into the Masseter and Gluteal Muscle on Pain, Edema and Trismus After Impacted Lower Third Molar Surgery? A Randomized Clinical Trial.* J Oral Maxillofac Surg, 2021. **79**(9): p. 1829-1836.
30. Gopinath KA, C.M., Arun V., *Comparative evaluation of submucosal and intravenous dexamethasone on postoperative sequelae following third molar surgery: A prospective randomized control study.* Int J Oral Care Res, 2017. **5**(3): p. 191-195.

31. Gozali, P., et al., *Decreased post-operative pain using a sublingual injection of dexamethasone (8 mg) in lower third molar surgery.* J Dent Anesth Pain Med, 2017. **17**(1): p. 47-53.
32. Grossi, G.B., et al., *Effect of submucosal injection of dexamethasone on postoperative discomfort after third molar surgery: a prospective study.* J Oral Maxillofac Surg, 2007. **65**(11): p. 2218-26.
33. Ibikunle, A.A., W.L. Adeyemo, and A.L. Ladeinde, *Effect of submucosal or oral administration of prednisolone on postoperative sequelae following surgical extraction of impacted mandibular third molar: A randomized controlled study.* Niger Med J, 2016. **57**(5): p. 272-279.
34. Imon AA, A.M., Costa PD, Sabrina F, Haque N., *Comparative study for effectiveness of oral dexamethasone for assessment of pain, swelling and trismus following impacted 3rd molar surgery.* SVOA Dentistry, 2021. **2**(1): p. 20-24.
35. Kang, S.H., et al., *Effect of preoperative prednisolone on clinical postoperative symptoms after surgical extractions of mandibular third molars.* Aust Dent J, 2010. **55**(4): p. 462-7.
36. Khalida B, F.M., Muntaha ST, Khan K., *Effect of submucosal injection of dexamethasone on post-operative swelling and trismus following impacted mandibular third molar surgery.* Pakistan Oral & Dental Journal, 2017. **37**(2): p. 231-234.
37. Larsen, M.K., et al., *Short-term Haematological Parameters Following Surgical Removal of Mandibular Third Molars with Different Doses of Methylprednisolone Compared with Placebo. A Randomized Controlled Trial.* J Oral Maxillofac Res, 2020. **11**(2): p. e3.
38. Larsen, M.K., et al., *Efficacy of methylprednisolone on pain, trismus and quality of life following surgical removal of mandibular third molars: a double-blind, split-mouth, randomised controlled trial.* Med Oral Patol Oral Cir Bucal, 2021. **26**(2): p. e156-e163.
39. Lim, D. and W.C. Ngeow, *A Comparative Study on the Efficacy of Submucosal Injection of Dexamethasone Versus Methylprednisolone in Reducing Postoperative Sequelae After Third Molar Surgery.* J Oral Maxillofac Surg, 2017. **75**(11): p. 2278-2286.
40. Majid, O.W., *Submucosal dexamethasone injection improves quality of life measures after third molar surgery: a comparative study.* J Oral Maxillofac Surg, 2011. **69**(9): p. 2289-97.
41. Majid, O.W. and W.K. Mahmood, *Effect of submucosal and intramuscular dexamethasone on postoperative sequelae after third molar surgery: comparative study.* Br J Oral Maxillofac Surg, 2011. **49**(8): p. 647-52.
42. Majid, O.W. and W.K. Mahmood, *Use of dexamethasone to minimise post-operative sequelae after third molar surgery: comparison of five different routes of administration.* Oral Surgery, 2013. **6**(4): p. 200-208.
43. Mojsa, I.M., et al., *Effect of submucosal dexamethasone injection on postoperative pain, oedema, and trismus following mandibular third molar*

surgery: a prospective, randomized, double-blind clinical trial. Int J Oral Maxillofac Surg, 2017. **46**(4): p. 524-530.
44. Nair, R.B., et al., *Effect of submucosal injection of dexamethasone on postoperative discomfort after third molar surgery: a prospective study.* J Contemp Dent Pract, 2013. **14**(3): p. 401-4.
45. Pansard, H.B., et al., *The Impact of Prior Use of Corticosteroid to Dental Extraction on Oral Health-Related Quality-of-Life and Clinical Outcomes: A Randomized Clinical Trial.* J Oral Maxillofac Surg, 2020. **78**(12): p. 2153.e1-2153.e9.
46. Pedersen, A., *Decadronphosphate in the relief of complaints after third molar surgery. A double-blind, controlled trial with bilateral oral surgery.* Int J Oral Surg, 1985. **14**(3): p. 235-40.
47. Prashar, D.V., et al., *A comparative evaluation of the effect of diclofenac sodium with and without per-orally administered methylprednisolone on the sequelae of impacted mandibular third molar removal: A cohort randomized double-blind clinical trial.* Indian J Dent, 2016. **7**(1): p. 11-6.
48. Sahu, S., et al., *Comparative evaluation of efficacy and latency of twin mix vs 2% lignocaine HCL with 1:80000 epinephrine in surgical removal of impacted mandibular third molar.* J Family Med Prim Care, 2020. **9**(2): p. 904-908.
49. Saravanan, K., et al., *A Single Pre Operative Dose of Sub Mucosal Dexamethasone is Effective in Improving Post Operative Quality of Life in the Surgical Management of Impacted Third Molars: A Comparative Randomised Prospective Study.* J Maxillofac Oral Surg, 2016. **15**(1): p. 67-71.
50. Schmelzeisen, R. and J.C. Frölich, *Prevention of postoperative swelling and pain by dexamethasone after operative removal of impacted third molar teeth.* Eur J Clin Pharmacol, 1993. **44**(3): p. 275-7.
51. Selimović, E., et al., *Prevention of trismus with different pharmacological therapies after surgical extraction of impacted mandibular third molar.* Med Glas (Zenica), 2017. **14**(1): p. 145-151.
52. Simone, J.L., et al., *Comparative analysis of preemptive analgesic effect of dexamethasone and diclofenac following third molar surgery.* Braz Oral Res, 2013. **27**(3): p. 266-71.
53. Tiigimae-Saar, J., E. Leibur, and T. Tamme, *The effect of prednisolone on reduction of complaints after impacted third molar removal.* Stomatologija, 2010. **12**(1): p. 17-22.
54. Vegas-Bustamante, E., et al., *Efficacy of methylprednisolone injected into the masseter muscle following the surgical extraction of impacted lower third molars.* Int J Oral Maxillofac Surg, 2008. **37**(3): p. 260-3.
55. Zerener, T., et al., *Clinical comparison of submucosal injection of dexamethasone and triamcinolone acetonide on postoperative discomfort after third molar surgery.* Quintessence Int, 2015. **46**(4): p. 317-26.
56. Locker, D. and M. Grushka, *The impact of dental and facial pain.* J Dent Res, 1987. **66**(9): p. 1414-7.

CHAPTER 4: GENDER-AFFIRMING HORMONE THERAPY FOR INDIVIDUALS WITH GENDER DYSPHORIA BELOW 26 YEARS OF AGE: A SYSTEMATIC REVIEW AND META-ANALYSIS

Anna Miroshnychenko, Sara Ibrahim, Yetiani Roldan, Chan Kulatunga-Moruzi, Steven Montante, Rachel Couban, Gordon Guyatt, Romina Brignardello-Petersen

Published: Archives of Disease in Childhood Journal [February 12, 2025]

ABSTRACT

Objective: In this systematic review and meta-analysis, we assess and summarize the certainty of the evidence about the effects of gender-affirming hormone therapy (GAHT) in individuals experiencing gender dysphoria (GD).

Methods: We searched MEDLINE, Embase, PsycINFO, Social Sciences Abstracts, LGBTQ+ Source, and Sociological Abstracts through September 2023. We included studies comparing GAHT to no GAHT in individuals under 26 years of age experiencing GD. Outcomes of interest included psychological and physical. Pairs of reviewers independently screened articles, abstracted data, and assessed the risk of bias in included studies. We performed meta-analyses and assessed the certainty of the evidence using the GRADE approach.

Results: We included 24 studies. Comparative observational studies (n=9) provided mostly very low certainty evidence regarding gender dysphoria, global function, and depression. One of nine comparative observational studies reported that the odds of depression may be lower (OR 0.73 [95% CI 0.61 to 0.88], n (number of studies) =1, low certainty) in individuals who received GAHT compared to those who did not. Before-after studies (n=13) provided very low certainty evidence about gender dysphoria, global function, depression, and BMD. One of two (n=2) case series studies provided high certainty evidence that the proportion of individuals with cardiovascular

events 7-109 months after receiving GAHT was 0.04 (95% CI 0.03 to 0.05, n = 1, high certainty).

Conclusion: *There is considerable uncertainty about the effects of GAHT, and we cannot exclude the possibility of benefit or harm. Methodologically rigorous prospective studies are needed to understand the effects of this intervention.*

Key messages:

1. *What is known on this topic:* Previously published evidence syntheses addressing the effects of GAHT in individuals experiencing GD are methodologically limited.
2. *What this study adds:* This publication addresses the effects of GAHT in individuals experiencing GD, while adhering to the highest methodological standards for conducting and reporting a systematic review and meta-analysis, and assessing the risk of bias in each included study and the certainty of the evidence for each outcome of interest.
3. *How this study might affect research, practice, and policy*: The evidence from this systematic review and meta-analysis can be used to inform individuals experiencing GD and considering GAHT, clinicians involved in their care as well as clinical practice guideline developers, policy makers and stakeholders who make decisions about treatment related to gender dysphoria.

INTRODUCTION

Gender dysphoria (GD) refers to the intense distress caused by feelings of incongruence between one's birth-assigned sex and gender identity [1]. Individuals who experience persistent GD may seek hormonal and surgical interventions to align their physical bodies with their internal or expressed gender and alleviate this distress [2].

Hormonal treatments for GD in youth include gonadotropin releasing hormone analogues (GnRHa) and gender-affirming hormone therapy (GAHT). The former, GnRHa (puberty blockers), may be administered as early as Tanner Stage 2 [3], followed by GAHT in adolescence to induce and maintain the desired secondary sex characteristics. Hormone therapies include the administration of testosterone for natal females (NF) to create a masculinized appearance, and estrogen in conjunction with GnRHa for natal males (NM) to produce a feminized appearance. Early interventions with puberty blockers followed by GAHT is believed to result in better physical outcomes aligned with the desired gender [4, 5], though some individuals may receive only GAHT.

A high-quality SR is needed to overcome methodological limitations in this field. This SR and meta-analysis aimed to summarize the effects of GAHT in individuals with GD under the age of 26.

METHODS

We report this SR and meta-analysis following the guidance of the preferred reporting items for systematic reviews and meta-analyses (PRISMA) checklist (Appendix 1). We registered the protocol in PROSPERO (registration ID: CRD42023452171).

Eligibility criteria

For eligibility criteria, see Table 1.

Table 1. Eligibility criteria.

Types of studies	We included randomized controlled trials, comparative observational studies, and before-after studies addressing the intervention and comparison of interest. We also included case series addressing the intervention of interest in special instances. We did not find any RCTs and included all eligible comparative observational and before-after studies. As for case series, if an outcome of interest was not reported in the eligible comparative observational or before-after studies, we included all eligible case series studies addressing that outcome. We included studies published in full, and in English language.

Population	We included individuals under 26 years, who were diagnosed, experienced, self-identified, or were identified by a parent as having GD, gender identity disorder or gender incongruence, or who identified as transgender or non-binary. To be as inclusive as possible, we included all studies where the mean age of participants was below 26 years. We decided to include individuals below 26 years of age because the definition of youth, the target population of this review, is commonly defined as extending into the mid-twenties [6, 7].
Intervention	We included studies assessing the effects of GAHT. We defined GAHT as stated by the authors or as the use of feminizing hormones in an individual assigned male at birth or as the use of masculinizing hormones in an individual assigned female at birth.
Comparator	The comparator of interest was no GAHT (e.g., psychological therapy, no treatment). In case series studies, a comparator group was not necessary.
Outcomes	We included studies reporting on any of the following outcomes if follow up was short term (≤6 months) or long-term (≥1 year): gender dysphoria, completed suicides, global function, depression, sexual dysfunction from physiological perspective (i.e., lack of erection, dyspareunia, problems related

	to dry and degenerated mucosal tissue, anorgasmia), bone mineral density (BMD), and cardiovascular events.

Information sources

With the assistance of an information specialist (RC), we searched MEDLINE, Embase, PsycINFO, Social Sciences Abstracts, Contemporary Women's Issues, LGBTQ+ Source, Sociological Abstracts, Studies on Women, Gender Abstracts, and Google Scholar from inception to September 2023. This search was part of an umbrella search for a related SR [8]. All search strategies are in Appendix 2.

Study selection

Two reviewers (SI, YR), using Covidence software (https://www.covidence.org/) and following training and calibration exercises, independently screened titles and abstracts, and full texts of potentially eligible studies. A third reviewer (AM) resolved conflicts. The study selection for this SR was completed in tandem with another related SR at the abstract and full-text stages [8].

Data collection

For data collection, see Appendix 3.

Risk of bias in individual studies

We assessed the risk of bias using a modified version of the Cochrane risk of bias tool for non-randomized studies of interventions (ROBINS-I) for each study design. For details, see Appendix 4 and 5.

Data synthesis

Although study authors used various observational study designs, we classified studies according to how the data were analyzed for this review. See Appendix 6.

For dichotomous outcomes, we summarized the effect of interventions using odds ratios in comparative observational and before-after studies, and proportions (i.e., number of events per number of participants in the study group) in case series studies. For continuous outcomes, we summarized the effects of interventions using mean difference in comparative observational studies (i.e., difference in scores between the study groups), mean change in before-after studies (i.e., difference in scores before and after intervention), and mean in case series.

Since the study authors did not provide correlation coefficients, we imputed a moderate correlation coefficient (r=0.5) when calculating mean change. We calculated 95% confidence intervals (CI) around all estimates.

We conducted meta-analysis using a random-effects model when appropriate, as determined by subject area experts (CKM, SM), for studies addressing the same outcome and with no clinical heterogeneity between them (i.e., study design, population, intervention/comparator, outcome definition). When studies reported outcomes using different scales, we calculated the standardized mean change for before-after studies. If we could not perform a meta-analysis, we summarized the evidence across studies. We used the *meta* and *metafor* packages in R Studio Version 4.2 for analyses.

Certainty of the evidence

We assessed the certainty of the evidence following grading of recommendations assessment, development, and evaluation (GRADE) approach [9]. For details, see Appendix 7. We assessed the certainty in the causal effects of GAHT on the outcomes of interest rather than the association between GAHT as an exposure. We followed GRADE guidance and principles to address questions about interventions using observational studies. This process involves clarifying the question (target of certainty), defining the intent of the question (causality), and assessing the certainty of the evidence under those parameters.

When assessing risk of bias for each outcome, we rated down the certainty of evidence from observational studies by up to three levels due to prognostic imbalance. In case

series, outcomes requiring a comparison group (e.g., GD, completed suicides, global function, depression, sexual dysfunction, BMD) were rated down three levels due to the absence of such a group. However, outcomes not requiring a comparison group (e.g., cardiovascular events linked to gender-affirming hormones) were not rated down, as these events were specific to intervention recipients.

To minimize value judgments, we used a null effect threshold (1 for relative measures, and 0 for absolute measures, mean differences, or mean changes) to rate the certainty of any benefit or harm (of any magnitude) from receiving GAHT over not receiving GAHT. We did not define a minimally important difference to determine whether an effect was clinically meaningful or important.

Subgroup and sensitivity analyses

For subgroup and sensitivity analyses, see Appendix 8.

Management of conflicts of interest

For management of conflicts of interest, see Appendix 9. Other SRs under the described agreement include SRs about the effects of social gender transition, mastectomy [10], chest binding and genital tucking, and puberty blockers (all submitted for publication).

RESULTS

After screening 6,736 titles and abstracts for this SR and another related [8], we included 24 studies. Figure 1 shows the study search and selection process. We present reasons for exclusion at the full-text screening stage (n=311) with references in Appendix 10.

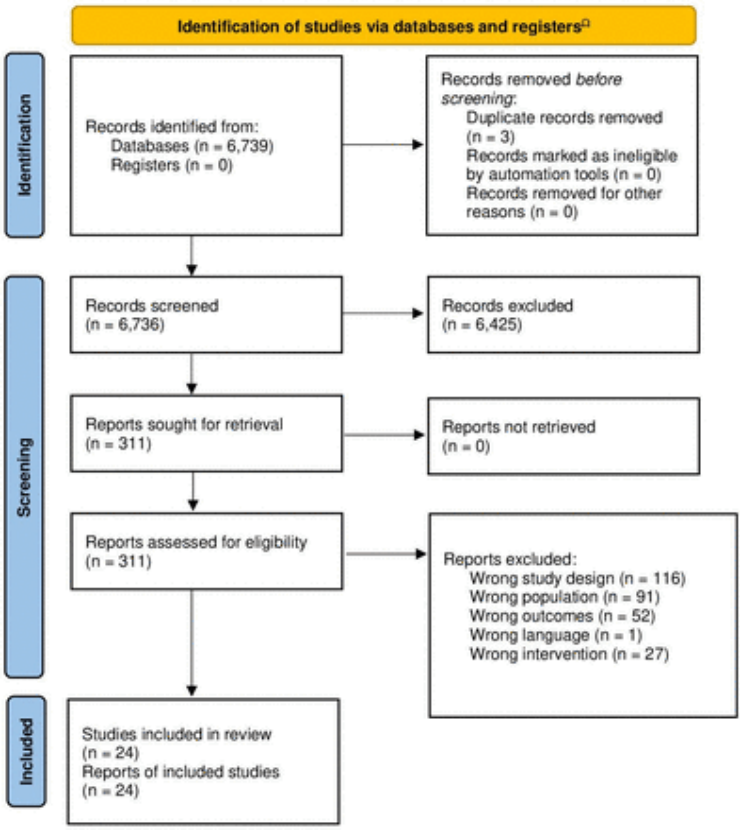


Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-analyses 2020 flow diagram for new systematic reviews which included searches of databases and register only. Ω This was an umbrella search completed for two related systematic reviews and meta-analyses. 24 studies were included in this systematic review. The studies that were in another review are part of the studies excluded for wrong

intervention. *Ten of 27 studies excluded for wrong intervention were included in another review.

Characteristics of included studies

Of 24 included studies, 9 were comparative observational [11-19], 13 were before-after [20-32], and 2 were case series [33, 34] (Figure 1). Thirteen studies included NMs and NFs, and 11 included NFs only.

The mean (SD) age of participants at the time of GAHT ranged from 15.1 (1.8) to 25.1 (4.8). We present characteristics of included studies in Appendix 11. Appendix 12 describes outcome measurement instruments used in the studies and their interpretability.

Risk of bias in included studies

Across comparative observational studies, the domains commonly judged as serious or critical risk of bias were confounding, missing data, and deviation from intended intervention (i.e., administration of co-interventions). Before-after studies were at serious or critical risk of bias due to missing data and deviation from intended intervention (i.e., administration of co-interventions). In addition to lacking a comparison group, case series were at critical risk of bias due to measurement of the outcome (Appendix 13).

Effects of interventions

We describe the effects of the interventions for each study design. Tables 1-3 provide summary of findings tables, and appendix 14 displays forest plots of meta-analysis. If sex-specific data were available, we included separate data points for NMs and NFs in each meta-analysis (Appendix 14). When studies reported data for both groups and no important heterogeneity was found, we presented a single combined effect estimate.

1. **Comparative observational studies**

See Table 2 for summary of findings table.

Gender dysphoria (GD): Current GD, using the Gender Distress Scale ranging from 1 to 5, may be lower (MD (mean difference) 0.4 lower [95% CI 0.24 lower to 0.16 higher], number of studies (n) = 1, very low certainty) in NMs and NFs who received GAHT compared to those who did not; however, we are very uncertain about the causal effect of the intervention on GD [15].

Global function: A meta-analysis suggests that global function, measured within the last 12 to 24 months, may be higher (standardized mean difference (SMD) 0.87 higher, [95% CI 0.25 lower to 2 higher], n = 2, very low certainty) in NMs and NFs who

received GAHT compared to those who did not; however, we are very uncertain about the causal effect of the intervention on global function [35, 36].

Depression: Eight studies reported this outcome using seven different measurement instruments. Due to variability in instruments, time points, and reporting, we could not include all studies in a single meta-analysis.

A meta-analysis suggests that depression, measured within the last 12 months, may be lower (SMD 0.3, [95% CI 0.85 lower to 0.25 higher], n = 2, very low certainty) in NMs and NFs who received GAHT compared to those who did not; however, we are very uncertain about the causal effect of the intervention on depression [16, 36]. See Appendices 15 and 16 for low to very low certainty evidence about depression from studies not pooled with this evidence.

Table 2. Gender-affirming hormone therapy vs no gender-affirming hormone therapy: evidence from comparative observational studies.

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with no gender-affirming hormone therapy	Risk with gender-affirming hormone therapy				

Gender Dysphoria, current (no follow-up) assessed with: participant reported Gender Distress Scale, higher scores indicate higher gender distress Scale from: 1 to 5	The mean gender Dysphoria, current (no follow-up) was **4.17**	MD **0.4 lower** (0.24 lower to 0.16 higher)	-	146 (1 non-randomised study)[1]	⨁◯◯◯ Very low [a, b]	The evidence is very uncertain about the effect of gender-affirming hormone therapy on gender dysphoria (no follow-up) in natal males and natal females.
Global Function, Long Term Follow-Up assessed with: participant reported, various scales [Symptom Checklist-90 Revised (SCL-90-R) Global Severity Index, The Children's Global Assessment Scale (CGAS)], higher scores indicate better global function follow-up: range 12 months to 24 months [c]	-	SMD **0.87 SD higher** (0.25 lower to 2 higher)	-	125 (2 non-randomised studies)[2,3]	⨁◯◯◯ Very low [d, e, f]	The evidence is very uncertain about the effect of gender-affirming hormone therapy on global function at long term follow-up in natal males and natal females.
Depression, Long Term Follow-Up assessed with: participant reported, various scales [Symptom Checklist-90 Revised (SCL-90-R) Depression Domain, Children's Depression Inventory (CDI)], higher scores represent worse depression follow-up: mean 12 months [c]	-	SMD **0.3 SD lower** (0.85 lower to 0.25 higher)	-	154 (2 non-randomised studies)[2,4]	⨁◯◯◯ Very low [f, g, h]	The evidence is very uncertain about the effect of gender-affirming hormone therapy on depression at long term follow-up in natal males and natal females.
Other Outcomes - not measured [i]	-	-	-	-	-	-

*__The risk in the intervention group__ (and its 95% confidence interval) is based on the assumed risk in the comparison group and the __relative effect__ of the intervention (and its 95% CI).

CI: confidence interval; **MD:** mean difference; **OR:** odds ratio; **SMD:** standardised mean difference

GRADE Working Group grades of evidence
High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.
Moderate certainty: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
Low certainty: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.
Very low certainty: we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

Explanations

a. Rated down three levels due to critical risk of bias because of lack of adjustment for important confounders (i.e., psychiatric interventions, mental health comorbidities, socioeconomic status, and family support) and missing data (i.e., 41.71% provided outcome data).
b. Rated down one level for imprecision as the confidence intervals cross the threshold of no effect (i.e., MD=0), suggesting both a possibility of a benefit or harm in the outcome.
c. Long Term Follow-Up: outcome measured at ≥ 12 months follow-up.
d. Rated down three levels due to critical risk of bias because of lack of adjustment for important confounders in the two included studies (i.e., psychiatric interventions, mental health comorbidities, socioeconomic status, and family support) and missing data in one of the two included studies (i.e., 37% provided outcome data).
e. Statistically, there was considerable heterogeneity with $I2=88\%$ and $p<0.01$. However, we did not rate down for inconsistency as this heterogeneity could be explained by the fact that one of the two included studies measured the outcome only in natal female participants, while the other study measured the outcome in natal female and male participants.
f. Rated down one level for imprecision as the confidence intervals cross the threshold of no effect (i.e., SMD=0), suggesting both a possibility of a benefit or harm in the outcome.
g. Rated down three levels due to critical risk of bias because of lack of adjustment for important confounders in the two included studies (i.e., psychiatric interventions, mental health comorbidities, socioeconomic status, and family support) and serious risk of bias due to deviation of intended intervention in one of the included studies (28.26% of the participants in the no gender-affirming hormone therapy group were receiving puberty blockers or spironolactone as mono-therapy).
h. Statistically, there was moderate heterogeneity with $I2=63\%$ and $p=0.03$. However, we did not rate down for inconsistency as this can be explained because one of the two included studies measured the outcome only in natal female participants, while the other study measured in natal female and male participants.
i. Outcomes not measured: death by suicide, sexual dysfunction from a physiological perspective (i.e., lack of erection, dyspareunia, problems related to dry and degenerated mucosal tissue, anorgasmia), bone density, cardiovascular events.

References

1.AMV, Fontanari, F, Vilanova, MA, Schneider, I, Chinazzo, BM, Soll, K, Schwarz, MIR, Lobato, A, Brandelli, Costa. Gender affirmation is associated with transgender and gender nonbinary youth mental health improvement. LGBT Health; 2020.
2.H, Filipov, Y, Kavla, S, Şahin, ME, Gökler, Ş, Turan. The effects of gender-affirming hormone therapy on body satisfaction, self-esteem, quality of life, and psychopathology in people with female-to-male gender dysphoria. Transgender Health; 2023.
3.I, Becker-Hebly, S, Fahrenkrug, F, Campion, H, Richter-Appelt, M, Schulte-Markwort, C, Barkmann. Psychosocial health in adolescents and young adults with gender dysphoria before and after gender-affirming medical interventions: a descriptive study from the Hamburg Gender Identity Service European Child and Adolescent Psychiatry; 2021.
4.C, Grannis, WI, Mattson, SF, Leibowitz, L, Nahata, D, Chen, JF, Strang, H, Thobe, JA, Indyk, EE, Nelson. Expanding upon the relationship between gender-affirming hormone therapy, neural connectivity, mental health, and body image dissatisfaction. Psychoneuroendocrinology; 2023.

2. Before-after studies

See Table 3 for summary of findings table.

Gender dysphoria (GD): Meta-analysis suggested that GD, measured within the last 6 months with the Gender Preoccupation and Stability Questionnaire ranging from 14 to 70, was lower (SMD 0.26 lower [95% CI 1.64 lower to 1.13 higher], n = 2, very low certainty) in NFs after receiving GAHT compared to before, although we are very uncertain about the causal effect of the intervention on GD [21, 23].

Global function: Three studies reported global function using 3 different measures at 2 different timepoints.

Global function, measured within the last 6 months, may be higher (SMD 0.25 higher [95% CI 0.09 higher to 0.4 higher], n = 2, very low certainty) in NFs after receiving GAHT compared to before, although we are very uncertain about the causal effect of the intervention on global function [23, 27]. See Appendices 15 and 16 for very low certainty evidence about global function from studies not pooled with this evidence.

Depression: Four studies reported this outcome using 4 different scales. Due to variability in measurement instruments, timepoints, and reporting, we could not include all studies in a single meta-analysis.

A meta-analysis suggested that depression, measured within 18 to 24 months, may be lower (SMD 0.41 lower [95% CI 0.65 lower to 0.17 lower], n = 2, very low certainty) in NMs and NFs after receiving GAHT compared to before, although we are very uncertain about the causal effect of the intervention on depression [20, 22]. See

Appendices 15 and 16 for very low certainty evidence about depression from studies not pooled with this evidence.

Sexual dysfunction: A study reported a linear regression analysis with no statistically significant change in sexual dysfunction (i.e., vagina dryness or itch) reported by the NFs after 6 months of receiving GAHT (b = -0.01, 95% CI -0.09, 0.8) and after 12 months of receiving GAHT (b= 0.053, 95% CI: -0.03, 0.13) compared to before the intervention. This evidence was rated as low certainty; therefore, we are very uncertain about the causal effect of the intervention on sexual dysfunction [28].

Bone mineral density (BMD): Six studies reported lumbar spine BMD, 3 studies reported femoral neck BMD, and 3 studies reported hip BMD using z-scores and g/cm^2.

Lumbar spine BMD, measured within the last 12 to 36 months with g/cm^2 may be higher (0.01 higher [95% CI 0 higher to 0.01 higher], n=2, very low certainty) in NFs receiving GAHT compared to before, although we are very uncertain about the causal effect of the intervention on lumbar spine BMD [29, 32].

Femoral neck BMD, measured within the last 12 months assessed with the DXA, z-scores ranging from -3 to 3, may not change (MC 0 [95% CI 0.01 lower to 0], n=1, very low certainty) in NFs after receiving GAHT compared to before, although we are very uncertain about the causal effect of the intervention on femoral neck BMD [32].

Hip BMD, measured within the last 12 to 36 months with g/cm^2 was higher (0.01 higher [95% CI 0.01 higher to 0.01 higher], n=1, very low certainty) in NFs receiving GAHT compared to before, although we are very uncertain about the causal effect of the intervention on hip BMD [32].

See Appendices 15 and 16 for very low certainty evidence about BMD from studies not pooled with this evidence.

Table 3. Gender-affirming hormone therapy vs no gender-affirming hormone therapy: evidence from before-after studies.

Outcomes	**Anticipated absolute effects*** (95% CI)		№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk with no gender-affirming hormone therapy	**Risk with gender-affirming hormone therapy**			
Gender Dysphoria, Short Term Follow-up assessed with: participant reported, Gender Preoccupation and Stability Questionnaire, higher scores indicate higher levels of gender dysphoria Scale from: 14 to 70 follow-up: mean 6 months[a]	-	standardized mean change **0.26 lower** (1.64 lower to 1.13 higher)	36 (1 non-randomised study)[1,2]	⨁◯◯◯ Very low[b,c]	The evidence is very uncertain about the effect of gender-affirming hormone therapy on gender dysphoria at short term follow-up in natal females.

Global Function, Short Term Follow-up assessed with: participant reported, various scales [RAND Short Form-36 (SF-36) Health Survey, Symptom Checklist-90 Revised (SCL-90-R) Global Severity Index], higher scores indicate better global function Scale from: 0 to 100 follow-up: mean 6 months[a]	-	standardized mean change **0.25 higher** (0.09 higher to 0.4 higher)	73 (2 non-randomised studies)[1,3]	⨁◯◯◯ Very low[c,d,e,f]	The evidence is very uncertain about the effect of gender-affirming hormone therapy on global function at short term follow-up in natal females.
Depression, Long Term Follow-up assessed with: participant reported, various scales [Beck Depression Inventory, Hospital Anxiety and Depression Scale (HADS)], higher scores indicate worse depression follow-up: range 18 months to 24 months[g]	-	standardized mean change **0.41 lower** (0.65 lower to 0.17 lower)	389 (2 non-randomised studies)[4,5]	⨁◯◯◯ Very low[h,i]	The evidence is very uncertain about the effect of gender-affirming hormone therapy on depression at long term follow-up in natal males and females.
Sexual Dysfunction (i.e., Vaginal Dryness or Itch), Long Term Follow-Up assessed with: participant report of symptoms follow-up: mean 12 months[g]	In 193 participants, a linear regression analysis showed that there was no change from baseline in symptoms of vaginal dryness or itch after receiving GAHT (b= 0.053, 95% CI: -0.03, 0.13). [j,k]		193 (1 non-randomised study)[6]	⨁◯◯◯ Very low[l]	The evidence is very uncertain about the effect of gender-affirming hormone therapy on sexual dysfunction (i.e., vaginal dryness or itch) at long term follow-up in natal females.
Sexual Dysfunction (i.e., Vaginal Dryness or Itch), Short Term Follow-Up assessed with: participant report of symptoms follow-up: mean 6 months[a]	In 193 participants (i.e., natal females), a linear regression analysis showed that there was no change from baseline in symptoms of vaginal dryness or itch after receiving GAHT (b= -0.01, 95% CI: -0.09, 0.8). [j,k]		193 (1 non-randomised study)[6]	⨁◯◯◯ Very low[l]	The evidence is very uncertain about the effect of gender-affirming hormone therapy on sexual dysfunction (i.e., vaginal dryness or itch) at short term follow-up in natal females.

Bone Mineral Density - Femoral Neck, Long Term Follow-up assessed with: Dual-energy x-ray absorptiomety (DXA), z-scores Scale from: -3 to 3 follow-up: mean 12 months[g]	The mean bone Mineral Density - Femoral Neck, Long Term Follow-up was **0.84**	mean change **0** (0.01 lower to 0)	199 (1 non-randomised study)[7]	⨁◯◯◯ Very low[m]	The evidence is very uncertain about the effect of gender-affirming hormone therapy on bone mineral density - femoral neck at long term follow-up in natal females.
Bone Mineral Density - Hip, Long Term Follow-up assessed with: Dual-energy x-ray absorption (DXA), g/cm2 follow-up: range 12 months to 36 months[g]	The mean bone Mineral Density - Hip, Long Term Follow-up was **0.95**	mean change **0.01 higher** (0.01 higher to 0.01 higher)	199 (1 non-randomised study)[7]	⨁◯◯◯ Very low[m]	Gender-affirming hormone therapy may increase bone mineral density - hip, at long term follow-up slightly in natal females.
Bone Mineral Density - Lumbar Spine, Long Term Follow-up assessed with: Dual-energy x-ray absorption (DXA), g/cm2 follow-up: range 12 months to 36 months[g]	The mean bone Mineral Density - Lumbar Spine, Long Term Follow-up was **1.04**	mean change **0.01 higher** (0 to 0.01 higher)	234 (2 non-randomised studies)[7,8]	⨁◯◯◯ Very low[m]	Gender-affirming hormone therapy may increase bone mineral density - lumbar spine, at long term follow-up slightly in natal females.
Other Outcomes - not measured[n]	-	-	-	-	-

*__The risk in the intervention group__ (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: confidence interval; **MD:** mean difference; **OR:** odds ratio; **SMD:** standardised mean difference

GRADE Working Group grades of evidence
High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.
Moderate certainty: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
Low certainty: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.
Very low certainty: we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

Explanations

a. Short Term Follow-Up: outcome measured at ≤ 6 months follow-up.
b. Rated down three levels due to risk of bias stemming from prognostic imbalance associated with the observational study design and critical risk of bias due to missing data (i.e., 46.75% provided outcome data).
c. Rated down one level for imprecision because the optimal information size (OIS=200) was not met. Low sample size importantly increases the risk of random error.
d. Rated down two levels due to risk of bias stemming from prognostic imbalance associated with the observational study design and critical risk of bias due to missing data in one of the two included studies (i.e., 46.75% provided outcome data).
e. Statistically, there was considerable heterogeneity with I2=94% and p<0.01. However, we did not rate down for inconsistency as the overall effect estimate was not importantly affected by the studies contributing to statistical heterogeneity.
f. Rated down one level for indirectness because one of the two included studies reports the outcome only for natal females.
g. Long Term Follow-Up: outcome measured at ≥ 12 months follow-up.
h. Rated down three levels due to risk of bias stemming from prognostic imbalance associated with the observational study design, as well as critical and serious risk of bias due to missing data in the two included studies (i.e., 20% and 69% of participants, respectively, provided outcome data).
i. Statistically, there was considerable heterogeneity with I2=100% and p<0.01. However, we did not rate down for inconsistency as the overall effect estimate was not importantly affected by the studies contributing to statistical heterogeneity.
j. GAHT: gender-affirming hormone therapy.
k. In the linear mixed model, time was as added categorical variable to detect changes in symptom scores between 0-3 months, 0-6 months, and 0-12 months of GAHT. Differences in changes in symptom scores between different administration forms were corrected for baseline differences to avoid regression to the mean. An increase or decrease in symptom scores of 0.2 was considered clinically relevant.
l. Rated down three levels due to risk of bias stemming from prognostic imbalance associated with the observational study design and critical risk of bias due to concerns with measurement of the outcome (i.e., subjective and self-reported outcome).
m. Rated down three levels due to risk of bias stemming from prognostic imbalance associated with the observational study design and critical risk of bias due to missing data (i.e., 48% of participants provided outcome data).
n. Other outcomes: gender dysphoria, sexual dysfunction from physiological perspective (i.e., lack of erection, dyspareunia, anorgasmia), cardiovascular events.

References

1.L, Foster, Skewis, I, Bretherton, SY, Leemaqz, JD, Zajac, AS, Cheung. Short-term effects of gender-affirming hormone therapy on dysphoria and quality of life in transgender individuals: a prospective controlled study. Frontiers in Endocrinology; 2021.
2.M, Ascha, DC, Sasson, R, Sood, JW, Cornelius, JM, Schauer, A, Runge, AL, Muldoon, N, Gangopadhyay, L, Simons, D, Chen, JF, Corcoran, SW, Jordan. Top surgery and chest dysphoria among transmasculine and nonbinary adolescents and young adults. JAMA pediatrics; 2022.
3.Ş, Turan, C, Aksoy,Poyraz, NG, Usta,Sağlam, ÖF, Demirel, Ö, Haliloğlu, P, Kadıoğlu, A, Duran. Alterations in body uneasiness, eating attitudes, and psychopathology before and after cross-sex hormonal treatment in patients with female-to-male gender dysphoria. Archives of Sexual Behavior; 2018.
4.D, Chen, J, Berona, YM, Chan, D, Ehrensaft, R, Garofalo, MA, Hidalgo, SM, Rosenthal, AC, Tishelman, J, Olson-Kennedy. Psychosocial functioning in transgender youth after 2 years of hormones. The New England Journal of Medicine; 2023.
5.Z, Aldridge, S, Patel, B, Guo, E, Nixon, W, Pierre, Bouman, GL, Witcomb, J, Arcelus. Long-term effect of gender-affirming hormone treatment on depression and anxiety symptoms in transgender people: A prospective cohort study. Andrology; 2021.
6.D, van,Dijk, MJHJ, Dekker, EB, Conemans, CM, Wiepjes, EGM, de,Goeij, KA, Overbeek, AD, Fisher, M, den,Heijer, G, T'Sjoen. Explorative prospective evaluation of short-term subjective effects of hormonal treatment in trans people-results from the European network for the investigation of gender incongruence. The Journal of Sexual Medicine; 2019.
7.CM, Wiepjes, MC, Vlot, M, Klaver, NM, Nota, CJ, de,Blok, RT, de,Jongh, P, Lips, AC, Heijboer, AD, Fisher, T, Schreiner, G, T'Sjoen, M, den,Heijer. Bone mineral density increases in trans persons after 1 year of hormonal treatment: a multicenter prospective observational study. Journal of Bone and Mineral Research : The Official Journal of the American; 2017.
8.P, van,Kesteren, P, Lips, W, Deville, C, Popp-Snijders, H, Asscheman, J, Megens, L, Gooren. The effect of one-year cross-sex hormonal treatment on bone metabolism and serum insulin-like growth factor-1 in transsexuals. The Journal of Clinical Endocrinology and Metabolism; 1996.

3. Case series

See Table 4 for summary of findings table. One of the before-after studies reported data about death by suicide only after the intervention and we classified it as case series for that outcome [22].

Death by suicide: As retrieved from medical records, death by suicide within 24 months of receiving GAHT occurred in 2 of 315 NMs and NFs (0.6%); proportion 0.006 (95% CI 0.001 to 0.018, n = 1, very low certainty). We are very uncertain about the effects of GAHT on death by suicide [22].

Cardiovascular events: As retrieved from medical records, cardiovascular events within 7 to 109 months of receiving GAHT occurred in 151 of 3875 NFs (3.9%); proportion 0.04 (95% CI 0.03 to 0.05, n = 1, high certainty) [33].

Cardiovascular events: As retrieved from medical records, cardiovascular events within 26 months of receiving GAHT occurred in 3 of 1893 NFs (0.2%); proportion 0.00 (95% CI 0.00 to 0.01, n = 1, moderate certainty) [34].

Table 4. Gender-affirming hormone therapy vs no gender-affirming hormone therapy: evidence from case series.

Outcomes	Anticipated absolute effects* (95% CI)				Comments

	Risk with no gender-affirming hormone therapy	Risk with gender-affirming hormone therapy	Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	
Death by Suicide, Long Term Follow-up assessed with: medical records follow-up: mean 24 months [a]	No comparison group available	**6 per 1,000** (1 to 18)	**proportion 0.006** (0.001 to 0.018)	315 (1 non-randomised study)[1]	⨁◯◯◯ Very low [b]	The evidence is very uncertain about the effect of gender-affirming hormone therapy on death by suicide at long term follow-up in natal males and females.
Cardiovascular Events, Long Term Follow-Up assessed with: medical records, number of events follow-up: range 7 months to 109 months [a, c]	No comparison group available	**40 per 1,000** (30 to 50)	**proportion 0.04** (0.03 to 0.05)	3875 (1 non-randomised study)[2]	⨁⨁⨁⨁ High[f]	The proportion of natal females experiencing cardiovascular events at long term follow-up is 40 per 1,000.
Cardiovascular Events, Long Term Follow-Up assessed with: medical records, number of participants with an event follow-up: mean 26 months [a, d]	No comparison group available	**0 per 1,000** (0 to 10)	**proportion 0.00** (0.00 to 0.01)	1893 (1 non-randomised study)[3]	⨁⨁⨁◯ Moderate [ef]	The proportion of natal females experiencing cardiovascular events at long term follow-up is 1 per 1,000.
Other Outcomes - not measured [g]	-	-	-	-	-	-

*__The risk in the intervention group__ (and its 95% confidence interval) is based on the assumed risk in the comparison group and the __relative effect__ of the intervention (and its 95% CI).

CI: confidence interval

GRADE Working Group grades of evidence
High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.
Moderate certainty: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
Low certainty: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.
Very low certainty: we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

Explanations
a. Long Term Follow-Up: outcome measured at ≥ 12 months follow-up. b. Rated down three levels for risk of bias due to lack of a comparison group. c. Cardiovascular events include: stroke, myocardial infarction, and venous thromboembolism. d. Cardiovascular events include: thromboembolism. e. Rated down one level for indirectness because this study included natal males only. f. We did not rate down for risk of bias because this outcome does not need a comparison group, as the study participants can only experience this outcome if they have received the intervention. g. Other outcomes not measured: gender dysphoria, global function, depression, sexual dysfunction from physiological perspective (i.e., lack of erection, dyspareunia, problems related to dry and degenerated mucosal tissue, anorgasmia), bone mineral density.
References
1.D, Chen, J, Berona, YM, Chan, D, Ehrensaft, R, Garofalo, MA, Hidalgo, SM, Rosenthal, AC, Tishelman, J, Olson-Kennedy. Psychosocial functioning in transgender youth after 2 years of hormones. The New England Journal of Medicine; 2023. 2.NM, Nota, CM, Wiepjes, CJM, de,Blok, LJG, Gooren, BPC, Kreukels, M, den,Heijer. Occurrence of acute cardiovascular events in transgender individuals receiving hormone therapy. Circulation; 2019. 3.M, Pyra, I, Casimiro, L, Rusie, N, Ross, C, Blum, K, Keglovitz, Baker, A, Baker, J, Schneider. An observational study of hypertension and thromboembolism among transgender patients using gender-affirming hormone therapy. Transgender Health; 2020.

DISCUSSION

This SR and meta-analysis synthesized the available evidence regarding the effects of GAHT in young individuals with GD. Comparative observational studies provided mostly very low certainty evidence for GD, global function, and depression. One study provided low certainty evidence that depression may be lower in NMs and NFs who received GAHT compared to those who did not. Before-after studies provided very low certainty evidence. Case series provided very low certainty evidence on death by suicide and high to moderate certainty evidence for cardiovascular events.

Although some may view our modification of the ROBINS-I tool as a limitation, we strongly believe it produced similar conclusions than if we had used the original tool

or alternatives, such as the Newcastle-Ottawa scale [37]. Given the widespread methodological limitations in this field, any risk of bias tool would yield similar conclusions. Comparative observational and before-after studies were at serious or critical risk of bias due to missing data and deviation from intended intervention (i.e., administration of co-interventions). Case series, which lack a comparison group, were at critical risk of bias due to measurement of the outcome. These studies should only be used to generate hypothesis for more rigorous study designs, such as prospective cohorts.

The target question of this SR and of the decision-makers considering these interventions is: what are the effects of GAHT? In the absence of randomized controlled trials or comparative observational studies, case series and before-after studies provide the best available evidence to answer this question. While these study designs answer single-group questions (e.g., what is the functional status among people who received GAHT), they are limited in assessing intervention effects (e.g., whether functional status is better in people who received GAHT than those who did not). We accounted for these limitations, and assessed the certainty of the available evidence following current methodological standards [38].

We rated down the certainty of the evidence mostly because of risk of bias and imprecision, often resulted from an insufficient sample size, for most outcomes and study designs. We did not find evidence about sexual dysfunction in NMs.

The overarching theme from this and other SRs on GAHT is the lack of high-quality evidence for individuals with GD. Unlike this SR, other reviews did not assess the certainty of evidence for each outcome.

Taylor et al included studies with individuals 18 years and below, rating most studies as low to moderate quality using the Newcastle-Ottawa Scale. They found limited evidence on GD, body satisfaction, psychological and cognitive outcomes, and infertility [39]. Doyle et al reported on psychosocial functioning changes after GAHT among transgender individuals of all ages. They concluded that risk of bias, assessed with the Newcastle-Ottawa Scale, varied among studies. Small sample sizes and unadjusted confounders limited the ability to draw causal inferences [40].

Van Leerdam et al concluded that GAHT may reduce GD, body dissatisfaction, and uneasiness, subsequently improving psychological well-being and quality of life in transgender individuals of all ages [41]. They rated the evidence as low to moderate in quality, based on longitudinal cohort and cross-sectional studies, without clarifying their rating methods. Chew et al suggested that GAHT helps adolescents achieve intended physical effects, with limited evidence on its psychosocial and cognitive impact [42]. Further, a SR by Connelly et al concluded that current data are insufficient to determine GAHT's impact on blood pressure in transgender individuals [43].

Across all these SRs, the findings highlight methodological limitations, low-quality evidence, and important gaps in evidence.

The evidence about the effects of GAHT in individuals under the age of 26 experiencing GD is predominantly very low certainty, with lack of moderate and high certainty evidence about the effects of this intervention. This information is crucial for patients, caregivers, clinicians, guideline developers, and policymakers involved in treatment decisions. Beyond evidence certainty, decision-making should consider other factors, including the magnitude and consequences of potential benefits and harms, patients' and caregivers' values and preferences, resource use, feasibility, acceptability, and equity [44]. Guideline developers and policy makers must transparently state which and whose values they prioritize when developing treatment recommendations and policies.

Strengths and limitations of the review process

This SR and meta-analysis has multiple strengths. We rigorously followed the highest methodological standards, we assessed the risk of bias for each study using the ROBINS-I tool, and evaluated the certainty of the evidence for each outcome using the GRADE approach. A limitation of our review is the inclusion of only English-language studies, though we do not expect this to fundamentally alter our conclusions. Due to feasibility considerations, we prioritized specific outcomes and could not address others that may be important to readers, such as regret, anxiety, pelvic pain, or cancers (e.g., breast, gynecological, prostate, and colon cancer).

CONCLUSION

The best available evidence reporting on the effects of GAHT in individuals experiencing GD ranged from moderate to high certainty for cardiovascular events in case series studies, and mostly very low certainty for the outcomes of GD, global function, depression, sexual dysfunction, BMD, and death by suicide in comparative observational, before-after and case series studies. We did not find evidence on NM sexual dysfunction. The evidence found in this SR and meta-analysis does not exclude the possibility of benefit or harm upon receipt of GAHT. Prospective studies yielding higher certainty evidence are needed to understand the short- and long-term effects of GAHT.

Supplemental material

A supplemental appendix to this article is available online.

Conflicts of interest statement

Dr. Romina Brignardello-Petersen and Anna Miroshnychenko provided methodological expertise for the SEGM initiative to summarize and appraise the quality of publications related to gender medicine for the SEGM online platform, and for this work they received financial compensation from SEGM. This work was completely independent from the systematic review and meta-analysis.
Indirect financial conflicts of interest:
Manuscript authors do not have indirect financial conflicts of interest.

2. Financial conflicts of interest (as reported by the protocol authors who were not part of the evidence synthesis team at the time of their participation in the generation of the question):

Direct financial conflicts of interest:

E. Abbruzzese is a contributing author for the Society for Evidence-based Gender Medicine online platform and received financial compensation from SEGM.

Indirect financial conflicts of interest:

Dr. William Malone's fee for publishing a research article as “open access” was compensated by SEGM.

3. Other disclosures (manuscript authors):

Manuscript authors do not have other disclosures.

3. Other disclosures (as reported by the protocol authors who were not part of the evidence synthesis team at the time of their participation in the generation of the question):

Affiliations:

Dr. William Malone is a board member of SEGM.

Expressed opinions:

Dr. William Malone has expressed opinions about gender affirmation interventions for adolescents and young adults in The Journal of Clinical Endocrinology and Metabolism, The Lancet, Child and Adolescent Health, and Medscape.

Funding statement

The Society for Evidence-Based Gender Medicine (SEGM), the sponsor, commissioned this work from McMaster University. This systematic review is part of a large research project funded through a research agreement between the Society for Evidence-based Gender Medicine (SEGM), the sponsor, and McMaster University. None of the team members received financial compensation directly from SEGM to conduct this systematic review and meta-analysis.

Management of conflicts of interest

This SR is part of a research project funded through a research agreement between the Society for Evidence-based Gender Medicine (SEGM), the sponsor, and McMaster University. None of the researchers conducting this SR and meta-analysis received financial compensation from the sponsor to conduct this work. The SR and meta-analysis research question was designed through a collaboration between SEGM members, and the methods team based at McMaster University. The rest of the SR and meta-analysis processes (i.e., search and study selection, data extraction, data analyses, manuscript writing, approval of final draft of manuscript) were conducted by independent researchers who do not have any financial or intellectual conflicts of interest or disclosures, and the methods team based at McMaster University (the group of authors). The methods team was solely responsible for the synthesis and interpretation of results, and for drawing conclusions. To minimize bias, a methodologist who was not involved in the data collection, synthesis, and interpretation (GG) ensured that results interpretation was consistent with the findings. This

manuscript was drafted by the methods team and approved by all authors, and the sponsor did not have any say nor reviewed its content.

Ethics approval statement

The ethics approval was not required.

REFERENCES

1. American Psychiatric Association, *The Diagnostic and Statistical Manual of Mental Disorders (5th ed DSM-5)*, ed. A. American Psychiatric and D.S.M.T.F. American Psychiatric Association. 2013, Arlington, VA: American Psychiatric Association.
2. Trans Care BC. 2024; Available from: https://www.transcarebc.ca/.
3. Hembree, W.C., *Guidelines for pubertal suspension and gender reassignment for transgender adolescents.* Child Adolesc Psychiatr Clin N Am, 2011. **20**(4): p. 725-32.
4. Biggs, M., *The Dutch Protocol for Juvenile Transsexuals: Origins and Evidence.* Journal of Sex & Marital Therapy, 2023. **49**(4): p. 348-368.
5. Cohen-Kettenis, P.T. and F. Pfäfflin, *The DSM diagnostic criteria for gender identity disorder in adolescents and adults.* Arch Sex Behav, 2010. **39**(2): p. 499-513.
6. Mercy, J.A. and A.M. Vivolo-Kantor, *The Center for Disease Control and Prevention's (CDC) Youth Violence Prevention Centers: Paving the Way to Prevention.* J Prim Prev, 2016. **37**(2): p. 209-14.
7. United Nations. *Youth*. Available from: https://www.un.org/en/global-issues/youth.
8. Miroshnychenko A, et al., *Puberty blockers for individuals with gender dysphoria below 26 years of age: A systematic review and meta-analysis.* Submitted for publication, 2024.
9. Guyatt, G.H., et al., *GRADE: an emerging consensus on rating quality of evidence and strength of recommendations.* BMJ, 2008. **336**(7650): p. 924-926.
10. Miroshnychenko, A., et al., *"Mastectomy for individuals with gender dysphoria below 26 years of age: A systematic review and meta-analysis".* Plast Reconstr Surg, 2024.
11. Achille, C., et al., *Longitudinal impact of gender-affirming endocrine intervention on the mental health and well-being of transgender youths:*

preliminary results. International Journal of Pediatric Endocrinology, 2020. **2020**: p. 8.
12. Becker-Hebly, I., et al., *Psychosocial health in adolescents and young adults with gender dysphoria before and after gender-affirming medical interventions: a descriptive study from the Hamburg Gender Identity Service.* European Child & Adolescent Psychiatry, 2021. **30**(11): p. 1755-1767.
13. Cantu, A.L., et al., *Changes in Anxiety and Depression from Intake to First Follow-Up Among Transgender Youth in a Pediatric Endocrinology Clinic.* Transgender Health, 2020. **5**(3): p. 196-200.
14. Filipov, H., et al., *The Effects of Gender-Affirming Hormone Therapy on Body Satisfaction, Self-Esteem, Quality of Life, and Psychopathology in People with Female-to-Male Gender Dysphoria.* Transgender Health, 2023. **8**(2): p. 168-174.
15. Fontanari, A.M.V., et al., *Gender Affirmation Is Associated with Transgender and Gender Nonbinary Youth Mental Health Improvement.* Lgbt Health, 2020. **7**(5): p. 237-247.
16. Grannis, C., et al., *Expanding upon the relationship between gender-affirming hormone therapy, neural connectivity, mental health, and body image dissatisfaction.* Psychoneuroendocrinology, 2023. **156**: p. 106319.
17. Green, A.E., et al., *Association of Gender-Affirming Hormone Therapy With Depression, Thoughts of Suicide, and Attempted Suicide Among Transgender and Nonbinary Youth.* Journal of Adolescent Health, 2022. **70**(4): p. 643-649.
18. Olsavsky, A.L., et al., *Associations Among Gender-Affirming Hormonal Interventions, Social Support, and Transgender Adolescents' Mental Health.* Journal of Adolescent Health, 2023. **72**(6): p. 860-868.
19. Tordoff, D.M., et al., *Mental Health Outcomes in Transgender and Nonbinary Youths Receiving Gender-Affirming Care.* JAMA Network Open, 2022. **5**(2): p. e220978.
20. Aldridge, Z., et al., *Long-term effect of gender-affirming hormone treatment on depression and anxiety symptoms in transgender people: A prospective cohort study.* Andrology, 2021. **9**(6): p. 1808-1816.
21. Ascha, M., et al., *Top Surgery and Chest Dysphoria Among Transmasculine and Nonbinary Adolescents and Young Adults.* JAMA Pediatr, 2022. **176**(11): p. 1115-1122.
22. Chen, D., et al., *Psychosocial Functioning in Transgender Youth after 2 Years of Hormones.* N Engl J Med, 2023. **388**(3): p. 240-250.
23. Foster Skewis, L., et al., *Short-Term Effects of Gender-Affirming Hormone Therapy on Dysphoria and Quality of Life in Transgender Individuals: A Prospective Controlled Study.* Front Endocrinol (Lausanne), 2021. **12**: p. 717766.
24. Haraldsen, I.R., et al., *Cross-sex pattern of bone mineral density in early onset gender identity disorder.* Horm Behav, 2007. **52**(3): p. 334-43.
25. Kaltiala, R., et al., *Adolescent development and psychosocial functioning after starting cross-sex hormones for gender dysphoria.* Nord J Psychiatry, 2020. **74**(3): p. 213-219.

26. Schagen, S.E.E., et al., *Bone Development in Transgender Adolescents Treated With GnRH Analogues and Subsequent Gender-Affirming Hormones.* J Clin Endocrinol Metab, 2020. **105**(12): p. e4252-63.
27. Turan, Ş., et al., *Alterations in Body Uneasiness, Eating Attitudes, and Psychopathology Before and After Cross-Sex Hormonal Treatment in Patients with Female-to-Male Gender Dysphoria.* Arch Sex Behav, 2018. **47**(8): p. 2349-2361.
28. van Dijk, D., et al., *Explorative Prospective Evaluation of Short-Term Subjective Effects of Hormonal Treatment in Trans People-Results from the European Network for the Investigation of Gender Incongruence.* J Sex Med, 2019. **16**(8): p. 1297-1309.
29. van Kesteren, P., et al., *The effect of one-year cross-sex hormonal treatment on bone metabolism and serum insulin-like growth factor-1 in transsexuals.* J Clin Endocrinol Metab, 1996. **81**(6): p. 2227-32.
30. Vlot, M.C., et al., *Gender-Affirming Hormone Treatment Decreases Bone Turnover in Transwomen and Older Transmen.* J Bone Miner Res, 2019. **34**(10): p. 1862-1872.
31. Wiepjes, C.M., et al., *Bone Safety During the First Ten Years of Gender-Affirming Hormonal Treatment in Transwomen and Transmen.* J Bone Miner Res, 2019. **34**(3): p. 447-454.
32. Wiepjes, C.M., et al., *Bone Mineral Density Increases in Trans Persons After 1 Year of Hormonal Treatment: A Multicenter Prospective Observational Study.* Journal of Bone and Mineral Research, 2017. **32**(6): p. 1252-1260.
33. Nota, N.M., et al., *Occurrence of Acute Cardiovascular Events in Transgender Individuals Receiving Hormone Therapy.* Circulation, 2019. **139**(11): p. 1461-1462.
34. Pyra, M., et al., *An Observational Study of Hypertension and Thromboembolism Among Transgender Patients Using Gender-Affirming Hormone Therapy.* Transgend Health, 2020. **5**(1): p. 1-9.
35. Becker-Hebly, I., et al., *Psychosocial health in adolescents and young adults with gender dysphoria before and after gender-affirming medical interventions: a descriptive study from the Hamburg Gender Identity Service.* Eur Child Adolesc Psychiatry, 2021. **30**(11): p. 1755-1767.
36. Filipov, H., et al., *The Effects of Gender-Affirming Hormone Therapy on Body Satisfaction, Self-Esteem, Quality of Life, and Psychopathology in People with Female-to-Male Gender Dysphoria.* Transgend Health, 2023. **8**(2): p. 168-174.
37. Wells, G.A.W., G.; Shea, B.; Shea, B.; O'Connell, D.; Peterson, J.; Welch, Losos, M.; Tugwell, P.; Ga, S.W.; Zello, G.A.; & Petersen, J.A.;, *The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomized studies in meta-analyses. Ottawa Hospital Research Institute.* 2014.
38. Brignardello-Petersen, R. and G.H. Guyatt, *Assessing the Certainty of the Evidence in Systematic Reviews: Importance, Process, and Use.* Am J Epidemiol, 2024.

39. Taylor, J., et al., *Masculinising and feminising hormone interventions for adolescents experiencing gender dysphoria or incongruence: a systematic review.* Arch Dis Child, 2024.
40. Doyle, D.M., T.O.G. Lewis, and M. Barreto, *A systematic review of psychosocial functioning changes after gender-affirming hormone therapy among transgender people.* Nature Human Behaviour, 2023. **7**(8): p. 1320-1331.
41. van Leerdam, T.R., J.D. Zajac, and A.S. Cheung, *The Effect of Gender-Affirming Hormones on Gender Dysphoria, Quality of Life, and Psychological Functioning in Transgender Individuals: A Systematic Review.* Transgend Health, 2023. **8**(1): p. 6-21.
42. Chew, D., et al., *Hormonal Treatment in Young People With Gender Dysphoria: A Systematic Review.* Pediatrics, 2018. **141**.
43. Connelly, P.J., et al., *Transgender adults, gender-affirming hormone therapy and blood pressure: a systematic review.* Journal of Hypertension, 2021. **39**(2).
44. Schünemann, H., Brożek J, Guyatt G, Oxman A, editors., *GRADE handbook for grading quality of evidence and strength of recommendations*. 2013: The GRADE Working Group.

CHAPTER 5: MASTECTOMY FOR INDIVIDUALS WITH GENDER DYSPHORIA BELOW 26 YEARS OF AGE: A SYSTEMATIC REVIEW AND META-ANALYSIS

Anna Miroshnychenko, Yetiani M Roldan, Sara Ibrahim, Chan Kulatunga-Moruzi, Kristen Dahlin, Steven Montante, Rachel Couban, Gordon Guyatt, Romina Brignardello-Petersen

Published: Journal of Plastic Surgery [September 10, 2024]

ABSTRACT

Background: Gender dysphoria (GD) refers to psychological distress associated with the incongruence between one's sex and one's gender. In response to GD, birth-registered females may choose to undergo mastectomy. In this systematic review, we summarize and assess the certainty of the evidence on the effects of mastectomy.

Methods: We searched MEDLINE, Embase, PsycINFO, Social Sciences Abstracts, LGBTQ+ Source, and Sociological Abstracts through June 20, 2023. We included studies comparing mastectomy to no mastectomy in birth-registered females under 26 years of age with GD. Outcomes of interest included psychological and psychiatric outcomes, and physical complications. Pairs of reviewers independently screened articles, abstracted data, and assessed risk of bias of the included studies. We performed meta-analysis and assessed the certainty of the evidence using the GRADE approach.

Results: We included 39 studies. Observational studies (n=2) comparing mastectomy to chest binding provided very low certainty evidence for the outcome of GD. One observational study comparing mastectomy to no mastectomy provided very low certainty evidence for the outcomes global functioning and suicide attempts, and low certainty evidence for the outcome non-suicidal self-injury (aOR 0.47 [95% CI 0.22 to 0.97]). Before-after (n=2) studies provided very low certainty evidence for all outcomes. Evidence from case series (n=34) studies ranged from high to very low certainty.

Conclusion: *Case series studies demonstrated high certainty evidence for the outcomes of death, necrosis, and excessive scarring; however, these studies are limited in methodological quality. In comparative and before-after studies the evidence ranged from low to very low certainty, and therefore does not exclude the possibility of benefit or harm.*

INTRODUCTION

The concept of gender dysphoria (GD), introduced in *The Diagnostic and Statistical Manual of Mental Disorders Fifth Edition (DSM-5),* denotes psychological distress arising from incongruence between one's sex and one's gender identity [1]. A similar condition of "Gender Incongruence" exists in the World Health Organization's International Classification of Diseases, Eleventh Edition (ICD-11), referring to discordant birth and experienced gender, with or without distress. GD includes "Gender Dysphoria in Childhood" and "Gender Dysphoria in Adolescents and Adults". In both, the incongruence must persist for at least six months causing distress or impairment in social, school, or occupational functioning, and at least two symptoms [1].

In response to GD, birth-registered females may opt for mastectomy. The primary surgical goal is to achieve a masculine aesthetic chest. Although additional surgical modalities are incorporated for chest masculinization, subcutaneous mastectomy is considered paramount for both optimal aesthetic outcomes and the comprehensive surgical treatment of GD [2].

Some studies suggest improvement in GD, psychological symptoms, quality of life, and sexual function by combining hormonal and surgical interventions [2, 3]. Studies examining transgender male chest reconstruction include surgeon-reported outcomes (e.g., surgical complications), patient-reported outcomes (e.g., psychological well-being), and preoperative breast morphology-based treatment approaches to minimize complications [2, 4-6].

As interest in gender-affirming surgery grows, an understanding of outcomes is essential for counseling patients seeking these treatments during the decision-making process to ensure informed consent. During these conversations, patients often consider suicidality, GD recurrence, other psychiatric diagnoses, post-operative complications, and satisfaction [7, 8], ideally evaluating these outcomes in the short- and long-term. Knowing the best available evidence about the effects of mastectomy in gender reassignment is essential to inform patients, develop practice guidelines, and identify areas for future research.

Currently, there is no systematic review (SR) and meta-analysis assessing the psychological and physical effects of mastectomy in individuals with GD that follows the latest methodological guidance for conducting such publications. The aim of this SR and meta-analysis was to summarize the evidence about the psychological and physical effects of mastectomy in individuals with GD under 26 years of age.

METHODS

We report this SR and meta-analysis following the guidance of the preferred reporting items for systematic reviews and meta-analyses (PRISMA) checklist [9] (Appendix, Supplemental Digital Content 1). We registered the protocol in PROSPERO (registration ID: CRD42022324741).

Eligibility criteria

The eligibility criteria are detailed in Table 1.

Table 1. Eligibility criteria.

Types of studies	We included randomized clinical trials, comparative observational studies, and before-after studies addressing the intervention and comparison of interest. We also included case series addressing the intervention of interest. We included peer reviewed publications in English language.
Population	We included individuals under 26 years, who were diagnosed with, experienced, self-identified, or were identified by a parent as having GD, gender identity disorder or gender incongruence, or who identified as transgender or non-binary. We included studies with participants of mean age below 26 years at the time of mastectomy. We selected this age threshold because individuals 25 years of age and below represent children and adolescents, and adolescents or youth are commonly thought to extend to the mid-twenties. For instance, the Centers for Disease Control and Prevention (CDC) and the United Nations (UN) regard youth as 14-24 years old. In order to be inclusive of individuals who are slightly above 25, approaching 26, we decided to include all individuals below 26 years of age. Based on clinical expertise, there is no important difference in surgical complications (i.e., physical complications) between individuals who are below 26

	years and those between 27 to 30 years. Therefore, we included studies with participants having a mean age below 30 years at the time of mastectomy, reporting on physical complications.
Intervention	We included studies assessing the effects of mastectomy (all types) and/or breast reduction (all types). We defined mastectomy as stated by authors or as a surgery to remove all breast tissue. We defined breast reduction as stated by authors or as removal of skin and tissue from the breast, followed by reshaping and elevating to create a smaller breast size.
Comparator	The comparator of interest was no mastectomy (e.g., psychological therapy, chest binding). We defined breast binding as stated by authors or as a process of compressing the breast tissue to achieve a flatter chest contour.
Outcomes	We included studies reporting on the following outcomes at any time point: quality of life, global functioning, depression, death by suicide, physical complications (i.e., death, necrosis, postsurgical persistent pain, persistent numbness, excessive scarring), shortness of breath, rib pain, and back pain, GD/incongruence, suicide attempts, non-suicidal self-harm, inpatient psychiatric hospitalizations, substance use/abuse changes or incidence, regret, and satisfaction with chest/body.

Information sources and search strategy

With support from an information specialist (RC), we searched in MEDLINE, Embase, PsycINFO, Social Sciences Abstracts, Contemporary Women's Issues, LGBTQ+ Source, Sociological Abstracts, Studies on Women, Gender Abstracts, and Google Scholar from inception to April 2023. Search strategies are included in Appendix, Supplemental Digital Content 2.

Study selection

Using Covidence software (https://www.covidence.org/), a pair of trained reviewers (SI, YR) independently screened titles and abstracts, and full texts of studies identified as eligible. A third reviewer (AM) resolved conflicts.

Data collection

For each eligible study, a pair of trained reviewers (SI, YR) independently extracted data using a standardized, piloted data extraction form. Reviewers collected information on study characteristics and participants, interventions, and outcomes of interest. Within a single study, if multiple time points were reported, we chose the longest. Reviewers discussed and resolved discrepancies, involving a third reviewer (AM) when necessary.

Risk of bias of individual studies

For each eligible study and outcome, a pair of trained reviewers (SI, YR) used a modified version of the Cochrane risk of bias tool for non-randomized studies of interventions (ROBINS-I) [10] and assessed studies as (a) low risk of bias; (b) moderate risk of bias; (c) high risk of bias; or (d) critical risk of bias, across several domains (Table, Supplemental Digital Content 3). For randomized clinical trials, we planned to use the revised Cochrane risk of bias tool for randomized trials [11]. Reviewers discussed and resolved discrepancies, involving a third reviewer (AM) when necessary.

Data synthesis

Although study authors used various observational study designs, we classified studies according to how their data was analyzed for this review. We considered studies to be comparative observational if they reported outcome data on an intervention group compared to an independent group, before-after design if the study presented outcome data collected prior to and after the intervention in the same group of people, and case series if it only provided data after the intervention in one group of people.

To summarize intervention effects, we used odds ratios for dichotomous outcomes in comparative observational and before-after studies, and proportions in case series. For continuous outcomes, we employed mean difference in comparative observational studies, mean change in before-after studies, and mean in case series. When calculating mean change, due to lack of data from study authors, we imputed a moderate correlation coefficient (r=0.5). We calculated 95% confidence intervals (CI) around all

estimates. When studies reported the same outcome using varied scales, we converted these data to the scale most frequently reported [12].

When clinically appropriate according to experts (KD, CMK, SM), we conducted meta-analyses of studies addressing the same outcome using a random-effects model and weighting the studies according to the inverse of their variance. We conducted these analyses using the *meta* [13] and *metafor* [14] packages in R Studio Version 4.2.

Certainty of the evidence

We were interested in addressing a causation question (i.e., what are the effects of mastectomy and/or breast reduction). We assessed the certainty of the evidence using the grading of recommendations assessment, development, and evaluation (GRADE) approach [14]. For each comparison and outcome, a pair of methodologists with experience in GRADE (SI, YR) rated each domain independently, resolving discrepancies by consulting a third methodologist (AM). We rated the certainty as high, moderate, low, or very low. All bodies of evidence started as high certainty [15], and could be rated down due to risk of bias, inconsistency, indirectness, imprecision, and publication bias; or rated up when a large magnitude of effect was found, a dose-response relationship, or when all plausible confounders or other biases increased our confidence in the estimated effect [16].

Even if the studies met the mean age eligibility criteria (below 26 years), when addressing indirectness, we rated down the certainty of the evidence if less than 60% of participants were below 26 years of age at the time of undergoing mastectomy or

breast reduction for all outcomes (except physical complications) and rated down the certainty of the evidence if less than 60% of participants were below 30 years of age for physical complications outcomes.

Following GRADE guidance when risk of bias at the study level is assessed using the ROBINS-I tool, in observational comparative studies, we rated down the certainty of the evidence up to three levels due to the risk of prognostic imbalance at the outcome level [15]. In case series, for outcomes needing a comparison group to assess the effect of mastectomy (i.e., quality of life, depression, death by suicide, body satisfaction, chest satisfaction), we rated down three levels due to the lack of a comparison group. For outcomes not needing a comparison group (i.e., death, necrosis, persistent numbness, scarring, postsurgical persistent pain, regret), we did not rate down for this reason.

We used a null effect threshold to rate the certainty of the benefit or harm of undergoing mastectomy over breast reduction or no mastectomy. We used GRADEpro [17] to create summary of findings tables.

Subgroup and sensitivity analyses

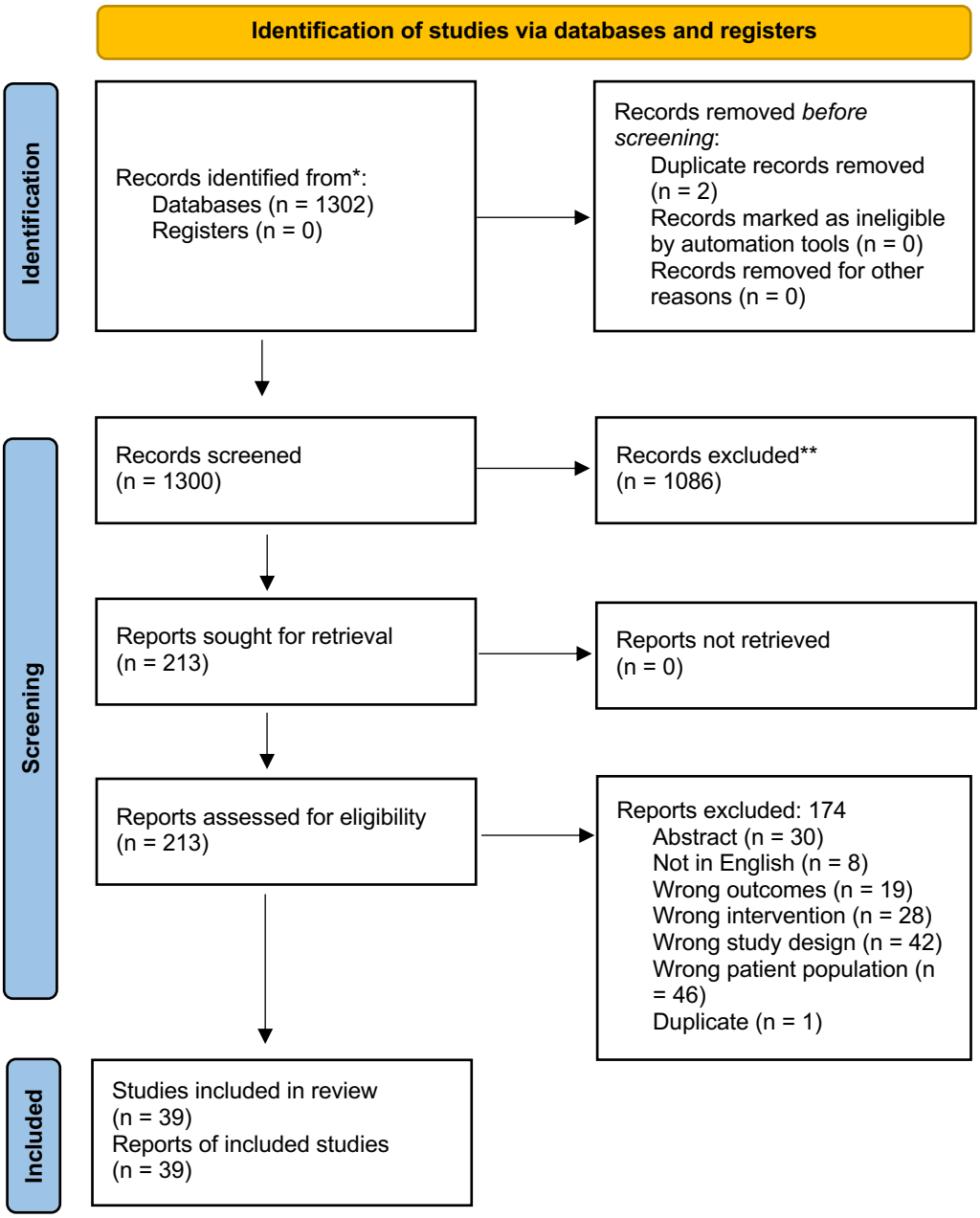
We planned the following subgroup analyses if at least two studies per subgroup were available: early onset of puberty vs late onset of puberty, childhood-onset GD vs pubertal/post-pubertal onset GD; recent-onset of GD vs several years long GD; psychiatric/neuropsychiatric diagnosis vs uncomplicated by psychiatric/neuropsychiatric diagnosis; clinical vs community populations; diagnosed

vs self-reporting of GD; transmasculine versus other identifications; gynephilic vs others (i.e., androphilic). We planned to use the Instrument for Credibility of Effect Modification Analyses (ICEMAN) to assess the credibility of subgroup effects [18]. Due to data availability, we could not perform subgroup analyses. To calculate mean change, we conducted sensitivity analyses by imputing correlation coefficients (r=0.2 and r=0.8) as study authors did not provide them.

RESULTS

After screening 1302 titles and abstracts, we included 39 studies. Figure 1 shows the results of the selection process. We present reasons for exclusion at the full text screening stage (n=138) in Figure 1 and Table, Supplemental Digital Content 4.

Figure 1. Study identification and selection flowchart.



Characteristics of included studies

Of 39 included studies, 3 were comparative observational, 2 were before-after, and 34 were case series (Figure 1) [19-57]. Due to the complexity and limitations on reporting, we could not classify comparative observational studies as cross-sectional or cohort.

The mean (SD) age of participants at the time of surgery ranged from 17.2 (1.54) to 29.1 (9.5), and the median (range) age ranged from 16 (12-17) to 28 (21-49). We present characteristics of included studies in Table, Supplemental Digital Content 5. Table, Supplemental Digital Content 6 describes outcome measurement scales and their interpretability.

Risk of bias in included studies

Across comparative observational studies, the domain most frequently judged as serious or critical risk of bias, was confounding. Before-after studies were at critical risk of bias due to confounding, representativeness of the sample (i.e., only a proportion of eligible participants were included) and missing outcome data. Case series studies were at critical risk of bias due to confounding, representativeness of the sample (i.e., only a proportion of eligible participants were included), missing outcome data, and measurement of outcome (Appendix, Supplemental Digital Content 7).

Effects of interventions

We describe the effects of the interventions according to study design and comparisons. Only one study presented data at multiple time points, the effect estimates were not considered to be importantly different between these time points [34]. Appendix, Supplemental Digital Content 8 displays forest plots of meta-analyses.

1. Comparative observational studies

1.1 Mastectomy versus chest binding

We could not perform meta-analyses. There was very low certainty evidence for GD related to the chest [42], and for body and chest satisfaction, [54]. See Table, Supplemental Digital Content 9 for details.

2.2 Mastectomy versus no mastectomy

We could not perform meta-analyses. There was very low certainty evidence about psychological distress (surrogate outcome for global functioning), and suicide attempts [50]. Low certainty evidence suggested that individuals who underwent mastectomy had 0.47 times the odds of non-suicidal self-injury (measured with an item from the Aoteorana/New Zealand Youth' 12 survey) than those who did not (aOR 0.47 [95% CI 0.22 to 0.97], low certainty) [50]. See Table, Supplemental Digital Content 10 for details.

2. Before-after studies

2.1 Mastectomy compared to no mastectomy (i.e., no intervention): We could not perform any meta-analyses. There was very low certainty evidence about quality of life, depression, GD related to the chest, GD, postsurgical persistent pain, chest satisfaction, and body satisfaction. See Table, Supplemental Digital Content 11.

Sensitivity analyses using different imputed correlation coefficients provided similar results (Appendix, Supplemental Digital Content 12).

3. Case series (addressing mastectomy)

3.1 Outcomes needing a comparison group to assess the effect of mastectomy:

There was very low certainty evidence about quality of life, depression, death by suicide, body satisfaction and chest satisfaction. See details in Table, Supplemental Digital Content 13.

3.2 Outcomes not needing a comparison group to assess the effect of mastectomy:

3.2.1 Death: A meta-analysis of five studies found that the number of people who undergo mastectomy that results in death is 0 per 1,000 (proportion 0 [95% CI 0 to 0], high certainty) [24, 26, 30, 45, 48].

3.2.2 Necrosis: A meta-analysis of eight studies found that the number of people who undergo mastectomy with complete and partial nipple/areola/nipple-areola complex/nipple graft necrosis is 30 per 1,000 (proportion 0.03 [95% CI 0.01 to 0.07], high certainty) [25, 26, 29, 33, 34, 37, 45, 46]. A meta-analysis of 13 studies showed

that the number of people who undergo mastectomy with complete nipple/areola/nipple-areola complex/nipple graft necrosis is 20 per 1,000 (proportion 0.02 [95% CI 0.01 to 0.03], high certainty) [19, 20, 27, 28, 31, 35, 36, 38-40, 44, 47, 48]. A meta-analysis of 11 studies suggested that the number of people who undergo mastectomy with partial nipple/areola/nipple-areola complex/nipple graft necrosis may be 20 per 1,000 (proportion 0.02 [95% CI 0.01 to 0.04], low certainty) [20, 23, 32, 39, 41, 44, 47-49, 52, 56].

A meta-analysis of three studies showed that the number of people who undergo mastectomy with complete flap necrosis is 10 per 1,000 (proportion 0.01 [95% CI 0.00 to 0.03], high certainty) [36, 41, 48]. A single study suggested that the number of people who undergo mastectomy with partial flap necrosis is 0 per 1,000 (proportion 0.00 [95% CI 0.00 to 0.04], low certainty) [48].

When reported per breast, a meta-analysis of three studies showed that the number of people who undergo mastectomy with complete nipple/areola/nipple-areola complex/nipple graft necrosis is 20 per 1,000 (proportion 0.02 [95% CI 0.00 to 0.07], high certainty), and the number of people who undergo mastectomy with partial nipple/areola/nipple-areola complex/nipple graft necrosis may be 40 per 1,000 (proportion 0.04 [95% CI 0.00 to 0.29], low certainty) [22, 43, 57].

3.2.3 Scarring: A meta-analysis of 12 studies suggested that the number of people who undergo mastectomy with hypertrophic scarring is 50 per 1,000 (proportion 0.05 [95% CI 0.03 to 0.08], high certainty) [21, 22, 28, 36, 39, 40, 45-47, 49, 51, 56].

The evidence reporting on fat necrosis, complete and partial nipple/areola/nipple-areola complex/nipple graft necrosis (reported per breast), have partial/full tissue loss (reported per breast) persistent numbness, keloid scarring, unclassified scarring, postsurgical persistent pain, and regret was very low certainty. See details in Table, Supplemental Digital Content 13.

DISCUSSION

This SR and meta-analysis synthesizes and appraises the best available evidence about the effects of mastectomy or chest masculinization. Comparative observational studies provided low to very low certainty evidence. Before-after studies provided very low certainty evidence. In case series studies, the certainty of the evidence ranged from high to very low for all outcomes.

Risk of bias in comparative observational studies were at critical risk of confounding. Before-after studies were at moderate to serious risk of bias due to confounding and missing outcome data. Case series studies were at moderate to critical risk of bias related to representativeness of the sample, missing outcome data, measurement of outcomes, and lack of a comparison group (at the outcome level). The results from case series studies have limited applicability to the larger GD population. Therefore, these findings should be used strictly for hypothesis generation for studies of stronger methodological design.

Risk of bias, indirectness, and imprecision were the primary reasons for rating down the certainty of the evidence for most outcomes, comparisons, and study designs.

Indirectness often resulted from less than 60% of participants being of the eligible mean age of 26 years and having full support of caregivers, which we believe does not represent the majority of individuals undergoing gender-affirming treatments. Imprecision often resulted from an insufficient sample size.

We did not find data for the outcomes of shortness of breath, rib pain, back pain, inpatient psychiatric hospitalization, and substance use/abuse. These outcomes are likely important to patients with GD and clinicians involved in their care. Therefore, emerging studies evaluating mastectomy in this population should assess these outcomes. Further, due to feasibility concerns, we were unable to include the following outcomes that were considered relevant to clinical practice: revision due to complications, sexual function/satisfaction, internalizing disorders, externalizing disorders, eating disorders, suicidal ideation, relationships (family/peers), academic/occupational functioning, subsequent surgical/medical transition, age of commencement of puberty blockers and/or cross-sex hormones. Future SRs assessing the effects of mastectomy in individuals with gender dysphoria should consider these outcomes when devising review methodology.

This is the first SR and meta-analysis to include psychological and physical outcomes following mastectomy in individuals experiencing GD. In 2019, a SR assessing the effects of mastectomy in transgender individuals was published and included 22 articles and 2447 patients [58]. This review found limited evidence about aesthetic satisfaction and surgical complications (i.e., nipple necrosis, hematoma, and hypertrophic scarring) and stated that evidence regarding psychological outcomes is

missing, perhaps due to lack of validated tools to measure these outcomes in this patient population. In 2022, a SR assessing the effects of gender-affirming surgery included 35 publications assessing mastectomy in transgender individuals [59]. Similar to the review by Cohen et al, this review presented limited data about cosmetic satisfaction and surgical complications (i.e., complication rate, hematoma, and rate of reoperation) and did not report data on the psychological outcomes. Both of the aforementioned SRs did not assess the risk of bias in the included publications and the certainty of the evidence for each outcome of interest. Interpretation of the results of both SRs would be highly dependent on the assessment of risk of bias and certainty of the evidence due to various methodological limitations (i.e., heterogeneity) across the included studies.

Given the predominantly very low certainty evidence on mastectomy effects in individuals below 26 years of age with GD, factoring in both the available evidence and the values and preferences of patients and caregivers is important when making practice recommendations, as well as clinical and policy decisions for this group. Guideline developers and policy makers should be transparent about whose values they are prioritizing when making recommendations and policy decisions. Other considerations may include resource use, acceptability, feasibility, and equity [60].

Strengths and limitations of the review process

This SR and meta-analysis has multiple strengths. We rigorously followed the highest methodological standards throughout the process. We assessed the risk of bias for each study and the certainty of the evidence for each outcome using current methods. We

performed analyses and interpreted results following GRADE guidance. A limitation was the inclusion of only English-published studies; however, we do not anticipate important changes to our conclusions by adding studies in other languages.

CONCLUSION

The best available evidence reporting the effects of mastectomy in individuals with GD ranged from high to very low certainty. High certainty evidence from prospective cohort studies and, if ethical, randomized controlled trials, are needed to understand the short- and long-term effects of mastectomy in individuals with GD on physical and psychological outcomes. Higher certainty evidence would be tremendously useful to individuals with GD considering chest masculinization; clinicians and surgeons involved in their care; guideline developers; and policy makers and stakeholders who make decisions about treatment related to GD.

Supplemental material

A supplemental appendix to this article is available online.

Conflicts of interest statement

Dr. Romina Brignardello and Anna Miroshnychenko provided methodological expertise for the SEGM initiative to summarize and appraise the quality of publications related to gender medicine for the SEGM online platform and received financial compensation from SEGM. Dr. Romina Brignardello-Petersen presented preliminary

results at the conference titled *International Perspectives of Evidence Based Treatment for Gender Dysphoric Youth (New York, 2023)*.

Dr. Chan Kulatunga-Moruzi and Dr. Kristen Dahlin's fee to participate in the conference titled *International Perspectives of Evidence Based Treatment for Gender Dysphoric Youth (New York, 2023)* was waived.

Dr. Gordon Guyatt was a speaker at the conference titled *International Perspectives of Evidence Based Treatment for Gender Dysphoric Youth (New York, 2023)*.

Dr. Yetiani M Roldan, Sara Ibrahim, Dr. Steven Montante and Rachel Couban have no conflicts of interest to report.

E. Abbruzzese is a contributing author for the Society for Evidence-based Gender Medicine online platform and received financial compensation from SEGM.

Dr. William Malone's fee for publishing a research article as "open access" was compensated by SEGM. Dr. William Malone has expressed opinions about gender affirmation interventions for adolescents and young adults in The Journal of Clinical Endocrinology and Metabolism, The Lancet, Child and Adolescent Health, and Medscape.

Funding statement

This systematic review is part of a large research project funded through a research agreement between the Society for Evidence-based Gender Medicine (SEGM), the

sponsor, and McMaster University. None of the team members received financial compensation directly from SEGM to conduct this work.

Management of conflicts of interest statement

We conducted this SR is part of a research agreement between the Society for Evidence-based Gender Medicine (SEGM), the sponsor, and McMaster University. None of the researchers conducting this SR and meta-analysis received direct financial compensation from the sponsor to conduct this work. The SR and meta-analysis research question was designed through a collaboration between SEGM representatives, independent researchers, and the methods team based at McMaster University. Independent researchers who do not have any financial or intellectual conflicts of interest or disclosures, and the methodology team based at McMaster University (the group of authors) conducted the rest of the SR and meta-analysis processes (i.e., search, data extraction, data analyses, manuscript writing, approval of final draft of manuscript). The methods team was solely responsible for the synthesis and interpretation of results, and for drawing conclusions. To minimize bias, a methodologist who was not involved in the data collection, synthesis, and interpretation (GG) ensured that results interpretation was consistent with the findings. This manuscript was drafted by the methods team and approved by all authors, and the sponsor did not have any say in its content.

Ethics approval statement

The ethics approval was not required.

REFERENCES

1. American Psychiatric Association, *The Diagnostic and Statistical Manual of Mental Disorders (5th ed DSM-5)*, ed. A. American Psychiatric and D.S.M.T.F. American Psychiatric Association. 2013, Arlington, VA: American Psychiatric Association.
2. Monstrey, S., et al., *Chest-wall contouring surgery in female-to-male transsexuals: a new algorithm.* Plast Reconstr Surg, 2008. **121**(3): p. 849-859.
3. Murad, M.H., et al., *Hormonal therapy and sex reassignment: a systematic review and meta-analysis of quality of life and psychosocial outcomes.* Clin Endocrinol (Oxf), 2010. **72**(2): p. 214-31.
4. Ammari, T., et al., *Female-to-Male Gender-Affirming Chest Reconstruction Surgery.* Aesthetic Surgery Journal, 2018. **39**(2): p. 150-163.
5. Bluebond-Langner, R., et al., *Top Surgery in Transgender Men: How Far Can You Push the Envelope?* Plast Reconstr Surg, 2017. **139**(4): p. 873e-882e.
6. Wolter, A., et al., *Sexual reassignment surgery in female-to-male transsexuals: an algorithm for subcutaneous mastectomy.* J Plast Reconstr Aesthet Surg, 2015. **68**(2): p. 184-91.
7. Dy, G.W., et al., *Patient reported outcome measures and quality of life assessment in genital gender confirming surgery.* Transl Androl Urol, 2019. **8**(3): p. 228-240.
8. Park, R.H., et al., *Long-term Outcomes After Gender-Affirming Surgery: 40-Year Follow-up Study.* Ann Plast Surg, 2022. **89**(4): p. 431-436.
9. Page, M.J., et al., *PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews.* BMJ, 2021. **372**: p. n160.
10. Sterne, J.A., et al., *ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions.* BMJ, 2016. **355**: p. i4919.
11. Sterne, J.A.C., et al., *RoB 2: a revised tool for assessing risk of bias in randomised trials.* BMJ, 2019. **366**: p. l4898.
12. Thorlund, K., et al., *Pooling health-related quality of life outcomes in meta-analysis-a tutorial and review of methods for enhancing interpretability.* Res Synth Methods, 2011. **2**(3): p. 188-203.
13. Schwarzer, G., *meta: General Package for Meta-Analysis.* 2015(R package meta version 5.1-0).
14. Viechtbauer, W., *metafor: Meta-Analysis Package for R.* 2021(R package metafor version 3.0-2).
15. Schünemann, H.J., et al., *GRADE guidelines: 18. How ROBINS-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence.* J Clin Epidemiol, 2019. **111**: p. 105-114.
16. Guyatt, G.H., et al., *GRADE guidelines: 9. Rating up the quality of evidence.* J Clin Epidemiol, 2011. **64**(12): p. 1311-6.
17. GRADEpro GDT., *GRADEpro Guideline Development Tool*. 2022, McMaster University and Evidence Prime.

18. Schandelmaier, S., et al., *Development of the Instrument to assess the Credibility of Effect Modification Analyses (ICEMAN) in randomized controlled trials and meta-analyses.* Cmaj, 2020. **192**(32): p. E901-e906.
19. Ascha, M., et al., *Top Surgery and Chest Dysphoria Among Transmasculine and Nonbinary Adolescents and Young Adults.* JAMA Pediatr, 2022. **176**(11): p. 1115-1122.
20. Ayyala, H.S., et al., *A Three-Step Technique for Optimal Nipple Position in Transgender Chest Masculinization.* Aesthet Surg J, 2020. **40**(11): p. Np619-np625.
21. Berry, M.G., R. Curtis, and D. Davies, *Female-to-male transgender chest reconstruction: a large consecutive, single-surgeon experience.* J Plast Reconstr Aesthet Surg, 2012. **65**(6): p. 711-9.
22. Bustos, S.S., et al., *The Nipple Split Sharing vs. Conventional Nipple Graft Technique in Chest Wall Masculinization Surgery: Can We Improve Patient Satisfaction and Aesthetic Outcomes?* Aesthetic Plastic Surgery, 2020: p. 1-9.
23. Caro, C., et al., *Color Doppler Sonography Assisted Subcutaneous Mastectomy with Inferior Pedicled Nipple-Areola Complex in Female-to-Male Transsexuals: A Retrospective Cohort Analysis.* Aesthetic Plast Surg, 2022.
24. Cuccolo, N.G., et al., *Mastectomy in Transgender and Cisgender Patients: A Comparative Analysis of Epidemiology and Postoperative Outcomes.* Plast Reconstr Surg Glob Open, 2019. **7**(6): p. e2316.
25. Donato, D.P., et al., *Female-to-Male Chest Reconstruction: A Review of Technique and Outcomes.* Ann Plast Surg, 2017. **79**(3): p. 259-263.
26. Ederer, I.A., et al., *A Single-Center 10-Year Experience of 180 Transmasculine Patients Undergoing Gender-Affirming Mastectomy While Continuing Masculinizing Hormone Replacement Therapy.* Aesthetic Plast Surg, 2023. **47**(3): p. 946-954.
27. Elfering, L., et al., *Combining total laparoscopic hysterectomy and bilateral salpingo-oophorectomy with subcutaneous mastectomy in trans men: The effect on safety outcomes.* Int J Transgend Health, 2020. **21**(2): p. 138-146.
28. Frederick, M.J., A.E. Berhanu, and R. Bartlett, *Chest Surgery in Female to Male Transgender Individuals.* Ann Plast Surg, 2017. **78**(3): p. 249-253.
29. Gold, D., et al., *Combined Mastectomy and Laparoscopic Hysterectomy with Salpingo-Oophorectomy in Transgender Men: a Cohort Study.* Reprod Sci, 2021. **28**(12): p. 3515-3518.
30. Jolly, D., E.R. Boskey, and O. Ganor, *Racial Disparities in the 30-Day Outcomes of Gender-affirming Chest Surgeries.* Ann Surg, 2023. **278**(1): p. e196-e202.
31. Junn, A., et al., *Hockey stick incision: a modified technique for chest wall masculinization.* Ann Transl Med, 2021. **9**(7): p. 600.
32. Kääriäinen, M., et al., *Chest-wall contouring surgery in female-to-male transgender patients: A one-center retrospective analysis of applied surgical techniques and results.* Scand J Surg, 2017. **106**(1): p. 74-79.

33. Kamali, A., et al., *Improved Surgical Outcome with Double Incision and Free Nipple Graft in Gender Confirmation Mastectomy.* Plast Reconstr Surg Glob Open, 2021. **9**(7): p. e3628.
34. Kaur, M.N., et al., *Health state utility values in patients undergoing chest masculinization surgery.* J Plast Reconstr Aesthet Surg, 2023. **81**: p. 26-33.
35. Kelly-Schuette, K.A., et al., *Masculinizing chest surgery for gender affirmation: A retrospective study of outcomes and patient reported satisfaction.* Am J Surg, 2022. **223**(3): p. 577-580.
36. Lane, M., et al., *Gender Affirming Mastectomy Improves Quality of Life in Transmasculine Patients: A Single-center Prospective Study.* Ann Surg, 2023. **277**(3): p. e725-e729.
37. Lang, C.L., et al., *Low Risk of Persistent Pain, Sensory Disturbances, and Complications Following Mastectomy After Gender-Affirming Surgery.* Transgend Health, 2021. **6**(4): p. 188-193.
38. Marinkovic, M. and R.S. Newfield, *Chest reconstructive surgeries in transmasculine youth: Experience from one pediatric center.* International Journal of Transgenderism, 2017. **18**(4): p. 376-381.
39. McEvenue, G., et al., *Female-to-Male Gender Affirming Top Surgery: A Single Surgeon's 15-Year Retrospective Review and Treatment Algorithm.* Aesthet Surg J, 2017. **38**(1): p. 49-57.
40. Naides, A.I., et al., *Chest Masculinization Technique and Outcomes in 72 Double-incision Chest-contouring Procedures with Free Nipple Grafting.* Plast Reconstr Surg Glob Open, 2021. **9**(3): p. e3459.
41. Namba, Y., T. Watanabe, and Y. Kimata, *Mastectomy in female-to-male transsexuals.* Acta Med Okayama, 2009. **63**(5): p. 243-7.
42. Olson-Kennedy, J., et al., *Chest Reconstruction and Chest Dysphoria in Transmasculine Minors and Young Adults: Comparisons of Nonsurgical and Postsurgical Cohorts.* JAMA Pediatr, 2018. **172**(5): p. 431-436.
43. Perez-Alvarez, I.M., et al., *Incidence of Complications in Chest Wall Masculinization for the Obese Female-to-Male Transgender Population: A Case Series.* World J Plast Surg, 2021. **10**(2): p. 14-24.
44. Robinson, I.S., et al., *Perioperative Hormone Management in Gender-Affirming Mastectomy: Is Stopping Testosterone before Top Surgery Really Necessary?* Plastic and Reconstructive Surgery, 2023. **151**(2).
45. Rothenberg, K.A., et al., *Gender-Affirming Mastectomy in Transmasculine Patients: Does Obesity Increase Complications or Revisions?* Ann Plast Surg, 2021. **87**(1): p. 24-30.
46. Rothenberg, K.A., W.M.Y. Tong, and K.M. Yokoo, *Early Experiences With the Buttonhole Modification of the Double-Incision Technique for Gender-Affirming Mastectomies.* Ann Plast Surg, 2018. **81**(6): p. 642-645.
47. Sır, E. and Y. Tuluy, *Evaluation of Life Improvement in Trans Men After Mastectomy: A Prospective Study Using the TRANS-Q.* Aesthetic Plast Surg, 2022. **46**(5): p. 2556-2561.
48. Stein, M.J., et al., *Surgical and patient-reported outcomes following double incision and free nipple grafting for female to male gender affirmation: does*

obesity make a difference? J Plast Reconstr Aesthet Surg, 2021. **74**(8): p. 1743-1751.
49. Stojanovic, B., et al., *One-Stage Gender-Confirmation Surgery as a Viable Surgical Procedure for Female-to-Male Transsexuals.* J Sex Med, 2017. **14**(5): p. 741-746.
50. Tan, K.K.H., et al., *Unmet need for gender-affirming care as a social determinant of mental health inequities for transgender youth in Aotearoa/New Zealand.* J Public Health (Oxf), 2023. **45**(2): p. e225-e233.
51. Tang, A., et al., *Gender-Affirming Mastectomy Trends and Surgical Outcomes in Adolescents.* Annals of Plastic Surgery, 2022. **88**(4).
52. Top, H. and S. Balta, *Transsexual Mastectomy: Selection of Appropriate Technique According to Breast Characteristics.* Balkan Med J, 2017. **34**(2): p. 147-155.
53. van de Grift, T.C., et al., *Surgical Indications and Outcomes of Mastectomy in Transmen: A Prospective Study of Technical and Self-Reported Measures.* Plast Reconstr Surg, 2017. **140**(3): p. 415e-424e.
54. van de Grift, T.C., et al., *Subcutaneous Mastectomy Improves Satisfaction with Body and Psychosocial Function in Trans Men: Findings of a Cross-Sectional Study Using the BODY-Q Chest Module.* Plast Reconstr Surg, 2018. **142**(5): p. 1125-1132.
55. van de Grift, T.C., et al., *Body Image in Transmen: Multidimensional Measurement and the Effects of Mastectomy.* The Journal of Sexual Medicine, 2016. **13**(11): p. 1778-1786.
56. Wolf, Y. and S. Kwartin, *Classification of Transgender Man's Breast for Optimizing Chest Masculinizing Gender-affirming Surgery.* Plast Reconstr Surg Glob Open, 2021. **9**(1): p. e3363.
57. Wolter, A., et al., *Subcutaneous mastectomy in female-to-male transsexuals: Optimizing perioperative and operative management in 8 years clinical experience.* J Plast Reconstr Aesthet Surg, 2018. **71**(3): p. 344-352.
58. Cohen, W.A., et al., *Female-to-Male Transgender Chest Contouring: A Systematic Review of Outcomes and Knowledge Gaps.* Ann Plast Surg, 2019. **83**(5): p. 589-593.
59. Oles, N., et al., *Gender Affirming Surgery: A Comprehensive, Systematic Review of All Peer-reviewed Literature and Methods of Assessing Patient-centered Outcomes (Part 1: Breast/Chest, Face, and Voice).* Ann Surg, 2022. **275**(1): p. e52-e66.
60. Schünemann, H., Brożek J, Guyatt G, Oxman A, editors., *GRADE handbook for grading quality of evidence and strength of recommendations*. 2013: The GRADE Working Group.

CHAPTER 6: DISCUSSION AND OPPORTUNITIES FOR FUTURE RESEARCH

OBJECTIVE AND OVERVIEW

The objective of this thesis is to illustrate and discuss the standard, intermediate and advanced syntheses methods used to inform decision making about the use of opioids and analgesics for acute dental pain and gender-affirming treatments for gender dysphoria. This thesis is based on four peer-reviewed publications that illustrate the application of these methods: two publications assessing the effects of analgesics for the management of acute dental pain and two publications assessing the effects of interventions to manage gender dysphoria. This discussion summarizes the standard, intermediate, and advanced methods used in each of these publications, highlights important findings of these publications, and suggests a future direction for research about acute dental pain and gender dysphoria.

CLASSIFICATION OF METHODS

There is no widely agreed upon definition of standard methods in systematic reviews. For this thesis, we defined "standard methods" as any methods for conducting an evidence synthesis that are taught in the HRM 743 Systematic Review Methods course. We chose the HRM 743 course, as it was developed by Dr. David Sacket and Dr. Gordon Guyatt, who coined the term evidence-based medicine. The course continues to be updated and taught by the faculty members within the Health Research Methodology program at McMaster University who are world-renowned experts in conducting evidence syntheses. The methods taught within the HRM 743 course align with all methods mentioned in the Methodological Expectations of Cochrane

Intervention Reviews [1] and PRISMA Checklist [2] globally used to monitor reporting quality of systematic reviews and meta-analyses. Further, the course teaches methods that minimize risk of bias in systematic reviews and meta-analyses measured by the AMSTAR 2 [3] and ROBIS [4] tools. Most researchers around the world use these tools when designing, conducting, and reporting systematic reviews and meta-analyses. We define "intermediate methods" as any methods for conducting an evidence synthesis that are not explicitly taught with the lectures of the HRM 743 Systematic Reviews course but may be used by some students in the course with the guidance of their mentors. We define "advanced methods" as any methods for conducting an evidence synthesis that are beyond the scope of the HRM743 Systematic Reviews course. Further, we define advanced methods as any methods that employ artificial intelligence (AI). We define non-standard methods as any intermediate and advanced methods.

EVIDENCE SYNTHESIS: ACUTE DENTAL PAIN

According to the published statistics, acute pain associated with tooth extraction or following tooth extraction has been and continues to be one of the most common symptoms managed by dental professionals. In the United States of America, acute dental pain was estimated to affect 30% of the population each year, while in Canada it affects 11.7% each year [5-7]. Due to the high prevalence of dental pain among North Americans and the abundance of therapeutics available to manage acute dental pain, we anticipated to find a large number of prospective studies, specifically randomized controlled trials (RCTs), assessing the effectiveness of these therapeutics.

As there are many interventions available to manage acute dental pain, we aimed to determine the comparative effectiveness of the most frequently used analgesics, including acetaminophen, NSAIDs, and opioids, through a systematic review and network meta-analysis. In terms of continuous outcomes, in collaboration with the clinical experts, we included pain relief, total pain relief (TOTPAR), summed pain intensity difference (SPID), and global efficacy rating. With respect to the dichotomous outcomes, we included neurological and gastrointestinal adverse effects. In this systematic review and network meta-analysis, we only included randomized controlled trials.

Publication 1. Acute postoperative pain due to dental extraction in the adult population: A systematic review and network meta-analysis

Standard Methods

The first evidence synthesis, a systematic review and network meta-analysis, compared the effectiveness of acetaminophen, non-steroidal anti-inflammatory drugs (NSAIDs), and opioids to each other and placebo when managing acute postoperative pain. We used standard methods when searching databases to find relevant articles for inclusion. We searched databases including Medline, EMBASE, CENTRAL, and US Clinical Trials registry. In terms of data synthesis, according to standard methods, when studies reported the same outcome using scales with different ranges, we converted data to the scale most commonly reported before conducting analyses. Further, when standard deviation (SD) was not reported, we calculated SD using standard error, confidence

intervals, means, and sample sizes. With respect to risk of bias, for each eligible trial and outcome, pairs of reviewers, following training and calibration exercises, independently used the Cochrane tool to assess the risk of bias in randomized controlled trials (RoB 2.0). We used standard methods when reporting conflicts of interest by stating any conflicts of interest of manuscript authors at the end of the manuscript, according to the guidelines of the manuscript journal.

Intermediate Methods

The purpose of the first publication was to support the process of informing a clinical practice guideline, which was considered an intermediate method. The research question of the first publication aimed to compare multiple interventions to each other and to placebo, which was also considered an intermediate method. In terms of data synthesis, when standard error, confidence intervals, means, and sample sizes were not reported, we imputed SD by choosing a median SD of three studies with similar means. In terms of risk of bias, we used intermediate methods when we changed the response options to better capture educated guesses whether the bias was probably low or probably high. We changed response options from (1) low, (2) high, (3) some concerns to (1) low, (2) probably low, (3) probably high, (4) high (See Appendix 1 for sample of risk of bias assessment). With respect to certainty of the evidence, in collaboration with a research panel consisting of 20 professionals, we selected and used a single threshold approach to rate the certainty that there was an important effect, using a threshold of 10% of the length of the scale as the minimally important difference for continuous outcomes and 10% of baseline risk for dichotomous outcomes.

Advanced Methods

For the first publication, we used advanced methods for data synthesis by performing frequentist network meta-analysis (NMA) for all outcomes. We used Grading of Recommendations Assessment, Development and Evaluation (GRADE) guidance for network meta-analyses and used an automated tool to rate the certainty of the evidence for direct, indirect, and network estimates of effect. For direct estimates, we assessed risk of bias, inconsistency, indirectness, publication bias, and imprecision. For indirect estimates, we assessed transitivity and imprecision. For network estimates, we assessed incoherence and imprecision. In terms of data interpretation, to draw conclusions, we classified interventions in groups from the most to the least effective by considering the estimates of effect and the certainty of the evidence. For data presentation, we used tables to organize interventions from the most to the least effective that include the estimates of effect and the certainty of the evidence ratings using a colour coding system (See Appendix 2). Green, yellow, and red colours indicated high and moderate certainty evidence, and light, moderate, and dark grey indicated low and very low certainty evidence. Red and dark grey colours indicated that an intervention is no better than placebo, yellow and moderate grey colours indicated that an intervention is better than placebo but no better than other interventions, and green and light grey colours indicated that an intervention is better than placebo and some alternatives. See Table 1 and 2 for a summary of standard, intermediate and advanced methods used to assess the comparative effectiveness of acetaminophen, non-steroidal anti-inflammatory drugs, and opioids in adults with acute postoperative pain due to dental extraction.

Table 1. The standard, intermediate, and advanced methods used to assess the comparative effectiveness of acetaminophen, non-steroidal anti-inflammatory drugs, and opioids in adults with acute postoperative pain due to dental extraction.

Systematic review stage	**Method**	**Type of method**
Purpose of the review	To support the process of informing a clinical practice guideline	Intermediate
Research question	Multiple comparison question	Intermediate
Search/Data collection	Searched Medline, EMBASE, CENTRAL, and US Clinical Trials Registry	Standard
Data synthesis	When studies reported the same outcome using scales with different ranges, converted data to the scale most commonly reported When standard deviation (SD) was not reported, calculated SD using standard error, confidence intervals, means, and sample sizes	Standard
	When standard error, confidence intervals, means, and sample sizes were not reported, imputed SD by choosing a median SD of 3 studies with similar means	Intermediate
	Performed frequentist NMAs for all outcomes	Advanced
Risk of bias	For each eligible trial and outcome, pairs of reviewers, following training and calibration exercises, independently used RoB 2.0 tool to assess risk of bias in randomized controlled trials	Standard
	Changed the response options to better capture educated guesses whether the bias was probably low or probably high Changed response options from (1) low, (2) high, (3) some concerns to (1) low, (2) probably low, (3) probably high, (4) high	Intermediate
Certainty of the evidence	In collaboration with a research panel consisting of 20 professionals, selected and used a single threshold approach to rate the certainty that there was an important effect, using a threshold of 10% of the length of the scale as the minimally important difference for continuous outcomes and 10% of the baseline risk for dichotomous outcomes	Intermediate

	Used GRADE guidance for network meta-analyses and used an automated tool to rate the certainty of the evidence for direct, indirect, and network estimates of effect	Advanced
Data interpretation	To draw conclusions, classified interventions in groups from the most to the least effective by considering the estimates of effect and the certainty of the evidence	Advanced
Data presentation	Using a color-coding system, organized interventions from the most to the least effective by considering the estimates of effect and the certainty of the evidence ratings	Advanced
Reporting and management of conflicts of interest	Reported any conflicts of interest of manuscript authors at the end of the manuscript	Standard

Table 2. The non-standard methods used to assess the comparative effectiveness of acetaminophen, non-steroidal anti-inflammatory drugs, and opioids and its corresponding standard methods.

Evidence syntheses methods	**Standard methods (not used in this publication)**	**Non-standard methods (used in this publication)**
Purpose of the review	• To answer an effectiveness question	• To support the process of informing a clinical practice guideline
Research question	• Single comparison question	• Multiple comparison question
Analysis	• Conducting a pairwise meta-analysis to assess the effectiveness of one therapeutic to another or placebo	• Conducting a network meta-analysis to assess the comparative effectiveness of multiple therapeutics to each other or placebo
Risk of bias	• Use of original response options: (1) low, (2) high, (3) some concerns	• Changing response options from (1) low, (2) high, (3) some concerns to (1) low, (2) probably low, (3) probably high, (4) high in order to better capture educated guesses whether the bias was probably low or probably high
Certainty of the evidence assessment	• Using GRADEpro to rate the certainty of the evidence from a pairwise meta-analysis	• Using GRADE guidance for direct, indirect, and network meta-analysis and using an automated tool to rate the

	• Selecting and using a single threshold approach to rate the certainty that there was an important effect by searching the literature and/or consulting an expert in the field	certainty of the evidence for direct, indirect and network estimates of effect • In collaboration with a research panel, selecting and using a single threshold approach to rate the certainty that there was an important effect, using a threshold of 10% of the length of the scale as the minimally important difference for continuous outcomes and 10% of the baseline risk for dichotomous outcomes
Data interpretation/ presentation	• Using statements to report findings of the pairwise comparisons that include the estimates of effect and the certainty of the evidence ratings • Using the GRADE summary of findings tables to report the estimates of effect with the certainty of the evidence ratings for each pairwise comparison of one intervention to another or to placebo	• Classifying the interventions into groups from the most to the least effective by considering both, the estimates of effect and the certainty of the evidence ratings • Using tables to organize interventions from the most to the least effective that include the estimates of effect and the certainty of the evidence ratings using a colour coding system

Publication 2. Corticosteroids for managing acute pain subsequent to surgical extraction of mandibular third molars: A systematic review and meta-analysis

Standard Methods

The second evidence synthesis, a systematic review and meta-analysis, assessed the comparative effectives of various corticosteroids to manage acute pain subsequent to surgical extraction of mandibular third molars. In terms of data synthesis, using

standard methods, for dichotomous outcomes, we summarized the effect of interventions using odds ratios, and risk difference when the incidence of the outcome was low across studies. For continuous outcomes, we used mean difference. When studies reported the same outcome using scales with different ranges, we converted data to the scale most commonly reported before conducting analyses in accordance with standard methods. When SD was not reported, we calculated SD using standard error, confidence intervals, means, and sample sizes. According to standard methods, we conducted random-effects meta-analyses using the Review Manager software. With respect to risk of bias, for each eligible trial and outcome, pairs of reviewers, following training and calibration exercises, independently used the RoB 2.0 tool to access the risk of bias in randomized controlled trials. In terms of the certainty of the evidence, we used a single threshold approach to rate the certainty that there was an important effect using the null as the threshold. To interpret data, we used statements to report findings of the pairwise comparisons that include the estimates of effect and the certainty of the evidence ratings. To present data, we used the GRADE summary of findings tables to report the estimates of effect with the certainty of the evidence ratings for each pairwise comparison of the intervention to placebo. Finally, we used standard methods when stating any conflicts of interest of manuscript authors at the end of the manuscript, according to the guidelines of the manuscript journal.

Intermediate Methods

As an intermediate method, the purpose of the second publication was to support the process of informing a clinical practice guideline. In terms of data synthesis, using

intermediate methods, we performed the search through the Epistemonikos database by creating a matrix of evidence that displays all of the systematic reviews addressing the question of interest and the studies that these reviews have included. With respect to data synthesis, we used intermediate methods when imputing SD by choosing a median SD of three studies with similar means when standard error, confidence intervals, means, and sample sizes were not reported. According to intermediate methods, we changed the response options to better capture educated guesses whether the bias was probably low or probably high. We changed the response options from (1) low, (2) high, (3) some concerns to (1) low, (2) probably low, (3) probably high, (4) high (See Appendix 3). To assess the certainty of the evidence, after consultation with a research panel consisting of 20 professionals, when the effect estimate was close to the null effect, we rated the certainty of a trivial effect using a threshold of 10% of the baseline risk for dichotomous outcomes and 10% of the scale range for continuous outcomes.

Advanced Methods

For the second publication, we searched for the evidence using the Living Overview of Evidence (LOVE) platform that incorporates artificial intelligence, in addition to creating a matrix of evidence through the Epistemonikos database (See Appendix 4). The LOVE platform organizes systematic reviews and randomized controlled trials into topics with the assistance of artificial intelligence and methodologists. When using the LOVE platform, systematic reviewers select an appropriate topic (e.g., third molar) aligned with their research question, and then further specify their selection criteria

(e.g., corticosteroids as intervention) by choosing among the available options that the platform provides. Then, the platform presents a list of potentially relevant publications. See Table 3 and 4 for a summary of standard, intermediate, and advanced methods used to assess the clinical effectiveness of corticosteroids for the management of acute pain subsequent to surgical extraction of mandibular third molars.

Table 3. The standard, intermediate, and advanced methods used to assess the clinical effectiveness of corticosteroids for the management of acute pain subsequent to surgical extraction of mandibular third molars.

Systematic review stage	**Method**	**Type of method**
Purpose of the review	To support the process of informing a clinical practice guideline	Intermediate
Research question	Single comparison question	Standard
Search/Data collection	Performed the search through the Epistemonikos database by creating a matrix of evidence	Intermediate
	Performed the search through the LOVE platform using artificial intelligence	Advanced
Data synthesis	For dichotomous outcomes, summarized the effect of interventions using odds ratios, and when the incidence of the outcome was low across studies, used the risk difference For continuous outcomes, used mean difference When studies reported the same outcome using scales with different ranges, converted data to the scale most commonly reported When SD was not reported, calculated SD using standard error, confidence intervals, means, and sample sizes Conducted random-effects meta-analyses using the Review Manager software	Standard
	In rare instances, when standard error, confidence intervals, means, and sample sizes were not	Intermediate

	reported, imputed SD by choosing a median SD of 3 studies with similar means	
Risk of bias	For each eligible trial and outcome, pairs of reviewers, following training and calibration exercises, independently used the RoB 2.0 tool to assess the risk of bias in randomized controlled trials	Standard
	Changed the response options to better capture educated guesses whether the bias was probably low or probably high Changed response options from (1) low, (2) high, (3) some concerns to (1) low, (2) probably low, (3) probably high, (4) high	Intermediate
Certainty of the evidence	Used a single threshold approach to rate the certainty that there was an important effect using the null as the threshold	Standard
	After consultation with a research panel consisting of 20 professionals, when the effect estimate was close to the null effect, rated the certainty of a trivial effect using a threshold of 10% of the baseline risk for dichotomous outcomes and 10% of the scale range for continuous outcomes	Intermediate
Data interpretation	Used statements to report findings of the pairwise comparisons that include the estimates of effect and the certainty of the evidence ratings	Standard
Data presentation	Used the GRADE summary of findings tables to report the estimates of effect with the certainty of the evidence ratings for each pairwise comparison of the intervention to placebo	Standard
Reporting and management of conflicts of interest	Reported any conflicts of interest of manuscript authors at the end of the manuscript	Standard

Table 4. The non-standard methods used to assess the clinical effectiveness of corticosteroids and its corresponding standard methods.

Evidence syntheses methods	**Standard methods (not used in this publication)**	**Non-standard methods (used in this publication)**
Purpose of review	• To answer an effectiveness question	• To support the process of informing a clinical practice guideline

Search	• Searching scientific databases (including the Cochrane Library), and references of individual studies	• Searching through Epistemonikos database and LOVE platform o Epistemonikos ▪ Cochrane Database of Systematic Reviews, Pubmed/MEDLINE, EMBASE, CINAHL, PsycINFO, LILACS, DARE, HTA Database, Campbell Database, JBI Database of Systematic Reviews and Implementation Reports, EPPI-Centre Evidence Library) • Synthesized a matrix of evidence, including all relevant systematic reviews and their included studies o Living Overview of Evidence Platform (LOVE) ▪ PubMed, EMBASE, and CENTRAL • Using artificial intelligence, identified individual studies aligned with the research question of interest
Data synthesis	• When standard deviation (SD) is not reported, calculating SD using standard error, confidence intervals, means, and sample sizes	• When standard error, confidence intervals, means, and sample sizes are not reported, imputing SD by choosing a median SD of 3 studies with similar means
Risk of bias	• Using the original response options: (1) low, (2) high, (3) some concerns	• Changing the response options from (1) low, (2) high, (3) some concerns to (1) low, (2) probably low, (3) probably high, (4) high to better capture educated guesses whether the bias was probably low or probably high
Certainty of the evidence	• Selecting and using a single threshold approach to rate the certainty that there	• After consultation with a research panel of clinical professionals, when the point estimate is close to the null effect, rating the certainty as having a

	was an important effect by searching the literature and/or consulting an expert in the field	trivial effect using a threshold of 10% of the baseline risk for dichotomous outcomes and 10% of the scale range for continuous outcomes

EVIDENCE SYNTHESIS: GENDER DYSPHORIA

The American Psychiatric Association added the diagnosis of "gender identity disorder" to the third *Diagnostic and Statistical Manual of Mental Disorders* (DSM-3) in 1980 [8]. In 2013, this diagnosis was reconceptualized as "gender dysphoria" in the *DSM-5* [8].

As the diagnosis of "gender dysphoria" only entered the *DSM-5* in 2013, we did not expect to find a large amount of high-quality evidence about the effects of gender-affirming treatments in individuals experiencing gender dysphoria. Therefore, we conducted a series of systematic reviews and meta-analyses, each assessing the effects of an intervention for the management of gender dysphoria compared to no such intervention, and we included all study designs to be comprehensive.

Publication 3. Gender-affirming hormone therapy for individuals with gender dysphoria below 26 years of age: A systematic review and meta-analysis

Standard Methods

The purpose of the third publication was to answer a question about clinical effectiveness of gender-affirming hormone therapy compared to no such intervention in individuals experiencing gender dysphoria, which we regarded as a standard method.

The research question was a single comparison question. Further, we searched relevant databases with the assistance of a research librarian. In terms of data synthesis, according to standard methods, for dichotomous outcomes, we summarized the effects of interventions using the odds ratio in comparative observational and before-after studies. For continuous outcomes, we summarized the effects of interventions using mean difference in comparative observational studies, and mean change in before-after studies. Following standard methods, when studies reported the same outcome using scales with different ranges, we converted these data to the scale most commonly reported before conducting analyses. Further, when clinically appropriate according to experts, we conducted random-effects meta-analyses using the Review Manager software. In terms of the risk of bias, we used the Cochrane risk of bias tool for non-randomized studies of interventions (ROBINS-I) for each study design. Following GRADE guidance, when addressing the risk of bias as a domain of GRADE for each outcome, we rated the certainty of the evidence from observational studies up to three levels for the risk of prognostic imbalance. Further, we used the null effect threshold to rate the certainty that there was a benefit or a harm of receiving gender-affirming hormone therapy over not receiving it. In order to interpret data, we used statements to report findings of the pairwise comparisons that include the estimates of effect and the certainty of the evidence ratings. To present data, we used the GRADE summary of findings tables to report the estimates of effect with the certainty of the evidence ratings for the pairwise comparisons of interest.

Intermediate Methods

We used intermediate methods to include the case series studies in the second publication. In terms of data synthesis, for dichotomous outcomes of case series studies, we summarized the effects of interventions using proportions, and for continuous outcomes, using means. In terms of the risk of bias, we developed guidance for the assessment of each risk of bias domain of the ROBINS-I tool that is specific to assessing gender-affirming hormone therapy in individuals experiencing gender dysphoria (See Appendix 5). Following GRADE guidance, for case series studies, we rated down the certainty of the evidence three levels due to the lack of a comparison group for all outcomes that required a comparison group. However, we did not rate down three levels if an outcome did not need a comparison group (e.g., cardiovascular events linked to receipt of gender-affirming hormone therapy). According to intermediate methods, we reported conflicts of interest of the protocol and the manuscript authors. We followed intermediate methods when we reported conflicts of interest of the systematic review protocol and the manuscript authors.

Advanced Methods

According to advanced methods, for data synthesis, when researchers did not provide correlation coefficients, to calculate mean change, we used a correlation coefficient of 0.5 and conducted sensitivity analyses by imputing correlation coefficients of r=0.2 and r=0.8 (See Appendix 6). For the third publication, due to the controversial nature of the subject matter, we devised a plan for minimization and management of conflicts of interest before initiating the systematic review process (See Appendix 7). See Table 5 and 6 for the standard, intermediate, and advanced methods used to summarize the

effects of gender-affirming hormone therapy in individuals under 26 years of age with gender dysphoria.

Table 5. The standard, intermediate, and advanced methods used to summarize the effects of gender-affirming hormone therapy for individuals with gender dysphoria below 26 years of age.

Systematic review stage	Method	Type of method
Purpose of the review	To answer a question about clinical effectiveness	Standard
Research question	Single comparison question	Standard
Search/Data collection	Searched in MEDLINE, PsycINFO, Social Sciences Abstracts, Contemporary Women's Issues, LGBTQ+ Source, Sociological Abstracts, Studies on Women, Gender Abstracts, and Google Scholar	Standard
	Included case series studies	Intermediate
Data synthesis	For dichotomous outcomes, summarized the effects of interventions using the odds ratio in comparative observational and before-after studies For continuous outcomes, summarized the effects of interventions using mean difference in comparative observational studies, mean change in before-after studies When studies reported the same outcome using scales with different ranges, converted data to the scale most commonly reported When clinically appropriate according to experts, conducted random-effects meta-analyses using Review Manager software	Standard
	For dichotomous outcomes, summarized the effects of interventions using proportions in case series studies For continuous outcomes, summarized the effects of interventions using means in case series studies	Intermediate
	When researchers did not provide correlation coefficients, when calculating mean change, used	Advanced

	r=0.5 and conducted sensitivity analyses by imputing correlation coefficients, r=0.2 and r=0.8	
Risk of bias	Used the Cochrane risk of bias tool for non-randomized studies of interventions (ROBINS-I) for each study design	Standard
	Developed guidance for the assessment of each risk of bias domain of the ROBINS-I tool that is specific to assessing gender-affirming hormone therapy in individuals experiencing gender dysphoria	Intermediate
Certainty of the evidence	Following GRADE guidance, when addressing the risk of bias as a domain of GRADE for each outcome, we rated down the certainty of the evidence from observational comparative studies up to three levels for the risk of prognostic imbalance Used the null effect threshold to rate the certainty that there was a benefit or a harm of gender-affirming hormone therapy over not receiving it	Standard
	Following GRADE guidance, for case series studies, rated down three levels due to lack of a comparison group for outcomes that needed a comparison group, and did not rate down three levels for outcomes that did not need a comparison group	Intermediate
Data interpretation	Used statements to report findings of the pairwise comparisons that include the estimates of effect and the certainty of the evidence ratings	Standard
Data presentation	Used the GRADE summary of findings tables to report the estimates of effect with the certainty of the evidence ratings for the pairwise comparisons of interest	Standard
Reporting and management of conflicts of interest	Reported conflicts of interest of the protocol and the manuscript authors at the end of the manuscript	Intermediate
	Devised a plan for minimization and management of conflicts of interest	Advanced

Table 6. The non-standard methods used to summarize the effects of gender-affirming hormone therapy and its corresponding standard methods.

Evidence syntheses methods	**Standard methods**	**Non-standard methods**
Search/Data collection	• Exclusion of case series studies	• Inclusion of case series studies

Data synthesis	• Exclusion of case series studies from the analysis • Not calculating mean change if correlation coefficient is not provided by the study authors; or using r=0.5 when researchers do not provide a correlation coefficient, without conducting sensitivity analyses with correlation coefficients, r=0.2 and r=0.8	• For dichotomous outcomes, summarizing the effects of interventions using proportions in case series studies For continuous outcomes, summarizing the effects of interventions using the mean in case series studies • When researchers do not provide correlation coefficients, when calculating mean change, using r=0.5 and conducting sensitivity analyses by imputing correlation coefficients r=0.2 and r=0.8
Risk of bias	• Using the generic guidance of the risk of bias tool to assess each risk of bias domain	• Developing guidance for the assessment of each risk of bias domain that is specific to assessing gender-affirming hormone therapy in individuals experiencing gender dysphoria
Certainty of the evidence	• Not assessing the certainty of the evidence derived from the case series studies	• Following GRADE guidance, for case series studies, rating down three levels due to the lack of a comparison group for outcomes that need a comparison group, and not rating down for outcomes that do not need a comparison group
Reporting and management of conflicts of interest	• Not devising a plan for minimization and management of conflicts of interest • Reporting conflicts of interest of manuscript authors	• Devising a plan for minimization and management of conflicts of interest • Reporting conflicts of interest of the protocol and manuscript authors

Publication 4: Mastectomy for individuals with gender dysphoria below 26 years of age: A systematic review and meta-analysis

Standard Methods

According to standard methods, the purpose of the fourth publication was to answer a question about clinical effectiveness of gender-affirming mastectomy compared to breast reduction or no mastectomy. The research question was a single comparison question. We used standard methods to search relevant databases with the assistance of a research librarian. In terms of data synthesis, according to standard methods, we summarized intervention effects using odds ratios for dichotomous outcomes in comparative observational and before-after studies. For continuous outcomes, we used mean difference in comparative observational studies, and mean change in before-after studies. When studies reported the same outcome using scales with different ranges, we converted data to the scale most commonly reported before conducting analyses. Further, when clinically appropriate according to experts, we conducted random-effects meta-analyses using the Review Manager software. In terms of the risk of bias, we used the ROBINS-I tool for each study design. Following GRADE guidance, when addressing risk of bias as a domain of GRADE for each outcome, we rated down the certainty of the evidence from observational comparative studies up to three levels for the risk of prognostic imbalance. We used a null effect threshold to rate the certainty of the benefit or harm of undergoing mastectomy over breast reduction or no mastectomy. To interpret data, we used statements to report findings of the pairwise comparisons that include the estimates of effect and the certainty of the evidence ratings. To present data, we used the GRADE summary of findings tables to report the estimates of effect with the certainty of the evidence ratings for each pairwise comparison.

Intermediate Methods

When searching the databases for relevant literature, we included case series studies. In terms of data synthesis, for dichotomous outcomes of case series studies, we summarized the effects of interventions using proportions, and for dichotomous outcomes using means. We developed guidance for the assessment of each risk of bias domain that is specific to assessing gender-affirming mastectomy in individuals experiencing gender dysphoria (See Appendix 8). Following GRADE guidance, for case series studies, we rated down the certainty of the evidence three levels due to the lack of a comparison group for outcomes that needed a comparison group (e.g., quality of life). However, we did not rate down the evidence three levels if an outcome needed a comparison group (e.g., scarring due to undergoing gender-affirming mastectomy). According to intermediate methods, we reported conflicts of interest of the protocol and the manuscript authors.

Advanced Methods

According to advanced methods, for data synthesis, when researchers did not provide correlation coefficients, to calculate mean change, we used a correlation coefficient of 0.5 and conducted sensitivity analyses by imputing correlation coefficients of r=0.2 and r=0.8 (See Appendix 9). With respect to conflicts of interest, due to the controversial nature of the subject matter, we devised a plan for minimization and management of conflicts of interest prior to initiating the systematic review process (See Appendix 6). See Table 7 and 8 for a summary of the standard, intermediate, and

advanced methods used to summarize the effects of gender-affirming mastectomy in individuals with gender dysphoria below 26 years of age.

Table 7. The standard, intermediate, and advanced methods used to summarize the effects of gender-affirming mastectomy in individuals with gender dysphoria below 26 years of age.

Systematic review stage	**Method**	**Type of method**
Purpose of the review	To answer a question about clinical effectiveness	Standard
Research question	Single comparison question	Standard
Search/Data collection	Searched in MEDLINE, Embase, PsycINFO, Social Sciences Abstracts, Contemporary Women's Issues, LGBTQ+ Source, Sociological Abstracts, Studies on Women, Gender Abstracts, and Google Scholar	Standard
	Included case series studies	Intermediate
Data synthesis	To summarize intervention effects, used odds ratios for dichotomous outcomes in comparative observational and before-after studies For continuous outcomes, we used mean difference in comparative observational studies, mean change in before-after studies When studies reported the same outcome using scales with different ranges, converted data to the scale most commonly reported When clinically appropriate according to experts, conducted random-effects meta-analyses using the Review Manager software	Standard
	For dichotomous outcomes, summarized the effects of interventions using proportions in case series studies For continuous outcomes, summarized the effects of interventions using means in case series studies	Intermediate
	When researchers did not provide correlation coefficients, when calculating mean change, used	Advanced

	r=0.5 and conducted sensitivity analyses by imputing correlation coefficients, r=0.2 and r=0.8	
Risk of bias	Used the Cochrane risk of bias tool for non-randomized studies of interventions (ROBINS-I) for each study design	Standard
	Developed guidance for the assessment of each risk of bias domain that is specific to assessing gender-affirming mastectomy in individuals experiencing gender dysphoria	Intermediate
Certainty of the evidence	Following GRADE guidance, when addressing risk of bias as a domain of GRADE for each outcome, we rated down the certainty of the evidence from observational comparative studies up to three levels for the risk of prognostic imbalance Used a null effect threshold to rate the certainty of the benefit or harm of undergoing gender-affirming mastectomy compared to breast reduction or no mastectomy	Standard
	Following GRADE guidance, for case series studies, rated down three levels due to lack of a comparison group for outcomes that needed a comparison group, and did not rate down for outcomes that did not need a comparison group	Intermediate
Data interpretation	Used statements to report findings of the pairwise comparisons that include the estimates of effect and the certainty of the evidence ratings	Standard
Data presentation	Used the GRADE summary of findings tables to report the estimates of effect with the certainty of the evidence ratings for each pairwise comparison of one intervention to another or to placebo	Standard
Reporting and management of conflicts of interest	Reported conflicts of interest of the protocol and the manuscript authors at the end of the manuscript	Intermediate
	Devised a plan for minimization and management of conflicts of interest	Advanced

Table 8. The non-standard methods used to summarize the effects of gender-affirming mastectomy and its corresponding standard methods.

Evidence syntheses methods	**Standard methods**	**Non-standard methods**
Search/Data collection	• Exclusion of case series studies	• Inclusion of case series studies

Data synthesis	• Exclusion of case series studies from the analysis • Not calculating mean change if correlation coefficient is not provided by the study authors; or using r=0.5 when researchers do not provide a correlation coefficient, without conducting sensitivity analyses with correlation coefficients, r=0.2 and r=0.8	• For dichotomous outcomes, summarizing the effects of interventions using proportions in case series studies For continuous outcomes, summarizing the effects of interventions using the mean in case series studies • When researchers do not provide correlation coefficients, when calculating mean change, using r=0.5 and conducting sensitivity analyses by imputing correlation coefficients r=0.2 and r=0.8
Risk of bias	• Using the generic guidance of the risk of bias tool to assess each risk of bias domain	• Developing guidance for the assessment of each risk of bias domain that is specific to assessing gender-affirming mastectomy in individuals experiencing gender dysphoria
Certainty of the evidence	• Not assessing the certainty of the evidence derived from the case series studies	• Following GRADE guidance, for case series studies, rating down three levels due to the lack of a comparison group for outcomes that need a comparison group, and not rating down for outcomes that do not need a comparison group
Reporting and management of conflicts of interest	• Not devising a plan for minimization and management of conflicts of interest • Reporting conflicts of interest of manuscript authors	• Devising a plan for minimization and management of conflicts of interest • Reporting conflicts of interest of the protocol and the manuscript authors

FUTURE OPPORTUNITIES FOR RESEARCH TO MANAGE THE ACUTE DENTAL PAIN AND GENDER DYSPHORIA

After conducting evidence syntheses about acute dental pain and gender dysphoria, several methodological limitations about this evidence were apparent. In this section, we describe these methodological limitations and potential solutions to address these limitations in future research.

Evidence-based medicine (EBM) seeks to integrate clinical judgment on the part of the physician, the best available evidence obtained through robust clinical research methodology, as well as patients' values and preferences. According to EBM, randomized controlled trials (RCTs) or systematic reviews and meta-analyses or network meta-analyses are considered higher on the hierarchy of evidence than observational studies when aiming to estimate a causal effect of an intervention on outcomes of interest. RCTs control for systematic error [9]. One aim of RCTs is to minimize bias in the selection of participants and allocation of participants to the trial arms through randomization and concealment of the allocation sequence, respectively [9]. Another goal of RCTs is to maintain prognostic balance by blinding participants, caregivers, data collectors, data analysts and outcome assessors, and by conducting the trial through the intention-to-treat analysis [9]. These methodological aspects of RCTs aim to minimize selection bias, confounding, and bias in assessment of outcomes, thereby strengthening any conclusions made about the relationship between the difference in outcomes at the end of the study and the difference in treatment during the study [9].

For the evidence synthesis about acute dental pain, consisting of one systematic review and network meta-analysis and three systematic reviews and meta-analyses, we chose

to include RCTs only. From our perspective, RCTs are the best study design to assess the comparative effectiveness of interventions to manage the acute dental pain and are frequently conducted for this purpose. The systematic review and network meta-analysis demonstrated that, based on high and moderate certainty, NSAIDs with or without acetaminophen result in better pain-related outcomes than opioids with or without acetaminophen or placebo. Further, based on low- and very low-certainty evidence, most interventions were classified as no more harmful than placebo for most selected adverse effects. These are important findings that provide further support to minimize or discontinue prescription of opioids for the management of acute pain, thereby reducing the possibility of addiction and death by overdose. Due to feasibility concerns, we were only able to assess central nervous system adverse effects and gastrointestinal adverse effects at the longest available time point. We found limited data for the adverse effects such as dysphagia, diarrhea, dyspepsia, constipation, and syncope. Future primary studies and systematic reviews should consider including these adverse effects for assessment, as well as consider collecting data at longer follow up points such as 6 months or 1 year.

Although most evidence included in the systematic review and network meta-analysis was rated as moderate to high certainty, several included RCTs were found to be at risk of bias. For instance, several RCTs reporting pain relief, total pain relief (TOTPAR), and summed pain intensity difference (SPID) outcomes were at risk of bias due to missing outcome data (i.e., up to 20%) and did not report the type and magnitude of measure of variability (i.e., standard deviation, standard error, etc.). Future efforts to

minimize the loss of outcome data should be made. Further, researchers should prioritize reporting type of measure of variability (i.e., standard deviation, standard error) and its magnitude for each measure of central tendency (i.e., mean, median).

In terms of the systematic review and meta-analysis assessing the effectiveness of corticosteroids compared to placebo, low and very low certainty evidence suggested that there is a trivial difference in postoperative pain intensity and postoperative infection when corticosteroids administered orally, submucosally, or intramuscularly are compared with placebo in patients undergoing third-molar extractions. This evidence indicates that corticosteroids may not be necessary when managing pain subsequent to surgical third-molar extraction. Due to feasibility concerns, we only included adverse effects such as postoperative infection, alveolar osteitis, mood alteration, and gastrointestinal adverse effects. From these outcomes, we only found data about postoperative infection, alveolar osteitis, and gastrointestinal adverse effects. Only data about postoperative infection was low certainty; data about alveolar osteitis and gastrointestinal adverse effects was very low certainty. Future primary studies and systematic reviews should assess these adverse effects and other patient-important long-term adverse effects related to the use of corticosteroids to manage acute dental pain.

Some RCTs included in the systematic review and meta-analysis assessing the effectiveness of corticosteroids compared to placebo were limited in terms of the randomization process, such that the authors did not describe whether the allocation sequence was random and concealed until the participants were recruited and assigned

to the interventions of interest. Further, in several RCTs, the researchers did not clearly describe whether and how blinding of participants and the outcome assessors was done. The lack of blinding in an RCT is known to introduce performance bias and lead to an overestimate of the treatment effect. Future RCTs assessing the effects of corticosteroids and other treatments to manage acute dental pain must carefully consider and report all aspects of the randomization process, including the random generation of a concealed allocation sequence and blinding of both, the participants and the outcome assessors involved in the RCT.

In terms of the systematic reviews and meta-analyses assessing the effects gender-affirming treatments, most evidence was derived from before-after and case series studies, which are limited in methodology. This evidence ranged from high to very low certainty. These systematic reviews and meta-analyses were the first to assess both, the risk of bias of each individual study and the certainty of the evidence of each outcome of interests while assessing gender-affirming interventions in individuals experiencing gender dysphoria. As such, these reviews are important for patients considering gender-affirming interventions, clinicians treating these patients, and policy makers and other end users making decisions about these treatments at the global scale.

Although both, the RCTs and observational studies were eligible, we did not find RCTs assessing the effects of gender-affirming hormone therapy or gender-affirming mastectomy in individuals experiencing gender dysphoria. One reason for not conducting RCTs to assess the effects of gender-affirming interventions may be that these RCTs would be prone to biases, applicability issues, and ethical concerns. The

purpose of puberty blockers, gender-affirming hormone therapy, and surgery is to inhibit or produce visible bodily changes. As such, meaningful effects on patient-important outcomes, such as psychological well-being and quality of life, may not be possible until these individuals experience the inhibition or creation of visible physical changes following their treatment [10]. The robust methodological design of an RCT relies on blinding of the researchers and study participants. However, it would not be possible to blind individuals receiving gender-affirming interventions, as these interventions result in visible physical changes that would be noticeable to the researchers and study participants. In case of surgery, blinding of both the study participant and surgeon would not be possible. This lack of blinding may lead to performance bias, as the participant or surgeon who trusts or mistrusts in the effect of a specific intervention may unconsciously or intentionally perceive an enhanced or reduced treatment effect, respectively [11].

Further, once participants become aware of their assigned group, a large proportion of the study participants are likely to crossover to the intervention group, withdraw from the study or pursue alternative sources of gender-affirming interventions, potentially leading to attrition bias [10]. Withdrawing from the study and non-compliance to the study protocol would most likely occur among adolescents who have other means of securing gender-affirming care or experience more severe bodily gender dysphoria. If individuals with severe bodily gender dysphoria withdraw from the control arm and not from the intervention arm, the results of the study may be biased away from the

null, resulting in an overestimate of the treatment effect, as more improvement may be detected in the individuals in the intervention arm compared to the control arm.

In terms of applicability, the results of an RCT may not be generalizable to the entire population of individuals with gender dysphoria, because youth who have alternative means of accessing gender-affirming care are unlikely to participate in an RCT and risk being randomized to the control group [12-14]. Similarly, adolescents who are white, socioeconomically privileged, live in areas with more gender-affirming care providers, and have strong parental support are less likely to participate in an RCT [15-17]. In contrast, adolescents who have less access to resources and support are more likely to participate in an RCT. Participation could also be impacted by the severity of individual's gender dysphoria, as those who experience greater distress about their body are more likely to seek out guaranteed access to puberty blockers, hormone therapy and/or surgeries [10]. Therefore, the results of an RCT may only be applicable to a subgroup of individuals with gender dysphoria, likely those with less resources and parental support, and less severe gender dysphoria.

Puberty blockers and gender-affirming hormones may be time-sensitive in terms of administration because waiting to administer these interventions may hinder their effect at a later time [18-20]. Therefore, withholding these interventions, as in the control group of an RCT, could be considered unethical. While withholding these interventions may negatively impact the possible benefit of them, administering these interventions may cause unknown harm due to the limited evidence about the effects of these intervention in this patient population. The current best available evidence

reporting on the effects of puberty blockers and gender-affirming hormones in individuals experiencing gender dysphoria mostly comes from before-after and case series studies and ranges from high to very low certainty. This is an important finding to consider when undergoing treatments such as puberty blockers, gender-affirming hormones and/or gender-affirming surgery. With respect to gender-affirming hormones, of 24 relevant studies that we found, one low-certainty comparative observational study suggested that the odds of depression may be lower (OR 0.73 [95% CI 0.61 to 0.88], low certainty) in individuals who received gender-affirming hormone therapy compared to those who did not. Further, one case series study provided high certainty evidence that the proportion of individuals with cardiovascular events 7-109 months after receiving GAHT was 40 per 1,000 (40 per 1,000 [95% CI 30 to 50], high certainty).

With respect to gender-affirming mastectomy, evidence ranged from high to very low certainty. Of 39 included studied, one low-certainty comparative observational study indicated that individuals who underwent mastectomy may have reduced odds of non-suicidal self-harm compared to individuals who did not (aOR 0.47 [95% CI 0.22 to 0.97], low certainty). Further, case series studies reported that 0 per 1,000 (0 per 1,000 [95% CI 0 to 0], high certainty) individuals die as a result of gender-affirming mastectomy, 30 per 1,000 (30 per 1,000 [95% CI 10 to 70], high certainty) experience necrosis and 50 per 1,000 (50 per 1,000 [95% CI 30 to 80], high certainty) experience excessive scarring. Therefore, the aforementioned results, although limited, seem to suggest both, possible benefit and harm associated with undergoing gender-affirming

hormone therapy and mastectomy. Future research of high quality is needed to better understand the effects of these interventions.

Due to the aforementioned challenges posed by the RCTs, well-designed observational studies have been preferred to RCTs by many clinicians and researchers studying the effects of gender-affirming treatment. Observational studies, such as a prospective cohort study, can reduce the risks associated with withdrawal and non-adherence by selecting a group of individuals who have undergone the intervention of interest and another group who have yet to undergo the intervention of interest but are planning to in the future. As observational studies do not directly intervene in the treatment of the individual, these studies may include individuals who would be unlikely to participate in an RCT, thus increasing the generalizability of the study results.

Although observational studies are vulnerable to unmeasured confounding, methodological and statistical tools can be used to ascertain and limit the risk of unknown and unmeasured confounders and control for known and measured confounders such as race, socioeconomic status, psychological comorbidities, and parental support. Observational studies, such as prospective cohort studies, are typically considered less equipped at inferring causality compared to RCTs due to a higher risk of bias arising from confounding, selection of participants, and measurement of outcomes. However, advanced methods, such as an instrumental variable analysis, may be used to estimate the causal effect of treatments on outcomes in the population of interest when conducting an observational study [21].

Future research in gender affirming care should consider prioritizing well-designed observational studies of prospective nature. For example, if we were to conduct a prospective cohort study assessing the effects of mastectomy, we could include individuals under the age of 26 who were diagnosed with, experienced, self-identified, or were identified by a parent as having gender dysphoria, gender identity disorder or gender incongruence, or who identified as transgender or non-binary. We could specify that we are interested in participants attending gender identity or gender-affirming health clinics that are available to be included in our study. If possible, to increase the power of the study to detect a difference in selected outcomes between the two groups if the difference exists, we could include all relevant clinics in Canada. The intervention of interest would be all types of gender-affirming mastectomy, and the comparator would be no gender-affirming mastectomy.

In terms of outcomes, we would aim to include all outcomes important to patients receiving the intervention of interest, and may include physical complications (i.e., hypertrophic scarring, keloid scarring, persistent numbness, persistent pain, necrosis, death), satisfaction with chest and body, regret, quality of life, global function, gender dysphoria or incongruence, depression, death by suicide, suicide attempts, non-suicidal self-harm, and any other outcomes considered important after deliberation with a team of clinical experts and patients. In terms of the outcomes, we could consult the experts in measurement of these outcomes to ensure we choose instruments with high validity, reliability, and responsiveness, while being easy and efficient to complete. In order to collect long-term outcome data, that is currently missing in published literature, we

could complete outcome assessments at baseline, 6 months, 1 year, 2 years and 5 years. These data would be important to inform patients considering gender-affirming mastectomy.

At baseline, in both groups, we could collect all outcomes, except for physical complications. We could also collect demographic data such as age, gender identity, race and ethnicity, diagnosis of gender dysphoria, age at diagnosis of gender dysphoria, psychiatric comorbidities (e.g., attention deficit hyperactivity disorder, obsessive-compulsive disorder, bipolar disorder), overall psychological functioning, physical comorbidities (e.g., hormonal imbalance, infertility, hypothyroidism) socioeconomic status, and family and peer support. The aforementioned variables are known confounders in our population of interest; therefore, collecting these data will assist in controlling for these confounders in the analysis. Further, at baseline, in both groups, we would collect data pertaining to receipt of any co-interventions to manage gender dysphoria, specifically data related to psychotherapy, social gender transition, puberty blockers, hormone therapy, and other gender-affirming treatments. These co-interventions are also possible confounders that will need to be accounted for in the analysis. As the few prospective cohort studies that we included in our systematic reviews and meta-analyses about gender dysphoria were rated as having high risk of bias due to confounding, it would be critically important to diligently address all known confounders at the analysis stage.

At baseline, in the gender-affirming mastectomy group, we could collect information pertaining to the treatment protocol, immediately before, during and after surgery. Data

pertaining to the variability in the treatment protocol among the included clinics could assist in explaining any potential heterogeneity in results. At baseline, in the control group, we could ask participants whether they desire to receive gender-affirming mastectomy, and if so, when they plan to undergo this procedure. At 6 months, 1 year, 2 years, and 5 years, we could assess all outcomes using the selected tools in both groups. This prospective study design could contribute high certainty evidence assessing patient-important outcomes of interest in individuals experiencing gender dysphoria that is currently missing from the literature.

CONCLUSION

We used standard, intermediate, and advanced methods to create evidence syntheses to assess interventions for the management of acute dental pain and gender dysphoria. In terms of advanced methods, as numerous RCT were available to assess the comparative effectiveness of acetaminophen, NSAIDs, and opioids, we performed a systematic review and network meta-analysis. We used GRADE guidance for network meta-analyses and used an automated tool to rate the certainty of the evidence for direct, indirect, and network estimates of effect. For interpretation and clarity of presentation, due to the abundance of evidence, we classified the interventions from the most to the least effective in accordance with the magnitude of effect and the certainty of the evidence. We organized this information into tables using a colour coding system that considered the certainty of the evidence. Because numerous outdated systematic reviews and meta-analyses about the effects of corticosteroids have been published, we searched the Epistemonikos database and the LOVE platform that utilizes artificial

intelligence. Research about gender dysphoria may be subject to controversy, therefore we devised a plan for minimization and management of conflicts of interest to ensure integrity while performing this work. The rapid expenditure of dental and gender medicine fields is challenging researchers to continually develop methods for conducting evidence syntheses and use these methods appropriately to yield authentic results.

REFERENCES

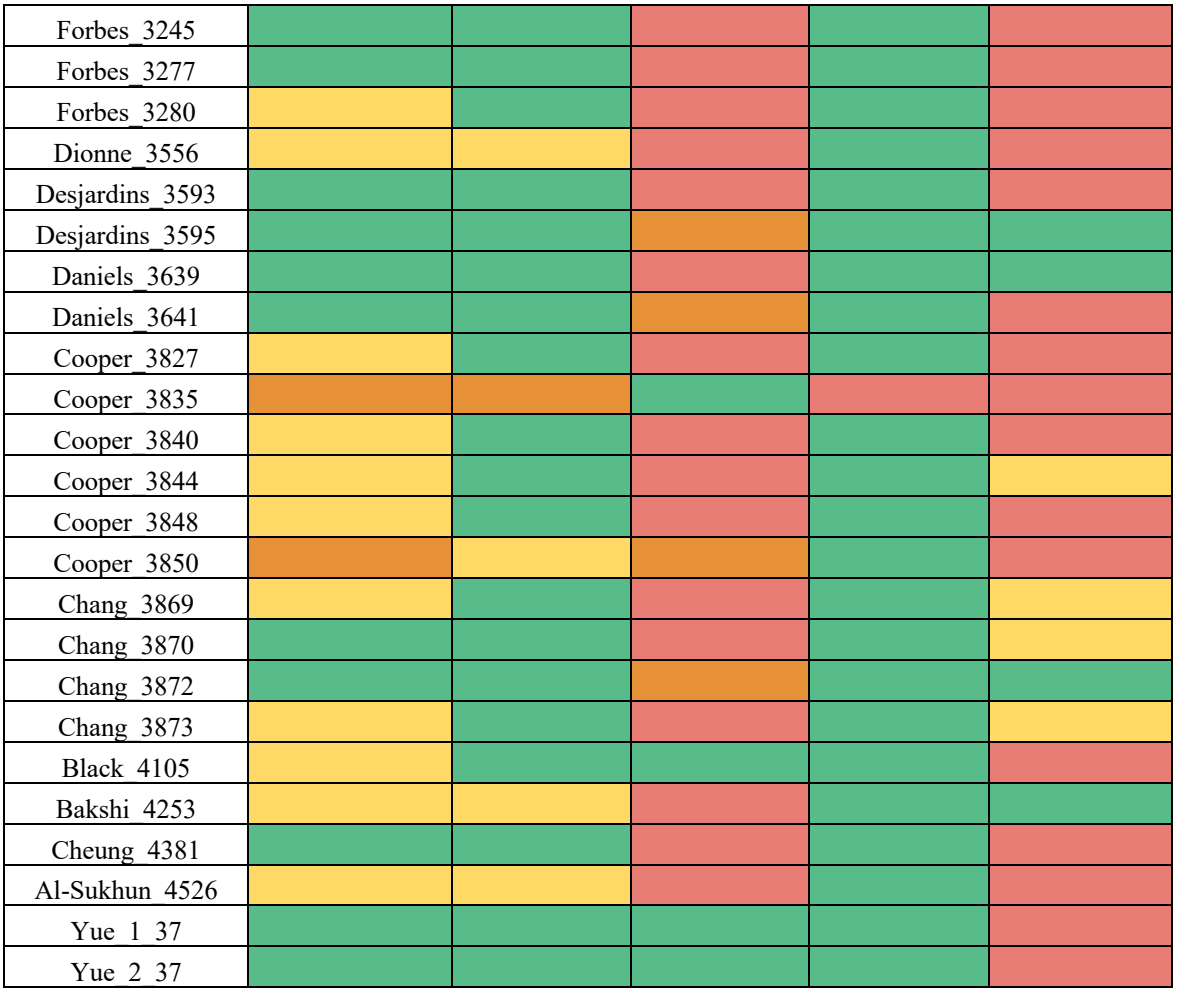
1. Higgins, J., Lasserson, T., Thomas, J., Flemyng, E., Churchill, R., *Methodological Expectations of Cochrane Intervention Reviews*. Cochrane: London.
2. Page, M.J., et al., *The PRISMA 2020 statement: an updated guideline for reporting systematic reviews.* Bmj, 2021. **372**: p. n71.
3. Shea, B.J., et al., *AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both.* BMJ, 2017. **358**: p. j4008.
4. Whiting, P., et al., *ROBIS: A new tool to assess risk of bias in systematic reviews was developed.* J Clin Epidemiol, 2016. **69**: p. 225-34.
5. Lipton, J.A., J.A. Ship, and D. Larach-Robinson, *Estimated prevalence and distribution of reported orofacial pain in the United States.* J Am Dent Assoc, 1993. **124**(10): p. 115-21.
6. Locker, D. and M. Grushka, *The impact of dental and facial pain.* J Dent Res, 1987. **66**(9): p. 1414-7.
7. Locker, D. and M. Grushka, *Prevalence of oral and facial pain and discomfort: preliminary results of a mail survey.* Community Dent Oral Epidemiol, 1987. **15**(3): p. 169-72.
8. Naz Khan, F., *A History of Transgender and Gender Diverse Health Care: From Medical Mistreatment to Gender-Affirmative Health Care*, in *Transgender and Gender Diverse Health Care: The Fenway Guide*, A.S. Keuroghlian, J. Potter, and S.L. Reisner, Editors. 2022, McGraw Hill: New York, NY.
9. Kennel, K.A., B.A. Swiglo, and V.M. Montori, *CHAPTER 69 - Evidence-Based Osteoporosis Care*, in *Osteoporosis (Third Edition)*, R. Marcus, et al., Editors. 2008, Academic Press: San Diego. p. 1629-1649.
10. Ashley, F., et al., *Randomized-controlled trials are methodologically inappropriate in adolescent transgender healthcare.* International Journal of Transgender Health, 2024. **25**(3): p. 407-418.

11. Probst, P., et al., *Blinding in randomized controlled trials in general and abdominal surgery: protocol for a systematic review and empirical study.* Systematic Reviews, 2016. **5**(1): p. 48.
12. Horton, C., *"It felt like they were trying to destabilise us": Parent assessment in UK children's gender services.* International Journal of Transgender Health, 2023. **24**(1): p. 70-85.
13. Horton, C., *Experiences of Puberty and Puberty Blockers: Insights From Trans Children, Trans Adolescents, and Their Parents.* Journal of Adolescent Research, 2022. **39**(1): p. 77-103.
14. Carlile, A., E. Butteriss, and A.P. Sansfaçon, *"It's like my kid came back overnight": Experiences of trans and non-binary young people and their families seeking, finding and engaging with clinical care in England.* International Journal of Transgender Health, 2021. **22**(4): p. 412-424.
15. Tordoff, D.M., et al., *Factors Associated with Time to Receiving Gender-Affirming Hormones and Puberty Blockers at a Pediatric Clinic Serving Transgender and Nonbinary Youth.* Transgender Health, 2022. **8**(5): p. 420-428.
16. Lett, E., et al., *Ethnoracial inequities in access to gender-affirming mental health care and psychological distress among transgender adults.* Social Psychiatry and Psychiatric Epidemiology, 2022. **57**(5): p. 963-971.
17. Everhart, A.R., L. Ferguson, and J.P. Wilson, *Construction and validation of a spatial database of providers of transgender hormone therapy in the US.* Social Science & Medicine, 2022. **303**: p. 115014.
18. Cohen-Kettenis, P.T., et al., *Puberty suppression in a gender-dysphoric adolescent: a 22-year follow-up.* Arch Sex Behav, 2011. **40**(4): p. 843-7.
19. de Vries, A.L., et al., *Puberty suppression in adolescents with gender identity disorder: a prospective follow-up study.* J Sex Med, 2011. **8**(8): p. 2276-83.
20. Steensma, T.D., et al., *Factors associated with desistence and persistence of childhood gender dysphoria: a quantitative follow-up study.* J Am Acad Child Adolesc Psychiatry, 2013. **52**(6): p. 582-90.
21. Walker, V., et al., *Reading and conducting instrumental variable studies: guide, glossary, and checklist.* BMJ, 2024. **387**: p. e078093.

Appendix 1: Pain relief risk of bias assessment for the systematic review and meta-analysis to assess the comparative effectiveness of acetaminophen, non-steroidal anti-inflammatory drugs, and opioids in adults with acute postoperative pain due to dental extraction.

Low
Probably low
Probably high
High

Author_Study ID	Randomization	Deviations from the intended intervention	Missing outcome data	Measurement of outcome	Selection of the reported results
Pain Relief at 6 Hours					
Zelenakas_10					
Sunshine_287					
Ziccardi_403					
Skoglund_418					
Seymour_574					
Schwaratz_622					
Schou_630					
VanDyke_685					
Qi_1102					
Olson_1334					
Mehlisch_1736					
Mehlisch_1737					
Mehlisch_1743					
Morrison_2161					
Moore_2191					
Moore_2197					
Kyselovic_2275					
Kubitzek_2309					
Kiersch_2412					
Malmstrom_2464					
Malmstrom_2465					
Malmstrom_2466					
Malmstrom_2467					
Malmstrom_2468					
Kellstein_2518					
Hersh_2859					
Hersh_2863					
Hersh_2868					
Giglio_3103					
Gay_3155					
Fricke_3213					
Fricke_3215					
Forbes_3244					



Appendix 2: Summary of benefit outcomes compared with placebo (no treatment).

	Pain relief	TOTPAR (Total Pain Relief)	SPID (Summed Pain Intensity Difference)	Global Efficacy Rating	Rescue Analgesia
Time point	6 hours	6 hours	6 hours	6 hours	6 hours
Scale	0 (none) – 4 (complete) [c]	(0–24)- higher better [d]	18 points- higher better [e]	0 (poor) – 4 (excellent) [c]	
Thresholds	-0.4, 0.4	-2.4, 2.4	-1.8, 1.8	-0.4, 0.4	-8, 8
Placebo[a]	0.62	4.1	0.345	0.69	80 per 100
Ibuprofen 200-400 mg plus Acetaminophen 500-1,000 mg	1.68 (1.06 to 2.31)	11.07 (8.23 to 13.91)	4.41 (5.78 to 3.04)	-	-55.60 (-70.27 to -31.22)
Oxycodone 5 mg or Codeine 60 mg	0.10 (-0.06 to 0.25)[b]	1.13 (0.17 to 2.09)[b]	0.78 (0.02 to 1.55)	0.23 (-0.14 to 0.61)	-3.64 (-20.49 to 7.57)

Acetaminophen 650 mg plus Oxycodone 10 mg	1.19 (0.85 to 1.54)	7.91 (6.49 to 9.32)	5.54 (5.26 to 6.02)	1.76 (1.35 to 2.18)	-45.18 (-62.93 to -22.10)
Ibuprofen 400 mg (fast acting or acid)	1.31 (1.17 to 1.45)	8.65 (7.82 to 9.48)	5.58 (4.85 to 6.31)	1.47 (1.27 to 1.68)	-43.01 (-49.50 to -36.02)
Tramadol 37.5 mg plus Acetaminophen 325 mg	0.01 (-0.34 to 0.36)[b]	-	-	-	-
Acetaminophen 500-1,000 mg	0.42 (0.23 to 0.62)	4.20 (3.30 to 5.09)	2.95 (2.31 to 3.60)	0.85 (0.65 to 1.06)	-24.00 (-32.02 to -16.30)
Acetaminophen 600-650 mg plus Codeine 60 mg	0.49 (0.27 to 0.71)	5.03 (4.04 to 6.03)	2.92 (2.32 to 3.53)[b]	0.98 (0.72 to 1.25)	-21.20 (-32.13 to -11.10)[b]
Naproxen 400-440 mg	1.44 (1.07 to 1.80)	8.47 (6.15 to 10.79)	5.27 (3.50 to 7.03)[b]	-	-51.49 (-64.71 to -33.31)
Ibuprofen 200 mg plus Hydrocodone 5 mg	-	-	-	-	-
Hydrocodone 5 mg plus Acetaminophen 300-325 mg	-	-	-	-	-

a The expected risk of each outcome with placebo is reported in the grey row. Numbers in the coloured cells are the estimated mean differences (95% CI) or risk differences (95% CI) per 100 patients when compared to placebo.
b The best estimate of effect was obtained from direct evidence.
Empty cells: there was no evidence for the specific intervention.
TOTPAR: total pain relief, SPID: sum of pain intensity differences
c We used this scale range as it was the most reported scale for this outcome among the included studies.
d The range of possible scores ranged from 0 to 24.
e The range of possible scores ranged from -6 to 12, a total length of 18 points.

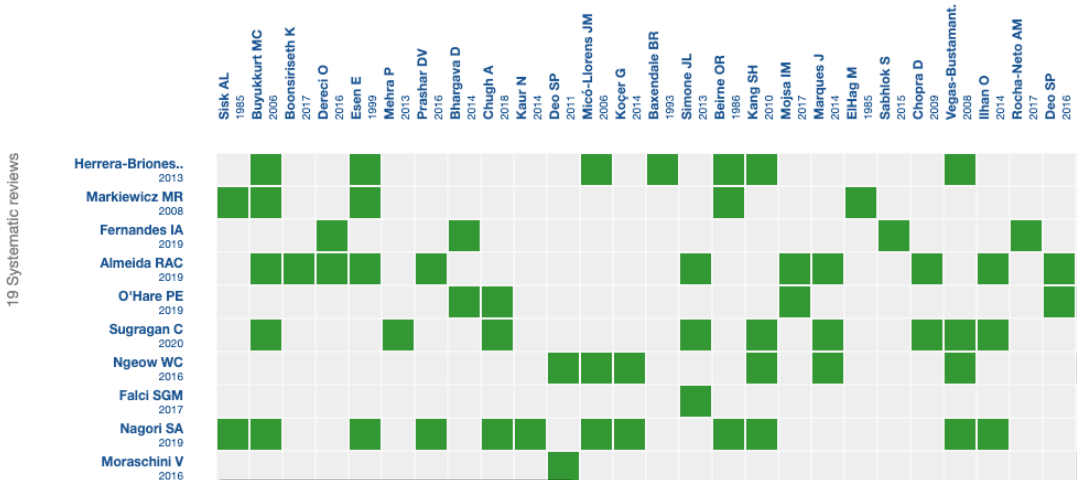
Legend		
	BENEFIT OUTCOMES	
	High/Moderate certainty evidence	Low/Very low certainty evidence
AMONG THE BEST	Better than placebo and some alternatives	May be better than placebo and some alternatives
INTERMEDIATE	Better than placebo, but no better than any alternatives	May be better than placebo, but no better than any alternatives
AMONG THE WORST	No better than placebo	May be no better than placebo

Appendix 3. Risk of bias assessment for the systematic review and meta-analysis to assess the effects of corticosteroids for managing acute pain subsequent to surgical extraction of mandibular third molars.

Low
Probably low
Probably high
High

Study ID	Randomization	Deviations from the intended intervention	Missing outcome data	Measurement of outcome	Selection of the reported results
Pain at 6 Hours					
Afkan_2018A					
Afkan_2018B					
Bauer_2012					
Bortoluzzi_2013 A					
Bortoluzzi_2013 B					
Mojsa_2017A					
Mojsa_2017B					
Simone_2013					
Buyukkurt_2006					
Vegas-Bustamante_2008					

Appendix 4. Sample image of the Epistemonikos matrix of evidence for the systematic review and meta-analysis to assess the clinical effectiveness of corticosteroids for the management of acute pain subsequent to surgical extraction of mandibular third molars. Nineteen systematic reviews were found. The rows list the systematic reviews, and the columns list primary studies (randomized controlled trials) included in each systematic review. This image includes a sample of systematic reviews found and a sample of primary studies included in each systematic review.



Appendix 5. Assessment of risk of bias in observational studies (Gender-affirming hormone therapy): Modified version of ROBINS-I.

This modified version of the ROBINS-I tool is intended for assessment of risk of bias in observational studies focusing on gender-affirming hormone therapy. This modified version includes ROBINS-I domains that are important (as determined by methodologists and clinical specialists in this field) for capturing risk of bias in comparative studies, case series, and before and after studies focusing on gender-affirming hormone therapy.

Domains used to assess risk of bias in each study designs.

Domain	**Low**	**Moderate**	**Serious**	**Critical**
Domains used to assess risk of bias in comparative studies				
Confounding	Adjusted for the following confounders: 1) Psychiatric interventions AND 2) One or more of the following: a. Mental health condition comorbidities b. Socioeconomic status c. Family support/function	Adjusted for the following confounders: 1) Two or all of the following: a. Mental health condition comorbidities b. Socioeconomic status c. Family support/function AND Did not adjust for the following: 1) Psychiatric interventions	Adjusted for the following confounders: 1) One of the following: a. Mental health condition comorbidities b. Socioeconomic status c. Family support/function AND Did not adjust for the following: 1) Psychiatric interventions	Did not adjust for any of the following confounders: 1) Psychiatric interventions 2) Mental health condition comorbidities 3) Socioeconomic status 4) Family support/function

Classification of intervention	Gender-affirming hormone therapy was recorded prospectively or from medical records	Asked participants to recall whether they received gender-affirming hormone therapy	Assumed that the participants received gender-affirming hormone therapy based on clearly reported information in the manuscript	Assumed that participants received gender-affirming hormone therapy based on vaguely reported information in the manuscript
Deviation from intended interventions	No cointerventions received or important cointerventions[1] were equally balanced between the study groups	-	-	Important cointerventions were unbalanced between the study groups
Missing data	More than 90% of patients who started the study provided outcome data	Between 90-70% of patients who started the study provided outcome data	Between 70-50% of patients who started the study provided outcome data	Less thank 50% of patients who started the study provided outcome data
Measurement of outcome	Outcome was measured with an appropriate tool[2] and all in the same way in both groups	Outcome was measured in the same way in both groups, but it was not clear whether the tool was appropriate for this outcome	Outcome was measured with a tool that was appropriate for this outcome, but it was unclear whether it was used in the same way in both groups	Outcome was measured with a tool that was not appropriate for this outcome, and the tool was not used in the same way in both groups
Domains used to assess risk of bias in case series without follow up				
Representativeness of the sample	Included all consecutive patients, eg., visiting a clinic over a specific period	Included all consecutive patients but one sample characteristic (e.g., receipt of psychotherapy, supportive family, high socioeconomic status) was related to the prognosis after gender-affirming hormone therapy	Included all consecutive patients but multiple sample characteristics (e.g., receipt of psychotherapy, supportive family, high socioeconomic status) were related to the prognosis after gender-affirming hormone therapy	Included a highly selected sample based on specific characteristics outlined by the authors that were related to the prognosis after gender-affirming hormone therapy
Classification of intervention	Gender-affirming hormone therapy were recorded prospectively or from medical records	Asked participants to recall whether they received gender-affirming hormone therapy	Assumed that the participants received gender-affirming hormone therapy based on clearly reported information in the manuscript	Assumed that participants received gender-affirming hormone therapy based on vaguely reported information in the manuscript
Deviation from intended interventions	None of the participants received cointerventions[1] that could influence measured outcomes	Less than 50% of patients received cointerventions[1] that could influence measured outcomes	Between 50% and 90% of patients received cointerventions[1] that could influence measured outcomes	More than 90% of patients received cointerventions[1] that could influence measured outcomes
Missing data	More than 90% of included	Between 90-70% of included participants	Between 70-50% of included participants	Less than 50% of included participants

	participants provided outcome data	provided outcome data	provided outcome data	provided outcome data
Measurement of outcome	Outcome was measured with an appropriate tool[2]	–	–	Outcome was measured with a tool that was not appropriate for this outcome
Domains used to assess risk of bias in before-after studies				
Representativeness of the sample	Included all consecutive patients, e.g., visiting a clinic over a specific period	Included all consecutive patients but one sample characteristic (e.g., receipt of psychotherapy, supportive family, high socioeconomic status) was related to the prognosis after gender-affirming hormone therapy	Included all consecutive patients but multiple sample characteristics (e.g., receipt of psychotherapy, supportive family, high socioeconomic status) were related to the prognosis after gender-affirming hormone therapy	Included a highly selected sample based on specific characteristics outlined by the authors that were related to the prognosis after gender-affirming hormone therapy
Classification of intervention	Gender-affirming hormone therapy were recorded prospectively or from medical records	Asked participants to recall whether they received gender-affirming hormone therapy	Assumed that the participants received gender-affirming hormone therapy based on clearly reported information in the manuscript	Assumed that participants received gender-affirming hormone therapy based on vaguely reported information in the manuscript
Deviation from intended interventions	None of the participants received cointerventions[1] that could influence measured outcomes throughout the duration of the study	Less than 50% of patients received cointerventions[1] that could influence measured outcomes throughout the duration of the study	Between 50% and 90% of patients received cointerventions[1] that could influence measured outcomes throughout the duration of the study	More than 90% of patients received cointerventions[1] that could influence measured outcomes throughout the duration of the study
Missing data	More than 90% of patients who started the study provided outcome data	Between 90-70% of patients who started the study provided outcome data	Between 70-50% of patients who started the study provided outcome data	Less thank 50% of patients who started the study provided outcome data
Measurement of outcome	Outcome was measured with an appropriate tool[2] and in the same way before and after the intervention	Outcome was measured in the same way before and after the intervention, but it was not clear whether the tool was appropriate for this outcome	Outcome was measured with a tool that was appropriate for this outcome, but it was unclear whether it was used in the same way before and after the intervention	Outcome was measured with a tool that was not appropriate for this outcome, and the tool was not used in the same way before and after the intervention

1. Psychiatric interventions (e.g., psychotherapy), medical interventions (e.g., gender-affirming surgery).
2. The tool was designed to measure the outcome of interest and has questions judged to be sensible by our content experts (KD, CKM, and SM). For example, the outcome of depression may be measured with PROMIS Depression Scale or ascertained by a licensed psychologist/psychiatrist.

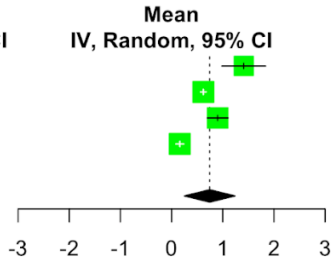
Appendix 6. Sensitivity analysis by correlation coefficient for the systematic review and meta-analysis assessing the effects of gender-affirming hormone therapy for individuals with gender dysphoria below 26 years of age. A sample of one outcome (bone mineral density - hip) is shown below.

Bone Mineral Density - Hip

Bone Mineral Density (z-score) - Hip Following at Least 1 Year (Mean Change, r = 0.2)

Study	Mean	SD	Total	Weight	Mean IV, Random, 95% CI
Schagen 2020A_NF	1.41	0.4868	5	22.1%	1.41 [0.98; 1.84]
Schagen 2020B_NF	0.62	0.1794	37	26.3%	0.62 [0.56; 0.68]
Schagen 2020A_NM	0.90	0.3264	10	25.3%	0.90 [0.70; 1.10]
Schagen 2020B_NM	0.16	0.2047	26	26.3%	0.16 [0.08; 0.24]
Total (95% CI)			**78**	**100.0%**	**0.74 [0.25; 1.24]**

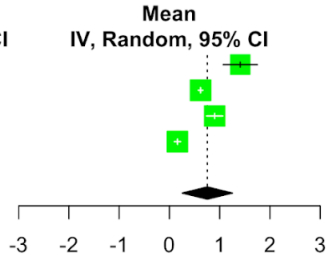
Heterogeneity: $Tau^2 = 0.2415$; $Chi^2 = 120.65$, df = 3 (P < 0.01); $I^2 = 98\%$



Bone Mineral Density (z-score) - Hip Following at Least 1 Year (Mean Change, r = 0.5)

Study	Mean	SD	Total	Weight	Mean IV, Random, 95% CI
Schagen 2020A_NF	1.41	0.3896	5	23.1%	1.41 [1.07; 1.75]
Schagen 2020B_NF	0.62	0.1439	37	25.8%	0.62 [0.57; 0.67]
Schagen 2020A_NM	0.90	0.2599	10	25.2%	0.90 [0.74; 1.06]
Schagen 2020B_NM	0.16	0.1638	26	25.8%	0.16 [0.10; 0.22]
Total (95% CI)			**78**	**100.0%**	**0.75 [0.25; 1.26]**

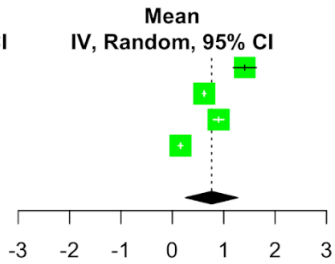
Heterogeneity: $Tau^2 = 0.2527$; $Chi^2 = 188.56$, df = 3 (P < 0.01); $I^2 = 98\%$



Bone Mineral Density (z-score) - Hip Following at Least 1 Year (Mean Change, r = 0.8)

Study	Mean	SD	Total	Weight	Mean IV, Random, 95% CI
Schagen 2020A_NF	1.41	0.2579	5	24.2%	1.41 [1.18; 1.64]
Schagen 2020B_NF	0.62	0.0960	37	25.4%	0.62 [0.59; 0.65]
Schagen 2020A_NM	0.90	0.1690	10	25.1%	0.90 [0.80; 1.00]
Schagen 2020B_NM	0.16	0.1084	26	25.3%	0.16 [0.12; 0.20]
Total (95% CI)			**78**	**100.0%**	**0.76 [0.26; 1.27]**

Heterogeneity: $Tau^2 = 0.2643$; $Chi^2 = 431.45$, df = 3 (P < 0.01); $I^2 = 99\%$



Appendix 7. Management of conflicts of interest for the systematic reviews and meta-analyses summarizing the effects of gender-affirming treatments.

This SR is part of a research project funded through a research agreement between the Society for Evidence-based Gender Medicine (SEGM), the sponsor, and McMaster University. None of the researchers conducting this SR and meta-analysis received financial compensation from the sponsor to conduct this work. The SR and meta-analysis research question was designed through a collaboration between SEGM members, and the methods team based at McMaster University. The rest of the SR and meta-analysis processes (i.e., search and study selection, data extraction, data analyses, manuscript writing, approval of final draft of manuscript) were conducted by independent researchers who do not have any financial or intellectual conflicts of interest or disclosures, and the methods team based at McMaster University (the group of authors). The methods team was solely responsible for the synthesis and interpretation of results, and for drawing conclusions. To minimize bias, a methodologist who was not involved in the data collection, synthesis, and interpretation (GG) ensured that results interpretation was consistent with the findings. This manuscript was drafted by the methods team and approved by all authors, and the sponsor did not have any say nor reviewed its content.

Appendix 8. Assessment of Risk of Bias in Observational Studies (Mastectomy): Modified Version of ROBINS-I

This modified version of the ROBINS-I tool is intended for assessment of risk of bias in observational studies focusing on mastectomy. This modified version includes ROBINS-I domains that are important (as determined by methodologists and clinical specialists in this field) for capturing risk of bias in comparative studies, case series, and before and after studies focusing on mastectomy. Appendix 8 contains 3 tables.

Table 1. Domains used to assess risk of bias in comparative studies.

Domain	Low	Moderate	Serious	Critical
Confounding	Adjusted for the following confounders: 1) Psychiatric interventions AND 2) One or more of the following: a. Mental health condition comorbidities b. Socioeconomic status c. Family support/function	Adjusted for the following confounders: 1) Two or all of the following: a. Mental health condition comorbidities b. Socioeconomic status c. Family support/function AND Did not adjust for the following: 1) Psychiatric interventions	Adjusted for the following confounders: 1) One of the following: a. Mental health condition comorbidities b. Socioeconomic status c. Family support/function AND Did not adjust for the following: 1) Psychiatric interventions	Did not adjust for any of the following confounders: 1) Psychiatric interventions 2) Mental health condition comorbidities 3) Socioeconomic status 4) Family support/function
Deviation from intended interventions	No cointerventions received or important cointerventions[1] were equally balanced between the study groups			Important cointerventions were unbalanced between the study groups
Missing data	More than 90% of patients who started the study provided outcome data	Between 90-70% of patients who started the study provided outcome data	Between 70-50% of patients who started the study provided outcome data	Less thank 50% of patients who started the study provided outcome data
Measurement of outcome	Outcome was measured with an appropriate tool[2] and all in the same way in both groups	Outcome was measured in the same way in both groups, but it was not clear whether the tool was appropriate for this outcome	Outcome was measured with a tool that was appropriate for this outcome, but it was unclear whether it was used in the same way in both groups	Outcome was measured with a tool that was not appropriate for this outcome, and the tool was not used in the same way in both groups

1. Psychiatric interventions (eg., psychotherapy), medical interventions (eg., hormone/puberty blockers, gender-affirming hormones, gender-affirming surgery)
2. The tool was designed to measure the outcome of interest and has questions judged to be sensible by our content experts (KD, CKM, and SM). For example, the outcome of depression was measured with PROMIS Depression Scale or ascertained by a licensed psychologist/psychiatrist.

Table 2. Domains used to assess risk of bias in case series without follow up.

Domain	Low	Moderate	Serious	Critical

Representativeness of the sample	Included all consecutive patients, eg., visiting a clinic over a specific period	Included all consecutive patients but one sample characteristic (eg., receipt of psychotherapy, supportive family, high socioeconomic status) was related to the prognosis after gender-affirming mastectomy	Included all consecutive patients but multiple sample characteristics (eg., receipt of psychotherapy, supportive family, high socioeconomic status) were related to the prognosis after gender-affirming mastectomy	Included a highly selected sample based on specific characteristics outlined by the authors that were related to the prognosis after gender-affirming mastectomy
Deviation from intended interventions	None of the participants received cointerventions[1] that could influence measured outcomes	Less than 50% of patients received cointerventions[1] that could influence measured outcomes	Between 50% and 90% of patients received cointerventions[1] that could influence measured outcomes	More than 90% of patients received cointerventions[1] that could influence measured outcomes
Missing data	More than 90% of included participants provided outcome data	Between 90-70% of included participants provided outcome data	Between 70-50% of included participants provided outcome data	Less than 50% of included participants provided outcome data
Measurement of outcome	Outcome was measured with an appropriate tool[2]			Outcome was measured with a tool that was not appropriate for this outcome

1. Psychiatric interventions (eg., psychotherapy), medical interventions (eg., hormone/puberty blockers, gender-affirming hormones, gender-affirming surgery)
2. The tool was designed to measure the outcome of interest and has questions judged to be sensible by our content experts (KD, CKM, and SM). For example, the outcome of depression was measured with PROMIS Depression Scale or ascertained by a licensed psychologist/psychiatrist.

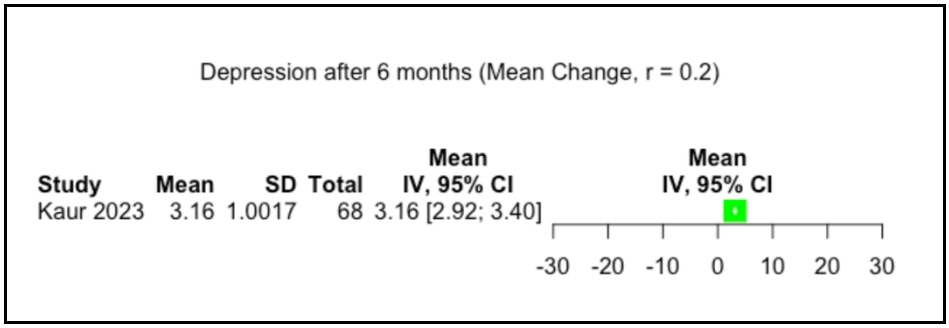
Table 3. Domains used to assess risk of bias in before/after studies.

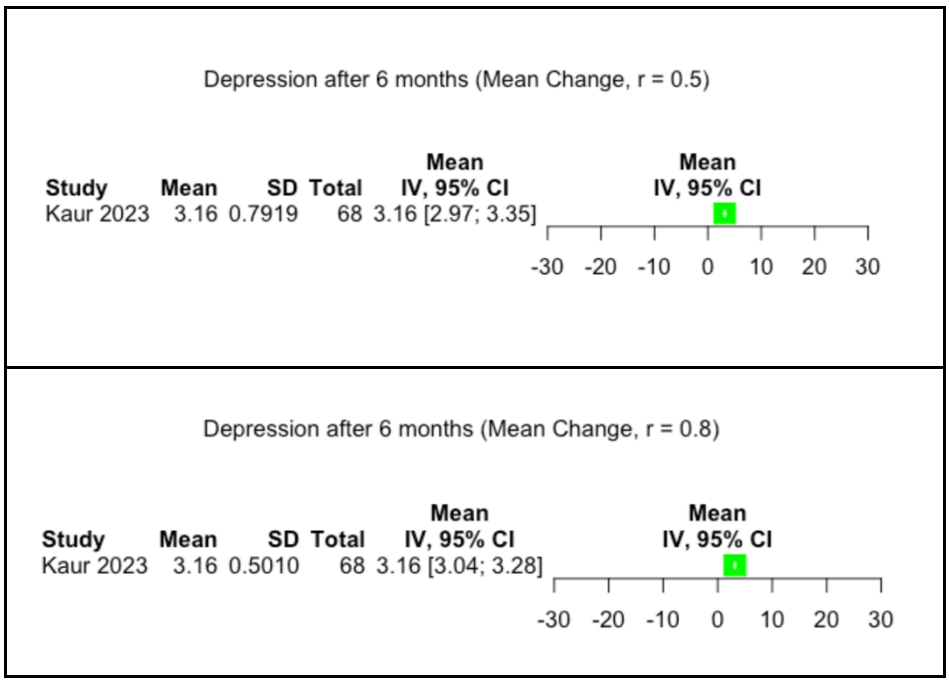
Domain	Low	Moderate	Serious	Critical
Representativeness of the sample	Included all consecutive patients, eg., visiting a clinic over a specific period	Included all consecutive patients but one sample characteristic (eg., receipt of psychotherapy, supportive family, high socioeconomic status) was related to the prognosis after mastectomy	Included all consecutive patients but multiple sample characteristics (eg., receipt of psychotherapy, supportive family, high socioeconomic status) were related to the prognosis after mastectomy	Included a highly selected sample based on specific characteristics outlined by the authors that were related to the prognosis after mastectomy
Deviation from intended interventions	None of the participants received cointerventions[1] that	Less than 50% of patients received cointerventions[1] that	Between 50% and 90% of patients received	More than 90% of patients received cointerventions[1] that

	could influence measured outcomes throughout the duration of the study	could influence measured outcomes throughout the duration of the study	cointerventions[1] that could influence measured outcomes throughout the duration of the study	could influence measured outcomes throughout the duration of the study
Missing data	More than 90% of patients who started the study provided outcome data	Between 90-70% of patients who started the study provided outcome data	Between 70-50% of patients who started the study provided outcome data	Less thank 50% of patients who started the study provided outcome data
Measurement of outcome	Outcome was measured with an appropriate tool[2] and in the same way before and after the intervention	Outcome was measured in the same way before and after the intervention, but it was not clear whether the tool was appropriate for this outcome	Outcome was measured with a tool that was appropriate for this outcome, but it was unclear whether it was used in the same way before and after the intervention	Outcome was measured with a tool that was not appropriate for this outcome, and the tool was not used in the same way before and after the intervention

1. Psychiatric interventions (eg., psychotherapy), medical interventions (eg., hormone/puberty blockers, gender-affirming hormones, gender-affirming surgery)
2. The tool was designed to measure the outcome of interest and has questions judged to be sensible by our content experts (KD, CKM, and SM). For example, the outcome of depression was measured with PROMIS Depression Scale or ascertained by a licensed psychologist/psychiatrist.

Appendix 9. Sensitivity analysis by correlation coefficient for the systematic review and meta-analysis assessing the effects of mastectomy in individuals with gender dysphoria below 26 years of age. A sample of one outcome (depression after 6 months) is shown below.




Depression after 6 months (Mean Change, r = 0.5)
Study Mean SD Total Mean IV, 95% CI
Kaur 2023 3.16 0.7919 68 3.16 [2.97; 3.35]
Mean IV, 95% CI
-30 -20 -10 0 10 20 30
Depression after 6 months (Mean Change, r = 0.8)
Study Mean SD Total Mean IV, 95% CI
Kaur 2023 3.16 0.5010 68 3.16 [3.04; 3.28]
Mean IV, 95% CI
-30 -20 -10 0 10 20 30