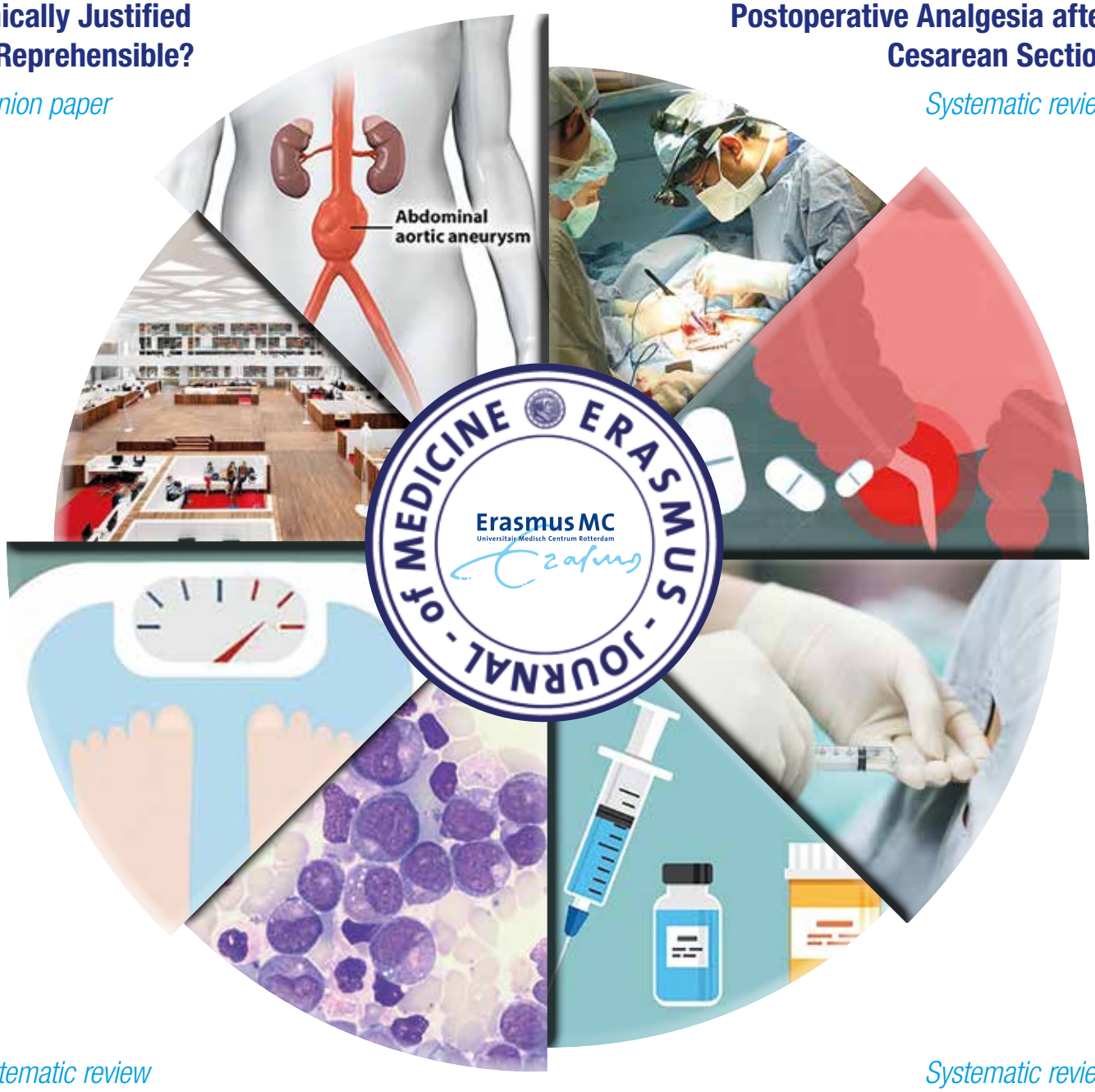
EJM
Erasmus Journal of Medicine
October 2020 • 15th issue
Erasmus Journal of Medicine: independent scientific journal

Screening for Abdominal Aortic Aneurysm in Men aged 65+: Ethically Justified or Reprehensible?

Opinion paper

The Effectiveness of Intrathecal versus Epidural or i.v. PCA Morphine as Postoperative Analgesia after Cesarean Section

Systematic review



Systematic review

Elevated Serum 2-Hydroxyglutarate Levels in Acute Myeloid Leukemia patients with IDH1/2 Mutations

Systematic review

Nonoperative Treatment for Acute Uncomplicated Appendicitis in Children to Reduce Complications as Recurrent Appendicitis

Colophon

The Erasmus Journal of Medicine (EJM) is a scientific magazine by and for students, especially students of Erasmus MC University Medical Center Rotterdam. It was initiated by the MFVR (the medical students' organization of Erasmus MC). We strive to release the journal twice a year. It is published on paper (750 copies) and on the EJM website (link below).

The main purpose EJM is to encourage medical and research master students to conduct research (empirical studies or systematic reviews), report on this research, and become acquainted with the professional publishing process either as authors, reviewers or editors. A secondary purpose is to make the results of excellent student-driven research known to others.
The Journal accepts articles describing original research, systematic reviews, extended abstracts (summaries of recently conducted studies), calls from research projects for students to participate, opinion papers written by students, editorial comments, case reports, clinical lessons, clinical images, and letters to the editor.
The Erasmus Journal of Medicine is funded by Erasmus MC, through an unrestricted grant.

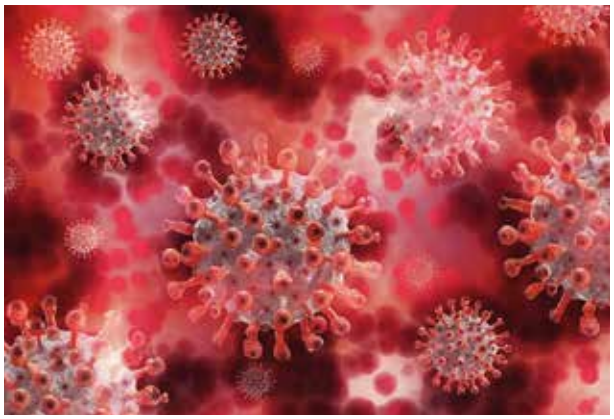
Publisher
Erasmus University Medical Center Rotterdam.

Formatting & Printing
BS Producties, Monnickendam

Correspondence
Erasmus Journal of Medicine
S.B. (Schandra) Lachman, editorial assistant
Room Ae-232
PO Box 2040
3000 CA Rotterdam
E-mail: ejm@erasmusmc.nl

Website
www.eur.nl/erasmusmc/onderwijs/extracurriculair/erasmus-journal-medicine
Like us on www.facebook.com/erasmusjournalofmedicine

Foreword



We are pleased to present this 15th issue of the *Erasmus Journal of Medicine*. The Corona virus had a major impact on the educational and research programs within Erasmus MC. Nevertheless, we have received several nice articles, written by our medical students, which we were happy to accept. The topics that are being addressed vary largely, ranging from the influence of pregnancy on weight loss after bariatric surgery to screening for abdominal aortic aneurysm, and everything in between. We trust, you will enjoy reading.

The current issue appears at the start of a new academic year. Especially our first year bachelor students should be aware of the value of the *Erasmus Journal of Medicine* as an extra-curricular educational instrument, aiming to awake and develop research skills. The doctor of the future not only applies state-of-the art knowledge, but is, ideally, also capable to shift the border between the known and the unknown. That is why we encourage (even our bachelor) students to combine education with research, and to publish their research findings. *The Erasmus Journal of Medicine* provides a low-threshold platform in this respect.

We appreciate that Covid-related measures will continue to influence socio-educational life of our students, despite attempts to soften the effects. However, speaking of 'unknowns', Covid provides many opportunities for medical research, as the SARS-CoV-2 virus affects multiple aspects of human health, which are yet not understood. Several of our (master) students are involved in Covid-related research at Erasmus MC. They are cordially invited to publish their findings in our Journal!

Prof. Maarten Frens, pro-dean Education

Prof. Eric Boersma, chair of the editorial board

Contents

Erasmus Journal of Medicine

FOREWORD 4

EDITORIAL COMMENT

The Effectiveness of Intrathecal versus Epidural or i.v. PCA Morphine as Postoperative Analgesia after Cesarean Section 6
T. Williams, D. Williams, R. Mijnster

OPNION PAPER

Screening for Abdominal Aortic Aneurysm in Men aged 65+: Ethically Justified or Reprehensible? 7
R. Riet

SYSTEMATIC REVIEW

Does Subsequent Pregnancy Influence the Weight Loss of Bariatric Surgery? 10
D. Koornneef, M.Verkouteren, M.Dinkelberg

Elevated Serum 2-Hydroxyglutarate Levels in Acute Myeloid Leukemia patients with IDH1/2 Mutations: a Systematic Review and Meta-analysis 17
N. Guchelaar, L. van den Berg, B.Wouters

Nonoperative Treatment for Acute Uncomplicated Appendicitis in Children to Reduce Complications as Recurrent Appendicitis: a Systematic Review and Meta-analysis 24
H. Koese, L. van Krimpen, L. Oudshoorn

The Effectiveness of Intrathecal versus Epidural or i.v. PCA Morphine as Postoperative Analgesia after Cesarean Section 32
I. de Lange, A. van Opstal, C. van der Marel

The Relation between the 24h Urinary Cortisone Concentration and Obesity: a Systematic Review and Meta-analysis 40
M. Kerolles, A. Alkemade, J. van Osch, B. van der Voorn

INSTRUCTIONS FOR EJM AUTHORS

Instructions for EJM authors 47

The template for authors 49

INSTRUCTIONS FOR EJM REVIEWERS

Advice to the reviewers of EJM 50

The Effectiveness of Intrathecal versus Epidural or i.v. PCA Morphine as Postoperative Analgesia after Cesarean Section

T.D.R.E. Williams[a, b], D.T.R.E. Williams[a,b], R.J.M. Mijnster BSc[a, b]
[a] Medical student, Erasmus MC University Medical Center Rotterdam, the Netherlands
[b] Editorial Board of the Erasmus Journal of Medicine

In this issue of the Erasmus journal of Medicine, a systematic review by De Lange et al. determined the effectiveness of intrathecal (IT) morphine compared to epidural (EP) morphine or intravenous Patient Controlled Analgesia (i.v. PCA) morphine for postoperative analgesia after cesarean section (CS). Up until this moment, a lack of research into this topic on effectiveness and occurrence of adverse events prevents the development of a standardized guideline. This study found no difference in the effectiveness of IT morphine compared to EP morphine and i.v. PCA. IT morphine, however, seems to be more effective than i.v. PCA as postoperative analgesia after a CS. [1]

IT morphine is still considered to be the golden standard for the administration of neuraxial opioids for post-CS-pain management. The main reason for this is the lower dosage that can be administrated through IT morphine, which results in a lower chance of neonatal drug transfer.

A study showed that the optimal dose of IT morphine for all patients has not been determined. EP morphine, on the other hand, has an optimal dose of 2 to 4 mg. [2] So, the likelihood of giving higher dosages than necessary can increase the risk of developing side-effects with the use of IT morphine. Morphine is a dangerous analgetic and there is little flexibility in the dosage that can be used. Therefore, the absence of an optimal dose of IT morphine influences the safety of it. In clinical practices this may translate in dosages with lots of variation and consequently negative side-effects. [3]

All the included studies were of high quality, only one of the included studies was of moderate quality. The definition of the VAS scores, however, was different among the included studies. For instance, one study used a 0-100 VAS, another study used a 1-4 VAS and another study looked at the 0-100 VAS score throughout the 24 hours observation period. This causes a higher (statistical) heterogeneity between studies. Because of this, only four studies were used in the meta-analysis, which is a relatively small amount of studies. All these differences may have influenced the results of this study. We believe that the conclusion of this study might have been different if all studies used a consistent definition of the VAS-scale. If the circumstances were kept the same, one type of morphine administration might have shown statistically better results than another type of morphine administration. Overall, this systematic review is a relevant study which explored the effectiveness and side-effects of three often used methods of morphine administration.

References

1. *I. de Lange and A.E. van Opstal et al. The effectiveness of intrathecal versus epidural or i.v. PCA morphine as postoperative analgesia after cesarean section 2020*
2. *Sutton CD, MD, Carvalho, Brendan, MBBCh, FRCA. Optimal Pain Management After Cesarean Delivery. Anesthesiology Clinics 2016;35(1):107-124.*
3. *Lasagna, Louis, M.D, Beecher, HK, M.D. The Optimal Dose Of Morphine. JAMA Network 1954;156(3):230–234.*

Screening for Abdominal Aortic Aneurysm in Men aged 65+: Ethically Justified or Reprehensible?

Renske van Riet[a]
[a]*Medical student, Erasmus MC University Medical Center Rotterdam, The Netherlands*

Correspondence: Renske van Riet, e-mail: 445318rr@student.eur.nl

Dilemmas concerning screening

"I have to choose between surgery, which can be life-threatening, or no surgery, which can also be life-threatening," says a man who was diagnosed with an abdominal aortic aneurysm (AAA) after screening.[1]

An AAA is a weakening of the aortic wall due to chronic degeneration and inflammation. A ruptured aneurysm can have serious consequences. In the Netherlands, an AAA rupture has a mortality rate of approximately 50%.[2,3]

In Sweden, England and America, an ultrasound screening for AAA has been introduced for men aged 65+ in order to diagnose and treat AAAs at an early stage. Men this age are most at risk for AAA.[2,4]

Before AAA screening can be introduced in the Netherlands, a number of criteria must be met. For example, there must be evidence of effectiveness, the benefits have to outweigh the disadvantages and there must be a balance between the costs and net benefits of screening.[5]

Screening also entails disadvantages such as overdiagnosis and overtreatment. Overdetection and overtreatment of aneurysms that would not have caused any symptoms or would not have been the cause of a person's death, could lead to physical and mental consequences.[6-8]

A number of ethical dilemmas play a role in the introduction of AAA screening. Do the health benefits outweigh the harmful consequences? Can patients make their own informed decision to participate or will their interests be ignored, leading to paternalism? Paternalism is an interference with a person's liberty of action, expressly for the purpose of promoting their welfare, treating them as less than moral equals. At last: does offering screening to men aged 65+ lead to an optimal use of resources in health care or does this result in an uneven distribution of these resources?

The question is: "Is it ethically justified to offer men aged 65 and over a one-time ultrasound screening for AAA in the Netherlands?"

Medical context of AAA

An AAA is an infrarenal aortic diameter of 3 cm or more.[9] Significant risk factors for the development of AAA are: family history of AAA, smoking, male gender and old age.[2,4] AAAs often remain undiagnosed because the majority is asymptomatic. In some patients, an aneurysm rupture will occur, leading to symptoms as abdominal pain, passing out, rapid heart rate and shock. Ruptures are acute and usually fatal (59% to 83% dies before reaching the hospital).[10] In the Netherlands, the mortality rate of AAA ruptures is lower than in other countries, mostly because the risk of dying during or after surgery is lower and partly because more patients reach the hospital in time, since health care is well organized and the distance to reach hospitals is limited.[3]

Patients are eligible for surgery when the AAA has reached a diameter of 55 mm. There are two options for elective surgery, namely the insertion of a pant prosthesis into the aortic bifurcation (open surgery) or the insertion of a stent into the aorta (endovascular aortic repair, EVAR).[11]

According to a recently published report of the Dutch 'Gezondheidsraad', the prevalence of AAA in the Netherlands is 1-2% for men aged 65 and over, which is three to four times higher than the prevalence in women.[3]

Ultrasound is the standard screening method for AAA, due to the sensitivity and specificity of 98%.[12] It is a relatively inexpensive, non-invasive screening method.[9] In the Netherlands, AAA screening costs €4817 for men aged 65+ per year of life gained. There is a 70% chance that screening will be cost-effective if a cost-effectiveness threshold of €20.000 is assumed.[13]

Effectiveness of AAA screening

A reduction in disease-specific mortality was found by four randomized studies.[14-17] After a single screening of men aged ≥ 65 years, a significant reduction in AAA-related mortality of 43% was found after 3-5 years of follow-up. After 13-15 years of follow-up, this reduction was 34%.[18] The greatest effect on survival gains can be seen in men aged 65 to 75, with an absolute risk reduction of AAA-related mortality of 0.16% and a number needed to screen of 625.[19] No reduction in overall mortality was found. [20,21]

When looking at AAA screening in women, there is insufficient evidence for the effectiveness of screening.[14,21]

Individual patient care in the Netherlands

Interestingly, by individual patient care, more AAA patients

are identified in the Netherlands compared to countries with a screening program for AAA. In the Netherlands, AAAs are detected by the GP or a specialist, for example in patients with a high risk of developing AAA by smoking or having (a family member with) cardiovascular diseases. In the past 20 years, the amount of preventive surgeries is strongly increased in the Netherlands and is higher than in other countries, even higher than in countries with a screening program for AAA, like England and Sweden. Since the fact the indication for surgery and the prevalence of AAA is nearly the same in these countries, the Dutch 'Gezondheidsraad' concludes that the detection rate of AAA in the Netherlands is higher.[3]

Negative consequences of AAA screening

AAA screening doubles the number of preventive surgeries, of which the mortality rate is 1,7%.[3,20,22] Complications such as myocardial infarction, renal failure and respiratory failure occur in 3-5% of the surgeries.[3]

Preventive surgery done after AAA screening in the Netherlands, will prevent a maximum of approximately 120-130 extra deaths per year. However, approximately 10-12 men will die because of the consequences of screening and 30 men will get serious complications.[3]

Psychological effects and quality of life

The psychological effects of being diagnosed with an AAA by screening are less easy to quantify than the physical consequences. It has been shown that moderate psychological damage is caused by an AAA diagnosis after screening. Patients can experience shock, fear, insecurity, regret and the burden to prevent others (especially family members) from worrying. However, the current quantitative evidence is insufficient to estimate the frequency and severity of these psychological consequences precisely.[8]

Data on the influence of AAA screening on quality of life vary widely. According to Lindholt et al., the quality of life was 5% lower in men diagnosed with a narrow AAA after screening compared to the control group.[23] This decrease in quality of life is mainly due to the psychosomatic stress caused by less trust in one's own body. According to Bath et al., diagnosing men with AAA by screening results in some mental health problems up to a year after diagnosis, after which recovery takes place again.[24]

Screening criteria

In 1968, The World Health Organization published criteria as a guideline for the introduction of screening methods.[5] These criteria must be met before AAA screening can be introduced in the Netherlands. The AAA screening seems to comply with the criteria for screening. First, AAA is an important health problem, with a prevalence of 1-2% in men aged 65+ in the Netherlands.[3] It is a condition which can lead to serious consequences if an aneurysm rupture occurs.[10] When an AAA is detected early, elective surgery is designated to prevent AAA from rupturing, but only when the risk of rupture is thought to outweigh the operative risk.[2] Furthermore, ultrasound is a non-invasive and inexpensive screening method, so screening in the Netherlands appears to be cost-effective for men aged 65 and over.[9,13]

Ethical aspects

From a utilitarian perspective, screening is acceptable if it results in a net benefit for the population.[25] A number of ethical aspects have to be taken into consideration to determine whether the AAA screening complies with this. The question is whether the screening is proportional: does the health gain offset the negative consequences of screening? Does offering screening to asymptomatic men lead to paternalism, by interfering with their actions and decisions to increase their welfare?

Beneficence vs. non-maleficence

First of all, it has to be taken into consideration if the introduction of AAA screening is proportional, meaning that the positive effects outweigh the negative consequences. On the one hand, screening provides a minimal health gain, but on the other hand AAA results in a real risk of health loss.

A physician has the duty to promote the health of patients wherever possible. Offering a single ultrasound screening to men aged 65 and over leads to some health benefits. A reduction in AAA-related mortality is seen.[18,19] However, the absolute risk reduction is small and there is no evidence that the screening reduces the overall mortality.[20,21]

Screening can also lead to false positive results, resulting in unnecessary examinations, possible health damage and patient anxiety. This is contrary to the principle of non-maleficence. This principle means that there is an obligation not to inflict harm on others.

There is also lead-time bias, since an earlier diagnosis of AAA leads to a longer period of 'being sick'.

Moreover, to avoid a single death with screening, 4 men are diagnosed with an aneurysm that otherwise would never have been detected or would never have caused health problems.[7] Overdiagnosis leads to overtreatment, from which psychological and physical complications, and in some cases even death, can result.

Autonomy vs. paternalism

Secondly, there is a discrepancy between autonomy and motivating people to participate in screening. In the case of screening, autonomy means that patients must be able to make their own, well-informed decision whether or not to participate. Moreover, this decision must be entirely voluntary. It is important that the patient is provided with all the information they need to make this decision.[26] People must be informed about the consequences, such as follow-up examinations, psychological consequences, complications and the mortality risk of treatment. However, the frequency and severity of the psychological effects caused by screening have not been sufficiently quantified, which makes it impossible to inform people sufficiently about this.[8]

From a paternalistic perspective, one can state that the government is obliged to protect the well-being of its citizens. Offering a screening for AAA to prevent a rupture could be part of this. However, offering a screening is done without anyone having asked for it and forces people to make a choice. This can also cause people to feel obliged to participate in the screening. Often, one of the screening goals is to get as many participants as possible, while fully informing potential participants about the risks is insufficient. In this case, it could be stated that the interests of the patient are ignored.

Justice and cost-effectiveness

Finally, it is important that the benefits of screening outweigh the costs. Justice in screening means that offering screening must lead to an optimal use of available health care resources. The introduction of screening may mean that fewer resources, such as money and the deployment of doctors, are available for other forms of health care.[26]

In addition, according to the principle of justice, equal cases must be treated equally. However, this does not mean that everyone should be treated the same. Offering screening to only a certain part of the population can be justified if this group runs a demonstrably increased risk of the disorder. AAA screening only appears to be effective in men aged 65 and older, partly because this group has the highest risk factors.[21] Although cost-effectiveness analyses vary widely, AAA screening in the Netherlands therefore only appears to be cost-effective in this target group.[13] This makes offering screening to other target groups unlikely to be effective.

Conclusions

In conclusion, there is insufficient evidence for the effectiveness of AAA screening in men aged 65 and over in the Netherlands. For this reason, it is not ethically justified to offer this type of screening. Although the screening results in a reduction of AAA-related mortality, this health gain is very limited and it has several adverse consequences. This relative health gain does not outweigh the damage participants may experience as a result of screening, what makes the added value of screening too limited to compensate the negative consequences.

Before the screening eventually can be introduced in the future, more extensive cost-effectiveness analyses are needed. In addition, more quantitative data must be obtained on the physical and mental consequences of AAA screening. Only by providing sufficient information on all important aspects and risks of screening, it is possible that potential participants are able to make an informed decision about their participation.

References

1. *Pettersson M, Hansson A, Brodersen J, Kumlien C. Experiences of the screening process and the diagnosis abdominal aortic aneurysm among 65-year-old men from invitation to a 1-year surveillance. J Vasc Nurs 2017 Jun;35(2):70-77.*
2. *Eckstein HH, Bockler D, Flessenkamper I, Schmitz-Rixen T, Debus S, Lang W. Ultrasonographic screening for the detection of abdominal aortic aneurysms. Dtsch Arztebl Int 2009 Oct;106(41):657-663.*
3. *Gezondheidsraad. Bevolkingsonderzoek naar aneurysma van de abdominale aorta (AAA). 2019:2-3-5.*
4. *Vardulaki KA, Walker NM, Day NE, Duffy SW, Ashton HA, Scott RA. Quantifying the risks of hypertension, age, sex and smoking in patients with abdominal aortic aneurysm. Br J Surg 2000 Feb;87(2):195-200.*
5. *Wilson J, Jungner G. Principles and practice of screening for disease. World Health Organization. Public Health Papers 1968(34).*
6. *Huber TS, Wang JG, Derrow AE, Dame DA, Ozaki CK, Zelenock GB, et al. Experience in the United States with intact abdominal aortic aneurysm repair. J Vasc Surg 2001 Feb;33(2):304-10; discussion 310-1.*
7. *Johansson M, Hansson A, Brodersen J. Estimating overdiagnosis in screening for abdominal aortic aneurysm: could a change in smoking habits and lowered aortic diameter tip the balance of screening towards harm? BMJ 2015 Mar 3;350:h825.*
8. *Cotter AR, Vuong K, Mustelin L, Yang Y, Rakhmankulova M, Barclay CJ, et al. Do psychological harms result from being labelled with an unexpected diagnosis of abdominal aortic aneurysm or prostate cancer through screening? A systematic review. BMJ Open 2017 Dec 12;7(12):e017565-2017-017565.*
9. *Moll FL, Powell JT, Fraedrich G, Verzini F, Haulon S, Waltham M, et al. Management of abdominal aortic aneurysms clinical practice guidelines of the European society for vascular surgery. Eur J Vasc Endovasc Surg 2011 Jan;41 Suppl 1:S1-S58.*
10. *Chaikof EL, Brewster DC, Dalman RL, Makaroun MS, Illig KA, Sicard GA, et al. The care of patients with an abdominal aortic aneurysm: the Society for Vascular Surgery practice guidelines. J Vasc Surg 2009 Oct;50(4 Suppl):S2-49.*
11. *Lederle FA, Kane RL, MacDonald R, Wilt TJ. Systematic review: repair of unruptured abdominal aortic aneurysm. Ann Intern Med 2007 May 15;146(10):735-741.*
12. *Lindholt JS, Vammen S, Juul S, Henneberg EW, Fasting H. The validity of ultrasonographic scanning as screening method for abdominal aortic aneurysm. Eur J Vasc Endovasc Surg 1999 Jun;17(6):472-475.*
13. *Spronk S, van Kempen BJ, Boll AP, Jorgensen JJ, Hunink MG, Kristiansen IS. Cost-effectiveness of screening for abdominal aortic aneurysm in the Netherlands and Norway. Br J Surg 2011 Nov;98(11):1546-1555.*
14. *Ashton HA, Gao L, Kim LG, Druce PS, Thompson SG, Scott RA. Fifteen-year follow-up of a randomized clinical trial of ultrasonographic screening for abdominal aortic aneurysms. Br J Surg 2007 Jun;94(6):696-701.*
15. *Lindholt JS, Sorensen J, Sogaard R, Henneberg EW. Long-term benefit and cost-effectiveness analysis of screening for abdominal aortic aneurysms from a randomized controlled trial. Br J Surg 2010 Jun;97(6):826-834.*
16. *McCaul KA, Lawrence-Brown M, Dickinson JA, Norman PE. Long-term Outcomes of the Western Australian Trial of Screening for Abdominal Aortic Aneurysms: Secondary Analysis of a Randomized Clinical Trial. JAMA Intern Med 2016 Dec 1;176(12):1761-1767.*
17. *Thompson SG, Ashton HA, Gao L, Buxton MJ, Scott RA, Multicentre Aneurysm Screening Study (MASS) Group. Final follow-up of the Multicentre Aneurysm Screening Study (MASS) randomized trial of abdominal aortic aneurysm screening. Br J Surg 2012 Dec;99(12):1649-1656.*
18. *Ali MU, Fitzpatrick-Lewis D, Kenny M, Miller J, Raina P, Sherifali D. A systematic review of short-term vs long-term effectiveness of one-time abdominal aortic aneurysm screening in men with ultrasound. J Vasc Surg 2018 Aug;68(2):612-623.*
19. *Fleming C, Whitlock EP, Beil TL, Lederle FA. Screening for abdominal aortic aneurysm: a best-evidence systematic review for the U.S. Preventive Services Task Force. Ann Intern Med 2005 Feb 1;142(3):203-211.*
20. *Guirguis-Blake JM, Beil TL, Senger CA, Whitlock EP. Ultrasonography screening for abdominal aortic aneurysms: a systematic evidence review for the U.S. Preventive Services Task Force. Ann Intern Med 2014 Mar 4;160(5):321-329.*
21. *Cosford PA, Leng GC. Screening for abdominal aortic aneurysm. Cochrane Database Syst Rev 2007 Apr 18;(2):CD002945. doi(2):CD002945.*
22. *Stather PW, Sidloff D, Dattani N, Choke E, Bown MJ, Sayers RD. Systematic review and meta-analysis of the early and late outcomes of open and endovascular repair of abdominal aortic aneurysm. Br J Surg 2013 Jun;100(7):863-872.*
23. *Lindholt JS, Vammen S, Fasting H, Henneberg EW. Psychological consequences of screening for abdominal aortic aneurysm and conservative treatment of small abdominal aortic aneurysms. Eur J Vasc Endovasc Surg 2000 Jul;20(1):79-83.*
24. *Bath MF, Sidloff D, Saratzis A, Bown MJ, UK Aneurysm Growth Study investigators. Impact of abdominal aortic aneurysm screening on quality of life. Br J Surg 2018 Feb;105(3):203-208.*
25. *Johansson M, Jorgensen KJ, Brodersen J. Harms of screening for abdominal aortic aneurysm: is there more to life than a 0.46% disease-specific mortality reduction? Lancet 2016 Jan 16;387(10015):308-310.*
26. *de Beaufort I, Hilhorst M, Vandamme S, van de Vathorst S, Schermer M. De Kwestie; Praktijkboek ethiek in de gezondheidszorg. . 2nd ed. Den Haag: Boom Lemma Uitgevers; 2011. p. 201.*

Systematic Review

Does Subsequent Pregnancy Influence the Weight Loss of Bariatric Surgery?

D.J.M. Koornneef[a], M.T.A.M. Verkouteren[b]* and M.T. Dinkelberg[c]**

[a] *Health science student, Erasmus University, Rotterdam, The Netherlands*
[b] *Medical student, Erasmus University Medical Center, Rotterdam, The Netherlands*
[c] *Psychology student, Erasmus University, Rotterdam, The Netherlands*
* *These authors contributed equally to this work.*

Abstract

***Objective*: Several factors may induce a difference in weight loss between pregnant and nonpregnant women who have undergone bariatric surgery. This study was performed to better understand the influence of pregnancy on weight loss of bariatric surgery.**

***Methods*: Using our inclusion and exclusion criteria we searched for articles on the academic database PubMed. Studies that measured at least one outcome on weight loss between pregnant and non-pregnant women after bariatric surgery were included. Moreover, only women who delivered a life-born baby were considered in this study. Studies that focused on other organisms than humans, studies not in English, conference papers, review articles and studies without full text available were excluded from the review. The study defined two separate groups: the pregnant group which consisted of women who had been pregnant during the study and the nonpregnant group which consisted of women who not had been pregnant during the study. A significant effect means that the pregnant group lost less weight in comparison to the nonpregnant group. The retrieved articles were independently screened by all the researchers and this was followed by the screening of full-text articles. Thereafter, we assessed the studies using the ErasmusAge quality assessment tool, based on study design, study size, exposure, outcome and adjustments.**

***Results:* A total of seven studies were included in this systematic review. The following four outcome measures for weight loss were compared: body mass index (BMI), current weight, excess weight loss (EWL%) and total weight loss. Therefore, all the studies that were carried out had different results as well because of the different follow-up the studies have. Five studies measured BMI after bariatric surgery multiple times in the follow-up. Four results within the five studies indicated a significant effect. The outcome current weight had one significant effect out of three results, the other two results were non-significant. The outcome EWL% has been investigated in six studies in which it has been measured multiple times. Of the nine results, three were significant. For the total weight loss, one study found a significant result and another found a non-significant result.**

***Conclusion:* There may be a negative influence of pregnancy on the maximal postoperative weight loss after bariatric surgery. However, our study suggests that this is only a short-term effect as two studies with longer follow-up (≥ 60 months) did not show a significant effect in the maximal postoperative weight loss within pregnant and nonpregnant groups. Thus, the advice is to wait two months with stabilized weight or to wait twelve months after surgery, before pregnancy. Furthermore, it is recommended to raise awareness among women about the effects of bariatric surgery in the reproductive age and achieving the maximal postoperative weight loss.**

Keywords

Bariatric surgery; pregnancy; weight loss

Introduction

The prevalence of obesity is a worldwide increasing problem and is affecting approximately one in five women of reproductive age.[1] People with a body mass index (BMI) above 30 kg/m2 are classified as obese. Obesity has a negative effect on the hypothalamic pituitary adrenal axis and the menstrual cycle. [2, 3] Moreover, obesity can cause adverse pregnancy outcomes such as reduced development of the oocyte concerning smaller size or stillbirth.[2, 3] As a result, some obese women are less fertile.[4] Weight loss through bariatric surgery can increase the fertility rate.[5, 6] At the same time, bariatric surgery has some negative effects on maternal and foetal outcomes.[7, 8] Therefore, the recommendation for women who want to be pregnant after bariatric surgery is to wait two months with stabilized weight, or to wait twelve months before they become pregnant.[9] A lot of women become pregnant unexpectedly before or after this waiting advice.[10]

Many studies have researched the effect of bariatric surgery on pregnancy, maternal and foetal health. However, there are only a few studies that researched the effect of pregnancy on outcomes of bariatric surgery. Nevertheless, this can an important is-

sue because of the hormonal influence of a pregnancy. This can change the eating behaviour of pregnant women. Women also have a faster metabolism during the pregnancy which can influence their eating behaviour and this may affect the strict eating regime after bariatric surgery.[11] Secondly, the gastric band of pregnant women who have undergone a gastric band surgery is usually loosened during the pregnancy.[12] This means that there is no gastric restriction anymore and women can therefore experience a longer time before feeling sated, which induces less restricted eating.[13]

Due to these reasons, women who have been pregnant and have undergone bariatric surgery may lose less weight in comparison to women who have never been pregnant during the study and have undergone bariatric surgery. In this case it would be important to better inform women about bariatric surgery in the reproductive age and achieving the maximal postoperative weight loss. Due to these reasons, we postulate that pregnancy has a negative effect on the outcome of bariatric surgery regarding weight loss.

Therefore, to better understand the influence of pregnancy on weight loss after bariatric surgery, this systematic review was performed.

Methods

Search strategy

We searched for articles on the academic database PubMed up until the 3rd of October 2019. The following search strategy was used:

("Pregnancy"[Mesh]) AND ("WeightLoss"[Mesh] OR "Weight Reduction"[tiab] OR "Weight Loss"[tiab]) AND ("Bariatric surgery"[Mesh] OR "Bariatric surgery"[tiab] OR "Metabolic Surgery"[tiab] OR "Bariatric Surgical Procedure"[tiab] OR "Stomach Stapling"[tiab]).

Inclusion criteria

Studies were required to have at least one outcome measure that focused on weight loss between pregnant and non-pregnant women after bariatric surgery. Moreover, only women who delivered a life-born baby were included.

Exclusion criteria

Studies that met the following criteria were excluded from our review:

i. Studies that focused on other organisms than humans;
ii. Studies not in English;
iii. Conference papers;
iv. Review articles;
v. Studies without full text available.

Outcome measures

This systematic review has compared BMI, current weight, excess weight loss % (EWL%), and total weight loss% (Table 3). Two groups were defined: the pregnant group which consisted of women who had been pregnant during the study. Then, the nonpregnant group which consisted of women who not had been pregnant during the study. However, it is unknown if pregnancy could have occurred before bariatric surgery. Furthermore, we distinguished short- and long-term results. Short term was defined as < 60 months after bariatric surgery and long term as ≥ 60 months after bariatric surgery. In this study, a significant effect means that the pregnant group lost less weight in comparison with the nonpregnant group.

Data extraction methodology

Once the literature search was completed, the retrieved articles were independently screened by all the researchers, based on the title and abstract. In doing so, the in- and exclusion criteria were applied. This was followed by the screening of full-text articles, which was done by all authors. Disagreement on the inclusion of an article was resolved by discussion between all the authors.

Study analysis

The features of each study were gathered and described in a table that was designed a priori. This included the author's name, year of publication, country, type of study, aim, patient characteristics including age at time of pregnancy and/or bariatric surgery, the number of patients and the weight loss of the patients, in- and exclusion criteria, time of follow-up, type of surgery, outcomes, limitations and conclusions. The quality of each study was assessed with the ErasmusAge quality assessment tool 2013. This scoring system is based on five items concerning study size, study design, exposure and outcome measurement and whether key confounders were controlled. A total of 0-10 points can be scored, where 10 is representing the highest quality (Table 1).

Table 1 - ErasmusAge quality assessment tool 2013

	Study design:	Study size:	Exposure:	Outcome:	Adjustments:	Total:
Froylich et al.	1	0	2	2	2	7
Rottenstreich et al.	0	1	2	2	2	7
Papastathi et al.	1	1	2	2	1	7
Paham et al.	1	1	2	2	0	6
Alatische et al.	1	0	2	2	1	6
Musella et al.	1	1	2	1	0	5
Haward et al.	1	2	2	2	1	8

Results

Search strategy

Our initial search provided 176 studies. In our final narrative analysis seven studies were included and analyzed. These articles were published between 2011 and 2018. The flow diagram of the study selection is shown in Figure 1. All characteristics of the different studies are shown in Table 2.

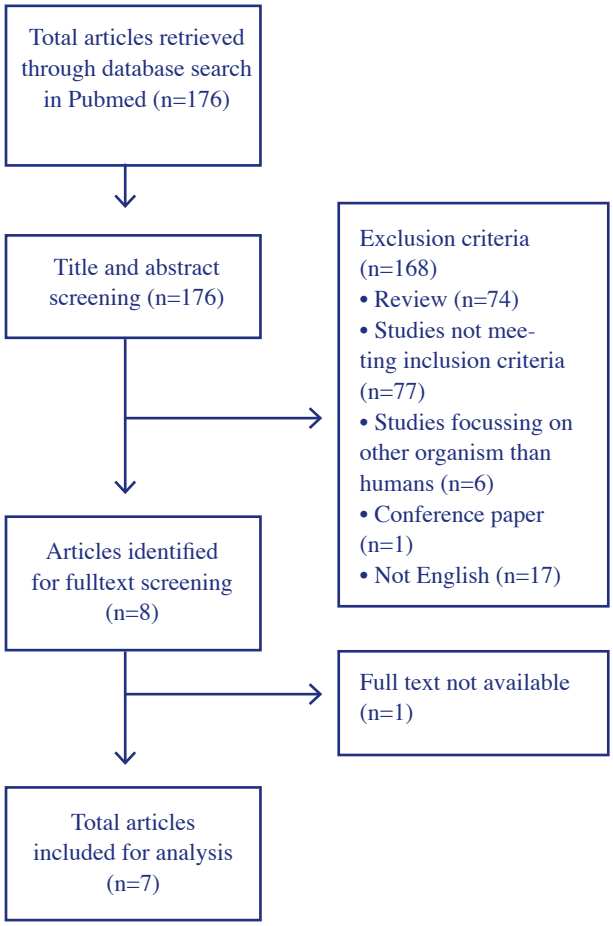
Quality assessment

Each study received points for study design, study size, exposure, outcome and adjustment (Table 1.) The points were based on limitations the researches formed. More detailed information about the limitations are presented in the appendix.

Type of bariatric surgery

A total of 1662 patients were included in this systematic review. The type of bariatric surgery varied between the studies. The following types of surgery were distinguished: adjustable gastric banding, (Roux-en-Y) gastric bypass, sleeve gastrectomy, intragastric balloon insertion, and others. More than half of the patients (55%) underwent adjustable gastric banding. Gastric bypass and sleeve gastrectomy were also frequently used (respectively 29% and 14%). Two groups had a small number of patients (1%) concerning intragastric balloon insertion and other types of bariatric surgery.

Figure 1- Flow diagram of the search strategy and study selection
***Inclusion criteria: studies that have at least one outcome measure that focused on weight loss between pregnant and nonpregnant women after bariatric surgery and women who delivered a life-born baby.**



BMI

BMI is defined by the following formula: weight (kg)/length (m)2. It was used as an indicator of weight loss. The BMI of in total 1270 patients was discussed in five studies.[14-18] Four studies found a significant effect, showing a significant lower BMI in the nonpregnant group, which indicates more weight loss.[14-17] The study of Froylich et al. showed an average BMI of 36.1 kg/m2 in the pregnant group versus 32.9 kg/m2 in the nonpregnant group after 43.9 months (p<.001).[14] Papastathi, et al. had also found a significant effect (p<.0001).[15] In this study, the pregnant group had an average BMI of 35.4 kg/m2 versus 31.1 kg/m2 in the nonpregnant group after 36 months.[15] Furthermore, pregnancy had a significant negative influence on the BMI (p=.025) in the study of Pham et al..[16] The average BMI in the pregnant group was 37.3 kg/m2 versus 35.1 kg/m2 in the nonpregnant group. This result was only found after two years follow-up. After five years follow-up, the differences were nonsignificant (p=.960). In the pregnant group, the average BMI was 37.3 kg/m2 versus 37.4 kg/m2 in the nonpregnant group. The study of Musella et al. had a follow up of 30 months and found a significant result (p<.001). [17] The BMI in the pregnant group was 43.9 kg/m2 versus 34.2 kg/m2 in the nonpregnant group. In the study of Haward et al., two results were non-significant. Before the surgery, the pregnant group had a BMI of 45.1 kg/m2 versus 44.4 kg/m2 in the nonpregnant group (p=.31).[18] After six and half years, the BMI was 36.5 kg/m2 in the pregnant group and in the nonpregnant group, the average was 35.8 kg/m2. This was also non-significant (p=.47).

Current weight (kg)

Current weight is defined as the weight at a particular moment. In this case, it is at the end of the follow up of the studies. Only two studies have used current weight as outcome measure and 675 patients were included. In these studies, only one significant result was found. This means that the pregnant group weighted significantly more than the nonpregnant group. Pham et al. found a significant effect at two years (p=.018).[16] The pregnant group had an average current weight of 102.5 kg and the nonpregnant group had an average current weight of 96.1 kg. However, at five years, this effect disappeared (p=.693). The pregnant group had 103.5 kg as current weight versus 101.8 kg in the nonpregnant group. Haward et al. also found a non-significant effect (p=.65).[18] The pregnant group had a current weight of 99.2 kg versus 97.6 kg in the nonpregnant group after six and a half years.

Excess weight loss % (EWL%)

Excess weight loss is an outcome measure which is measured by the following formula: (initial weight [kg]–current weight [kg])/(initial weight [kg]–ideal weight [kg])×100, with an ideal weight (kg)=25×(height [m])2. This method was used in six studies, with a total of 1552 patients. A significant effect means that the nonpregnant group had a significant higher EWL% than the pregnant group. The significant effects were found in Froylich et al., Papastathi et al. and Pham et al..[14-16] Froylich et al. mentioned an EWL% of 53% in the pregnant group and 68% in the nonpregnant group (p<.001).[14] This means that there is more weight loss in the nonpregnant group. Papasthati et al. found this

Table 2 - Main characteristics of all included studies

Authors	Title	Year of publication	Country of research	Type of study	Time of follow-up	Nonpregnant (n=)	Pregnant (n=)	ErasmusAge Score (points)*
Froylich et al.	The effect of pregnancy before and/or after bariatric surgery on weight loss	2016	United States United	Cohort study, retrospective	43.9 months	92	total: 62, before surgery: 34, after surgery: 24, before and after surgery: 4	7
Rottenstreich et al.	The long-term effect of pregnancy on weight loss after sleeve gastrectomy	2018	States/ Isreal	Cross-sectional case-control study	62.4 months	80	80	7
Papastathi et al.	Impact of pregnancy on weight loss and quality of life following gastric banding	2015	France	Cohort, prospective	36 months	270	61	7
Pham et al.	Does pregnancy influence long-term results of bariatric surgery?	2015	France	Cohort, retrospective	24 and 60 months	519	84	6
Alatishe et al.	Bariatric surgery in women of childbearing age	2013	England	Cohort, retrospective	30 months	211	21	6
Musella et al.	Effect of bariatric surgery on obesityrelated infertility	2012	Italy	Restrospective	30 months	41	69	5
Haward et al.	Does pregnancy increase the need for revisional surgery after laparoscopic adjustable gastric banding?	2011	Australia	Cohort study, prospective	78 months or 24 months or 36 months	189	189	8

For further information see the appendix

result as well (p<.0001).[15] The pregnant group had an EWL% of 64.7 %, while the nonpregnant group had an EWL of 43.6%. Pham et al. found similar results.[16] The study found a significant effect of EWL% on the short term (p=.002). The pregnant group had an average EWL% of 37.3% versus 35.1% in the nonpregnant group after two years. However, this effect diminished and was non-significant (p=.644). Then the EWL% was 45.9% in the pregnant group and 49.9% in the nonpregnant group. Only Haward et al. found no significant effect on both the long-term and short-term.[18] This study distinguished between weight loss after two, three and six and a half years. Two years after bariatric surgery, the pregnant group had an EWL% of 46.2% versus 53.9% in the nonpregnant group (p=.76). Furthermore, there was no significant effect at three years (p=.76). In the pregnant group the EWL% was 47.9% and in the nonpregnant group the EWL% was 47.7%. At six and a half years the pregnant group had an EWL% of 44.0% and 45.8% in the nonpregnant group (p=.56). Also, Rottenstreich et al. found no significance after 62.4 months (p=.77).[19] The EWL% was 74% in the pregnant group and 72% in the nonpregnant group. Finally, Alatische et al. found no effect (p-value not available).[20] The pregnant group had an EWL% of 70.4% versus 70% in the nonpregnant group after 30 months.

Total weight loss %

The last outcome measure is total weight loss. This was measured by: (current weight–start weight)/start weight) x 100%. Only two studies have looked at this outcome measure, which makes a total of 314 patients. Rottenstreich et al. found a non-significant effect after 62.4 months (p=.77).[19] The pregnant group had a total weight loss of 31% versus 30% in the nonpregnant group. However, Froylich et al. found a different result.[14] The pregnant group had a total weight loss of 24% versus 31% in the nonpregnant group after 43.9 months (p<.001). This is a significant result, which means that the pregnant group had less total weight loss than the nonpregnant group.

Discussion

This systematic review of seven articles and a total of 1662 patients shows the influence of pregnancy on weight loss after bariatric surgery. All these studies were compared by weight loss outcomes using BMI, current weight, EWL% and total weight loss %.

Outcomes

We researched five articles about the BMI. Four articles showed a significant difference between the pregnant and nonpregnant group within a range of 3062.4 months. The study of Haward et al. did not show a significant difference between these groups after 6,5 years.[18] The significant difference of the BMI of Pham et al. at two years disappeared after five years.[16] It is important to keep an eye on the fact that BMI is depending on muscles and length, which influences the real weight. Therefore, BMI is unreliable in comparison with other outcome measures.

The current weight was discussed in two studies. The analyses of these articles suggest that there is a significant difference in current weight (kg) after two years, but this difference fade away after five and six and a half years.[16,18]

In six studies, the EWL% was reported. Three studies showed a significant difference in the time range of 36 months till 60 months.[14-16] One study did not show a significant difference after five years.[19] Remarkably, two studies did not show a significant difference of the EWL% after two/two and a half years.[18,20] We first assumed that this was because in this study, only patients were included who had undergone gastric banding while in the other studies often more types of bariatric surgery were included. Nevertheless, the study of Papastathi et al. did show a significant difference and they also only included patients who underwent gastric banding.[15] Based on this, it is concluded that the kind of bariatric surgery has no impact on the influence of pregnancy on weight loss. It is still unclear what caused the non-significant difference at two years in this study, but it may be related to the generalization of the research group or sampling error.

Table 3 - Weight loss among patients who were pregnant and matched nonpregnant patients

	Froylich et al. (14)			Rottenstreich et al. (19)			Papastathi et al. (15)			Pham et al. (16)			Alatishe et al. (20)			Musella et al. (17)			Haward et al. (18)		
	Pregnant	Nonpregnant	Pregnant vs. nonpregnant	Pregnant	Nonpregnant	Pregnant vs. nonpregnant	Pregnant	Nonpregnant	Pregnant vs. nonpregnant	Pregnant	Nonpregnant	Pregnant vs. nonpregnant	Pregnant	Nonpregnant	Pregnant vs. nonpregnant	Pregnant	Nonpregnant	Pregnant vs. nonpregnant	Pregnant	Nonpregnant	Pregnant vs. nonpregnant
BMI* (kg/m2) (N=1270)	3.66 yrs*: 36.1 ± 6.6	3.66 yrs: 32.9 ± 6.9	3.66 yrs: p< .001	-	-	-	3 yrs: 35.4	3 yrs: 31.1	3 yrs: p<.0001	2 yrs: 37.3 ± 6.3; 5 yrs: 37.3 ± 6.9	2 yrs: 35.1 ±7.8; 5 yrs: 37.4 ±8.9	2 yrs: p=.025 5 yrs: p=.960	-	-	-	2.5 yrs: 34.2± 2.4	2.5 yrs; 41.5 ± 2.8	2.5 yrs: p=.001	6.5 yrs; 36.5 ± 7.3	6.5 yrs: 35.8 ± 7.7	6.5 yrs: p=.47
CW* (kg) (N = 675)	-	-	-	-	-	-	-	-	-	2 yrs: 102.5 ± 19.1;	2 yrs: 96.1 ±21.9;	2 yrs: p=.018	-	-	-	-	-	-	6.5 yrs: 99.2 ± 20.3	6.5 yrs: 97.6 ± 21.9	6.5 yrs: p= .65
EWL%* (N= 1552)	3.66 yrs: 53.0 ± 25.0	3.66 yrs: 68.0 ± 26.0	3.66 yrs: p< .001	5.2 yrs: 74%	5.2 yrs: 72%	5.2 yrs: p= .77	3 yrs: 43.6	3 yrs: 43.6	3 yrs: p<.0001	5 yrs: 103.5 ± 20.7 2 yrs: 45.9 ± 24.6%; 5 yrs: 47.7 ± 27.7%	5 yrs: 101.8 ± 25.2 2 yrs: 56.9 ± 28.6% 5 yrs: 49.9 ± 28.9%	5 yrs: p=.693 2 yrs: p =.002 5 yrs: p =.644	2.5 yrs: 70.4%	2.5 yrs: 70.0%	2.5 yrs: Non sig. (p=available	-	-	-	2 yrs: 46,2 ±26.4; 3 yrs: 47.9 ± 29.2; 6.5 yrs: 44.0 ± 28.6	2 yrs: 53.9 ± 44.4; 3 yrs: 47.7 ± 34.3; 6.5 yrs: 45.8 ± 32.7	2 yrs: p=.08 3 yrs: p=.76; 6,5 yrs: p =.56
TWL%*(N = 314)	3.66 yrs: 24.0 ±11.0	3.66 yrs: 31.0±12. 0	3.66 yrs: p< .001	5.2 yrs: 31%	5.2 yrs: 30%	5.2 yrs: p= .77	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

BMI = body mass index; CW= current weight; EWL% = percentage excess weight loss; TWL% = percentage total weight loss

yrs = years, 3.66 yrs = 43.9 months, 5.2 yrs = 62.4 months, 3 yrs = 36 months, 2 yrs = 24 months, 5 yrs = 60 months, 2.5 yrs = 30 months, 6.5 yrs = 78 months

An analysis of two studies reported the total weight loss %. One study of Froylich et al. represented a significant difference after 43.9 months.[14] Rottenstreich et al. did not show a significant difference after 62.4 months.[19] This indicates that the effect disappeared by time.

Study differences

It is important to consider the differences between the studies. The differences in studies are: outcome measures, patients' characteristics, time of pregnancy after bariatric surgery duration of the study and sample size differed between the included studies. The seven studies used different weight outcome measures. Because of the difference in outcome measures, not all results were comparable. Due to the fact that in six of the seven studies EWL% is measured, the EWL% is considered as the most appropriate outcome measure.

There are several factors that have an influence on the validity, such as patients' characteristics. The pregnant group was significantly younger than the nonpregnant group in some studies, but not every study took this into account.[15,16,18,20] In the study of Piers et al., the basal metabolic rate (BMR) was significantly lower in older subjects.[21] This means they had a lower energy consumption, which may indicate less weight loss. However, we do not think this would have changed the results, because Piers et al. study had a larger range of age within their subjects in comparison with the women in our study.[21] All women in our study were in reproductive age and differ less in age range (data not shown).

Remarkably, the study of Froylich et al. showed a significant difference.[14] However, their pregnant study group consisted of patients who became pregnant within three years after or three years before the bariatric surgery. After univariate analysis, this study showed that the only influencing factor for significance was pregnancy before surgery. Thus, women who were pregnant before bariatric surgery lost less weight in comparison with the nonpregnant group. This indicates an effect of pregnancy before bariatric surgery on the maximal postoperative weight loss. In this study, it is unclear what the time range was between moment of pregnancy and surgery. If women became pregnant relatively shortly after bariatric surgery, fluctuating hormone levels may still be influencing the fat cell, which might influence the weight loss. If there was a long period between pregnancy and surgery, weight loss is probably more influenced by the lower physical activity.[22] Furthermore, women could have less motivation and focus on themselves after pregnancy and more focus on their child. These reasons could explain why pregnancy before bariatric surgery would have a negative effect. Still, further investigation is recommended to better understand the effect of pregnancy before surgery. Presumably, women have to wait with surgery if they were pregnant in a certain time range.

The duration of the studies varied between 24 months follow-up and 78 months follow-up, which seemed to influence the results. Studies with a longer follow-up (≥ 5 years) did not show a significant difference between the pregnant group and nonpregnant group while four of six studies with a shorter follow up did represent a significant difference as mentioned above. Because of the elaborated results, we believe that pregnancy can have a negative effect on the maximal postoperative weight loss for short term, but this effect attenuated over time.

Furthermore, the sample size of the studies differed. The study of Alatische et al. investigated a small pregnant group in comparison with the other studies.[20] In addition, Musella et al. had a small control group (n=41, pregnant group n=69).[17] This may have had an effect on the external validity of the results.

Long term versus short term effects

Another point of discussion is the long- and short-term effects of obesity. There are studies which assume that the pregnant group lose less weight on short term compared with the nonpregnant group. In that manner, women who have been pregnant have been more corpulent for a longer time. They have been exposed to the negative influences of obesity concerning, among others, cardio-vascular risks, high blood pressure and inflammation for a longer time than the nonpregnant group. [23-25] The consequence could be that the pregnant group has more complications and side effects of being obese for a longer time. This has not been discussed in this systematic review, but it could be clinically important. To show more clearly whether pregnant women have more side effects during this longer obese time and what consequences are related to this, more research concerning this topic is warranted.

Study limitations

Our study itself had some limitations. The included articles were restricted to articles written in English and that were full text free for Erasmus MC. Therefore, eighteen studies were unreadable for the authors. Also, we may have missed articles that matched our criteria due to the use of only one database (Pubmed). It may be possible that missed studies had different weight loss results. Moreover, the quality assessment (Table 1) in the results is subjective. It is based on the ErasmusAge tool 2013 but was applied subjectively by the researchers.

Furthermore, a meta-analysis was not performed due to the heterogeneity of the data. We recommend to do a statistic analysis such as a meta-analysis, to confirm our thoughts and to strengthen the result this systematic review has.

Conclusion

There may be a negative influence of pregnancy on the maximal postoperative weight loss after bariatric surgery. However, our study suggests that this only takes place on the short, as two studies with longer follow-up (≥ 60 months) did not show a significant effect in the maximal postoperative weight loss within pregnant and nonpregnant groups. This means that our hypothesis is partly substantiated by the results. Thus, based on our result, the current advice for women in the reproductive age who want to undergo a bariatric surgery suffices. This means that women have to wait two months with stabilized weight or have to wait twelve months after surgery, before they become pregnant. Adding to the current advice, it is recommended to raise awareness of the importance of knowledge about bariatric surgery in the reproductive age and achieving the maximal postoperative weight loss.

References

1. Zera C, McGirr S, Oken E. Screening for obesity in reproductive-aged women. Prev Chronic Dis. 2011;8(6):A125.
2. Broughton DE, Moley KH. Obesity and female infertility: potential mediators of obesity's impact. Fertil Steril. 2017;107(4):840-7.
3. Leddy MA, Power ML, Schulkin J. The impact of maternal obesity on maternal and fetal health. Rev Obstet Gynecol. 2008;1(4):170-8.
4. Talmor A, Dunphy B. Female obesity and infertility. Best Pract Res Clin Obstet Gynaecol. 2015;29(4):498-506.
5. Kassir R, Goiset MP, Tiffet O. Pregnancy Following Bariatric Surgery-Medical Complications and Management. Obes Surg. 2016;26(11):2785.
6. Dilday J, Derickson M, Kuckelman J, Reitz C, Ahnfeldt E, Martin M, et al. Sleeve Gastrectomy for Obesity in Polycystic Ovarian Syndrome: a Pilot Study Evaluating Weight Loss and Fertility Outcomes. Obes Surg. 2019;29(1):93-8.
7. Akhter Z, Rankin J, Ceulemans D, Ngongalah L, Ackroyd R, Devlieger R, et al. Pregnancy after bariatric surgery and adverse perinatal outcomes: A systematic review and meta-analysis. PLoS Med. 2019;16(8):e1002866.
8. Rottenstreich A, Elazary R, Goldenshluger A, Pikarsky AJ, Elchalal U, Ben-Porat T. Maternal nutritional status and related pregnancy outcomes following bariatric surgery: A systematic review. Surg Obes Relat Dis. 2019;15(2):324-32.
9. Mahawar KK. Pregnancy and bariatric surgery. Minerva Chir. 2017;72(6):538-45.
10. Weiss HG, Nehoda H, Labeck B, Hourmont K, Marth C, Aigner F. Pregnancies after adjustable gastric banding. Obes Surg. 2001;11(3):303-6.
11. Ladyman SR, Augustine RA, Grattan DR. Hormone interactions regulating energy balance during pregnancy. J Neuroendocrinol. 2010;22(7):805-17.
12. Dixon JB, Dixon ME, O'Brien PE. Pregnancy after Lap-Band surgery: management of the band to achieve healthy weight outcomes. Obes Surg. 2001;11(1):59-65.
13. Doherty C, Maher JW, Heitshusen DS. Prospective investigation of complications, reoperations, and sustained weight loss with an adjustable gastric banding device for treatment of morbid obesity. J Gastrointest Surg. 1998;2(1):102-8.
14. Froylich D, Corcelles R, Daigle CR, Kirwan JP, Brethauer SA, Schauer PR. The effect of pregnancy before and/or after bariatric surgery on weight loss. Surg Obes Relat Dis. 2016;12(3):596-9.
15. Papastathi C, Disse E, Berthiller J, Laville M, Gouillat C, Robert M. Impact of Pregnancy on Weight Loss and Quality of Life Following Gastric Banding. Obes Surg. 2016;26(8):1843-50.
16. Quyen Pham T, Pigeyre M, Caiazzo R, Verkindt H, Deruelle P, Pattou F. Does pregnancy influence long-term results of bariatric surgery? Surg Obes Relat Dis. 2015;11(5):1134-9.
17. Musella M, Milone M, Bellini M, Sosa Fernandez LM, Leongito M, Milone F. Effect of bariatric surgery on obesity-related infertility. Surg Obes Relat Dis. 2012;8(4):445-9. 18. Haward RN, Brown WA, O'Brien PE. Does pregnancy increase the need for revisional surgery after laparoscopic adjustable gastric banding? Obes Surg. 2011;21(9):1362-9.
18. Rottenstreich A, Shufanieh J, Kleinstern G, Goldenshluger A, Elchalal U, Elazary R. The long-term effect of pregnancy on weight loss after sleeve gastrectomy. Surg Obes Relat Dis. 2018;14(10):1594-9.
19. Alatishe A, Ammori BJ, New JP, Syed AA. Bariatric surgery in women of childbearing age. Qjm. 2013;106(8):717-20.
20. Piers LS, Soares MJ, McCormack LM, O'Dea K. Is there evidence for an age-related reduction in metabolic rate? J Appl Physiol (1985). 1998;85(6):2196-204.
21. van Raaij JM. Energy requirements of pregnancy for healthy Dutch women. Eur J Obstet Gynecol Reprod Biol. 1995;61(1):7-13.
22. Wiebe N, Stenvinkel P, Tonelli M. Associations of Chronic Inflammation, Insulin Resistance, and Severe Obesity With Mortality, Myocardial Infarction, Cancer, and Chronic Pulmonary Disease. JAMA Netw Open. 2019;2(8):e1910456.
23. Riaz H, Khan MS, Siddiqi TJ, Usman MS, Shah N, Goyal A, et al. Association Between Obesity and Cardiovascular Outcomes: A Systematic Review and Meta-analysis of Mendelian Randomization Studies. JAMA Netw Open. 2018;1(7):e183788.
24. Dong Y, Jan C, Zou Z, Dong B, Wang Z, Yang Z, et al. Effect of Overweight and Obesity on High Blood Pressure in Chinese Children and Adolescents. Obesity (Silver Spring). 2019;27(9):1503-12.

Appendix

Table 1 - ErasmusAge quality assessment tool 2013

	Study design:	Study size:	Exposure:	Outcome:	Adjustments:	Total:
Froylich et al.	1	0	2	2	2	7
Rottenstreich et al.	0	1	2	2	2	7
Papastathi et al.	1	1	2	2	1	7
Paham et al.	1	1	2	2	0	6
Alatische et al.	1	0	2	2	1	6
Musella et al.	1	1	2	1	0	5
Haward et al.	1	2	2	2	1	8

The points were based on limitations the researches formed. The limitations used are described below.

Study design:

If the study is a cross sectional data selection the study received zero points. Studies with a cohort as study design received one point. There were no randomized control trials in this review, so no study received two points.

Study size:

The studies with less than 60 patients in the pregnant group after bariatric surgery received zero points. If studies included between 60 and 150 pregnant patients, they received one point. Studies with a research group larger than 150 patients in the pregnant group after bariatric surgery received two points.

Exposure:

Every study had a pregnancy group, because of this all the studies received two points

Outcome:

If a study had no weight loss outcome measure, the study received zero points. The studies who had one weight loss outcome measure received one point. Studies with more than two weight loss outcomes measures received two points.

Adjustment:

If the studies did not match on characteristics, they received zero points. If studies matched on characteristic, but had a significance difference in age, they received 1 point. The studies who had no significance difference and matched on characteristics, received 2 points.

Elevated Serum 2-Hydroxy-glutarate Levels in Acute Myeloid Leukemia patients with IDH1/2 Mutations:

a Systematic Review and Meta-analysis

N.A.D. Guchelaar[a], L. van den Berg[a]* and dr. B.J. Wouters[b]*

[a] Medical students, Erasmus University Medical Center, Rotterdam, The Netherlands
[b] Department of Hematology, Erasmus University Medical Center, Rotterdam, The Netherlands
** These authors contributed equally to this work.*

Abstract

Somatic mutations in isocitrate dehydrogenase 1 and 2 (IDH1 and IDH2) occur in 10-20% cases of acute myeloid leukemia (AML). IDH1 and IDH2 are genes expressing key metabolic enzymes and convert isocitrate to alpha-ketoglutarate (αKG). Somatic point mutations in IDH change the enzyme in a way that a NADPH-dependent conversion of αKG to the 'oncometabolite' D-2-hydroxyglutarate (2-HG) occurs. The aim of this systematic review and meta-analysis is to investigate to what extent serum 2-HG levels are elevated in AML patients with IDH1/2 mutations compared to IDH1/2 wildtype patients. We performed a literature search and carried out a meta-analysis of published studies. The primary outcome was serum 2-HG levels. Six published articles are retrieved, describing 693 patients. The meta-analysis shows that AML patients with IDH1/2 mutations have significantly higher serum 2-HG levels with a mean difference of 17.45 μmol/L [95% CI 6.73, 28.16]) as compared to IDH1/2 wildtype patients. Further research on the clinical application of measuring serum 2-HG levels as a predictor of treatment response or prognosis in AML is needed.

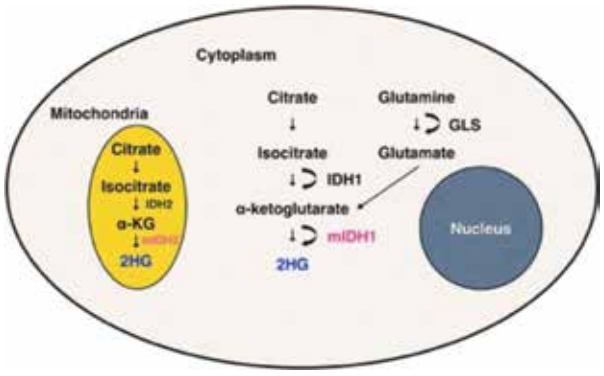
Keywords

acute myeloid leukemia; isocitrate dehydrogenase; 2-hydroxyglutarate; biomarker; survival

Introduction

Acute myeloid leukemia (AML) is a clonal hematopoietic malignancy of progenitor cells with substantial diversities in biological and clinical characteristics (1). AML has an incidence rate of 3.7 cases per 100,000 in Europe and is the most common type of acute leukemia in adults (2). Unfortunately, AML is a disease with a high mortality and poor survival rates (2). Given the poor prognosis, the need for better treatment options is imperative. Beside age, comorbidities and white blood cell counts (WBC), several characteristic gene mutations have shown to be predictive for the prognosis of AML. A number of gene mutations in leukemia cells that contribute to oncogenesis, for instance nucleophosmin (NPM1), have been discovered in recent years (3). Somatic heterozygous mutations in isocitrate dehydrogenase 1 and 2 (IDH1 and IDH2) occur in 10-20% AML-cases and are commonly associated with a normal karyotype (4, 5). Suggesting that IDH mutations are not directly related to AML pathogenesis (6) IDH mutations are not only found in AML, but also in myelodysplastic syndromes and glioma (7-10) IDH1 and IDH2 are genes expressing key metabolic enzymes, and convert isocitrate to alpha-ketoglutarate (αKG). Within the group of IDH mutations, there are different hot spots for IDH1 and IDH2 mutations, i.e. IDH1 R132, IDH2 R172 and IDH2 R140 (3). Somatic point mutations in IDH change the enzyme in a way that a NADPH-dependent conversion of αKG to the 'oncometabolite' D-2-hydroxyglutarate (2-HG) occurs, see Figure 1. (6, 11-13). 2-HG inhibits a number of enzymes whereby tumor growth is stimulated, cellular differentiation is impaired and tumor suppressor proteins are disabled (14-16).

Figure 1. Intracellular conversion of citrate to 2-Hydroxyglutarate by Isocitrate Dehydrogenase mutations (37).



IDH: Isocitrate Dehydrogenase; 2HG: 2-Hydroxy-glutarate; αKG: Alpha-Ketoglutaric acid; GLS: Glutaminase

Inhibition of mutated IDH proteins with specific IDH1 and IDH2 inhibitors leads to a decrease of 2-HG serum levels (17) and induces cellular differentiation of leukemia cells (18). As such ivosidenib and enasidenib inhibit respectively mutated IDH1 and IDH2 leading to cell differentiation in AML patients and resulting in clinical response (19). In Europe, application of these IDH1 and IDH2 inhibitors is still investigated and evaluated by the European Medicines Agency (EMA) before complete implementation in practice is possible (20). Other targeted inhibitors of mutant IDH1 and IDH2 nowadays are developed.

To date, several studies have shown increased levels of serum 2-HG in IDH1/2 mutated AML patients as compared to wildtype patients but series were small and results differed substantially. Therefore, we performed a systematic review and meta-analysis to evaluate the precise extent to which serum 2-HG levels are elevated in an extensive group of IDH1/2 mutated AML patients compared to IDH1/2 wildtype AML patients.

Methods

Research Question

The research question was devised using the Population, Intervention, Comparison and Outcome (PICO) elements (21) and follows: 'To what extent are the blood serum 2-Hydroxyglutarate (2-HG) levels elevated in acute myeloid leukemia patients (AML) with IDH1/2 mutations compared to IDH1/2 wildtype AML patients?'

Search Strategy

For this systematic review and meta-analysis electronic literature searches were conducted using the databases of Embase, Medline Ovid, Web of Science, Cochrane Central and Google Scholar on 7th of June 2019. The publications were restricted to those in English language. The detailed search strategy is presented in Appendix 1. Based upon the selection criteria, titles and abstracts of the search results were screened by two researchers independently and eventually the full-text articles screening was carried out. A manual search of the reference lists of the selected studies was also conducted for additional relevant articles.

Data Extraction and Selection Criteria

For this systematic review, eligible studies met the following inclusion criteria: (1) measurement of blood serum 2-HG levels at diagnosis in patients with AML; (2) IDH1/2 mutations were found in the case group using molecular diagnostics and the control group did not have the IDH1/2 mutations confirmed by the molecular diagnostics; (3) serum 2-HG levels were measured using mass spectrometry, as this is considered as the most reliable measuring method (22); (4) serum 2-HG levels were reported as a median value with upper and lower limits (range). The exclusion criteria were as follows: (1) animal studies, letters and reviews; (2) serum 2-HG values were not categorized into specific mutations but combined with mutations other than IDH.

Unit Conversions

We present the serum 2-HG levels in μmol/L to compare studies among each other. If studies reported the results in a different unit of measurement, such as ng/mL, the values were converted using the molecular mass of 2-HG (MW =148.114 g/mol) (23).

Statistical Analysis

In all studies the serum 2-HG levels were reported as medians with lower and upper limits (range). This made it impossible to perform a meta-analysis because no statistical tests were available to analyze the difference between medians. Thus, we decided to estimate the mean and standard deviation using a method by Hozo et al. for each individual study (24). Subsequently, we performed a meta-analysis using Review Manager Software 5.3 empowered by Cochrane Library. We constructed a forest plot using the random effects model to determine mean differences and corresponding confidence intervals (CI) for serum 2-HG levels. Heterogeneity between included studies was determined using I2 statistics. A funnel plot was also constructed. Values of $p < 0.05$ were considered statistically significant.

Quality Assessment

Quality assessment was done individually by the two researchers and in case of discrepancy a final decision was made through consensus. We modified the QUADAS-2 tool (25) as it did not completely match the required criteria, because it only assesses diagnostic accuracy studies. The adjusted quality scoring system was based on three categories and a maximum score of five points could be given (figure 2). First, we assessed the patient selection by respectively rating the age variation, the male/female ratio, the cytogenetic risk and the IDH mutated/IDH wildtype patient ratio. The second category required the studies to have specified inclusion and exclusion criteria. The final category assessed the reproducibility by determining whether the serum 2-HG determinations were done with the same analytical method and whether the patients were all measured at the same time during treatment. By applying the quality assessment method, it was possible to rate the suitability of the selected articles based on criteria important for our research question. Articles that scored 0 to 1 point were considered to be of poor quality, 2 to 3 points were of moderate quality and 4 to 5 points were of good quality. Details of the quality assessment are available in Appendix 2.

Figure 2. Quality assessment.

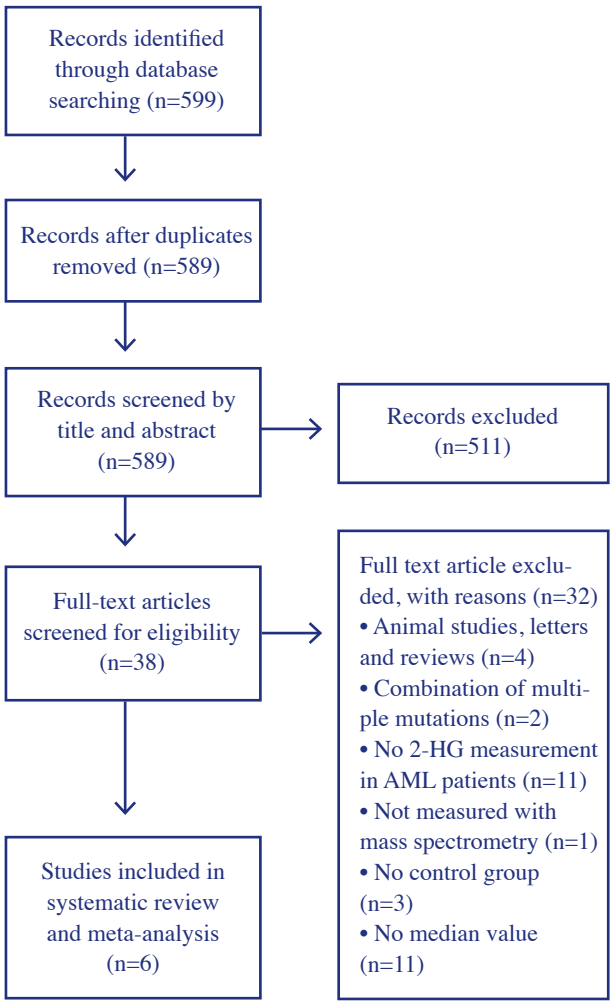
1. Patient selection?	**0 / 1 / 2**
2. Inclusion and exclusion criteria?	**0 / 1**
3. Reproducibility?	**0 / 1 / 2**

Results

Literature Search

The literature search resulted in a total of 589 unique articles. After screening the articles by title and abstract, 551 articles were excluded. The remaining 38 articles were fully analysed for eligibility using the inclusion and exclusion criteria. This resulted in an exclusion of 32 articles. For instance, multiple mutations and not only IDH mutations were studied in two articles (4, 26). Eventually, six studies were included. Figure 3 shows how the articles were selected for this systematic review.

Figure 3- Flowchart of the study selection.



Study Characteristics

Studies were published from 2012 to 2019 and conducted in Germany, United States and France. The number of patients in the studies varied between eight and 223. This resulted in a total pool of 693 patients, from which 236 patients were IDH1/2 mutated and 457 patients were IDH1/2 wildtype. The details of the six included studies are shown in Table 1. Some studies reported the results in μmol/L (27-29) and some studies in ng/mL (17, 30, 31). We converted the results in nl/mL to μmol/L, as described in the method section.

Table 1 - Summary of included studies measuring serum 2-HG levels.

Author	Journal	Total patients	IDH1/2 mutated	IDH1/2 wildtype	Disease	Quality Assessment
Balss et al. (2016)	Leukemia	143	59	84	AML	4
Brunner et al. (2019)	Cancer	195	50	145	AML	4
DiNardo et al. (2013)	Blood	223	62	161	AML	4
Fathi et al. (2012)	Blood	42	9	33	AML	3
Janin et al. (2014)	J Clin Oncol	82	53	29	AML	4
Willekens et al. (2015)	Blood	8	3	5	MS	3

Abbreviations: IDH: Isocitrate Dehydrogenase; AML, acute myeloid leukemia; MS, myeloid sarcoma.

Serum 2-HG Levels

In all included publications measurement of serum 2-HG levels of IDH1/2 mutated patients and of IDH1/2 wildtype patients was performed at diagnosis. A significant mean difference in serum 2-HG levels between the IDH mutated and IDH wildtype groups (MD 118.08 μmol/L [CI 25.18, 210.97], MD 31.90 μmol/L [CI 0.65, 63.15] and MD 19.87 μmol/L [CI 6.70, 33.04]) was shown in three studies. (resp. (17, 27, 31)) The other three studies did not show a significant mean difference in median serum 2-HG levels between the IDH mutated and IDH wildtype groups (MD 8.24 μmol/L [CI -22.72, 39.20], MD 9.00 μmol/L [CI -0.72, 18.72] and MD 20.00 μmol/L [CI -0.43, 40.43]). (resp. (28-30)) Although the results were not statistically significant, the difference was found to be in the same direction. Figure 4 shows the forest plot of the mean differences in serum 2-HG levels between the IDH mutated and IDH wildtype groups. There was a significant mean difference in serum 2-HG level between both groups (MD 17.45 μmol/L [95% CI 6.73, 28.16], p = 0.001). The heterogeneity in the articles was moderate (I2 = 39%). Figure 5 shows the funnel plot of the included studies.

We used the method of Hozo et al. (24) to estimate the means and standard deviations from the reported median values. This study showed that the median can be used to estimate the mean when the sample size is larger than 25. For smaller populations a distinct formula was devised. Means in Fathi et al. (17) and Willekens et al. (28) were determined using this formula.

We also performed a meta-analysis without these two small studies to test the effect of this formula (figure 5). In this case, the meta-analysis showed a significant mean difference in serum 2-HG level between both groups as well (MD 19.92 μmol/L [95% CI 10.03, 29.81], p < 0.0001.

Figure 4- Forest plot: serum 2-HG levels of IDH mutated patients compared to IDH wildtype patients, outcome: Mean Difference.

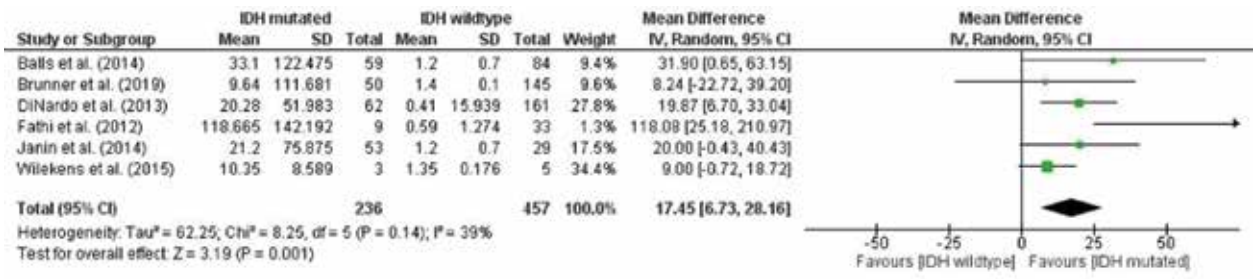
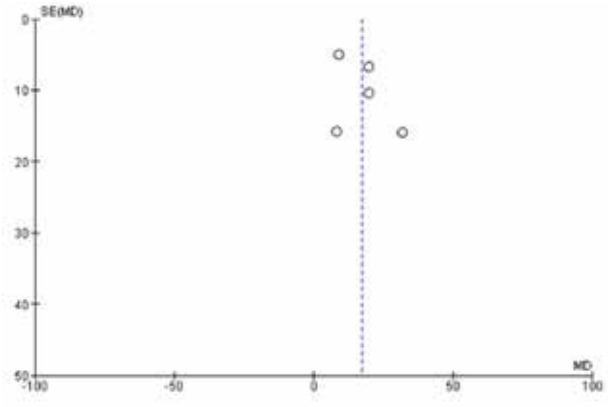


Figure 5- Funnel plot of the included studies.



Differences between IDH Hot Spots

Four studies analyzed the differences in serum 2-HG levels between the three different IDH mutation hot spots (27, 29-31). IDH hot spots included were IDH1 R132, IDH2 R172 and IDH2 R140. Janin et al. (29) reported that total 2-HG serum levels in IDH2 R172 were increased compared to IDH2 R140 and IDH1 R132 (respectively $p = 0.02$ and $p = 0.04$). However, Balss et al. (27) found that 2-HG serum levels in patients with IDH2 R172 were lower than in patients with IDH1 R132 and IDH2 R140 mutations, with a trend to significance ($p = 0.06$). Brunner et al. (30) and DiNardo et al. (31) found no difference in 2-HG levels among patients with IDH mutations based on specific mutation hot spots.

Serum Levels at Complete Remission

In three studies measurement of the serum levels of 2-HG at complete remission (CR) was performed. The studies reported the difference between the serum levels at diagnosis and the serum levels at CR. Janin et al. (29) showed that median serum 2-HG levels at CR (1.3 μmol/L) were significantly lower than at diagnosis (22.4 μmol/L) ($p < 0.01$). DiNardo et al. (31) calculated a median change in 2-HG level from diagnosis to remission of 20.85 μmol/L (range, +0.29 to –90.22 μmol/L). Decreased serum 2-HG levels were also described by Fathi et al. (17) Median serum 2-HG levels at baseline were 12 58 μmol/L and in all evaluated patients who reached complete remission, serum 2-HG levels decreased to <3.38 μmol/L by day 30 and <1.35 μmol/L by day 60.

Prognostic Value

The prognostic value of serum 2-HG was studied in three articles. They followed the patients over a longer time to determine the progression free survival (PFS) and overall survival (OS). In the study of Janin et al. (29) AML patients were treated with standard AML therapy with the addition of gemtuzumab ozogamicin (GO). They found that the serum 2-HG levels (> 2 μmol/L vs ≤ 2 μmol/L) at CR had an impact on the progression free survival as the HR is 4.37 ([1.14-16.8], $p = 0.032$). In the study of Balss et al. (27) patients were treated with chemotherapy followed by allogeneic stem cell transplant. They also found a significant HR for event free survival with 2-HG as a prognostic variable (HR = 1.32 [1.02-1.72], $p = 0.04$). They did not find a significant impact on the OS. Finally, the study of DiNardo et al. (31), where patients were treated with induction chemotherapy followed by consolidation and autologous transplant, also found a non-significant HR of 0.72 ([0.49 – 1.06], $p = 0.09$) for the OS.

Cut Off Value

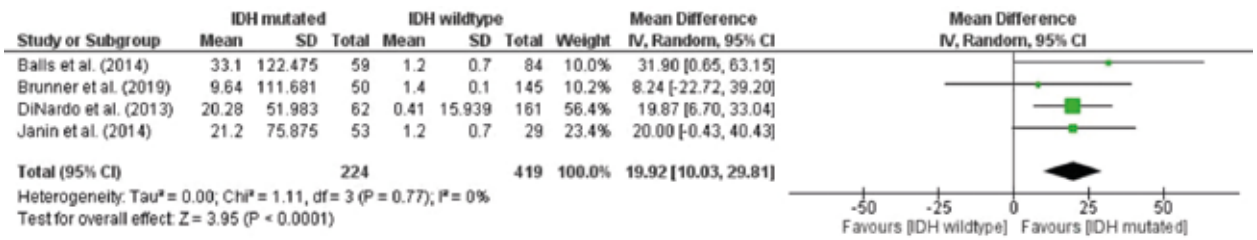
A cut off value at which serum 2-HG level could be useful as a biomarker test to distinguish between AML patients with or without IDH1/2 mutations was tried to be determined in three studies. Janin et al. (29) found a cut-off value of 2.0 μmol/L, where they achieved 100% sensitivity and 89% specificity. DiNardo et al. (31) identified an optimal cut off value of 4.73 μmol/L where the sensitivity and specificity were 86.9% and 90.7%, respectively. A cut off value determined by Brunner et al. (30) of 3.61 μmol/L was reported, with a sensitivity of 80.4% and a specificity of 98.8%.

Discussion

In this systematic review and meta-analysis, we aimed to investigate to what extent blood serum 2-Hydroxyglutarate levels are elevated in all AML patients so far described in literature with IDH1/2 mutations compared to IDH1/2 wildtype AML patients. The meta-analysis of 6 clinical studies showed that blood serum 2-HG levels are significantly higher with a mean difference of 17.45 μmol/L in AML patients with an IDH1/2 mutation in comparison to IDH1/2 wildtype AML patients.

Furthermore, Janin et al. (29) showed that serum 2-HG levels at complete remission are of prognostic value for progression free survival (PFS). Serum 2-HG levels are not clearly different for the three different IDH mutation hot spots (IDH1 R132, IDH2 R172 and IDH2 R140). Finally, a diagnostic cut off value was tried to be determined for identifying an IDH mutation. However, such a conclusive diagnostic cut off value is not yet determined. The cut off values found in the three studies varied substantially among each other.

Figure 6- Forest plot: serum 2-HG levels of IDH mutated patients compared to IDH wildtype patients, outcome: Mean Difference sub-analysis.



The included studies in this review and meta-analysis were rated with a quality assessment. A total of four studies were of good quality as they scored 4 points. The remaining two studies were of moderate quality with a total score of 3 points. This was caused by a smaller study population. In the quality assessment, one of the assessed categories was the study population in order to determine if the study populations were representative. We tested on age variation, male/female ratio, cytogenetic risk profile and IDH mutated/IDH wildtype ratio, to estimate the risk of confounders. One study (17) did not report on details of the included patients. The other studies had homogenous study populations. One study (28) only included patients with myeloid sarcoma instead of patients with acute myeloid leukemia. We included this study as MS is considered a manifestation of AML (1). The total included patients highly differed between the studies. Willekens et al. (28) only included eight patients, whereas the largest study (31) included a total of 223 patients. The heterogeneity in the articles was moderate (I2 = 39%). A funnel plot was constructed to assess the risk of publication bias. One study (17) deviated from the funnel plot, as the mean difference is much larger than the other studies. Apart from that study, there is no indication for publication bias.

All outcomes of the individual studies were given in medians with lower and upper limits. This was done because the data were skewed as there was not a normal distribution. However, this made it impossible to perform a meta-analysis as there were no statistical tests available to perform a meta-analysis with medians with lower and upper limits. Therefore, we considered the possibilities to convert the medians to means with standard deviations. We used the study of Hozo et al. (24) as we found it the most used method to obtain a reliable approximation of the mean. This made it possible to perform a meta-analysis with a scientifically reliable method. They found that the median in groups of more than 25 participants is a good estimation of the mean. This was the case in all but two studies. In the remaining two studies with less than 25 participants a formula given by Hozo et al. (24) was used to estimate the mean. We also did a sub-analysis without the two studies with less than 25 participants to test the effect of this tool on the outcome. However, we found that the mean difference in serum 2-HG levels still remained significantly higher in favor of the IDH1/2 mutated group. It is to be noted that in groups of more than 25 participants with a skewed distribution, the median either over-or underestimates the mean.

We contacted the authors of two articles (8, 12) that did not report median values if they could provide us with the median value. We did not receive an answer, however. We may have missed some usable data by excluding these articles.

We only included studies that measured the serum 2-HG levels by mass spectrometry. We did not however distinguish between different types of mass spectrometry. This could have caused a larger difference between the measured serum 2-HG levels as different studies used other forms of mass spectrometry.

As a secondary outcome three studies (27, 29, 31) investigated the prognostic value of serum 2-HG levels and followed the patients over a longer time to determine the progression free survival (PFS) and overall survival (OS). They found that serum 2-HG levels are a prognostic variable for progression free survival, but not for overall survival. However, treatments differed between studies which could have affected the results. There was a slight difference between induction therapy given to the patients in each study, which could be a potential cause of the differences between serum 2-HG levels.

Clinical Relevance

IDH1/2 mutations are found in approximately 20% of the AML patients (4, 5). Previous studies have shown that serum 2-HG levels in AML patients are elevated when an IDH1/2 mutation is present. However, patient series were small and results differed substantially. This systematic review shows the mean difference of serum 2-HG levels between IDH1/2 mutated and IDH1/2 wildtype AML patients in an extensive group. However, some included studies were small and the mean values were estimated on the basis of reported medians. Despite the fact that this could have affected the results, our findings can serve as a guide for further research before measuring serum 2-HG levels could be implemented in practice as a minimal invasive and fast way to monitor IDH1/2 targeted therapies.

Measuring serum 2-HG levels could also be of value to monitor and predict the efficacy of IDH inhibitors (32). Targeted inhibitors of mutant IDH1 and IDH2, like enasidenib and ivosidenib, are nowadays developed and are already used in practise after approval of the FDA (33-35). A recent study of Stein et al (36) showed that serum 2-HG levels decreased after treatment with enasidenib. However, 2-HG reductions were significantly greater when CR was achieved. Therefore, measuring serum 2-HG levels could be of great value to monitor the therapy effect through for instance measuring Minimal Residual Disease (MRD) or drug resistance (29). However, before measuring serum 2-HG levels could be implemented in the clinic a safe threshold should be determined. Based on our findings, we have not yet found a cut-off value that could be used as a threshold. Further research in a larger group of patients is needed to determine a safe threshold where a differentiation between IDH mutated patients and IDH wildtype patients can be made. Based on the three studies (29-31) that estimated a cut-off value, we expect the cut-off value to be between approximately 2-4 μmol/L.

In conclusion, serum 2-HG levels are significantly elevated with a mean difference of 17.45 μmol/L in acute myeloid leukemia patients with IDH1/2 mutations compared to IDH1/2 wildtype patients. More research is needed to investigate the potential for a clinical application of serum 2-HG measurement in AML patients and an exact diagnostic cut off point needs to be determined before it could be used for diagnostic purposes.

Acknowledgements, The authors would like to thank W.M. Bramer (Biomedical information specialist, Erasmus MC) for his assistance with the search strategy.

No potential conflict of interest was reported by the authors.

References

1. Caligiuri M, Levi MM, Kaushansky K, Lichtman MA, Prchal J, Burns LJ, et al. Williams Hematology, 9E: McGraw-Hill Education; 2015.
2. Visser O, Trama A, Maynadie M, Stiller C, Marcos-Gragera R, De Angelis R, et al. Incidence, survival and prevalence of myeloid malignancies in Europe. European journal of cancer (Oxford, England : 1990). 2012;48(17):3257-66.
3. Mardis ER, Ding L, Dooling DJ, Larson DE, McLellan MD, Chen K, et al. Recurring mutations found by sequencing an acute myeloid leukemia genome. The New England journal of medicine. 2009;361(11):1058-66.
4. Wiseman DH, Small HF, Wilks DP, Waddell ID, Dennis MW, Ogilvie DJ, et al. Elevated plasma 2-hydroxyglutarate in acute myeloid leukaemia: association with the IDH1 SNP rs11554137 and severe renal impairment. Br J Haematol. 2014;166(1):145-8.
5. Sellner L, Capper D, Meyer J, Langhans CD, Hartog CM, Pfeifer H, et al. Increased levels of 2-hydroxyglutarate in AML patients with IDH1-R132H and IDH2-R140Q mutations. Eur J Haematol. 2010;85(5):457-9.
6. Ward PS, Patel J, Wise DR, Abdel-Wahab O, Bennett BD, Coller HA, et al. The common feature of leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity converting alpha-ketoglutarate to 2-hydroxyglutarate. Cancer cell. 2010;17(3):225-34.
7. Kosmider O, Gelsi-Boyer V, Slama L, Dreyfus F, Beyne-Rauzy O, Quesnel B, et al. Mutations of IDH1 and IDH2 genes in early and accelerated phases of myelodysplastic syndromes and MDS/myeloproliferative neoplasms. Leukemia. 2010;24(5):1094-6.
8. Lin P, Luo Y, Zhu S, Maggio D, Yang H, Hu C, et al. Isocitrate dehydrogenase 2 mutations correlate with leukemic transformation and are predicted by 2-hydroxyglutarate in myelodysplastic syndromes. J Cancer Res Clin Oncol. 2018;144(6):1037-47.
9. Yan H, Parsons DW, Jin G, McLendon R, Rasheed BA, Yuan W, et al. IDH1 and IDH2 mutations in gliomas. The New England journal of medicine. 2009;360(8):765-73.
10. Parsons DW, Jones S, Zhang X, Lin JC-H, Leary RJ, Angenendt P, et al. An integrated genomic analysis of human glioblastoma multiforme. Science (New York, N Y). 2008;321(5897):1807-12.
11. Dang L, White DW, Gross S, Bennett BD, Bittinger MA, Driggers EM, et al. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. Nature. 2009;462(7274):739-44.
12. Gross S, Cairns RA, Minden MD, Driggers EM, Bittinger MA, Jang HG, et al. Cancer-associated metabolite 2-hydroxyglutarate accumulates in acute myelogenous leukemia with isocitrate dehydrogenase 1 and 2 mutations. J Exp Med. 2010;207(2):339-44.
13. Pollyea DA, Kohrt HE, Zhang B, Zehnder J, Schenkein D, Fantin V, et al. 2-Hydroxyglutarate in IDH mutant acute myeloid leukemia: predicting patient responses, minimal residual disease and correlations with methylcytosine and hydroxymethylcytosine levels. Leukemia & lymphoma. 2013;54(2):408-10.
14. Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature. 2010;466(7310):1129-33.
15. Kaelin WG, Jr. Cancer and altered metabolism: potential importance of hypoxia-inducible factor and 2-oxoglutarate-dependent dioxygenases. Cold Spring Harbor symposia on quantitative biology. 2011;76:335-45.
16. Dang L, Su S-SM. Isocitrate Dehydrogenase Mutation and (R)-2-Hydroxyglutarate: From Basic Discovery to Therapeutics Development. Annual review of biochemistry. 2017;86:305-31.
17. Fathi AT, Sadrzadeh H, Borger DR, Ballen KK, Amrein PC, Attar EC, et al. Prospective serial evaluation of 2-hydroxyglutarate, during treatment of newly diagnosed acute myeloid leukemia, to assess disease activity and therapeutic response. Blood. 2012;120(23):4649-52.
18. Wang F, Travins J, DeLaBarre B, Penard-Lacronique V, Schalm S, Hansen E, et al. Targeted inhibition of mutant IDH2 in leukemia cells induces cellular differentiation. Science (New York, N Y). 2013;340(6132):622-6.
19. Stein EM, DiNardo CD, Pollyea DA, Fathi AT, Roboz GJ, Altman JK, et al. Enasidenib in mutant IDH2 relapsed or refractory acute myeloid leukemia. Blood. 2017;130(6):722-31.
20. Agency EM. Withdrawal Assessment Report Idhifa. 2019.
21. Stone PW. Popping the (PICO) question in research and evidence-based practice. Applied nursing research : ANR. 2002;15(3):197-8.
22. Poinsignon V, Mercier L, Nakabayashi K, David MD, Lalli A, Penard-Lacronique V, et al. Quantitation of isocitrate dehydrogenase (IDH)-induced D and L enantiomers of 2-hydroxyglutaric acid in biological fluids by a fully validated liquid tandem mass spectrometry method, suitable for clinical applications. J Chromatogr B Analyt Technol Biomed Life Sci. 2016;1022:290-7.
23. 2-Hydroxyglutaric acid [Internet]. 2004.
24. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. BMC Medical Research Methodology. 2005;5(1):13.
25. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Annals of internal medicine. 2011;155(8):529-36.
26. Wang JH, Chen WL, Li JM, Wu SF, Chen TL, Zhu YM, et al. Prognostic significance of 2-hydroxyglutarate levels in acute myeloid leukemia in China. Proc Natl Acad Sci U S A. 2013;110(42):17017-22.
27. Balss J, Thiede C, Bochtler T, Okun JG, Saadati M, Benner A, et al. Pretreatment d-2-hydroxyglutarate serum levels negatively impact on outcome in IDH1-mutated acute myeloid leukemia. Leukemia. 2016;30(4):782-8.
28. Willekens C, Micol J-B, Poinsignon V, Quivoron C, Saada V, Broutin S, et al. Serum 2-Hydroxyglutarate Level Can Predict IDH2 Mutation in Myeloid Sarcoma. Blood. 2015;126(23):3835-.
29. Janin M, Mylonas E, Saada V, Micol JB, Renneville A, Quivoron C, et al. Serum 2-hydroxyglutarate production in IDH1- and IDH2-mutated de novo acute myeloid leukemia: a study by the Acute Leukemia French Association group. J Clin Oncol. 2014;32(4):297-305.
30. Brunner AM, Neuberg DS, Wander SA, Sadrzadeh H, Ballen KK, Amrein PC, et al. Isocitrate dehydrogenase 1 and 2 mutations, 2-hydroxyglutarate levels, and response to standard chemotherapy for patients with newly diagnosed acute myeloid leukemia. Cancer. 2019;125(4):541-9.
31. DiNardo CD, Propert KJ, Loren AW, Paietta E, Sun Z, Levine RL, et al. Serum 2-hydroxyglutarate levels predict isocitrate dehydrogenase mutations and clinical outcome in acute myeloid leukemia. Blood. 2013;121(24):4917-24.
32. Chowdhury R, Yeoh KK, Tian Y-M, Hillringhaus L, Bagg EA, Rose NR, et al. The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. EMBO reports. 2011;12(5):463-9.
33. U.S. Food & Drug Administration. FDA approves ivosidenib for relapsed or refractory acute myeloid leukemia 2018, October 18 [Available from: https://www.fda.gov/Drugs/InformationOnDrugs/ApprovedDrugs/ucm614128.htm.
34. U.S. Food & Drug Administration. FDA granted regular approval to enasidenib for the treatment of relapsed or refractory AML 2017, January 8 [Available from: https://www.fda.gov/drugs/informationondrugs/approveddrugs/ucm569482.htm.
35. DiNardo CD, Stein EM, de Botton S, Roboz GJ, Altman JK, Mims AS, et al. Durable Remissions with Ivosidenib in IDH1-Mutated Relapsed or Refractory AML. New England Journal of Medicine. 2018;378(25):2386-98.

36. *Stein EM, DiNardo CD, Fathi AT, Pollyea DA, Stone RM, Altman JK, et al. Molecular remission and response patterns in patients with mutant-IDH2 acute myeloid leukemia treated with enasidenib. Blood. 2019;133(7):676-87.*
37. *Natsumeda M, Igarashi H, Nomura T, Ogura R, Tsukamoto Y, Kobayashi T, et al. Accumulation of 2-hydroxyglutarate in gliomas correlates with survival: a study by 3.0-tesla magnetic resonance spectroscopy. Acta neuropathologica communications. 2014;2:158.*

Appendix 1

The following databases were searched:
- Embase
- Medline Ovid
- Web of Science
- Cochrane Central
- Google Scholar

Embase (434 items found)
('hematologic malignancy'/de OR 'leukemia'/exp OR (leukemi* OR leukaemi* OR erythroleukem* OR erythroleukaem* OR aml OR ((hemato* OR haemato*) NEAR/3 (malignan* OR neoplas* OR cancer*))):ab,ti) AND ('isocitrate dehydrogenase'/de OR 'isocitrate dehydrogenase (NADP)'/de OR 'isocitrate dehydrogenase (NAD)'/de OR 'isocitrate dehydrogenase 1'/de OR 'isocitrate dehydrogenase 2'/de OR 'idh1 gene'/de OR 'idh2 gene'/de OR ((isocitr* NEAR/3 dehydrogenas*) OR idh OR midh OR idh1 OR midh1 OR idh2 OR midh2):ab,ti) AND ('2 hydroxyglutaric acid'/de OR 'glutaric acid derivative'/de OR 'glutaric acid'/de OR (hydroxy-glutar* OR hydroxyglutar* OR Hg OR 2hg OR αhg OR d2hg OR r2hg OR l2hg OR glutaric-acid* OR glutarate*):ab,ti)

Medline Ovid (162 items found)
(Hematologic Neoplasms / OR exp Leukemia/ OR (leukemi* OR leukaemi* OR erythroleukem* OR erythroleukaem* OR aml OR ((hemato* OR haemato*) ADJ3 (malignan* OR neoplas* OR cancer*))).ab,ti.) AND (Isocitrate Dehydrogenase/ OR ((isocitr* ADJ3 dehydrogenas*) OR idh OR midh OR idh1 OR midh1 OR idh2 OR midh2).ab,ti.) AND (alpha-hydroxyglutarate.nm. OR Glutarates/ OR (hydroxy-glutar* OR hydroxyglutar* OR Hg OR 2hg OR d2hg OR r2hg OR l2hg OR glutaric-acid* OR glutarate*).ab,ti.)

Web of science (304 items found)
TS=(((leukemi* OR leukaemi* OR erythroleukem* OR erythroleukaem* OR aml OR ((hemato* OR haemato*) NEAR/2 (malignan* OR neoplas* OR cancer*)))) AND (((isocitr* NEAR/2 dehydrogenas*) OR idh OR midh OR idh1 OR midh1 OR idh2 OR midh2)) AND ((hydroxy-glutar* OR hydroxyglutar* OR Hg OR 2hg OR αhg OR d2hg OR r2hg OR l2hg OR glutaric-acid* OR glutarate*)))

Cochrane Central (17 items found)
((leukemi* OR leukaemi* OR erythroleukem* OR erythroleukaem* OR aml OR ((hemato* OR haemato*) NEAR/3 (malignan* OR neoplas* OR cancer*))):ab,ti) AND (((isocitr* NEAR/3 dehydrogenas*) OR idh OR midh OR idh1 OR midh1 OR idh2 OR midh2):ab,ti) AND ((hydroxy-glutar* OR hydroxyglutar* OR Hg OR 2hg OR αhg OR d2hg OR r2hg OR l2hg OR glutaric-acid* OR glutarate*):ab,ti)

Google Scholar
leukemia|leukaemia|erythroleukemia|erythroleukaemia|aml "isocitrate dehydrogenase"|"isocitric acid dehydrogenase"|idh|midh|idh1|midh1|idh2|midh2 hydroxyglutarate|"hydroxyglutaric acid"|Hg|2hg|αhg|d2hg|r2hg|l2hg|"glutaric-acid"|glutarate

Appendix 2

Explanation Quality Assessment

(1) Patient selection: age variation is well distributed and representative for AML; male/female ratio is representative for AML; cytogenetic risk profiles between patients are not significantly different; IDH mutated/IDH wildtype ratio is representative.

- 0 points: no items are sufficient.
- 1 point: 1 or 2 items are sufficient.
- 2 points: 3 or 4 items are sufficient.

(2) Inclusion and exclusion criteria are specified.

- 0 points: inclusion and exclusion criteria are not specified.
- 1 point: inclusion and exclusion criteria are specified.

(3) Reproducibility of the method: serum 2-HG determinations were done with the same analytical method at all patients. Also, the point of measurement was for all patients at the same time in treatment.

- 0 points: no items are fulfilled.
- 1 point: 1 item is fulfilled.
- 2 points: all items are fulfilled.

Articles that score 0 to 1 points are considered to be of poor quality, 2 to 3 points are of moderate quality and 4 to 5 points are of good quality.
The maximum score is 5 points.

Systematic Review

Nonoperative Treatment for Acute Uncomplicated Appendicitis in Children to Reduce Complications as Recurrent Appendicitis:

a Systematic Review and Meta-analysis

Hilde Koese[a], Lydia van Krimpen[a] and Laura Oudshoorn[a]

[a] Medical student, Erasmus MC, Rotterdam, The Netherlands

Abstract

In the paediatric population, appendectomy remains the standard approach for treating acute appendicitis, but in recent years, interest in antibiotics as an alternative to appendectomy has been growing. In this systematic review, we aimed to provide an overview of the current literature on the clinical effectiveness of nonoperative management (NOM) of acute uncomplicated appendicitis in children.

A systematic literature search on PubMed was conducted to find available literature from the period August 2017 to September 2019. Only studies written in English or Dutch, which compared nonoperative treatment to appendectomy in children with acute uncomplicated appendicitis, were included. The quality of the articles was assessed using a quality scoring tool. Additionally, we performed a meta-analysis on the recurrence and complications rates.

Six studies were included in this systematic review. Only one of the six studies suggested a significant difference in the subsequent appendectomy rate between the NOM and appendectomy group. Our meta-analysis showed 11.2% (95%CI (0.2-20.8%)) of the patients treated nonoperatively had a recurrence compared to 5.4% (95%CI (0.4-6.3%)) of the patients who had an appendectomy. Complications occurred in 16 of the 169 patients (9.5%) who were treated nonoperatively compared to 16 of 186 patients (8.6%) who were treated with appendectomy. The difference in both rates between the two groups was not significant.

We found no significant difference in NOM or appendectomy as treatment for acute paediatric uncomplicated appendicitis with recurrence rate, subsequent appendectomy rate and complication rate as the outcomes. On the basis of our findings we can conclude that nonoperative management (NOM) for acute uncomplicated appendicitis in children is not inferior to an appendectomy.

Keywords

Paediatrics, acute appendicitis, nonoperative management

Introduction

Acute appendicitis is an inflamed and painful (cecal) appendix. With a lifetime risk of 7-8% and a peak incidence in the teenage years, acute appendicitis is a frequent cause of abdominal pain in children (1,2). Moreover, acute appendicitis is the most common surgical emergency among children (1). Traditionally, removal of the inflamed appendix, named appendectomy, is the standard treatment for acute appendicitis in both children and adults (3). Although appendectomy is an effective treatment and generally well tolerated, it is a surgical intervention and is therefore associated with mortality and morbidity, such as intestinal obstruction and wound infection (4,5). Readmission for complication(s) after appendectomy is needed in 7.4% of the children (4).

However, the necessity of an appendectomy can be debated when the appendicitis is uncomplicated (appendicitis without an abscess, and not appendicolith or perforated or gangrenous appendicitis). In recent years, interest in antibiotics as an alternative to appendectomy has been growing. In various studies adults with uncomplicated appendicitis have been treated with nonoperative management (NOM), antibiotics, and this was compared to appendectomy as conservative treatment. These studies have shown the benefits of NOM in comparison to appendectomy (5-9). With conservative treatment, anaesthesia and surgery including the risks of associated complications can be avoided.

In the paediatric population, appendectomy remains the standard approach for treating acute appendicitis. To date, a few randomized controlled trials that investigated conservative treatment as alternative to appendectomy have been published. In recent years, some systematic reviews and meta-analysis have been published. However, the most recent systematic review is dates from August 2017 (10). In previous systematic reviews, very few studies were included that focused only on children, and a significant amount of studies were retrospective reviews (10,11). Some studies about appendectomy versus NOM of appendicitis have recently been published after the latest systematic review. On account of this recent studies, another new evidence-based view can be found about the treatment of an uncomplicated appendicitis in children in this systematic review.

With the knowledge of the successful conservative treatment of acute appendicitis in adults, the aim of our study was to provide an overview of the current literature on the clinical effectiveness of NOM of acute appendicitis in children. In this systematic review, we analysed the recurrence rate defined as the recurrence of appendicitis or a related diagnosis and the subsequent appendectomy rate defined as the need for appendectomy after initial antibiotic treatment. In addition, we focused on the complications of both treatments of acute appendicitis.

Methods

We searched the database PubMed using the search term: ("Appendicitis/drug therapy"[Mesh] OR "Anti-Bacterial Agents"[Mesh] OR "Conservative Treatment"[Mesh] OR nonoperati*[tiab] OR non-operati*[tiab] OR nonsurg*[tiab] OR non-surg*[tiab] OR antibioti*[tiab]) AND ("Child"[Mesh] OR "Pediatrics"[Mesh] OR child*[tiab] OR pediatri*[tiab] OR infant*[tiab] OR neonat*[tiab] OR teenage*[tiab]) AND ("Appendicitis/surgery"[Mesh] OR "Appendectomy"[Mesh] OR appende*[tiab]) AND (uncomplic*[tiab] OR unperfora*[tiab] OR nonperfora*[tiab] OR non-perfora*[tiab] OR "simple"[tiab] OR "acute"[tiab]) NOT "Abscess"[Mesh] NOT "Preoperative Period"[Mesh] NOT "Antibiotic Prophylaxis"[Mesh] .

We performed the search on September 19, 2019. The results of this search were independently assessed by three authors on inclusion and exclusion criteria by reading the titles and abstracts.

Study selection

Our review is limited to the treatment of appendicitis in children (aged 18 years and younger) and original studies to improve the reliability of the conclusion of this systematic review. We included articles written in English or Dutch. Additionally, only studies which compared surgical intervention to antibiotic treatment in children who had a diagnosis of acute uncomplicated appendicitis were included. Furthermore, the articles had to be original research articles.

Moreover, we excluded studies that focused on appendectomy versus antibiotics in patients with complicated appendicitis or studies which did not have a separate group of paediatric patients with uncomplicated appendicitis, or studies that looked at antibiotic as surgical prophylaxis, diagnostic tests for detecting appendicitis or reducing symptoms. We also excluded articles that were not full text available for Erasmus MC students.

Quality assessment

After including and excluding studies, we used a quality scoring tool based on the Newcastle-Ottawa Scale and GRADE criteria (Appendix 1) developed for the purpose of this systematic review to assess the quality of the residual articles and exclude unreliable studies. We scored each article according to ten questions in our quality scale. Articles that scored 6 or less points were considered to have a low quality. Those articles were excluded from the selection.

Outcomes

Our primary outcome is the recurrence rate with recurrent appendicitis defined as diagnosis of appendicitis after completing initial treatment with antibiotics, a diagnosis of stump appendicitis (inflammation of the remaining appendiceal stump) after appendectomy, or a diagnosis potentially related to appendicitis (e.g., abdominal pain, (complicated) appendicitis, colitis, intestinal obstruction, diarrhoea, vomiting, dehydration), during the follow-up period.

Our secondary outcomes are subsequent appendectomy and the rate of complications or adverse outcomes of treatment. Subsequent appendectomy is defined as the need for appendectomy after initial treatment, during the follow-up period. Even after an appendectomy, children can get a second appendicitis, because the remaining stump can be infected.

Figure 1 - Flow diagram of article selection

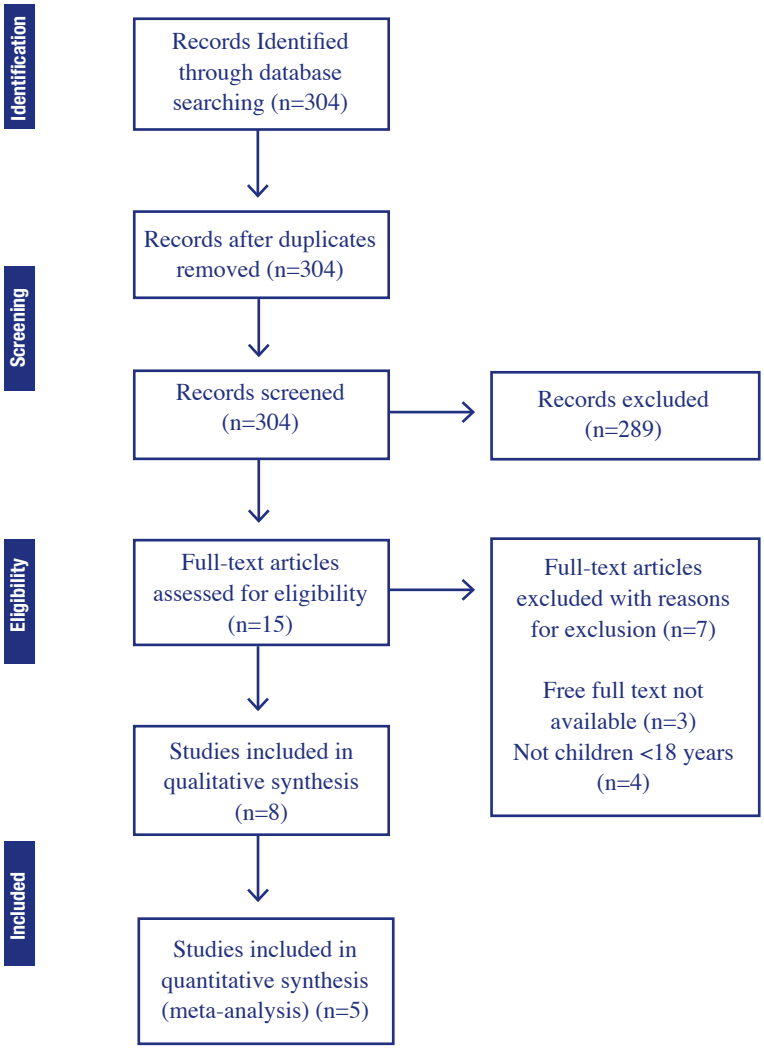


Table 1 - Characteristics of the studies

Authors and year	Country	Study design	Included patients	Patient category (years	Follow up (years	Intervention	Controle	Quality assessment
Bachur et al. (2017) (11)	United States of America	Retrospective study	99001	<19	1	NOM	Appendectomy	9/10
Hall et al. (2017) (12)	United Kingdom	Prospective, multicentre randomised study	106	3-15	1	Active observation (AO) - Broad spectrum IV AB - Reviewed every 3 months for a year	Interval appendectomy (IA) - Open or laparoscopic - Reviewed after 6 weeks and 1 year	8/10
Minnec et al. (2016) (13)	United States of America	Prospective patient-preference cohort study	102	7-17	1	NOM - 24 hours of inpatient observation while receiving IV AB (piperacillin sodium-tazobactam sodium or ciprofloxacin hydrochloride and metronidazole hydrochloride if allergic) - After that 10 days of treatment with oral AB - Follow up: at 2 to 5 days, 10 to 14 days, 30 days, 6 months, and 1 year after discharge.	Appendectomy - Laparoscopic within 12 hours - Follow up: at 30 days and 1 year	8/10
Svensson et al. (2015) (14)	Sweden	Pilot Randomized Controlled Trial	50	5-15	1	NOM - IV meropenem (10 mg/kg × 3 per 24 hours) and metronidazole (20 mg/kg × 1 per 24 hours) for at least 48 hours - Child clinically well and tolerating oral intake: oral ciprofloxacin (20 mg/kg × 2 per 24 hours) and metronidazole (20 mg/kg × 1 per 24 hours) for another 8 days	Appendectomy - Preoperative AB prophylaxis with 20 mg/kg of metronidazole - Open or laparoscopic	9/10
Gorter et al. (2017) (16)	The Netherlands	Follow-up study	44	7-17	1,25-3	Initially nonoperative treatment - IV administration of AB with in-hospital monitoring, diet restriction and pain medication as needed.	Immediate appendectomy	6/10
Lee et al. (2017) (18)	United States of America	Prospective patient-preference cohort study	83	3-17	1	NOM - IV AB (ceftriaxone/metronidazole or ciprofloxacin /metronidazole if penicillin allergic - Tolerating a diet and having completed at least one dose of IV AB -> oral amoxicillin-clavulanate, ciprofloxacin/Flagyl, or cefdinir/Flagyl for a total of 10 days - Follow up: at 10 to 14 days, 30 days, 3 months, and 1 year after discharge.	Appendectomy - Laparoscopic - Initiation of IV AB - Follow-up: at 30 days, 3 months and 1 year	8/10

NOM: nonoperative management. IV: intravenous. AB: antibiotics

Hypothesis

Our hypothesis is that the recurrence rate of appendicitis and the subsequent appendectomy rate in children who received NOM of acute appendicitis will be low, so the majority of this group will be successfully treated with antibiotics. Besides, we expect that the number of complications will be higher in the operative management group.

Statistical Analysis

In order to enlarge the clinical relevance and to bypass possible bias, we pooled the data from the five studies we included. We exclude studies which did not report results for both groups and also for both outcomes. We performed a meta-analysis on two specific outcomes, namely recurrence and complication rate. Excel Common license Attribution 3.0 Unported (CC BY 3.0) program was used to perform the meta-analysis. By comparing these datasets for the outcomes we selected, we could analyse the incidence rate of each outcome for both appendectomy and NOM. The pooled data were used to generate Forest plots which show the recurrence and complications rate by study and by group. In addition, the forest plots provide an overview of those pooled data for the recurrence and complication rate.

Results

Literature search

Our search in PubMed resulted in 304 publications. Eight studies met our inclusion criteria (Figure 1).

Among these eight studies, one study is a pilot randomized controlled trial (12), one study is a prospective multicentre randomised study (13), two are prospective patient-preference cohort studies (14,15), another two studies are non-randomized prospective studies (16,17), one nonrandomized follow-up study (18), and one study is retrospective (11).

Study characteristics

The results of the quality scoring tool are shown in Table 1. Of the eight included studies, six had a high-quality score (six points or more), whereas two studies had a low-quality score (five points or less). We excluded those two studies.

Main characteristics of the included studies are shown in Table 1. The number of patients in the studies varied from 44 to 99001. Median age differed from 8 to 14 years. All six studies reported data on children treated with antibiotics and children treated with appendectomy.

Table 2 - Comparison of study characteristics for the recurrence rate in the NOM versus appendectomy treatment group

Author(s)	Sample size (NOM group; appendectomy group)	Follow-up period	NOM group	Appendectomy group	p-value
Bachur et al.	99001 (4190; 61522)	1 year	2628 (29.9%)	18613 (18.6%)	N.A.*
Hall et al.	106 (51; 44)	1 year	6 (15.4%)	2 (4.5%)	N.A.
Minneci et al.	102 (37; 65)	1 year	1 (2.7%)	8 (12.3%)	0.15
Svensson et al.	50 (24; 26)	1 year	2 (8.3%)	0 (0%)	0.23
Gorter et al.	44 (25; 19)	1.25-3 years	3 (12%)	0 (0%)	N.A.
Lee et al.	83 (51; 32)	1 year	9 (17.6%)	0 (0%)	N.A.

**Not available/applicable/announced*

Table 2 shows the 6 studies that examined the recurrence rate of appendicitis or a related diagnosis. In Bachur et al. 2628 patients in the NOM group and 18613 patients in the appendectomy group visit the emergency department once or more in the one-year follow-up period. In Hall et al. six of the fifty-one children received antibiotics during the study period developed recurrent acute appendicitis, and two of the forty-four children in the appendectomy group developed recurrent appendicitis prior to their planned appendectomy. Minneci et al. shows the rates of complicated appendicitis. In Minneci et al. the success rate of nonoperative management was 75.7% (95% CI, 58.9%-88.2%) at 1 year (28 of 37 children) and in Svensson et al. the success rate was similar in each group [nonoperative treatment group 22/24 (92%) vs appendectomy group 26/26 (100%); P=0.23]. Gorter et al. and Lee at al. show the rate of recurrent appendicitis. None of the studies found a significant difference between the NOM and appendectomy group, but only two of the six studies showed a p-value.

Three studies (12,14,15) had a control group of children undergoing laparoscopic appendectomy, children in the control group of another study (13) underwent open or laparoscopic appendectomy, and the other two studies (18,19) did not describe the type of (immediate) appendectomy in the control group. Five studies (12-15,19) had a one year follow-up and one had a follow-up of 1.25 to 3 years (18). The recurrence rate of appendicitis or a related diagnosis and the subsequent appendectomy rate was analysed in all six studies. Data about the complications rate or adverse outcomes were reported in five studies (12-15,18).

Table 2 shows the six studies that examined the recurrence rate of appendicitis or a related diagnosis. In Minneci et al. (13) the success rate of NOM was 75.7% (95% CI:58.9%-88.2%) at 1 year (28 of 37 children). None of the studies found a significant difference for the rate of recurrent appendicitis, the number emergency department visits or the rate of complicated appendicitis between the NOM and appendectomy group, but only two of the six studies provided a p-value. However, in Gorter et al. (9) the 95% confidence interval of the 12% recurrence rate is 4-30% implying a significant difference.

Subsequent appendectomy

Table 3 shows the six studies that examined the number of subsequent appendectomies. None of the studies found a significant difference in the subsequent appendectomy rate for the NOM and appendectomy group, but only one of the six studies provided a p-value. The prevalence rates of a needed subsequent appendectomy in the NOM group vary from 5.4% to 37.5%, compared to 0% to 4.5% in the group of children who underwent appendectomy.

Table 3 - Comparison of study characteristics for subsequent appendectomy in the NOM versus appendectomy treatment group

Author(s)	Sample size (NOM group; appendectomy group)	Follow-up period	NOM group	Appendectomy group	p-value
Bachur et al.	99001 (4190; 61522)	1 year	1032 (24.6%)	N.A.	N.A.
Hall et al.	106 (52; 44)	1 year	8 (15.4%)	2 (4.5%)	N.A.
Minneci et al.	102 (37; 65)	1 year	2 (5.4%)	0 (0%)	0.15
Svensson et al.	50 (24; 26)	1 year	9 (37.5%)	0 (0%)	N.A.
Gorter et al.	44 (25; 19)	1.25-3 years	6 (24%)	0 (0%)	N.A.
Lee et al.	83 (51; 32)	1 year	8 (15.7%)	0 (0%)	N.A.

Table 3 shows the 6 studies that examined the number of subsequent appendectomies. Bachur et al., Hall et al., Svensson et al., Gorter et al., and Lee et al. show the number of subsequent appendectomies for recurrent appendicitis or abdominal pain. Minneci et al. shows the rate of appendicitis-related surgery or other invasive procedure within 30 days. None of the studies found a significant difference between the NOM and appendectomy group, but only one of the six studies showed a p-value.

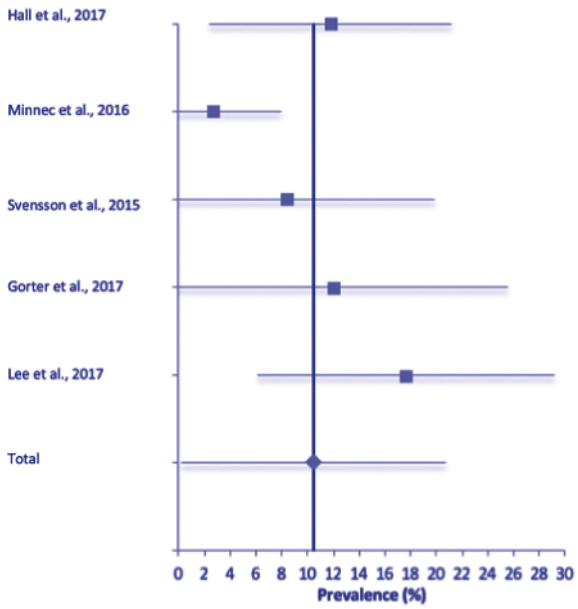
Table 4 - Comparison of study characteristics for the complication rate in the NOM versus appendectomy treatment group

Author(s)	Sample size (NOM group; appendectomy group)	Follow-up period	NOM group	Appendectomy group	p-value
Hall et al.	106 (51; 44)	1 year	2 (3.9%)	3 (6.8%)	N.A.
Minneci et al.	102 (37; 65)	1 year	0 (0%)	5 (7.7%)	N.A.
Svensson et al.	50 (24; 26)	1 year	2 (8.3%)	0 (0%)	0.23
Gorter et al.	44 (25; 19)	1.25-3 years	3 (12%)	2 (10.5%)	N.A.
Lee et al.	83 (51; 32)	1 year	9 (17.6%)	6 (18.8%)	N.A.

Table 4 shows the five studies that examined the incidence of complication or adverse outcomes after treatment with antibiotics and (subsequent) appendectomy. Complications such as allergic reaction, readmission for gastroenteritis, urinary tract infection, and reoperation, occurs in all five studies. Bachur et al. does not report data about complications. None of the studies found a significant difference between the NOM and appendectomy group, but only one of the five studies showed a p-value.

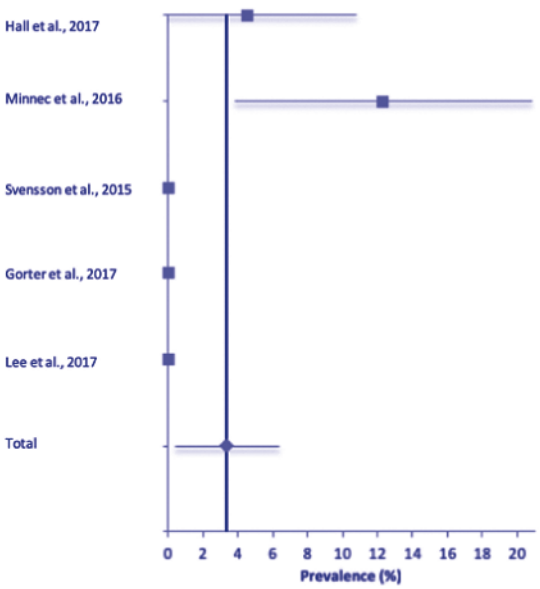
Figure 2- The recurrence rate for antibiotics and appendectomy

Recurrence rate NOM



Forest plot of pooled recurrence rate in the NOM group

Recurrence rate appendectomy



Forest plot of pooled recurrence rate in the appendectomy group

Complications

Table 4 shows the five studies that examined the complication rate or adverse outcomes after treatment with antibiotics and (subsequent) appendectomy. Complications such as allergic reaction, readmission for gastroenteritis, urinary tract infection, and reoperation, occurred in all five studies. Bachur et al. (11) does not report data about complications. None of the studies found a significant difference between the NOM and appendectomy group, but only one of the five studies provided a p-value. However, in Gorter et al. (9) the 95% confidence interval of the 12% complication rate in the NOM group is 4-30% and 95% confidence interval of the 11% complication rate in the appendectomy group is 3-31%, implying no significant difference.

Meta-analysis

Recurrence rate

A total of five studies consisting of a total of 385 patients have been incorporated in the meta-analyse. Within the NOM group, 21 of the total 188 (11.2%) patients had a recurrence. The recurrence rate in the appendectomy group was 10 out of 186 (5.4%) patients. The recurrence rates are shown in figure 2. The difference in overall recurrence rate between both groups is not significant for respectively the NOM and appendectomy group (95% CI [0.2-20.8%] and [0.4-6.3%]).

Complications

A total of five studies consisting of 385 patients has been incorporated in the meta-analysis. Patients treated nonoperatively had a total of sixteen events of complications of the total 169 (9.5%). In the appendectomy group, there was a total of sixteen complications from the total of 186 (8.6%). The complication rates are shown in figure 3. The overall complication rate was lower in the appendectomy group, though the difference between both groups was not significant (95% CI [9.6-11.4%] and [7.8-9.7%]) for respectively the NOM and appendectomy group.

The complications rate for NOM ranges from 0.0% to 28.1% (resp. 14 and 15). For appendectomy, these values vary from 0.0% to 18.8% (resp. 12 and 15). An overview of the values is shown in the pooled data analyses, in the figure the total prevalence. The mean complication rate for NOM is 10.5% (95%CI: 9.55-11.4%), in comparison to the appendectomy group where the percentage is 8.76% (95%CI: 7.84-9.67%). In conclusion there is no significant difference in complication rate between both groups

Discussion

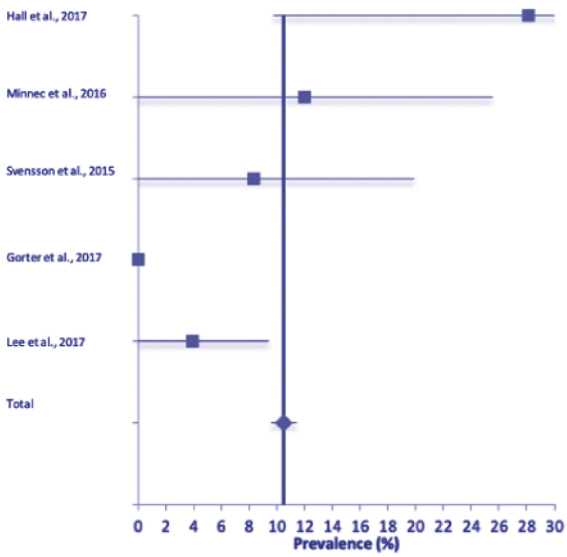
In this systematic review we evaluated six studies to compare conservative and surgical treatment of acute uncomplicated appendicitis in children. There was no significant difference between therapies.

Firstly, the recurrence rate was higher for NOM than for appendectomy, though this difference in recurrence rate was not significant. Secondly, the complication rate was approximately the same for the NOM and appendectomy group. Thirdly, subsequent appendectomy occurred more frequent after conservative treatment, but differed not significantly from treatment with appendectomy.

Due to the partly the same CI for the pooled data analyses, we can conclude that the complication rate for both treatments is not significantly different. We can say that the pooled data slightly increases the complication rate. In both cases, one study showed significantly a higher prevalence of complications, so the overall complication rate is higher as well.

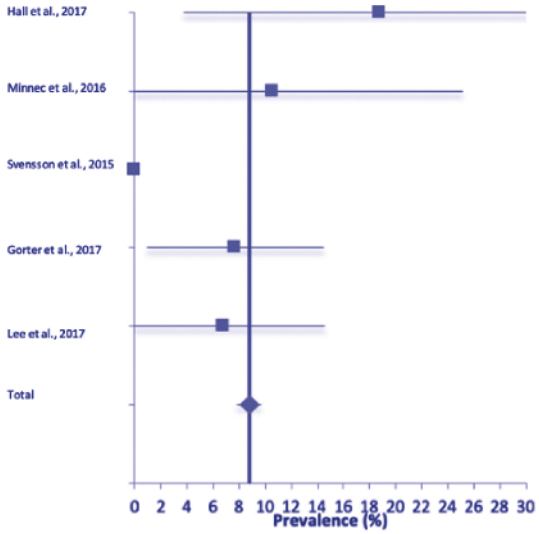
Figure 3- The complications for antibiotics and appendectomy

Complications NOM



Forest plot of pooled analysis of complication rate in the NOM group

Complications appendectomy



Forest plot of pooled analysis of complication rate in the appendectomy group

One of the previous reviews about the treatment of appendicitis in children studied the difference in outcomes (20). They concluded that there was a broad heterogeneity for the outcomes and definitions of outcomes of systematic review and randomized controlled trials about this subject. Therefore, we carefully selected articles with the same outcomes in order to compare those studies and draw a conclusion in this systematic review. Other reviews can partly be compared to our results although they did not include studies with a large population (10,18,21,22). In our review, we included a study with a large group of participants which provides our review of stronger power (11). Our findings are conform to the results in another review which found that there are no significant differences in occurrence of outcome between NOM and appendectomy (10).

Subsequent appendectomy

When interpreting the findings, it should be noted that the recurrence rate and subsequent appendectomy rate is always low in the group of patients treated with appendectomy. This can be explained by the fact that the appendix of these patients is removed. As a result, the recurrence rate tends to be greater in the group of patients treated nonoperatively. Our review showed a recurrence rate that is greater in the NOM group, but the difference between the two groups is not significant. In addition, the complication rate differs not significantly between both groups. Therefore, NOM does not seem to be inferior to an appendectomy in children with acute uncomplicated appendicitis. For future cases of appendicitis, we advise to provide evidence-based information for both children and parent about possible treatments and potential complications, so they can choose the treatment they prefer.

However, there is a stump of the appendix remaining after appendectomy, which could become inflated as well. People with stump appendicitis need reoperation to remove the remaining inflated stump. Table 3 shows 5.4% of those who underwent appendectomy get stump appendicitis and needed a subsequent appendectomy. However, because of the small population of the study 5.4% corresponds to only two children (14). So, the rate of subsequent appendectomy is particularly low, but these two children are exposed to all risks of surgery again.

Four studies showed 0% needed subsequent appendectomy. Despite their particularly high recurrence rate, Bachur et al (11) unfortunately did not report data about the number of subsequent appendectomies.

Limitations of the included studies

We included six studies. Despite the high-quality score of these studies, there are some limitations. For example, the randomization of children with acute appendicitis for treatment with NOM or appendectomy was absent. Because appendectomy is the worldwide gold standard treatment for acute appendicitis, the young patients or their parents have the right to choose the way of therapy themselves (3). Not randomizing the patients can cause a bias in the review because children who are in more pain maybe choose for a surgery because of the instant pain relief. On the other hand, children who cannot withstand the idea of surgery choose for antibiotics.

None of the studies included in our meta-analysis reported children below three years of age. Moreover, results were not reported according to age categories. Children younger than five years have an increased risk of perforated appendicitis (23). Therefore, nonoperative treatment for those children may be more hazardous and less efficacious than for older infants.

In addition, the median age is ten years or more in four of the five studies, and children of that age are structurally more similar to adults. Taking this into account, the complication rate we found in this review might be an underestimation and not representative for children of any age, but only for the older aged children.
One outcome, subsequent appendectomy was not included in our meta-analysis because of the debatable cases of subsequent appendectomy in the appendectomy group. As expected, in almost all studies, no subsequent appendectomies were mentioned in the appendectomy group, though one study mentioned two cases of the appendectomy group who had undergone subsequent appendectomy (13). Further research led us to the conclusion that those two cases were caused by the fact that two children refused the initial appendectomy. Later on, an appendectomy was needed due to a reinfection of the appendix. These cases were defined as subsequent appendectomy, but it did not fit our definition of subsequent appendectomy, so we decided to exclude this outcome from the pooled evidence analysis.

Strengths of the systematic review

This extensive review was possible because we have chosen to limit ourselves to children only. In addition, many studies have been done to analyse the effect of NOM on treatment outcomes in adults with appendicitis, but the optimum treatment in children with acute uncomplicated appendicitis is still unclear. This allows us to make a specific recommendation on this subject. Our review included a selection of high rated studies and a total group size that is representative.

Limitations of the systematic review

There are several limitations of our systematic review. These limitations prevent us from drawing strong conclusions. Firstly, we studied a limited number of articles and treatment outcomes. Studying only these three outcomes results in an incomplete view about all the other outcomes that can occur during/after treatment of acute appendicitis in children. So, it urges to conduct another study that includes other outcomes before a definite advice can be formulated about how to treat children with acute uncomplicated appendicitis. However, studying these three specific outcomes accurately resulted in a reliable conclusion.
Secondly, we focused on a follow-up of one year after treatment because these intervals were used in the most articles we found. This is a limitation of our study because a longer follow-up of more than three years could have given other significant results, which we overlook now. We can imagine recurrent appendicitis will occur more frequent after nonoperative treatment when the children are followed for more than three years compared to one year. However, today there are no studies with a follow-up longer than three years.
Thirdly, the use of different definitions of nonoperative treatment (also see table 1) between the studies may affect our findings. For instance, active observation reviewing patients each three months during one year, hospital observation and intravenous antibiotics for a minimum of 24 hours, intravenous meropenem and metronidazole for at least 48 hours, intravenous antibiotics ceftriaxone/metronidazole (resp. 13, 14, 12, 15). This can lead to differences in effectiveness of nonoperative treatment. Therefore, in future studies it is important that the same policy of nonoperative management will be used.
Finally, we did not take account of confounders. For example, incomplete adjustment by our included studies for confounders such as comorbidities in and exposures to the children (for example nutrition) may have played a role in the findings. Moreover, the studies vary in study design and were performed in different countries. We think that the different countries could have a big difference in policy. For example, the procedure of appendectomy and antibiotic prophylaxis, and the aftercare, are different between countries (24,25). The high variation in study design and country makes that it is more difficult to compare the six studies. When all the studies were randomized controlled trials and were performed in the same country, we would have good-quality, comparable studies and probably be able to provide a more reliable result.

Further research

In future studies, we recommend to research additional outcomes in addition to the three outcomes in this review. Additional outcomes such as the length of stay, disability days, health-related quality of life, days of pain medication, number of patients who develop a complicated or perforated appendicitis, imaging studies of the abdomen after discharge and health costs are especially of interest.
Moreover, in future studies we could use a follow-up longer than three years, to detect later complications (such as adhesion obstruction) and more cases of recurrent appendicitis and subsequent appendectomy.
When sufficient evidence has been generated, paediatric patients together with their parents and clinicians are able to make a well-informed decision.
In addition, the exact pathogenesis of appendicitis is still unresolved. It is thus not surprising that it remains unclear whether it is better to treat children with uncomplicated appendicitis conservatively or surgically. To resolve this issue, further analysis of pathophysiological aspects of appendicitis, more specifically in children, is necessary.
One could suggest that the antibiotic treatment is less invasive for children. Therefore, our recommendation is to include the more patient-centred outcomes for further research instead of the clinical outcomes. The experience of children with the treatment of antibiotics can prove significantly different, because this systematic review states that there is no difference in clinical outcomes for the treatment of uncomplicated appendicitis in children.

Conclusions

This systematic review provides a high-level summary for treatment of a paediatric appendicitis. On the basis of our results, we can conclude antibiotic treatment is not inferior to appendectomy for acute uncomplicated appendicitis in children. For future cases, we recommend doing more research to prove treatment with antibiotics is as effective as appendectomy for all three; medical outcomes, patient-related outcomes and cost-effectiveness in children with uncomplicated appendicitis.

References

1. *Addiss DG, Shaffer N, Fowler BS, Tauxe RV. The epidemiology of appendicitis and appendectomy in the United States. Am J Epidemiol 1990 Nov;132(5):910-925*
2. *Rothrock SG, Pagane J. Acute appendicitis in children: emergency department diagnosis and management. Ann Emerg Med 2000 Jul;36(1):39-51*
3. *Schneuer FJ, Adams SE, Bentley JP, Holland AJ, Huckel Schneider C, White L, et al. A population-based comparison of the postoperative outcomes of open and laparoscopic appendicectomy in children. Med J Aust 2018 Jul 16;209(2):80-85*
4. *Rollins KE, Varadhan KK, Neal KR, Lobo DN. Antibiotics Versus Appendicectomy for the Treatment of Uncomplicated Acute Appendicitis: An Updated Meta-Analysis of Randomised Controlled Trials. World J Surg 2016 Oct;40(10):2305-2318*
5. *Varadhan KK, Neal KR, Lobo DN. Safety and efficacy of antibiotics compared with appendicectomy for treatment of uncomplicated acute appendicitis: meta-analysis of randomised controlled trials. BMJ 2012 Apr 5;344:e2156*
6. *Styrud J, Eriksson S, Nilsson I, Ahlberg G, Haapaniemi S, Neovius G, et al. Appendectomy versus antibiotic treatment in acute appendicitis. a prospective multicenter randomized controlled trial. World J Surg 2006 Jun;30(6):1033-1037*
7. *Hansson J, Korner U, Khorram-Manesh A, Solberg A, Lundholm K. Randomized clinical trial of antibiotic therapy versus appendicectomy as primary treatment of acute appendicitis in unselected patients. Br J Surg 2009 May;96(5):473-481*
8. *Vons C, Barry C, Maitre S, Pautrat K, Leconte M, Costaglioli B, et al. Amoxicillin plus clavulanic acid versus appendicectomy for treatment of acute uncomplicated appendicitis: an open-label, non-inferiority, randomised controlled trial. Lancet 2011 May 7;377(9777):1573-1579*
9. *Gorter RR, The SML, Gorter-Stam MAW, Eker HH, Bakx R, van der Lee JH, et al. Systematic review of nonoperative versus operative treatment of uncomplicated appendicitis. J Pediatr Surg 2017 Aug;52(8):1219-1227*
10. *Xu J, Adams S, Liu YC, Karpelowsky J. Nonoperative management in children with early acute appendicitis: A systematic review. J Pediatr Surg 2017 Sep;52(9):1409-1415*
11. *Bachur RG, Lipsett SC, Monuteaux MC. Outcomes of Nonoperative Management of Uncomplicated Appendicitis. Pediatrics 2017 Jul;140(1):10.1542/peds.2017-0048. Epub 2017 Jun 2*
12. *Hall NJ, Eaton S, Stanton MP, Pierro A, Burge DM, CHINA study collaborators and the Paediatric Surgery Trainees Research Network. Active observation versus interval appendicectomy after successful non-operative treatment of an appendix mass in children (CHINA study): an open-label, randomised controlled trial. Lancet Gastroenterol Hepatol 2017 Apr;2(4):253-260*
13. *Minneci PC, Mahida JB, Lodwick DL, Sulkowski JP, Nacion KM, Cooper JN, et al. Effectiveness of Patient Choice in Nonoperative vs Surgical Management of Pediatric Uncomplicated Acute Appendicitis. JAMA Surg 2016 May 1;151(5):408-415*
14. *Svensson JF, Patkova B, Almstrom M, Naji H, Hall NJ, Eaton S, et al. Nonoperative treatment with antibiotics versus surgery for acute nonperforated appendicitis in children: a pilot randomized controlled trial. Ann Surg 2015 Jan;261(1):67-71*
15. *Minneci PC, Hade EM, Lawrence AE, Saito JM, Mak GZ, Hirschl RB, et al. Multi-institutional trial of non-operative management and surgery for uncomplicated appendicitis in children: Design and rationale. Contemp Clin Trials 2019 Aug;83:10-17*
16. *Gorter RR, van der Lee JH, Heijsters FACJ, Cense HA, Bakx R, Kneepkens CMF, et al. Outcome of initially nonoperative treatment for acute simple appendicitis in children. J Pediatr Surg 2018 Sep;53(9):1849-1854*
17. *Steiner Z, Buklan G, Gutermacher M, Litmanovitz I, Landa T, Arnon S. Conservative antibiotic treatment for acute uncomplicated appendicitis is feasible. Pediatr Surg Int 2018 Mar;34(3):283-288*
18. *Lee SL, Spence L, Mock K, Wu JX, Yan H, DeUgarte DA. Expanding the inclusion criteria for nonoperative management of uncomplicated appendicitis: Outcomes and cost. J Pediatr Surg 2017 Oct 9*
19. *Lin KB, Lai KR, Yang NP, Chan CL, Liu YH, Pan RH, et al. Epidemiology and socioeconomic features of appendicitis in Taiwan: a 12-year population-based study. World J Emerg Surg 2015 Sep 17;10:42-015-0036-3. eCollection 2015*
20. *Hall NJ, Kapadia MZ, Eaton S, Chan WW, Nickel C, Pierro A, et al. Outcome reporting in randomised controlled trials and meta-analyses of appendicitis treatments in children: a systematic review. Trials 2015 Jun 17;16:275-015-0783-1*
21. *Markar SR, Pinto D, Penna M, Karthikesalingam A, Bulathsinghala BK, Kumaran K, et al. A comparative international study on the management of acute appendicitis between a developed country and a middle income country. Int J Surg 2014;12(4):357-360*
22. *Kessler U, Mosbahi S, Walker B, Hau EM, Cotton M, Peiry B, et al. Conservative treatment versus surgery for uncomplicated appendicitis in children: a systematic review and meta-analysis. Arch Dis Child 2017 Dec;102(12):1118-1124*
23. *Gomes CA, Abu-Zidan FM, Sartelli M, Coccolini F, Ansaloni L, Baiocchi GL, et al. Management of Appendicitis Globally Based on Income of Countries (MAGIC) Study. World J Surg 2018 Dec;42(12):3903-3910*
24. *Ruffolo C, Fiorot A, Pagura G, Antoniutti M, Massani M, Caratozzolo E, et al. Acute appendicitis: what is the gold standard of treatment? World J Gastroenterol 2013 Dec 21;19(47):8799-8807*

Appendix

Appendix 1- Quality assessment

Criteria- articles with a score ≤ 6 were seen as weak articles and were excluded	Score
1. The study aim is clearly formulated	1
2. A hypothesis has been formulated	1
3. The allocation of patients into the intervention and control group is randomized	1
4. It is clearly described how the groups are classified	1
5. There is clear description of the treatment in both intervention group and control group	1
6. The intervention and control group are comparable or it has been corrected for	1
7. The number of included patients is more than or equal to 50	1
8. The study has one or more of the following as (primary or secondary) outcome: - Length of stay (LOS) - ED visits / hospitalizations - Recurrent appendicitis / frequency of appendectomy after discharge from the index visit - Disability days / treatment associated disability / days to normal activity, complications	1
9. The description of the method is clear and replicable for other researchers	1
10. The results of all outcomes are clearly shown	1

Systematic Review

The Effectiveness of Intrathecal versus Epidural or i.v. PCA Morphine as Postoperative Analgesia after Cesarean Section

I. de Lange[a] and A.E. van Opstal[a] Supervisor: C. van der Marel[b]

[a] Medical student, Erasmus MC University Medical Center Rotterdam, the Netherlands
[b] Dept. of Anesthesiology, Sophia Children's Hospital Rotterdam, the Netherlands

Correspondence: I. de Lange, email: 467376il@student.eur.nl, student id: 467376

Abstract

Objective: **The primary aim of this review is to determine the effectiveness of intrathecal (IT) morphine compared to epidural morphine (EP) or intravenous Patient Controlled Analgesia (i.v. PCA) morphine for postoperative analgesia after cesarean section (CS).**

Methods: **We searched five databases including studies on IT, EP and i.v. PCA morphine after cesarean section, which compare either IT and EP or IT and i.v. PCA morphine.**

Eligible articles were screened based on our exclusion criteria. We also assessed the found articles on quality using a self-altered Newcastle Ottawa Scale. We collected the patients' maternal age, American Society of Anesthesiologist (ASA) classification, weight and parity and looked at mean pain scores and incidence of side-effects. After data extraction, we performed a meta-analysis.

Results: **Of the 964 found articles, nine articles were included. One article considered i.v. PCA morphine versus IT morphine and all articles investigated EP morphine versus IT morphine. We calculated the Standardized Mean Differences (SMD's) for the visual analogue scale (VAS) outcomes of the studies which compared IT and EP morphine, as the use of the VAS differed per study. I.V. PCA morphine was associated with a higher pain score than IT morphine [0.14 (0.06) vs. 0.04 (0.01)].**

Our meta-analysis showed no significant difference in pain score between patients who received EP or IT morphine: SMD -0.05 (-0.30;0.19). This was overall supported by the outcomes of the articles which were not included in our meta-analysis.

Discussion: **There is no difference in the effectiveness of IT morphine compared to EP morphine as postoperative analgesia after CS based on our systematic review. IT morphine seems to be more effective than i.v. PCA morphine as postoperative analgesia after CS. However, more research on EP and i.v. PCA morphine compared to IT morphine is necessary to draw a definitive conclusion.**

Keywords

Morphine, Post-cesarean analgesia, Intrathecal, Epidural, i.v. Patient Controlled Analgesia

Introduction

The incidence of cesarean section (CS) for childbirth has increased in the last few decades. Of all childbirths, 18.6% are currently performed using CS.[1] Postoperative analgesia is important after CS to provide the most optimal conditions for the mother to bond with and take care of her child. It also contributes to early ambulation and discharge, supporting the mother in her recovery. Morphine is an effective drug for postoperative analgesia, which effects can last up to many hours after injection during spinal or epidural anesthesia.[2,3] For the administration of morphine, epidural (EP) and intrathecal (IT) analgesia and intravenous patient-controlled analgesia (i.v. PCA) are common methods.[4] IT administration of morphine requires a small dosage for long-lasting and adequate patient-satisfaction and is found to be cost-effective.[5] However, there is no clearly superior method of morphine administration when comparing IT, EP and i.v. PCA in effectiveness and occurrence of side-effects.[6] Therefore, the primary aim of this review is to determine the effectiveness of IT morphine compared to EP or i.v. PCA morphine as postoperative analgesia after CS. Our secondary outcome is the occurrence of side-effects with IT versus EP or i.v. PCA morphine as postoperative treatment after CS.

Methods

Search Strategy

On the 1st of May 2020 we searched the following databases: Embase, Medline ALL, Web of Science Core Collection, Cochrane Central Register of Controlled Trials and Google Scholar. We searched for English articles, excluding non human research and conference abstracts. We used the following key terms for the search string: Morphine, Intrathecal, Epidural, Pa-

tient-Controlled Analgesia and Cesarean section. All full search strings are attached in Appendix I.

Study selection

We independently reviewed the titles and abstracts of the found articles for eligibility. We used the following exclusion criteria:

- Reviews
- Not considering IT administration of morphine
- Not comparing EP or i.v. PCA to IT administration
- Comparing IT morphine to EP/ i.v. PCA analgesia other than morphine
- Not using a pain score as outcome
- Not considering patients undergoing CS
- Duplicates

We chose not to select articles based on their study type.

Quality assessment

To assess the quality of the articles, we altered the Newcastle-Ottawa scale (NOS) for cohort studies to be relevant for our review. [7] For our full quality assessment scale, see Appendix II. The maximum amount of points to be received is seven, wherein 1-3 points is rated low quality, 4-5 points moderate quality and 6-7 points high quality.

Looking at selection, one point was given if the IT cohort was similar to the average post-CS patient, if the EP/ i.v. PCA cohort was drawn from the same community as the IT cohort and when the ascertainment of administration type was from secured records or structured interviews.

As for comparability, studies received a point when there were no significant differences in baseline demographics between the study groups or significant differences were accounted for and did not influence the study results.

Lastly, we assessed outcome. Studies received a point when assessment of pain score came from self-reports or record linkage, when the follow-up was at least 24 hours after CS and when there was complete follow up or description of those lost provided.

Data extraction

We looked at the different patient characteristics in the study cohorts, including maternal age, the American Society of Anesthesiologist (ASA) classification, weight and parity. We extracted the mean pain scores and the incidence of pruritus and postoperative nausea and vomiting (PONV) in the IT, EP and i.v. PCA groups.

Analysis

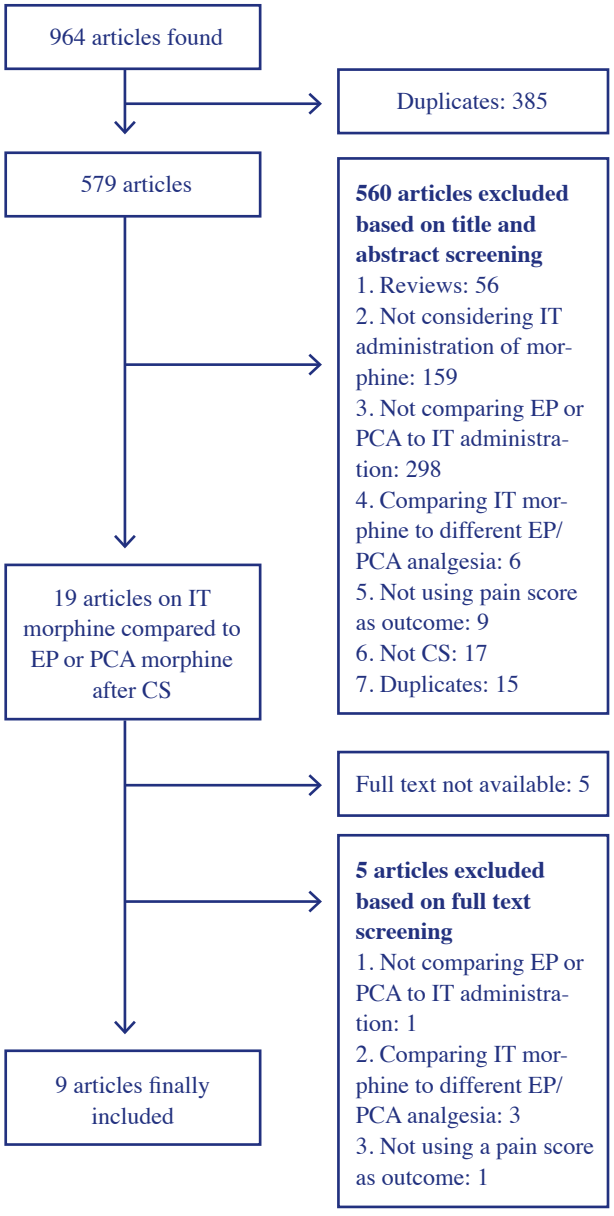
We calculated the standardized mean differences (SMD) for the primary outcome using a SMD calculator.[8] For both the primary and secondary outcome, we used the Meta-Essentials: Workbooks for meta-analysis to perform the meta-analysis on the found data.[9] We based the model of the meta-analysis on the percentage of heterogeneity of our studies. If the I2 percentage was above 50%, we used a random effects model. If the percentage was under 50%, we used a fixed effects model.

Results

Study search and selection

The search produced 964 articles, of which 385 were duplicates. After applying the exclusion criteria, 19 articles remained. 5 articles were not available to us in full text. After full text screening, 9 articles were finally included. For the full search selection, see figure 1.

Figure 1 - Flowchart of selection process. IT = intrathecal, EP = epidural, PCA = patient-controlled analgesia, CS = cesarean section



Description of studies

In table 1a and 1b, the baseline characteristics of the included studies are reported. The studies were published between 1988 and 2016. All of them were prospective cohort studies with a number of patients ranging from 35 to 949.

Different patient characteristics were described for. Average maternal age ranged from 26 to 33 across the studies. The included American Society of Anesthesiologist (ASA) score was mostly ASA I-II. Caranza et al. only included patients with ASA score I and Kaufner et al. included patients with ASA score I-III. Bloor et al. did not specify the ASA scores. The parity of the patient differed between studies. Average maternal weight varied from 65 to 83.8.

One study investigated the effectivity of i.v. PCA morphine and all studies investigated the effectivity of EP and IT morphine. All studies used a Visual Analogue Scale (VAS) as measurement for postoperative pain, but they differed slightly. For example, Chadwick et al. used a 0-100 mm VAS, Lim et al. used a 1-4 VAS and Dualé et al. looked at the 0-100 VAS score throughout the 24 hours observation period and calculated the area under the curve (AUC). Different side-effects were investigated, but all studies looked at pruritus and postoperative nausea and vomiting (PONV). Therefore, we choose to analyze these two side effects only. The dosage of morphine which patients received differed per study.

Table 1a - Baseline characteristics of the included studies

Studies	Type of study	Number of patients	Location	Administration type	Type of results
Chadwick et al. 1988 [12]	Prospective cohort study	399	United States	EP/IT	VAS 0-100. Pruritus, PONV, urinary retention, respiratory depression
Lim et al. 2005 [13]	Prospective cohort study	949	Singapore	EP/IT/PCA	VAS 0-4. Pruritus, PONV, backache, headache
Dualé et al. 2003 [14]	Prospective cohort study	53	France	EP/IT	VAS 0-100. Consumption of IV morphine. Sedation, pruritus, PONV
Bloor et al. 2000 [15]	Prospective cohort study	60	United Kingdom	EP/IT	Pain/pruritus/PONV: VRS (verbal rating scale) 0 (nil), 1 (mild), 2 (moderate) tot 3 (severe)
Caranza et al. 1999 [16]	Prospective cohort study	55	United Kingdom	EP/IT	VAS 0-100. Pruritus, PONV, respiratory depression.
Eskander et al. 1994 [17]	Prospective cohort study	35	United Kingdom	EP/IT	VAS. Pruritus, nausea, vomiting.
Hallworth et al. 1999 [18]	Prospective cohort study	48	United Kingdom	EP/IT	VAS 0-100. VRS 0-3. Pruritus, ventilation frequency, PONV.
Kaufner et al. 2016 [19]	Prospective cohort study	199	Germany	EP/IT	VAS 0-100. Nausea, pruritus.
Sarvela et al. 2002 [20]	Prospective cohort study	146	Finland	EP/IT	VAS 0-10. Pruritus, PONV, respiratory depression

Table 1b - Baseline characteristics of the included studies - Continued

Studies	Maternal age	ASA	Parity	Maternal weight	Dosage of morphine
Chadwick et al. 1988 [12]	IT: 26 ± 5 EP: 26 ± 6	I-II	IT: 17% primagravida 70% multigravida EP: 38% primigravida 50% multigravida	IT: 77 ± 19 EP: 82 ± 24	IT: 0.4 ± 0.1 mg EP: 4.3 ± 0.5 mg
Lim et al. 2005 [13]	Unknown	I-II	Unknown	Unknown	IT: 0.1 mg EP: 3-4 mg PCA: 1 mg bolus, maximum 8-12 mg/h
Dualé et al. 2003 [14]	IT: 32 (29-32) EP: 32.5 (28-35.5)	I-II	IT: 1 (0-1) EP: 1 (0-2)	IT: 65 (61.5-72.5) EP: 69.5 (62.5-79.5)	IT: 0.075 mg EP: 2 mg
Bloor et al. 2000 [15]	IT: 31.7 ± 5.2 EP: 31.4 ± 5.2	Unknown	IT: 1 (0-3) EP: 1 (0-5)	IT: 83.8 ± 15.7 EP: 76.5 ± 15.9	IT: 0.3 mg EP: 3 mg
Caranza et al. 1999 [16]	IT: 30.19 ± 3.83 EP: 31.04 ± 4.78	I	IT: 46% primiparous EP: 34% primiparous	IT: 69.9 ± 9.5 EP: 71.0 ± 10.1	IT: 0.2 mg EP: 3 mg
Eskander et al. 1994 [17]	IT and EP similar	I-II	Unknown	IT and EP similar	IT: 0.75 mg EP: 3.3 mg
Hallworth et al. 1999 [18]	IT: 31.3 (21-40) EP: 31.6 (20-40)	I-II	IT: 1.4 ± 1.1 EP: 1.5 ± 1.0	IT: 79.3 ± 15.1 EP: 76.1 ± 11.7	IT: 0.25 mg EP: 5 mg
Kaufner et al. 2016 [19]	IT: 31 ± 5.4 EP: 31.7 ± 6.3	I-III	IT: 1 (0-1) EP: 1 (0-1)	IT: 82.8 ± 20.2 EP: 83.2 ± 14.2	IT: 0.1 mg EP: 3 mg
Sarvela et al. 2002 [20]	IT: 32 ± 6 EP: 33 ± 4	I-II	IT: 2 (1-5) EP: 2 (1-6)	IT: 79 ± 11 EP: 76 ± 10	IT: 0.2 mg EP: 3 mg

All studies scored high quality in our quality assessment, except for one study. Lim et al. scored moderate quality. There were no articles which scored low quality. For the full outcome of our quality assessment, see Appendix III.

IT versus i.v. PCA morphine

Only one study investigated the effectivity and side effects of i.v. PCA morphine. Lim et al. compared pain scores at rest and on movement when pain was managed with IT, EP or i.v. PCA morphine. The mean pain score was significantly higher in the i.v. PCA group compared to the IT group at rest: i.v. PCA 0.14 (0.06) vs. IT 0.04 (0.01) and on movement: i.v. PCA 0.84 (0.11) vs. IT 0.26 (0.1).

Lim et al. also investigated the occurrence of pruritus, PONV, backache and headache in the IT, EP and i.v. PCA group. Pruritus occurred significantly more in the IT group than in the i.v. PCA group (50% vs. 21%). The other side effects did not occur significantly more in either group.

IT versus EP morphine - VAS score

Because only four of the included studies have comparable VAS results, we only analysed these four in our meta-analysis. Because the studies used VAS scales differently, we calculated the SMD's in order to perform the meta-analysis. The calculated SMD's are given in table 2. For the calculation of the SMD's, see Appendix IV. There was little chance of heterogeneity between these four studies, because the I2 value is 27.98%. Therefore, we used a fixed effects model for this meta-analysis. All four studies showed no significant difference in VAS between EP and IT administration of morphine. When pooled together, the studies therefore did not show a significant difference in pain management: SMD -0.05 (-0.30;0.19) (see figure 2).

The other five articles did not provide sufficient information for inclusion in the meta-analysis, but did discuss the VAS results of the IT and EP groups. Bloor et al. categorised pain as 'nil', 'mild', 'moderate' and 'severe' and found no significant difference in pain scores between the IT and EP group. Hallworth et al. stated that there were no significant differences in VAS at any of their measured time intervals. Caranza et al. also measured VAS at different points in time and found no significant VAS differences between their IT and EP group. The fourth study, Eskander et al., observed significantly lower VAS scores in the IT group compared to the EP group on four different time points. Lastly, Sarvela et al. found significantly lower VAS scores at 24 hours after surgery in the EP group, but found no significant difference in VAS scores between the groups within the first 21 hours after surgery.

Table 2 - Outcome measures of VAS

Studies	IT group (mean)	EP group (mean)	VAS (SMD)	Confidence interval (95%)	I^2
1. Chadwick et al.	23	24	-0.0433	(-0.02396;0.153)	
2. Lim et al.	0.04	0.05	-0.8295	(-0.5469;1.1121)	
3. Dualé et al.	205	165	0.2477	(-0.2986;0.7839)	
4. Kaufner et al.	10	20	-0.8099	(-1.6974;0.0775)	
5. Total			-0.05	(-0.30;0.19)	27.98%

Table 3 - Outcome measures of pruritus

Studies	IT group (n)	EP group (n)	OR	Confidence interval (95%)	I^2
1. Chadwick et al.	155	136	1.60	(1.02;2.50)	
2. Lim et al.	422	30	0.72	(0.41;1.27)	
3. Dualé et al.	9	13	0.65	(0.21;2.01)	
4. Bloor et al.	29	20	14.50	(1.64;128.09)	
5. Hallworth et al.	21	19	1.84	(0.37:9.14)	
6. Caranza et al.	23	20	3.45	(0.79;15.02)	
7. Eskander et al.	13	6	4.77	(1.08;21.10)	
8. Sarvela et al.	42	35	3.60	(1.17;11.05)	
9. Total	714	279	1.87	(0.87;4.05)	63,57%

IT versus EP morphine - Side-effects

The odds ratios (OR) and incidence of pruritus in the IT and EP group per study are given in table 3. In four studies, patients in the IT group experienced significantly more pruritus than patients in the EP group. However, the other four studies did not find a significant difference in the occurrence of pruritus between the IT and EP groups. When pooled together, we did not find a significant difference in the occurrence of pruritus between the groups: OR 1.87 (0.87;4.05) (see figure 3). There was a chance of heterogeneity between these eight studies, because the I2 value is 63.57%. Therefore, we used a random effects model for this meta-analysis. Kaufner et al. recorded the incidence of pruritus in the IT and EP group at several points in time and found no significant difference between the groups.

Table 4 shows the OR's and occurence of PONV in the patient groups per study. Only Lim et al. shows that patients in the EP group experienced significantly more PONV than the patients in the IT group (OR 0,32 (0,17; 0,57)). The other seven studies did not find a significant difference between the groups. We used a random effects model for this meta-analysis, as the I2 value is 74.07%. When pooled together, we did not find a significant difference in the occurrence of PONV between the EP and IT group: OR 0.91 (0.42;1.98).

Figure 2 - Forest plot of the VAS scores (SMD) of IT group compared to EP group per study. The positive horizontal axis represents higher VAS scores in the IT group and the negative horizontal axis represents higher VAS scores in the EP group.

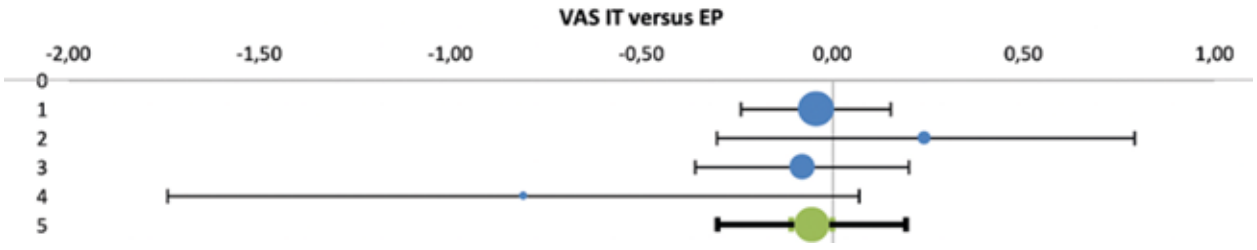


Table 4 - Outcome measures of PONV

Studies	IT group (n)	EP group (n)	OR	Confidence interval (95%)	I^2
1. Chadwick et al.	65	49	1.47	(0.95;2.29)	
2. Lim et al.	131	19	0.32	(0.17;0.57)	
3. Dualé et al.	11	10	1.41	(0.46;4.39)	
4. Bloor et al.	5	7	0.60	(0.16;2.23)	
5. Hallworth et al.	1	6	0.13	(0.01;1.26)	
6. Caranza et al.	19	12	3.85	(1.20;12.33)	
7. Eskander et al.	7	7	0.91	(0.22;3.70)	
8. Sarvela et al.	13	13	1.09	(0.44;2.70)	
9. Total	252	123	0.91	(0.42;1.98)	74.07%

Discussion

We performed a review and meta-analysis, comparing the effectiveness and occurrence of side effects in post-CS patients who received either IT, EP or i.v. PCA morphine. We compared IT to EP administration and IT to i.v. PCA administration.

There was a significantly higher mean VAS score in the i.v. PCA group compared to the IT group in one study, which also found significantly more occurrence of pruritus in the IT group compared to the i.v. PCA group. Other side effects, including PONV, did not show a significant difference in the occurrence between the groups.

When looking at VAS scores, our meta-analysis showed no significant differences in the effectiveness of IT morphine compared to EP morphine. Most of the other reviewed studies support these findings. Looking at side-effects, our analysis shows no significant difference in the occurrence of pruritus in the IT groups compared to the EP groups. We draw the same conclusion for PONV.

Based on our findings, we cannot recommend one type of morphine administration over the other when looking at effectiveness and side-effects.

Current guidelines recommend neuraxial opioids for post-CS pain management [10] , in which intrathecal morphine is the gold standard. [11] Our meta-analysis showed no difference between EP and IT morphine effectiveness and side-effects, but the guideline preferes intrathecal morphine because of lower dosages and therefore lower chance of neonatal drug transfer. [11]

Eskander et al. was the only study to find significantly lower VAS scores in the IT group. However, this study had a small sample size, which might have been a limitation. Concluding from the results of our quality assessment scale, we can say that all articles were at least of moderate quality. So, we conclude that there were no limitations because of the quality of the articles.

Figure 3 - Forest plot of the occurrence of pruritus of the IT group compared to EP group per study. Effect size >1 represents more occurrence in the IT group and effect size <1 represents more occurrence in the EP group.

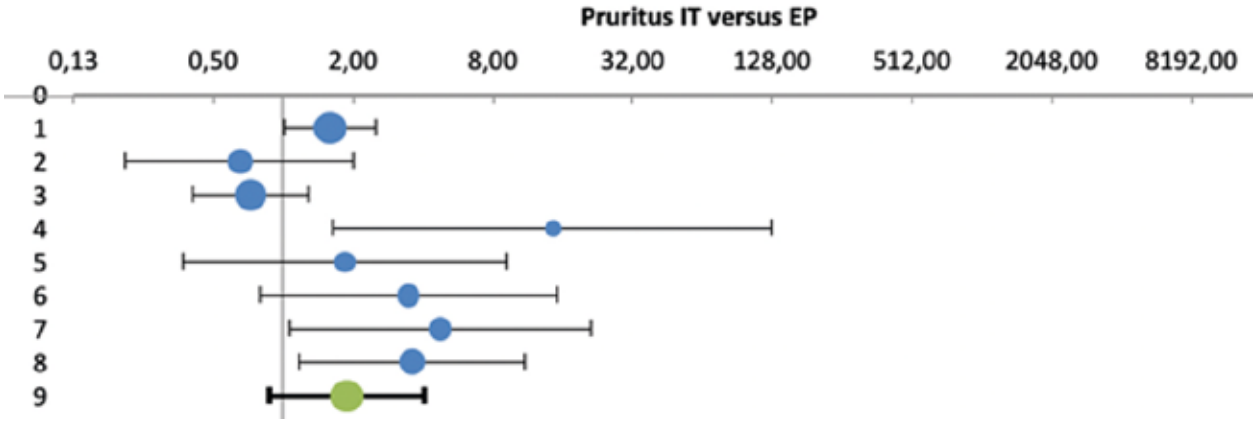
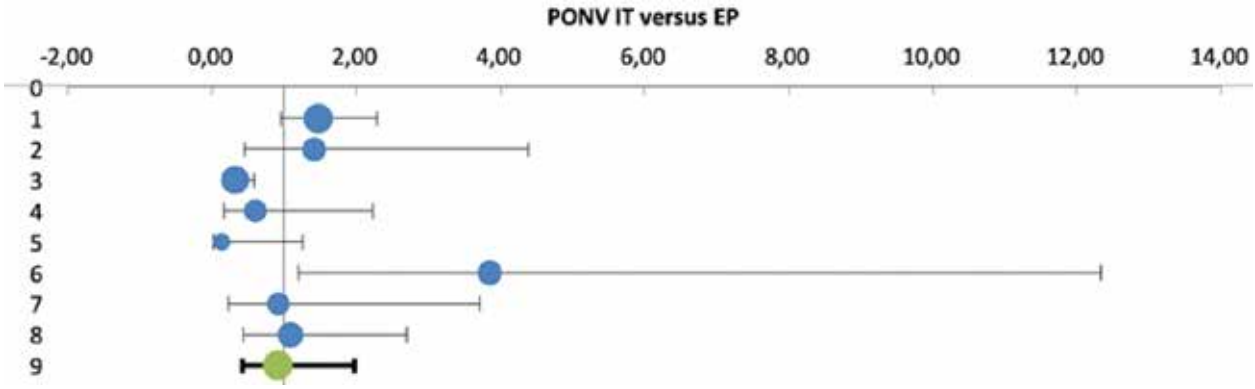


Figure 4 - Forest plot of the occurrence of PONV of the IT group compared to EP group per study. Effect size >1 represents more occurrence in the IT group and effect size <1 represents more occurrence in the EP group.



Our review has several strengths: we used multiple databases to find our set of included articles, we performed three meta-analysis and we altered the NOS to comply with our subject.
However, our article also had several limitations. Despite screening a considerable amount of articles, the amount of articles which we finally included is relatively small. We could not include several of these articles into our meta-analysis, because a limited amount of data was given. Also, VAS scores are used differently in different studies, which made it more difficult to compare.
Based on our findings which show no significant difference in VAS scores and side-effects between the EP and IT groups, we conclude that neither IT and EP morphine is preferred over the other as postoperative analgesia after CS.
More research is necessary to determine if IT morphine is actually superiorly effective to i.v. PCA morphine. We also recommend more research to be done on comparing IT and EP morphine, in order to able to advise on an administration type for clinical practice. This is based on the fact that we were able to include only a small amount of articles, despite our large set of found articles. It is also advisable to research more severe side effects, such as respiratory depression in the mother and the child (via breastfeeding) after administration of IT, i.v. PCA or EP morphine. There were some articles which investigated the incidence of respiratory depression, but to limited extent. The occurrence of these events can weigh more heavily in the recommendation for an administration type of morphine as postoperative pain treatment. We believe prospective cohort studies, similar to the ones included, will be appropriate.

Acknowledgement

We wish to thank Elise Krabbendam from the Erasmus MC Medical Library for developing and updating the search strategies.

References

1. *Betran AP, Ye J, Moller AB et al. The Increasing Trend in Caesarean Section Rates: Global, Regional and National Estimates: 1990-2014. PLoS One. 2016;11:e0148343.*
2. *Dahl Jørgen BMDPD, Jeppesen Inge SMD, Jørgensen HMD et al. Intraoperative and Postoperative Analgesic Efficacy and Adverse Effects of Intrathecal Opioids in Patients Undergoing Cesarean Section with Spinal Anesthesia : A Qualitative and Quantitative Systematic Review of Randomized Controlled Trials. Anesthesiology: The Journal of the American Society of Anesthesiologists. 1999;91:1919-.*
3. *Saldman Lawrence JMDE, Cousins Michael JMDFFARACS, Mather Laurence EMSPD. Intrathecal and Epidural Administration of Opioids. Anesthesiology: The Journal of the American Society of Anesthesiologists. 1984;61:276-310.*
4. *Vercauteren M. Analgesia after Caesarean delivery in the 21st century 2009. Available from: https://www.esahq.org/~/media/ESA/Files/Refresher%20Courses/2009/Analgesia%20after%20Caesarean%20section%20in%20the%2021st%20century%20(2009).ashx.*
5. *Gerancher JC, Floyd H, Eisenach J. Determination of an effective dose of intrathecal morphine for pain relief after cesarean delivery. Anesth Analg. 1999;88:346-51.*
6. *Van de Velde M. What is the best way to provide postoperative pain therapy after caesarean section? Curr Opin Anaesthesiol. 2000;13:267-70.*
7. *Wells GA, Shea B, O'Connell D et al. Coding manual for case-control studies. Available from: http://www.ohri.ca/programs/clinical_epidemiology/nos_manual.pdf*
8. *David B. Wilson PD. Practical Meta-Analysis Effect Size Calculator. 2010.*
9. *Suurmond R, van Rhee H, Hak T. Introduction, comparison, and validation of Meta-Essentials: A free and simple tool for meta-analysis. Res Synth Methods. 2017;8:537-53.*
10. *ACOG Committee Opinion. Postpartum pain management. Obstet Genecol 2018;132.*
11. *Sutton CD, MD, Carvalho, Brendan, MBBCh, FRCA. Optimal Pain Management After Cesarean Delivery. Anesthesiology Clinics 2016;35(1):107-124.*
12. *Chadwick HS, Ready LB. Intrathecal and epidural morphine sulfate for post-cesarean analgesia--a clinical comparison. Anesthesiology. 1988;68:925-9.*
13. *Lim Y, Jha S, Sia AT et al. Morphine for post-caesarean section analgesia: intrathecal, epidural or intravenous? Singapore Med J. 2005;46:392-6.*
14. *Duale C, Frey C, Bolandard F et al. Epidural versus intrathecal morphine for postoperative analgesia after Caesarean section. Br J Anaesth. 2003;91:690-4.*
15. *Bloor GK, Thompson M, Chung N. A randomised, double-blind comparison of subarachnoid and epidural diamorphine for elective caesarean section using a combined spinal-epidural technique. Int J.Obstet Anesth. 2000 Oct;9(4):233-237.*
16. *Caranza R, Teyapalan I, Buggy DJ. Central neuraxial opioid analgesia after caesarean section: Comparison of epidural diamorphine and intrathecal morphine. Int J Obstet Anesth 1999;8(2):90-93.*
17. *Eskander WF, Harvey PB, Kestin IG. Subarachnoid diamorphine compared with extradural diamorphine for postoperative pain relief following elective caesarean section. International journal of obstetric anesthesia 1994;3(3):173-.*
18. *Hallworth SP, Fernando R, Bell R, Parry MG, Lim GH. Comparison of intrathecal and epidural diamorphine for ellective Caesarean section using a combined spinal-epidural technique. Br J Anaesth 1999;82(2):228-232.*
19. *Kaufner L, Heimann S, Zander D, Weizsäcker K, Correns I, Sander M, et al. Neuraxial anesthesia for pain control after cesarean section: A prospective randomized trial comparing three different neuraxial techniques in clinical practice. Minerva Anestesiol 2016;82(5):514-524.*
20. *Sarvela J, Halonen P, Soikkeli A, Korttila K. A double-blinded, randomized comparison of intrathecal and epidural morphine for elective cesarean delivery. Anesth Analg 2002;95(2):436-440.*

Appendix I

Embase – 221 articles
('cesarean section'/exp OR (cesarea* OR cesaria* OR caesarea* OR caesaria* OR ((abdom*) NEAR/3 (deliver*))):ab,ti,kw) AND ('diamorphine'/de OR (diamorphine* OR morphin*):ab,ti,kw) AND (('intrathecal drug administration'/de AND ('epidural anesthesia'/de OR 'patient controlled analgesia')) OR ((intrathecal* AND (epidural* OR extradural* OR peridural* OR epi-dural* OR extra-dural* OR peri-dural* OR ((patient* OR self) NEAR/3 (analges* OR anasthes* OR anesthes* OR anaesthes*))))):ab,ti,kw) NOT ((animal/exp OR animal*:de OR nonhuman/de) NOT ('human'/exp)) AND [english]/lim NOT ([Conference Abstract]/lim)

Medline – 128 articles
(exp Cesarean Section/ OR (cesarea* OR cesaria* OR caesarea* OR caesaria* OR ((abdom*) ADJ3 (deliver*))).ab,ti,kw.) AND (Morphine/ OR (diamorphine* OR morphin*).ab,ti,kw.) AND ((Injections, Spinal/ AND (Injections, Epidural/ OR Analgesia, Epidural/ OR Anesthesia, Epidural/ OR Analgesia, Patient-Controlled/)) OR ((intrathecal* AND (epidural* OR extradural* OR peridural* OR epi-dural* OR extra-dural* OR peri-dural* OR ((patient* OR self) ADJ3 (analges* OR anasthes* OR anesthes* OR anaesthes*))))).ab,ti,kw.) NOT ((animal/ OR animal*:de OR nonhuman/) NOT (human/)) AND english.la. NOT (news OR congres* OR abstract* OR book* OR chapter* OR dissertation abstract*).pt.

Cochrane – 183 articles
((cesarea* OR cesaria* OR caesarea* OR caesaria* OR ((abdom*) NEAR/3 (deliver*))):ab,ti,kw) AND ((diamorphine* OR morphin*):ab,ti,kw) AND (((intrathecal* AND (epidural* OR extradural* OR peridural* OR epi-dural* OR extra-dural* OR peri-dural* OR ((patient* OR self) NEAR/3 (analges* OR anasthes* OR anesthes* OR anaesthes*))))):ab,ti,kw)

Web of Science – 332 articles
TS=(((cesarea* OR cesaria* OR caesarea* OR caesaria* OR ((abdom*) NEAR/2 (deliver*)))) AND ((diamorphine* OR morphin*)) AND (((intrathecal* AND (epidural* OR extradural* OR peridural* OR epi-dural* OR extra-dural* OR peridural* OR ((patient* OR self) NEAR/2 (analges* OR anasthes* OR anesthes* OR anaesthes*)))))) NOT ((animal* OR rat OR rats OR mouse OR mice OR murine OR dog OR dogs OR canine OR cat OR cats OR feline OR rabbit OR cow OR cows OR bovine OR rodent* OR sheep OR ovine OR pig OR swine OR porcine OR veterinar* OR chick* OR zebrafish* OR baboon* OR nonhuman* OR primate* OR cattle* OR goose OR geese OR duck OR macaque* OR avian* OR bird* OR fish*) NOT (human* OR patient* OR women OR woman OR men OR man))) AND DT=(Article OR Review) AND LA=(English)

Google Scholar – 100 articles (top-100)
cesarean|cesarian|caesarean|caesarian morphine intrathecal epidural|extradural|peridurall"patient|self analgesia"

Appendix II

NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE, COHORT STUDIES
Altered by Iris de Lange and Aletta van Opstal.
Note: a study can be awarded a maximum of one point for each numbered item within the categories.

Selection

1) Representativeness IT group
a) Truly representative of the average post-CS patient — 1 point
b) Somewhat representative of the average post-CS patient — 1 point
c) Selected group of users
d) no description of the derivation of the cohort

2) Selection of EP/ i.v. PCA group
a) Drawn from the same community as the IT cohort — 1 point
b) Drawn from a different source
c) No description of the derivation of the EP/ i.v. PCA cohort

3) Ascertainment of administration type
a) Secure records — 1 point
b) Structured interview — 1 point
c) Written self-report
d) No description

Comparability

4) Comparability of IT group and EP/ i.v. PCA group
a) No significant differences in baseline demographics between the study groups — 1 point
b) Significant difference(s) in baseline demographics between the study groups accounted for and no influence on the study results — 1 point
c) Significant difference(s) in baseline demographics between the study groups not accounted for/accounted for and influence on the study results
d) No description

Outcome

5) Assessment of pain score/side-effects
a) Self-report — 1 point
b) Record linkage — 1 point
c) No description

6) Follow-up at least 24h after CS
a) Yes — 1 point
b) No

7) Adequacy of follow up of cohorts
a) Complete follow up - all subjects accounted for — 1 point
b) Subjects lost to follow up - those lost accounted for — 1 point
c) Subjects lost to follow up - those lost not accounted for
d) No statement

Maximum of 7 points:
1-3 points = low quality; 4-5 points = moderate quality; 6-7 points = high quality.

Appendix III

Table 5 - Outcome of quality assessment

Studies	1	2	3	4	5	6	7	Total	Low/ moderate/ high
Chadwick et al.	a	a	a	c	a	a	a	6	high
Lim et al.	a	a	b	d	a	a	d	5	moderate
Dualé et al.	a	a	a	a	a	a	b	7	high
Bloor et al.	b	a	a	a	a	a	d	6	high
Hallworth et al.	a	a	a	a	a	a	b	7	high
Caranza et al.	a	a	a	a	a	a	b	7	high
Eskander et al.	a	a	a	a	a	a	d	6	high
Sarvela et al.	a	a	a	a	a	a	b	7	high
Kaufner et al.	b	a	a	a	a	a	b	7	high

Appendix IV

Table 6 - Standardized mean difference (SMD) calculation

Chadwick et al.	Mean	SD	N	SMD (95% CI)
Spinal	23*	25	200	-0.0433 (-0.02396;0.153)
Epidural	24*	21	199	

** Calculated from 0-100 VAS scale where 100 represents perfect analgesia*

Table 7 - Standardized mean difference (SMD) calculation

Dualé et al.	Mean	SD	N	SMD (95% CI)
Spinal	205	233.75	25	0.2427 (-0.2986;0.7839)
Epidural	156	52.5	28	

Table 8 - Standardized mean difference (SMD) calculation

Lim et al.	Mean	SD	N	SMD (95% CI)
Spinal	0.04	0.01	850	-0.8295 (-0.5469;-1.1121)
Epidural	0.05	0.03	52	

Table 9 - Standardized mean difference (SMD) calculation

Kaufner et al.	Mean	SD	N	SMD (95% CI)
Spinal	10	13	20	-0.8099 (-1.6974;0.0775)
Epidural	20	10	7	

Systematic Review

The Relation between the 24h Urinary Cortisone Concentration and Obesity:

a Systematic Review and Meta-analysis

M. Kerolles[a], A. Alkemade[a], J. van Osch[a], B. van der Voorn[b]

[a]Medical student, Erasmus MC, Rotterdam, The Netherlands.
[b]Dr. Erasmus MC, Rotterdam, The Netherlands.

Correspondence: Makarious Kerolles, email: 464983mk@student.eur.nl

Abstract

Introduction: **Obesity is related to an elevated serum cortisol level but interpreting cortisol levels proves to be difficult. Assessment of a combination of (urinary) cortisol and cortisone could offer a more detailed image of the Hypothalamic-Pituitary-Adrenal (HPA) axis regulation. Approximately half of the patients with obesity show HPA-axis disbalances. The aim of this systematic review is therefore to investigate the value of 24-hour urine cortisone levels and the urinary cortisol/cortisone ratio in otherwise, apart from diabetes mellitus, healthy individuals with obesity.**

Methods: **We used PubMed to search the database of MEDLINE for studies measuring the urine cortisone concentration. We defined the urinary cortisone concentration as our primary outcome parameter and the urinary cortisol/cortisone ratio as our secondary outcome parameter. We did a meta-analysis of the regression between Body Mass Index (BMI) and urinary cortisone concentration and/or the urinary cortisol/cortisone ratio to determine the presence of a correlation between these different markers.**

Results: **We included 5 articles with a total of 314 subjects. The regression analysis showed a significant relation between obesity and the 24-hour urinary cortisone concentrations in otherwise, apart from diabetes mellitus, healthy obese adults ($p < 0{,}01$), but no significant relation between obesity and the 24-hour urinary cortisol/cortisone ratio ($p = 0{,}36$).**

Conclusions: **These results indicate that cortisone concentrations in urine could be used as an easy and non-invasive way to identify HPA axis disbalances caused by obesity in otherwise, apart from diabetes mellitus, healthy individuals.**

Keywords

Obesity, cortisone, urine analysis, adrenal cortex hormones

Introduction

Since 1975 the incidence of obesity has nearly tripled. In 2016, 39% of the world population was overweight and 13% of the population was obese, making obesity a serious upcoming worldwide health problem.[1]

Evidence shows that obesity is associated to an elevated serum cortisol level.[2] Some studies have found an association between high cortisol levels and the prevalence of comorbidities, such as components of the metabolic syndrome [3-5], a greater risk of cardiovascular diseases and depression.[6,7] Accordingly, it seems favorable to measure serum cortisol levels in obese patients, firstly to diagnose underlying causes of obesity and secondly to monitor the effects of obesity. However, although associations between serum cortisol and comorbidities have been found in population studies or certain subgroups, evidence-based cut-off values with clinical consequences are very hard to develop for such a dynamic parameter as serum cortisol.[8]

Measuring cortisol levels can be done in numerous ways but proves to be difficult. Measuring cortisol levels through venipuncture could cause a stress reaction which makes cortisol levels rise.[9] Long-period cortisol levels could be measured by hair analysis, but this is not appropriate for all patients with short hair, as a minimum hair length of three centimeters is needed and even though it is a good indicator for chronic stress, abnormalities in the cortisol day rhythm cannot be identified.[10] Moreover, although more and more used in research over the world, hair steroid analyses can only be performed in few endocrinology laboratories and no clinical cut-off values yet exist other than for Cushing's Syndrome.[8] Also, measuring saliva cortisol poses some issues. Saliva cortisol is a good technique to monitor the cortisol day-rhythm but could also give a wrong impression of cortisol levels if not measured during a complete day.[11]

The golden standard for identifying imbalances in the HPA-axis at this moment is measuring cortisol in 24-hour urine. The level of cortisol in 24-hour urine is only slightly elevated if the patient has underlying pathology in the Hypothalamic-Pituitary-Adrenal (HPA) axis. Although it is a good technique for identifying

big deflections in the HPA axis, it might not be sensitive for subtle changes in the HPA axis.[12]
Cortisone is a non-active hormone, produced from cortisol by 11beta-hydroxysteroid dehydrogenase (11β-HSD).[13] Cortisol and cortisone can be converted in one another by 11β-HSD type 1 and 2, but not into any other active metabolite than cortisol. This makes that cortisone represents the bio-inactive pool of cortisol. Together these two hormones represent the 'total glucocorticoid bioactivity', unlike other metabolites in the cortisol synthesis that still contribute to the production of other glucocorticoid hormones in the adrenal gland.[14,15] This pool of cortisol and cortisone functions as a reinforcement of cortisol measurement to give a complete image of the HPA axis function, as reported by Nomura et al. [16] However, serum cortisone levels are much lower than serum cortisol levels, which makes it harder for measurement and comparison.[17] For hair cortisone measurement the same problems arise as with hair cortisol measurements.[10] An easy, non-invasive measurement of cortisone could be through saliva. Our initial plan was to study salivary cortisone excretion, but unfortunately our PubMed search did not deliver enough useful material to write a systematic review (search: "Cortisone"[Mesh] AND "Saliva"[Mesh], filters: 'Language: English', performed at 01-10-2019).
Accordingly, in order to explore the use of monitoring cortisone to interpret disbalances in the HPA axis, we investigated the use of 24-hour urine cortisone. As the 24-hour cortisol value is the standard for monitoring cortisol, cortisone could easily additionally be measured. Next to this, other bodily materials pose some issues.
As there are many issues with measuring cortisol levels, we were wondering if testing another metabolite could enhance the function of 24-hour urine cortisol levels as a marker for HPA axis disbalances in patients with obesity. For this review, we will discuss the available literature about the possibilities of using cortisone as an indicator of HPA axis disbalances. In individuals with obesity this combination of tests could help with identifying underlying pathology or monitor interventions.
We have therefore chosen to investigate the potential of 24-hour urine cortisone levels in relation to BMI as an indicator for raised cortisol levels in order to find if cortisone measurements could help in identifying and monitoring imbalances in the HPA axis in otherwise, apart from diabetes mellitus, healthy obese adults.

Methods

Search strategy
For this study we used PubMed to search the database of MEDLINE with the following search terms: "Overweight" [Mesh] AND "Cortisone" [Mesh]. We conducted the search on October 1, 2019.

Inclusion criteria
We included studies that reported data about urine cortisone in adults with obesity. We only included studies written in English and performed on humans. We used the filters provided by PubMed to filter the studies found by our search terms for studies not written in English or performed on humans. The search was not limited by publication date.

Exclusion criteria
We excluded 1) studies that did not report any data about the direct relation between urine cortisone and obesity, 2) studies that included patients having an underlying pathology of the HPA-axis (including pregnancy) and 3) studies that were an intervention trial involving medication.

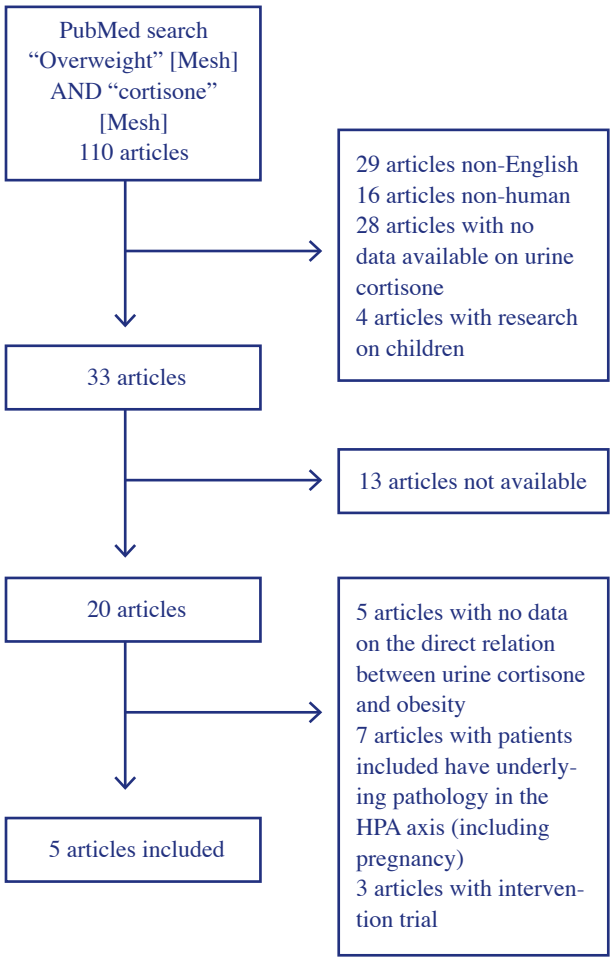
Study selection
We independently assessed the titles and abstracts of all studies found with the search terms in order to include or exclude the articles according to our inclusion and exclusion criteria. We assessed the full text articles if necessary.

Quality assessment
The quality of the included studies was independently assessed by three reviewers using an adjusted form of the Newcastle-Ottawa Quality Assessment Form for Cohort Studies. All articles were assessed in three domains (selection, comparability and outcome). Not all items of the quality assessment were suitable for all articles due to the different study designs, so certain items were left out. Based on the total score of all domains each article was given a GOOD, FAIR or POOR (see appendix).

Data extraction
Three reviewers independently extracted the following data from the included articles: baseline characteristics including age, sex and BMI of the participants, study design, urinary free cortisol, urinary free cortisone and the urinary cortisol/cortisone ratio. We defined the urinary cortisone concentration as our primary outcome parameter. We defined the urinary cortisol/cortisone ratio as our secondary outcome parameter. If the urinary cortisol/cortisone ratio was not mentioned by a study, we calculated the urinary cortisol/cortisone ratio by dividing the urinary free cortisol by the urinary free cortisone. We used the urinary 5α-tetrahydrocortisol (5αTHF) plus the urinary 5β-tetrahydrocortisol (5βTHF) divided by the urinary tetrahydrocortisone (THE) ((5αTHF + 5βTHF)/THE) when studies mentioned neither the urinary cortisol/cortisone ratio, nor the urinary free cortisone and the urinary free cortisol (see discussion for elaboration). We used a meta-analysis to determine the presence of a significant correlation between the BMI and the urinary cortisol/cortisone ratio or the urinary cortisone.

Statistical analysis
We used Interactive Data Language (IDL) 7.0 (2008) retrieved from (www.exelisvis.com/ProductsServices/IDL.aspx) to perform our data analysis. We have entered the data of the BMI, cortisone and cortisol concentrations and performed a linear regression analysis. A 95% confidence interval was selected.

Figure 1 - Flowchart of the PubMed search



Results

Search

The PubMed search resulted in 110 articles. 33 articles remained after applying our inclusion criteria through the selected PubMed filters. 5 articles remained after applying our exclusion criteria (figure 1).

Baseline characteristics

Our included studies were published between 1998 and 2013. The included studies have a total of 314 participants with an age that varied from 26,5 to 59,1 years (table 1).

Quality assessment

The quality assessment of these articles resulted in FAIR's for Andrew et al. 1998, Mussig et al. 2009 and Rask et al. 2013, GOOD for Mussig et al. 2008 and POOR for Rask et al. 2001 (see appendix for the full table).[18-22] We did decide to include Rask et al. 2001 in our results, even though it was given a POOR.[22] We were not able to score all the criteria of the quality assessment in Rask et al. 2001 due to missing data, which resulted in a low score.[22] To see how much Rask et al. 2001 influenced the results, we conducted a sensitivity analysis.[22] This showed that the omission of the data from Rask et al. 2001 had no effect on the significance of both outcome measures.[22] The p-value of the correlation between BMI and the cortisone concentration remained the same, the p-value of the correlation between BMI and the cortisol/cortisone ratio went from 0,38 to 0,5.

BMI versus urinary cortisone and the urinary cortisol/cortisone ratio

Table 3 shows the BMI, urinary cortisone, urinary cortisol and urinary cortisol/cortisone ratio of all participants. In order to determine if a significant correlation between BMI and the urinary cortisone was present, we did a linear regression analysis in (figure 2A). The slope in the trend line was 0,00234 (95% confidence interval (CI), 0,0014 to 0,0033). The p-value of the regression was <0,01. This suggests that the urine of obese subjects contains significantly more cortisone than the urine of the lean subjects. Additionally, we performed a regression analysis of the correlation between BMI and the urinary cortisol/cortisone ratio (figure 2B). The slope in the trend line was -0,00494 (95% CI, -0,016 to 0,0062). The p-value was 0,38. The significant higher cortisone concentration and the unchanged cortisol/cortisone ratio suggest that the cortisol concentration

Table 1 - Baseline characteristics and study designs

Study	Study design	Age in years (control group)	BMI in kg/m2 (control group)	Number of participants (control group)	Age in years (obese group)	BMI kg/m² (obese group)	Number of participants (obese group)
Rask et al. 2013 [20]	Prospective cohort study	42,4 (± 8,7)	30,3 (± 5,1)	31*	40,0 (±8,5)	44,6 (± 4,5)	31*
Müssig et al. 2009 [19]	Case control study	31,5 (± 5)	22,1 (± 1,8)	20	40 (± 13)	45,3 (± 8,9)	59
Müssig et al. 2008 [21]	Case control study	30,3 (± 1,0)	22,3 (± 0,3)	30	39,3 (± 1,4)	45,5 (± 1,1)	72
Rask et al. 2001 [22]	Prospective cohort study	Lowest tertile of BMI: 46,8 (± 8,7) Middle tertile of BMI: 49,6 (± 8,5)	Lowest tertile of BMI: 22,9 (± 1,4) Middle tertile of BMI: 26,4 (± 0,7)	Lowest tertile of BMI: 11 Middle tertile of BMI: 11	Highest tertile of BMI: 51,9 (± 12,1)	Highest tertile of BMI: 31,7 (± 4,0)	Highest tertile of BMI: 12
Andrew et al. 1998 [18]	Cross- sectional study	Males: 51,9 (± 2,6) Females: 52,0 (± 2,5)	Males: 26,6 (± 3,4) Females: 25,2 (± 4,1)	Males: 31 Females: 37	N/A	N/A	N/A

*N/A = Not available. *Participants in the control group are the same participants as in the obese group but after weight loss.*

Figure 2a - Regression analysis of the BMI plotted against cortisone concentration in 24h urine

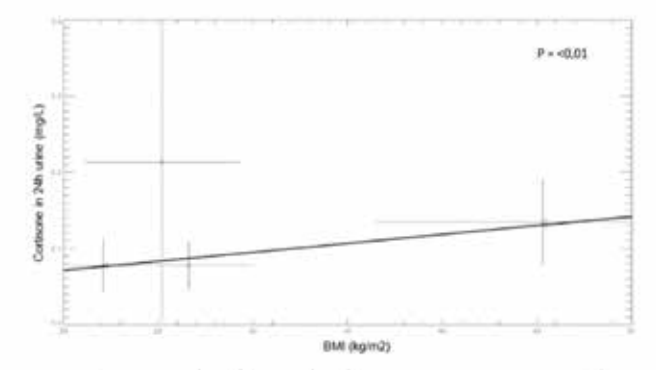
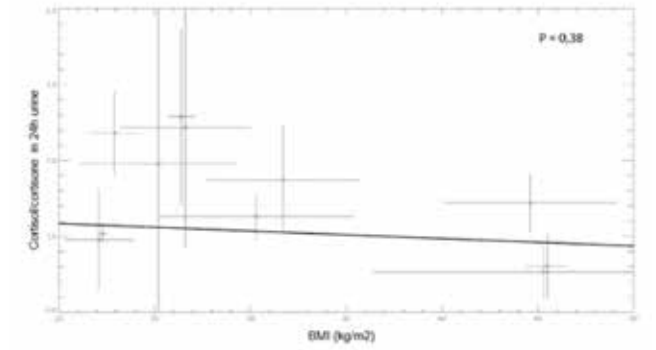


Figure 2b - Regression analysis of the BMI plotted against cortisol/ cortisone ratio in 24h urine



was elevated as well. This might indicate that 11β-HSD1 and 11β-HSD2 are more active in obese patients. The literature does not provide a clear picture of cortisol elevation at a higher BMI, although it is thought that a high cortisol level is more common in people with obesity.[2,23-25]

Discussion

Conclusion

In this study we tried to determine if the urinary cortisone is a relevant parameter to identify and monitor underlying HPA axis disbalances in subjects with obesity. Our analysis showed a significant relation between obesity and the urinary cortisone concentrations in otherwise, apart from diabetes mellitus, healthy adults with obesity, but no significant relation between the urinary cortisol/cortisone ratio and obesity. These results indicate that cortisone concentrations in urine could be used in addition to cortisol concentrations in urine to determine HPA axis disbalance in relation with obesity. It is important to determine the biological glucocorticoid activity (and thus the clinical burden) and to understand where the problem in the HPA-axis occurs. After all, a HPA-axis disbalance can result from differences in production, metabolism and breakdown. Cortisone has the potential to be a more stable predictor for HPA-axis disturbances and might therefore offer additional information to cortisol in urinary analyses of patients with obesity.[8] If these measurements proof to be useful in the clinical practice, cortisone levels could be measured relatively easy because the equipment for measuring cortisone and cortisol is the same and the measurements can be done simultaneously. While more and more patients with obesity get offered therapy all over the country, this might be a feasible predictor in personalized diagnostics and treatment.

Limitations of our study

The measuring methods in the included studies differed. Rask et al. 2013 used gas liquid chromatography (GLC), Müssig et al. 2009 and Müssig et al. 2008 used a radioimmunoassay (RIA), Andrew et al. 1998 used electron impact gas chromatography/mass spectrometry and Rask et al. 2001 used gas chromatography and electron impact mass spectrometry.[18-22] A recent study reviewed different methods of measuring steroids and concluded that mass spectrometry is more reliable compared to immunoassays.[26] It is therefore expected that the data from Müssig et al. 2009 and Müssig et al. 2008 is less accurate.[19,21] Furthermore, we assumed a linear correlation between the cortisone concentration and cortisol/cortisone ratio. To confirm you would ideally do a cross-sectional cohort study with a chi-squared regression analysis but given the limited amount of data this was not feasible. We were not able to take all baseline characteristics into consideration for our meta-analysis. We advise future studies to take them into account.

Only Mussig et al. 2008 corrected their data for age and gender, whereas the others have not.[21] The data may therefore not be representative. Furthermore, Mussig et al. wrote two articles, one in 2008 and one in 2009. Both articles use a study population of obese patients from the same hospital and it is therefore likely that this is partly the same group of patients, although this is not further discussed in the articles. To see if the exclusion of one of these articles influenced the results, we did a sensitivity analysis in which Mussig et al. 2009 (the article with the lowest quality assessment) was omitted.[19] The analysis showed that the omission has no influence on the significance of the regressions. The p-value of the correlation between BMI and the cortisone concentration remained the same, the p-value of the correlation between BMI and the cortisol/cortisone ratio went from 0,38 to 0,91.

Rask et al. 2001 was valued as POOR in the quality assessment, mainly due to a lack of: 1) correction for gender, 2) follow-up and 3) 'Demonstration that outcome of interest was not present at start of study' (see appendix).[22] We nevertheless decided to include this study, because the raw data Rask et al. 2001 provided was useful for our analysis.[22] However, there is a chance that the data from this study is not representative. In addition, Rask et al. 2001 used different metabolites to measure the cortisol/cortisone ratio, namely (5αTHF + 5βTHF) / THE.[22] We used this ratio after discussion, because research showed that these two ratios do match.[27]

In Andrew et al 1998 the data for the 24-hour urinary cortisol and cortisone was provided, as was the cortisol/cortisone ratio.[18] However, the given ratio did not match the outcome of dividing the cortisol and cortisone concentrations given in the article. Because Andrew et al. gave no explanation about this difference, we decided to use our manually calculated ratio, since

Table 2 - Quality assessment of the included articles

Adjusted Newcastle-Ottawa Quality Assessment Form for Cohort Studies					
Article	**Andrew et al. 1998 [18]**	**Mussig et al. 2008 [21]**	**Mussig et al. 2009 [18]**	**Rask et al. 2001 [22]**	**Rask et al. 2013 [19]**
SELECTION					
Representativeness of the exposed cohort	1	1	1	1	1
a) Truly representative (one star)					
b) Somewhat representative (one star)	x	x	x	x	
c) Selected group					x
d) No description of the derivation of the cohort					
Selection of the non-exposed cohort	N/A	1	1	N/A	N/A
a) Drawn from the same community as the exposed cohort (one star)		x	x		
b) Drawn from a different source					
c) No description of the derivation of the non exposed cohort	-			-	-
Ascertainment of exposure	1	1	1	1	1
a) Secure record (e.g., surgical record) (one star)	x	x	x	x	x
b) Structured interview (one star)					
c) Written self report					
d) No description					
e) Other					
Demonstration that outcome of interest was not present at start of study	1	1	0	0	1
a) Yes (one star)	x	x			x
b) No			x	x	
SELECTION TOTAL	3	4	3	2	3
COMPARABILITY					
Comparability of cohorts on the basis of the design or analysis controlled for confounders	1	1	0	0	0
a) The study controls for age, sex and marital status (one star)					
b) Study controls for other factors (one star)	x	x			
c) Cohorts are not comparable on the basis of the design or analysis controlled for confounders			x	x	x
COMPARABILITY TOTAL	1	1	0	0	0

OUTCOME					
Assessment of outcome	1	1	1	1	1
a) Independent blind assessment (one star)					
b) Record linkage (one star)	x	x	x	x	x
c) Self report					
d) No description					
e) Other					
Was follow-up long enough for outcomes to occur	N/A	N/A	N/A	N/A	1
a) Yes (one star)					x
b) No					
Indicate the median duration of follow-up and a brief rationale for the assessment above:	N/A	N/A	N/A	N/A	Two years after gastric bypass
Adequacy of follow-up of cohorts	N/A	1	N/A	N/A	1
a) Complete follow up- all subject accounted for (one star)					
b) Subjects lost to follow up unlikely to introduce bias- number lost less than or equal to 20% or description of those lost suggested no different from those followed. (one star)		x			x
c) Follow up rate less than 80% and no description of those lost					
d) No statement					
OUTCOME TOTAL	1	2	1	1	3
OVERAL TOTAL	4	7	4	3	6
QUALITY (0-3 = POOR / 4-6 = FAIR / >6 = GOOD)	FAIR	GOOD	FAIR	POOR	FAIR

Table 3 - BMI, urinary cortisone, urinary cortisol and urinary cortisol/cortisone ratio of the participants

Study	BMI (kg/m2)	Cortisone (mg/24h urine)	Cortisol (mg/24h urine)	Cortisol/cortisone
Rask et al. 2013 [20]	30,3 (± 5,1)	N/A	N/A	0,63 (± 0,15)
	44,6 (± 4,5)	N/A	N/A	0,72 (± 0,19)
Müssig et al. 2009 [19]	22,1 (± 1,8)	0,0772 (± 0,0347)	0,0369 (± 0,0196)	0,4780 (± 0,48)*
	45,3 (± 8,9)	0,135 (± 0,056)	0,036 (± 0,019)	0,2667 (± 0,18)*
Müssig et al. 2008 [21]	22,3 (± 0,3)	0,0776 (± 0,0062)	0,0401 (± 0,0036)	0,5168 (± 0,06)*
	45,5 (± 1,1)	0,1319 (± 0,0091)	0,0399 (± 0,0028)	0,3025 (± 0,21)*
Rask et al. 2001 [22]	22,9 (± 1,4)	N/A	N/A	1,18 (± 0,28)**
	26,4 (± 0,7)	N/A	N/A	1,29 (± 0,58)**
	31,7 (± 4,0)	N/A	N/A	0,87 (± 0,37)**
Andrew et al. 1998 [18]	25,2 (± 4,1)	0,213 (± 0,208)	0,218 (± 0,144)	1,023 (± 1,15)*
	26,6 (± 3,4)	0,078 (± 0,031)	0,095 (± 0,049)	1,218 (± 0,79)*

*N/A = Not available. *Calculated based on the urinary cortisol and cortisone. **(5αTHF + 5βTHF) / THE*

this was the same method as we used on the other studies where they only provided the urinary cortisone and cortisol concentrations.[18]

In addition, we only searched the MEDLINE database for articles. We have done a snowball search with the references of the included articles and nothing relevant came out of this. Nevertheless, it is possible that we may have missed articles that were only available through other databases such as EMBASE and Cochrane. Also, by only including articles written in English we may have missed articles that were written in another language.

The final point of criticism of our research relates to the analysis of the cortisol/cortisone ratio. The study of Mussig et al. 2008 has the smallest standard error. The difference between the cortisol/cortisone ratio in lean and obese subjects is significant in the analysis of Mussig et al. 2008.[21] However, when we added the data from the other included studies in our analysis, this correlation disappeared. The study of Mussig et al. 2008 is the most reliable, based on both the number of participants and our quality assessment.[21] It is therefore likely to assume a correlation between the urinary cortisol/cortisone ratio and the BMI, although our analysis does not show it.

Recommendations

We suggest further research into this possible connection between urinary cortisone, cortisol and obesity with a cross-sectional cohort study with a larger number of healthy (e.g. no HPA disbalances), adult participants and complete correction for baseline characteristics to increase the power of the study. With the BMI, urinary cortisone and cortisol concentration, a chi-squared regression analysis can be performed to more accurately analyze the relationship between them. We additionally suggest mass spectrometry as the measuring method of the urinary cortisone and cortisol concentration given it is the most reliable method for steroids.[23] We also suggest further research in the relation between the salivary cortisone/cortisol ratio and obesity, as collecting saliva is more practical compared to collecting 24-hour urine. Eventually these investigations could lead to implementation of salivary and urinary cortisone as a diagnostic test or monitoring tool in the clinical guidelines for obesity, not only at academic level but also by a general practitioner.

References

1. *Obesity and Overweight. 19-02-2018; Available at: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight.*
2. *Schorr M, Lawson EA, Dichtel LE et al., Cortisol Measures Across the Weight Spectrum. J Clin Endocrinol Metab 2015 Sep ; 1 00: 3313-3321.*
3. *Hackett RA, Kivimaki M, Kumari M et al., Diurnal Cortisol Patterns, Future Diabetes, and Impaired Glucose Metabolism in the Whitehall II Cohort Study. J Clin Endocrinol Metab 2016 Feb ; 101: 619-625.*
4. *Martinez-Aguayo A, Campino C, Baudrand R et al., Cortisol/cortisone ratio and matrix metalloproteinase-9 activity are associated with pediatric primary hypertension. J Hypertens 2016 Sep ; 34: 1808-1814.*
5. *Trevino Villarreal DC, Lopez Guevara V, Ramirez Lopez LE et al., Relationship of serum cortisol and metabolic syndrome components, dietary intakes and anxiety disorder in children at 8 to 12 years of age with obesity. Nutr Hosp 2012 Sep-Oct ; 27: 1562-1568.*
6. *Nakajima Y, Yamada M, Taguchi R et al. Cardiovascular complications of patients with aldosteronism associated with autonomous cortisol secretion. J Clin Endocrinol Metab 2011 Aug ; 96: 2512-2518.*
7. *Pariante CM, Lightman SL. The HPA axis in major depression: classical theories and new developments. Trends Neurosci 2008 Sep ; 31: 464-468.*
8. *Savas M, Wester V, Rijke Y, et al., Hair Glucocorticoids as a Biomarker for Endogenous Cushing's Syndrome: Validation in Two Independent Cohorts. Neuroendocrinology 2019 Feb 13 ; 109: 171-178.*
9. *Bellitti P, Valeriano R, Gasperi M et al., Cortisol and heart rate changes in first- and fourth-time donors. Vox Sang 1994 ; 67: 42-45.*
10. *Binz TM, Braun U, Baumgartner MR et al., Development of an LC-MS/MS method for the determination of endogenous cortisol in hair using (13)C3-labeled cortisol as surrogate analyte. J Chromatogr B Analyt Technol Biomed Life Sci 2016 Oct 15 ; 1033-1034: 65-72.*
11. *Kristenson M, Garvin P, Lundberg U. The Role of Saliva Cortisol Measurement in Health and Disease. 2011 ; Bentham Science Publishers.*
12. *Kidambi S, Raff H, Findling JW. Limitations of nocturnal salivary cortisol and urine free cortisol in the diagnosis of mild Cushing's syndrome. Eur J Endocrinol 2007 Dec ; 157: 725-731.*
13. *Pavlovic R, Cannizzo FT, Panseri S et al., Tetrahydro-metabolites of cortisol and cortisone in bovine urine evaluated by HPLC-ESI-mass spectrometry. J Steroid Biochem Mol Biol 2013 May ; 135: 30-35.*
14. *Auzeby A, Bogdan A, Touitou Y. Evidence for a new biologic pathway of androstenedione synthesis from 11-deoxycortisol. Steroids 1991 Jan ; 56(1): 33-36.*
15. *Boucher E, Provost PR, Tremblay Y. Ontogeny of adrenal-like glucocorticoid synthesis pathway and of 20alpha-hydroxysteroid dehydrogenase in the mouse lung. BMC Res Notes 2014 Mar 1 ; 7: 119-0500-7-119.*
16. *Nomura S, Fujitaka M, Jinno K, et al., Clinical significance of cortisone and cortisone/cortisol ratio in evaluating children with adrenal diseases. Clin Chim Acta. 1996 ; 256: 1–11.*
17. *Lewicka S, von Hagens C, Hettinger U et al., Cortisol and cortisone in human follicular fluid and serum and the outcome of IVF treatment. Hum Reprod 2003 Aug ; 18: 1613-1617.*
18. *Andrew R, Phillips DI, Walker BR. Obesity and gender influence cortisol secretion and metabolism in man. J Clin Endocrinol Metab 1998 May ; 83: 1806-1809.*
19. *Mussig K, Remer T, Fritsche A et al., Urinary free cortisone, but not cortisol, is associated with urine volume in severe obesity. Steroids 2009 Sep ; 74: 742-745.*
20. *Rask E, Simonyte K, Lonn L et al., Cortisol metabolism after weight loss: associations with 11 beta-HSD type 1 and markers of obesity in women. Clin Endocrinol (Oxf) 2013 May ; 78: 700-705.*
21. *Mussig K, Remer T, Haupt A et al., 11beta-Hydroxysteroid Dehydrogenase 2 Activity is Elevated in Severe Obesity and Negatively Associated with Insulin Sensitivity. Obesity (Silver Spring) 2008 Jun ; 16: 1256-1260.*
22. *Rask E, Olsson T, Soderberg S et al., Tissue-specific dysregulation of cortisol metabolism in human obesity. J Clin Endocrinol Metab 2001 Mar ; 86: 1418-1421.*
23. *Abraham SB, Rubino D, Sinaii N et al., Cortisol, obesity, and the metabolic syndrome: a cross-sectional study of obese subjects and review of the literature. Obesity (Silver Spring) 2013 Jan ; 21: E105-17.*
24. *Lopez-Aguilar I, Ibarra-Reynoso LDR, Malacara JM. Association of Nesfatin-1, Acylated Ghrelin and Cortisol with Scores of Compulsion, Food Addiction, and Binge Eating in Adults with Normal Weight and with Obesity. Ann Nutr Metab 2018 ; 73: 54-61.*
25. *Tenk J, Matrai P, Hegyi P et al., In Obesity, HPA Axis Activity Does Not Increase with BMI, but Declines with Aging: A Meta-Analysis of Clinical Studies. PLoS One 2016 Nov 21 ; 11: e0166842.*
26. *Zendjabil M, Chellouai Z, Abbou O. Role of mass spectrometry in steroid assays. Ann Endocrinol (Paris) 2016 Feb ; 77: 43-48.*
27. *Mongia A, Vecker R, George M, et al., Role of 11betaHSD type 2 enzyme activity in essential hypertension and children with chronic kidney disease (CKD). J Clin Endocrinol Metab 2012 Oct ; 97: 3622-3629.*

Instructions for EJM authors

General

The instructions that follow have several purposes. First, we want to make life easy for you, the authors, and for the editors and peer reviewers, the layout (prepress) people, and the journal readers.

The section Authors instructions storyline, on the website (www.erasmusmc.nl/erasmusjournalofmedicine) will help you to organize your article in a logical, credible and readable way. This will help you - it tells you what goes where—and, thus, save you time. It will help the editors and peer reviewers—they will easily see the credibility and relevance of your work— and, thus, save them from writing rejection letters. And, it will help readers to quickly and easily read and understand your work and see its value.

The section entitled Formatting Instructions will help you as well; the basic idea is to keep the formatting as simple as possible, so you can focus on content and not get involved with layout. The language editor and the prepress people will also be able to more efficiently do their jobs. Please follow these instructions.

Please be aware that we will have to return papers that do not conform to these instructions to the authors.

What you can enter

Research news - Research articles describe one study or analysis, usually from an elective research project or one of the masters programs. Number of words: max. 3500 + 4 figures or tables.

Extended abstracts - Extended abstracts consist of a condensed presentation of final or preliminary results of a study. Extended abstracts can concern ongoing research that is not yet published elsewhere which is comparable with a congress presentation thus does not require copyright transfer. An extended abstract can also be submitted after publication in another Journal if possible with extra figures, this does require proper referencing. Number of words: 350 words + 1 figure or table.

Research papers - Here researchers or teachers describe ongoing research projects at the Erasmus Medical centre for which they want to invite students to participate. Number of words: 350.

Systematic reviews - A systematic review is a literature review focused on a research question that tries to identify, appraise, select and synthesize all high quality research evidence relevant to that question in a quantitative way. Systematic reviews of high-quality randomized controlled trials are crucial to evidence-based medicine, and are considered very important by the editorial board of EJM. Besides health interventions, systematic reviews may concern clinical tests, public health interventions, social interventions, adverse effects, and economic evaluations. Number of words: 3000 + 3 figures or tables.

Opinion papers - These are papers that reflect the opinion of the author on a scientific topic. The author should be clear where evidence ends and personal opinion starts. A paper typically has a length of about 1000 words.

Clinical lesson/question - A clinical lesson should present a scenario and a concrete related question about a disease or condition, the article should elaborate on possible approaches or treatment options for this disease or condition. Conclusion should provice a solid evidence based conclusion on the preferred approach or treatment. Number of words: 1000 + 1 figure or table.

Case reports - A case report consists of the initial presentation, medical history, examination, tests performed, eventual outcome and discussion on the case backed up by scientific literature. Number of words: 900 + 1 figure or table.

Clinical quiz - A clinical quiz should present a scenario and a concrete related question about the disease or condition, preferably accompanied by a clinical image, and four plausible treatment options or courses of action. Conclusion should elaborate on which is the correct option and why. Number of words: 600 + 1 figure or table.

Clinical images - Clinical images should present a typical abnormality on a photograph/imaging tests of a patient or on an additional investigation. It must be accompanied by an elaboration on the clinical diagnosis. Number of words: 350 + 1 figure. Make sure that the patient is not identifiable or that the data presented traceable to the patient. Additionally, written consent should be obtained from presented patient. We expect the author to refer to scientific literature to back up their case presentations.

Comments - In this section editors, or faculty staff, as well students are invited to write a short critical comment on a paper, putting it into perspective for a broader medical public readership. Number of words: 350.

Letters to the editor - The editorial board encourages students to write a letter to the editor to comment on published papers, or on the journal in general. These will be published on the website of the journal. Letters should not exceed 200 words and may be abbreviated by the editor.

The review process

Papers may be submitted to the editorial office. Please indicate which author will act as corresponding author. We expect this author to maintain contact with the other authors and to speak and decide on their behalf.

Each paper will be assigned to a team consisting of a managing editor and an associate editor. Each submitted paper will be checked for compliance with the author instructions. If this is not the case, the paper may be returned to the author.

Instructions for EJM authors

When the paper is taken into review, it will be sent out to two external reviewers, a student and a staff member of Erasmus MC. Based upon these reviewers comments, their recommendations and the opinion of the editorial team, a decision will be made: reject, major revision, minor revision, accept with or without minor changes.

The paper will then be returned to the corresponding author, along with the recommendation. We try to return papers within 3 weeks after submission. When a paper is rejected, it cannot be resubmitted, but we encourage resubmissions when we recommend major or minor changes to a paper. Resubmitted paper will be reviewed again by the same reviewers and editorial team.

Before a paper can be accepted for publication, we will need a statement that the staff member that supervised your work agrees with the submission of your paper. Moreover, we need a signed Copyright Transfer Agreement (CTA) and a signed Conflict of Interest statement. When your research project involves patients or volunteers, we need a statement in the paper that the research protocol has been reviewed by a Medical Ethics Committee. Failure to provide this information at an early stage of the submission may impair the review process.

When a paper is accepted for publication, it will often be forwarded to our language editing and restructuring editors. They will each in turn give recommendations and ask the author adapt the paper accordingly. When this phase is completed, the paper will be forwarded to the publisher. Page proofs will be sent to the author for a final check.

Formatting instructions

Entry format - Papers should be submitted by email, to ejm@erasmusmc.nl. Word 2007 files are preferred for the initial submission. The file should include all figures and tables.

Title page - The title page should clearly identify the authors, the institute where the research project was carried out, as well as the staff member who supervised the project. The corresponding author name (first name and family name), email address, student id, should be clearly indicated. In case of multiple authors, state functions and departments only in superscript in alphabetical order.

Example:
First name A.G. Family name[a] and First name W.F. Family name[a] Supervisor: First name R. Lastname[b]
[a] Medical students, Erasmus MC University Medical Center Rotterdam, the Netherlands
[b] Dept. of Internal Medicine, Erasmus MC University Medical Center Rotterdam, the Netherlands
Correspondence: First name A.G. Family name, email: FirstnameFamilyname@me.com.

Structure - Please use the following sections in all papers (except in comments and opinion papers): Abstract, Introduction, Methods, Results, Discussion, References, Tables, Figures.

References - Number references in order of appearance. References should have the following format:
Rothwell, P. M. Medical and surgical management of symptomatic carotid stenosis. Int J Stroke. 2006; 1: 140-149. (I.e. year;vol:ppp-ppp) In case of more than 3 authors, name the first 3 and insert "et al.". Limit the number of references to 30. References should appear in the text as follows: "… treatment is of proven benefit.[1]"

Tables and figures - Tables and illustrations (both numbered in Arabic numerals) should be prepared on separate pages. Number tables and figures separately and consecutively. Tables require a heading and figures a legend, also prepared on a separate page and should be formatted with a text editor (example). Figures should be submitted electronically. B/w half-tone and color illustrations must have a final resolution of 300 dpi after scaling, line drawings one of 800-1,200 dpi (jpg and tiff is an acceptable format). Please note that all color-figures will be converted to gray tones. Please adapt graphs to suit this format, i.e. make use of dotted and dashed lines and hatched bars instead of colored items.. The final submission should contain figures as JPG or TIFF files.

Page layout
- Standard margins
- no headers or footers
- no columns
- left align (ragged right)
- font: 12pt Arial
- single line spacing
- main headings 14 pt bold; subheading 12 point italic
- indent every paragraph, except after headings, tables, bulleted lists or figures

Other formatting
- number all tables and figures sequentially
- place tables and figures at the end of article; insert captions at correct locations in body text
- no text boxes
- no footnotes or end notes
- do not submit figures with text as drawing objects (they cannot be edited)
- limit the use of italics and do not use italics for simple emphasis; do not italicize quotations; quotation marks are sufficient
- do not use italics for commonly understood Latin expressions such as "in vitro"
- use italics for other foreign words, such as expressions in Dutch
- no "sub-paragraphs"
- no hyphenation (afbreking)

Language
US English spelling and punctuation

The template for authors

Introduction

1. *What is the health-related problem that your research helps to solve?*

2. *What is your strategy to solve the problem?*

3. *What is your research question/hypothesis?*
   Whether a question or a hypothesis, state it in terms of 2 items:
   - variables: the measurable/observable independent and outcome variables that you measured/observed and
   - relationships: the relationships between those variables that your data analyses were designed to determine.

4. *The core concept of the methods you used to answer the research question*
   Briefly describe the core concept of the methods at the end of the Introduction section. This helps readers to understand the complex details that are then presented in the Methods section

Methods section

Organize the details of the Methods section under subheadings. Possible subheadings:

What was studied and study design (subheading)

Describe the details of
- what was studied: sample from a patient/animal population, and
- the design of the study: case-series, cohort study, case-control study, randomized trial, etc.

Data collection (subheading)

Describe the details of how the data was collected/observed

Note
Observable variables will be credible only if qualified observers and validated instruments were used to assess them. Examples of observable variables include patient symptoms, subject responses to open interviews/ questionnaires, ultrasound/MRI/CT images, assessments of articles in a literature review etc. In such cases, build credibility in the Methods section; report "who" observed and interpreted the data. For example, "An experienced radiologist interpreted the images."

Note
When reporting on decisions/judgments that were made, use the "we" form—take responsibility for what you did.

Note
The Methods section reports historical facts and must be in past tense.

Data analysis (subheading)

Results section

5. *The core concept of the Results*
   Briefly describe the core concept of the results in a short paragraph at the beginning of the Results section. This helps readers to understand the details that follow. Note just as in the Methods section, this section reports historical facts and must be in past tense.

Then organize the details of your Results under sub-headings, for example:

Patient/animal characteristics
Data
Statistical results

Discussion section

Structure your Discussion to focus on 4 core concepts (6, 7, 8, and 9 below).

6. *The answer to your research question*
   Present this right at the top of the Discussion section—the very first sentence, a present tense statement that expresses—to the best of your knowledge—how the world works as related to your research question/hypothesis. It is a direct answer to the question/hypothesis stated in the Introduction.
7. *Support that answer?*
   a) how your factual findings, (expressed in past tense), support your answer.
   b) relating the findings of others to your answer.
   c) theoretical considerations that support your answer.

Limitations (subheading)

8. *The limitations to that answer*
   Focus explicitly on limitations related to possible confounders:
   - sample size
   - specific locations/medical centers of your study,
   - possible ethnic/cultural variables,
   - uncontrolled patient/subject characteristics and
   - underlying assumptions.

Conclusions (subheading)

The Conclusion is not a summary, but should focus on the consequences of your work. Structure this subsection using separate paragraphs that state 2 main messages (9 and 10)

9. *What are the practical/theoretical consequences of your answer?*
   The value—relevance— of your work: how it helps to solve the problem described at the beginning of the Introduction.
10. *What is a next step to help solve the original problem?*
    - a new research question to be answered
    - a refinement of the present study to reduce limitations
    - a protocol to implement the findings in the clinic

Advice to the reviewers of EJM

For the convenience of our future contributors and our readers, we publish here the advice we give to our reviewers.

In the process of reviewing a paper, please refer to the following points:

- Your first step should be to evaluate your relationship with the authors. To ensure the credibility of the process, reviewers should not have a conflict of interest with the authors. If this is a case, the paper should be appointed to other reviewers. Please keep us informed whether conflict of interest is an issue for you as an appointed reviewer.
- Is this work relevant and interesting for EJM?
- Are the objectives appropriate and clearly stated?
- Are the data valid?
- Are the conclusions valid and properly supported?
- Is the already existing work described adequately?
- Paper structure/organization; is this logical?
- Does abstract clearly convey meaning of the paper?
- Is the paper well written and can be easily understood? (Please keep in mind that students don't have the experience to reed throughout the paper very quickly and to understand everything in a research paper at the first glance)
- Are all sections really needed, or could they be shortened?
- Is the science reliable? Please, be aware of ethical issues such as plagiarism!

Comments should be detailed and specific. Mentoring the authors includes helping authors improve their paper under review even if these papers will/could not be accepted for publication in our journal. By careful reviewing, you will help improving the quality of papers published elsewhere too. Avoid vague complaints and provide appropriate citations if authors are unaware of the relevant work.

Please consider a manuscript received for reviewing as a confidential document and do not discuss the content of this paper with others. To maintain the validity of this process, you should never contact the authors about the paper under review.

The review process serves two important goals: providing guidance to the authors to improve the quality of their paper, and providing the editor or editorial board with valuable recommendations regarding the acceptance or rejection of the peer-reviewed papers (along the whole spectrum of major revision- minor revision- rejection). So it is important that you give comments to the authors, and to the editor in separate sections. Please use the provided form, because this makes life easier for you, the editor and the authors.

EJM is committed to rapid editorial decisions and publication. We request that reviewers return their comments within the time indicated at invitation. If any unanticipated difficulties arise that may prevent you from submitting the review on time, contact us by sending an email to the editorial office at ejm@erasmusmc.nl. You are welcome to contact us if you have any questions.

For more information about guidelines for the review process, please visit our website: www.erasmusmc.nl/ejm. We also recommend you to view the presentations of the EJM workshop on our website. Here you can find instructions about how to scan through a paper and grab its essence, and how to structure your comments to the authors and to the editor.

Januari 2017, Editorial board of Erasmus Journal of Medicine.


ERASMUS - JOURNAL - of MEDICINE
Erasmus MC
Universitair Medisch Centrum Rotterdam
Erasmus Journal of Medicine