EJM
Erasmus Journal of Medicine

June 2019 • 13th issue

Erasmus Journal of Medicine: independent scientific journal

Elective Bariatric Surgery in Obese Patients with Type 2 Diabetes Mellitus

Opinion

Initiatives to improve cultural competency among medical students

Systematic review

Systematic review

Clinical outcomes of post-cardiac arrest exposure to hyperoxia

Systematic review

Does mutant NPM1 MRD predict relapse in adult AML?

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 (1250 copies) and on the EJM website (www.erasmusjournalofmedicine.nl).

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, Hilversum

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

Website
www.erasmusjournalofmedicine.nl
Like us on www.facebook.com/erasmusjournalofmedicine

The important first step

Cliché as it may sound, every journey starts with a first step. Even the greatest scientists started with their first little experiment and wrote their first article. At Erasmus MC, we are aware that tomorrow's scientists are today's students, and we motivate them to take that first step early in their career. Obviously, in order to become a successful scientist, distinctive talents and personality traits, including curiosity, an open mind and a positive critical attitude are a prerequisite. But being a scientist also requires education, craftsmanship, dedication and hard work.

At Erasmus MC, annually, over 1000 systematic reviews, opinions, Minor essays and Master theses are being prepared by our students. In order to familiarize them with the process of scientific writing, they are invited to submit these products for publication in Erasmus Journal of Medicine (EJM). Their manuscript will then be reviewed by fellow students and staff reviewers. The entire editorial management process is the hands of student editors, who are tutored by staff researchers.

As the current issue of EJM demonstrates, this process has again proven successful. A broad variety of topics are addressed, ranging from a discussion of the (cost-)effectiveness of elective bariatric surgery in obese patients with type 2 diabetes mellitus to the question whether soft-drinks with sugar should be taxed. We do hope that you will enjoy reading this work as much as we and our colleagues enjoyed tutoring the students. And fellow students: do not hesitate to submit your valuable research! It might be the first step in a wonderful scientific career.

Prof. Hans van Leeuwen, dean

Prof. Eric Boersma, chair of the editorial board

Erasmus Journal of Medicine

Too much oxygen?

Paul van Daele, MD PhD
Department of Immunology, staff editorial board EJM,
Erasmus Medical Center

In the current issue of the journal, Zwaan et al describe the potential detrimental effects of over-liberal oxygen supplementation to patients after cardiac arrest. They mention that hyperoxia, defined as a PaO2 above 300 mmHg, is associated with poorer outcome, both in mortality and neurological morbidity.

Although oxygen itself is vital to sustain life, breathing oxygen at higher than normal partial pressure may lead to hyperoxia and can cause oxygen toxicity or oxygen poisoning. Both acute and chronic oxygen toxicity can occur. The acute toxicity (very high dose for relatively short time as seen for instance in divers (1)) manifests generally with central nervous system (CNS) effects, while chronic toxicity (lower dose of oxygen for longer periods for instance in an ICU-setting) has mainly pulmonary effects.

Patients at risk for oxygen toxicity are those that are on hyperbaric oxygen therapy, patients exposed to prolonged high levels of oxygen, premature infants and underwater divers. Oxygen toxicity is probably caused by oxygen-derived free radicals. Oxygen is particularly toxic for CNS, lungs and eyes (neonates).

Fortunately, both in children and adults, oxygen-induced pulmonary toxicity is often reversible. For infants, those who have survived following an incidence of bronchopulmonary dysplasia, will ultimately recover near-normal lung function, since lungs continue to grow during the first 5–7 years, but they remain vulnerable to respiratory infections.

Also oxygen-induced retinopathy is often reversible but in advanced stages the outcome is less favorable. (2)

Whether the effects of high dose oxygen on the brain are reversible is less well known. Hyperbaric oxygen can induce seizures which are classified as brief, generalized tonic-clonic seizures and they are believed to cause no residual cognitive damage, although sound data in humans are lacking. In a mouse model it was shown that hyperbaric oxygen caused a transient decrement in cognitive function. (3) Remarkably high dose oxygen may also have beneficial effects in acute severe traumatic brain injury. (4) This appears to be in contrast to the results from Zwaan et al.

So even for oxygen Paracelsus's paradigm holds true: "All things are poisons, for there is nothing without poisonous qualities. It is only the dose which makes a thing poison."

References:

1. *Wingelaar TT, van Ooij P-JAM, van Hulst RA. Oxygen Toxicity and Special Operations Forces Diving: Hidden and Dangerous. Front Psychol [Internet]. 2017 Jul 25 [cited 2019 May 2];8:1263. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28790955*
2. *Cooper JS, Shah N. Oxygen Toxicity [Internet]. StatPearls. 2019 [cited 2019 May 2]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28613494*
3. *Domachevsky L, Rachmany L, Barak Y, Rubovitch V, Abramovich A, Pick CG. Hyperbaric oxygen-induced seizures cause a transient decrement in cognitive function. Neuroscience [Internet]. 2013 Sep 5 [cited 2019 May 3];247:328–34. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23732232*
4. *Daly S, Thorpe M, Rockswold S, Hubbard M, Bergman T, Samadani U, et al. Hyperbaric Oxygen Therapy in the Treatment of Acute Severe Traumatic Brain Injury: A Systematic Review. J Neurotrauma [Internet]. 2018 Feb 15 [cited 2019 May 3];35(4):623–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29132229*

Diagnostic Accuracy Of Plasma NT-ProBNP For Diagnosing Acute Systolic Heart Failure In Adult Patients Presenting With Dyspnea At The Emergency Department: *a systematic review and Meta-Analysis*

Daniëlle Gebhard, medical student
David de Groot, medical student
Thomas Woo, medical student
Editorial Board of the Erasmus Journal of Medicine

In this issue of the Erasmus Journal of Medicine, a systematic review and meta-analysis by Kemps et al (1) provides an insight in the usage of N-terminal prohormone of brain natriuretic peptide (NT-proBNP) in diagnosing Acute Systolic Heart Failure (ASHF). With a rising incidence of acute heart failure, optimizing the diagnostic process of ASHF is of great clinical relevance. Due to its high sensitivity, as has been shown in this meta-analysis, NT-proBNP can function as a rule-out biomarker for ASHF. This means that normal NT-proBNP serum levels make ASHF unlikely as the cause of acute dyspnea.

By searching the latest available literature, Kemps et al found eight studies eligible for inclusion in their meta-analysis to calculate the sensitivity and specificity of NT-proBNP as biomarker for diagnosing ASHF. Using eight studies to calculate the sensitivity and specificity, the authors were able to make reliable estimates. However, there are several limitations to this meta-analysis. For example, the included studies had a different approach on the biochemical measurement of NT-proBNP in patient plasma samples, and not all systolic heart failure diagnoses in the included articles were confirmed by an echocardiography. Furthermore, we think that the fact that the included articles did not investigate the differences between men and women is also a limitation to this study.

The importance of the influence of gender in heart disease was recently addressed by prof. dr. Angela Maas and dr. Janneke Wittekoek in the Dutch media.(2) Some research has been done concerning this matter. It seems that female heart failure patients tend to have more comorbidities. However, females tend to have a better prognosis in terms of mortality and hospitalization risk compared to men.(3) Women tend to develop heart failure at an older age compared to men, and heart failure with preserved ejection fraction is more common in women.(4) Also, studies have shown that serum levels of BNP in healthy adults are significantly higher in women compared to men.(5, 6) According to Kim et al (7) the NT-proBNP level is a more reliable marker in the prediction of long-term mortality and heart failure readmission in men than in women. Generally, it seems that ASHF tends to be different in female patients. This raises the question whether the potential use of NP-proBNP in the diagnosis of ASHF also depends on gender.

In the current systematic review, the included studies had an equal distribution in gender. However, no subgroup analysis has been conducted for the sensitivity and specificity of the biomarker for men and women separately. Future research is required to improve the diagnosis of heart failure in both male and female patients using NT-proBNP as biomarker to minimize the number of false-positives and false-negatives. Furthermore, the difference in usability of NT-proBNP between the sexes should be a subject for future research, due to the known differences between male and female patients with ASHF.

References:

1. *Kemps N, Lodder L, Lubbers V, van der Does Y. Diagnostic Accuracy Of Plasma NT-ProBNP For Diagnosing Acute Systolic Heart Failure In Adult Patients Presenting With Dyspnea At The Emergency Department: A Systematic Review And Meta-Analysis. 2019.*
2. *Te weinig cardiologen hebben verstand van het vrouwenhart https://nos.nl/nieuwsuur/artikel/2214218-te-weinig-cardiologen-hebben-verstand-van-het-vrouwenhart.html. 2018.*
3. *Savarese G, D'Amario D. Sex differences in heart failure. Sex-Specific Analysis of Cardiovascular Function: Springer; 2018. p. 529-44.*
4. *Bozkurt B, Khalaf S. Heart failure in women. Methodist DeBakey cardiovascular journal. 2017;13(4):216.*
5. *Redfield MM, Rodeheffer RJ, Jacobsen SJ, Mahoney DW, Bailey KR, Burnett JC. Plasma brain natriuretic peptide concentration: impact of age and gender. Journal of the American College of Cardiology. 2002;40(5):976-82.*
6. *Wang TJ, Larson MG, Levy D, Leip EP, Benjamin EJ, Wilson PW, et al. Impact of age and sex on plasma natriuretic peptide levels in healthy adults. The American journal of cardiology. 2002;90(3):254-8.*
7. *Kim H-L, Kim M-A, Choi D-J, Han S, Jeon E-S, Cho M-C, et al. Gender Difference in the Prognostic Value of N-Terminal Pro-B Type Natriuretic Peptide in Patients With Heart Failure - A Report From the Korean Heart Failure Registry (KorHF) - Circulation Journal. 2017:CJ-16-1345.*

Elective Bariatric Surgery in Obese Patients with Type 2 Diabetes Mellitus

Ricardo M. Bianchi[a] and Jesse N. Robert[a]
[a]Medical student, Erasmus Medical Center Rotterdam, The Netherlands
Correspondence: Ricardo M. Bianchi, email: 424702rb@student.eur.nl
Jesse N. Robert, email: 424713jr@student.eur.nl

Abstract

Introduction: **Global prevalence of Diabetes Mellitus has increased substantially in the last few decades. More treatment options for patients with Type 2 Diabetes Mellitus have been tested to achieve glycemic control or even remission. Multiple studies also showed that surgical bariatric interventions are superior to medication management in the treatment of Type 2 Diabetes Mellitus.**
Research question: **Should bariatric surgery become the standard treatment in obese patients with Type 2 Diabetes Mellitus instead of conventional medication treatment?**
Methods: **Literature search was performed to investigate the effectiveness, adverse events, cost-efficacy and logistics of both types of treatment, as well as other treatments and ethical aspects.**
Results: Although bariatric surgery is more effective in terms of remission of Type 2 Diabetes Mellitus, the adverse events, cost-efficacy and logistics are inferior to those of conventional medication treatment. Moreover, implementing bariatric surgery as standard treatment has to overcome some ethical considerations however. Medication regimen is not fully optimized yet due to a poor adherence rate.
Conclusion: **Bariatric surgery should not become the standard treatment in patients suffering from Type 2 Diabetes Mellitus.**

Introduction

The global prevalence of Diabetes Mellitus (DM) has risen from 4.7% in 1980 to 8.5% in 2014, or from 108 to 422 million individuals respectively.[1] DM, which is diagnosed when fasting plasma glucose is equal to or above 7.0 mmol/L, is responsible for approximately a two-fold increased risk of coronary heart disease, stroke, and deaths from a cardiovascular cause such as heart failure, cardiac arrhythmia, sudden death, hypertensive disease and aortic aneurysms.[2,3]
There are two main forms of DM. Type 1 Diabetes Mellitus (T1DM) results in insulin deficiency due to auto immune-mediated destruction of pancreatic β-cell islets, while Type 2 Diabetes Mellitus (T2DM) is caused by defect(s) in insulin secretion, almost always with a major contribution from insulin resistance.[4] T2DM is the most common form and is associated with obesity and decreased physical activity.[5] About 52% of the patients with diagnosed DM are obese, which is defined as having a Body Mass Index (BMI) of at least 30 kg/m2. Moreover, 83% of those patients are overweight, which is defined as having a BMI of at least 25 kg/m2.[6] While patients with T1DM need the use of insulin to survive, treatment options for T2DM are more versatile.[4] The American Diabetes Association (ADA) has stated in their latest annual version of the Standards of Medical Care in Diabetes that a pharmacologic agent, for example Metformin, is the preferred initial method in the treatment of T2DM, but lifestyle management and weight control also play an important role in controlling T2DM.[7,8] In The Netherlands however, lifestyle intervention and weight control are considered the initial treatment methods in controlling T2DM.

In the last decade more treatment options for patients with T2DM have been tested to achieve glycemic control or even remission of T2DM, including weight control programs and additional medicines like Empagliflozin and Omarigliptin.[9-11] Multiple studies also showed that surgical bariatric interventions are superior to medication management in the treatment of T2DM.[12-15] Bariatric surgery not only helps controlling T2DM, but is also associated with an significant risk reduction in macrovascular outcome, myocardial infarction, stroke, cardiovascular events and mortality compared to non-surgical interventions.[16,17] Our goal is to determine whether bariatric surgery should become the standard treatment in obese patients with Type 2 Diabetes Mellitus, as opposed to conventional medication treatment. To answer this question, we performed a literature search on PubMed between November 26th 2018 and January 4th 2019 and selected articles that had their main focus on effectiveness, adverse events, cost-efficacy and logistics of bariatric surgery or conventional medication treatment, as well as other treatments and ethical aspects.

Effectiveness

To evaluate the possibility to accept bariatric surgery as standard treatment for patients suffering from T2DM, it is worth investigating the effectiveness of both bariatric surgery and conventional treatment, in this case lifestyle management and medication. It has been reported that complete T2DM remission, which is defined as lowering glycated hemoglobin (HbA1c) to normal limits and fasting blood glucose (FBG) levels below 100 mg/

dl (5.6 mmol/L) for at least one year without the use of anti-diabetes medication, occurred in about 70% of patients that underwent bariatric surgery.[18,19] In that same study, just over 12% of post-bariatric patients achieved partial remission, defined as HbA1c below 6.5% or 48 mmol/mol and FBG levels of 100–125 mg/dl or 5.6–6.9 mmol/L for a minimum of 1 year without pharmacological treatment.[18] However, about 12% of patients with partial or complete remission of T2DM also experienced relapse.[19] Medication regimen on the other hand, showed no remission after 2 years at all.[20] According to the bottom-up, prevalence-based European-wide CODE-2 study, which studied economic burden of Diabetes Mellitus in Europe, conventional treatment only achieves adequate glycemic control in 31% of the investigated patients suffering from T2DM and thus, is far from optimal.[21] Multiple studies though, have addressed poor adherence to medication treatment and insulin therapy in patients with T2DM.[22-30] One study found that insulin therapy unwillingness was common. Nearly 30% of the investigated individuals reported being unwilling to take insulin if prescribed, suggesting an adherence rate of only 70% in all eligible patients.[31] A recent meta-analysis reported a mean rate of poor medication adherence of 38%, where the patients with good adherence had significant fewer hospitalization events and less all-cause mortality events.[32]

Adverse events

Although bariatric surgery leads to higher remission rates of T2DM, it is a more invasive treatment option with adverse events and additional risks depending on the performed type of bariatric surgery. Either performing a Sleeve Gastrectomy (SG) or Roux en Y Gastric Bypass (RYGB), both which are types of bariatric surgery, results in at least one major complication in 2.0% of the patients, with leakage as most common complication. With the use of Laparoscopic Adjustable Gastric Banding (LAGB), another bariatric procedure, the median complication rate is 42.7% with slippage and pouch dilatation as most common complication.[34] Despite all complications, a meta-analysis from 2017 showed that bariatric surgery is associated with a low perioperative mortality rate (0.18%) and a long-term reduction in all-cause mortality of 41% in patients receiving bariatric surgery compared to non-operated obese controls.[35] Medication treatment however, also has side-effects. Sulphonylureas are known for its major adverse effect: hypoglycemia. Moreover, cases of acute pancreatitis have been reported with the use of glucagon-like peptide 1 (GLP-1) receptor agonists, while skin rash is prevalent after the use of dipeptidyl peptidase 4 (DPP-4) inhibitors. Urinary tract infection and genitourinary tract candidiasis are commonly reported side-effects of Sodium-glucose co-transporter 2 (SGLT-2) inhibitors.[36] Thiazolidinediones (TZDs) could reduce bone density, leading to an increased risk of bone fractures.[37] Therefore, it is questionable whether the risk of complications after bariatric surgery outweighs the adverse events of conventional medication treatment.

Cost efficacy

In the US, medical costs of people with diagnosed diabetes are about 2.3 times higher than in the absence of diabetes. Considered institutional care, outpatient care and outpatient medications and supplies, total costs solely attributed to the presence of DM are 9.601 US dollars per capita per year. Indirect burden of diabetes and mortality costs attributed to diabetes account for another 89.9 billion and 19.9 billion US dollars respectively. In total, those numbers have led to a 26% increase in economic costs of diabetes from 2012 to 2017 after adjusting for inflation. The main reasons for this rise are increased prevalence of diabetes and the increased cost per person with diabetes.[38] Estimates of total costs solely attributed to bariatric surgery however, seem to be between 14.000 and 15.000 US dollars per capita.[39] The incremental cost-effectiveness ratio (ICER), which is a statistic used to quantify the expenses of a healthcare intervention, varied from over 150.000 US dollars per Quality Adjusted Life Year (QALY) 36 months after bariatric surgery to around 90.000 US dollars per QALY, guaranteeing cost-effectiveness with a probability of 98% 6 years after bariatric surgery.[40] In comparison, the cost efficacy of metformin, a drug widely used in the prevention and treatment of T2DM, is calculated to be around 30.000 US dollars per QALY.[41] Terranova et al. stated in their review that other cost efficacy analyses also consistently demonstrated that additional years of lives gained through bariatric surgery may be obtained at a reasonable and affordable cost. This cost will likely differ between countries. Furthermore, they stated the use of bariatric surgery may save money in a relatively short period of time, even after a required initial economic investment.[42] Unfortunately, no data was found on the European evaluation of cost-efficacy of both bariatric surgery and conventional medication treatment.

Feasibility

A question that needs to be raised, is if elective bariatric surgery can be performed on all obese patients with T2DM without strain on the health care system. One study estimated the workload for surgeons to perform bariatric surgery on all roughly 22 million obese Americans. It would take 5500 bariatric surgeons doing 400 cases a year each for 10 years to attempt to surgically treat every obese American. Considering the country's current surgical and health care resources, these numbers cannot be achieved.[43] Instead, it could be more beneficial to address obesity and diabetes with long-term, concerted policy efforts worldwide. This includes lifestyle changes like healthy eating, regulation of food supply, public education, but also motivation and incentives in various societies. Other efforts such as healthy commuting through walking or biking first need gradual change of the current infrastructure.[44]

Other treatment strategies

Before bariatric surgery can be considered the preferred method in treatment and cure of T2DM, other treatment strategies need to be discussed. So far, treatment through medication has been mentioned. Another possible treatment strategy is lifestyle management, most notably in the form of dietary control.[45] Previous studies have shown that a low-calorie diet results in a reduction of plasma glucose levels of the patients being observed.[46-48] According to a study performed by Lim and colleagues, the mechanism behind this effect lies in the reduction of fat content in liver and pancreas, as well as in normalization of β-cell function.[46] During that study, participants were given a diet

of 600 kcal/day for eight weeks. As the participants were given this diet, fat content in the liver decreased by 9.9% and fat content in the pancreas by 1.8% after eight weeks. After the eighth week, fat content was comparable to that of participants in the control group. Moreover, hepatic insulin sensitivity increased and plasma glucose levels decreased by 3.3 mmol/L to similar levels found in the non-diabetic control group. This decrease in pancreas fat content caused more insulin secretion and thus a return in β-cell function. These effects were observed after one week of taking the low-calorie diet and showed a normalization of glucose levels and hepatic insulin sensitivity in a short period of time.[46] However, no significant increase of peripheral insulin sensitivity was found. Other studies showed a decrease in HBA1c levels after dietary control, further suggesting that a low-calorie diet can lead to a remission of T2DM and normalized plasma glucose levels.[45,48,49] These studies showed that the weight of participants also decreased as triglyceride content in the liver and pancreas decreased.[45-49]

Ethical aspects

So far, we have discussed the practical implications and considerations of implementing bariatric surgery in the treatment of T2DM. However, there are also ethical aspects related to this discussion. First of all, a trend towards more invasive interventions has been described previously, even within well-regulated surgical settings.[50] This could be explained by the fact that private for profit bariatric centers may let non-medical motives interfere with medical intentions. Also, conflicts of interest can occur as surgeons and scientists hold positions in or are paid by commercial companies providing products for bariatric surgery.[51] In conclusion, bariatric surgery seems to be at least influenced by stake holder interests.

Secondly, informed consent regarding bariatric surgery can be quite challenging. Etiology and mechanisms of T2DM, as well as outcomes of various interventions is complex and not fully understood yet, even to specialists caring for patients undergoing bariatric treatment. It might therefore be a difficult task to communicate this to patients. Moreover, it is difficult to inform patients about the risk, side-effects and expected consequences of bariatric treatment, such as long-term nutritional, medical and psychological consequences as they are, again, not fully understood by healthcare professionals.[50]

Thirdly, T2DM is frequently referred to as a "lifestyle disease". It is often considered to be self-inflicted resulting from lack of self-control and is subject to prejudice.[52] Making bariatric surgery the standard treatment method could lead to a change in perception about the disease. People may see T2DM as a disease that could simply be cured by means of surgery and may not see lifestyle as an important factor. However, T2DM is much more complex than this. Surgery alone would not cure this disease and there are many factors involved.[53] The perception of T2DM as merely a disease cured by surgery could create expectations that could not be met by health professionals.

Although surgery may make T2DM more "biological" and reduce this prejudice, it can be seen as a surgical solution to the medical implications of a social problem.[51] In this perspective one could ask whether it is ethically justified to perform surgery in a healthy abdomen.

Lastly, bariatric surgery is not always compensated by the healthcare insurance companies nowadays. This has the potential to induce inequality to the access of appropriate healthcare, since not every patient has the financial means to make up for the costs themselves.[54] Thus, changes in the current healthcare insurance system are crucial to ensure that appropriate healthcare remains accessible to every patient before bariatric surgery can be considered the standard treatment.

Authors' opinion

In this article we summarized the effectiveness of bariatric surgery versus conventional medication treatment in obese patients suffering from T2DM, as well as cost efficacy of both treatment strategies and adverse events. We described the potential of lifestyle management in the form of dietary control as another treatment strategy in the remission of T2DM and evaluated the potential of implementing bariatric surgery as standard treatment. Finally, we touched upon the most relevant ethical aspects and logistics regarding the implementation of elective bariatric surgery.

We acknowledge the potential of bariatric surgery in the treatment of T2DM in obese patients, especially because of the growing number of articles showing the effectiveness of this treatment method. On the other hand, less invasive methods like conventional medication treatment are not yet optimized due to poor adherence. We believe that an improvement of medication adherence could lead to better health-related outcomes than conventional medication shows us nowadays.

Although both treatment methods have proven to be cost effective, it has been established that the cost efficacy of bariatric surgery is lower than that of conventional medication. If surgery were to become the standard treatment method for T2DM, it would lead to increased financial costs. Instead, we believe it would be financially beneficial to invest in the improvement of medication adherence.

Besides conventional medication, another treatment strategy is lifestyle intervention in the form of dietary control. There is evidence to suggest that a strict low-calorie diet can produce similar results to those found in bariatric surgery patients.

We are aware that both conventional medication and lifestyle intervention have some challenges to overcome in terms of adherence. In the next article of this series, we will evaluate the possibility to improve adherence to these treatment strategies.

Finally, bariatric surgery still has to take into account some ethical aspects before it can be implemented as the standard treatment of T2DM. Stake holder interests need to be more transparent to prevent conflicts of interest and to guarantee the best outcomes for each individual patient. Creation of guidelines could ease the process of informed consent and consensus should be reached about the justification of performing surgery in a healthy abdomen. Other points to be raised are the potential shift in perception about the curability of T2DM and access to the surgery itself across the entire patient population.

In conclusion, conventional medication and bariatric surgery are treatment options with similar outcomes. Considering all aspects involved, we believe that bariatric surgery should not become the standard treatment in patients suffering from Type 2 Diabetes Mellitus.

References

1 *NCD Risk Factor Collaboration (NCD-RisC), Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. Lancet. 2016 ; 387: 1513-1530.*

2 *World Health Organization. Global report on diabetes. 2016.*

3 *Emerging Risk Factors Collaboration, Sarwar, N., Gao, P., Seshasai, S.R. et al., Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. Lancet. 2010 ; 375: 2215-2222.*

4 *World Health Organization. (1999). Definition, diagnosis and classification of diabetes mellitus and its complications: report of a WHO consultation. Part 1, Diagnosis and classification of diabetes mellitus. Geneva: World Health Organization. http://www.who.int/iris/handle/10665/66040. Assessed on January 9th, 2019.*

5 *Zimmet, P.Z., Diabetes epidemiology as a tool to trigger diabetes research and care. Diabetologia. 1999 ; 42: 499-518.*

6 *Cheung, B.M., Ong, K.L., Cherny, S.S. et al., Diabetes prevalence and therapeutic target achievement in the United States, 1999 to 2006. Am J Med. 2009 ; 122: 443-453.*

7 *American Diabetes Association, Pharmacologic Approaches to Glycemic Treatment: Standards of Medical Care in Diabetes-2018. Diabetes Care. 2018 ; 41: 73-85.*

8 *Lean, M.E., Leslie, W.S., Barnes, A.C. et al., Primary care-led weight management for remission of type 2 diabetes (DiRECT): an open-label, cluster-randomised trial. Lancet. 2018 ; 391: 541-551.*

9 *O'Neil, P.M., Miller-Kovach, K., Tuerk, P.W. et al., Randomized controlled trial of a nationally available weight control program tailored for adults with type 2 diabetes. Obesity. 2016 ; 24: 2269-2277.*

10 *Zhang, Y.J., Han, S.L., Sun, X.F. et al., Efficacy and safety of empagliflozin for type 2 diabetes mellitus: Meta-analysis of randomized controlled trials. Medicine. 2018 ; 97: e12843.*

11 *Wang, X., Li, X., Qie, S. et al., The efficacy and safety of once-weekly DPP-4 inhibitor omarigliptin in patients with type 2 diabetes mellitus: A systemic review and meta-analysis. Medicine. 2018 ; 97: e11946.*

12 *Ribaric, G., Buchwald, J.N., McGlennon, T.W., Diabetes and weight in comparative studies of bariatric surgery vs conventional medical therapy: a systematic review and meta-analysis. Obes Surg. 2014 ; 24: 437-455.*

13 *Khorgami, Z., Shoar, S., Saber, A.A. et al., Outcomes of Bariatric Surgery Versus Medical Management for Type 2 Diabetes Mellitus: a Meta-Analysis of Randomized Controlled Trials. Obes Surg. 2018.*

14 *Maleckas, A., Venclauskas, L., Wallenius, V. et al., Surgery in the treatment of type 2 diabetes mellitus. Scand J Surg. 2015 ; 104: 40-47.*

15 *Muller-Stich, B.P., Senft, J.D., Warschkow, R. et al., Surgical versus medical treatment of type 2 diabetes mellitus in nonseverely obese patients: a systematic review and meta-analysis. Ann Surg. 2015 ; 261: 421-429.*

16 *Fisher, D.P., Johnson, E., Haneuse, S. et al., Association Between Bariatric Surgery and Macrovascular Disease Outcomes in Patients With Type 2 Diabetes and Severe Obesity. JAMA. 2018 ; 320: 1570-1582.*

17 *Kwok, C.S., Pradhan, A., Khan, M.A. et al., Bariatric surgery and its impact on cardiovascular disease and mortality: a systematic review and meta-analysis. Int J Cardiol. 2014 ; 173: 20-28.*

18 *Buse, J.B., Caprio, S., Cefalu, W.T. et al., How do we define cure of diabetes? Diabetes Care. 2009 ; 32: 2133-2135.*

19 *de Oliveira, V.L.P., Martins, G.P., Mottin, C.C. et al., Predictors of Long-Term Remission and Relapse of Type 2 Diabetes Mellitus Following Gastric Bypass in Severely Obese Patients. Obes Surg. 2018 ; 28: 195-203.*

20 *Mingrone, G., Panunzi, S., De Gaetano, A. et al., Bariatric surgery versus conventional medical therapy for type 2 diabetes. N Engl J Med. 2012 ; 366: 1577-1585.*

21 *Liebl, A., Mata, M., Eschwege, E. et al., Evaluation of risk factors for development of complications in Type II diabetes in Europe. Diabetologia. 2002 ; 45: 23-28.*

22 *Gibson, T.B., Song, X., Alemayehu, B. et al., Cost sharing, adherence, and health outcomes in patients with diabetes. Am J Manag Care. 2010 ; 16: 589-600.*

23 *Hong, J.S., Kang, H.C., Relationship between oral antihyperglycemic medication adherence and hospitalization, mortality, and healthcare costs in adult ambulatory care patients with type 2 diabetes in South Korea. Med Care. 2011 ; 49: 378-384.*

24 *Encinosa, W.E., Bernard, D., Dor, A., Does prescription drug adherence reduce hospitalizations and costs? The case of diabetes. Adv Health Econ Health Serv Res. 2010 ; 22: 151-173.*

25 *Jha, A.K., Aubert, R.E., Yao, J. et al., Greater adherence to diabetes drugs is linked to less hospital use and could save nearly $5 billion annually. Health Aff. 2012 ; 31: 1836-1846.*

26 *Lau, D.T., Nau, D.P., Oral antihyperglycemic medication nonadherence and subsequent hospitalization among individuals with type 2 diabetes. Diabetes Care. 2004 ; 27: 2149-2153.*

27 *Ho, P.M., Rumsfeld, J.S., Masoudi, F.A. et al., Effect of medication nonadherence on hospitalization and mortality among patients with diabetes mellitus. Arch Intern Med. 2006 ; 166: 1836-1841.*

28 *Zhu, V.J., Tu, W., Rosenman, M.B. et al., Nonadherence to Oral Antihyperglycemic Agents: Subsequent Hospitalization and Mortality among Patients with Type 2 Diabetes in Clinical Practice. Stud Health Technol Inform. 2015 ; 216: 60-63.*

29 *White, T.J., Vanderplas, A., Chang, E., The costs of non-adherence to oral antihyperglycemic medication in individuals with diabetes mellitus and concomitant diabetes mellitus and cardiovascular disease in amanaged care environment. Disease Management & Health. 2004 ; 12: 181-188.*

30 *Edelman, S., Pettus, J., Challenges associated with insulin therapy in type 2 diabetes mellitus. Am J Med. 2014 ; 127: 11-16.*

31 *Polonsky, W.H., Fisher, L., Guzman, S. et al., Psychological insulin resistance in patients with type 2 diabetes: the scope of the problem. Diabetes Care. 2005 ; 28: 2543-2545.*

32 *Khunti, K., Seidu, S., Kunutsor, S. et al., Association Between Adherence to Pharmacotherapy and Outcomes in Type 2 Diabetes: A Meta-analysis. Diabetes Care. 2017 ; 40: 1588-1596.*

33 *Li, R.A., Fisher, D.P., Dutta, S. et al., Bariatric surgery results: reporting clinical characteristics and adverse outcomes from an integrated healthcare delivery system. Surg Obes Relat Dis. 2015 ; 11: 1119-1125.*

34 *Shen, X., Zhang, X., Bi, J. et al., Long-term complications requiring reoperations after laparoscopic adjustable gastric banding: a systematic review. Surg Obes Relat Dis. 2015 ; 11: 956-964.*

35 *Cardoso, L., Rodrigues, D., Gomes, L. et al., Short- and long-term mortality after bariatric surgery: A systematic review and meta-analysis. Diabetes Obes Metab. 2017 ; 19: 1223-1232.*

36 *Davoren, P., Glucose-lowering medicines for type 2 diabetes. Aust Fam Physician. 2015 ; 44: 176-179.*

37 *Riche, D.M., King, S.T., Bone loss and fracture risk associated with thiazolidinedione therapy. Pharmacotherapy. 2010 ; 30: 716-727.*

38 *American Diabetes Association, Economic Costs of Diabetes in the U.S. in 2017. Diabetes Care. 2018 ; 41: 917-928.*

39 *Doble, B., Wordsworth, S., Rogers, C.A. et al., What Are the Real Procedural Costs of Bariatric Surgery? A Systematic Literature Review of Published Cost Analyses. Obes Surg. 2017 ; 27: 2179-2192.*

40 *Klebanoff, M.J., Chhatwal, J., Nudel, J.D. et al., Cost-effectiveness of Bariatric Surgery in Adolescents With Obesity. JAMA Surg. 2017 ; 152: 136-141.*

41 *Herman, W.H., Hoerger, T.J., Brandle, M. et al., The cost-effectiveness of lifestyle modification or metformin in preventing type 2 diabetes in adults with impaired glucose tolerance. Ann Intern Med. 2005 ; 142: 323-332.*

42 *Terranova, L., Busetto, L., Vestri, A. et al., Bariatric surgery: cost-effectiveness and budget impact. Obes Surg. 2012 ; 22: 646-653.*

43 *Richards, N.G., Beekley, A.C., Tichansky, D.S., The economic costs of obesity and the impact of bariatric surgery. Surg Clin North Am. 2011 ; 91: 1173-1180.*

44 *Azizi, F., Bariatric surgery for obesity and diabetes. Arch Iran Med. 2013 ; 16: 182-186.*

45 *Doshmangir, P., Jahangiry, L., Farhangi, M.A., The effectiveness of theory- and model-based lifestyle interventions on HbA1c among patients with type 2 diabetes: a systematic review and meta-analysis. Public Health. 2018 ; 155: 133-141.*

46 *Lim, E.L., Hollingsworth, K.G., Aribisala, B.S. et al., Reversal of type 2 diabetes: normalisation of beta cell function in association with decreased pancreas and liver triacylglycerol. Diabetologia. 2011 ; 54: 2506-2514.*

47 *Steven, S., Lim, E.L., Taylor, R., Population response to information on reversibility of Type 2 diabetes. Diabet Med. 2013 ; 30: e135-138.*

48 *Sarathi, V., Kolly, A., Chaithanya, H.B. et al., High rates of diabetes reversal in newly diagnosed Asian Indian young adults with type 2 diabetes mellitus with intensive lifestyle therapy. J Nat Sci Biol Med. 2017 ; 8: 60-63.*

49 *Terranova, C.O., Brakenridge, C.L., Lawler, S.P. et al., Effectiveness of lifestyle-based weight loss interventions for adults with type 2 diabetes: a systematic review and meta-analysis. Diabetes Obes Metab. 2015 ; 17: 371-378.*

50 *Pinkney, J.H., Johnson, A.B., Gale, E.A., The big fat bariatric bandwagon. Diabetologia. 2010 ; 53: 1815-1822.*

51 *Hofmann, B., Hjelmesaeth, J., Sovik, T.T., Moral challenges with surgical treatment of type 2 diabetes. J Diabetes Complications. 2013 ; 27: 597-603.*

52 *Puhl, R.M., Heuer, C.A., The stigma of obesity: a review and update. Obesity. 2009 ; 17: 941-964.*

53 *Wu, Y., Ding, Y., Tanaka, Y. et al., Risk factors contributing to type 2 diabetes and recent advances in the treatment and prevention. Int J Med Sci. 2014 ; 11: 1185-1200.*

54 *Martin, M., Beekley, A., Kjorstad, R. et al., Socioeconomic disparities in eligibility and access to bariatric surgery: a national population-based analysis. Surg Obes Relat Dis. 2010 ; 6: 8-15.*

A tax on soft drinks: Yes or No?

Ralph Heijmans, BSc[a], Supervisor: Mandy van Hoek, MD, PhD[b]
[a] Medical student, Erasmus MC University Medical Center Rotterdam, the Netherlands
[b] Internist-diabetologist, Department of Internal Medicine, Erasmus MC University Medical Center, Rotterdam, the Netherlands
Correspondence: Ralph Heijmans, 424048, e-mail: 424048rh@student.eur.nl.

Introduction

On January the first of this year, the Dutch government has increased the reduced VAT from 6% to 9%.[1] As a result, prices of all supermarket goods, including healthy products, have been raised. The increase in price of products such as fruit and vegetables caused a great amount of discussion and led to the question whether a tax on unhealthy products alone would have been more appropriate. Soft drinks are among the most consumed unhealthy products. The popularity of soft drinks has increased enormously in recent decades.[2] Soft drinks are part of the daily routine for many of us. On average, a Dutch person drinks one glass of soft drink per day. Simultaneously, the prevalence of obesity has increased.[3] The consumption of soft drinks could be a major contributing factor to this phenomenon as they increase the chance of weight gain.[4,5] In addition, research has shown that soft drinks are associated with many other health risks.[4-14] Consequently, it is not surprising that various interventions were investigated to discourage the use of soft drinks. The introduction of a tax on soft drinks might be suitable. However, it is not an easy decision as the introduction of this tax entails both questions on the effectiveness and ethical objections. The following question therefore arises: Should a tax on soft drinks be introduced in the Netherlands?

Effects of sugar sweetened beverages on health

In order to decide whether a tax on soft drinks should be implemented, it is important to gather insight into the harmful health effects of both sugar sweetened beverages (SSBs) and artificially sweetened (diet) beverages (ASBs).

First, the consumption of SSBs has a harmful effect on blood glucose and insulin levels.[14] Studies have shown that the addition of excess energy to a regular diet by SSBs results in a mean difference of 0.12 mmol/L and 4.70 pmol/L in fasting blood glucose and insulin, respectively.[14]

The use of SSBs is also associated with several disorders. Type 2 diabetes mellitus (T2DM) is one of them.[6-8] Research has shown that the incidence of T2DM increases with 13% per unit of consumed SSB per day.[6] This percentage was found after adjustment for body mass index and is therefore independent of excess body weight. The risk of being overweight is also increased. Studies show a positive association between the use of SSBs and increased body mass index in both adults and children.[4] In line, the consumption of SSBs was associated with a 18% increase of obesity risk.[5] Studies also indicate that there is a higher chance of acquiring hypertension.[9] In addition, there appears to be an effect on cardiovascular disease, although this may be due to indirect effects through the aforementioned diseases.[10,11] Lastly, the consumption of SSBs may affect kidney and liver. It is associated with increased risk of chronic kidney disease by 58% and nonalcoholic fatty liver disease by 55%.[12,13]

Effects of artificially sweetened beverages on health

Artificially sweetened beverages (ASBs) are considered to be healthier alternatives to SSBs. Studies have shown that artificial sweeteners, in contrast to SSBs, indeed do not affect blood glucose levels after consumption.[15] However, another study suggests that the energy avoided from substituting SSBs with ASBs is compensated for at subsequent meals, resulting in a similar total energy intake.[16] Furthermore, associations between ASBs and several disorders have been found.[5,6,9,10] The consumption of SSBs is associated with a 59% increase of obesity risk and a 8% increase in risk of T2DM.[5,6] ASBs are also presumed to increase the risk of stroke and hypertension, in comparable amounts as SSBs.[9,10] Contrarily, no associations were found for chronic kidney disease and nonalcoholic fatty liver disease.[12,13]

It is worth noting that the amount of research regarding ASBs is limited and results are conflicting, possibly due to reverse causality, where high risk consumers tend to switch to ASBs to prevent the predisposed disease (such as obesity). Therefore, more research is needed to further clarify the long-term effects of ASBs.

Soft drink popularity

When deciding on the implementation of a soft drink tax, the scale upon which soft drinks are being consumed, should be considered.

Soft drinks are very popular in Europe and America. In the last thirty years, both the amount of soft drinks consumed as well as the size of one consumption have increased enormously.[2] The energy intake derived from SSBs even increased by 135%.[2] Simultaneously, the consumption of dairy products has decreased.[2] The average American drinks 600 ml of sugar sweetened beverages per day.[2] In the Netherlands, an average person drinks 86.4 liters of soft drinks per year, which amounts to 240 ml per day.[17] Fructose is a sugar that is widely used as a sweetener in soft drinks. It is therefore no surprise that SSBs are the main source of fructose, which accounts for 9% of the total daily energy intake in the Netherlands.[18]

Effectiveness and efficiency of a soft drink tax

The effectiveness and efficiency of the intervention should also be taken into account in the decision-making process of soft drink tax implementation. In other words: to what extent does the implementation of this tax succeed in reducing soft drink consumption and preventing health hazards.

First of all, there is sufficient evidence that reducing the use of SSBs will lead to a reduction in the risk of obesity and obesity-related diseases such as type 2 diabetes.[19] Research has also shown that the prevention of weight gain in the long term, for example by reducing the consumption of SSBs, is more important in prevention of obesity than weight reduction in the short term.[19] This proves that reducing the use of SSBs will have a positive impact on public health.

The question remains whether a tax will actually lead to a reduction of SSBs consumption. Studies regarding the price elasticity of soft drinks show that a 10% tax can lead to a decrease of 8% to 10% in sale.[20] If the tax is only applicable to SSBs, this decrease could be even greater as some people will switch to ASBs instead.[21] It is expected that a tax of 1 (dollar) cents per 30 ml will give a minimum reduction of 20 kcal per person per day.[21] This is sufficient to achieve weight loss and to reduce the risk of obesity.[21] It is also expected that this effect will be even greater for people whose consumption of soft drinks is high. Their behavior will be more sensitive to a change in price.[21] Another study predicts that the introduction of a 20% tax will lead to a 1.3% decrease of obesity prevalence.[22] Although these results were based on predictive models, the effects of introducing a soft drink tax have also been investigated in a field study. In a Californian city, a tax of 1 (dollar) cent per 30 ml of sugared soft drink was introduced. This led to a 21% decrease in SSBs consumption while consumption in comparable cities rose by 4%.[23]

Furthermore, the effects of tax implementation in the past should also be considered. For example, tobacco taxation has proven to be an efficient way to reduce smoking and improve public health.[24]

Ethical dilemma

To decide whether a tax on soft drinks should be implemented does raise an ethical dilemma. On the one hand, the measure has a positive influence on public health, while on the other hand, the measure opposes the respect for autonomy and it could possibly be considered as an act of paternalism. The question therefore remains: Should a tax on soft drinks be introduced in the Netherlands?

Beneficence

In health care, beneficence is an important principle. This principle entails that, as a medical practitioner, one must always try to promote people's health.[25] Research has shown that reducing consumption of SSBs will lead to a decrease in the risk of obesity and obesity-related diseases and a tax on SSBs could achieve the intended decrease in consumption.[19-23] By promoting public health, a tax on SSBs complies with the principle of beneficence and should therefore be implemented based on this principle.

Non-maleficence

Complementary with beneficence, is the principle of non-maleficence. This principle states that the actions of a health care provider should do no harm. However, it also entails that no harm should be done to an individual through neglect.[25] During recent years, healthy foods and beverages have been consistently more expensive than less healthy ones, with the Dutch reduced VAT rate widening this gap even further.[1,26] As a result, an overall unhealthy lifestyle is promoted, which subsequently increases the total burden of disease. In addition, it may also exacerbate social inequalities in health.[26] Neglecting the discouraging environment significantly harms population health and is therefore a violation of the principle of non-maleficence. A tax on SSBs could help equalize prices of healthy and less healthy consumer goods, changing the negative incentive which is currently present.

Respect for autonomy

Arguments, however, can also be made against the implementation of the tax. In health care, respect for the opinions, choices and lifestyle of a person is an important value.[25] This implies that everyone should be able to make uncoerced choices based on their own concepts and views. Important to note here is, that while everybody is responsible for their own choices, a person has to be competent enough to make these choices.[25] This principle is called respect for individual autonomy. The implementation of a soft drink tax opposes this principle. By raising prices, the government indirectly influences the choices of customers. After all, due to the implementation of the tax, a person can buy less soft drinks with the same salary and the demand for soft drinks will decrease.[20] Consumption of soft drinks is therefore reduced by a decision by the government, not by a decision of the individual. Free choice for unhealthy products is thus limited and part of the responsibility is shifted towards the government.

One could consider the implementation an act of paternalism. This implies that the government does what it deems best for all people, without taking into account their individual preferences.[25] In this case, a tax would be introduced to promote health, without considering whether people prefer to live a healthy life and what they are willing to sacrifice to achieve this.

Weak paternalism

Sometimes, however, paternalism is justified. Care providers may ignore the will of an individual if there is no deliberate choice.[25] This is called weak paternalism and this principle is applicable to this situation. Firstly, in this case, the deliberate choice to consume soft drinks is strongly influenced by social circumstances. Therefore, one might question whether it is a free choice at all.[27] As mentioned previously, studies have shown that, on average, unhealthy products are cheaper than healthy products. [26] Thus, a tax on soft drinks does not hinder the own choice of an individual, but simplifies the choice for healthy products by a positive influence on the social conditions.[27]

Secondly, making a well-considered healthy choice is a challenge in itself.[28] It requires sufficient knowledge about which products are healthy, based upon scientific research. Although there is a general idea about health and healthy products, few will be aware, for example, that a glass of soft drink per day increases the risk of diabetes by 13%.[6] In addition, when making healthy decisions, a conflict is present between short term and long term interests.[28] In the short term, unhealthy products provide benefits as they are considered to be tasty and are inexpensive.[26] The negative consequences, however, become

apparent in the long term, which further complicates making a healthy decision.

Thirdly, consumption of soft drinks by children is a major problem.[2] It is the government's job to protect them from health risks because they cannot make an informed choice themselves and they are largely influenced by their parents' lifestyle.

Conclusion

Based on the considerations above, a tax on sugar sweetened beverages should be implemented in the Netherlands. It is clear that the widespread use of SSBs has negative effects on public health. It has been shown that a tax on SSBs can achieve the intended reduction of soft drink consumption and can thus positively influence public health. Thereby, it complies with the principle of beneficence. In addition, neglecting the price inequality between healthy and less healthy products significantly harms population health and is therefore a violation of the principle of non-maleficence. Although the implementation opposes the respect for autonomy, such a form of paternalism is justified in this circumstance.

A tax on artificially sweetened beverages, however, is not justified as of yet, as future studies are required to further elucidate the potential harmful effects of ASBs consumption.

Nevertheless, a remark should be made regarding our conclusion. SSBs are often only a small part of an overall unhealthy lifestyle. Therefore, it is even more important to focus on reducing the consumption of other unhealthy products and breaking unhealthy habits. This could be achieved by lowering taxes on healthy products, improving education about healthy habits and products, increasing the availability of healthy choices over unhealthy ones and reducing the amount of hidden unhealthy contents of frequently used products.

References

1 *Rijksoverheid Belastingplan 2019. 2018.*

2 *Nielsen S.J., Popkin B.M. Changes in beverage intake between 1977 and 2001. Am J Prev Med. 2004; 27: 205-210.*

3 *Collaboration N.C.D.R.F. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128.9 million children, adolescents, and adults. Lancet. 2017; 390: 2627-2642.*

4 *Malik V.S., Pan A., Willett W.C., et al. Sugar-sweetened beverages and weight gain in children and adults: a systematic review and meta-analysis. Am J Clin Nutr. 2013; 98: 1084-1102.*

5 *Ruanpeng D., Thongprayoon C., Cheungpasitporn W., et al. Sugar and artificially sweetened beverages linked to obesity: a systematic review and meta-analysis. QJM. 2017; 110: 513-520.*

6 *Imamura F., O'Connor L., Ye Z., et al. Consumption of sugar sweetened beverages, artificially sweetened beverages, and fruit juice and incidence of type 2 diabetes: systematic review, meta-analysis, and estimation of population attributable fraction. BMJ. 2015; 351: h3576.*

7 *Greenwood D.C., Threapleton D.E., Evans C.E., et al. Association between sugar-sweetened and artificially sweetened soft drinks and type 2 diabetes: systematic review and dose-response meta-analysis of prospective studies. Br J Nutr. 2014; 112: 725-734.*

8 *Malik V.S., Popkin B.M., Bray G.A., et al. Sugar-sweetened beverages and risk of metabolic syndrome and type 2 diabetes: a meta-analysis. Diabetes Care. 2010; 33: 2477-2483.*

9 *Kim Y., Je Y. Prospective association of sugar-sweetened and artificially sweetened beverage intake with risk of hypertension. Arch Cardiovasc Dis. 2016; 109: 242-253.*

10 *Narain A., Kwok C.S., Mamas M.A. Soft drinks and sweetened beverages and the risk of cardiovascular disease and mortality: a systematic review and meta-analysis. Int J Clin Pract. 2016; 70: 791-805.*

11 *Keller A., Heitmann B.L., Olsen N. Sugar-sweetened beverages, vascular risk factors and events: a systematic literature review. Public Health Nutr. 2015; 18: 1145-1154.*

12 *Cheungpasitporn W., Thongprayoon C., O'Corragain O.A., et al. Associations of sugar-sweetened and artificially sweetened soda with chronic kidney disease: a systematic review and meta-analysis. Nephrology (Carlton). 2014; 19: 791-797.*

13 *Wijarnpreecha K., Thongprayoon C., Edmonds P.J., et al. Associations of sugar- and artificially sweetened soda with nonalcoholic fatty liver disease: a systematic review and meta-analysis. QJM. 2016; 109: 461-466.*

14 *Choo V.L., Viguiliouk E., Blanco Mejia S., et al. Food sources of fructose-containing sugars and glycaemic control: systematic review and meta-analysis of controlled intervention studies. BMJ. 2018; 363: k4644.*

15 *Nichol A.D., Holle M.J., An R. Glycemic impact of non-nutritive sweeteners: a systematic review and meta-analysis of randomized controlled trials. Eur J Clin Nutr. 2018; 72: 796-804.*

16 *Tey S.L., Salleh N.B., Henry J., et al. Effects of aspartame-, monk fruit-, stevia- and sucrose-sweetened beverages on postprandial glucose, insulin and energy intake. Int J Obes (Lond). 2017; 41: 450-457.*

17 *the Dutch Associaton for Soft Drinks W.a.J. Basic statistical information. 2016.*

18 *Sluik D., Engelen A.I., Feskens E.J. Fructose consumption in the Netherlands: the Dutch National Food Consumption Survey 2007-2010. Eur J Clin Nutr. 2015; 69: 475-481.*

19 *Hu F.B. Resolved: there is sufficient scientific evidence that decreasing sugar-sweetened beverage consumption will reduce the prevalence of obesity and obesity-related diseases. Obes Rev. 2013; 14: 606-619.*

20 *Andreyeva T., Long M.W., Brownell K.D. The impact of food prices on consumption: a systematic review of research on the price elasticity of demand for food. Am J Public Health. 2010; 100: 216-222.*

21 *Brownell K.D., Farley T., Willett W.C., et al. The public health and economic benefits of taxing sugar-sweetened beverages. N Engl J Med. 2009; 361: 1599-1605.*

22 *Briggs A.D., Mytton O.T., Kehlbacher A., et al. Overall and income specific effect on prevalence of overweight and obesity of 20% sugar sweetened drink tax in UK: econometric and comparative risk assessment modelling study. BMJ. 2013; 347: f6189.*

23 *Falbe J., Thompson H.R., Becker C.M., et al. Impact of the Berkeley Excise Tax on Sugar-Sweetened Beverage Consumption. Am J Public Health. 2016; 106: 1865-1871.*

24 *Chaloupka F.J., Straif K., Leon M.E., et al. Effectiveness of tax and price policies in tobacco control. Tob Control. 2011; 20: 235-238.*

25 *de Beaufort I.H., M. Vandamme, S. van de Vathorst, S. Schermer, M. De Kwestie. Amsterdam: Boom Uitgevers Amsterdam, 2011.*

26. *Jones N.R., Conklin A.I., Suhrcke M., et al. The growing price gap between more and less healthy foods: analysis of a novel longitudinal UK dataset. PLoS One. 2014; 9: e109343.*

27 *Borden D.P. Innovative policies under Bloomberg's 'New' Public Health. Hastings Cent Rep. 2014; 44: 6-7.*

28 *Jacobson P.D., Parmet W.E. Defending public health regulations: the message is the medium. Hastings Cent Rep. 2014; 44: 4-6.*

Diagnostic Accuracy of Plasma NT-ProBNP for Diagnosing Acute Systolic Heart Failure in Adult Patients Presenting with Dyspnea at the Emergency Department:
a systematic review and Meta-Analysis

Naomi Kemps[a], Lucas Lodder[a], Veerle Lubbers[a], Yuri van der Does[b]
[a] Medical student, Erasmus Medical Center, Rotterdam, The Netherlands
[b] MD, PhD, Supervisor, Department of Emergency Medicine, Erasmus Medical Center, Rotterdam, The Netherlands
Correspondence: Lucas Lodder, l.lodder.1@erasmusmc.nl

Abstract
Objective: **To meta-analyze the sensitivity and specificity of NT-proBNP for ruling in and ruling out acute systolic heart failure (ASHF) at the emergency department (ED). Methods: The databases of Medline and Embase were systematically screened for English records published until September 31st, 2018 using a predefined protocol. Usage of sensitivities and specificities, dyspnea as the main complaint at the ED, judication for ASHF and usage of standard diagnostic cut-off values for ruling in and ruling out ASHF were the main inclusion criteria. Data extraction was performed using a predefined protocol. Quality was assessed using an eight point scoring tool.**
Results: **8 studies were included in the meta-analysis. The pooled sensitivity and specificity of the rule-in cut-off value are 89.4% (95%-CI: 86.8%-91.5%) and 71.7% (95%-CI: 45.3%-88.6%), respectively. The pooled sensitivity and specificity of the rule-out cut-off value are 97.7% (95%-CI: 95.5%-98.9%) and 55.9% (95%-CI: 36.8%-73.4%), respectively. Quality adjustment had no influence on the results of the meta-analysis.**
Conclusions: **The sensitivity and specificity of the rule-in cut-off value are 89.4% and 71.7%, respectively. Regarding the ruling out cut-off value, the analysis showed a sensitivity and specificity of 97.7% and 55.9%, respectively. These findings provide insight into the diagnostic usage of NT-proBNP in diagnosing ASHF, but progress still has to be made to improve its accuracy.**

Introduction
Dyspnea is one of the top ten main reasons for adults to present to an emergency department.[1] Dyspnea can be a sign of both life threatening and non-life threatening pathologies, making the diagnostic approach of undifferentiated dyspnea clinically challenging. Acute heart failure is one of the possible conditions associated with dyspnea in emergency care.[2] With a rising incidence and an increase in medical costs, acute heart failure is becoming more and more of a health problem worldwide.[3] Therefore, accurate and rapid diagnosis of acute heart failure upon presentation in the emergency department is important.
Aminoterminal B type natriuretic peptides (NT-proBNP) are the prohormones of BNP that increase in plasma concentration when cardiomyocytes are strained.[4] Previously conducted studies have found a three to ten times higher plasma level of natriuretic peptides in patients experiencing acute heart failure compared to a healthy control group.[5-7] Moreover, the amount of circulating natriuretic peptides is twice as high in acute heart failure as opposed to chronic heart failure.[8] The gold standard test of diagnosing acute systolic heart failure (ASHF) is by use of echocardiography, but this form of imaging at the emergency department cannot always be rapidly used, for example, due to the need of specialized handling.[9] NT-proBNP could pose as a possibly accurate and easier diagnostic tool for the diagnosis of ASHF in the emergency setting.
The 2012 European Society of Cardiology guidelines suggest cut-off values for ruling out acute heart failure of 300 pg/ml or less for NT-proBNP.[2] Often used threshold values for ruling in acute heart failure are 450 pg/ml, 900 pg/ml and 1800 pg/ml for patients younger than 50 years, between 50 and 75 years and older than 75 years, respectively.[5] Roberts et al. previously reported on the diagnostic accuracy of natriuretic peptides in the acute care setting, based on a meta-analysis of records published until January, 2014.[10] In this systematic review and

meta-analysis, we aim to evaluate the diagnostic accuracy of plasma NT-proBNP measurement for diagnosing ASHF in adult dyspnea patients presenting to the emergency department (ED), expressed as sensitivities and specificities. With this, we also want to include eligible studies published after publication of the article by Roberts et al., as to capture all available studies on the diagnostic value of NT-proBNP.

Methods

This article was written in accordance with the 2009 PRISMA guidelines and checklist for performing systematic reviews and meta-analyses.[11]

Inclusion and Exclusion Criteria

Previously published studies that evaluated the diagnostic value of NT-proBNP to either rule in or rule out ASHF in the ED were considered eligible if 1) the diagnostic value of the plasma NT-proBNP measurement was expressed as sensitivity and specificity analyses or the article supplied data that could be used to calculate sensitivity and specificity manually; 2) adult patients (≥ 18 years of age) enrolled in the study presented at the ED with acute dyspnea as their main complaint and were diagnosed before discharge or hospital admission; 3) data on the total number of adult dyspnea patients, the total number of ASHF patients, sensitivity and specificity analyses, the way NT-proBNP was assessed, threshold values used for ruling in and / or ruling out ASHF and the method of judicating ASHF diagnoses were available and 4) ASHF in patients with elevated NT-proBNP was judicated by use of echocardiography (left ventricular ejection fraction < 50%) by either the emergency physician or a cardiologist or by expert evaluation of medical records. Studies were included for meta-analyzing if the NT-proBNP threshold for ruling in ASHF was 450 pg/ml, 900 pg/ml and 1800 pg/ml for patients younger than 50 years, between 50 and 75 years and older than 75 years, respectively, or if the threshold for ruling out ASHF was 300 pg/ml. Articles written in any other language than English, review articles, duplicate articles and articles without full-text access were excluded from the meta-analysis.

Data Extraction

Data extracted from eligible studies included the study size, country, mean age of participants, sex of participants, the mean presentation at the ED, the study outcomes, the used cut-off values for ruling in and ruling out acute systolic heart failure, the assay used to measure plasma NT-proBNP, the study's definition of ASHF, the means of judication of ASHF diagnoses and sensitivity and specificity analyses. In addition, information on the amount of true positive, false positive, true negative and false negative NT-proBNP test results was used to manually calculate the sensitivity and specificity if the study had no specific analysis for these variables.

Quality Assessment

We assessed the quality of included articles with an eight point scoring tool designed for this study. Cardiac history of included patients, the amount of missing data, the definition and presence of inclusion and exclusion criteria and the way of judicating systolic heart failure diagnoses were taken into account. Articles that scored 5 or more points were considered to have a high quality. The quality scoring tool is shown in table 1.

Search Strategy

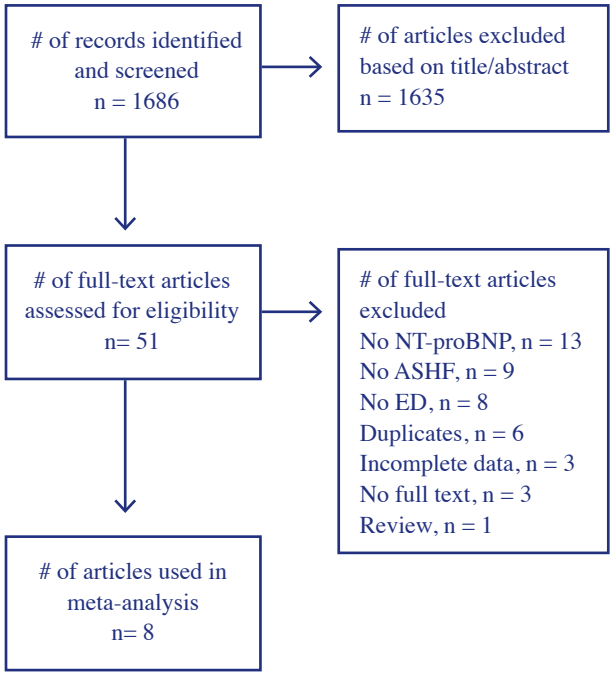
We conducted a search of the electronic databases of Medline and Embase by use of PubMed and the Cochrane Library respectively. For PubMed, the following search query was used: ("pro-brain natriuretic peptide (1-76)" [Supplementary Concept] OR "pro-brain natriuretic peptide" [TIAB] OR "amino-terminal pro-brain natriuretic peptide" [TIAB] OR "aminoterminal pro-brain natriuretic peptide" [TIAB] OR "proBNP" [TIAB] OR "pro-BNP" [TIAB] OR "NTproBNP" [TIAB] OR "NT-proBNP" [TIAB] OR "NT-pro-BNP" [TIAB]) AND (("Heart Failure, Systolic/analysis" [Mesh] OR "Heart Failure, Systolic/blood" [Mesh] OR "Heart Failure, Systolic/diagnosis" [Mesh] OR "Heart Failure, Systolic/enzymology" [Mesh] OR "Heart Failure, Systolic/metabolism" [Mesh] OR "Heart Failure, Systolic/statistics and numerical data" [Mesh]) OR (("Heart failure" [TIAB] OR "Cardiac failure" [TIAB] OR "Heart decompensation" [TIAB] OR "Myocardial failure" [TIAB]) AND ("Systolic" [TIAB] OR "Acute" [TIAB]))).

For the Cochrane Library, we used the following search query: ("pro-brain natriuretic peptide" :ti,ab OR "amino-terminal pro-brain natriuretic peptide" :ti,ab OR "aminoterminal pro-brain natriuretic peptide" :ti,ab OR "proBNP" :ti,ab OR "pro-BNP" :ti,ab OR "NTproBNP" :ti,ab OR "NT-proBNP" :ti,ab OR "NT-pro-BNP" :ti,ab) AND (("heart failure" :ti,ab OR "cardiac failure" :ti,ab OR "heart decompensation" :ti,ab OR "myocardial failure" :ti,ab) AND ("systolic" :ti,ab OR "acute" :ti,ab)). The end of the search date was set to September 31st, 2018.

Table 1 - Scoring tool for the quality assessment of eligible studies. ASHF = acute systolic heart failure, ED = emergency department.

Criterium		Available points
Availability of cardiac history	Available for each individual patient	2
	Available for patients as a group	1
	Absent or incomplete	0
Inclusion and exclusion criteria are defined in study		2
Amount of missing data in study	All data accounted for	2
	One or two missing variables	1
	Three or more missing variables	0
Method of ASHF diagnosis	Gold standard test (echocardiography)	2
	Expert evaluation	1
	Based on ED presentation	0
Maximal score		**8**

Figure 1- Flow chart of the selection of eligible articles.



Statistical Analysis

Diagnostic sensitivity and specificity analyses were calculated using OpenMetaAnalyst, the 64 bit Windows 8 version, updated October 2012.[12] We performed analyses separately for ruling in and ruling out cut-off values of NT-proBNP. Forest plots were constructed on the basis of a random effects model according to DerSimonian-Laird.[13] Heterogeneity among eligible studies was assessed with I2 statistics. Heterogeneity between studies was considered statistically significant when $P < 0,05$. The study flow chart was created with the 5th edition of Cochrane's 2014 Review Manager.[14] We adjusted for the study qualities by comparing the pooled results to the results of high quality studies alone.

Results

Study Selection

Figure 1 illustrates the selection process of eligible articles. As shown in the flow chart, our database search with the previously described query resulted in a total of 1686 articles. After screening of all abstracts, 51 articles remained for full text analysis. Of these articles, 13 were excluded because they did not assess the diagnostic value of NT-proBNP, 9 articles were excluded for the reason that they made use of other heart failure types than ASHF in their assessment, 8 articles were excluded because they were not (entirely) conducted at the ED and 3 were excluded for lack of a threshold specification. In addition, 6 duplicate studies, 3 articles without full text access and 1 review were removed. We assessed 51 full-text articles on relevancy for our research question. Application of the predefined inclusion and exclusion criteria lead to 8 remaining articles for our meta-analysis.[15-22]

Article Description

Table 2 shows the characteristics of the included articles. Studies were mostly conducted in Europe or other western countries. The study size ranged from 96 to 1461 enrolled patients. In all studies, there was a near equal distribution of the sexes. As defined in the inclusion criteria, the main presentation at the emergency department was acute dyspnea (acute new onset dyspnea or an acute exacerbation of chronic dyspnea), sometimes accompanied by signs of peripheral edema. All studies measured plasma NT-proBNP levels, but with different types of bioche-

Table 2 - Characteristics and extracted data of eligible studies. Missing data is indicated with [M]. SD = standard deviation, IQR = interquartile range, ASHF = acute systolic heart failure, LVEF = left ventricular ejection fraction, LV = left ventricle.
*** No further specification of the used assay for measuring plasma NT-proBNP.**

Author	Year	Country	Study size	Mean age (SD) or median age (IQR)	Male (%)	Cardiac History (n)	Mean presentation	Sensitivity and specificity	Cut-off	Assay	ASHF definition	ASHF diagnosis
Stoica et al.	2018	Romania	127	75,96 (11,11)	[M]	[M]	Dyspnea	84,8%, 75,5%	Rule-in	Hospital Standard*	LVEF < 40%	Transthoracic echocardiography
Januzzi et al.	2018	North America	1461	56,4 (15,7)	51,9	[M]	Dyspnea	91,5%, 38,6%	Rule-in	Cobase 601	LVEF	Transthoracic echocardiography
Darche et al.	2017	Germany	312	[M]	59	[M]	Dyspnea	89,5%, 76,1%	Rule-in	Stratus CS Acute Care	LV function	Expert evaluation
Sartini et al.	2017	Italy	236	79,98 (12,13)	49,6	74	Dyspnea	97,4%, 27,7%	Rule-out	Hospital Standard*	[M]	Expert evaluation
Martinez et al.	2016	Spain	96	[M]	46	[M]	Dyspnea	86,4%, 76,7%	Rule-in	Hospital Standard*	LVEF	Transthoracic echocardiography
Hoffmann et al.	2015	Germany	401	67 (18-96)	51	[M]	Dyspnea / peripheral edema	96,1%, 48,2%	Rule-out	Hospital Standard*	[M]	Expert evaluation
Behnes et al.	2014	Germany	212	67 (65-69)	50	[M]	Dyspnea / peripheral edema	97,4%, 83,1%	Rule-out	Hospital Standard*	LVEF	Transthoracic echocardiography
Januzzi et al. I	2006	Multiple (Western)	655	68,4 (15,9)	51	[M]	Dyspnea	90,1%, 84,3%	Rule-in	Elecsys Roche	LVEF < 50%	Transthoracic echocardiography
Januzzi et al. II								99,2%, 60,1%	Rule-out			

Table 3 - Quality assessment of the eligible studies based on the eight point quality scoring tool.

Study (year)	Criterium	Score
Stoica et al. (2018)	A	1
	B	2
	C	0
	D	2
Total Score		5
Januzzi et al. (2018)	A	1
	B	2
	C	1
	D	2
Total Score		6
Darche et al. (2017)	A	1
	B	2
	C	1
	D	2
Total Score		6
Sartini et al. (2017)	A	0
	B	2
	C	1
	D	1
Total Score		4
Martinez et al. (2016)	A	1
	B	2
	C	1
	D	2
Total Score		6
Hoffmann et al. (2015)	A	0
	B	2
	C	0
	D	1
Total Score		3
Behnes et al. (2014)	A	1
	B	2
	C	1
	D	2
Total Score		6
Januzzi et al. (2006)	A	1
	B	2
	C	1
	D	2
Total Score		6

A. Availability of cardiac history, B. Inclusion and exclusion criteria, C. Missing data, D. Method of ASHF diagnosis

mical assays. 3 studies used the standard rule-out cut-off value (< 300 pg/ml), 4 studies used the standard rule-in cut-off value (> 450 pg/ml, > 900 pg/ml and > 1800 pg/ml) and 1 study used both cut-offs. Most studies defined systolic heart failure as an impaired left ventricular heart function (LVEF < 50%). ASHF was diagnosed by either expert evaluation or transthoracic echocardiography. Table 2 shows the endpoints of the studies we used for our meta-analysis. The sensitivity ranged from 84.8% to 99.2% and the specificity ranged from 27.7% to 84.3%. These percentages were either available in the articles or calculated manually.

Quality Assessment

The result of the quality scoring tool is shown in table 3. Of the eight included studies, six had a high quality scoring, whereas two studies had a low quality scoring. Studies that were deemed of 'low quality' scored particularly low on providing a cardiac history for its included patients or had a relatively large amount of missing data. Exclusion of the studies with low qualities in a separate meta-analysis did not lead to a significant result in both the sensitivity and the specificity analysis. Therefore, we did not exclude studies with a low quality, resulting in a final analysis of eight studies in total.

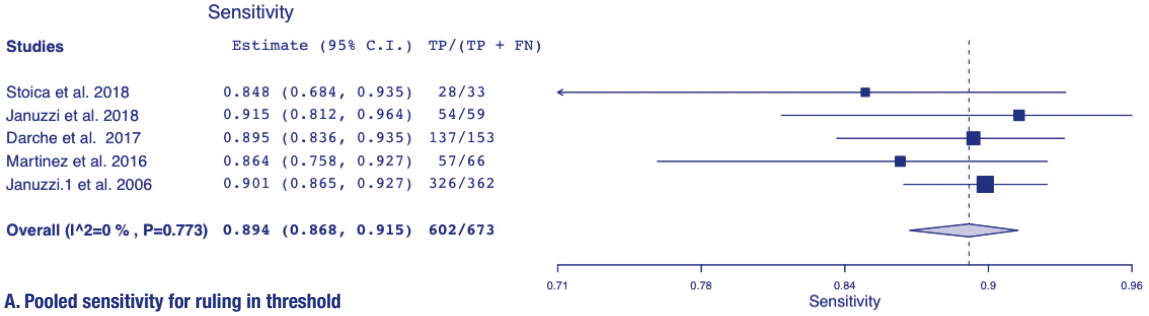
Meta-Analysis

Figure 2a depicts the sensitivity of the studies that used the standard rule-in cut-off value. The pooled sensitivity of these studies is 89.4% (95%-CI: 86.8%-91.5%). Figure 2b shows the specificity of the studies that used the standard rule-in cut-off value. The pooled specificity of these studies is 71.7% (95%-CI: 45.3%-88.6%). Figure 2c illustrates the sensitivity of the studies that used the standard rule-out cut-off value. The pooled sensitivity of these studies is 97.7% (95%-CI: 95.5%-98.9%). The last meta-analysis in figure 2d shows the specificity of the studies that used the standard rule-out cut-off value. The pooled specificity in this analysis is 55.9% (95%-CI: 36.8%-73.4%). Heterogeneity analyses showed significant results for both specificity calculations (P < 0.001), but not for the sensitivity analyses.

Conclusion and Discussion

In our study, we meta-analyzed the diagnostic accuracy of NT-proBNP in diagnosing acute systolic heart failure in adult dyspnea patients presenting to the emergency department. The meta-analysis was performed separately for ruling in and ruling out criteria. For the ruling in criteria, we calculated a pooled sensitivity and specificity of 89.4% and 71.7%, respectively. Regarding the ruling out criteria, the pooled sensitivity and specificity

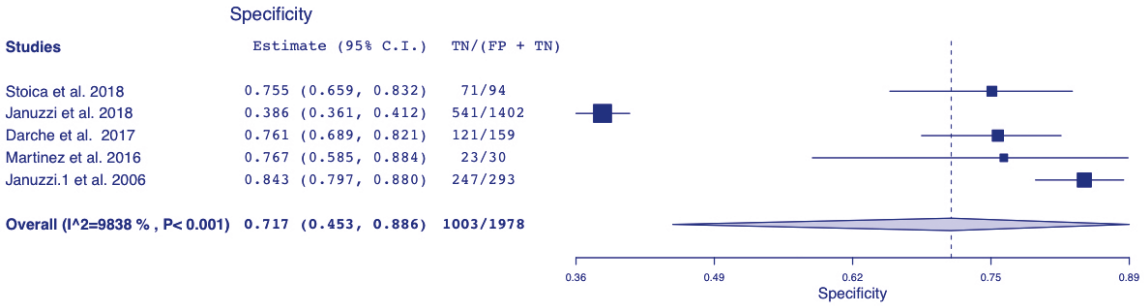
Figure 2a - Meta-analysis of included articles calculating the pooled sensitivity and specificity of both ruling in and ruling out cut-off values of NT-proBNP in diagnosing ASHF at the ED.



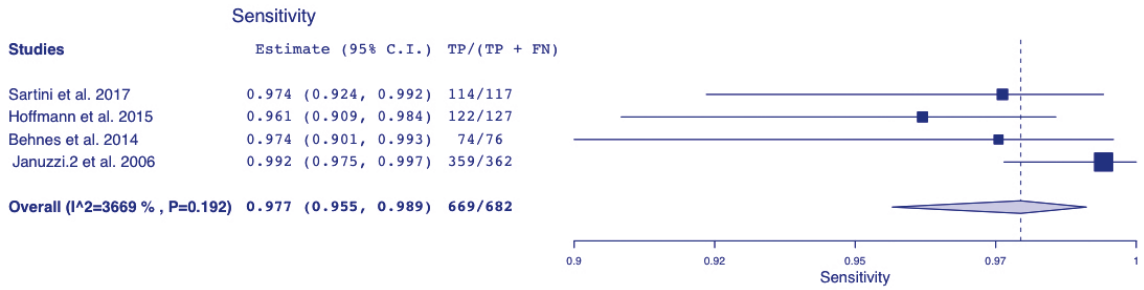
A. Pooled sensitivity for ruling in threshold

Figure 2b,c,d - Meta-analysis of included articles calculating the pooled sensitivity and specificity of both ruling in and ruling out cut-off values of NT-proBNP in diagnosing ASHF at the ED.

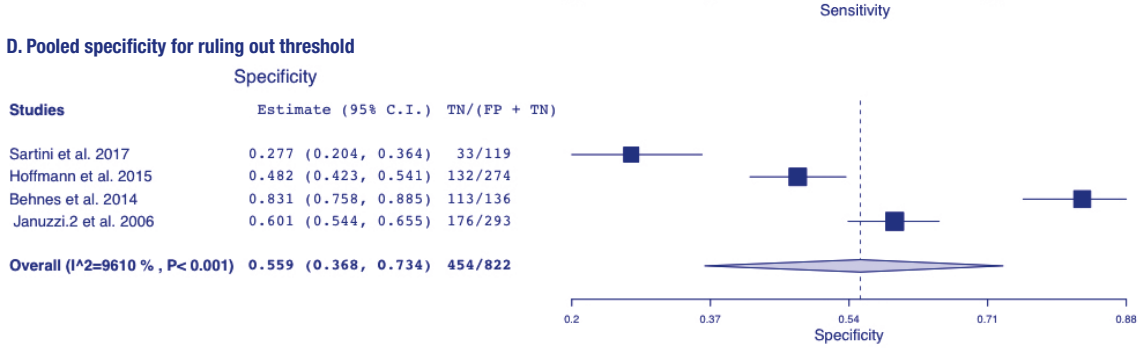
B. Pooled specificity for ruling in threshold



C. Pooled sensitivity for ruling out threshold



D. Pooled specificity for ruling out threshold



CI = confidence interval, TP = true positive, FN = false negative, FP = false positive, TN = true negative, ASHF = acute systolic heart failure, ED = emergency department.

were 97.7% and 55.9%, respectively. Quality adjustment had no statistically significant impact on the result of our analyses.

Our study has a few limitations. First, our study only made use of specific threshold values for ruling in and ruling out ASHF at the ED. The cut-off values of 300 pg/mL, 450 pg/mL, 900 pg/mL and 1800 pg/mL were used in most potentially eligible articles. However, during the study inclusion, we encountered a number of variating thresholds as ruling in or ruling out criteria. Data from these studies could not be used in our meta-analysis. Our study does not answer the question of which cut-off value for NT-proBNP provides the highest sensitivity and specificity in the researched population.

Second, not all included studies supplied detailed information about the prevalence of a previous cardiac history in the enrolled patients. Natriuretic peptides have been shown to be elevated in different types of cardiac disease, such as pulmonary hypertension, pulmonary embolism and renal failure.[23,24] Previous cardiac stress could have caused higher levels of circulating NT-proBNP, therefore increasing the possibility of a patient meeting the ruling in threshold. Since the presence of a full medical history record was unavailable in all studies, this could have caused a higher number of false positive test results.

Third, the included studies had a different approach on the biochemical measurement of NT-proBNP in patient plasma samples. Due to the differences between commercially available NT-proBNP assays, usage of multiple assays cannot be reduced to a singular cut-off value in the diagnostic process of ASHF.[25] The variability in biomarker detection might have impacted the validity of our meta-analysis.

Finally, not all systolic heart failure diagnoses in the included articles were judicated by means of echocardiography, which is the gold standard test for this disease.[9] Some of the analyzed articles based their working diagnosis on expert evaluation of the patient's history, the clinical presentation, routine blood testing and the likelihood of heart failure being the underlying pathology. A possible bias could have been the availability of the NT-proBNP result to the treating physician that diagnosed the patients. This could have led to a higher number of ASHF diagnoses, altering our results to a possibly higher accuracy than we would have found if this would have been taken into consideration in the articles. Other diagnostic approaches could differ in accuracy, potentially leading to more missed cases of heart failure or more incorrectly diagnosed patients.

The pursuit of the highest possible accuracy of diagnostic tests in medical practice is of vital importance for multiple reasons. Inaccurate testing could lead to high rates of falsely diagnosed patients and missed diagnoses. A high false positive rate increases medical expenses for treatment of healthy patients and a high false negative rate poses a threat to patient wellbeing, as these patients might not receive the treatment they need. In addition, false positive diagnoses could cause emotional stressors for the involved patients, which do not contribute to patient wellbeing. In order to reduce health risks and to be cost efficient, medical tests should have both a high sensitivity and a high specificity.

Our study calculated the pooled sensitivity and specificity of often used cut-off values for NT-proBNP in diagnosing ASHF at the ED. Our findings provide insight into the diagnostic value of this biomarker in heart failure. For the ruling in as well as the ruling out threshold, progress still has to be made to improve the diagnostic accuracy of the test. Especially the specificity for ruling out ASHF has room for improvement, since a specificity of 55.9% would mean that 44.1% of patients were falsely tested positive for heart failure. Furthermore, in our analyses, we found statistically significant heterogeneity between the studies used in the calculation of the diagnostic specificity of NT-proBNP. This suggests that the calculated specificities might have been influenced by the wide range of results in the investigated articles, expressing the need for further research on this matter.

We only meta-analyzed the accuracy of the most commonly used cut-off values for NT-proBNP, but several of the excluded articles made use of other thresholds. Whether the investigated threshold can be declared the optimal cut-off value for acute systolic heart failure diagnosis remains uncertain. Therefore, in future research, investigation of an optimally accurate cut-off value of ruling in or ruling out ASHF at the ED is necessary.

References

1 *Niska R., Bhuiya F., Xu J. National Hospital Ambulatory Medical Care Survey: 2007 emergency department summary. Natl Health Stat Report. 2010; 26: 1-31.*

2 *Perk J., De Backer G., Gohlke H., et al. European Guidelines on cardiovascular disease prevention in clinical practice (version 2012). The Fifth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of nine societies and by invited experts). Eur Heart J. 2012; 33: 1635-1701.*

3 *McMurray J.J., Stewart S. Epidemiology, aetiology, and prognosis of heart failure. Heart. 2000; 83: 596-602.*

4 *Nakagawa O., Ogawa Y., Itoh H., et al. Rapid transcriptional activation and early mRNA turnover of brain natriuretic peptide in cardiocyte hypertrophy: evidence for brain natriuretic peptide as an emergency cardiac hormone against ventricular overload. J Clin Invest. 1995; 96: 1280-1287.*

5 *Maisel A.S., Krishnaswamy P., Nowak R.M., et al. Rapid measurement of B-type natriuretic peptide in the emergency diagnosis of heart failure. N Engl J Med. 2002; 347:161-167.*

6 *Maisel A.S., Koon J., Krishnaswamy P., et al. Utility of B-natriuretic peptide as a rapid, point-of-care test for screening patients undergoing echocardiography to determine left ventricular dysfunction. Am Heart J. 2001; 141: 367-374.*

7 *Rodseth R.N., Padayachee L., Biccard B.M. A meta-analysis of the utility of pre-operative brain natriuretic peptide in predicting early and intermediate-term mortality and major adverse cardiac events in vascular surgical patients. Anaesthesia. 2008; 63: 1226-1233.*

8 *Porapakkham P., Porapakkham P., Zimmet H., et al. B-type natriuretic peptide-guided heart failure therapy: A meta-analysis. Arch Intern Med. 2010; 170: 507-514.*

9 *Paulus W.J., Tschope C., Sanderson J.E., et al. How to diagnose diastolic heart failure: a consensus statement on the diagnosis of heart failure with normal left ventricular ejection fraction by the Heart Failure and Echocardiography Associations of the European Society of Cardiology. Eur Heart J. 2007; 28: 2539-2550.*

10 *Roberts E., Ludman A.J., Dworzynski K., et al. The diagnostic accuracy of the natriuretic peptides in heart failure: systematic review and diagnostic meta-analysis in the acute care setting. BMJ. 2015; 350: h910.*

11 *Moher D., Liberati A., Tetzlaff J., et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009; 339: b2535.*

12 *Wallace B. Closing the Gap between Methodologists and End-Users: R as a Computational Back-End. Journal of Statistical Software. 2012; 49.*

13 *DerSimonian R., Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986; 7: 177-188.*

14 *Cochrane collaboration. Review Manager. 2014; 5,3.*

15 *Stoica A., Sorodoc V., Lionte C., et al. Acute cardiac dyspnea in the emergency department: diagnostic value of N-terminal prohormone of brain natriuretic peptide and galectin-3. J Int Med Res. 2018: 300060518798257.*

16 *Januzzi J. L., Chen-Tournoux A.A., Christenson R.H., et al. N-Terminal Pro-B-Type Natriuretic Peptide in the Emergency Department: The ICON-RELOADED Study. J Am Coll Cardiol. 2018; 71: 1191-1200.*

17 *Darche F.F., Baumgartner C., Biener M., et al. Comparative accuracy of NT-proBNP and MR-proANP for the diagnosis of acute heart failure in dyspnoeic patients. ESC Heart Fail. 2017; 4: 232-240.*

18 *Sartini S., Frizzi J., Borselli M., et al. Which method is best for an early accurate diagnosis of acute heart failure? Comparison between lung ultrasound, chest X-ray and NT pro-BNP performance: a prospective study. Intern Emerg Med. 2017; 12: 861-869.*

19 *Gil Martinez P., Mesado Martinez D., Curbelo Garcia J., et al. Amino-terminal pro-B-type natriuretic peptide, inferior vena cava ultrasound, and biolectrical impedance analysis for the diagnosis of acute decompensated CHF. Am J Emerg Med. 2016; 34: 1817-1822.*

20 *Hoffmann U., Espeter F., Weiss C., et al. Ischemic biomarker heart-type fatty acid binding protein (hFABP) in acute heart failure - diagnostic and prognostic insights compared to NT-proBNP and troponin I. BMC Cardiovasc Disord. 2015; 15: 50-015.*

21 *Behnes M., Brueckmann M., Lang S., et al. Connective tissue growth factor (CTGF/CCN2): diagnostic and prognostic value in acute heart failure. Clin Res Cardiol. 2014; 103: 107-116.*

22 *Januzzi J.L., van Kimmenade R., Lainchbury J., et al. NT-proBNP testing for diagnosis and short-term prognosis in acute destabilized heart failure: an international pooled analysis of 1256 patients: the International Collaborative of NT-proBNP Study. Eur Heart J. 2006; 27: 330-337.*

23 *Vickery S., Price C.P., John R.I., et al. B-type natriuretic peptide (BNP) and amino-terminal proBNP in patients with CKD: relationship to renal function and left ventricular hypertrophy. Am J Kidney Dis. 2005; 46: 610-620.*

24 *Leuchte H.H., Neurohr C., Baumgartner R., et al. Brain natriuretic peptide and exercise capacity in lung fibrosis and pulmonary hypertension. Am J Respir Crit Care Med. 2004; 170: 360-365.*

25 *Collin-Chavagnac D., Dehoux M., Schellenberg F., et al. Head-to-head comparison of 10 natriuretic peptide assays. Clin Chem Lab Med. 2015; 53: 1825-1837.*

Clinical outcomes of post-cardiac arrest exposure to hyperoxia: *a systematic review*

Robert R. Zwaan[a], Pieter J.P. Verploegh[a] and Reon Togo[a]
Supervisor: Marco Goeijenbier[b]
[a]Medical students, Erasmus Medical Center Rotterdam, The Netherlands
[b] Md PhD, Dept. of Internal Medicine, Erasmus Medical Center
Correspondence: Reon Togo, email: 424013rt@student.eur.nl

Abstract

Objective: **Due to liberal oxygen supplementation following resuscitation from cardiac arrest (CA), hyperoxia is a common complication seen in the intensive care unit (ICU). There is increasing evidence that hyperoxia can be harmful. We aimed to systematically review the current literature on the association between post-CA exposure to hyperoxia and adverse clinical outcomes.**
Methods: **We searched PubMed/EMBASE for observational studies investigating the effects of post-CA hyperoxia on clinical outcomes. We decided to apply a hyperoxia threshold of 300 mmHg in line with earlier published reports.**
Results: **10 articles met the selection criteria and were included. Five of the 10 articles showed a significant association between post-CA hyperoxia and mortality. Four articles also looked at neurological function as an outcome, with one of them showing a significant association with post-CA hyperoxia. Four articles additionally investigated secondary outcomes (such as length of stay in hospital, improved organ function and functional status at discharge), with two of these studies reporting a significant association.**
Conclusion: **We have found strong evidence supporting the relationship between hyperoxia after resuscitation and increased mortality and poor neurological outcome in CA patients. There was limited evidence for the other outcomes. However, the evidence is restricted to heterogenous, observational studies with mixed results and important methodological issues.**

Introduction

Cardiac arrest (CA) is a common but life-threatening condition that requires immediate medical care. Sudden CA is a leading cause of death worldwide, with approximately 350.000 cases every year in the U.S.[1] The mortality rate of CA remains high, despite improvements in the post-resuscitation management of patients.[2-4] Several interventions such as therapeutic hypothermia and hemodynamic monitoring are widely implemented to improve the post-CA prognosis.[5]

CA patients also regularly receive liberal oxygen supplementation in the intensive care unit (ICU) in an attempt to prevent hypoxia-related injury.[6,7] However, there is increasing evidence that hyperoxia may also be harmful. High concentrations of oxygen in arterial blood can induce vasoconstriction and increased production of reactive oxygen species (ROS).[8] Hyperoxia can therefore contribute to a complex combination of pathophysiological processes following resuscitation called Post-Cardiac Arrest Syndrome (PCAS), including systemic inflammation, myocardial dysfunction and brain injury.[9]

In various experimental studies, animals exposed to hyperoxia following cardiac arrest had worse neurological outcomes, compared to the normoxic group.[10-12] However, clinical studies on the effects of hyperoxia in patients after CA have given conflicting results.[13,14] Studies used varying criteria to define hyperoxia, although most studies applied a predetermined partial pressure of arterial oxygen (PaO2) of 300 mmHg as a boundary to differentiate hyperoxia from normoxia.[15] One study published in 2011 analyzed PaO2 as a continuous variable and found a linear relationship between increasing O2 tensions and mortality.[16]

We therefore conducted a systematic review of observational studies to determine whether hyperoxia after return of spontaneous circulation (ROSC) from CA in adults is associated with poor clinical outcomes, using the cut-off PaO2 value of 300 mmHg for hyperoxia in line with previous studies.[15]

Methods

Literature search

On the 28th of September 2018, we searched PubMed/EMBASE for suitable studies. We used the following combination of search terms: ("Hyperoxia"[MeSH] OR "Hyperoxia"[All Fields]) AND ("Heart arrest"[MeSH] OR "Cardiac arrest"[All Fields]). We also cross-checked our search results with the "relevant articles" section of PubMed and the reference lists of the articles we found.

Selection criteria

We first screened the eligible articles using the title and abstract. We included articles where hyperoxia was observed in adult CA patients in the hours after ROSC. We defined hyperoxia as a PaO2 ≥ 300 mmHg, since the majority of the studies used this cut-off value.

We applied the following exclusion criteria: 1) CA patients with only normoxia and/or hypoxia, 2) participants did not have CA prior to hyperoxia, 3) no clinical outcomes reported (survival,

neurological function etc.), 4) pediatric patients (< 18 years of age), 5) not an original study/trial (reviews, systematic reviews, editorial comments, letters or case reports), 6) studies on animals / in vitro models, 7) non-English articles. Furthermore, we only included articles with full texts that were free and available online through the Erasmus Medical Centre. We did not apply any restrictions on the date of publication for our search. For the remaining articles after the initial screening, we assessed their full texts, according to the aforementioned criteria. We resolved any disagreements by a third person (see Figure 1).

Data extraction

For each of the articles, we gathered general information such as the sample size (N), study population, age and percentage of male participants, and how the arterial blood gas (ABG) data was recorded and assessed. We also collected data on the prevalence of hyperoxia within the cohort and on clinical outcomes.

Quality assessment

We assessed the methodological quality of the included articles using the Newcastle-Ottawa scale (NOS), which is a validated scale for assessing cohort studies.[17] The NOS consists of three categories: selection, comparability and outcome (see Appendix 1).[17] The included studies were individually assessed by the three reviewers and disagreements were resolved by consensus-based discussion.

Results

Study selection

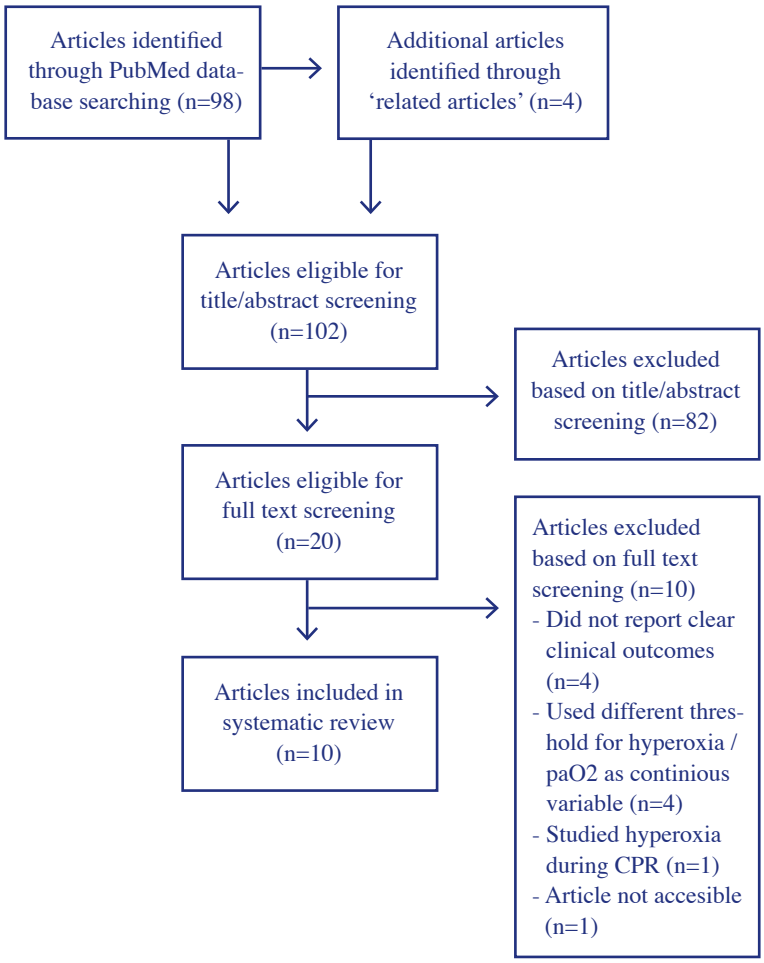
Our systematic search in PubMed resulted in 98 articles. We found four additional articles through the 'related articles' function in PubMed. Of the 102 articles eligible for screening, we excluded 82 articles using our exclusion criteria after screening the titles and abstracts. After full-text assessment using the same set of criteria, 10 more articles were excluded. This resulted in 10 articles for the systematic review (see Figure 1).

Description of studies

Table 1 summarizes the general characteristics of the included studies. We included ten studies in total. Five studies were conducted in North America, three in Oceania, one in Europe and one in Asia. Six studies used retrospective cohorts, two used prospective cohorts, and two studies performed a retrospective analysis of a prospective cohort. Regarding the study population, five studies only included patients with an Out of Hospital Cardiac Arrest (OHCA), one only included patients with an In Hospital Cardiac Arrest (IHCA). Three studies included both OHCA and IHCA patients, and one of the studies included Intensive Care Unit Cardiac Arrest (ICUCA) as well as OHCA and IHCA patients. The sample size of the studies varied between 119 and 12108. The mean age of the included patients varied between 59-66 years. The percentage of male patients varied between 54-80%, although this value was missing in Nelskylä et al.[26]

Table 2 shows the prevalence of hyperoxia in the included studies, while Tables 3a, 3b and 3c summarize the association of hyperoxia with mortality/survival, neurological function and other outcomes, respectively. The results were usually presented as an odds ratio (OR), relative risk (RR) or as percentages, where the hyperoxic group was compared to the normoxic group. Table 4 shows the result of the NOS quality assessment of the included studies. We awarded all 10 studies with either 7 or 8 stars, which can be interpreted as good quality.

Figure 1- Flow diagram of the literature search for studies included in the systematic review



Prevalence of hyperoxia

Nine of the included studies provided data on the prevalence of hyperoxia. Only Johnson et al. did not provide a value for the prevalence of hyperoxia.[21] There was a large variation in the prevalence of hyperoxia across the studies, ranging from as low as 3% to as high as 41% (see Table 2).

Mortality/Survival

All 10 of the included studies investigated the relationship between hyperoxia and mortality, although the significance of the outcomes differed between the studies.

First of all, Kilgannon et al. showed that hyperoxia is significantly associated with increased hospital mortality (adjusted OR 1.8; 95% CI 1.5-2.2; $p < 0.001$).[18] Secondly, Elmer et al. found a 17% decrease in odds of survival to hospital discharge for every hour the patient was exposed to hyperoxia (adjusted OR 0.83; 95% CI 0.69-0.99; $p = 0.04$).[19] Bellomo et al. also reported a 20% increased hospital mortality in patients exposed

Table 1 - Characteristics of the included studies

Study (Author) (Publication year) [Ref #]	Country	Study design	Total N	Study population	Age (mean/ median yrs)	Male (%)	Timing / assessment of ABG
Roberts et al. (2018) [23]	U.S.	Multicenter prospective cohort	280	OHCA, IHCA	59	64	First 6 hr - 1st ABG 1 +/- 2 hr after ROS - 2nd ABG 6 +/- 2 hr after ROSC
Johnson et al. (2017) [21]	U.S.	Multicenter retrospective cohort	544	OHCA, IHCA	61	56	At 1, 6, 12, 24, and 48 hr after CA
Helmerhorst et al. (2015) [24]	The Netherlands	Multicenter retrospective cohort	5258	OHCA	66	70	First 24 hr, worst paO2
Elmer et al. (2015) [19]	U.S.	Retrospective analysis of single-center prospective cohort	184	OHCA, IHCA	60	54	Hourly for 24 hr after CA
Ihle et al. (2013) [25]	Australia/ New Zealand	Retrospective cohort	584	OHCA (ventricular fibrillation)	63	80	First 24 hr, worst PaO2
Nelskylä et al. (2013) [26]	Australia	Retrospective analysis of prospective cohort	119	OHCA, IHCA, ICUCA	61	-	First 24 hr, highest PaO2
Bellomo et al. (2011) [20]	Australia/ New Zealand	Retrospective cohort	12108	OHCA	64	64	First 24 hr, worst PaO2
Kilgannon et al. (2010) [18]	U.S.	Retrospective cohort	6326	OHCA	64	54	First 24 hr, first PaO2
Wang et al. (2017) [22]	U.S./Canada	Prospective cohort	9176	OHCA	64	65	First 24 hr, first & last PaO2
Oh et al. (2014) [27]	South Korea	Retrospective cohort	792	IHCA	64	65	1st ABG 10 min after ROSC 2nd ABG 60-120 min after ROSC

*OHCA = Out-of-hospital cardiac arrest, *IHCA = In-hospital cardiac arrest, *ICUCA = Intensive care unit cardiac arrest, *ABG = Arterial blood gas, *TH = Therapeutic hypothermia

to hyperoxia after adjusting for illness severity (adjusted OR 1.2; 95% CI 1.0-1.5; p = 0.04).[20] When the authors additionally applied Cox proportional hazards modelling of survival, sensitivity analyses using deciles of hypoxemia, time period matching and hyperoxia defined as PaO2 > 400 mmHg, the association no longer reached statistical significance.[20] However, their new PaO2 threshold of 400 mmHg was not relevant for this particular review.[20]

Two of the studies also investigated if the timing of the hyperoxia exposure within the post-resuscitation period affected the outcome.[21,22] Johnson et al. looked at the effect of hyperoxia on mortality for 5 different time periods (1, 6, 12, 24, and 48 hours after CA), but only found a significant association between hyperoxia and decreased odds of survival at 12 hours after CA (adjusted OR 0.17; 95% CI 0.03-0.89; p = 0.032).[21] Wang et al. looked at the first and last PaO2 measurement in the first 24 hours after ROSC.[22] The presence of final or any hyperoxia was associated with an increased in-hospital mortality, with adjusted ORs of 1.6 (95% CI 1.26-2.04) and 1.25 (95% CI 1.11-1.41), respectively.[22] This was not the case for initial hyperoxia (adjusted OR 1.1; 95% CI 0.97-1.26).[22]

In contrast to these findings, Roberts et al., Helmerhorst et al., Nelskylä et al., and Oh et al. did not find a significant association between hyperoxia and mortality/survival.[23-27] Ihle at al. claimed that there was no significant association between hyperoxia and hospital mortality but did find an association with increased ICU mortality.[25]

Table 2 - Prevalence of hyperoxia

Study (Author) [Ref #]	Prevalence of hyperoxia (%)
Roberts et al. [23]	38
Johnson et al. [21]	--
Helmerhorst et al. [24]	3
Elmer et al. [19]	36
Ihle et al. [25]	6
Nelskylä et al. [26]	41
Bellomo et al. [20]	11
Kilgannon et al. [18]	18
Wang et al. [22]	26.5
Oh et al. [27]	11.6

Neurological outcome

Four of the included studies investigated the association between hyperoxia and neurological outcome, of which only Roberts et al. showed significant results.[19,21,23,27]

Firstly, Roberts et al. found after adjustment for potential baseline and post-CA confounders, that hyperoxia was an significant independent predictor of poor neurological function at hospital discharge (defined as a modified Rankin Scale (mRS) > 3), with an adjusted relative risk of 1.23 (95% CI 1.11-1.35; p < 0.001).[23] Secondly, the authors showed that hyperoxia was significantly associated with a 32% increased risk of early neurological injury (defined as a Full Outline of Unresponsiveness (FOUR) score ≤ 6 at 72 hours after ROSC), after analysis using listwise deletion (adjusted RR 1.32; 95% CI 1.03-1.69; p = 0.026).[23] Thirdly, the same study showed that hyperoxia was associated with a 3% increase in risk of poor neurological outcome per hour of exposure (adjusted RR 1.03; 95% CI, 1.02-1.05), therefore demonstrating the cumulative effect of hyperoxia as well.[23] Additionally, Roberts et al. also examined this association across other thresholds for hyperoxia (PaO2 > 100, 150, 200, 250, 350 and 400 mmHg). The significant association with poor neurological outcome only began at a threshold of ≥ 300 mmHg.[23] Despite the significant findings of Roberts et al., the three remaining studies did not show a significant association between hyperoxia and poor neurological outcome.[19,21,27] Johnson et

Table 3a - Mortality/survival

Study (Author) [Ref #]	Reported outcome	Reported OR/RR*	95% confidence interval	p-value
Roberts et al. [23]	In-hospital mortality	Adjusted RR 1.24	0.99-1.55	0.06
Johnson et al. [21]	Survival	Adjusted OR 0.17	0.03-0.89	0.032
Helmerhorst et al. [24]	Hospital mortality	Adjusted OR 1.13	0.81-1.57	0.46
Elmer et al. [19]	Survival to hospital discharge	Adjusted OR 0.83	0.69-0.99	0.04
	(per hour of hyperoxia exposure)			
Ilhe et al. [25]	- Hospital mortality	- Adjusted OR 1.2	- 0.51-2.82	- 0.83
	- ICU mortality	- 35% (vs. 32%)	--	- 0.042
Nelskylä et al. [26]	30-day survival	45% (vs. 36%)	--	0.313
Bellomo et al. [20]	- Hospital mortality	- Adjusted OR 1.2	- 1.0-1.5	- 0.04
	- Hospital mortality**	- Adjusted OR 1.0	- 0.8-1.2	- 0.71
Kilgannon et al. [18]	Hospital mortality	Adjusted OR 1.8	1.5-2.2	<0.001
Wang et al. [22]	- Hospital mortality (initial hyperoxia)	- Adjusted OR 1.1	- 0.97-1.26	--
	- Hospital mortality (final hyperoxia)	- Adjusted OR 1.6	- 1.26-2.04	--
	- Hospital mortality (any hyperoxia)	- Adjusted OR 1.25	- 1.11-1.41	--
Oh et al. [27]	Survival	Adjusted OR 1.03	0.31-3.40	0.96

* OR = Odds ratio, RR = Relative risk
** Hospital mortality after applying Cox proportional hazards modelling of survival, sensitivity analyses using deciles of hypoxemia, time period matching and hyperoxia defined as PaO2 > 400 mmHg. Although the threshold of 400 mmHg made this outcome irrelevant to our study, we decided to report it nevertheless since it formed the main conclusion of the authors.

Table 3b - Neurological outcome

Study (Author) [Ref #]	Reported outcome	Reported OR/RR	95% confidence interval	p-value
Roberts et al. [23]	- Neurological function at hospital discharge (mRS score* > 3)	- Adjusted RR 1.23	- 1.11-1.35	- <0.001
	- Early neurological injury (FOUR score** ≤ 6)	- Adjusted RR 1.32	- 1.03-1.69	- 0.026
	- Neurological outcome (per hour of hyperoxia exposure)	- Adjusted RR 1.03	- 1.02-1.05	--
Johnson et al. [21]	Neurological outcome at hospital discharge	No significant association (data not available)	--	--
Helmerhorst et al. [24]	--	--	--	--
Elmer et al. [19]	Neurological outcome (CPC score*** ≥ 3)	No significant association (data not available)	--	--
Ilhe et al. [25]	--	--	--	--
Nelskylä et al. [26]	--	--	--	--
Bellomo et al. [20]	--	--	--	--
Kilgannon et al. [18]	--	--	--	--
Wang et al. [22]	--	--	--	--
Oh et al. [27]	Neurological outcome	Adjusted OR 1.07	0.30-3.84	0.91

* mRS = modified Rankin Scale (0, no symptoms; 1, no significant disability; 2, slight disability; 3, moderate disability; 4, moderate severe disability; 5, severe disability; 6, death).
** FOUR = Full Outline of Unresponsiveness. A score of brain injury used for comatose patients with 4 components: eye responses, motor responses, brainstem reflexes, and respiration pattern (ranges from 0 to 16 with lower scores indicating worse injury).
*** CPC = Pittsburgh Cerebral Performance Category (1-2, good neurological outcome; 3-5, poor neurological outcome).

Table 3c - Other outcomes

Study (Author) [Ref #]	Reported outcome	Reported OR/RR	95% confidence interval	p-value
Roberts et al. [23]	--	--	--	--
Johnson et al. [21]	--	--	--	--
Helmerhorst et al. [24]	--	--	--	--
Elmer et al. [19]	Organ function at 24 hours	Unadjusted OR 1.00	0.89-1.12	0.97
Ilhe et al. [25]	Median hospital length of stay	7.6 days (vs. 8.9 days)	--	0.013
Nelskylä et al. [26]	- ICU length of stay	- 3.6 days (vs. 3.7 days)	--	- 0.474
	- ICU discharge	- 53% (vs. 46%)	--	- 0.430
	- Hospital discharge	- 41% (vs. 34%)	--	- 0.468
Bellomo et al. [20]	--	--	--	--
Kilgannon et al. [18]	Hospital discharge as functionally independent*	29% (vs. 38%), proportional difference 9%	3-15%	0.002
Wang et al. [22]	--	--	--	--
Oh et al. [27]	--	--	--	--

* Functional independence is defined as the ability to live at home and complete activities of daily living (ADLs) without assistance.

Table 4: Methodological assessment of included studies with NOS

Study (Author) [Ref #]	Selection				Comparability	Outcome			Total
	Representativeness of the exposed cohort	Selection of the non-exposed cohort	Ascertainment of exposure	Demonstration that outcome of interest was not present at start of study	Comparability of cohorts on the basis of the design or analysis controlled for confounders	Assessment of outcome	Was follow-up long enough for outcomes to occur	Adequacy of follow-up of cohorts	score (stars)
Roberts et al. [23]	*	*	*	*	*	*	*	*	8
Johnson et al. [21]	*	*	*	*	*	*	*	*	8
Helmerhorst et al. [24]	*	*	*	*	*	*	*	*	8
Elmer et al. [26]	*	*	*	*	*	*	*	*	8
Ilhe et al. [25]	*	*	*	*	*	*	*	*	8
Nelskylä et al. [26]	*	*	*	-	*	*	*	*	7
Bellomo et al. [19]	*	*	*	-	*	*	*	*	7
Kilgannon et al. [20]	*	*	*	*	*	*	*	*	8
Wang et al. [18]	*	*	*	*	*	*	*	*	8
Oh et al. [27]	*		*	*	*	*	*	*	8

al. found no significant association between hyperoxia measured at five different time intervals within the first 48 hours after CA and neurological outcome at hospital discharge.[21] Elmer et al. did not find a significant association between hyperoxia and the Pittsburgh Cerebral Performance Category (CPC) score for neurological outcome.[19] At last, Oh et al. showed no significant association between hyperoxia and neurological outcome as well, with an adjusted OR of 1.07 (95% CI 0.30-3.84; p = 0.91).[27]

Other outcomes

Besides the most commonly evaluated outcomes, mortality/survival and neurological outcome, four of the included studies examined other clinically relevant outcomes.[18,19,25,26]

To begin with, Nelskylä et al. and Ihle et al. both looked at the length of stay, but had contradictory findings.[25,26] Ilhe et al. found that hyperoxia was associated with a shorter median hospital length of stay, compared to the normoxia group (7.6 vs. 8.9 days; p = 0.013).[25] On the other hand, hyperoxia was not significantly associated with a difference in length of stay at the ICU unit in Nelskylä et al. (3.6 vs. 3.7 days; p = 0.474).[26] In addition to this, Nelskylä et al. showed no significant association between hyperoxia exposure and difference in ICU or hospital discharge.[26]

Kilgannon et al. also investigated the effect of hyperoxia on functional status at hospital discharge.[18] Among hospital survivors, patients with hyperoxia had a significantly lower chance of being discharged as functionally independent (29% vs. 38%; p = 0.002).[18]

Lastly, Elmer et al. concluded that there was no significant association between hyperoxia and improved organ function at 24 hours after CA (unadjusted OR 1.00; 95% CI 0.89-1.12; p = 0.97).[19]

Discussion

In this systematic review we aimed to evaluate the current evidence regarding the association between exposure to hyperoxia following resuscitation and adverse clinical outcomes in CA patients.

Findings

We included observational cohort studies investigating the effect of post-resuscitation hyperoxia on mortality, neurological function as well as some other outcomes. All 10 studies looked at the mortality (most commonly in-hospital), with five of them showing a significant association. Four of the studies looked at neurological outcome as an end-point, with Roberts et al. being the only one to identify a significant association between hyperoxia and poor neurological status.[23] On some occasions, an association between hyperoxia and clinical outcomes was lost after adjusting for multiple variables. Only a few studies investigated the association between hyperoxia and other outcomes, such as length of stay in hospital, improved organ function and functional status at discharge. Therefore, the evidence was insufficient for us to be able to draw a conclusion.

A recent systematic review and meta-analysis of observational studies by Patel et al. found post-arrest hyperoxia to be associated with in-hospital mortality.[28] However, the authors warned that these results should be interpreted with caution because there was significant heterogeneity between studies.[28]

Review strengths/limitations

This review has several strengths. By only including the articles that defined hyperoxia as a PaO2 value greater than 300 mmHg, we aimed to reduce the heterogeneity between the studies that stemmed from the use of a different threshold. A study by Kilgannon et al. however showed that PaO2 as a continuous varia-

ble had a significant linear relationship with mortality.[16] This could be an explanation for the inconsistency of the results, since dichotomization of the variable could lead to loss of valuable information and therefore power.

The methodological quality in all included studies was high after assessment using the NOS, meaning that the risk of bias was small. We also conducted this systematic review according to the PRISMA statement to ensure that our study is thorough and replicable.[29]

We would like to address some important limitations in the included studies that may explain the inconsistent results. The observational nature of the studies mean that findings only represent an association, and therefore do not imply causation. There were substantial differences in the sample size of the different studies included. Considering the relatively low prevalence of hyperoxia, this meant that there was not always sufficient power to identify significant results. The large variation in the prevalence of hyperoxia suggests that there may have been differences in the baseline characteristics of the participants or in the post-resuscitation care between ICU's in different hospitals. For instance, some studies only included OHCA patients, while others included both OHCA and IHCA patients for a more pragmatic approach. Additionally, varying percentages of patients underwent other post-CA interventions, including therapeutic hypothermia and coronary revascularization, leading to differences in the treatment received. Some studies did not adequately adjust for certain confounding variables, such as the timing and duration of hyperoxia exposure, CPR duration and initial cardiac rhythm, which are known to have an influence on the outcomes. This makes the significance of their conclusions more uncertain. Most studies depended on a single ABG measurement obtained at varying points within the first 24 hours after ROSC, while it is definitely possible that the values can fluctuate during this period. A single reading also does not provide information on the duration of the exposure to hyperoxia. Different studies assessed the ABG's differently as well, selecting either the first, highest or worst PaO2 reading. According to a recent cohort study, the incidence of hyperoxia in the ICU as well as the strength of the association with worse outcome, varied depending on the method of ABG measurement applied, with mean/median PaO2 displaying the strongest relationship with outcome.[30] Several studies did acknowledge this problem and took multiple readings into consideration.[21,22]

Although this systematic review is limited to studies that defined hyperoxia as a PaO2 of greater than 300 mmHg, it remains unclear what the optimum lower and upper limits for PaO2 is in clinical practice. In fact, there is very little evidence to support the use of this single cut-off value, which may lead to an under- or overestimation of the hyperoxia incidence. However, Roberts et al. found that poor neurological outcome only began at a PaO2 of ≥ 300 mmHg, thus supporting our choice for this particular threshold.[23]

Conclusion

Based on the results of this systematic review, we can conclude that there may be an association between hyperoxia after resuscitation and adverse clinical outcomes in CA patients. The relationship is a complex one, and to a certain extent depends on various baseline factors such as the cause of cardiac arrest, post-CA care received and duration of hyperoxia.

However, the evidence is restricted to heterogenous, observational studies with mixed results and methodological flaws. It is therefore difficult to comment on the implications for the current clinical practice, and medical professionals can best follow the current post-resuscitation guidelines for optimal oxygen therapy. According to Callaway et al., targeting an arterial oxygen saturation (SaO2) of 94-98%, using a controlled oxygenation strategy, can help avoid potentially harmful exposure to both hypoxia and hyperoxia.[15]

Since the brain is most susceptible to reperfusion injury in the period immediately following ROSC, future studies may benefit from protocol-directed ABG measurements at specified time points in the early post-resuscitation period to avoid inconsistencies and achieve a higher level of evidence.[23]

References

1 *Benjamin EJ, Virani SS, Callaway CW, Chamberlain AM, Chang AR, Cheng S, et al. Heart Disease and Stroke Statistics-2018 Update: A Report From the American Heart Association. Circulation. 2018; 137:e67-e492.*

2 *Carr BG, Kahn JM, Merchant RM, Kramer AA, Neumar RW. Inter-hospital variability in post-cardiac arrest mortality. Resuscitation. 2009; 80:30-34.*

3 *Berdowski J, Berg RA, Tijssen JG, Koster RW. Global incidences of out-of-hospital cardiac arrest and survival rates: Systematic review of 67 prospective studies. Resuscitation. 2010; 81:1479-1487.*

4 *Nichol G, Thomas E, Callaway CW, Hedges J, Powell JL, Aufderheide TP, et al. Regional variation in out-of-hospital cardiac arrest incidence and outcome. JAMA. 2008; 300:1423-1431.*

5 *Sunde K, Pytte M, Jacobsen D, Mangschau A, Jensen LP, Smedsrud C, et al. Implementation of a standardised treatment protocol for post resuscitation care after out-of-hospital cardiac arrest. Resuscitation. 2007; 73:29-39.*

6 *Llitjos JF, Mira JP, Duranteau J, Cariou A. Hyperoxia toxicity after cardiac arrest: What is the evidence? Ann Intensive Care. 2016; 6:23.*

7 *Helmerhorst HJ, Schultz MJ, van der Voort, P H, Bosman RJ, Juffermans NP, de Jonge E, et al. Self-reported attitudes versus actual practice of oxygen therapy by ICU physicians and nurses. Ann Intensive Care. 2014; 4:23.*

8 *Damiani E, Donati A, Girardis M. Oxygen in the critically ill: friend or foe? Curr Opin Anaesthesiol. 2018; 31:129-135.*

9 *Nolan JP, Neumar RW, Adrie C, Aibiki M, Berg RA, Bottiger BW, et al. Post-cardiac arrest syndrome: epidemiology, pathophysiology, treatment, and prognostication. A Scientific Statement from the International Liaison Committee on Resuscitation; the American Heart Association Emergency Cardiovascular Care Committee; the Council on Cardiovascular Surgery and Anesthesia; the Council on Cardiopulmonary, Perioperative, and Critical Care; the Council on Clinical Cardiology; the Council on Stroke. Resuscitation. 2008; 79:350-379.*

10 *Hazelton JL, Balan I, Elmer GI, Kristian T, Rosenthal RE, Krause G, et al. Hyperoxic reperfusion after global cerebral ischemia promotes inflammation and long-term hippocampal neuronal death. J Neurotrauma. 2010; 27:753-762.*

11 *Marsala J, Marsala M, Vanicky I, Galik J, Orendacova J. Post cardiac arrest hyperoxic resuscitation enhances neuronal vulnerability of the respiratory rhythm generator and some brainstem and spinal cord neuronal pools in the dog. Neurosci Lett. 1992; 146:121-124.*

12 *Zwemer CF, Whitesall SE, D'Alecy LG. Cardiopulmonary-cerebral resuscitation with 100% oxygen exacerbates neurological dysfunction following nine minutes of normothermic cardiac arrest in dogs. Resuscitation. 1994; 27:159-170.*

13 *Lee BK, Jeung KW, Lee HY, Lee SJ, Jung YH, Lee WK, et al. Association between mean arterial blood gas tension and outcome in cardiac arrest patients treated with therapeutic hypothermia. Am J Emerg Med. 2014; 32:55-60.*

14 *Vaahersalo J, Bendel S, Reinikainen M, Kurola J, Tiainen M, Raj R, et al. Arterial blood gas tensions after resuscitation from out-of-hospital cardiac arrest: associations with long-term neurologic outcome. Crit Care Med. 2014; 42:1463-1470.*

15 *Callaway CW, Donnino MW, Fink EL, Geocadin RG, Golan E, Kern KB, et al. Part 8: Post-Cardiac Arrest Care: 2015 American Heart Association Guidelines Update for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. Circulation. 2015; 132:465-482.*

16 *Kilgannon JH, Jones AE, Parrillo JE, Dellinger RP, Milcarek B, Hunter K, et al. Relationship between supranormal oxygen tension and outcome after resuscitation from cardiac arrest. Circulation. 2011; 123:2717-2722.*

17 *Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Available at: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp. Accessed 11/3, 2018.*

18 *Kilgannon JH, Jones AE, Shapiro NI, Angelos MG, Milcarek B, Hunter K, et al. Association between arterial hyperoxia following resuscitation from cardiac arrest and in-hospital mortality. JAMA. 2010; 303:2165-2171.*

19 *Elmer J, Scutella M, Pullalarevu R, Wang B, Vaghasia N, Trzeciak S, et al. The association between hyperoxia and patient outcomes after cardiac arrest: analysis of a high-resolution database. Intensive Care Med. 2015; 41:49-57.*

20 *Bellomo R, Bailey M, Eastwood GM, Nichol A, Pilcher D, Hart GK, et al. Arterial hyperoxia and in-hospital mortality after resuscitation from cardiac arrest. Crit Care. 2011; 15:R90.*

21 *Johnson NJ, Dodampahala K, Rosselot B, Perman SM, Mikkelsen ME, Goyal M, et al. The Association Between Arterial Oxygen Tension and Neurological Outcome After Cardiac Arrest. Ther Hypothermia Temp Manag. 2017; 7:36-41.*

22 *Wang HE, Prince DK, Drennan IR, Grunau B, Carlbom DJ, Johnson N, et al. Post-resuscitation arterial oxygen and carbon dioxide and outcomes after out-of-hospital cardiac arrest. Resuscitation. 2017; 120:113-118.*

23 *Roberts BW, Kilgannon JH, Hunter BR, Puskarich MA, Pierce L, Donnino M, et al. Association Between Early Hyperoxia Exposure After Resuscitation From Cardiac Arrest and Neurological Disability: Prospective Multicenter Protocol-Directed Cohort Study. Circulation. 2018; 137:2114-2124.*

24 *Helmerhorst HJ, Roos-Blom MJ, van Westerloo DJ, Abu-Hanna A, de Keizer NF, de Jonge E. Associations of arterial carbon dioxide and arterial oxygen concentrations with hospital mortality after resuscitation from cardiac arrest. Crit Care. 2015; 19:348.*

25 *Ihle JF, Bernard S, Bailey MJ, Pilcher DV, Smith K, Scheinkestel CD. Hyperoxia in the intensive care unit and outcome after out-of-hospital ventricular fibrillation cardiac arrest. Crit Care Resusc. 2013; 15:186-190.*

26 *Nelskyla A, Parr MJ, Skrifvars MB. Prevalence and factors correlating with hyperoxia exposure following cardiac arrest--an observational single centre study. Scand J Trauma Resusc Emerg Med. 2013; 21:35.*

27 *Oh YT, Kim YH, Sohn YD, Park SM, Shin DH, Hwang SY, et al. Early hyperoxemia may not increase mortality after cardiac arrest: a pilot study. Clin Exp Emerg Med. 2014; 1:28-34.*

28 *Patel JK, Kataya A, Parikh PB. Association between intra- and post-arrest hyperoxia on mortality in adults with cardiac arrest: A systematic review and meta-analysis. Resuscitation. 2018; 127:83-88.*

29 *Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ. 2009; 339:b2700.*

30 *Helmerhorst HJ, Arts DL, Schultz MJ, van der Voort, P H, Abu-Hanna A, de Jonge E, et al. Metrics of Arterial Hyperoxia and Associated Outcomes in Critical Care. Crit Care Med. 2017; 45:187-195.*

Appendix 1: Newcastle-Ottawa Scale (NOS) assessment form for cohort studies

Note: A study can be given a maximum of one star for each numbered item within the Selection and Outcome categories. A maximum of two stars can be given for Comparability.

Selection

1) Representativeness of the exposed cohort
   a) Truly representative (one star)
   b) Somewhat representative (one star)
   c) Selected group
   d) No description of the derivation of the cohort
2) Selection of the non-exposed cohort
   a) Drawn from the same community as the exposed cohort (one star)
   b) Drawn from a different source
   c) No description of the derivation of the non exposed cohort
3) Ascertainment of exposure
   a) Secure record (e.g., surgical record) (one star)
   b) Structured interview (one star)
   c) Written self report
   d) No description
   e) Other
4) Demonstration that outcome of interest was not present at start of study
   a) Yes (one star)
   b) No

Comparability

1) Comparability of cohorts on the basis of the design or analysis controlled for confounders
   a) The study controls for age, sex and marital status (one star)
   b) Study controls for other factors (one star)
   c) Cohorts are not comparable on the basis of the design or analysis controlled for confounders

Outcome

1) Assessment of outcome
   a) Independent blind assessment (one star)
   b) Record linkage (one star)
   c) Self report
   d) No description
   e) Other
2) Was follow-up long enough for outcomes to occur
   a) Yes (one star)
   b) No
3) Adequacy of follow-up of cohorts
   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)
   c) Follow up rate less than 80% and no description of those lost
   d) No statement

Interpretation of the NOS (good, fair, and poor):

- Good quality: 3-4 stars in selection domain AND 1-2 stars in comparability domain AND 2-3 stars in outcome domain
- Fair quality: 2 stars in selection domain AND 1-2 stars in comparability domain AND 2-3 stars in outcome domain
- Poor quality: 0-1 star in selection domain OR 0 stars in comparability domain OR 0-1 stars in outcome domain

Initiatives to improve cultural competency among medical students: *a systematic review*

*Jennifer Sat Yung Chin-A-Woeng[a], *, Sara Khoshnaw[a], *, Rava Hossain[a], *, Thomas Pereira Horta[a], *, Juliette A.B. Hopman[b]*
[a] Medical student, Erasmus University Medical Center Rotterdam, the Netherlands
[b] Supervisor, Department of Medical Psychology and Psychotherapy, Erasmus MC University Medical Center Rotterdam, the Netherlands
Correspondence: Rava Hossain, e-mail: 465458rh@student.eur.nl
** These authors contributed equally to this work*

Abstract

Objective: **The objective of this systematic review is to determine the effectiveness of the initiatives designed to improve cultural competency among medical students, and to give an advice to the Erasmus Medical Center (MC) regarding to cultural competency education.**

Methods: **A systematic literature search in Pubmed electronic database for English language publications was conducted, reporting initiatives for cultural competency education and its effectivity. Cultural competency, used in this review, is a combination of knowledge, attitude and skills regarding to cross-cultural situations. The title, abstract and full text of the publications were reviewed for eligibility and appraised by a quality assessment.**

Results: **A total of 272 publications were retrieved, of which seven met the criteria. Positive results regarding to the students' attitude to cross cultural differences were found in four of the seven publications, positive results in knowledge in three of the seven publications and positive results in skills in two of the seven publications. In one publication no changes were found regarding to attitude, knowledge and/or skills.**

Conclusion: **More ways of teaching cultural competency are possible, initiatives must be combined to achieve the optimal outcome.**

Introduction

Cultural competency has become a very important ingredient for physicians.[1] Rotterdam is a multicultural city with more than 50% of the residents from foreign origin.[2] This high percentage is the result of the large migration of non-Western migrants to the Netherlands in the past decades (e.g. Turkish and Moroccan labor migrants and Surinamese) and the recent migration of Afghans and Iranians. The variation in ethnic background of the patient population may lead to difficult communication between doctor and patient.[1] Research has shown that physicians behave less affectively towards patients with an ethnic minority background and check less often whether these patients have understood their message, comparing to majority populations.[1] Evidence suggests that provider–patient communication is directly linked to patient satisfaction and adherence and subsequently to health outcomes.[3]

The main cause for this phenomenon may be the lack of knowledge of the physician about the patients' culture. Every culture is different from another in its daily habits, experience of illness and norms and values. Not knowing and not being aware of these differences can lead to health disparities.[4,5] To avoid these health disparities, education in cultural competency can play an essential role.

The definition of cultural competency, used in this review, is a combination of knowledge, attitude and skills regarding to cross-cultural situations.[6] Every culture contains its own ideas, customs and social behaviors, which differ between societies or particular groups of people.[7] For example, it can refer to characteristics as age, gender, sexual orientation, disability, religion, income level, race or ethnicity, education, geographical location or profession.[8] In this review we specifically focused on ethnic culture, due to the increasingly diverse ethnic cultures in the Netherlands.[9]

Described above, more and more medical faculties are considering to implement cultural competency education into their curriculum.

For example, the medical faculty of Vrije Universiteit (VU) Amsterdam has recently renewed its curriculum with the "leerlijn interculturalisatie". In this curriculum, art is used as a tool to improve cultural competency among medical students.[10]

Also the Erasmus MC wants their students to be as cultural competent as possible.[11] Occasionally attention is paid to cultural competency in the current curriculum. The Medical Psychology department has set up a "five-week Interculturality elective course" for second year students. During these weeks, topics will be discussed such as social-cultural differences with regard to health and care use, communication (including working with interpreters), basic knowledge of migration and migrants, philosophy of life and health care, prejudices and stereotypes, case studies and experiences in care. Furthermore, Erasmus MC has created a Student Working Group on Inclusion and Diversity (SWID). SWID wants to pay attention to the development of knowledge, skills and attitude formation of students in the field

of all forms of diversity. In this way, SWID hopes that the students who graduate from Erasmus MC will become team players who are capable of working in the healthcare of the future. This will be achieved due organizing activities and thinking about new education. In order to create an effective curriculum, knowledge about a framework that focuses on strategies to improve knowledge, attitudes and skills is needed. A previous review has shown that cultural competency training can improve this among medical trainees.[12] However, a good overview about what the effectiveness is of initiatives among medical students, is missing. The aim of this systematic review is to determine which initiatives concerning cultural competency are implemented among undergraduate medical students and what works. In this way we want to give an advice to the Erasmus MC about ideas for a new, more culturally competent, curriculum.

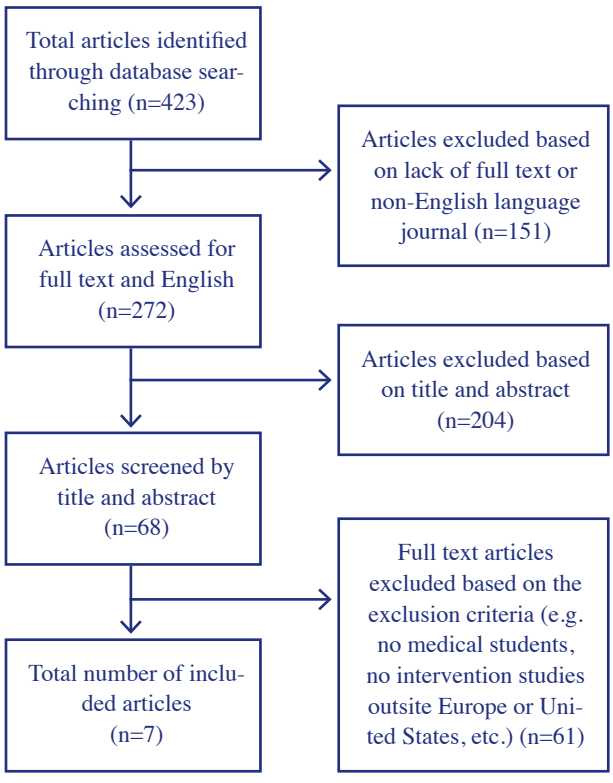
Methods

Search Strategy and Selection Criteria

On December 23nd 2018, a searched for publications available in the Erasmus Medical Centre Pubmed electronic database was done. A systematic search was conducted in PubMed using the following Medical Subject Headings (MeSH) search terms: ("Education, Medical"[Mesh] OR "Students, Medical"[Mesh]) AND "Cultural Competency"[Mesh]. Sequently, all found studies were considered that were written in English and available as full text version in the Erasmus Medical Centre Pubmed electronic database. Four independent authors have reviewed all publications on title and abstract. To determine whether the publication should be excluded, the authors examined the following exclusion criteria:

1. participants were not only medical students (e.g. residents, health professionals, nursing or dental students);
2. participating medical students did not study in Europe or the United States of America;
3. no intervention was included in the study; and
4. the effectiveness on cultural competency was not included as an outcome measure; please see the introduction for the definition of cultural competency.

Figure 1- Selection process of articles



The publications that met the criteria were divided into groups, i.e. group one and group two. Two authors individually read the full text publications of group one and the other two authors read the full articles of group two. The authors of each group then discussed whether the publications of their group were useful. Disagreements were resolved through discussion with a third reviewer (S. Khoshnaw and R. Hossain).

Finally, the reference lists of all included publications were analyzed for additional studies missed by our Pubmed search. However, all of the studies found in these reference lists met one or more exclusion criteria.

Inventory of effects

To assess the different aspects of cultural competency, we used the model of Betancourt.[14] This model shows that training in cross-cultural medicine can divided in three subterms focusing on attitude, knowledge and skills. All three subterms plays a crucial role, and needs to be supported by the other subterms. We examined the improvement the three subterms in the different publications.[13] In the following we define these subterms.

Attitude

The subterm 'attitude' refers to being aware of all outside influences on the patient. It is a combination of being respectful, sensitive and curious.[15,16]

Knowledge

The subterm 'knowledge' refers to having the basic knowledge about group differences regarding to disease incidence and prevalence, ethnopharmacology, and historical/cultural factors that could possibly shape health behavior.[17]

Skills

The subterm 'skills' refers to the ethnographic tools used in medical interviewing, taking into account the cross-cultural issues, social issues, and health beliefs.[18,19,20,21,22]

Results

Search Result

In total 272 publications were retrieved from Pubmed, of which seven complied with all of our criteria. Figure 1 summarizes the study selection process.

Characteristics of the included studies

Intervention characteristics of the included studies are summarized in table 1. The publications originated from two different countries, namely the United States of America and Germany. Of the seven publications, two were randomized controlled trials and the remaining five were either non-randomized trials or observational studies.

Table 1 - Study characteristics

Title	Authors	Study type	Journal	Year	Country	Program
Teaching cross-cultural communication skills online: a multi-method evaluation.[22]	Amy L. Lee et al	Randomized Controlled Trial	Family Medicine	April 2015	USA	Medical students attending the family medicine clerkships followed a 1-hour online reaching module about culture competency and health disparities. This module was combined with an assignment where students practiced using PACT questions with 10 of their patients and sequently writing a reflection.
Promoting medical competencies through international exchange programs: benefits on communication and effective doctor-patient relationships.[23]	Fabian Jacobs et al.	Case report	BMC Medical Education	March 2014	Germany	3rd year medical students followed an international exchange program to Ghana.
The deaf strong hospital program: a model of diversity and inclusion training for first-year medical students. [24]	Denise Thew et al	Case report	Academic Medicine	September 2012	USA	1st year medical students followed an educational program, in which the medical students themselves simulated as patients and deaf healthcare workers as doctors.
Responding to the challenges of teaching cultural competency.[25]	Paritosh Kaul et al	Case report	Medical Education	May 2010	USA	1st year medical students followed an introductory session on culture, health and illness.
Reflective practice enriches clerkship students' cross-cultural experiences.[26]	Desiree Lie et al	Case report	Journal of General Internal Medicine	December 2009	USA	Medical students, attending a four week family medicine clerkship, wrote an assignment which was followed by faculty-facilitated peer discussions.
Empowering Students With Cultural Knowledge: Randomized Controlled Trial of a Cultural competency Curriculum for Third-Year Medical Students.[27]	Inginia Genao et al	Randomized Controlled Trial	Journal of the National Medical Association	December 2009	USA	3rd year medical students followed a training where 7 core elements of cultural competency were highlighted.
A validated cultural competency curriculum for US pediatric clerkships.[28]	Angela P. Mihalic et al.	Case report	Elsevier	July 2009	USA	Medical students followed two interactive workshops, multimedia web cases and a cultural and linguistic competency pocket guide.

Beneficial effects

Attitude

Change in attitude as a result of a cultural competency initiative was assessed in four of the seven publications (57%): the 'Problem-Affect-Concern-Treatment' (PACT)-question model (Amy L. Lee et al.); the deaf strong hospital program (Denise Thew et al.); introductory session on culture, health and illness (Paritosh Kaul et al.); and the reflective practice curriculum (Desiree Lie et al.). All above named initiatives showed positive results regarding the students' attitudes towards cross-cultural differences.

In the PACT-question model positive themes emerged from the students' reflective writing, for example the ability to reflect on the students' own biases against cross-cultural patients.

The Deaf Strong Hospital (DSH) Program positively affected students' attitude and future behavior as physicians when interacting with patients who do not speak English.

The introductory session on culture, health and illness increased students' attitudes significantly (p=0.001) on both care and elicit respectively resulting in large (d=0.66) and moderate (d=0.44) effect sizes.[21]

The reflective practice curriculum achieved a more nuanced understanding of cross-cultural encounters among medical students after discussion. Self-rating of confidence in addressing cultural issues after the curriculum was high, at the scale from 1 till 4 it scored 3.17 ± SD 0.57.

Knowledge

Change in knowledge as a result of a cultural competency initiative was assessed in four of the seven publications (57%): the international exchange program (Fabian Jacobs et al.); the reflective practice curriculum (Desiree Lie et al.); a comprehensive cultural competency curriculum (Inginia Genao et al.); and a validated cultural competency curriculum (Angela P. Mihalic et al.).

The exchange program showed no significant positive impact on knowledge, comparing to students that went to Ethiopia with a representative control group.

The reflective practice curriculum improved cultural knowledge scores significantly 3.17±SD 0.57 (1–4).

The comprehensive cultural curriculum significantly empowered third-year medical students with cultural competency knowledge.

Table 2 - Results of interventions, per domain

Attitude		
Teaching Cross-Cultural Communication Skills Online.[20]	(Amy L. Lee et al. 2015) (Denise Thew et al. 2012)	A more patient-centered focus was achieved by the PACT model, including the ability to reflect upon the students' own biases.
The Deaf Strong Hospital Program: A model of Diversity and Inclusion Training for First-Year Medical Students.[22]	(Paritosh Kaul et al. 2010) (Desiree Lie et al. 2009).	90% of the participating students (strongly) agreed that the DSH program helped them to realize the importance of the cultural, linguistic, and communication issues in delivering health care to patients from different cultures. The program positively affected students' attitudes and future behaviors when interacting with patients who do not speak English.
Responding to the challenges of teaching cultural competency.[23]		Students' attitudes increased significantly (p=0.001) on both care and elicit.Respectively resulting in large (d=0.66) and moderate (d=0.44) effect sizes.
Reflective Practice enriches clerkship students' cross-cultural experiences.[24]		Students' achieved greater synthesis and more nuanced understanding of cross-cultural encounters after discussion. Self-rating of confidence in addressing cultural issues after the curriculum was high at 3.17 ± SD 0.57 (1-4).
Knowledge		
Promoting medical competencies through international exchange programs: benefits on communication and effective doctor-patient relationships.[21]	(Fabian Jacobs et al. 2014) (Desiree Lie et al. 2009) (Inginia Genao et al. 2009) (Angela P. Mihalic et al. 2009).	No significant results were found regarding to knowledge of a second language and/or outside medicine, comparing students that went to Ethiopia with a representative control group.
Reflective Practice enriches clerkship students' cross-cultural experiences.[24]		Cultural knowledge scores improved significantly.
Empowering students with cultural competency knowledge: randomized controlled trial of a cultural competency curriculum for third-year medical students.[25]		Third-year medical students in the intervention group were significantly empowered with cultural competency knowledge when compared to the control group.
A validated cultural competency curriculum for US pediatric clerkships.[26]		Pre- and post-knowledge test scores improved by 17% (p<0.0001).
Skills		
Teaching Cross-Cultural Communication Skills Online.[20]	(Amy L. Lee et al. 2015)	Students asked significantly more PACT questions during the communication-focused SPE, when receiving the intervention.
The Deaf Strong Hospital Program: A model of Diversity and Inclusion Training for First-Year Medical Students.[22]	(Denise Thew et al. 2012)	Former participants, who are now in the clinical orientated stage, reported that the apply lessons from the DSH program in their clinical practice now, including how to work with interpreters and use nonverbal methods of communication. iAlso, the program had influenced how they will interact with non/limited-English-speaking patients.

Through the validated cultural competency curriculum, the pre- and post-knowledge test scores were improved by 17% (p<0.0001).

Skills

Changes in skills, as a result of a cultural competency initiative, was assessed in two of the seven publications (29%): the PACT-question model (Amy L. Lee et al.); and the DSHP (Denise Thew et al.).

In the PACT-question model the participating students used more cross-cultural communication PACT questions compared to the control group in an observed standardized patient exercise. Former participants in the DSH program reported that they are applying lessons from the DSH program in their clinical practice now, including how to work with interpreters and use nonverbal methods of communication. The program has also had positive influences on the interaction with non/limited-English-speaking patients.

No article was found that had an improvement in all the three categories.

Discussion and conclusion

In the last decades, the community of Rotterdam is becoming more and more diverse through migration flows.[2] Future physicians should be prepared for patients with different ethnic background in the consulting room. That is why it is crucial to implement ''cultural competency education'' in the medical curriculum.

Different initiatives are found to teach students cultural competencies, varying from exchange programs to introductory session on culture, health and illness. Except for the study about the exchange program,[23] all of the studies that were included in this review reported improvement in at least one of the following components of cultural competency: attitude, knowledge and skills regarding to cultural competency. However, no initiative was found that had an improvement in all three aspects. For the optimal outcome, improvement in all aspects of cultural competence, it is necessary that different initiatives are implemented.

Limitations

This review has several limitations.

First of all, the MeSH-term that was used in the MEDLINE search "Cultural Competency" is introduced since 2008. That

is why only articles, published between 2008 till 2018, are included. Older articles might have been interesting to include in our review, since the principal behind cultural competency is relatively stable over time. Future research must show whether older interesting initiatives regarding to cultural competency have been missed.

Secondly, although the term "Cultural Competency", introduced in 2008, was used in the MEDLINE search, an article published in 2007 was found in the search result. Normally, articles published before 2008 would not be in the search result, since the MeSH-term was introduced later in time. This increased skepticism about the quality of the MeSH-term "Cultural Competency".

Thirdly, despite the broad term of cultural competency, only articles focused on ethnic related culture were included. Multiple cultural competency articles regarding to "other types of culture", e.g. disability culture or military culture were excluded. Therefore, this review is representative for initiatives specifically focused on improving cultural competency, regarding to ethnic culture.

Fourthly, the literature search was only conducted in PubMed, caused by the restrictive time that was given to write this systematic review. PubMed was chosen because of the relatively high quality of the articles. Due to this, other relevant articles can be missed. Future research must include different databases and a more elaborated search term

At least, initiatives implemented on European medical students were intended to be included, to be as representative as possible to the situation in Rotterdam. However, this resulted in only one publication from Europe. Therefore, studies performed in the United States of America were also decided to be included. In the United States of America research is done among medical students, who live in a diverse environment and who have a general Western Learning. A large limitation hereby is that, despite in both countries physicians are educated, the medical curriculum in the USA has a different structure than in the Netherlands. [10] Future research must be performed to find out whether this makes a difference.

Jernigan V.B.B et al also performed an interesting review on cultural competency training in medical education, which we compared by emphasizing differences.[9]

For example, this review named above included not only undergraduate medical students', but also 'graduated students. Only interventions implemented among undergraduate medical students were included in our review. Furthermore, where Jernigan V.B.B et al included articles from 2000 till 2015, this review included publications from 2008 till 2018.

Additionally, the objective is different. The other review wants to give an advice to all medical universities in the United States of America. The primary goal of our review is to use our gained knowledge to specifically improve the cultural competency education in the curriculum of the Erasmus MC. When positive results are achieved at the Erasmus MC, our next goals are to inspire other medical faculties in the Netherlands.

Advice to the Erasmus MC

We would like to give an advice to the board of directors of the Erasmus MC. Further research needs to be done on the possibilities for the implementation of cultural competency interventions into the medical curriculum of the Erasmus MC. In specific, attention should be spent on the Global exchange-program to Ethiopia, the DSH-program and the introduction lessons given.

No improvements regarding to cultural competency were found in the global exchange program of German students going on exchange to Ethiopia. The quality of this study was assessed as low. Third year medical students at the Erasmus MC have the opportunity to attend the minor global health in foreign countries to (a.o.) improve their cultural competency. Before these students go abroad, they have a five-week lecture course. The (long term) effectiveness of the minor G.H. on specifically cultural competence should be assessed, to find out which aspects are gained during the minor.

In the DSH-program, 40 deaf healthcare workers were needed to educate 100 first year medical students. Every year, 410 first year medical students begin their studies at Erasmus MC. So, when the same form would be implemented at the Erasmus MC, this means a huge labor intensity. The timing of the intervention was perfect, right before all the clinical lessons had even started. If the same timing could be achieved at the Erasmus MC, students are able to take this experience with them through all the other clinical lessons.

In the article regarding the introduction lesson, first year medical students were educated cultural competency by "clinical phase"-students, in collaboration with physicians. This resulted in a higher acceptance to cultural competency by the first-year medical students. The specific Student Workgroup Inclusion and Diversity at the Erasmus MC could help in facilitating this.

Acknowledgements

We want to thank J.A.B. Hopman for feedback on this systematic review. We would also like to thank P.W. Moorman for supporting and teaching us with the literature search.

References

1 *Seeleman, C., Selleger, V., Essink-Bot, M.L. et al., Teaching communication with ethnic minority patients: Ten recommendations. Med Teach. 2011; 33: 814-819.*

2 *World Population Review. Rotterdam Population 2018Available from: http://worldpopulationreview.com/world-cities/rotterdam-population/ Accessed January 2019*

3 *mmunication. Cancer Prev Control. 1999; 3: 25–30.*

4 *Berger, J.T., Culture and ethnicity in clinical care. Arch Intern Med.1998; 158: 2085–2090.*

5 *Hill, R.F., Fortenberry, J.D., Stein, H.F., Culture in clinical medicine. SouthMed J. 1990; 83: 1071–1080.*

6 *Business Dictionary. Definition. Available from: http://www.businessdictionary.com/definition/cultural-competency.htm. Accessed January 2019*

7 *Oxford living Dictionary. Culture. Available from: https://en.oxforddictionaries.com/definition/culture. Accessed January 2019*

8 *Substance Abuse and Mental Health Services Administration. Cultural Competence. Available from: https://www.samhsa.gov/capt/applying-strategic-prevention/cultural-competence. Accessed January 2019*

9 *The post online. WRR: multiculturele samenleving nog multiculturele, faalt nog steeds. Available from: https://politiek.tpo.nl/2018/05/29/wrr-multiculturele-samenleving-nog-multiculture-*

ler-faalt-nog-steeds/ Accessed January 2019

10 *Vrije Universiteit van Amsterdam. Available from: https://www.vu.nl/nl/studiegids/2018-2019/bachelor/g-k/geneeskunde/index.aspx. Accessed December 2018*

11 *Prof. dr. J.J.M. et al. Excellentie door diversiteit. Letter to Raad van Bestuur. 2017.*

12 *Jernigan, V.B.B, et al. An Examination of Cultural Competence Training in US Medical Education Guided by the Tool for Assessing Cultural Competence Training. J Health Dispar Res Pract. 2016; 9:150-167*

13 *Northern Virginia Community College. Evidence-based practice for nursing: levels of evidence. Available from: https://libguides.nvcc.edu/c.php?g=361218&p=2439383. Accessed on January 2019.*

14 *Betancourt, J.R., Cross-cultural Medical Education: Conceptual Approaches and Frameworks for Evalution. Acad. Med. 2003; 78: 560-569*

15 *Bobo, L., Womeodu, R.J., Knox, A.L., Jr. Principles of interculturalmedicine in an internal medicine program. Am J Med Sci. 1991; 302: 244–248.*

16 *Gonzalez-Lee, T., Simon, H.J., Teaching Spanish and cross-culturalsensitivity to medical students. West J Med. 1987; 146: 502–504.*

17 *Seeleman, C., Suurmond, J., Stronks, K., Cultural competence: a conceptual framework forteaching and learning. Med Educ. 2009; 43: 229-237.*

18 *Kent, J. Dezee, Anthony, R., Artino, D., Michael Elnicki, et al. Medical education in the United States of America - Anthony R Artino. 2012; 34: 521–525*

19 *Nunez, A.E., Transforming cultural competence into cross-cultural efficacy in women's health education. Acad Med. 2000; 75: 1071–1080.*

20 *Robins, L.S., Fantone, J.C., Hermann, J. et al., Improving cultural awareness and sensitivity training in medical school. Acad Med. 1998; 73: S31–S34.*

21 *Gamble, V., Under the shadow of Tuskegee: African-Americansand health care. Am J Public Health. 1997; 87: 1773-1778.*

22 *Schulman, K.A., Berlin, J.A., Harless, W. et al., The effect of race and sex on physicians' recommendations for cardiac catherization. N Engl J Med. 1999; 340: 618–626.*

23 *Lee, A.L., Mader, E.M., Morley, C.P., Teaching cross-cultural communication skills online: a multi-method evaluation. Fam Med. 2015; 47:302-308.*

24 *Jacobs, F., Stegmann, K., Siebeck, M., Promoting medical competencies through international exchange programs: benefits on communication and effective doctor-patient relationships. BMC Med Educ. 2014; 14: 43.*

25 *Dr. Denise Thew et al. The Deaf Strong Hospital Program: A Model of Diversity and Inclusion Training for First-Year Medical Students. Acad Med. 2012; 87: 1496–1500.*

26 *Kaul P, Guiton G. Responding to the challenges of teaching cultural competency. Med Educ. 2010; 44: 506.*

27 *Lie, D., Shapiro, J., Cohn, F. et al., Reflective practice enriches clerkship students' cross-cultural experiences. J.Gen Intern Med. 2010; 25 Suppl 2: S119-S125.*

28 *Genao, I., Bussey-Jones, J., St George, D.M. et al., Empowering students with cultural competence knowledge: randomized controlled trial of a cultural competence curriculum for third-year medical students. J Natl Med Assoc. 2009; 101: 1241-1246.*

Does mutant NPM1 MRD predict relapse in adult AML?

Sofie Engels[a], Ferayed Hok[a], Jitte Reinhard[a]. Supervisor: Peter J.M. Valk[b]
[a]Medical students, Erasmus University Medical Center Rotterdam, The Netherlands
[b]Dept. of Hematology, Erasmus University Medical Center Rotterdam, The Netherlands
Correspondence: Sofie Engels, email: 468293se@student.eur.nl
Ferayed Hok, email: 472985fh@student.eur.nl
Jitte Reinhard, email: 468042jr@student.eur.nl

Abstract
Objective: **Acute myeloid leukaemia (AML) is a form of blood cancer with a major biologic heterogeneity. The prognosis of patients with AML dependents on several factors and as such, may differ from each other. The aim of the study was to determine whether mutant Nucleophosmin 1 (NPM1) minimal residual disease (MRD) is a prognostic factor for AML relapse.**
Methods: **We systematically searched PubMed for articles reporting on an association between mutant NPM1 MRD and relapse or survival of adult AML patients before transplantation. A meta-analysis was performed with relapse as the primary endpoint.**
Results: **7 studies were included in the systematic review, including 1274 patients. In our meta-analysis we found an association between mutant NPM1 MRD and AML relapse; random effects model: HR 3.16, 95% CI (1.87-5.34)**
Conclusions: **Mutant NPM1 MRD is a prognostic factor for AML relapse after induction therapy.**

Keywords
Acute myeloid leukaemia, Minimal Residual Disease, NPM1 mutation, Relapse, Adults

Introduction
Acute myeloid leukaemia (AML) is the most common acute leukaemia in adults.[1] AML is caused by uncontrolled proliferation of clonal, undifferentiated cells of the hematopoietic system, which infiltrate bone marrow, blood and other tissues.[2] The majority of patients are older than 65 years.[3] Despite improvements in treatment of AML, the prognosis in elderly patients is poor.[3] 70% of patients 65 years and older will die of AML within 1 year of diagnosis.[4]
AML is characterized by a cytogenetic and molecular heterogeneity.[2] This biologic heterogeneity is illustrated by the variety of molecular aberrations which carry prognostic value, but the clinical application of this new information is currently limited.[2] Nucleophosmin 1 (NPM1) is a molecular marker that is used in clinical practice, such as WHO classification of AML.[2,5] The NPM1 mutation status is incorporated in the 2017 ELN risk stratification.[6] NPM1 mutant AML patients have a favourable outcome, compared with other aberrations, even in the presence of low level FLT3 ITD mutations.[6] 25-35% of AML patients are NPM1 MRD positive.[2] NPM1 is one of the most commonly mutated genes in AML, but the mechanism of leukemogenesis is currently unclear.[5] An important additional predictive factor is monitoring of minimal residual disease (MRD).[7] MRD is defined as detectable leukemic cells, although this patient is in complete remission according to current standards.[8] Residual leukemic cells can still be detected by very sensitive methods, that is what we call presence of MRD. MRD-monitoring in AML with NPM1 mutations has been well studied in recent years.[2]
The current therapy of AML consists of an induction therapy and consolidation therapy. Induction therapy is an intensive chemotherapy with the aim of achieving complete remission. A second induction course is also given. Post-remission therapy is essential when complete remission is achieved after induction therapy. The purpose of this therapy is to maintain remission. Post-remission therapy consists of conventional chemotherapeutic consolidation therapy and/or (allogeneic or autologous) stem cell transplantation. When patients are unlikely to have sustained complete remission with conventional chemotherapy consolidation, they are eligible for (autologous or allogeneic) stem cell transplantation.[2]
It seems that NPM1 mutant AML patients have a better prognosis, but MRD is associated with a worse prognosis. We want to investigate whether patients who are NPM1 MRD positive have a better prognosis than patients who are NPM1 MRD negative. The aim of this systematic review was to evaluate the association between MRD and relapse in adult patients with NPM1 AML.

Methods
Literature search
We conducted a systematic review and meta-analysis by performing a literature search on PubMed on December 11, 2018 using the following search: "Leukaemia, Myeloid, Acute/genetics"[Mesh] AND "Recurrence"[Mesh] AND NPM1[All Fields] AND ("Neoplasm, Residual/genetics"[Mesh] OR ("Mutation"[Mesh] AND "Neoplasm, Residual"[Mesh]) OR "Mutation/genetics"[MESH]).

Inclusion and exclusion criteria
We used exclusion criteria to screen the titles and abstracts. Articles that were not written in English were excluded. Also, reviews were excluded. We included articles in the systematic review after reading the full text. We included articles a) describing the association between NPM1 MRD before (allogeneic or

Table 1 - Quality assessment

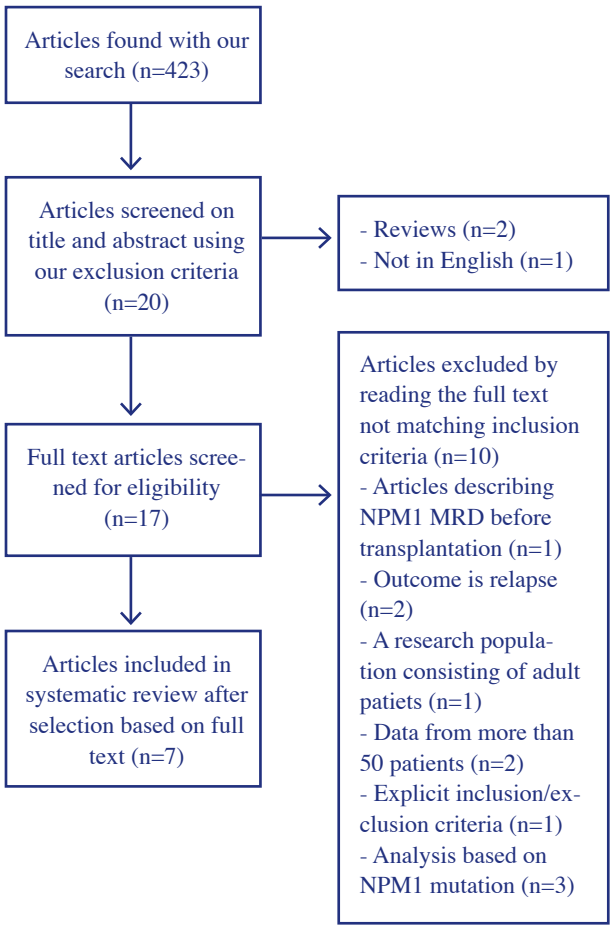
Criteria – Articles with a score <4 were seen as weak articles	Score
1. Representativeness of the exposed cohort	0/1
2. Selection of the non-exposed cohort meets the criteria	0/1
3. Ascertainment of exposure: laboratory measurements	0/1
Comparability	
4. Comparability of cohorts on the basis of the design or analysis controlled for confounders	0/1
Outcome	
5. Assessment of outcome: record linkage	0/1
6. Was follow-up long enough for outcomes to occur	0/1
7. 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	0/1

autologous) stem cell transplantation and relapse of adult AML patients (median age > 50 years), b) in which the analysis was based on NPM1 mutation and c) the article explicitly defined inclusion and exclusion criteria. In case of disagreement, consensus was reached through discussion.

Endpoints

Our primary endpoint was relapse, defined as no recurrence of AML. Relapse was measured from start of first line treatment. The secondary endpoint was overall survival.

Figure 1- Flowchart study selection



Quality assessment

Two independent researchers evaluated the quality of the included articles by analysing full articles according to a previously defined quality score table (Table 1).

We made our own quality assessment based on the Newcastle-Ottawa Scale.[9] This questionnaire contained seven criteria which focused on selection, comparability and outcome. If no consensus would be reached, disagreements would be resolved by a third researcher. Articles with a score of 4 or higher were considered strong articles.

Statistical analysis

We looked at the significance results of each study separately and compared them for signs of consistency in the outcome. Then, we performed a meta-analysis expressing the association between mutant NPM1 MRD and relapse-free survival as hazard ratio (HR) with 95% confidence intervals (CI). We used 'Review Manager Software 5.3' empowered by the Cochrane Library for this. A forest plot shows all hazard ratios. I^2-values describe study heterogeneity. If I-squared was > 50%, we controlled for unobserved heterogeneity by using a random effects model. The test was statistically significant if $p<0.05$.

Results

Literature search

Using the previously mentioned search, we found 20 articles (Figure 1), including two reviews that was excluded based on the set criteria. After reading the full text of every article in the result, we excluded another 10 articles as they did not meet the inclusion criteria, did not describe MRD in NPM1 mutated AML patients, did an analysis not based on NPM1 mutation, did not provide results with which to derive relapse-free survival or overall survival as outcome, did not have at least 50 patients in the cohort or did not provide sufficient exclusion and inclusion criteria for its study or a combination of the above mentioned. Eventually, we included 7 articles in our systematic review.

Study characteristics and quality appraisal

The characteristics of the 7 included studies are shown in Table 2. Measuring MRD in the studies was done either by standard PCR, RT-PCR, qPCR or RT-qPCR. A total of 1274 patients, median age 53.2 (range 6-80), were included in the studies. Median sample size was 158 (range 51-252). Median follow-up for survival varied between 0.8 to 4.9 years. All 7 articles were deemed of strong quality, the lowest rate was 5 out of 7, with 4 or more points defined as strong quality. For this review we looked at MRD measurements done after induction therapy and after stem-cell transplantation.

MRD after induction therapy

Positive MRD after induction therapy consistently shows higher risk of relapse. Six studies provided results showing that positive mutant NPM1 MRD was associated with a higher risk of relapse (Table 3), with hazard ratios ranging between 1.47 by Hubmann et al. (HR 1.47 (95% CI 1.17-1.84); p = 0.001) and 21.2 by Bill et al. (HR 21.1 (95% CI 4.9-91.6); p < 0.001).[10-13,15,16]

Bill et al. also found a higher overall survival (2.9 (95% CI 1.2-7.1); p = 0.020), was linked to negative mutant NPM1 MRD

Table 2 - Study characteristics

Study	Study population	Number of participants with NPM1 MRD (n)	Pre- or post-transplantation	Study outcome	Result quality assessment	Cut-off levels NPM1 mutant
Bill, et al. (2018) [10]	Median 61.6 years (32.6-73.9)	51	Pre-transplantation	- Cumulative incidence of relapse	7	ratio of 0,01
Ivey, et al. (2016) [11]	Median 50 years (6-68)	346	Pre- and post-transplantation	- Overall survival - Risk of relapse - Risk of death	6	ratio of 0,1
Kayser, et al. (2016) [12]	Median 53.3 years (17-73)	67	Pre-transplantation	- Relapse - Overall survival	6	1%
Krönke, et al. (2011) [13]	Median 49 years (19-61)	137	Pre- and post-transplantation	- Relapse - Death	6	200 NPM-1mut/104 ABL
Schnittger, et al. (2009) [14]	Median 58.9 years (20.1-79.3)	252	Pre-transplantation	- Event-free survival	5	3-log
Hubmann, et al. (2014) [15]	Median 57 years (18-80)	158	Pre-transplantation	- Relapse-free survival - Overall survival	6	ratio of 0,01
Shayegi, et al. (2013) [16]	Median 51 years (20-79)	155	Pre- and post-transplantation	- Disease-free - Overall survival	5	1%

Table 3 - Mutant NPM1 MRD after induction therapy

Study	Cumulative incidence of relapse (HR; 95% CI)	Event-free survival (Relative risk; 95% CI)	Overall survival (HR; 95% CI)	Death (HR; 95% CI)
Bill M, 2018 [10]	21.1 (4.9-91.6); p < 0.001		2.9 (1.2-7.1); p = 0.020	
Ivey A, 2016 [11]	4.80 (2.95-7.80); p < 0.001			4.38 (2.57-7.47); p < 001
Kayser S, 2016 [12]	9.03 (1.07-75.9); p = 0.04		0.70 (0.29-1.70); p = 0.02	
Krönke J, 2011 [13]	1.96 (1.55-2.47); p < 0.001			1.70 (1.36-2.14); p < 0.001
Schnittger S, 2009 [14]		1.307 (1.090-1.524); p = 0.006		
Hubmann M, 2014 [15]	1.47 (1.17-1.84); p = 0.001			
Shayegi N, 2013 [16]	3.41 (1.15-10.06); p = 0.03			

after induction, while Kayser et al. found a lower chance at overall survival (0.70 (95% CI 0.29-1.70); p = 0.02) associated with positive mutant NPM1 MRD after induction therapy.[10,12] Both Ivey et al. and Kronke et al. found a higher risk of death in NPM1 mutated AML patients with positive MRD after induction, respectively 4.38 (95% CI 2.57-7.47); p < 001 and 1.70 (95% CI 1.36-2.14); p < 0.001.[11,13]

Schnittger et al. showed that NPM1 mutated AML patients with negative MRD have a significantly higher chance at event-free survival then patients with positive MRD.[14] The results support the thesis that Minimal Residual Disease detection by NPM1 mutation carries prognostic value for relapse of AML.

MRD after consolidation therapy

Hubmann et al. show a significant higher risk of relapse after consolidation therapy (HR 1.39 (95% CI 1.10-1.77); p = 0.006) with positive MRD (Table 4).[15]

MRD after stem cell transplantation

According to Krönke et al., minimal residual disease by NPM1 mutation determined after stem-cell transplantation carries significant higher risk of relapse (HR 2.24 (95% CI 1.83 - 2.74); p< 0.001) and higher risk of death (HR 1.67 (95% CI 1.43-1.95); p<0.001).[13] Schnittger et al. show a higher chance at event-free survival with negative MRD, determined after stem-cell transplantation (Table 5).[14]

Table 4 - Mutant NPM1 MRD after consolidation therapy

Study	Relapse (HR; 95% CI)
Hubmann M, 2014 [15]	1.39 (1.10-1.77); p = 0.006

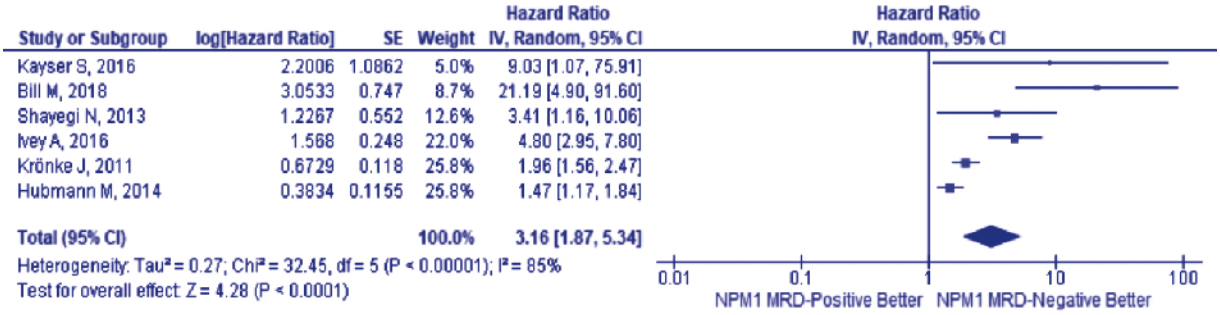
Table 5 - Mutant NPM1 MRD after SCT

Study	Relapse (HR; 95% CI)	Death (HR; 95% CI)	Event-free survival (Relative risk; 95% CI)
Krönke J, 2011 [13]	2.24 (1.83-2.74); p < 0.001	1.67 (1.43-1.95); p < 0.001	
Schnittger S, 2009 [14]			1.060 (1.075-1.113); p = 0.025
Shayegi N, 2013 [16]	1.84 (0.19-17.76); p = 0.6		

Meta-analysis

We found an overall association between minimal residual disease by mutant NPM1 after induction and relapse-free survival using our meta-analysis. The meta-analysis only included the six articles reporting on the incidence of relapse (Table 3) and shows significant prognostic value for MRD in mutant NPM1 AML for relapse-free survival, random effects model: HR 3.16. [95% CI 1.87-5.34] (Figure 2).[13,14,16] I^2 was 85%, suggesting serious heterogeneity between the studies.

Figure 2 - Meta-analysis: Relapse in patients with mutant NPM1 MRD compared to no NPM1 MRD after induction therapy. Outcome: hazard ratio of relapse



Discussion

The prognosis of AML patients shows great variety. In this study we did a literature review to assess the prognostic value of mutant NPM1 MRD.

MRD is a prognostic factor for mutant NPM1 AML after induction therapy. All articles showed a significant lower cumulative incidence of relapse or a higher cumulative incidence of event free survival, after induction therapy, for patients who are NPM1 MRD negative. In our meta-analyses, which included six articles.[7-10,12,13], we found a significant lower outcome for patients who are NPM1 MRD negative.

Bill M, Ivey A, Kayser S and Krönke J showed a significant higher overall survival rate or lower death rate after induction therapy for patients who are NPM1 MRD negative.[10-13,15,16]

Relevance of our conclusion

Our systematic review confirms the prognostic value of minimal residual disease by NPM1 mutation. This confirmation is important for both patients and clinicians. Patients can be better informed about their status and the chances of survival or relapse, while their doctors can make well-informed decisions concerning the patients' therapy. For example, patients who are NPM1 MRD negative, may receive less intensive therapy than patients who are NPM1 MRD positive.

Limitations of the research:

The articles used different NPM1 MRD cut-off levels, therefore it is difficult to compare the results with each other. This may have caused our high heterogeneity.

Schnittger, et al. provided 4 results after induction therapy.[14] Namely after 18-60, 61-120, 121-365 and >365 days.[6] We used the results between 121 days and 365 days, because this was comparable to the number of days in the other articles that discussed induction therapy.

For the cumulative incidence of relapse Shayegi, et al. compared multiple levels of NPM1/ABL MRD, instead of NPM1 MRD or no NPM1 MRD, which forced us to choose one level of NPM1 MRD for the review and meta-analysis.[16] We chose the NPM1 MRD level of 1% (compared with the reference) for further analysis, as this was also the cut-off level of three other included studies.[16]

Some patients had additional aberrations beside a NPM1 mutations, like FLT3. This can affect the prognosis of the patients. For instance, a patient with a FLT3 mutation has more risk to relapse. This can affect our outcome.

The median age of patients included in the articles used for this systematic review is 53 years, but there were also children included (range 6 - 80). Our research was aimed at investigating adult AML relapse and the conclusion should be considered as such. We didn't have the databases of the included studies and the results of the adults and children were not divided. Therefore, it was not possible to extract the results of the children. The median age of each population was at least 50 years. We acknowledge that the median age is lower than the majority of AML patients (older than 65), and may as such have an effect on our conclusion.

Whether mutant NPM1 MRD is a prognostic factor for paediatric AML, remains unclear.

The Newcastle-Ottawa scale was not fully relevant to the studies we were looking for. Our scale isn't researched and critiqued by the scientific community. Our scale is better, because we have criteria established in accordance with current knowledge and understanding of the diagnosis of AML.

Further research is needed to study how to adjust treatment when NPM1 MRD is negative. Also, it would be very interesting to investigate whether other genetic aberration and the MRD status does also have an impact on relapse or survival.

References

1 *Yamamoto J.F., Goodman M.T, Patterns of leukemia incidence in the United States by subtype and demographic characteristics, 1997-2002. Cancer Causes Control 2008; 19: 379-390.*

2 *Dohner H., Weisdorf D.J., Bloomfield C.D., Acute Myeloid Leukemia. N Engl J Med 2015; 373:1136-1152.*

3 *Shah A., Andersson T.M., Rachet B., et al., Survival and cure of acute myeloid leukaemia in England, 1971-2006: a population-based study. Br J Haematol 2013; 162: 509-516.*

4 *Meyers J., Yu Y., Kaye J.A., et al., Medicare fee-for-service enrollees with primary acute myeloid leukemia: an analysis of treatment patterns, survival, and healthcare resource utilization and costs. Appl Health Econ Health Policy 2013; 11: 275-286.*

5 *Heath E.M., Chan S.M., Minden M.D., et al., Biological and clinical consequences of NPM1 mutations in AML. Leukemia 2017; 31: 798-807.*

6 *Dohner H., Estey E, Grimwade D, Amadori S, Appelbaum FR, Buchner T, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. Blood 2017; 129: 424-447.*

7 *Grimwade D, Freeman SD. Defining minimal residual disease in acute myeloid leukemia: which platforms are ready for "prime time"? Blood 2014; 124: 3345-3355.*

8 *Paietta E. Assessing minimal residual disease (MRD) in leukemia: a changing definition and concept? Bone Marrow Transplant 2002; 29: 459-465.*

9 *National Center for Biotechnology Information, NCBI. Newcastle-Ottawa quality assessment Form for Cohort Studies.*

10 *Bill M., Grimm J., Jentzsch M., et al., Digital droplet PCR-based absolute quantification of pre-transplant NPM1 mutation burden predicts relapse in acute myeloid leukemia patients. Ann Hematol 2018; 97: 1757-1765.*

11 *Ivey A., Hills R.K., Simpson M.A., et al., Assessment of Minimal Residual Disease in Standard-Risk AML. N Engl J Med 2016; 374:422-433.*

12 *Kayser S., Benner A., Thiede C., et al., Pretransplant NPM1 MRD levels predict outcome after allogeneic hematopoietic stem cell transplantation in patients with acute myeloid leukemia. Blood Cancer J 2016; 6: e449.*

13 *Kronke J., Schlenk R.F., Jensen K.O., et al., Monitoring of minimal residual disease in NPM1-mutated acute myeloid leukemia: a study from the German-Austrian acute myeloid leukemia study group. J Clin Oncol 2011; 29: 2709-2716.*

14 *Schnittger S., Kern W., Tschulik C., et al., Minimal residual disease levels assessed by NPM1 mutation-specific RQ-PCR provide important prognostic information in AML. Blood 2009; 114: 2220-2231.*

15 *Hubmann M., Kohnke T., Hoster E., et al., Molecular response assessment by quantitative real-time polymerase chain reaction after induction therapy in NPM1-mutated patients identifies those at high risk of relapse. Haematologica 2014; 99: 1317-1325.*

16 *Shayegi N., Kramer M., Bornhauser M., et al., The level of residual disease based on mutant NPM1 is an independent prognostic factor for relapse and survival in AML. Blood 2013; 122: 83-92.*

Systematic Review

Differences in Eye Movements between patients with Fragile X Syndrome and patients with Autism Spectrum Disorder:

a systematic review

Ivana I.S.A. Barendse[a], Puck P.E. Brons[a] , Kyra Lubbers[b]
[a]Medical student, Erasmus MC, Rotterdam, The Netherlands
Correspondence: Ivana Barendse: 466375ib@student.eur.nl; Puck Brons: 470292pb@student.eur.nl
[b] Dept. of Child and Adolescent psychiatry / psychology, Erasmus University Medical Center Rotterdam, the Netherlands

Abstract

Objective: **The aim of this systematic review was to examine potential differences in eye movements in social situations between patients with Fragile X Syndrome (FXS) and patients with Autism spectrum disorder (ASD).**
Methods: **On December 12, 2018, we searched the PubMed database. Articles were included if they used social stimuli and measured the eye movements of the two groups.**
Results: **In total, six articles were included. Significant findings were: participants with FXS had a higher proportion of dwell time, fixated quicker on social stimuli, had a higher proportion of gaze to the face (whereas participants with ASD showed more gaze to objects), and spent a lower proportion of time at the eye region of the face compared to participants with ASD.**
Conclusions: **The results of the studies included in this systematic review seem to suggest that FXS patients tend to be more sensitive to social stimuli than ASD patients, and more likely to avoid eye contact. However, as the included studies used a wide variety of paradigms and the sample sizes are limited, the findings of this review should be interpreted with caution, and it is important that these studies are replicated. Nevertheless, the studies do show that eye tracking might be a valuable tool in clinical research as this non-invasive measure is sensitive to small behavioral differences between groups.**

Keywords

Fragile X Syndrome, Autism Spectrum Disorder, Eye Movements

Introduction

Fragile X Syndrome (FXS) is an inherited disorder, known as the most common cause of intellectual disability.[1] In FXS a repeat mutation in the FMR1 gene hinders the production of the protein FMRP, which is associated with the development of the synapses between neurons.[2,3] The reduced FMRP levels cause an impairment in the development of the nervous system and its functions.[2] It occurs in approximately 1 in 4000 males and 1 in 8000 females.[4] Because FXS is a genetic disorder in a X-linked dominant pattern, it is more common in boys and boys are more likely to have severe symptoms than girls.[5] Patients with FXS may present with social problems, behavioral problems, learning disabilities and developmental delay.[6] In addition, patients with FXS often experience symptoms like hyperactivity, repetitive behavior and impulsivity.[7]

Patients with FXS often have social and communicational problems, and many are diagnosed with autism spectrum disorder (ASD).[6] According to Turner et al, approximately 21% of the children (25,9% of boys) with FXS scored at or above the cut off for ASD.[4] Even though many patients with FXS also receive an ASD diagnosis, studies have reported differences in ASD characteristics between patients with FXS and ASD and patients with non-syndromic ASD. According to McDuffie et al, more severe deficits in complex mannerisms were seen in patients with FXS (with ASD) compared to patients with ASD only.[8] In contrast, Wolff et al. found that individuals with FXS (with ASD) had fewer deficits in social behavior compared to individuals with ASD only.[9] In that study, the social behavior was scaled by measuring the facial expression, gaze integration, social smile and other variables.

Currently, there is no effective treatment available for FXS. Two problems make it difficult to develop an effective treatment. First, there is a lot of variability in the FXS population in terms of the symptoms they experience, and we do not know which characteristics are specific to FXS and which characteristics are caused by for example the comorbid ASD or developmental delay. Second, it is unknown which of those characteristics are sensitive to treatment and which can be measured objectively to study the effectiveness of treatments.

Some researchers have studied characteristics of ASD by mo-

nitoring eye movements of patients in social situations.[10,11] With eye tracking, an individual's gaze and eye movement are followed by using an external device.[10] Eye tracking could be a useful tool to reveal differences between groups of patients, because it can give insight into the ways in which visual attention is distributed.[11] Using an eye tracker could provide insight in which characteristics are typical for FXS and it could show whether the ASD characteristics in individuals with FXS are similar to the ASD characteristics in individuals with ASD only. Similarities in those ASD characteristics could mean that individuals with FXS could benefit from the same treatment as individuals with ASD. Conversely, potential differences in the ASD characteristics between individuals with FXS and individuals with ASD would show that specialized treatment is needed. Because social impairments are an important characteristic of ASD, eye movements in social situations will be the focus of this review. However, it is important to note that this is just one of many characteristics that should be investigated in both groups. Furthermore, aberrant eye movements can never be used as the only biomarker to measure the effectiveness of a treatment. However, this way of measuring would be a good addition to the more subjective measures that are currently used in clinical practice (behavioral observations, parental- and self-reporting).

In this systematic review, we will answer the question what the differences are in eye movements in a social situation between patients with FXS (with or without ASD) and patients with ASD (without FXS). Thurman et al. used a depression, anxiety and mood scale to investigate psychiatric symptoms in individuals with FXS and ASD (without FXS).[12] He found that there was more hyperactive behavior and general anxiety in boys with FXS, compared to boys with ASD (without FXS). Knowing this, the hypothesis of differences in eye movements or gazing behavior in socials situations is that boys with FXS (with or without ASD) will respond less tot social stimuli compared to boys with ASD only, based on the anxiety.

The aim of our review is to find characteristics in eye movements, as an objective measure, which could act as a biomarker in the future. Such characteristics, typical for FXS, could help measure the effectiveness of potential treatments in the future. No systematic review has been published about this yet.

Methods

Search strategy

On December 12, 2018, a search was conducted in PubMed using search terms related to FXS, ASD and eye-movements. Since we are comparing patients with FXS (with/without autism) with patients who suffer from autism only, we wanted both of these terms to appear in our articles. Therefore, we used "AND". Furthermore, articles we included had to be English. Figure 1 shows our final search.

Figure 1 - Search term

(((“autism spectrum disorder”[MeSH Terms] OR “autism”[All Fields]) AND (“fragile x syndrome”[MeSH Terms] OR “fragile x”[All Fields])) AND (“eye movements”[MeSH] OR “eye tracking”[All Fields] OR “eye gaze”[All Fields] OR “eye movement”[All Fields]) AND English[lang])

Study selection

We scanned the articles on their titles and abstracts and discussed whether the articles were relevant for the review. Additionally, we scanned the reference lists of the selected articles for other potentially relevant articles. Studies were included if eye-movements of participants with FXS and participants with ASD were recorded when presented with social stimuli.

We composed the following exclusion criteria: articles were excluded if they compared FXS patients only with a control group (typically developing children), if they did not use eye-tracking or the participant's eye gaze as an outcome, if they used medication as an intervention, if they did not use FXS patients in their study, if they did not use patients with autism only in their study and if they used 'rapid eye movement' (REM) sleep as only outcome.

After our initial screening we did a full text screening to assess the eligibility of the articles we initially included.

Quality assessment

To determine the quality of the articles, we applied a checklist for measuring study quality as described in "The feasibility of creating a checklist for the assessment of the methodological quality both of randomized and non-randomized studies of health care interventions" by Sara H Downs and Nick Black.[13] This checklist consists of 27 questions that can be answered by either "yes" or "no" and some questions can be answered by "unable to determine". Question 5 can also be answered by "partially". For each question, the article could be rewarded with either one point (if the answer is yes) or zero points (if the answer is no or unable to determine). Some questions can be rewarded with more than one point. Furthermore, question 27 concerns the power of each study and therefore gets points according to the height of the power. Table 1 shows the checklist with its questions.

Results

Selection process

Our PubMed search produced 32 articles. After applying our exclusion criteria on the titles and abstracts, 7 articles were eligible. When reading the full text, one article had yet to be excluded as the participants in the ASD group in this article showed ASD symptoms, but did not have a formal ASD diagnosis. Figure 2 shows a flowchart of our study selection process.

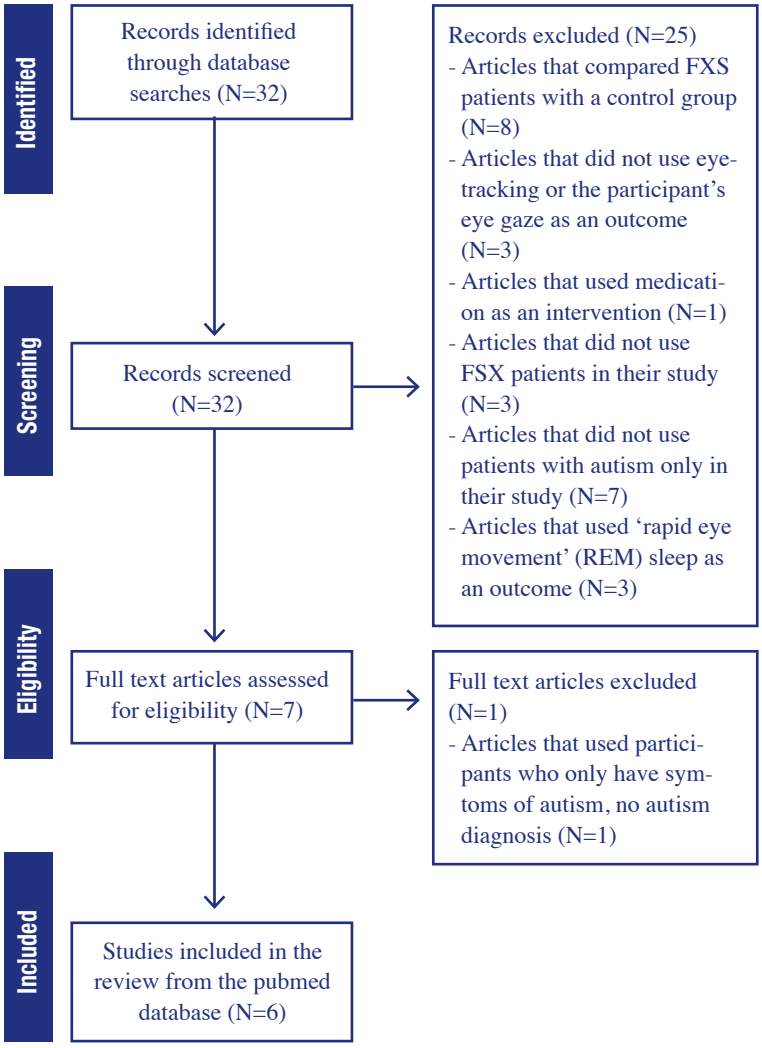
Study characteristics

The characteristics of the included articles are summarized in Table 2. Only Crawford et al. (2015) included one girl with FXS in their study.[11] The other studies exclusively used male participants.

Social stimuli

In 3 out of 6 studies social stimuli consisted of live interactions.[14-16] A video-recording was shown to the participants as a social stimulus in 2 of the studies. One of these studies showed social and non-social stimuli and the other showed a word-learning video performed by an actor.[10,17] The final study used photographs of human faces as social stimuli.[11]

Figure 2 - Flowchart of study selection



Measuring tools

3 out of 6 studies used an eye-tracker as a measuring tool.[10,11,17] In the other 3 studies trained raters analyzed participants' eye-movements using video recordings.[14-16]

Results and conclusions

In 4 studies, the reaction of the participant on social stimuli was the main focus.[10,14,15,17] This was specified as looking at the stimulus (person) or not, either by looking away or by looking at an object.

Friedman et al. had gaze avoidance as a percentage of the total video time as an outcome.[14] They did not find any statistical significant differences between the ASD and the FXS group ($p = 0.466$).

Crawford et al. (2016) observed the proportion of social dwell time of the total dwell time (social and non-social) for stimuli moving towards and stimuli moving past the viewer.[17] They found a higher proportion of dwell time in participants with FXS than in participants with ASD in both the moving towards and the moving past stimuli. They also found that participants with FXS fixated quicker on these stimuli compared to participants with ASD. However, ASD and FXS were not directly compared

Table 1 - Checklist for quality assessment

1. Is the hypothesis/aim/objective of the study clearly described?
2. Are the main outcomes to be measured clearly described in the Introduction or Methods section?
3. Are the characteristics of the patients included in the study clearly described?
4. Are the interventions of interest clearly described?
5. Are the distributions of principal confounders in each group of subjects to be compared clearly described?
6. Are the main findings of the study clearly described?
7. Does the study provide estimates of the random variability in the data for the main outcomes?
8. Have all important adverse events that may be a consequence of the intervention been reported?
9. Have the characteristics of patients lost to follow-up been described?
10. Have actual probability values been reported (e.g. 0.035 rather than <0.05) for the main outcomes except where the probability value is less than 0.001?
11. Were the subjects asked to participate in the study representative of the entire population from which they were recruited?
12. Were those subjects who were prepared to participate representative of the entire population from which they were recruited?
13. Were the staff, places, and facilities where the patients were treated, representative of the treatment the majority of patients receive?
14. Was an attempt made to blind study subjects to the intervention they have received?
15. Was an attempt made to blind those measuring the main outcomes of the intervention?
16. If any of the results of the study were based on "data dredging", was this made clear?
17. In trials and cohort studies, do the analyses adjust for different lengths of follow-up of patients, or in case-control studies, is the time period between the intervention and outcome the same for cases and controls?
18. Were the statistical tests used to assess the main outcomes appropriate?
19. Was compliance with the intervention/s reliable?
20. Were the main outcome measures used accurate (valid and reliable)?
21. Were the patients in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited from the same population?
22. Were study subjects in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited over the same period of time?
23. Were study subjects randomized to intervention groups?
24. Was the randomized intervention assignment concealed from both patients and health care staff until recruitment was complete and irrevocable?
25. Was there adequate adjustment for confounding in the analyses from which the main findings were drawn?
26. Were losses of patients to follow-up taken into account?
27. Did the study have sufficient power to detect a clinically important effect where the probability value for a difference being due to chance is less than 5%?

Table 2 Characteristics of the included studies

Authors	Sample Size (female)	Mean age	Social stimuli	Measuring tool	Outcome	Conclusions
Friedman et al, 2018 [15]	10 (0) individuals with FXS+ASD and 10 (0) individuals with idiopathic ASD.	Individuals with idiopathic ASD had a mean age of 13.45 years (SD=1.81) and individuals with FXS+ASD had a mean age of 12.31 years (SD=2.26).	The participant was sitting next to a trained examinator at a table while communicating for 10 minutes. This conversation was audio and video-taped.	Two trained coders coded the video recordings for eye gaze and gaze avoidance using Procoder video encoding software.	Gaze avoidance as a percentage of the total video time.	Individuals with ASD compared to individuals with FXS+ASD did not show significant differences in the percentage of gaze avoidance during the conversation (p=0.466 95% CI [0.04, 0.06].
Crawford et al, 2016 [11]	Study 1: 16 (1) individuals with ASD Study 2: 15 (0) individuals with FXS	Study 1: individuals with ASD had a mean age of 13.33 years (SD=0.62) Study 2: individuals with FXS had a mean age of 24.21 (SD=8.61)	In both studies, participants were presented with social and non-social video stimuli, showing the stimuli either moving toward or moving past the viewer.	An EyeLink 1000 Tower Mount system.	Proportion of social dwell time of the total dwell time (social and non-social) for stimuli moving towards and moving past the viewer and ratio of time to fixate on social to non-social stimuli.	Individuals with FXS had a higher proportion of dwell time in both the 'moving towards' and the 'moving past' situation than individuals with ASD. Individuals with FXS fixated quicker on social stimuli than on non-social stimuli in both the 'moving towards' as the 'moving past' situation than those with ASD. However, these findings are not statistically compared.
Crawford et al, 2015 [14]	13 (1) individuals with FXS and 15 (3) individuals with ASD.	Individuals with FXS had a mean age of 19.70 (SD=9.00) and individuals with ASD had a mean age of 11.00 (SD=3.48).	Participants were presented with two color photographs of faces that were adjacent to each other. There were 38 photographs in total, most of which displayed a neutral face. On a few photographs, the face had a happy or a disgusted expression. On each photograph, the face had an open mouth and a straigt-ahead gaze.	An EyeLink 1000 Tower Mount system.	Proportion of extra time spent looking at emotion compared to neutral face, proportion of time spent looking at the eye region of neutral faces and proportion of time spent looking at the mouth region of neutral faces.	No significant differences in the proportion of extra time spent looking at emotion between individuals with FXS and those with ASD. Individuals with FXS looked a significantly lower proportion of time at the eye region of the face in comparison with ASD individuals. No significant differences in proportion of time spent looking at the mouth region between both groups.
Benjamin et al, 2014 [10]	14 (0) individuals with FXS and 17 (0) individuals with ASD.	Individuals with FXS had a mean age of 7.68 (SD=1.85) and individuals with ASD had a mean age of 6.92 (SD=1.89).	Participants were shown video-recorded word-learning stimuli, performed by an actor, while they were seated in a chair or on a parent's lap in front of the video recorder, which was placed on a table-mounted eye-tracker.	A Tobii eye-tracker (either Tobii 1750 or Tobii 2150, using identical software).	Gaze proportion to target or face during target object presentation and gaze proportion to distracter or face during distracter object presentation. Follow-up paired t-tests compaired object gaze and face gaze.	FXS individuals showed a significantly greater proportion of gaze to the face than both the target object and the distracter object (p<0.01, p=0.04), while ASD individuals did not show any difference in this. Furthermore, ASD individuals showed significantly more gaze to both target and distracter objects (p=0.001, p=0.03) and less gaze to the face than FXS individuals.
Cohen et al, 1989 [17]	24 (0) individuals with FXS and 24 (0) individuals with non-FXS ASD.	Participants were subdivided into three age groups, each consisting of 8 individuals. For FXS individuals: early childhood (2-5 years) with a mean age of 4.0 (SE=0.3), middle childhood (6-12 years) with a mean age of 8.8 (SE=0.8) and adolescence (13-17 years) with a mean age of 14.6 (SE=0.5). For non-FXS ASD individuals: early childhood with a mean age of 4.1 (SE=0.2), middle childhood with a mean age of 8.8 (SE=1.0) and adolescence with a mean age of 15.1 (SE=0.5).	The subject interacted for 10 minutes with their parent while doing an object-centered task, such as reading a book. They were sitting side by side at a table, across from a one-way mirror.	Highly trained raters coded social gaze of the parent and child using the following coding system: AL, adult looks at child; CL, child looks at adult; BL, both look at each other; NL, neither looks at the other.	The effect of AL on the conditional probabilities of BL and CL in non-FXS ASD individuals compared to FXS individuals. Furthermore, the absolute probabilities of AL,BL and CL were compared in the different age groups.	Individuals with FXS did have a statistically significant difference in social gaze compared to individuals with non-FXS ASD: Individuals with FXS were sensitive to their parents' social gaze initiation (p<0.001), but they found eye contact aversive (p=0.034). In contrast, non-FXS ASD individuals were insensitive to their parents' social gaze initiation and found eye contacty aversive.
Cohen et al, 1988 [16]	12 (0) individuals with FXS, 7 (0) individuals with FXS+ASD, 10 (0) atypical individuals without FXS and 7 (0) individuals with ASD.	Individuals with FXS had a mean age of 7.1 (SD=2.4), individuals with FXS+ASD had a mean age of 7.0 (SD=2.9), atypical individuals had a mean age of 6.5 (SD=2.5) and individuals with ASD had a mean age of 7.0 (SD=2.8).	The participant sat in a room with a one-way mirror playing with material for 30 minutes. Social stimuli were: 10 minutes of noninteraction with the parent in the room, 10 minutes of parent-child interaction and 10 minutes of stranger-child interaction.	Two independent highly trained raters watched and the videotape.	Relative frequency of social gaze to avoidance and absolute frequency of social gaze.	Individuals with ASD showed a higher percentage in both relative frequency of social gaze to avoidance and absolute frequency of social gaze than individuals with FXS+ASD. These findings are not statistically compared.

Table 3 - Results of quality assessment

Question no.	Friedman et al.	Crawford et al. (2016)	Crawford et al. (2015)	Benjamin et al.	Cohen et al. (1989)	Cohen et al. (1988)
1	1	1	1	1	1	1
2	1	1	0	1	1	1
3	1	1	1	1	1	1
4	1	1	1	1	1	1
5	1	2	1	2	1	1
6	1	1	1	1	1	1
7	1	1	1	1	1	1
8	0	0	0	0	0	0
9	0	0	0	0	0	0
10	1	1	1	1	1	1
11	0	1	1	0	0	0
12	0	0	0	0	0	0
13	0	0	0	0	0	0
14	0	0	0	0	0	0
15	0	0	0	0	0	1
16	1	1	1	1	1	1
17	0	0	0	0	0	0
18	1	1	1	1	1	1
19	0	0	0	0	0	0
20	1	1	1	1	1	1
21	0	0	1	0	0	0
22	0	0	0	0	0	0
23	0	0	0	0	0	0
24	0	0	0	0	0	0
25	0	0	0	0	0	0
26	0	0	0	0	0	0
Total	11	13	12	12	11	12

in this study, because the participants of these groups were not matched on a number of participant characteristics. Therefore, these findings could not be statistically compared.

Benjamin et al. investigated gaze proportion to target or face during target object presentation and gaze proportion to distracter or face during distracter object presentation.[10] Their follow-up t-tests compared object gaze (target and distracter) to face gaze. Their study found that FXS individuals showed more gaze to the face than those with ASD during object presentation (target object: $p = 0.007$; distracter object: $p = 0.04$), whereas ASD individuals showed more gaze to the objects than those with FXS (target object: $p = 0.001$; distracter object: $p = 0.03$). Moreover, participants with FXS had a higher proportion of gaze to the face than to the objects (target object: $p < 0.001$; distracter object: $p = 0.04$) and participants with ASD distributed their gaze evenly to both the objects and the face. These findings were statistically significant.

Cohen et al. (1988) had relative frequency of social gaze to avoidance and absolute frequency of social gaze as an outcome.[15] In both relative frequency of social gaze to avoidance and absolute frequency of social gaze, a higher percentage was visible in participants with ASD than in those with FXS+ASD, though these groups were not directly compared in this study. Hence, we do not know whether this difference is statistically significant.

In one study, the part of the face that is most frequently viewed is examined.[11] In Crawford et al. (2015), the only significant outcome was the proportion of time spent looking at the eye region of neutral faces.[11] It was found that individuals with ASD spent a significantly higher proportion of time at the eye region of the face compared to individuals with FXS ($p = 0.001$). No significant differences were found in the proportion of time spent looking at emotional faces (happy or disgust) compared to looking at neutral faces (happy faces: $p = 0.683$; disgust faces: $p = 0.598$). Furthermore, the proportion of time spent looking at the mouth region on neutral faces did not significantly differ between individuals with FXS and individuals with ASD ($p = 0.49$).

The last included study, Cohen et al. (1989), examined the effect of 'adult looks at child' (AL) on the probability of the reaction of the child.[16] The reaction could either be 'child looks at adult' (CL) or 'both look at each other' (BL) or 'neither looks at the other' (NL). It was found that individuals with FXS were sensitive to the parents' social gaze initiation (AL) ($p < 0.001$), but the individuals with FXS had an aversive reaction in their eye contact ($p = 0.034$). In contrast, individuals with non-FXS ASD were insensitive to the parents' social gaze initiation (AL), but they did not find eye contact aversive.

Quality assessment

Two reviewers assessed the quality of the included articles. Table 3 shows the results of our quality assessment. For many questions, all articles are not rewarded with any points. This is due to the fact that these questions do not apply to the type of study that is performed. For example, none of the studies were randomized. We left out question 27 (concerning the power of the studies), because we were not able to calculate the power of all papers.

Cohen et al. (1988) blinded the raters so that they did not know the hypotheses of the study, but no other study did that.[15] Moreover, only Crawford et al. (2015) recruited their participants from the same source.[11] Others recruited their participants from an earlier study or they did not state how they recruited their participants.

In total, the article of Crawford et al. (2015) had the highest quality (13 points) and the article of Friedman et al. and that of Cohen et al. (1989) the lowest (11 points).[11,14,16]

Discussion

In this systematic review, we tried to answer the question what the differences are in eye movements in a social situation between patients with FXS (with or without ASD) and patients with ASD (without FXS).

Benjamin et al. showed that participants with FXS gazed more at the face in comparison with participants with ASD, who showed more gaze at the target or distracter object.[10] Though not statistically compared, Crawford et al. (2016) demonstrated that FXS individuals looked more at and fixated quicker on social stimuli than ASD individuals.[17] By contrast, Friedman et al. did not find any significant differences in social gaze between the two groups.[14] According to Crawford et al. (2015), individuals with FXS spent less time looking at the eye region of the face compared to individuals with ASD.[11] In the article of Cohen et al. (1989) it was found that individuals with FXS were statistically more sensitive to social gaze initiation by others compared to individuals with ASD.[16] Furthermore, they showed that individuals with FXS looked away when eye contact was made, whereas individuals with ASD did not. Lastly, Cohen et al. (1988) found that individuals with ASD showed more social gaze compared to individuals with FXS+ASD.[15]

To be able to analyze these results further, we divided them into

two subcategories: the reaction of the participants to social stimuli and the part of the face participants are more likely to look at.

First, the reaction of the participants to social stimuli was observed. 5 out of the 6 studies belong to this subcategory: Friedman et al., Crawford et al. (2016), Benjamin et al., Cohen et al. (1989) and Cohen et al. (1988).[10,14,15,16,17] Benjamin et al. and Cohen et al. (1989) significantly support that patients with FXS are more sensitive to social stimuli than patients with ASD.[10,16] This might be explained by the study of Wolff et al., which is mentioned earlier.[9] In that study, it is investigated whether there are differences in social behaviors in individuals with FXS+ASD compared to individuals with ASD-FXS. Ultimately, the study showed that individuals with FXS+ASD had fewer deficits in social behaviors than the ASD-FXS group. This might mean that individuals with FXS are more social than individuals with ASD only, because of the greater quality of social overtures, which is found in the study. The fact that individuals with FXS are more social, could explain why individuals with FXS tend to be more sensitive to social stimuli than patients with ASD.
Looking at the results described above, Benjamin et al. and Cohen et al. (1989) significantly support that patient with FXS are more sensitive to social stimuli than patients with ASD.[10,16] Crawford et al. (2016) stated that FXS individuals looked more at and fixated quicker on social stimuli than ASD individuals.[17] Yet, since the fact that these findings are not statistically compared, we do not know whether the finding is significant. The study of Friedman et al. and Cohen et al. (1988) also had outcomes that would allow us to say something about the reaction of the participants to social stimuli.[14,15] Though, there were no significant differences found between both groups in the study of Friedman et al.[14] Furthermore, we are not sure if the results investigated by Cohen et al. (1988) are significant, because those groups were not directly compared in their study so we extracted the results from the figure by visually observing it.[15]
Considering this, we tend to conclude that patients with FXS might be more sensitive to social stimuli than patients with ASD. This means that a patient with FXS is more likely to interact in a social situation. However, only 2 out of the 5 studies mentioned above significantly support this conclusion, so this makes it difficult to form a conclusion based on our results.[10, 16] Nevertheless, we still tend to conclude that there is a greater sensitivity to social stimuli in patients with FXS, based on the 2 articles which statistically investigated this combined with the background data we already read in the article of Wolff et al.[9,10,12,16] More research need to done in order to be able to confirm this in the future.

Second, two studies examined which part of the face is most preferred to look at.[11,15] Crawford et al. (2015) stated that patients with FXS look significantly less to the eyes than to other parts of the face compared to patients with ASD and Cohen et al. (1988) stated that patients with FXS look away after making eye contact with another person.[11,15] This might be explained by the study of Thurman et al.[12] In this study, it is found that boys with FXS had more general anxiety than boys with ASD only. This is also already mentioned in the introduction. The fact that boys with FXS are more scared could mean that they are more curious to keep an eye on anyone who is talking. But by the time the one who is talking looks at the patients, the individuals with FXS might be more likely to look away, because of the social fear. Ultimately, this makes it seem like patients with FXS might not enjoy making eye contact and are more likely to avoid eye contact, which could be a consequence of the fact that patients with FXS suffer from anxiety more frequently.

Our quality assessment showed that the article of Crawford et al. (2016) had the highest quality.[17] However, since the results that are relevant to our review are not statistically compared, we cannot value these results. Friedman et al. and Cohen et al. (1989) had the lowest quality.[14,16] Most of the articles did not state how they recruited their participants, so we were not able to tell if the participants were representative of the entire FXS and ASD population and if that had an influence on the results. Remarkable is that the quality scores of all studies are very similar: the difference in quality score between highest and lowest quality is only two points. Besides, the distribution of the scores in each study is almost similar, which makes it hard to distinguish in the values of the results of the different articles. Since the outcomes of the included studies are not completely similar, it is difficult to explain these different findings. Furthermore, not many articles have been published about this yet.
Based on the results, we have cautiously formed an answer to our research question: there seem to be differences in eye movements in a social situation between patients with FXS and patients with ASD. Patients with FXS tend to be more sensitive to social stimuli, but they seem to be more likely to avoid eye contact than patients with ASD.

This is the first systematic review that has focused on using eye-tracking as a biomarker to monitor treatments in patients with FXS. This subject is of great interest, because developing a treatment for FXS will be easier in the future by using an objectively measurable and sensitive biomarker.
However, there are also a few limitations in this systematic review. First, as described earlier, different social stimuli and measuring tools were used in the studies. 3 out of 6 studies used photos or videos as social stimuli, whereas the other 3 studies used live interacted social stimuli.[10,11,14,15,16,17] This is not remarkable, because there are no guidelines for measuring eye movements nor presenting social stimuli in this context. However, this does make it hard to compare those studies, because the social stimuli are presented in a completely different way. The participant could be more scared of a person than a video. There were also differences between the studies that used live interacted social stimuli. For example, in the study of Cohen et al. (1989), the participants did an object-centered task with their parents, whereas the participants in the study of Friedman et al. had to sit next to a stranger (examiner).[14,16] The fact that the participants had to communicate with an unknown person, could have made them more uncomfortable. This could have influenced the results. That is why the studies that used live interacted social stimuli are also hard to compare to each other.
In the future there should be a standardized approach for both social stimuli and measuring tools. That would make it easier to

compare the results from different studies. We suggest that studies in the future all should use an eye-tracker to measure the eye movements. An eye-tracker is a more objective measuring tool than a measurement by a rater and is less prone to making mistakes or missing movements. In addition, studies should all use real-life interactions as social stimuli, because this simulates reality the best. Via a video or a photograph, the patient might not react in the same way as in a real-life situation, because there is a certain distance between the patient and the stimulus.

Third, a majority of the included studies has not specified whether their participants with FXS had a comorbid ASD or not. If patients with FXS+ASD are compared to patients with ASD, it is possible to specifically examine the influence of FXS on social behavior in patients with ASD. Since this has not been specified in most of the studies, we cannot be certain that the characteristics we found are specific to FXS only.

Another limitation is that only boys with FXS are observed in all studies, except for the article of Crawford et al. (2015), that also observed one girl.[11] Therefore, our conclusion only applies to males.

Moreover, included studies for this systematic review all had a small sample size. Cohen et al. (1989) had the highest sample size of 48 participants, which is still very small.[16] Aside from the fact that a small sample size leads to a low statistical power, this could also mean that the samples are not representative of the entire FXS and ASD population.

Lastly, Crawford et al. (2016) was not able to statistically compare the findings that were relevant to our review and the results of Cohen et al. (1988) were only visible observed.[15, 17] This limits our review because this means that we cannot value their results.

The aim of this study was to examine whether eye-movements of patients with FXS contain characteristics that are specific for FXS in social situations/when presented with social stimuli. This may help determine whether eye-tracking could be considered as an option for measuring the effectiveness of potential treatments in the future. Despite the fact that our findings need to be interpreted cautiously, our systematic review has proved that there are differences in eye movements between patients with ASD and FXS and that eye tracking might be a useful tool to measure these differences. In order to prove our findings, more research with a standardized method needs to be performed.

References

1. *Riley C, Mailick M, Berry-Kravis E, Bolen J. The Future of Fragile X Syndrome: CDC Stakeholder Meeting Summary. Pediatrics. 2017;139(Suppl 3):S147-s52.*
2. *Kraan CM, Godler DE, Amor DJ. Epigenetics of fragile X syndrome and fragile X-related disorders. Dev Med Child Neurol. 2018.*
3. *Sitzmann AF, Hagelstrom RT, Tassone F, Hagerman RJ, Butler MG. Rare FMR1 gene mutations causing fragile X syndrome: A review. American journal of medical genetics Part A. 2018;176(1):11-8.*
4. *Turner G, Webb T, Wake S, Robinson H. Prevalence of fragile X syndrome. American journal of medical genetics. 1996;64(1):196-7.*
5. *Huddleston LB, Visootsak J, Sherman SL. Cognitive aspects of Fragile X syndrome. Wiley interdisciplinary reviews Cognitive science. 2014;5(4):501-8.*
6. *Ke JY, Chen CL, Chen YJ, Chen CH, Lee LF, Chiang TM. Features of developmental functions and autistic profiles in children with fragile X syndrome. Chang Gung medical journal. 2005;28(8):551-8.*
7. *Crawford H, Moss J, Stinton C, Singla G, Oliver C. Overactivity, impulsivity and repetitive behaviour in males with fragile X syndrome: contrasting developmental trajectories in those with and without elevated autism symptoms. Journal of intellectual disability research : JIDR. 2018;62(8):672-83.*
8. *McDuffie A, Thurman AJ, Hagerman RJ, Abbeduto L. Symptoms of Autism in Males with Fragile X Syndrome: A Comparison to Nonsyndromic ASD Using Current ADI-R Scores. Journal of autism and developmental disorders. 2015;45(7):1925-37.*
9. *Wolff JJ, Bodfish JW, Hazlett HC, Lightbody AA, Reiss AL, Piven J. Evidence of a distinct behavioral phenotype in young boys with fragile X syndrome and autism. Journal of the American Academy of Child and Adolescent Psychiatry. 2012;51(12):1324-32.*
10. *Benjamin DP, Mastergeorge AM, McDuffie AS, Kover ST, Hagerman RJ, Abbeduto L. Effects of labeling and pointing on object gaze in boys with fragile X syndrome: an eye-tracking study. Research in developmental disabilities. 2014;35(11):2658-72.*
11. *Crawford H, Moss J, Anderson GM, Oliver C, McCleery JP. Implicit Discrimination of Basic Facial Expressions of Positive/Negative Emotion in Fragile X Syndrome and Autism Spectrum Disorder. American journal on intellectual and developmental disabilities. 2015;120(4):328-45.*
12. *Thurman AJ, McDuffie A, Hagerman R, Abbeduto L. Psychiatric symptoms in boys with fragile X syndrome: a comparison with nonsyndromic autism spectrum disorder. Research in developmental disabilities. 2014;35(5):1072-86.*
13. *Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. Journal of epidemiology and community health. 1998;52(6):377-84.*
14. *Friedman L, Sterling A, Barton-Hulsey A. Gaze avoidance and perseverative language in fragile X syndrome and autism spectrum disorder: brief report. Developmental neurorehabilitation. 2018;21(2):137-40.*
15. *Cohen IL, Fisch GS, Sudhalter V, Wolf-Schein EG, Hanson D, Hagerman R, et al. Social gaze, social avoidance, and repetitive behavior in fragile X males: a controlled study. American journal of mental retardation : AJMR. 1988;92(5):436-46.*
16. *Cohen IL, Vietze PM, Sudhalter V, Jenkins EC, Brown WT. Parent-child dyadic gaze patterns in fragile X males and in non-fragile X males with autistic disorder. Journal of child psychology and psychiatry, and allied disciplines. 1989;30(6):845-56.*
17. *Crawford H, Moss J, Oliver C, Elliott N, Anderson GM, McCleery JP. Visual preference for social stimuli in individuals with autism or neurodevelopmental disorders: an eye-tracking study. Molecular autism. 2016;7:24.*

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