PALACKÝ UNIVERSITY OLOMOUC

Faculty of Medicine and Dentistry

Diagnostic test accuracy of existing practices identifying type 2 pre-diabetes in children: a systematic review of diagnostic test evidence

(Dissertation thesis)

PaedDr. Dagmar Tučková, Ph.D.

Supervisor: PhDr. Miloslav Klugar, Ph.D.

Olomouc 2021

Bibliographical identification

Author´s surname and first name: PaedDr. Dagmar Tučková, Ph.D.

Title of dissertation thesis: Diagnostic test accuracy of existing practices identifying type 2 pre-diabetes in children: a systematic review of diagnostic test evidence

Department: Department of Public Health, Faculty of Medicine and Dentistry, Palacký University in Olomouc

Supervisor: PhDr. Miloslav Klugar, Ph.D.

The year of presentation: 2021

I declare that I have written the dissertation by my own under the supervision of my supervisor PhDr. Miloslav Klugar, Ph.D. All the literature sources I have used are quoted properly and they are listed in a list of used literature.

In Olomouc, 13. 8. 2021 Signature

I would like to express my thanks to my supervisor PhDr. Miloslav Klugar, Ph.D., not only for the professional leading of my dissertation thesis, but also for his help and support throughout my doctoral studies. I would also like to thank to Dr. Abanub Riad as the second reviewer in the assessment of the studies, and his content expertise in epidemiology. I would like to thank also to Dr. Salman Hussain, Ph.D., the content expert for diabetology, for consultations on dialectological aspects of the thesis.

ABSTRACT

Introduction:

The prevalence of non-communicable diseases, cardiovascular diseases, and type 2 diabetes mellitus has increased in the past two decades. According to World Health Organisation, 422 million people worldwide suffer from diabetes. The number had arisen almost fourth times from 108 million people with diabetes in 1980. In 2016, diabetes was direct of 1.6 million deaths. Between 2000 and 2016, there was a 5% increase in premature mortality from diabetes. The need for a practical, accurate diagnostic test for paediatric patients is great due to the epidemic of childhood obesity in developed countries. This systematic review of diagnostic test accuracy is the synthesis of the best available evidence on the diagnostic test accuracy of alternative tests compared to the gold standard in diagnosing of pre-diabetes.

Review objective:

The original review objective was to identify all alternative tests currently in use for the diagnosis of type 2 pre-diabetes in children and establish their accuracy relative to this gold standard. The gold standard for the diagnosis of pre-diabetes was the measurement of fasting plasma glucose and the oral glucose tolerance test.

Inclusion criteria:

This thesis considered varying study designs including cross-sectional studies or diagnostic case-control studies. Cases included children up to 20 years at risk of pre-diabetes with defined characteristics: obesity, hypertension, low high density lipoprotein (HDL) levels, elevated triglyceride levels and glucose intolerance. As an index test, alternate diagnostic tests for pre-diabetes were considered. These tests included but not be limited to any non-fasting tests, HOMA-IR (Homeostatic Model Assessment for Insulin Resistance), measurements of serum glucose and insulin or HbA1c (glycated heameglobin). As a reference test, the measurement of FPG (Fasting plasma glucose) and OGTT (Oral glucose tolerance test) were considered.

Methods:

This systematic review was reported based on the Prefered Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) and followed JBI´s methodology for a systematic review of diagnostic test accuracy. A comprehensive search strategy included BMČ, CINAHL, Cinahl Trials, Cochrane Library, Current control trials, EMBASE, EmCare, ICTRP, Mednar, Ovid Medline, Pedro, ProQuest Dissertation, Scopus, WoS, PsychINFO, Ovid Nursing, COS Conference Papers, and Open Grey. Two independent reviewers screened, critically appraised eligible articles and extracted data using a standardised data extraction tool using JBI Data extraction sheet informed by the JBI System for the Unified Management, Assessment and Review of Information (SUMARI) software.

Data synthesis:

The authors completed a manual calculation and data transformation, so the meta-analyses were possible to be pooled using pooled effect sizes and confidence intervals of the measures provided. SROC (Summary Receive Operating Curve) plot of tests was used to express the results of meta-analyzes. The authors produced synthesised findings across the studies described in a narrative synthesis.

Results:

There were 24 studies that met the inclusion criteria. A total number of four studies was possible to be pooled in the meta-analyses and these studies had two reference tests OGTT and HOMA-IR and five index tests HOMA-IR, HbA1c, TyG (triglycerides), TG-HDL (triglycerids_high density lipoprotein) and FPG. HbA1c used three different cut off values 5.7 %; 5.8 % and 6.5 %; TyG used two different cut off values 8.5 mmol/L and 8.38 mmol/L and TG_HDL used also two different cut off values 2.22 mmol/L and 1.71 mmol/L. Separate meta-analyses were plotted as for different pairs of tests so for different cut off values. The meta-analyses showed a high level of heterogeneity.

The most accurate cut off point from all 24 included studies was proven in Kim´s study (2019). In this study, FPG as an index test and OGTT as a reference test for homogenous Korean population (aged 12.5+/-3.44, 52.1% girls, BMI (body mass index) not known) was used at cut off point ≥7.0 mmol/L with sensitivity 85.10% and specificity 100.00%.

The same cut off point was used in German population (aged 13.1+/-2.4, 55% girls, BMI 30.6+/-5.4 kg/m^2) in Ehehalt´s study (2017) with sensitivity 44.00% and specificity 99.60%. This test was included in the meta-analyses where it was indicated as the second most accurate test.

The most accurate cut off point for HbA1c as an index test and OGTT as a reference test was used in Ehehalt´s study for young German population (aged 13.1+/-2.4, 55% girls, BMI 30.6+/-5.4 kg/m^2). In this study, cut off point 6.5 % was used with sensitivity 84.00% and specificity 99.00% (Ehehalt, 2017). This result was the most accurate result raised from meta-analyses.

The third most accurate result based on the meta-analyses was TG_HDL as an index test with cut off point 1.71 mmol/L with 95.00% sensitivity and 69.00% specificity versus HOMA-IR as a reference test.

Discussion:

From the results of meta-analysis of the SR DTA, it was shown that the index test of HbA1c and a reference test OGTT with the cut off point was 6.5% is the most accurate result which was found out in the young German population (Ehehalt, 2017). The most accurate result from all 24 included studies is the one from Kim´s study with index test FPG at the level of cut off point ≥7.0 mmol/L. These two the most accurate tests are for specific populations, although at the baseline characteristic they have similar features (age, equally represented of both sexes,

obesity etc.). The only difference is in ethnicity, which was one of the predictors why there are differences in cut off points in most of the tests. Another hypothesis which raised on the results of the SR DTA is the need to divide the study groups according to ontogenetic development so that the results of the studies provide the most relevant scientific evidence.

Conclusions:

SR DTA has provided new insights into the accuracy of diagnostic tests in detecting pre-diabetes in children. SR DTA also provided evidence of the accuracy of the 7.0 mmol/L versus OGTT diagnostic FPG test in two studies in two different populations. From the results of SR DTA subsequently arise implications both for practice and for future research. Based on the results, three diagnostic tests are recommended for a practice (FPG at the level of cut off point ≥7.0 mmol/l versus OGTT, HbA1c at the level of cut off point 6.5 % versus OGTT, and TG_HDL at the level of cut off point 1.71 mmol/L versus HOMA-IR). Recommendations for further research should strictly follow the "Standards for Reporting Diagnostic accuracy studies" – STARD reporting guidelines. Emphasis should be placed on the specifics of the defined population and the distribution of age according to ontogenetic development.

Key words:

Evidence-based medicine, evidence-based public health, systematic review, evidence synthesis, diagnostic test accuracy, pre-diabetes, children, HOMA-IR, OGTT, HbA1c, FPG.

ABSTRAKT

Úvod:

Prevalence neinfekčních onemocnění, kardiovaskulárních onemocnění a diabetu mellitu 2. typu se v posledních dvou desetiletích zvýšila. Podle Světové zdravotnické organizace trpí diabetem 422 milionů lidí na celém světě. Počet se zvýšil téměř čtyřnásobně ze 108 milionů lidí s diabetem v roce 1980. V roce 2016 byl diabetes přímým důsledkem 1,6 milionu úmrtí. Mezi lety 2000 a 2016 došlo k 5% nárůstu předčasné úmrtnosti příčinnou diabetu. Potřeba praktického a přesného diagnostického testu pro dětské pacienty je vzhledem k epidemii dětské obezity ve vyspělých zemích velká. Systematické review přesnosti diagnostických testů je syntézou nejlepších dostupných důkazů o přesnosti diagnostických testů alternativních testů ve srovnání se zlatým standardem v diagnostice pre-diabetu.

Cíl systematického review:

Hlavním cílem systematického review bylo identifikovat všechny alternativní testy, které se v současné době používají pro diagnostiku pre-diabetu druhého typu u dětí, a stanovit jejich přesnost vzhledem ke „zlatému standardu“, který je používán. Zlatým standardem pro diagnostiku pre-diabetu je měření plazmatické glukózy nalačno a orální glukózový toleranční test.

Kritéria pro zahrnutí:

Tato disertační práce zahrnovala různé designy studií včetně průřezových studií nebo diagnostických případových studií. Případy zahrnovaly děti do 20 let s rizikem pre-diabetu s charakteristikami, které byly definovány následovně: obezita, hypertenze, nízká hladina HDL, zvýšená hladina triglyceridů a glukózová intolerance. Jako index test byl zvolen jakýkoliv alternativní diagnostický test na pre-diabetes. Tyto testy neměly žádné limity a zahrnovaly testy nalačno, HOMA-IR (homeostatický model inzulinové rezistence), měření hladiny glukózy v séru a inzulínu nebo HbA1c test (test glykovaného hemoglobinu). Jako referenční test bylo zvoleno měření FPG (plasmatická glukóza nalačno) a OGTT (orální test tolerance glukózy).

Metodika:

Systematické review bylo vytvořeno na základě Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) a v jeho tvorbě se postupovalo podle metodiky JBI pro tvorbu systematického review diagnostické přesnosti. Komplexní vyhledávací strategie zahrnovala BMČ, CINAHL, Cinahl Trials, Cochrane Library, Current control trials, EMBASE, EmCare, ICTRP, Mednar, Ovid Medline, Pedro, ProQuest Dissertation, Scopus, WoS, PsychINFO, Ovid Nursing, COS Conference Papers a Open Grey. Dva nezávislí hodnotitelé prověřili, kriticky zhodnotili vyhledané články a extrahovali data pomocí standardizovaného nástroje pro extrakci dat pomocí JBI kontrolního seznamu pro extrakci dat za použití softwaru JBI (System for the Unified Management, Assessment and Review of Information (SUMARI).

Syntéza dat:

Autoři dokončili ruční výpočet a transformaci dat, takže bylo možné pomocí velikostí efektů a intervalů spolehlivosti provést meta-analýzy. Pro vyjádření výsledků z meta-analýz byla použita SROC (Summary Receive Operating Curve) analýza testů. Autoři vytvořili syntetizovaná zjištění napříč studiemi popsanými v narativní syntéze.

Výsledky:

Kritéria pro zařazení splnilo 24 studií. Do meta-analýz bylo možné shromáždit celkový počet čtyř studií a tyto studie měly dva referenční testy OGTT a HOMA-IR a pět index testů HOMA-IR, HbA1c, TyG (triglyceridy), TG-HDL a FPG. U HbA1c byly použity tři různé mezní hodnoty 5.7 %; 5.8 % a 6.5 %; u TyG byly použity dvě různé mezní hodnoty 8.5 mmol/L a 8.38 mmol/L a u TG_HDL byly použity také dvě různé mezní hodnoty 2.22 mmol/L a 1.71 mmol/L. Byly provedeny samostatné meta-analýzy jak pro různé páry testů, tak pro různé mezní hodnoty. Meta-analýzy ukázaly vysokou úroveň heterogenity studií.

Nejpřesnější mezní bod ze všech 24 zahrnutých studií byl prokázán v Kimově studii (2019). V této studii byl použit FPG jako index test a OGTT jako referenční test pro homogenní korejskou populaci (ve věku 12.5 +/- 3.44, 52.1 % dívek, BMI (body mass index) nebylo uvedeno) v hraničním bodě ≥7.0 mmol/l se senzitivitou 85.10% a specificitou 100.00%.

Stejný mezní bod byl použit u německé populace (ve věku 13.1 +/- 2.4, 55 % dívek, BMI 30.6 +/- 5.4 kg/m2) v Ehehaltově studii (2017) se senzitivitou 44.00 % a specificitou 99.60 %. Tento test byl zařazen do meta-analýz, kde byl zhodnocen jako druhý nejpřesnější test.

Nejpřesnější mezní bod pro HbA1c jako index test a OGTT jako referenční test byl použit v Ehehaltově studii (2017) pro mladou německou populaci (ve věku 13.1 +/- 2.4, 55 % dívek, BMI 30.6 +/- 5.4 kg/m2). V této studii byl použit mezní bod 6.5 % se senzitivitou 84.00 % a specificitou 99.00% (Ehehalt, 2017). Tento výsledek byl nejpřesnějším výsledkem získaným z meta-analýz.

Třetím nejpřesnějším výsledkem na základě meta-analýz byl TG_HDL jako index test s hraničním bodem 1.71 mmol/L se senzitivitou 95.00 % a specificitou 69.00 % s HOMA-IR jako referenční test.

Diskuze:

Z výsledků meta-analýzy SR DTA vyplynulo, že index test HbA1c a referenční test OGTT s hraničním bodem byl 6.5% je nejpřesnějším výsledkem, který byl zjištěn u mladé německé populace (Ehehalt, 2017). Nejpřesnějším výsledkem ze všech 24 zahrnutých studií je ten z Kimovy studie s indexním testem FPG na úrovni mezního bodu ≥7.0 mmol/L. Tyto dva nejpřesnější testy jsou pro konkrétní populace, ačkoli na základní charakteristice mají podobné rysy (věk, stejná zastoupení obou pohlaví, obezita atd.). Jediným rozdílem je etnicita, která byla jedním z prediktorů, proč ve většině testu existují rozdíly v mezních bodech. Další hypotéza,

která vycházela z výsledků SR DTA, poukazuje na potřebu rozdělit zkoumané skupiny podle ontogenetického vývoje tak, aby výsledky studií poskytly nejrelevantnější vědecké důkazy.

Závěry:

SR DTA přineslo nové pohledy na přesnost diagnostických testů při detekci pre-diabetu u dětí.. SR DTA přineslo také důkaz o přesnosti diagnostického testu FPG ≥7.0 mmol/L versus OGTT ve dvou studiích u dvou různých populací. Z výsledků SR DTA následně vyplývají implikace jak pro praxi, tak pro budoucí výzkum. Na základě výsledků jsou pro praxi doporučeny tři diagnostické testy (FPG s mezní hodnotou ≥7.0 mmol/l versus OGTT, HbA1c s mezní hodnotou 6.5 % versus OGTT, a TG_HDL s mezní hodnotou 1.71 mmol/L versus HOMA-IR). Doporučení pro další výzkum poukazuje na potřebu striktně následovat standardy pro publikování diagnostických studií (STARD). Měl by se klást důraz na specifika definované populace a rozložení věku podle ontogenetického vývoje.

Klíčová slova:

Medicína založená na důkazech, veřejné zdravotnictví založené na důkazech, systematické review, syntéza důkazů, přesnost diagnostických testů, pre-diabetes, děti, HOMA-IR, OGTT, HbA1c, FPG.

CONTENTS

Table: 1 List of Abbreviations

Abbreviation	Definition
AUC	Area under the response curve
ADA	American Diabetes Association
BMI	Body Mass Index
CGM data	Continuous glucose monitoring
CINAHL	Cumulative Index to Nursing and Allied Health Literature
CMD	Cardiometabolic diseases
CMG	Continuous Glucose Monitoring
CVD	Cardiovascular diseases
DOR	Diagnostic odds ratio
DTA	Diagnostic Test Accuracy
DTA SR	Diagnostic Test Accuracy Systematic Review
EBHC	Evidence-based Healthcare
EBM	Evidence-based Medicine
EMBASE	Excerpta Medica dataBASE
FGIR	Fasting Glucose Insulin Ratio
FN	False Negative
FP	False Positive
FPF	False-positive Fraction
FPG	Fasting Plasma Glucose
FPI	Fasting Plasma Insulin
GRADE	Grading of Recommendations Assessment, Development and Evaluation
HbA1c	A haemoglobin A1c (HbA1c) test measures the amount of blood sugar (glucose) attached to haemoglobin.
HBSC	The Healthy Behaviour in School-aged children
HDL	High Density Lipoprotein
HOMA-IR	Homeostatic Model Assessment for Insulin Resistance
ICTRP	WHO International Clinical Trials Registry Platform
IDF	International Diabetes Federation
IFG	Impaired fasting glucose
IGT	Impaired glucose tolerance
ISI-composite	Insulin sensitivity index
JBI	Joanna Briggs Institute
JBI SUMARI	System for the Unified Management, Assessment and Review of Information
MEDLINE	Medical Literature Analysis and Retrieval System Online
MS	Metabolic Syndrome
NICE	National Institute for Health and Care Excellence

NPV	Negative Predictive Value
OGTT	Oral glucose tolerance test
PPV	Positive Predictive Value
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-analysis
PROSPERO	International prospective register of systematic reviews
QUADAS 2	Quality Assessment of Diagnostic Accuracy Studies
QUICKI	Quantitative insulin-sensitivity check index
RCT	Randomized Controlled Trial
ROC curve	Receiver operating characteristic
RPG	Random plasma glucose
SR	Systematic Review
SROC	Summary ROC
STARD	Standards for Reporting Diagnostic Accuracy Studies
SUMARI	System for the Unified Management, Assessment and Review of Information
T2DM	Type 2 diabetes mellitus
T2PDMC	Type 2 pre-diabetes mellitus
TrG_HDL	Triglyceride high density lipoprotein cholesterol ratio
TG_HDL	Triglyceride high density lipoprotein cholesterol ratio
TN	True Negative
TP	True Positive
TPF	True-positive fraction
TyG_HDL	Triglyceride glucose index
VLDL	Very low density lipoprotein
WHO	World Health Organization

INTRODUCTION

Chronic non-communicable diseases, especially cardiometabolic diseases (hereinafter CMD), cardiovascular diseases (hereinafter CVD) and diabetes mellitus (type-2: noninsulin-dependent, hereinafter T2DM), represent the leading cause of death all over the world and, at the same time, they represent extreme economic burden on health systems. The prevalence of CMD, CVD, and T2DM has increased in the past two decades. According to World Health Organisation, 422 million people worldwide have diabetes. The number had arisen almost fourth times from 108 million people with diabetes in 1980. In 2014, 8.5 % of adults aged 18 years and older had diabetes. In 2016, diabetes was the direct of 1.6 million deaths and in 2012 high blood glucose was the cause of another 2.2 million deaths. Between 2000 and 2016, there was a 5% increase in premature mortality from diabetes. In high-income countries the premature mortality rate due to diabetes decreased from 2000 to 2010 but then increased in 2010-2016. In lower-middle-income countries, the premature mortality rate due to diabetes increased across both periods[1].

T2DM was referred to as a diabetes of adults. It´s manifestation has started commonly after the age of 40), but due to the current sedentary lifestyle, it was also increasingly reported in children. Although thirty years ago, T2DM was considered to be very rare in children and adolescents, in the mid-1990s, investigators began to observe an increasing incidence of T2DM worldwide (Arslanian, 2002). The alarming fact is that recent reports indicate an increasing prevalence of T2DM in children and adolescents around the word in all ethnicities (American Indian, African-American, Asian, or Hispanic/Latino), even if the prevalence of obesity is not increasing any more. Therefore, the occurrence of T2DM has become an emerging clinical problem within paediatric practice.

T2DM is a disease caused by an imbalance between the secretion and effect of insulin on glucose metabolism. This metabolic disorder is characterized by elevated blood glucose levels with concomitant insulin resistance and relative insulin deficiency. Obesity is one of the primary causes of T2DM, especially in people who have an inherited predisposition to the disease. The increase of T2DM in children coincides in time with the obesity epidemic in the 1990s. Therefore, the American Diabetes Association (hereinafter ADA) recommends screening for T2DM the best in the age of 10 years or at the at the beginning of puberty onset, especially in children who are overweight or obese and have two additional risk factors (e.g. family history of T2DM, high blood pressure, a low level of HDL, or high level of triglycerides, high body mass index, inactivity etc.) The standard tests for T2DM identification include the diagnostic criteria for fasting plasma glucose (hereinafter FPG) measurement and oral glucose tolerance test (hereinafter OGTT). Normal FPG is stated as a fasting glucose ≤99 mg/dL. Indicator of impaired glucose tolerance (hereinafter IGT) and pre-diabetes is between 100 and 125 mg/dL (or greater) (7.0 mmol/L), a 2-hour plasma glucose level of 140-199 mg/dL (or

[1] The information were taken from the WHO statistics, available on https://www.who.int/news-room/fact-sheets/detail/diabetes [08-27-2020]

greater) (11.1. mmol/L) during OGTT, an Ac1 level of 6.5 % or more for diagnosis of diabetes mellitus, or a random plasma glucose level of 200 mg/dL (or greater) plus symptoms of polyuria, polydipsia, or unintentional weight loss[2]. These diagnostic criteria are used in everyday medical practice while identifying pre-diabetes, and they are considered to be "gold standard" in identification and screening of pre-diabetes (and we can call them reference tests). But there exist the alternative tests (we can call them "index tests", that are not currently recommended by the ADA. These screening tools could be useful in identification of pre-diabetes when we count on the diagnostic accuracy of these tests. That is the reason why we need to detect the diagnostic accuracy of the reference and index tests and compare them with each other to find out which of these tests brings the best and the most accurate results in identification of pre-diabetes in children. That is the aim of presented dissertation thesis via developed systematic review of the diagnostic accuracy because practice will largely benefit from knowing which of existing test is most accurately detecting children in the risk of pre-diabetes. The synthetized evidence from this review has a large potential to be used worldwide in clinical practice guidelines. That is the main reason, why the thesis is written in English language.

[2] The information were taken from the ADA, available from https://www.diabetes.org/a1c/diagnosis [08-27-2020]

THEORETICAL PART

1. Introduction to evidence-based medicine

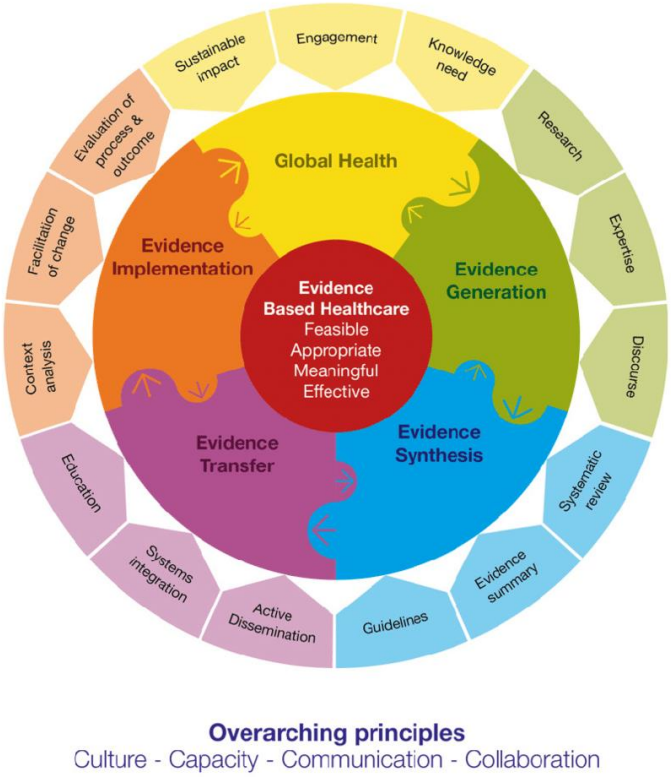
Evidence-based medicine (hereinafter EBM) remains a hot topic for clinicians, public health practitioners, purchasers, planners, and the public (Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996). A more detailed definition stated by Sackett and colleagues (1997) understands the EMB as the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence-based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research (Alan Pearson, Stannard, & Hu, 2012). However, it is not only medicine which is a key to provide the best care in everyday medical practice but e.g. healthcare professionals in all areas of medicine, their (non)clinical decision, policy makers, political leaders, research developers, and users of research who are involved into the evidence-based approach associated with every aspect of healthcare. That is why we talk more about Evidence-based Healthcare (EBHC) in this dissertation thesis. Evidence-based healthcare (hereinafter EBHC) as it is contemporarily conceived is based on the view that clinical decision should be based on the best available scientific evidence but recognizing patient´s preferences, the context of healthcare and the judgement of the clinicians (Alan Pearson, Wiechula, Court, & Lockwood, 2005). The word "E" has become the most common used expression that can be highly trusted because the evidence should be provided from the clinical research. Health professionals seek evidence to substantiate the worth of a very wide range of activities and interventions and thus the type of evidence needed depends on the nature of the activities and its purpose (Alan Pearson, Weeks, & Stern, 2011).

The model of evidence developed by the Joanna Briggs Institute (hereinafter JBI) consists of four major components of the EBHC cycle:

- Healthcare evidence generation,
- Evidence synthesis,
- Evidence (knowledge) transfer, and
- Evidence utilization.

Each of these components is modelled to incorporate their essential elements; and the achievement of improved global health is conceptualized as both the goal and end-point of any or all of the model components and the raison d´être and driver of EBHC (Alan Pearson, Daphne Stannard, et al., 2012). Using all of this information, health professionals are in position to make evidence informed decisions (Alan Pearson et al., 2011).

Figure 1: Conceptual model of EBHC[3]



Source: https://www.researchgate.net/figure/The-JBI-conceptual-model-for-evidence-based-healthcare_fig1_316328453

As it is seen, every step-in clinical inquiry process that is done with respect to specific population, cultures or environment, goes through an appraisal, synthesis and transformation of service delivery settings and professionals in healthcare. As the inception of the JBI, there has been a focus on ensuring that health professionals have access to information that addresses the different types of questions that may arise in clinical practice. This unique articulation of what constitutes evidence for decision-making was a first in the field at the time of the publication of the original model in 2005. The FAME Framework and this broader conceptualization of evidence is frequently cited and clearly resonates with those seeking to conduct research that is relevant to point of care decision-making.

The whole model of EBHC stands on four basic pillars:

- Feasibility (is the extent to which an activity is practical and practicable)
- Appropriateness (relates to the extent to which an intervention or activity fits with or is apt in a situation)
- Meaningfulness (refers to how an intervention or activity is experienced by the patient)
- Effectiveness (is the extent to which an intervention, when used appropriately, achieves the intended effect) (A. Pearson et al., 2012).

[3] (A. Pearson, Jordan, & Munn, 2012), (Alan Pearson, Loveday, & Holopainen, 2012)

When we transform these four pillars to the scale, we find out that the healthcare evidence includes questions related to the feasibility, appropriateness, and meaningfulness of healthcare practices and interventions, as well as their effectiveness and that such a span of knowledge interests necessarily demands a blend of evidence from research and clinical wisdom (Alan Pearson, Daphne Stannard, et al., 2012).

As such, we define evidence-based healthcare as clinical decision-making that considers the feasibility, appropriateness, meaningfulness and effectiveness of healthcare practices. The feasibility, appropriateness, meaningfulness and effectiveness of healthcare practices may be informed by the best available evidence, the context in which the care is delivered, the individual patient, and the professional judgment and expertise of the health professional (Jordan, Lockwood, Munn, & Aromataris, 2019).

1.1. Systematic review

More and more healthcare policy is being based on clear and comprehensive summaries of information collated through systematic reviews of the relevant literature (Khan, Kunz R., Kleijnen J., & Antes, 2011). The purpose of a systematic review is to evaluate and interpret all available evidence relevant to a particular question (Glaszious, Irwing, Bain, & Colditz, 2001).

A number of terms are used concurrently to describe the process of systematically reviewing and integrating research evidence, including “systematic review”, “meta-analysis”, research synthesis”, “overview”, or “pooling” (Egger, Smith, & Altman, 2007).

Based on that, we can define a systematic review (hereinafter SR) as a summary of the results of available carefully designed healthcare studies (controlled trials) that provides a high level of evidence on the effectiveness of healthcare interventions (Higgins & Green, 2008). SR can be broadly defined as a type of research synthesis that are conducted by reviewer groups with specialized skills, who set out to identify and retrieve international evidence that is relevant to a particular question or questions and to appraise and synthesize the results of this search to inform practice, policy, and in some cases, further research (Aromataris & Pearson, 2014) (Z. Munn et al., 2018)

Khan, Kunz and Antes (2003) define SR as a research article that identifies relevant studies, appraises their quality and summarizes their results using scientific methodology (Khan et al., 2011). According to the Cochrane Book, SR use explicit, systematic methods that are selected with a view to minimizing bias, thus providing more reliable findings from which conclusions can be drawn and decisions made (Higgins & Green, 2008). SR creates new knowledge from existing science and research, so it is named as secondary research (Evans & Pearson, 2001) (Glaszious et al., 2001) (Klugar, 2015).

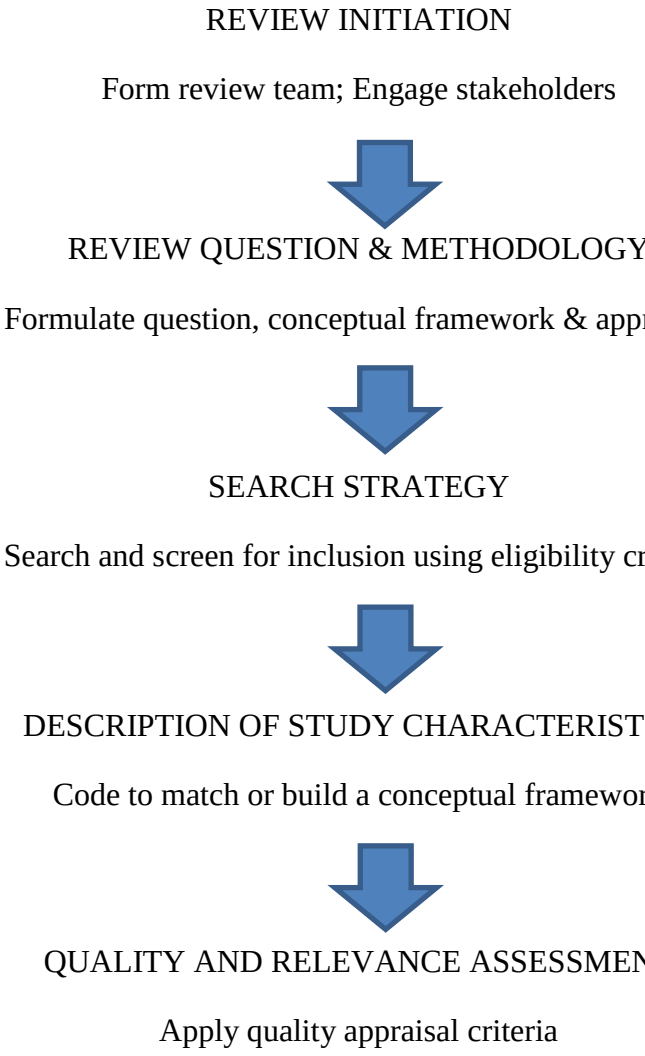
SR may be undertaken to confirm or refute whether or not current practice is based on relevant evidence, to establish the quality of that evidence, and to address any uncertainty or variation

in practice that may be occurring. SR may also identify gaps, deficiencies, and trends in the current evidence and can help underpin and inform future research in certain area. And finally, SR can be used to produce statements to guide clinical decision-making, the delivery of care, as well as policy development (Aromataris & Pearson, 2014).

The key characteristics of a systematic review are:

- A clearly stated set of objectives with pre-defined eligibility criteria for studies;
- An explicit, reproducible methodology;
- A systematic search that attempts to identify all studies that would meet the eligibility criteria;
- An assessment of the validity of the findings of the included studies, for example through the assessment of risk of bias; and
- A systematic presentation, and synthesis, of the characteristics and findings if the included studies (Higgins & Green, 2008).

Figure 2: Common stages in SR[4]



[4] (Gough, Oliver, & Thomas, 2017)

SYNTHESIS

Use conceptual framework, study codes & quality judgement

USING REVIEWS

Interpret & communicate findings with stakeholders

As it is shown, the issues of the comprehensiveness of the scientific evidence that is identified, the quality of each component of the studies included (excluded) and the general proof of evidence is made explicit.

Answering each type of question requires different study designs, and consequently different methods of systematic review. A thorough understanding of appropriate study types for each question is vital and will greatly assist the processes of findings, appraising and synthetizing studies from the literature (Glaszious et al., 2001). Based on the JBI Reviewer´s Manual, currently JBI has formal guidance for the following types of reviews:

- Systematic reviews of experience or meaningfulness;
- Systematic reviews of effectiveness;
- Systematic reviews of text and opinion/policy;
- Systematic reviews of prevalence and incidence;
- Systematic reviews of costs of a certain intervention process, or procedure;
- Systematic reviews of aetiology and risk;
- Systematic reviews of mixed methods;
- Systematic reviews of diagnostic test accuracy;
- Umbrella reviews;
- Scoping reviews (JBI, 2019)[5]

The SR represents the highest level of scientific evidence, especially if so called a meta-analysis it is carried out. The meta-analysis means to use the statistical methods to sum up the results of independent studies. Using of combining of all information from all relevant studies, the meta-analysis can provide an exact estimation of health care effects (Tučková & Klugar, 2015).

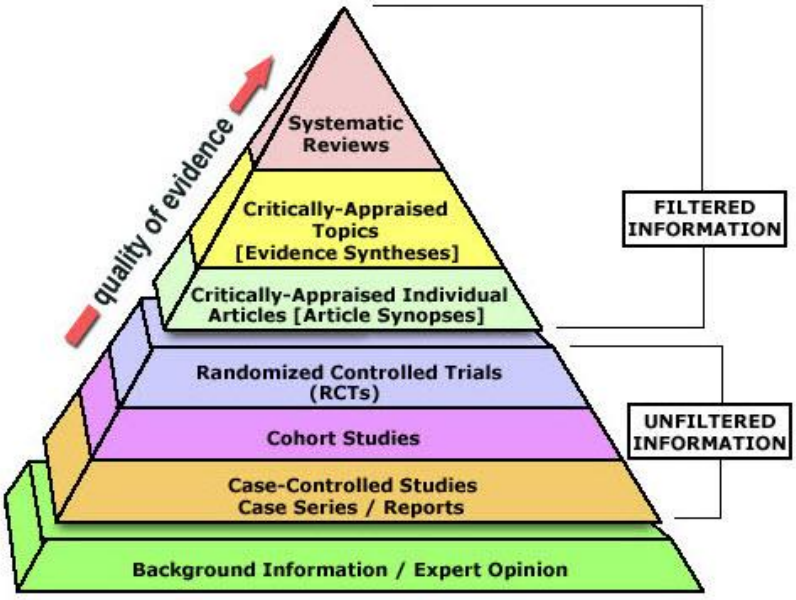
1.2. Study design and level of evidence

Based on the fact how the study is done we talk about a study design. Design of the study determines the validity of the observed effects, i. e. our confidence that results of a study are

[5] Compare to Glasziou, Irwing, Bain, Colditz´s (2001) table of clinical and public health questions, ideal study types and major appraisal issues that listed: Intervention, frequency/rate (burden of illness), aetiology and risk, prediction and prognosis, diagnostic accuracy and phenomena (Glaszious et al., 2001)

likely to approximate to the "truth" for the participants or patients' students depends on the soundness of its design. Ultimately the strength of review´s inferences depends on the integrity of the designs of the available studies (Khan et al., 2011). The way how to organize the different types of evidence level is an evidence pyramid (see Figure 3). This pyramid includes a variety of evidence types and levels. The levels of evidence pyramid provide a way to visualize both the quality of evidence and the amount of evidence available[6,7].

Figure 3:Level of evidence pyramid



Oxford Centre for Evidence-based Medicine sets out one approach to systematising this process for different question types in a following table 2:

Table: 2 Level of evidence[8]

Level	Therapy / Prevention, Aetiology / Harm	Prognosis	Diagnosis	Differential diagnosis / symptom prevalence study	Economic and decision analyses
1a	SR (with homogeneity*) of RCTs	SR (with homogeneity*) of inception cohort studies; CDR" validated in different populations	SR (with homogeneity*) of Level 1 diagnostic studies; CDR" with 1b studies from different clinical centres	SR (with homogeneity*) of prospective cohort studies	SR (with homogeneity*) of Level 1 economic studies

[6] EBM Pyramid and EBM Page Generator, copyright 2006 Trustees of Dartmouth College and Yale University. Produced by Jan Glover, David Izzo, Karen Odato and Lei Wang.
[7] Filtered resources appraise the quality of studies and often make a recommendation for practice (Walden University, 2020)
[8] (available from https://www.cebm.net/2009/06/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/)[8] Produced by Bob Phillips, Chris Ball, Dave Sackett, Doug Badenoch, Sharon Straus, Brian Haynes, Martin Dawes since November 1998. Updated by Jeremy Howick March 2009. (Phillips et al., 1998)

1b	Individual RCT (with narrow Confidence Interval"¡)	Individual inception cohort study with > 80% follow-up; CDR" validated in a single population	Validating** cohort study with good" " " reference standards; or CDR" tested within one clinical centre	Prospective cohort study with good follow-up****	Analysis based on clinically sensible costs or alternatives; systematic review(s) of the evidence; and including multi-way sensitivity analyses
1c	All or none§	All or none case-series	Absolute SpPins and SnNouts" "	All or none case-series	Absolute better-value or worse-value analyses " " " "
2a	SR (with homogeneity*) of cohort studies	SR (with homogeneity*) of either retrospective cohort studies or untreated control groups in RCTs	SR (with homogeneity*) of Level >2 diagnostic studies	SR (with homogeneity*) of 2b and better studies	SR (with homogeneity*) of Level >2 economic studies
2b	Individual cohort study (including low quality RCT; e.g., <80% follow-up)	Retrospective cohort study or follow-up of untreated control patients in an RCT; Derivation of CDR" or validated on split-sample§§§ only	Exploratory** cohort study with good" " " reference standards; CDR" after derivation, or validated only on split-sample§§§ or databases	Retrospective cohort study, or poor follow-up	Analysis based on clinically sensible costs or alternatives; limited review(s) of the evidence, or single studies; and including multi-way sensitivity analyses
2c	"Outcomes" Research; Ecological studies	"Outcomes" Research		Ecological studies	Audit or outcomes research
3a	SR (with homogeneity*) of case-control studies		SR (with homogeneity*) of 3b and better studies	SR (with homogeneity*) of 3b and better studies	SR (with homogeneity*) of 3b and better studies
3b	Individual Case-Control Study		Non-consecutive study; or without consistently applied reference standards	Non-consecutive cohort study, or very limited population	Analysis based on limited alternatives or costs, poor quality estimates of data, but including sensitivity analyses incorporating clinically sensible variations.
4	Case-series (and poor quality cohort and case-control studies§§)	Case-series (and poor quality prognostic cohort studies***)	Case-control study, poor or non-independent reference standard	Case-series or superseded reference standards	Analysis with no sensitivity analysis
5	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on economic theory or "first principles"

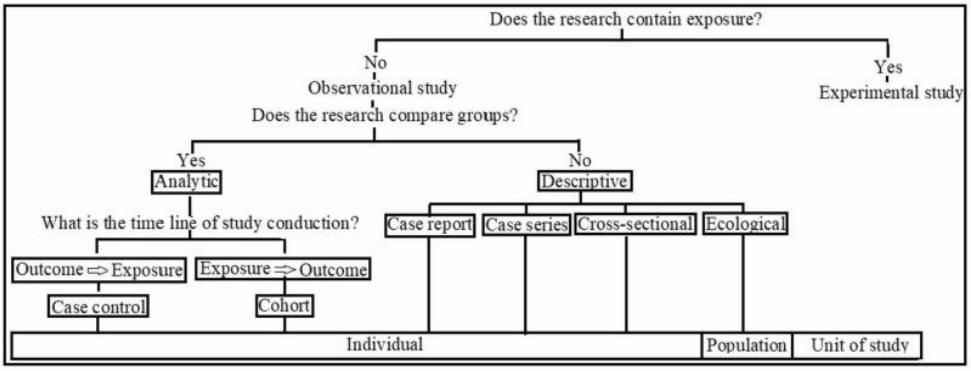
There are various levels or hierarchies of evidence; which level is appropriate depends upon the type of clinical question being asked. For intervention questions, the level of evidence ranks quantitative research designs (e.g. systematic reviews of randomized controlled trials) as providing levels of confidence that the studies will have reliable answer to these questions than designs with lower levels of confidence (e.g. descriptive studies) (Haynes, 2006; Melnyk & Fineout-Overholt, 2011). However, it should be noted that Heynes pyramid and reliance on study designs alone is obsolete. This use has been overcome by the GRADE WG, which still honors that some designs are more robust to systematic errors, but is aware of the fact that even a perfectly designed RCT can have many other problems and errors. Therefore, GRADE comes up with a certainty of evidence evaluation where robustness (based on designs) is only one of 8 factors influencing our certainty in evidence / studies

1. **Experimental study** = a comparative study in which the use of different intervention among participants is allocated by the researcher (Khan et al., 2011). They are designed to find new and more effective ways to diagnose and treat people with disease and they are more commonly associated with a drug therapy or diagnostic tests (Alan Pearson, Heather Loveday, et al., 2012).
    - Randomized controlled trials (hereinafter RCT) (with concealed allocation)
      The RCT is considered the gold standard for testing the efficacy of therapeutic interventions, particularly drugs (Alan Pearson, Heather Loveday, et al., 2012). Therefore the RCT is the most stringent way of determining whether a cause-effect relation exists between the intervention and the outcome (Sibbald & Roland, 1998). There are many different types of RCT and the design takes account of the order in which participants are exposed to the intervention, whether the investigator and/or participants are aware of the intervention, and whether the trial is attempting to demonstrate efficacy versus effectiveness[9] or superiority versus equivalence (Alan Pearson, Heather Loveday, et al., 2012). Randomization (with concealment of allocation sequence from caregivers) avoids bias because both known and unknown determinants of outcome, apart from the intervention, are usually equally distributed between the two groups of participants (Khan et al., 2011). The most commonly used RCT designs are parallel, crossover, factorial, and cluster designs (Alan Pearson, Heather Loveday, et al., 2012).
    - Experimental study without randomization (sometimes called quasi-experimental or quasi-randomized or pseudo-randomized studies)
      These studies encompass a broad range of non-randomized intervention studies and they are used when it is not logistically feasible or ethical to conduct the RCT (Harris et al., 2006). The allocation of participants to different interventions is managed by the researcher but the method of allocation falls short of genuine randomization, e.g.

[9] Effectiveness is the extent to which an intervention produces beneficial outcomes under ordinary day-to-day circumstances (Khan et al., 2011).
Efficacy is the ability of an intervention to produce the desired beneficial effect (Kim, 2013).
Efficiency (cost-effectiveness) is the extent to which the balance between input and output of intervention represents value for money (Khan et al., 2011).

alternate or even-odd allocation. Such methods fail to conceal the allocation sequence from caregivers (Khan et al., 2011).

Figure 4:Classification of observational study designs[10]



2. **Observational/analytic study (non-randomized studies)** = an experimental studies in which people are allocated to different interventions using methods that are not random (Higgins & Green, 2008). The most common uses of observational designs are to investigate the natural course of a disease or health risk or to observe changes before and after an event or intervention (Alan Pearson, Heather Loveday, et al., 2012). There are many types of non-randomized interventions, including cohort studies, case-control studies, controlled before-and-after studies, interrupted-time-series study and controlled trials that use inappropriate randomization strategies (Higgins & Green, 2008).

- Cohort studies

  A cohort study is the best method for tracking the natural history and incidence of a disease and may be retrospective or prospective (Alan Pearson, Heather Loveday, et al., 2012). A defined group of participants (the cohort) is followed over time, to examine associations between different interventions received and subsequent outcomes (Higgins & Green, 2008). But the cohort designs are not feasible where the disease incidence is rare or the latency to disease is long. Failure to follow-up a large number of study subjects likely introduces selection bias; for example, subjects with better or worse outcomes may be more likely to be followed up than others (differential loss in follow-up) (Wang & Attia, 2010).

- Case-control study

  A case-control study is usually retrospective and aim to assess whether a historical exposure to a risk factor in people with a disease is comparable to that of people who do not have the disease (Alan Pearson, Heather Loveday, et al., 2012). This type of study compares people with a specific outcome of interest ("cases") with people from the same sources' population but without that outcome ("control"), to examine the association between the outcome and prior exposure (e.g. having the intervention). This design is particularly useful when the outcome is rare (Higgins & Green, 2008).

[10] (Rezigalla, 2020)

The advantage of this design is also its biggest drawback: in assessing exposures retrospectively, cases may overreport exposures relative to controls (recall bias). Where and how to select the appropriate control group for a series of cases also may affect the study findings (potential selection bias) (Wang & Attia, 2010).

- Controlled before-and-after study

  The controlled before-and-after study design offers better evidence about intervention effectiveness than the other non-experimental designs. They are most useful in demonstrating the immediate impacts of short-term programs (Robson, Shannon, Goldenhar, & Hale, 2001). The observation is made before and after the implementation of an intervention, both in a group that receives the intervention and in a control group that does not (Higgins & Green, 2008).

- Interrupted-time-series study

  The interrupted-time-series study is a useful design with which to evaluate the longitudinal effects of interventions, through regression modelling. It is principally a tool for analysing observational data where full randomization, or case-control design is not affordable or possible (Kontopantelis, Doran, Springate, Buchan, & Reeves, 2015). In this type of study, multiple time points before and after the intervention (the "interruption") are observed. The design attempts to detect whether the intervention has had an effect significantly greater than any underlying trend over time (Higgins & Green, 2008).

3. **Descriptive study** = the simplest design describing the distribution of one or more variables, without regard to any causal or other hypothesis. It includes several types of studies, namely, cross-sectional studies, case series, case reports, and ecological studies (Aggarwal & Ranganathan, 2019). Main advantage of descriptive studies is that they are relatively inexpensive to conduct and other opportunity to identify associations that might then be explored using a controlled observational or experimental design. Their disadvantages are that samples are often self-selecting, responses rates may be low, and the results often have a number of plausible explanations making it difficult to infer causation (Alan Pearson, Heather Loveday, et al., 2012).

   - Cross-sectional studies

   The cross-sectional studies are used for a range of research questions including investigating the prevalence (frequency) of a particular condition at a point in time (Alan Pearson, Heather Loveday, et al., 2012). They collect information on interventions (past or present) and current health outcomes. They are very good for measuring the prevalence of a disease or of a risk factor in a population. Thus, these are very helpful in assessing the disease burden and healthcare needs. Sometimes, cross-sectional studies are repeated after a time interval in the same population (using the same subjects as were included in the initial study, or a fresh sample) to identify temporal trends in the occurrence of one or more

variables, and to determine the incidence of a disease (i.e., number of new cases) or its natural history (Aggarwal & Ranganathan, 2019).

- Case series and case reports (uncontrolled longitudinal study)

  The case series and case reports describe a number of individual cases of a disease or responses to an intervention without comparison to a control group (Alan Pearson, Daphne Stannard, et al., 2012) and refer to the description of a patient with an unusual disease or with simultaneous occurrence of more than one condition (Aggarwal & Ranganathan, 2019).

- Ecological studies

Ecological (also sometimes called as correlational) study design involves looking for association between an exposure and an outcome across populations rather than in individuals (Aggarwal & Ranganathan, 2019). In ecological studies, health outcomes are aggregates of individual health data, e.g. prevalence, incidence, rate of disease. Types of ecological studies are geographical, longitudinal, migration (Levin, 2006).

1.3. GRADE methodology

In the early 2000s, the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group developed a framework in which the certainty in evidence was based on numerous factors and not solely on study design which challenges the pyramid concept (G. H. Guyatt et al., 2008). The GRADE assessment system is a trustworthy and sensible approach which aims to transform guidance into practice. In the past, there have been many different systems that tried to address the challenge of writing the evidence and grading the recommendations. As a response to this increasing confusion, a working group was created in the year 2000 to develop a unified standard and sensible approach for a guideline development. The GRADE specifies an approach to framing questions, choosing outcomes of interest and rating their importance, evaluating the evidence, and incorporating evidence with considerations of values and preferences of patients and society to arrive at recommendations (G. Guyatt et al., 2011). The GRADE system can be use in a various range of clinical questions including diagnosis, screening, prevention, and therapy, and can be applied in field of public health and health systems questions. The advantages of the GRADE over other systems are:

- Development by a widely representative group of international guideline developers;
- Clear separation between quality of evidence and strength of recommendations;
- Explicit evaluation of the importance of outcomes of alternative management strategies;
- Explicit, comprehensive criteria for downgrading and upgrading quality of evidence ratings;
- Transparent process of moving from evidence to recommendations;
- Explicit acknowledgement of values and preferences;

- Clear, pragmatic interpretation of strong versus weak recommendations for clinicians, patients, and policy makers;
- Useful for systematic reviews and health technology assessment, as well as guidelines (G. H. Guyatt et al., 2008), (Langendam et al., 2013b) .

The GRADE approach specifies four levels of quality. The highest quality rating is for randomized trial evidence although even randomized trial evidence can have moderate, low or even very low quality of evidence. In the table 3, you can see the Levels of quality of a body of evidence in the GRADE approach (Higgins & Green, 2008)

Table: 3 Level of quality of a body of evidence in the GRADE approach

Underlying methodology	Quality rating
Randomized trials; or double-upgraded observational studies	High
Downgraded randomized trials; or upgraded observational studies	Moderate
Double-downgraded randomized trials; or observational studies	Low
Triple-downgraded randomized trials; or downgraded observational studies, or case series/case reports	Very low

The GRADE approach to rating the quality of evidence begins with the study designs and then address five reasons to possibly rate down the quality of the evidence (see Table 4: Factors that can reduce the quality of the evidence)[11] and three to possibly rate up the quality (see Table 5: Factors that can increase the quality of the evidence) (Langendam et al., 2013b) .

Table: 4 Factors that can reduce the quality of the evidence

Factor	**Consequence**
Limitation in study design or execution (risk of bias)	↓ 1 or 2 levels
Inconsistency of results	↓ 1 or 2 levels
Indirectness of the evidence	↓ 1 or 2 levels
Imprecision	↓ 1 or 2 levels
Publication bias	↓ 1 or 2 levels

[11] Compare with Higgins and Green (2014) Factors that may decrease the quality level of a body of evidence (Table 12.2 B, p 362) (Higgins & Green, 2008)

Table: 5 Factors that can increase the quality of the evidence

Factor	Consequence
Large magnitude of effect	↑ 1 or 2 levels
All plausible confounding would reduce the demonstrated effect or increase the effect if no effect was observed	↑ 1 level
Dose-response gradient	↑ 1 level

The GRADE approach uses two grades recommendations: "strong" and "weak". For guideline panel or others making recommendations to offer a strong recommendation they must be certain about the various factors that influence the strength of a recommendation. When a guideline panel is uncertain whether the balance is clear or when the relevant information about the various factors that influence the strength of a recommendation is not available, a guideline panel should be more cautious and in most instances it would opt to make a weak recommendation (Langendam et al., 2013b).

The GRADE provides a framework guiding through the critical components of the assessment in a structured way. By allowing to make the judgments explicit rather than implicit it ensures transparency and a clear basis for discussion (Langendam et al., 2013b) . The GRADE system is now a gold standard in guideline development and has been adopted by hundreds of organizations including Cochrane, WHO, NICE or JBI. Although it has its own limitations the recommendations that it provides are definite, clear, comprehensive, and pragmatics and they are followed by organizations all over the world engaging in systematic reviews development, health technology assessment or guidelines development.

For SR DTA, the use of the summary of finding table is more relevant, which is the phase in which the use of GRADE in the creation of a systematic review ends. The creation of recommendations is already a complex contextualized decision-making of guideline panels, which, based on the summary of finding, form recommendations. This approach has long been required by world leaders in EMB: Cochrane collaboration, JBI and Campbell. This requirement is also incorporated in the new PRISMA (Page et al., 2021).

1.3.1. Summary of findings

Systematic reviews of diagnostic studies should be accompanied by a summary of findings table, which should include the question being investigated, the index test, the reference test, the population, the estimates rate of true positives, false negatives, true negatives and false positives and the absolute difference between the index and reference tests for these values per 1000 patients, the sample size as well as the number of studies which contributed to the sample, the GRADE (Grading of Recommendations Assessment, Development and Evaluation) quality of evidence for each finding, and any comments (including decisions as to why the reviewers assigned the final GRADE ranking) (Campbell, Klugar, Ding, Carmody, Hakonsen, & Jadotte, 2015; Langendam et al., 2013a). Some reviews may include more than one 'Summary of findings' table, for example if the review addresses more than one major comparison, or

includes substantially different populations that require separate tables (e.g. because the effects differ or it is important to show results separately) (H. J. Schünemann, 2019).

1.3.2. Application of GRADE in systematic reviews of diagnostic test accuracy

The GRADE approach to grading the quality of evidence and strength of recommendations for diagnostic tests provides a comprehensive and transparent approach for developing these recommendations. Evidence from accuracy studies can be sufficient to make strong inferences about patient-important outcomes, when clinicians already have evidence from randomized trials showing that management of patients detected by a diagnostic test improves patient outcomes. However, this approach requires a clear understanding of the proposed place of a new test in a diagnostic pathway and its suggested benefits, as well as careful consideration of whether the patients detected by a new test are representative of the patients included in management trials (Lord, Irwig, & Simes, 2006). According to the GRADE approach, results that are TP, TN, FP, FN, or inconclusive, as well as the complications of a test and its cost (resource utilization) constitute the outcomes of a diagnostic accuracy study (Holger J Schünemann et al., 2008).

GRADE's four categories of certainty of evidence reflect a gradient of confidence in estimates of the effect of a diagnostic test strategy on patient-important outcomes. Study design and its quality of evidence from the trials or studies directly measuring patient-important outcomes is graded in the same way as for other interventions, and the initial grading based on study design can decrease because of other factors (Nasser & Fedorowicz, 2011). Therefore, the GRADE system provides additional quality criteria that can reduce the certainty of evidence about using diagnostic tests (limitation in study design and/or execution (risk of bias); indirectness of evidence; inconsistency of results; imprecision of results; publication bias) (Gopalakrishna et al., 2014).

Following the GRADE process, the overall certainty of evidence across outcomes is determined by the lowest grade of quality for any of the outcomes deemed critical (their importance for the decision was rated as 7, 8, or 9 on a 9-point scale) (Brożek et al., 2009). Therefore, the GRADE system provides a comprehensive and transparent approach for grading the quality of evidence and developing recommendations about diagnostic tests. Based on that, the GRADE approach asks guideline developers to make judgements about the relative importance of each outcome for making a recommendation explicit (Holger J Schünemann et al., 2007).

.

2. Diagnostic test accuracy systematic review

One of the types of systematic reviews that can be developed is diagnostic test accuracy systematic review (hereinafter DTA SR). Tests are used in everyday medical practice in all fields of medicine to screen for diagnose, grade, and monitor the progression of a certain disease. That is why this type of systematic review is the dynamically evolving part of evidence-based medicine. Diagnostic information is obtained from a multitude of sources, including imaging and biochemical technologies, histopathological, pathological and psychological investigations, laboratory and functional tests, and the signs and symptoms elicited during history-taking and clinical examinations. Each item of information obtained from these sources can be regarded as a result of a separate diagnostic or screening "test", whether it is obtained for the purpose of identifying diseases in sick people, or for detecting early disease in asymptomatic individuals (Egger et al., 2007). The condition of interest or target condition can refer to a particular disease or to any other identifiable condition that may prompt further clinical action; such as further diagnostic testing, or the initiation or cessation of treatment (White, Schultz, & Y., 2011). The main aim of the development of DTA SR is to produce estimates of performance based on all available evidence, to evaluate the quality of published studies, and to account for variation in findings between studies (Egger et al., 2007).

There exist two types of studies of the diagnostic test accuracy (hereinafter DTA). The first is the diagnostic case-control design, also sometimes called the "two gate design" where people having a certain disease come from another population than people without the disease. The second study design is cross-sectional, and involves all patients suspected to have the disease of interest who undergo both, reference and index test (Campbell, Klugar, Ding, Carmody, Hakonsen, & Jadotte, 2015). Therefore, the cross-sectional studies are preferred as a better/only evidence in DTA SR conclusion establishment, supported also by STARD (Standards for Reporting of Diagnostic Accuracy) reporting guidelines for primary diagnostic studies since 2003 (Bossuyt et al., 2003). However, the STARD initiative published the STARD statement in 2003. It was intended to help improve the transparency and completeness of reporting of diagnostic accuracy studies. STARD presented a checklist of 25 items that authors should address when reporting diagnostic accuracy studies (Korevaar et al., 2016).

2.1. Diagnostic test accuracy

Primary studies that examine the test performance are referred to as DTA studies and these studies compare a "new" test (or tests) to the best test (or method) that is currently available (White et al., 2011). The most commonly used measures of accuracy are the sensitivity and specificity of the test but other measures (e.g. predictive values, odds-ratio, receiver operating characteristic (ROC) curves) can also be used. Test accuracy is not a fixed property of a test. It can vary between patient subgroups, with their spectrum of disease, with the clinical setting, with the test interpreters, and may depend on the results of prior testing (MG, Deeks, Gatsonis, & Bossuyt, 2008).

The accuracy of index test is reported relative to the reference test in terms of sensitivity and specificity. Sensitivity is the probability that a person with the disease of interest will have a positive result, while specificity is the probability of a person without the condition having negative result (Campbell, Klugar, Ding, Carmody, Hakonsen, & Jadotte, 2015); in other words: people who receive a positive index test result but a negative reference test result are classifies as being "false positive" and those people who receive a negative index result and a positive reference test result are considered to be "false negative" due to the disagreement between the test results (White et al., 2011) (see table 6).

Table: 6 Description of patient classification for DTA studies[12]

Patient classification	Description of test results
True positive	Positive index test result Positive reference test result
True negative	Negative index test result Negative reference test result
False positive	Positive index test result Negative reference test result
False negative	Negative index test result Positive reference test result

The table can be displayed as 2x2 typical table of Description of patient classification for DTA studies (White et al., 2011) (Egger et al., 2007) (Campbell, Klugar, Ding, Carmody, Hakonsen, & Jadotte, 2015) (see Table 7)

Table: 7 Description of patient classification for DTA studies in 2x2 table

Index test Outcome	Reference positive	Reference negative	Total
Index test positive (T+)	True positives (TP)	False positives (FP)	Test positives (TP+FP)
Index test negative (T-)	False negatives (FN)	True negatives (TN)	Test negatives (FN+TN)
Total	Reference positives (TP+FN)	Reference negatives (FP+TN)	N (TP+FP+FN+TP)

Based on the tables 7 and 8, we can define a formula for calculating sensitivity as:

$$Sensitivity = \frac{True\ positives}{(True\ positives + False\ negatives)}$$

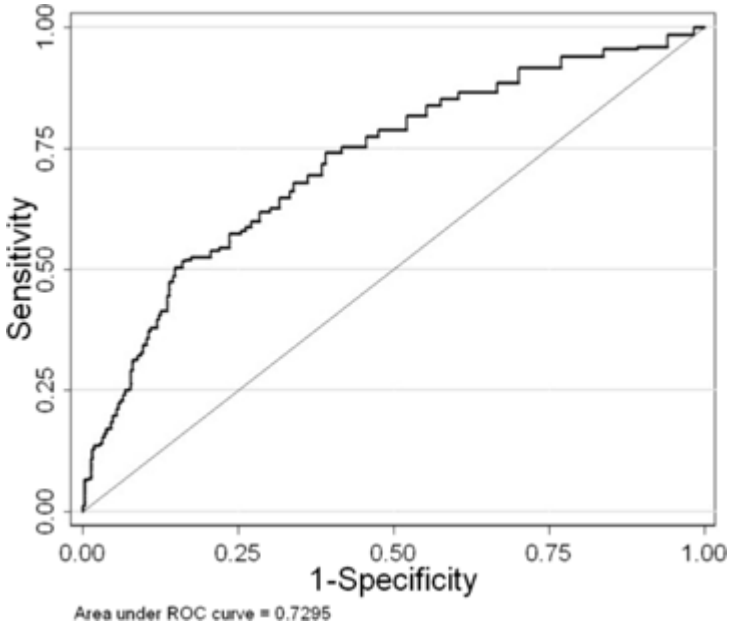
[12] (White et al., 2011)

Based on the tables 7 and 8, we can define a formula for calculating specificity as:

$$Specificity = \frac{True\ negatives}{(True\ negatives + False\ positives)}$$

A measure of test accuracy that brings together sensitivity and specificity is the diagnostic odds ratio, which is the ration of the odds of disease in test positives relative to the odds of disease in test negatives. Sensitivity and specificity have been identified as essential measures of diagnostic accuracy (Campbell, Klugar, Ding, Carmody, Hakonsen, Jadotte, et al., 2015), however, it depends on the level that has been chosen as the cut-off point for normal or abnormal. A different threshold can provide a different sensitivities and specificities. When several thresholds have been produced for a single set of data the diagnostic characteristics of the test can be illustrated graphically using a graph known as a receiver operating characteristic (ROC) curve of the true positives rate (sensitivity) against the false positive rate (1 – specificity) (Egger et al., 2007). The ROC curve is very common for evaluating the performance of diagnostic test that classify individual categories of those with and without a condition (Campbell, Klugar, Ding, Carmody, Hakonsen, Jadotte, et al., 2015).

Figure 5: A typical ROC curve[13]



The ROC analysis is used to plot the sensitivity (y-axis) against 1-specificity (x-axis) as the threshold values changes. This gives a visual representation of the relationship between sensitivity and specificity of a diagnostic test as the threshold value changes. This can be

[13] (Akobeng, 2007)

measured quantitatively by assessing the area under the curve (AUC). The AUC curve is a reflection of how good the test is at distinguishing between patients with a condition and those without condition (Akobeng, 2007), the AUC for a perfect test is 1.0, and a test with no differentiation between disorder and no disorder has an AUC of 0.5 (Campbell, Klugar, Ding, Carmody, Hakonsen, & Jadotte, 2015). The AUC serves as a single measure, independent of prevalence, that summarizes the discriminative ability of a test across the full range of cut-offs (Akobeng, 2007).

As it was mentioned above, other measures of DTA include predictive values and likelihood ratios but they are not the essential concepts when conducting DTA SR.

3. Specifics of conducting of diagnostic test accuracy systematic review

Every SR should be developed based on a previous published protocol in order to support the unbiased inclusion of studies and reporting findings (Klugar, 2016). A systematic review protocol describes the rationale, hypothesis, and planned methods of the intended review. It should be prepared before a review is started and used as a guide to carry out the review. For a better guidance in a protocol development, Preferred Reporting Items for Systematic Reviews and Meta-analysis (hereinafter PRISMA) can be used. PRISMA is an evidence-based minimum set of items for reporting in systematic reviews and meta-analyses. PRISMA focuses on the reporting of reviews evaluating randomized trials, but can also be used as a basis for reporting systematic reviews of other types of research, particularly evaluations of interventions[14]. PRISMA-P for protocol development was published in 2015. Detailed protocols should be made publicly available, and registered in a registry such as PROSPERO (Moher et al., 2015), Campbell Collaboration or Cochrane Collaboration depending on what the systematic review is focused on (Moher et al., 2015).

The items that are included in the checklist for PRISMA-P are contained in Appendix 1 of the dissertation thesis (Moher et al., 2015).

3.1. Review question/objectives

Every DTA SR should be developed based on the review question or objective. It is an essential step to make to undertake the best quality systematic review. The concrete acronym is used for the development of review question or objectives. In DTA SR, the mnemonic PIRD is recommended for setting out the key components of the SR. In this acronym:

P – stands for POPULATION (all participants who will undergo the diagnostic test

I – stands for INDEX TEST(s) (the diagnostic test(s) whose accuracy is being investigated in the review)

R – stands for REFERENCE TEST(s) (the "gold standard" test to which the results of the index test will be compared)[15]

D – stands for DIAGNOSIS OF INTEREST (it relates to what diagnosis is being investigated in DTA SR – disease, injury, disability or any other pathological condition) (Zachary Munn et al., 2018).

[14] The definition of PRISMA is available on http://www.prisma-statement.org/ [cited 09-04-2020]

[15] It is necessary to consider if multiple iterations of a test exist and who carries out or interprets the test, the conditions the test is conducted under and specific details regarding how the test will be conducted. It should be the best currently available for the diagnosis of the condition of interest (Munn, Stern, Aromataris, Lockwood, & Jordan, 2018)

This PIRD mnemonic will serve as a guide for the evaluation of studies to be included (or excluded) in a systematic review and helps in the development of sensitive (find as many studies as possible), minimize bias and be efficient search strategy).

3.2. Search strategy

Development of a comprehensive search strategy is one of the most important key elements of the scientific validity of DTA SR. The comprehensive search and complete identification of studies /papers has the primary importance in conducting DTA SR that includes detection of published and unpublished (grey literature) data. Based on the JBI methodology, the standard procedure is the development of three-step search strategy of an initial limited search to identify relevant keywords and indexing terms done in a major databases (such as MEDLINE or CINAHL), followed by a second thorough search across all included databases (general and specific subject databases are listed in Appendix 2), and then final review of the relevance lists of included studies (Campbell, Klugar, Ding, Carmody, Hakonsen, Jadotte, et al., 2015).

The three-step strategy according the JBI methodology includes:

Step 1: Identification of keywords and search terms – the aim is to locate some papers relevant to the DTA SR and determine whether those papers can provide any additional key word, index terms, or subject headings that may help in search of similar studies/papers.

Step 2: Conducting the search across the specified databases – the aim is to construct database-specific searches for each database included in the protocol of DTA SR.

Step 3: Reference list searching – the aim is to search in the reference lists of all studies included in the DTA SR for detecting additional studies (Campbell, Klugar, Ding, Carmody, Hakonsen, & Jadotte, 2015).

Except of published sources of a literature included to the DTA SR, we have to consider the sources of grey literature if we want to achieve a comprehensive systematic review search. To incorporate grey literature can be done in two ways: (1) as included items in these reviews and (2) as a means to identify relevant studies and publications for these projects (Godin, Stapleton, Kirkpatrick, Hanning, & Leatherdale, 2015). Exclusion or lack of the sources of grey literature can give the effect of publication bias and it may artificially amplify estimates of treatment effects, given the effects of publication bias (Hopewell, McDonald, Clarke, & Egger, 2007). Sources of grey literature include: thesis, dissertations, reports, blogs, technical notes, non-independent research, governmental documents standards, recommendations etc. (sources of grey literature are listed in Appendix 2).

3.3. Assessment of methodological quality (critical appraisal)

The quality of diagnostic studies is determined by the extent to which biases have been avoided (Glaszious et al., 2001) and the methods by which the study sample is recruited, the conduct of tests involved, blinding in the process of interpreting tests, and the completeness of the study report (Campbell, Klugar, Ding, Carmody, Hakonsen, & Jadotte, 2015). But even high-quality study will not be applicable for a certain DTA SR if the exact test used differs from the one to which you have local access or the test has been evaluated in a tertiary care setting while you are interested in primary care usage (Glaszious et al., 2001). However, the process of critical appraisal examines the methodology of a study against predefined criteria, with the aim of considering individual sources of risk of bias (Campbell, Klugar, Ding, Carmody, Hakonsen, Jadotte, et al., 2015). In an attempt to improve the scientific rigor and completeness of reporting, The Cochrane Collaboration established the Standards for the Reporting of Diagnostic Accuracy (STARD) initiative as a way of assessing study quality (White et al., 2011). STARD contains a checklist of 30 essential items and a diagram (please see Appendix 5) that can be used by authors, reviewers and other readers, to ensure that a report of a diagnostic accuracy study contains the necessary information. This explanatory document aims to facilitate the use, understanding, and dissemination of the checklist. The document contains a clarification of the meaning, rationale, and optimal use of each item on the checklist, as well as a short summary of the available evidence on bias and applicability. The STARD statement, checklist, flowchart, and this explanation and elaboration document should be useful resources to improve reporting of diagnostic accuracy studies. Complete and informative reporting can only lead to better decisions in health care (Bossuyt et al., 2003). The checklist published by the STARD research group is just one such instrument aimed at primary care. But one of the most used tools for examining diagnostic accuracy in tests is QUADAS (Quality Assessment of Diagnostic Accuracy Studies) tool which was released in 2011 (Willis & Quigley, 2011). QUADAS is a tool to assess the quality of diagnostic accuracy studies included in a systematic review, and a measurement, implying that its characteristics have to be evaluated (Whiting, Rutjes, et al., 2011). It consists of four key domains (see Table 8; the official document of tool recommended as a critical appraisal checklist for diagnostic studies by the JBI please see Appendix 3) covering patient selection (which addresses the risk of selection bias created by how patients were selected for the study), index tests (which addresses the risk of bias created by how the index test was conducted and interpreted), reference standard (which investigates the same for the reference test), and flow of patients through the study and timing of the index test(s) and reference standard ("flow and timing", which investigates the risk of bias attributable to the order in which the index and reference tests were conducted in the study).

Table: 8 QUADAS 2 signaling questions[16]

Critical appraisal questions
Domain 1: Patient selection
Was a consecutive or random sample of patients enrolled?
Was a case-control design avoided?
Did the study avoid inappropriate exclusions?
Domain 2: Index test
Were the index tests results interpreted without knowledge of the results of the reference standard?
If a threshold was used, was it prespecified?
Domain 3: Reference test
Is the reference standard likely to correctly classify the target condition?
Were the reference standard results interpreted without knowledge of the results of the index test?
Domain 4: Flow and timing
Was there an appropriate interval between index test and reference standard?
Did all patients receive the same reference standard?
Were all patients included in the analysis?

The critical appraisal questions are answer with the "yes", "no", "unclear" or "not applicable" options, where "yes" implying that the methodological feature is optimal; "no" meaning that the methodological feature is less than optimal with the potential of introducing bias or limiting its applicability (White et al., 2011). All studies included in the systematic review must be appraised by two independent reviewers. Any disagreement that arise between the reviewers must be resolved through discussion, or with a third reviewer. However, the DTA SR should aim either to exclude studies which do not meet the critical appraisal signaling questions and are susceptible to bias, or alternatively to include studies with a mixture of quality characteristics and explore the differences (Egger et al., 2007).

3.4. Data extraction

The aim of data extraction is to identify and extract relevant data which will be use in data synthesis. It is a process of sourcing and recording relevant results and details from the primary research studies (Campbell, Klugar, Ding, Carmody, Hakonsen, & Jadotte, 2015). There exist lots of standardized tools for data extraction which is appropriate to use (e.g. data extraction sheets from Cochrane Collaboration or JBI) (please see Appendix 4). Primary studies included to the DTA SR can have several outcomes but only the same type of data across all included studies, which are relevant to the review question/objectives, should be extracted. The decision threshold that was used to classify results as positive or negative is an item of data extraction unique to studies of diagnostic accuracy (Campbell, Klugar, Ding, Carmody, Hakonsen, Jadotte, et al., 2015). DTA SR are concerned with test results that can be presented in different formats as summarized in Table 11 (White et al., 2011).

[16] (Campbell, Klugar, Ding, Carmody, Hakonsen, Jadotte, et al., 2015)

Diagnostic test results are often defined on a continuous scale. A threshold is most occasionally defined below which test result could be negative or above which test result could be positive. That is the reason why all studies of DTA should be placed in a 2x2 table that classified patient test results and disease status (Campbell, Klugar, Ding, Carmody, Hakonsen, Jadotte, et al., 2015) and corresponds to STARD statement (please see Appendix 5). In this process, two different reviewers must extract data independently. If there any disagreement, the third reviewer must be involved. If it happens and some data are missing, the reviewers should contact the authors of the primary studies and ask them to provide missing data additionally.

3.5. Data synthesis

Data synthesis is a crucial part of DTA SR. Predictive values, likehood ratios, summary ROC, diagnostic odds ratio (DOR) and meta-regression are some approaches used in synthetizing diagnostic test accuracy studies depending on the initial relationships identified between sensitivity and specificity (White et al., 2011). Outcome data of the primary studies should be combined and reported via graphical representation, meta-analysis etc.

3.5.1. Graphical and tabular representation[17]

JBI uses two different major ways of graphical representation: a) forest plots (however, in order to present data on DTA, "paired" forest plots must be used – one for sensitivity, one for specificity); b) summary ROC (SROC) curves, which are graphs with 1-specificity on the *x*-axis and sensitivity on *y*-axes (Campbell, Klugar, Ding, Carmody, Hakonsen, Jadotte, et al., 2015)

A Cochrane review of DTA uses two main forms or graphical display, summary ROC plots and forest plots (Macaskill, Gatsonis, Deeks, Harbord, & Takwoingi, 2010).

To create these figures, the authors of DTA SR can use RevMan5 for each analysis that is specified or several other available software's.

3.5.1.1. Summary ROC plots

Summary ROC plots (hereinafter SROC) display the results of individual studies in ROC space, each study is plotted as a single sensitivity-specificity point (Macaskill et al., 2010). The simplest and most common used method for diagnostic meta-analysis is the Moses-Littenberg fixed effect method (Chappell, Raab, & Wardlaw, 2009). The method considers the relationship between the DOR and summary measure of diagnostic threshold, given by the product of the

[17] In Glasziou, Irwing, Bain and Colditz (2001), the graphical presentation is described as two types of plots: a) simple plot of sensitivity and specificity: it shows the sensitivity and specificity of each study with its confidence intervals (with the specificity for particular study shown alongside the sensitivity for that study); b) plot sensitivity against specificity: it is plot of sensitivity against specificity in ROC space, ideally showing the points as ovoids with an area proportional to the square root of the number of people on whom sensitivity and specificity have been calculated (Glaszious et al., 2001)

odds of true positive and the odds of false positive results (Egger et al., 2007). The threshold can vary according to sample size and, to indicate more precisely the precision of the estimates, point height may differ from point width, with these being respectively proportional to the number of diseased and control patients (Campbell, Klugar, Ding, Carmody, Hakonsen, Jadotte, et al., 2015).

3.5.1.2. Linked ROC plots

These plots are used in analysis of pairs of tests, where both tests have been evaluated in each study. The points are plotted as in a normal summary ROC plots, but the two estimates (one for each test) from each study are joined by a line (Macaskill et al., 2010)

3.5.1.3. Coupled forest plots

Coupled forest plots report the number of true positives and false negatives in diseased and true negatives and false positives in non-diseased participants in each study and the estimated sensitivity and specificity, together with confidence interval; they contain of two graphical sections: one depicting sensitivity, and one specificity (Macaskill et al., 2010).

3.5.2. A coupled forest plot (meta-analysis)

Meta-analysis of DTA studies is a method for increasing the level of validity by combining data from multiple studies. An analytic method used for this type of meta-analysis should estimate diagnostic accuracy with the least bias, incorporating various factors known to affect the results (Juneyoung Lee, Kim, Choi, Huh, & Park, 2015). The aim of the meta-analysis is to determine the magnitude of the effect of each primary study to obtain the total magnitude of the effect. The total magnitude of the effect is presented as point estimates and limits (Klugar, 2015). According to the JBI Reviewers´ Manual for DTA SR (2015), the authors of DTA SR need to define the kind of meta-analysis to perform. Questions to consider are:

- Should we estimate summary sensitivity and specificity?
- Should we compute a summary ROC curve?

This depends on the nature of the data available, and more exactly, whether the diagnostic threshold was the same across the selected primary studies. Inclusion of meta-analysis in a DTA SR is sufficient but not necessary. Whether or not meta-analysis should be conducted depends on a number of factors, chiefly the number and methodological quality of included primary studies and the heterogeneity of their findings of DTA as well as other features such as patients characteristics and methodologies (White et al., 2011).

3.5.2.1. Sensitivity analysis

Sensitivity analysis determines how different values of an independent variable affect a particular dependent variable under a given set of assumptions. In other words, sensitivity

analyses study how various sources of uncertainty in a mathematical model contribute to the model's overall uncertainty. This technique is used within specific boundaries that depend on one or more input variables (Saltelli et al., 2008). In diagnostic accuracy, we must distinguish between diagnostic sensitivity and diagnostic specificity. Diagnostic sensitivity refers to the patient population, while diagnostic specificity expresses the results of the method in relation to healthy individuals (Altman & Bland, 1994) However in practice, we always examine a mixed population, consisting of healthy and sick people, these two properties of the laboratory method can never be separated. In the following paragraphs, we will see how closely diagnostic sensitivity and specificity are related. Ideally, the laboratory method clearly separates the sick and healthy population - diagnostic sensitivity and specificity are both equal to 100%. However, this phenomenon is very rare. For most methods, the results of the population of healthy and sick individuals overlap to some extent: for some healthy people, the method gives false positive results, and for some patients we find false negative results. The decisive factor for the ratio of diagnostic sensitivity and specificity is the determination of such a test value from which the result will be considered positive (so-called "cut-off value", threshold) (Vickers, 2008).

Another important factor when assessing the sensitivity analysis is the existence of so called gold standard and its usage in research. The gold standard is the best single test (or a combination of tests) that is considered the current preferred method of diagnosing a particular disease (X). All other methods of diagnosing X, including any new test, need to be compared against this ′gold′ standard. The gold standard is different for different diseases (Parikh, Mathai, Parikh, Sekhar, & Thomas, 2008). When a cut-off point is used, sensitivity and specificity show an inverse relationship - as sensitivity increases, specificity decreases and vice versa. Estimation of the sensitivity and specificity requires the use of an appropriate unequivocal diagnostic method as a "gold standard". The selection of the appropriate level of sensitivity and specificity often depends upon the particular need. When screening for a disease or pathogen we require a reliable positive result with few false negatives and a reasonable number of false positives (within an economically justifiable level of rejection). This would require a test with a high sensitivity and reasonable specificity. On the other hand, if we need as few false positives as possible (e.g. to confirm a tentative diagnosis) a test with a high specificity and reasonable sensitivity is used. It is, however, important to note that the consequence of any diagnostic test with imperfect specificity (less than 100%) is that if a large number of tests are made on a single uninfected participant, there is a significant chance of finding a positive result (Fegan, 2000).

3.5.3. Narrative synthesis (synthesis without meta-analys)

Narrative syntheses refers to an approach to the systematic review and synthesis of findings from multiple studies that relies primarily on the use of words and text to summarise and explain the findings of the synthesis. Whilst narrative synthesis can involve the manipulation of statistical data, the defining characteristic is that it adopts a textual approach to the process of synthesis to ‘tell the story’ of the findings from the included studies. As used here ‘narrative synthesis’ refers to a process of synthesis that can be used in systematic reviews focusing on a wide range of questions, not only those relating to the effectiveness of a particular intervention

(Popay et al., 2006). A textual combination of data is often used when the included studies are dissimilar in terms of patients, methods, or data (JBI, 2019).

Four main domains were identified when conducting the marrative synthesis:

- Developing a theory of how the intervention works, why and for whom
- • Developing a preliminary synthesis of findings of included studies
- • Exploring relationships in the data
- • Assessing the robustness of the synthesis (Popay et al., 2006).

Narrative synthesis relies primarily on the use of words and text but tables are often included also to summarise and explain the findings of a synthesis proces. In narrative synthesis, we can textually describe the individual studies or we can textually describe the groups of studies.

- Textual descriptions of individual studies. Summaries of individual studies can be structured to provide details of the setting, participants, exposure, and outcomes, along with any other factors of interest (e.g. the income level of the users, age of users, previous experiences, attrition, length of follow-up, sample size);
- Textual descriptions of groups of studies. Based on relevant criteria (e.g. types of participants) included studies can be sub-grouped. Subsequently, commentaries summarizing key aspects of the studies in relation to the sub-group within which they were included are produced. In a final step, the scope, differences and similarities among studies are used to draw conclusions across the studies (Lucas, Baird, Arai, Law, & Roberts, 2007).

Where a narrative synthesis is undertaken to describe the included studies and their conclusions, it is important to discern how the evidence was weighted and whether conclusions were biased. It is recommended that the characteristics of the studies and the data extracted are emphasised and tables, graphs, and other diagrams are made use of to compare data (Lockwood & White, 2012). The narrative summary presents relevant data extracted from individual studies, as well as, where available, point estimates (a value that represents a best estimate of effects) and interval estimates.

3.5.4. Heterogeneity

We can define heterogeneity using the definition of Cooper (2009) who defines the heterogeneity as the extent to which observed effect sizes differ from one another. In meta-analysis, statistical tests allow for the assessment of whether the variability in observed effect size is greater than would be expected given chance (that is, sampling error alone) (Cooper, Hedges, & Valentine, 2019).

By heterogeneity in DTA SR, we mean variability in the properties of the included primary studies. Especially in DTA SR it is possible to find heterogeneity because between-study heterogeneity of DTA studies is generally larger than that of therapeutic/interventional studies.

It is mainly due to differences in study populations, procedures followed for carrying out tests, and the conditions or context of testing (Campbell, Klugar, Ding, Carmody, Hakonsen, Jadotte, et al., 2015). If there is evidence of a lack of heterogeneity in sensitivity and specificity across studies, two univariate meta-analyses for these measures using either fixed- or random-effects models could be considered. However, if sensitivity and specificity vary markedly and/or there is an evidence of a threshold effect between studies, summary points alone should be avoided, since the summary points such as summary sensitivity, specificity or DOR do not correctly reflect the variability between studies and may miss important information regarding heterogeneity between studies (Juneyoung Lee et al., 2015). Assessment of heterogeneity is a challenge in synthesis of diagnostic studies. There remain many discussions about interpretation of heterogeneity statistics and details. Subgroup analysis can be used to investigate potential sources of heterogeneity, however, when the extent and cause of heterogeneity cannot be explained, then narrative synthesis instead of meta-analysis should be conducted (Campbell, Klugar, Ding, Carmody, Hakonsen, Jadotte, et al., 2015).

3.6. Discussion of the results

The aim of this part of DTA SR is to discuss the results of the conducted SR and the limits of the primary studies which are included to the DTA SR. It is recommended to use flow chart conforming to the PRISMA as well as the DTA SR should be accompanied by a summary of findings table.

The following sections should be mentioned in the discussion:

- Summary of DTA SR results;
- problems related to the quality of research in the certain field of the research (e.g. insufficient indexation;
- Other issues of relevance;
- Implications for practice and research, including recommendation for the future;
- Potential limitations of the systematic review (such a narrow timeframe or other restrictions) (JBI, 2019).

The discussion does seek to establish a line of argument based on the findings regarding the comparison of diagnostic tests, or its impact on the diagnostic tools in the protocol. The discussion should also include a final overview of the results that address any limitations or issues arising from the results or conduct of the review (Campbell, Klugar, Ding, Carmody, Hakonsen, Jadotte, et al., 2015).

3.7. Implication for practice/Implication for research

Implications are typically two-pronged: implications for research or theory and implications for practice. Implications for practice involve discussing what the findings of DTA SR might influence the practice. Implications, like recommendations for further study, are some of the most important end components of DTA SR.

3.7.1. Implications for practice

Implications for a practice must be based on the documented results from the review findings; they are not just reviewer´s opinions. Where is the evidence of the DTA SR strong enough that should potentially influence the practice, appropriate recommendations should be made. These recommendations must be clear, concise and unambiguous (JBI, 2019). Grading of Recommendations Assessment, Development, and Evaluation (GRADE) offers a transparent and structured process for developing and presenting summaries of evidence and recommendations in health care (G. Guyatt et al., 2011). Guideline panels considering a diagnostic test should begin by clarifying its purpose. The purpose of a test under consideration may be for triage (to minimize use of an invasive or expensive test), replacement (of tests with greater burden, invasiveness, or cost), or add-on (to enhance diagnosis beyond existing tests) (Holger J Schünemann et al., 2008). Although a GRADE approach for diagnostic tests has been developed, providing guidance on how to translate accuracy data into a recommendation involving patient important outcomes requires much more consideration (Leeflang Mariska MG, Deeks, Takwoing Yemisi, & Petra, 2013).

3.7.2. Implications for research

Implications for further research follow from the results of DTA SR based on identified gaps, or on areas of weakness in the literature such as inappropriate tests used or methodological weakness. It can happen in some cases that a gap within the whole area will be discovered when conducting DTA SR (or generally any conducted SR), i.e. a SR will not find any relevant study that could be included in the SR during a process of SR development (Higgins & Green, 2008). Recommendations in this part must be clear, concise and unambiguous (JBI, 2019).

3.8. Final parts of DTA SR

Based on the JBI Reviewer´s Manual (JBI, 2019) the SR should have other parts to be complete. The parts are as follows:

- Conflicts of interest – a statement which either declares the absence of any conflicts of interest or which describes a specific or potential conflict of interest.
- Acknowledgements – a statement which mentions sources of external funding or the contribution of colleagues or institutions.
- References – a list with all reference in one of a full referencing style.
- Appendices – the parts with the critical appraisal and data extraction tools appended as appendices. These tools must match the criteria specified in the Inclusion Criteria and Critical Appraisal section.

All the parts mentioned in Chapter 3 Specifics of conducting of diagnostic test accuracy systematic review are necessary for the development of full DTA SR. To conduct DTA SR is an activity that requires a huge amount of work. An analysis of 37 meta-analysis done by Allen

and Olkin (1999) showed that the average hours for a review were 1 139 (median 1110) – or about 30 person-weeks of full-time work – but this ranged from 2016 to 2518 hours (Glaszious et al., 2001). Therefore, a development of DTA SR is time consuming and very challenging. But on the other hand, conducting a comprehensive DTA SR has a potential to be informative not only for researchers but it has relatively likely to have a tangible and substantive impact on policy and practice.

3.9. Conclusion of theoretical part

The field of evidence-based medicine is a dynamic and rapidly evolving field of medical research and medical practice. A well-asked clinical question, and especially a well-answered clinical question, can help healthcare professionals in the decision-making process, which concerns the most important thing that stands in the center of medicine, and that is the patient's health, protection, health improvement, prevention and all aspects related to patient's health. International worldwide organizations focusing on evidence-based medicine and evidence-based healthcare such as the Joanna Briggs Institute, Cochrane Collaboration, Campbell, etc. are leaders in this field. These organizations continue to develop and update procedures, tools and software that improve and make available the search for scientific evidence to all healthcare professionals who want to use this knowledge in their profession.

Therefore, asking a clinical question well and answering it well should be part of the equipment of every healthcare professional. Because this is the only way to improve the quality of health care not only for one specific patient, but it can also contribute to a change in established practice, which may not always be the best.

METHODOLOGICAL PART

The proposed systematic review was conducted in accordance with the Joanna Briggs Institute methodology for systematic reviews of diagnostic test accuracy. The protocol of the DTA SR was used to provide details about the DTA SR methodology (Tucková et al., 2017).

1. Introduction to the methodological part

Type 2 diabetes mellitus, which is caused by insulin resistance, (ADA, 2010) was in the past inaccurately named and known as non-insulin-dependent or adult-onset diabetes (WHO, 2016b). Almost 90% of people with diabetes suffer from type 2 diabetes (WHO, 2016b). The main cause for its development has been considered to be excess body weight and physical inactivity (WHO, 2009). When type 2 diabetes initially occurs, there may be no symptoms, as such it is usually diagnosed several years after onset when complications have already occurred.

In 2014, 420 million people worldwide had diabetes (Danaei et al., 2011). In 2012, diabetes caused 1.5 million deaths and higher-than-optimal blood glucose caused an additional 2.2 million deaths (WHO, 2016b), (WHO, 2009). According to American Diabetes Association, in 2018, 34.2 million American, or 10.5% of population, had diabetes. Of 34.2 million adults with diabetes, 26.8 million were diagnosed, and 7.3 were undiagnosed. About 210,000 Americans under age 20 are estimated to have diagnosed diabetes, approximately 0.25% of that population. In 2014-2015, the annual incidence of diagnosed diabetes in youth was estimated at 5,800 with type 2 diabetes mellitus. Diabetes was the seventh leading cause of death in the USA in 2017 (ADA, 2018).

In the past, type 2 diabetes mellitus was thought to be a metabolic disorder that only occurred in adults. However, its incidence among children has been rising in the past two decades, so much so that it is currently the main type of children's diabetes in some parts of the world (D'Adamo & Caprio, 2011), (Shaw, 2007), (WHO, 2016a). Although the focus is often on the United States of America (USA), childhood onset type 2 diabetes mellitus occurs in children of all races and in all parts of the world (Rosenbloom, Silverstein, Amemiya, Zeitler, & Klingensmith, 2009), (Dabelea et al., 2007).

Type 2 pre-diabetes mellitus in children (T2PDMC) is not as common in Europe as it is in the USA10; however, its prevalence is rising (WHO, 2016b). This trend is supported by a recent study conducted in Italy in 2011, which suggests there is a 12.4% prevalence of glucose metabolism alterations among overweight/ obese children or adolescents (D'Adamo & Caprio, 2011), whereas in 2002, the prevalence of T2PDMC and impaired glucose tolerance (IGT) in Italian children was only 0.5% and 5% (Goran, 2002). This research (D'Adamo & Caprio, 2011), (Shaw, 2007), (WHO, 2016a), (Rosenbloom et al., 2009), (Dabelea et al., 2007), (Haines, Wan, Lynn, Barrett, & Shield, 2007), (Goran, 2002) suggests that it is probable that

future generations will suffer from more chronic diabetic complications, such as cardiovascular disease, retinopathy, neuropathy and nephropathy and malignant neoplasms, as a consequence. The earlier age of onset also makes it likely that complications will concomitantly occur in younger patients. As such, T2P-DMC is an emerging public health problem (WHO, 2016b).

Tests and methods for the diagnosis of T2P-DMC are not applied in a standardized fashion between different countries or even within them. There are diagnostic tests for adults' type 2 pre-diabetes mellitus defined by World Health Organization (WHO) and the American Diabetes Association (ADA) (ADA, 2010), (WHO, 2016b). Both the WHO and ADA apply similar thresholds for IGT but use different cutoff values for impaired fasting glucose (IFG), which are the main indicators of type 2 pre-diabetes mellitus. The ADA also defines levels of glycated hemoglobin (HbA1c) for the diagnosis of type 2 pre-diabetes mellitus (ADA, 2010), (WHO, 2016a), (Tabak, Herder, Rathmann, Brunner, & Kivimaki, 2012), (WHO, 1999). Metabolic syndrome (MS), which was also once considered to occur only in adults, is now a recognized risk factor for developing T2-DMC (IDF, 2006). However, due to ontogenetic development and the differences in metabolic rate in children, it is difficult to establish criteria for identifying MS or type 2 prediabetes mellitus in this population.

When considering whether a patient suffers from MS, a minimum of three of five major criteria must be present; in adults, these are defined as: obesity (waist circumference 102 cm [40 in] in males and 88 cm [35 in] in females), hypertension (blood pressure 130/85 mmHg), fasting glucose >110 mg/dL, triglycerides 150 mg/dL and high density lipoproteins-cholesterol<40 mg/dL (Lam & LeRoith, 2012). The International Diabetes Federation (IDF) provides measurement values for the investigation of MS in children. Many of the criteria used to define the presence of MS, and MS itself, are risk factors for the development of type 2 pre-diabetes mellitus (Grundy, 2012). These values will be used to determine whether children are at risk of T2P-DMC in the inclusion criteria (IDF, 2006), (Lam & LeRoith, 2012).

The screening of at-risk children at the different ontogenetic stages (six to 10, 10–16 and >16 years) has been recommended to be carried out every two years or at the onset of puberty by the ADA (ADA, 2010). The gold standard for the diagnosis of pre-diabetes is the fasting plasma glucose (FPG) test carried out along with the oral glucose tolerance test (OGTT). Early diagnosis of pre-diabetes is crucial for early therapy, and the prevention of complications of diabetes that seriously affect public health all around the world. However, fasting (as required by the FPG test) can be difficult to implement in children, and the OGTT requires a two-hour period of waiting between the administration of glucose and the assessment of glucose tolerance. Both of these factors have been noted as impediments to the routine diagnosis of pre-diabetes (Brar, Mengwall, Franklin, & Fierman, 2014), (Gayoso-Diz et al., 2013). As such, there is a move toward the development of tests which can be more conveniently applied in children (Brar et al., 2014), (Gayoso-Diz et al., 2013), (S. Sharma & Fleming, 2012)

It is projected that by 2030 diabetic complications will be a leading cause of death in developed countries (Whiting, Guariguata, Weil, & Shaw, 2011). However, there are no guidelines for clinical practice or systematic reviews for child populations that investigate alternate tests for the investigation and diagnosis of pre-diabetes. We searched the JBI Database of Systematic

Review and Implementation Reports, PROSPERO database and Cochrane Library in February 2016, and found no developed systematic review or protocols on this topic. The need for a practical, accurate diagnostic test for pediatric patients is great due to the epidemic of childhood obesity in developed countries. As such, this systematic review of diagnostic test accuracy (Campbell, Klugar, Ding, Carmody, Hakonsen, & Jadotte, 2015), (Campbell, Klugar, Ding, Carmody, Hakonsen, Jadotte, et al., 2015) was being conducted to synthesize the best available evidence on the diagnostic test accuracy of alternative tests (that can be carried out more readily in children) compared to the gold standard ADA tests (glucose tolerance test, FPG). An additional aim of this systematic review was to identify which alternative tests are currently being utilized for the diagnosis of pre-diabetes as to date there has been no systematic attempt to identify and describe available alternatives to the FPG test and OGTT.

1.1. Review objective/question

The objective of this systematic review was to identify all alternative tests currently in use for the diagnosis of type 2 pre-diabetes mellitus in children and establish their accuracy relative to this gold standard. The gold standard for the diagnosis of pre-diabetes was the measurement of fasting plasma glucose and the oral glucose tolerance test.

1.2. Criteria for inclusion of the studies

The aim of the DTA SR was to find and identify the most relevant sources of available scientific evidence to answer the review question. The following section will determine the criteria for inclusion and/or exclusion of the studies.

1.2.1. Participants

The current review considered studies which included children up to 18 years of age at risk of developing T2P-DMC. At-risk children were defined as those with any of the following characteristics: obesity, hypertension, low HDL levels, elevated triglyceride levels and glucose intolerance. The IDF had set criteria for how the above conditions should be defined for different stages of ontogenetic development (six to 10, 10–16, and >16 years) which was applied.

Studies with participants over 18 years was excluded.

1.2.2. Index test

The current review considered studies that evaluate alternate (not currently recommended by the ADA) diagnostic tests for pre-diabetes as index tests. These included but not be limited to

any non-fasting tests, homeostatic model assessment of insulin resistance (a mathematical index that uses fasting glucose and insulin to measure insulin resistance), measurements of serum glucose and insulin, HbA1c and 1,5 anhydroglucitol.

1.2.3. Reference test

The reference test were the tests considered to make up the gold standard for the diagnosis of pre-diabetes by the ADA for the diagnosis of pre-diabetes. These were the measurement of FPG and the OGTT. Normal fasting glucose was defined as fasting glucose 99mg/dL, whereas fasting glucose between 100 and 125 mg/dL indicates IFG and pre-diabetes (higher values suggest diabetes) (Brar et al., 2014). Normal glucose tolerance was defined as glucose 139 mg/dL 2 h after glucose intake, whereas IGT was defined as a 2-h glucose level of 140–199 mg/dL (Brar et al., 2014).

1.2.4. Diagnosis of interest

The DTA SR considered studies that had the diagnosis of type 2 pre-diabetes mellitus as their diagnosis of interest.

1.2.5. Types of studies

The DTA SR considered diagnostic cross-sectional study designs for inclusion. Diagnostic case-control studies were also included; however, as they are at risk of overestimating the accuracy of tests (Campbell, Klugar, Ding, Carmody, Hakonsen, Jadotte, et al., 2015), they were only incorporated in data synthesis in case of a lack of cross-sectional studies.

Following the search, all identified citations were collected uploaded to the citation manager EndNote X9.2. The duplicates were removed. Titles and abstracts were screened by two independent reviewers for the assessment against inclusion/exclusion criteria (DT and AR). The third reviewer (MK) was used in case of discrepancies between the two reviewers. Potential, relevant studies were retrieved in a full text and imported into the JBI SUMARI (System for the Unified Management, Assessment and Review of Information, JBI, Adelaide, Australia). The full texts of selected studies were assessed by two independent reviewers (DT and AR) for the assessment against inclusion/exclusion criteria. Any disagreement that raised between the reviewers at each stage of the study selection process was resolved through discussion, or with a third reviewer.

The results of the search are presented in a Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) flow diagram insert citation to the PRISMA statement and include in the reference list (Moher et al., 2015) (please see Figure 6).

Reasons for exclusion of the full texts that did not meet the inclusion criteria are recorded and reported in this systematic review in the form of a summary table (See table 9). The complete list of 90 excluded studies is provided in the appendix 6.

Table: 9 Summary of reasons for exclusion

Reason for exclusion	**Number of excluded studies**
Wrong patient population	19
Wrong disease of interest	14
Wrong study design	47
Conference abstracts	10
Total	**90**

1.2.6. Search strategy

The search strategy used mainly subject headings and text words related to the issue which were tailored for each included database. The search strategy aimed to find both published and unpublished studies. A three-step search strategy was used in this review. An initial search was done in two databases: Ovid MEDLINE and Embase where terms such as ‘‘children’’, ‘‘HOMA’’, ‘‘HbA1c’’, ‘‘oral glucose tolerance test’’, ‘‘pre-diabetes mellitus type two’’ and ‘‘metabolic syndrome’’ was used. This initial search was followed by an analysis of the text words contained in the titles and abstracts. In addition, index terms describing articles was assessed. A second search using all identified keywords and index terms was then undertaken across all included databases. Third, the reference list of all identified papers and reports and articles was searched for additional studies. Studies published in all possible languages (if their titles and abstracts are available in English) was considered for inclusion in the systematic review. In those studies, in which their titles and abstracts were approved to be eligible for inclusion, the complete manuscript was translated. For this review, no time restriction was considered. The databases to be searched included: MedLine@ Ovid MEDLINE, Biomedica Czechoslovaca, Embase, Cochrane library, EMBASE, Emcare, CINAHL, Web of Science and Scopus, Pedro. The search for unpublished studies included: Open Grey, Current Controlled Trials, MedNar, ClinicalTrials.gov, Cos Conference Papers Index, International Clinical Trials Registry Platform of the World Health Organization, ProQuest, and Google Scholar. The initial key words to be used in the first search are presented in Appendix 7.

1.2.7. Information sources

Information sources is a list all information sources (e.g. electronic databases, contact with study authors etc.) The databases to be searched included insert databases with platforms as

appropriate. Sources of unpublished studies and grey literature to be searched included insert text, e.g. trial registers etc.

1.2.8. Assessment of methodological quality

Papers selected for retrieval were assessed by two independent reviewers (DT and AR) for methodological validity prior to inclusion in the review using the JBI diagnostic test accuracy review instrument (see Appendix 3) Critical appraisal checklist which was based on quality assessment of diagnostic accuracy studies (Campbell, Klugar, Ding, Carmody, Hakonsen, & Jadotte, 2015). Any disagreement that raised between the reviewers was resolved through discussion, or with a third reviewer (MK). All studies, regardless of their methodological quality, was included in the review. Analysis of sensitivity was performed to assess if the results were influenced by methodological quality.

1.2.9. Data extraction

Data were extracted from papers included in the review using the standardized data extraction instrument from the diagnostic test accuracy chapter in the JBI Reviewers' Manual (Campbell, Klugar, Ding, Carmody, Hakonsen, & Jadotte, 2015). Two reviewers (DT and AR) extracted data independently. If there was disagreement, a third reviewer (MK) involved. The data extracted included specific details about the populations, index tests and diagnosis of interest relevant to the review question and objectives. If there were data missing or incomplete, the reviewers contacted the authors or corresponding authors of the primary studies, but only one of them returned emails. The one who kindly responded was prof. prof. Noor Shafina Mohd Nor (shafinamohdnor@yahoo.com). All included studies were unfortunately presented without a standard 2x2 table showing TN, TP, FN and FP. So a third reviewer (MK) used several transformation methods to retrieve 2x2 values, which are necessary values for the Revman v 5.4 (Cochrane, 2020). However, the majority of studies were missing basic information completely, or they were missing them for specific thresholds. We calculated the data manually with the provided sensitivities, specificities, prevalence or likelihood ratios or an absolute number of included patients and number of positively diagnosed cases using MS Excel and diagnostic calculators from Cochrane (Cochrane, 2020) and Schwartz (Schwartz, 2014). We were able to manually calculate data from (Brar et al, 2014), (Ehehalt, 2017), (Garcia, 2019) and (Nam, 2018).

1.2.10. Data synthesis

Diagnostic data, where possible, were pooled in statistical meta-analysis. Data are presented graphically in two ways. Forest plots were used for sensitivity and specificity for each of the selected primary studies. This graph displays the means and confidence intervals (CI) for sensitivity/specificity on the level of 95% CI. These values are also expressed in numerical form. Moreover, the number of true positives, false positives, true negatives and false negatives

are also reported. Where possible, summary receiver operating characteristic (SROC) curves were created. The Multiple tests model for meta-analysis was be used. Initially, clinical heterogeneity was assessed by determining whether the studies are sufficiently similar to pool in terms of the inclusion criteria. The clinical heterogeneity is present within identified studies as because of the different populations in terms of ethnicity, gender and age so from the perspective of different cut of values. Different index and referenced tests are pooled separately according to different tests but also according to different thresholds. To reduce methodological heterogeneity three groups of Forest plots were created as well as three groups of SROC curves.

1.2.11. Assessing certainty of findings

The grading of Recommendations, Assessment, Development and Evaluation (GRADE) approach for grading the certainty of evidence was followed, and Summary of Finding (SoF) was using GRADE (Boon MH, Klugar M, & E, 2021). Due to high heterogeneity and small numbers of studies using similar tests is certainty of all results very low.

2. Results of the systematic review

2.1. Literature search and results

The process of the literature search in published and unpublished sources of literature will be described in this chapter. It will be also described how the studies that have been searched in the databases were organized using the citation manager EndNote X.9.

2.2. Explanation of the literature search

A systematic search of relevant studies was performed in 18 databases of both published and unpublished sources of literature on 20th March 2020. For the comprehensive search strategy please see Appendix 8.

The databases that were searched included MEDLINE, Embase (Elsevier), CINAHL (EBSCO), Web of Science Core Collection, Scopus, Emcare (Ovid), ProQuest Dissertations & Theses Global, Cochrane Library, Bibliographia medica Čechoslovaca and PEDro (Physiotherapy Evidence Database). Sources of unpublished studies and grey literature searched included MedNar, OpenGrey and clinical trials registers ClinicalTrials.gov, WHO International Clinical Trials Registry Platform (ICTRP) and Current control trials (ISRCTN registry).

Totally 7559 records were retrieved by the search and uploaded into EndNote X9.2. Deduplication of results was conducted according to the method described by Bramer et al. (Bramer, Giustini, de Jonge, Holland, & Bekhuis, 2016); and 2557 duplicates were detected. This led to 5002 records selected for the title and abstract screening. The results of the search are shown in the Table 10.

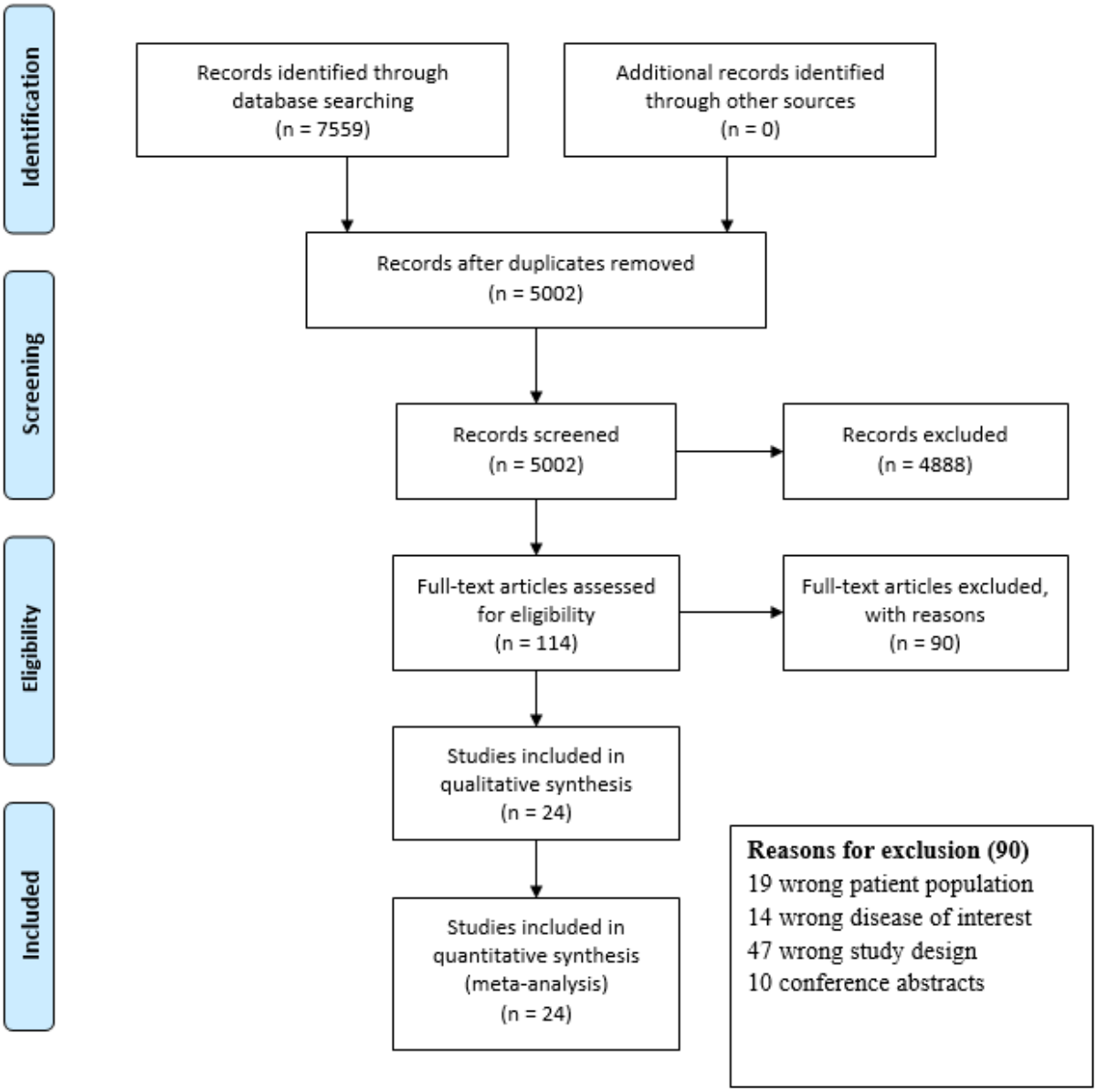
Table: 10 Systematic search results

Database	Total
BMČ	39
CINAHL	77
Cinahl Trials	599
Cochrane Library	41
Current control trials	84
EMBASE	654
EmCare	41
ICTRP	20
Mednar	635
Ovid Medline	619
Pedro	471
ProQuest Dissertation	9
Scopus	157
WoS	1511
PsychINFO	10
Ovid Nursing	24
COS Conference Papers	0
Open Grey	11
Total records found	**5002**

Two independent reviewers (DT and AR) analysed the number of 5002 studies at the title and abstract against the eligibility criteria. This phase resulted to 114 studies selected for a full text review for eligibility. The full text assessment was done again by two independent reviewers (DT and AR) using SUMARI (The System for the Unified Management, Assessment and Review Information (the JBI, Adelaide, Australia). The third reviewer (MK) was used as an arbiter in a case of any discrepancy between the two independent reviewers. The third reviewer provided an objective supervision and he was asked for arbitration in four cases (Lee, 2011; Lee, 2016; Lee, 2012). From the number of 114 studies, 90 full texts were excluded with a reason (19 studies had wrong population, 14 studies had wrong disease of interest, 47 studies had wrong study design, 10 records were conference abstract – for some of them the librarian was able to find title and abstract, but for some of them only the information about authors, name and year of a conference was possible to find). Two studies were additionally excluded after critical appraisal in SUMARI. Upon closer examination of the methodology of these two studies, it was found that they did not meet the inclusion criteria. Please, see the final PRISMA diagram (Figure 6) is on the next page:

Figure 6: PRISMA flow chart

Figure 6 Search and analysis flow diagram (Preferred Reporting Items for Systematic Reviews and Meta-analysis) Diagram adapted (Moher et al., 2009)



From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). *Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement.* PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

2.3. Description of the studies

In the chapter 2.3, the exploration of methodological and quality facet of the research using the JBI critical appraisal checklist will be explained.

2.3.1. Methodological quality of included studies

All the 27 included studies determined for narrative synthesis were recorded to JBI SUMARI software which contains the critical appraisal checklist modified and expanded from "Synthesizing evidence of diagnostic accuracy" (Campbell, Klugar, Ding, Carmody, Hakonsen, & Jadotte, 2015). Two independent reviewers (DT and AR) used the JBI critical appraisal tool for diagnostic test accuracy (Version 29 Aug 2017). If the study met the criteria, a "YES" was given and these were added together as a cumulative score of 1+ (for each "YES") up to a total possible score of 10 from total number of items 10. The higher the score of the individual studies, the more it was possible to point out the quality that could be expected from these types of studies assessing diagnostic accuracy. The main reason for performing this step is to assess what is the methodological quality of individual studies and if all studies might be pooled statistically together, or different methodological quality would be the source of methodological heterogeneity and such the studies with different levels of quality should be statistically pooled separately using sensitivity analyses. Table 11 shows the critical appraisal results. Where there is a "YES" answer that means that the condition was met. The higher score indicates the higher quality of the study. The studies in Table 11 are listed in alphabetical order. A total score of +1 items is displayed in the right column expressed as a percentage of "yes" answers out of the total number of answered items (Table 11).

Table: 11 Critical appraisal for DTA studies

Author, year	Q1 Was a consecutive or random sample of patients enrolled?	Q2 Was a case-control design avoided ?	Q3 Did the study avoid inappropriate exclusions?	Q4 Were the index results interpreted without knowledge?	Q5 If a threshold was used, was it pre-specified?	Q6 Is the reference standard likely to correctly classify the target condition?	Q7 Were the reference standard results interpreted without knowledge of the results of the index test?	Q8 Was there an appropriate interval between index test and reference standard?	Q9 Did all patients receive the same reference standard ?	Q 10 Were all patients including in the analysis ?	Total "YES" score (%)
Atabek, 2007	Unclear	Yes	Yes	Unclear	Yes	Yes	Unclear	Unclear	Yes	Yes	60
Brar, 2014	Yes	Yes	Yes	No	Yes	Yes	No	Yes	Yes	Yes	80
Bridges, 2016	Yes	Yes	Yes	Unclear	Yes	Yes	Unclear	Unclear	Yes	Yes	70
Ehehalt, 2017	Unclear	Yes	Yes	No	Yes	Yes	No	Yes	Yes	Yes	70
Galhardo, 2015	Yes	Yes	Yes	No	Yes	Yes	No	Yes	Yes	Yes	80
Garcia, 2019	No	Yes	Yes	Unclear	Yes	Yes	Unclear	Yes	Yes	Yes	70
Chan, 2015	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	80
Chan, 2016	No	Unclear	Yes	No	Yes	Yes	No	Yes	Yes	Yes	60
Kang, 2017	Yes	Yes	Yes	Unclear	Yes	Yes	Unclear	Yes	Yes	Yes	80
Kasturi, 2019	Yes	Yes	Yes	No	Yes	Yes	No	Yes	Yes	Yes	80
Keskin, 2005	Unclear	No	Yes	Unclear	Yes	Yes	Unclear	Yes	Yes	Yes	60
Kim, 2018	Yes	Yes	Yes	Unclear	Yes	Yes	Unclear	Yes	Yes	Yes	80
Kurtoğlu, 2010	Yes	Yes	Yes	Unclear	Yes	Yes	Unclear	Yes	Yes	Yes	80
Lee, 2019	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	100
Lee, 2012	Yes	Yes	Yes	Unclear	Yes	Yes	Unclear	Yes	Yes	Yes	80
Lee, 2011	Unclear	Yes	Yes	Unclear	Yes	Yes	Unclear	Unclear	Yes	Yes	60
Liang, 2015	Yes	Yes	Yes	Unclear	Yes	Yes	Unclear	Yes	Yes	No	70
Maffeis, 2010	Yes	Yes	Yes	Unclear	Yes	Yes	Unclear	Yes	Yes	Yes	80
Maldonado-Hernández, 2016	Unclear	Yes	Yes	Unclear	Yes	Yes	Unclear	Yes	Yes	Yes	70
Mutlu, 2013	Yes	Yes	Yes	Unclear	Yes	Yes	Unclear	Yes	Yes	Yes	80
Nam, 2017	Yes	Yes	Yes	No	Yes	Yes	No	Unclear	Yes	Yes	70
Noor, 2015	No	Yes	Yes	Yes	N/A	Yes	Yes	Yes	Yes	Yes	80
Pandey, 2017	Yes	Unclear	Yes	Unclear	Yes	Yes	Unclear	Yes	Yes	Yes	70
Puri, 2007	Unclear	Yes	Yes	No	Yes	Yes	Unclear	Yes	Yes	Yes	70
Shah, 2009	Yes	No	No	No	Yes	Yes	No	Yes	Yes	Yes	60
Sharma, 2012	Unclear	Yes	Yes	Unclear	Yes	Yes	Unclear	Yes	Yes	Yes	70
Tirabanchasak, 2015	Yes	Yes	Yes	Unclear	Yes	Yes	Unclear	Unclear	Yes	Yes	70
Overall quality per domain (%)	Yes 59, Unclear 29; No 11.1	Yes 85.1, Unclear .4, No 7.4	Yes 100, Unclear 0, No 0	Yes 12, Unclear 59.2, No 29.6	Yes 96.2, N/A 3.7	Yes 100, Unclear 0, No 0	Yes 11.1, Unclear 62.9, No 25.9	Yes 81.4, Unclear 18.5, No 0	Yes 100. Unclear 0, No 0	Yes 92.5, Unclear 0x, No 7.4	

Overall, the quality of included studies was good, as illustrated in Table 11. None of the 27 study was excluded based on the critical appraisal outcome. The assessment of methodological quality of the included studies was completed by two reviewers (DT and AR) within SUMARI

and the results transfer to excel. In this phase these two reviewers provided reasons for exclusion and then consensus was sought were disagreement occurred (DT and AR). There were as well the requirements of the third reviewer (MK) because the third author was assigned to provide abirritation. The studies the third reviewer arbitrated studies from Lee (2011), Lee (2012), Lee (2016), Shah (2009).

In Q1 about consecutive or random enrollment the answer "NO" was elected in 3 studies (Garcia, 2019; Chan, 2016; Noor, 2015) and answer "UNCLEAR" in 8 studies (Atabek, 2007; Ehehalt, 2017; Chan, 2015; Keskin, 2005; Lee, 2011; Maldonado-Hernández, 2016; Puri, 2007; Sharma, 2012). That could indicate to the patient selection bias using the data from the included studies.

One study received a rating "NO" (Keskin, 2005) and two studies received a rating "UNCLEAR" (Pandey, 2017; Chan, 2016) for Q2 whether a case-control design was avoided.

All 27 studies received in Q3 (avoiding of inappropriate exclusions) a rating "YES", except of Shah (2009).

In the most studies the interval between the index test and the reference test (Q4 and Q8) was unreported, but we believe that this fact is unlikely to undermine the reliability and validity of the results.

In most of the studies the threshold was pre-specified (Q5) and the reference standard likely to correctly classify the target condition (Q6).

In Q7 ("Where the reference standard results interpreted without the knowledge of the results of the index test?"), the results were reported rarely. Only in 3 studies (Chan, 2015; Lee, 2019; Noor, 2015), Q7 was answered clearly and understandably. The rest of studies received the rating "NO" (in total 7 studies) or "UNCLEAR" (in total 17 studies).

The evaluation shows that 29 studies were rated "YES" in Q9 whether all patients received the same reference standard, except of one "UNCLEAR" rating (Tirabachasak, 2015).

In 2 studies, participants were not included into the analysis (Q10):

- Chan (2015) – number of 12 participants (from total 118) were excluded due to incomplete CGM data;
- Liang (2015) – number of 42 participants (from total 1069) were excluded because the did not meet inclusion criteria (31 with difficult of blood sampling, 11 with a low birth weight, 12 diagnosed with early-onset T2DM, 9 with distress during BP monitoring, 20 with missing data in clinical or laboratory record, 10 refused to participate).

The lowest rating of studies was 6 "YES" answers (Atabek, 2007; Chan, 2016; Keskin, 2005; Lee, 2011; Shah, 2009), one study achieved a full evaluation (Lee, 2019).

If there was any disagreement about an item, both reviewers (DT and AR) discussed the differences and examined the reasons for the different answer to the question. After that they decided to assess the item upon a consensus. There was need for a third reviewer (MK) in a

critical appraisal of the studies. The studies the third reviewer arbitrated studies from Puri (2007) and Sharma (2012).

2.4. Included studies

Twenty-four studies appeared to provide data that could be extracted for the DTA SR.

Within these 24 studies, surprisingly only 2 of them followed STARD (Garcia, 2019 and Liang, 2015).

Based on the information provided by the studies´ authors regarding the study design, we found that:

- the design of the cross-sectional study was defined for 10 studies (Galhardo, 2014; Garcia, 2019; Chan, 2015; Kang, 2017; Lee, 2019; Liang, 2015; Maldonado-Hernández, 2016; Noor, 2015; Pandey, 2017; Sharma; 2012);
- the design with consecutive enrolment was stated in 7 studies (Atabek, 2017; Chan, 2016; Keskin, 2008; Kim, 2018; Kortoglu, 2010; Maffeis, 2010; Puri, 2007);
- the design with random sampling was states in 1 study (Bridges, 2016);
- the design of the retrospective chart view was defined for 4 studies (Brar, 2014; Mutlu, 2013; Nam, 2017; Tirabanchasak, 2015);
- the design of observational analysis was determined in 1 study (Ehehalt, 2017);
- the design of secondary analysis of randomized control trials was stated in 1 study (Kasturi, 2016);

In 6 studies, we needed to contact the primary authors or corresponding authors to provide us comments or original data from their studies. The list of contacted authors is followed:

- Dr. Perrin C. White, (perrin.white@utsouthwestern.edu>)
- Dr. Susanna Wiegand, (susanna.wiegand@charite.de>)
- Prof. Mehmet Keskin, (mkeskin@gantep.edu.tr>)
- Dr. Christine L. Chan (2 studies), (Christinel.chan@childrenscolorado.org>)
- Prof. Noor S. Mohd Nor (shafinamohdnor@yahoo.com>)

Only prof. Mohd Noor answered with apology that the set of original data are missing hence they are not able to provide original 2x2 table data that was needed.

A list of the 24 included studies is provided in Table 12. For the detailed description of the 24 included studies please Appendix 9

Table: 12 List of 24 included studies

No.	Study author & yeas	Country	Sample	Study type (enrolment)	Index test	Reference test	Aim of the study stated in a study
1.	Atabek 2007	Turkey	148	Consecutive enrolment	FGIR, HOMA-IR, QUICKI	OGTT	To compare Simple indices of insulin resistance calculated from fasting glucose and insulin levels with insulin sensitivity indices determined by OGTT (area under the response curve [$AUC_{insulin}$] and insulin sensitivity index [ISI-composite]) in obese children.
2.	Brar 2014	USA	149	A retrospective chart review of patients	HbA1c, HOMA-IR.	OGTT	1. To evaluate the accuracy of HbA1c and HOMA-IR as single screening tests for prediabctesit2dm in obese children and adolescents (compared with the OGTT criterion standard) and 2. To assess whether combining HbA1c with either fasting glucose or HOMA-IR increases the accuracy of diagnosing prediabetes/T2DM as confirmed by a positive OGTT result.
3.	Bridges 2016	USA	223	Random sampling, paediatric electronic medical records	TRG/HDL	HOMA-IR top quartile	To investigate the ability of TRG/HDL ration to assess IR in obese and overweight children.
4.	Ehehalt 2017	Germany	4848	An observational haemoglobin analysis	HbA1c	OGTT	To investigate the test properties of fasting plasma glucose, 2-h glucose, and hba1c levels for screening of type 2DM in asymptomatic or oligosymptomatic overweight and Obese children and adolescents living in Germany, and 2. To find appropriate cut-off values for the detection of manifest diabetes in children.
5.	Galhardo 2015	UK	266	A cross-sectional study	HbA1c, Fasting blood glucose, HOMA-IR, TG:HDL-ratio	OGTT	To assess hba1c as a screening tool for pre-diabetes and DM2 in high-risk obese children from a country with mostly Caucasian ethnicity.
6.	García 2019	Mexico	201	A prospective, comparative cross-sectional study	TyG, TG/HDL	HOMA-IR	To evaluate the sensitivity and specificity of TyG and TG/HDL for predicting IR.
7.	Chan 2015	USA	98	A cross-sectional study	HbA1c	OGTT	To examine whether glycosylated haemoglobin (hba1c)

							or the oral glucose tolerance test (OGTT) is a better predictor of free-living glycemia as measured by continuous glucose monitoring (CGM).
8.	Chan 2016	USA	117	Consecutive enrolment	1,5-anhydroglucitol, fructosamine, glycated albumin	OGTT, hba1c	To assess the ability of these three alternates non-fasting glycaemic markers to predict dysglycemia in obese youth as defined by the traditional screening tests hba1c and OGTT.
9.	Kang 2017	South Korea	231	A cross-sectional study	TyG, TG/HDL	HOMA-IR	To investigate the association between the triglycerides/glucose index (TyG index) and the homeostasis model assessment-estimated insulin resistance (HOMA-IR) in the prediction of insulin resistance (IR) among adolescents.
10.	Kasturi 2016	USA	93	A secondary analysis of a randomized controlled trial	OGTT – baseline, glucose peak >30 minutes, monophasic curve, 1-hr glucose 155 mg/dl	OGTT	To compare the reproducibility and diagnostic accuracy of these three morphological features of the OGTT glucose curve over a 6-week period.
11.	Keskin 2005	Turkey	57	Consecutive enrolment	HOMA-IR, QUICKI, FGIR	OGTT	To compare the HOMA, FGIR, and QUICKI methods for measuring insulin resistance, expressed by oral glucose tolerance test (OGTT) results, among obese children and adolescents.
12.	Kim 2018	South Korea	190	Consecutive enrolment	HbA1c, FPG	OGTT	To evaluate the correlation between plasma glucose and Hba1c and the diagnostic accuracy of hba1c as a screening tool to identify asymptomatic diabetes mellitus in children and adolescents with obesity or asymptomatic glucosuria.
13.	Kurtoğlu 2010	Turkey	268	Consecutive enrolment	HOMA-IR	OGTT	To determine HOMA-IR cut-off values in obese children and adolescents according to gender and pubertal status.
14.	Lee 2019	South Korea	9502	A nationally representative cross-sectional examination	HbA1c	FPG	To assess the extent of agreement between diagnoses based on FPG versus hba1c levels, to evaluate the diagnostic performance of hba1c, and to determine the optimal hba1c cut off values for diabetes and prediabetes in youths

							and young adults by using nationally representative data of Korea.
15.	Liang 2015	China	976	A cross-sectional study	HOMA-IR, TG/HDL-C	OGTT	To investigate the optimal cut offs of TG/HDL-C ratio, HOMA-IR and compare their accuracy to identify the MS in Chinese obese children.
16.	Maffeis 2010	Italy	563	Consecutive sampling	HOMA-IR, FPG, FSI	OGGT	To see whether one or more fasting glucose metabolism parameters were able to predict IGT in obese children and whether they could be suitable as screening tools for selecting obese children to be tested with OGTT.
17.	Maldonado-Hernández 2016	Mexico	133	Cross-sectional study	HOMA-IR, FPG	OGTT	To assess the use of the 13C-GBT for IR detection in adolescents through comparison with fasting and post-glucose stimulus IR surrogates.
18.	Mutlu 2013	Turkey	106	Medical records evaluated retrospectively	HbA1c	OGTT	To investigate whether hba1c and 1-hour glucose in OGTT are useful parameters for evaluation of glucose homeostasis in children and adolescents.
19.	Nam 2017	South Korea	389	Retrospectively reviewed the medical records	HbA1c	OGTT	To evaluate the diagnostic performance of hba1c and to compare the results with those of the OGTT, to determine the optimal cut off points for detection of prediabetes and diabetes in a large number of children and adolescents.
20.	Noor 2015	USA	225	Cross-sectional data	TyG index, TyG/HDL, 1/IF	Insulin-stimulated glucose disposal (Rd)	To assess the associations between the TyG index and *in vivo* insulin sensitivity, measured with the hyperinsulinemic–euglycemic clamp, in OB along the spectrum of glucose tolerance from normal to predm to diabetes; to compare the ability of TyG index, TG/HDL, and 1/IF in predicting insulin resistance in obese youth.
21.	Pandey 2017	India	526	Cross-sectional study	BMI, waist circumference	Not known	To find out the cut-off values of BMI and waist circumference for predicting pre-diabetes in adolescents in the Indian population.
22.	Puri 2007	USA	167	Consecutive enrolment	HOMA-IR	OGTT	To identify those obese minority youth at

							Greatest risk for having an abnormal oral Glucose tolerance test (OGTT) indicating impaired glucose tolerance (IGT) or type 2 Diabetes mellitus (DM2).
23.	Sharma 2012	USA	172	A full set of data of cross-sectional analysis	HOMA-IR, FPG	Hba1c	To compare the discriminating power of hba1c with other prediabetes diagnostic tests specifically in high-risk African American children, using a dual HPLC method that avoids confounding of hba1c levels due to the presence of genetic variants.
24.	Tirabanchasak 2015	Thailand	115	The study protocol; Data collected from the medical charts	FG, HOMA-IR	OGTT	1) describe fasting biochemical markers and fasting- or OGTT-derived indices of insulin resistance and secretion in obese youth; 2) identify the cut-off values of fasting glycaemic markers and insulin dynamic indices that could be used to predict IGT.

To enable this systematic review to be compared with relevant results across the studies the authors had to demonstrate that it was appropriate to include them in the comparison. It was necessary for included studies to report a similar methodology and provide data on diagnostic tests that could be compared based on their same cut off points using both, index test and reference test. This was not always prima facie. Only one study (Ehehalt, 2017) referred false positive, false negative number in STARD suggested 2x2 table, however true positive and true negative values were, where possible, manually calculated by a third reviewer (MK). The numbers from the rest of the studies had to be transformed and recalculated into the 2x2 table where possible in order to be able to perform the meta-analyses.

The index tests used in these 24 studies were as follows:

- HOMA-IR;
- FGIR;
- QUICKI;
- HbA1c;
- (proposed) TRH/HDL or TG:HDL_C ratio;
- (proposed) TyG;
- FPG ≥ 126 mg/dl (≥7.0 mmol/l);
- 2-h glucose ≥ 200 mg/dl (≥11.1 mmol/l);
- HbA1c ≥ 48 mmol/ mol (≥6.5%),

- HbA1c ≥ 39 mmol/ mol (≥5.7%);
- HbA1c ≥ 39 mmol/mol (≥5.7%);
- FPG ≥ 100 mg/dl (≥5.6 mmol/l);
- Fructosamin;
- Glycated albumin;
- 1.5-anhydroglucitol;
- Glucose peak > 30 minutes;
- Monophasic curve;
- 1-hr glucose 155 mg/dL;
- COMBO;
- FSI;
- % OD: adjusted percentage of oxidized ^{13}C-glucose dose at 180 minutes;
- 1/IF;
- BMI;
- Waist circumference;

The reference tests used in these 24 studies were as follows:

- OGTT;
- HOMA top quartile;
- HOMA-IR;
- 2h-glucose category;
- HbA1c;
- Fasting glucose 100 mg/dL,
- 2-hr glucose 140 mg/dL;
- FPG;
- FPI≥p90;
- 2h-OGTT;
- PI≥65 μU/ml;
- Insulin-stimulated glucose disposal (Rd)

Finally, from 16 studies of these 24 included studies, it was possible to make six pairs of the same tests that were supposed to be used to meta-analysis:

- HOMA-IR and OGTT;
- HbA1c and OGTT;
- TyG and HOMA-IR;
- TG_HDL and HOMA-IR;
- FPG and OGTT;
- TrG_HDL and OGTT.

The index tests used in the included studies for identification of diagnostic accuracy were indicated in studies as follows:

- HOMA-IR (Atabek, 2007; Brar, 2014; Galhardo, 2015; Keskin, 2005; Kortoglu, 2010; Maffeis, 2010; Puri, 2007; Liang, 2015);
- HbA1c (Brar, 2014; Ehehalt, 2017; Chan 2015; Mutlu, Nam, 2018; 2013; Puri, 2007);
- TyG (Garcia, 2019; Kang, 2017);
- TG_HDL (Garcia, 2019; Kang, 2017; Bridges, 2016);
- FGIR (Atabek, 2007; Keskin, 2005);
- FPG (Ehehal, 2017; Maffeis, 2010);
- TrG_HDL (Galhardo, 2015; Liang, 2015).

The reference tests used in the included studies for identification of diagnostic accuracy were indicated in studies as follows:

- OGTT (Atabek, 2007; Brar, 2014; Chan, 2015; Ehehalt, 2014; Galhardo, 2015; Keskin, 2005; Kortoglu, 2010; Maffeis, 2010; Nam, 2018; Puri, 2007; Liang, 2015);
- HOMA-IR (Bridges, 2016; Garcia, 2019; Kang, 2017);

After checking and comparing the individual cut off points for selected pairs of tests it was possible choose two identical cut off points in a pair of HbA1c and OGGT tests:

- Brar (2014) & Ehehalt (2017) – cut off point 5.7;
- Brar (2014) & Nam (2018) – cut off point 5.8;
- Brar (2014) & Chan (2015) – cut off point 5.9;

And identical cut off points in a pair of HOMA-IR and OGTT tests:

- Atabek (2007) & Brar (2014) – cut off point 2.7;
- Brar (2014) & Galhardo (2015) – cut off point 4.

No other identical cut off points were found in the other test pairs of included studies.

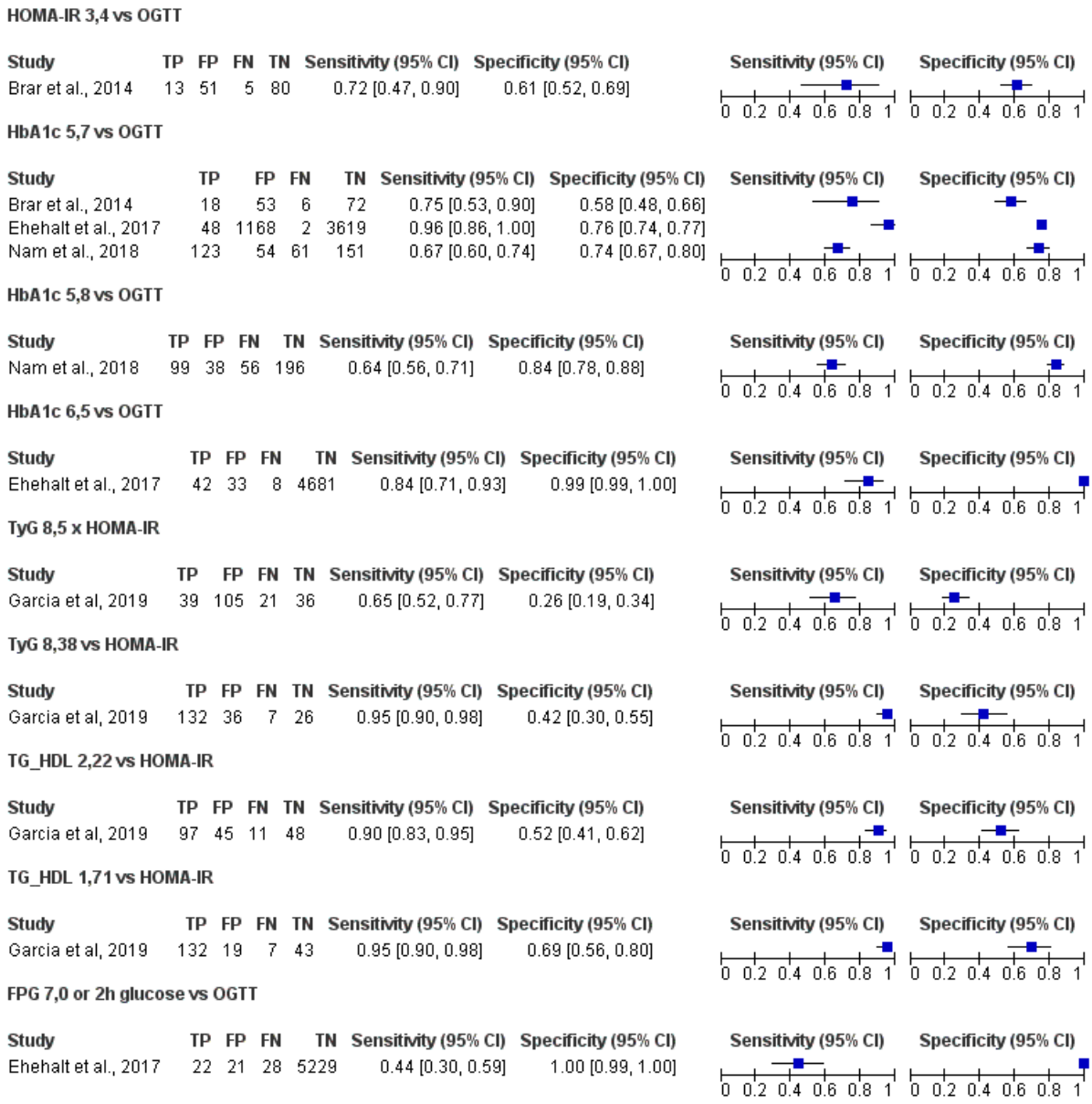
2.5. Meta-analyses of the results of the included studies

In this chapter, we pooled the manually calculated results from studies where manual calculation and data transformation were possible. Clinical and methodological heterogeneity is high for the presented results, so we have to interpret them very cautiously.

Total number of four studies was possible to be pooled in the meta-analyses and these studies had two reference tests OGTT and HOMA-IR and five index tests HOMA-IR, HbA1c, TyG, TG-HDL and FPG. HbA1c used three different cut off values 5.7; 5.8 and 6.5; TyG used two different cut off values 8.5 and 8.38 and TG_HDL used also two different cut off values 2.22 and 1.71. Separate meta-analyses were plotted as for different pairs of tests so for different cut off values. Total number of 9 meta-analyses are presented within the figure 7.

Figure 7: Forest plot of tests: 1) HOMA-IR 3.4 vs OGTT; 2) HbA1c 5.7 vs OGTT; 3) HbA1c 5.8 vs OGTT; 4) HbA1c 6.5 vs OGTT; 5) TyG 8.5 x HOMA-IR; 6) TyG 8.38 vs HOMA-IR; 7) TG_HDL 2.22 vs HOMA-IR; 8) TG_HDL 1.71 vs HOMA-IR; 9) FPG 7.0 or 2h glucose vs OGTT.

Figure 7: 9 meta-analyses forest plot of tests



The only meta-analysis had three studies in which HbA1c as an index test with a cut off of 5.7 vs OGTT as a reference test was compared. At first glance, it is clear from the meta-analysis that the result of the Ehehalt study (Ehehalt, 2017) is the most accurate. However, it should be noted that the individual studies that were included in the meta-analysis differ clinically (clinical heterogeneity). In the Ehehalt (2017) study 4848 overweight, obese, and extremely obese children and adolescents from Germany aged 7 to 17 years were included. In Brar´s study (2014), the number of participants was 149 aged 13.8+/-3.1. It was conducted in the USA and

it included 5 different ethnicities. Population in Nam´s study (2018) was 10 years and above with body mass index ≥ 85th percentile for age and gender and having two or more additional risk factors for diabetes, consistent with American Diabetes Association (ADA). So, if we assess the Ehehalt´s study (2017), Brar´s study (2014), Nam´s study (2018), the biggest difference is in ethnicity.

From the figure 7 for forest plot of the tests (HOMA-IR, HbA1c x OGTT; TyG, TG_HDL x HOMA-IR) we can see the results of the four studies included in meta-analysis. The results from Brar´s study where HOMA-IR as an index test and OGTT as a reference test were assessed at the level of cut off point 3.4 was 72.00% sensitivity and 61.00% specificity.

Three studies of Brar (2014), Ehehalt (2017) and Nam, (2018) searched an index test HbA1c on the level of cut off point 5.7 versus reference test OGTT. The Brar´s study (2014) had sensitivity 75.00% and specificity 58.00%. Ehehalt´s study (2017) had sensitivity 96.00% and specificity 76.00%. Nam´s study (2018) had sensitivity 67.00% and specificity 74.00%.

In Nam´s study (2018), cut of point of 5.8 was used for index test HbA1c versus reference test OGTT. The sensitivity was 64.00% and specificity was 84.00%.

Study of Ehehalt (2017), was used an index test HbA1c and reference test OGTT at the level of cut off point 6.5. The sensitivity was 84.00% and specificity was 99.00%.

Study from Garcia (2019) used TyG and TG_HDL as an index tests and HOMA-IR as a reference test at different levels of cut off points. The cut off point of 8.5 with an index test TyG and reference test had 65.00% sensitivity and 26.00% specificity. The same tests (TyG and HOMA-IR) was used at the level of cut off point 8.38 with 95.00% sensitivity and 42.00% specificity. In TG_HDL as an index test and HOMA-IR as a reference test, the cut off point 2.22 was used. The sensitivity was 90.00% and specificity 52.00%. The same tests (TG_HDL and HOMA-IR) was used at the level of cut off point 1.71 with 95.00% sensitivity and 69.00% specificity.

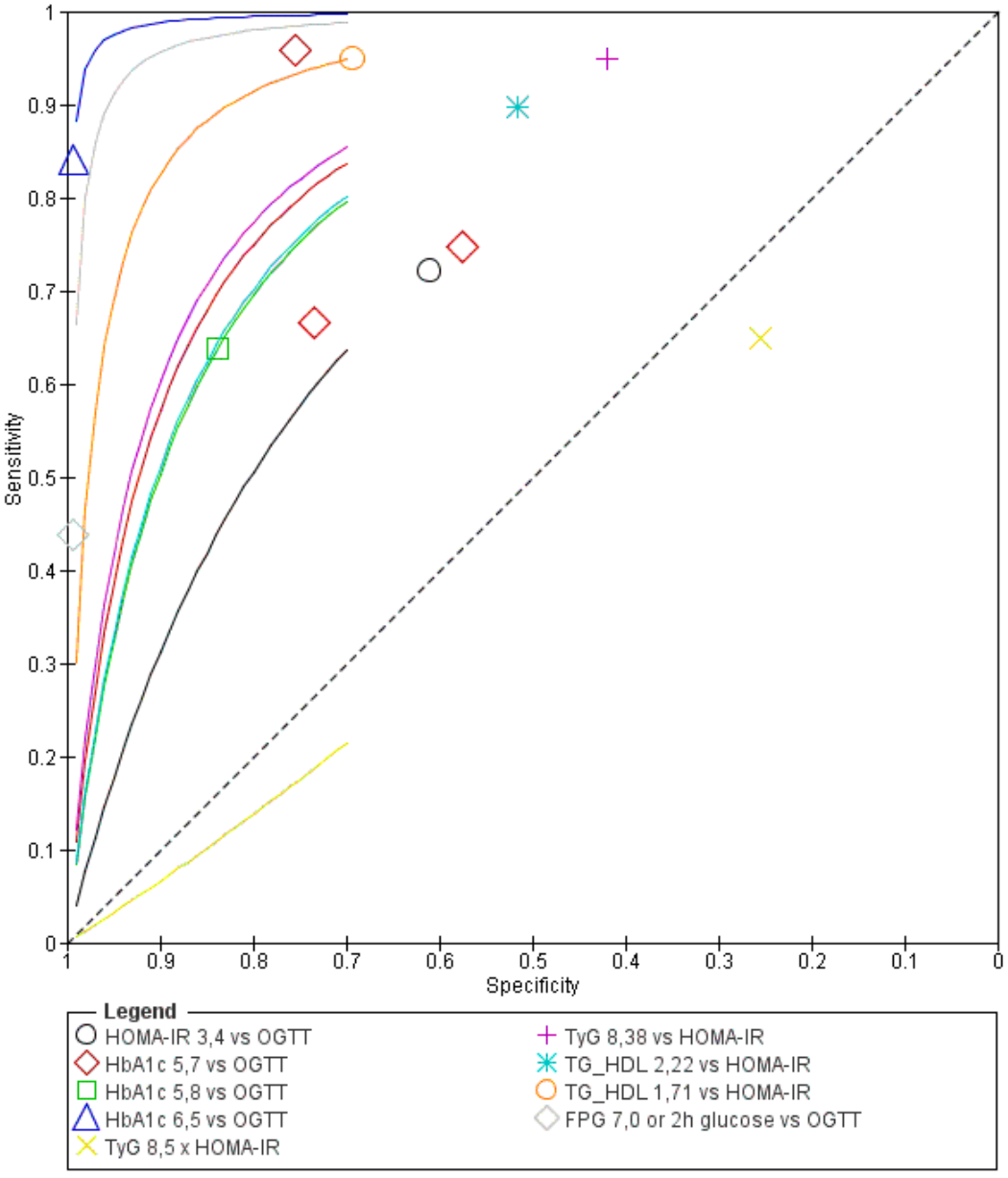
Ehehalt´s study (2017) searched for the sensitivity and specificity of FPG or 2h glucose as an index test and OGTT as a reference test. The cut off point was determined at the level of 7.0. The sensitivity was 44.00% and specificity was 99.60%.

The other comparisons were in only one study and so the rest of the meta-analyzes are single meta-analyzes. However, all results were analyzed by multiple test analysis using an SROC plot (please see Figure 8), which must be interpreted very carefully for all studies. This is due to the fact that we simply cannot state what we see – thus, the most accurate index test (of those analyzed here) is HbA1c with a cut off of 6.5 vs OGTT, and the second most accurate is FPG with a cut off point 7.0 vs OGTT, and the least accurate is TyG with cut off point 8.5 vs HOMA-IR. This is because each test is used in a given study on a different population, on a different ethnicity. And in addition, we compare different thresholds, different reference tests. Therefore, we created two subgroups of SROCs and meta-analyzes, which are divided according to reference tests (OGTT and HOMA-IR). Based on the above, it can be outlined that with a

certain dose of caution, HbA1c with a cut off point of 6.5 is indeed the most accurate index test of all the index tests included in the meta-analysis.

Figure 8: Summary ROC Plot of tests: 1) HOMA-IR 3.4 vs OGTT; 2) HbA1c 5.7 vs OGTT; 3) HbA1c 5.8 vs OGTT; 4) HbA1c 6.5 vs OGTT; 5) TyG 8.5 x HOMA-IR; 6) TyG 8.38 vs HOMA-IR; 7) TG_HDL 2.22 vs HOMA-IR; 8) TG_HDL 1.71 vs HOMA-IR; 9) FPG 7.0 or 2h glucose vs OGTT.

Figure 8: Summary ROC Plot of tests

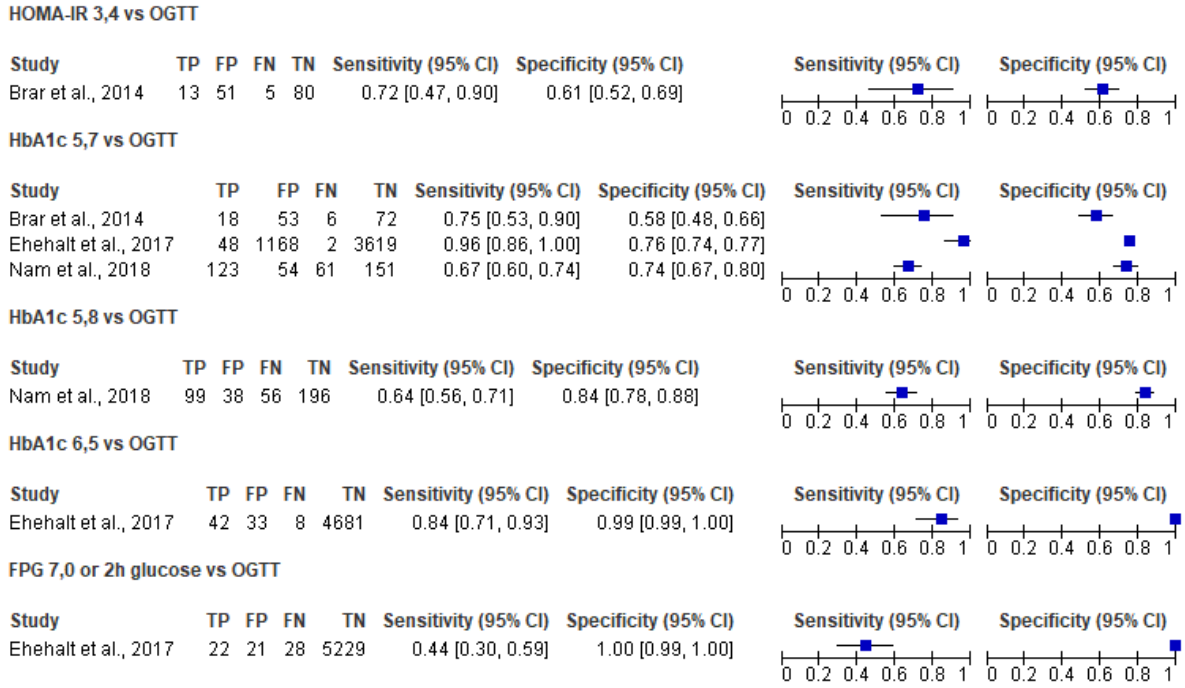


ROC curve plots the sensitivity (or true-positive fraction, TPF) versus 1-specificity (or false-positive fraction, FPF) of the tests. The theoretical best ROC curve has a square profile, yielding an AUC value of 1.0, indicating 100% sensitivity and specificity. A diagonal line from the lower left to the top-right corner would yield an AUC value of 0.5, indicating no ability to

discriminate between individual tests. These 9 tests had different ability to diagnose pre-diabetes. On the basis of ROC curves, the optimal sensitivity and specificity for and index test HbA1c were 84.00% sensitivity and 99.00% specificity with TP 42, FP 33, FN 8, TN 4681, a critical value of 6.5 versus a reference test OGTT. The curve for this test is displayed in dark blue tringle and it is seen that that it is at a value reaching the axis in point 1. As the second most accurate was test was evaluated an index test FPG at the level of cut off point 7.0 with sensitivity 44.00% and specificity 99.60% or 2-h glucose with TP 22, FP 21, FN 28 and TN 5229 versus OGTT as a reference test. The curve for this test is displayed in grey diamond and its value starts on the TPF axis around 0.67. The third an index test TG_HDL at the cut off point 1.71 with sensitivity 95.00% and specificity 69.00% with TP 132, FP 19, FN 7 and TN 43 versus HOMA-IR as a reference test. The curve for this test is displayed in orange circle and its value starts on the TPF axis at 0.3. Almost the same view had TyG as an index test at the cut off point 8.38 with sensitivity 95.00% and specificity 42.00% with TP 132, FP 36, FN 7 and TN 26. versus HOMA-IR as a reference test and HbA1c as an index test at the level of cut off point 5.7 versus OGTT as a reference test. Both of these curves displayed as red diamond (HbA1c x OGTT) and pink plus sign start at a value of approximately 0.11. The next two tests: HbA1c as an index test with the level of cut off point 5.8 with sensitivity 64.00% and specificity 84.00% with TP 99, FP 38, FN 56 and TN 196 versus OGTT as a reference test and TG_HDL as an index test with the level of cut off point 2.22 with sensitivity 90.00% and specificity 52.00% with TP 97, FP 45, FN 11 and TN 48 versus HOMA-IR as a reference test had the same view. Both of these curves displayed as light green square (HbA1c x OGTT) and turquoise star (TG_HDL x HOMA-IR start around 0.1 on the TPF axis and have an identical course. An index test HOMA-IR at the cut off point 3.4 with sensitivity 72.00% and specificity 61.00% with TP 13, FP 51, FN 5 and TN 80 versus OGTT as a reference test is closest to the diagonal curve, which is referred to as the “useless test” curve. It is displayed as a grey circle and it starts at about 0.05 of the TPF axis. The test which occurred under the AUC curve was TyG (index test) at the level of 8.5 with sensitivity 65.00% and specificity 26.00% with TP 39, FP 105, FN 21 and TN 36 versus HOMA-IR (reference test). It is marked as a yellow cross and its value is clearly below the diagonal curve.

Figure 9: Forest plot of tests: 1) HOMA-IR 3.4 vs OGTT; 2) HbA1c 5.7 vs OGTT; 3) HbA1c 5.8 vs OGTT; 4) HbA1c 6.5 vs OGTT; 9) FPG 7.0 or 2h glucose vs OGTT.

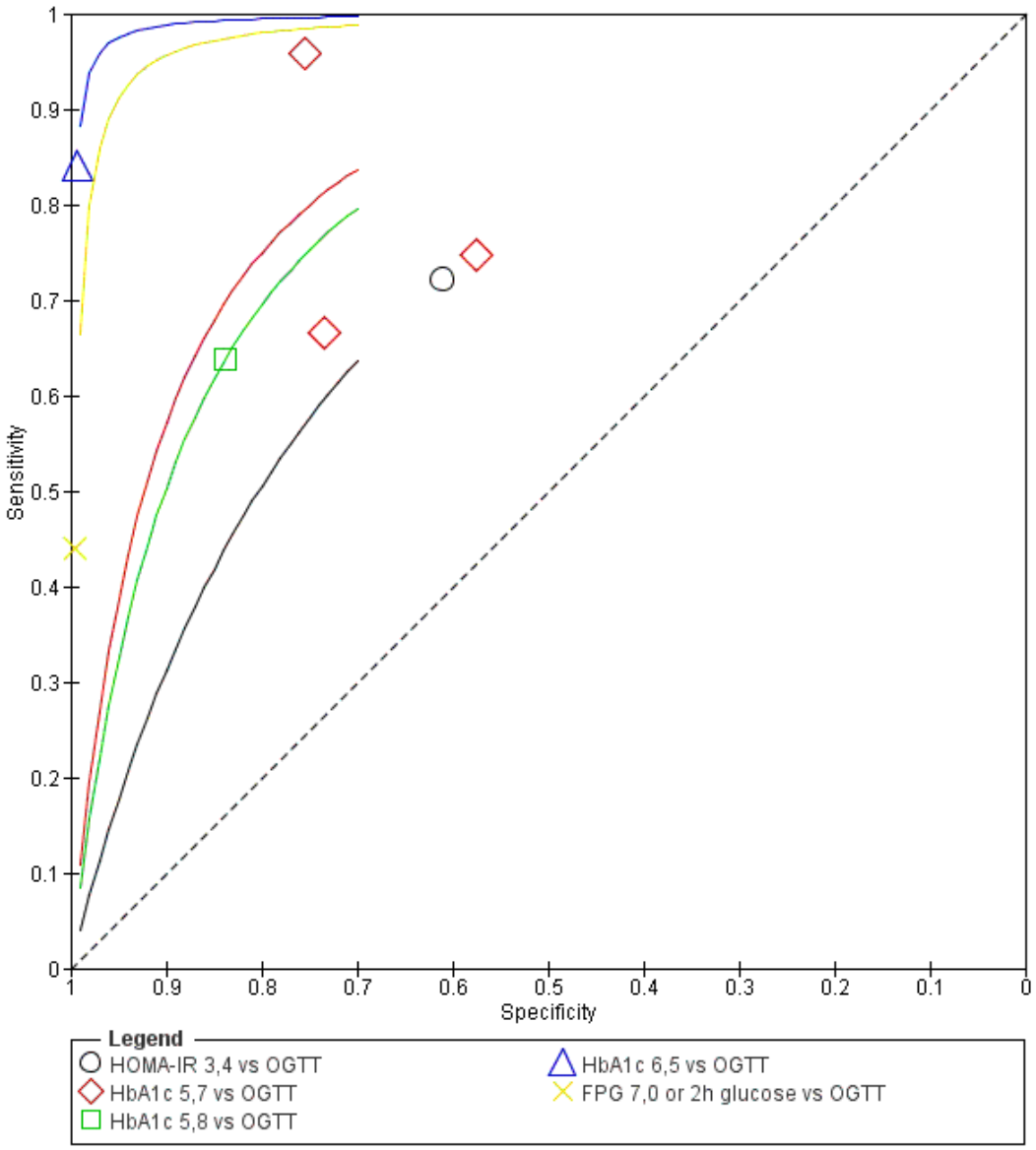
Figure 9: Sub meta-analysis (1) - forest plot of tests



In the sub meta-analysis (1) three different index tests (HOMA-IR, HbA1c and FPG 7.0 or 2-h glucose) and one reference test (OGTT) were used. Further, five different cut off points of these pairs of tests were used. Only in an index test HbA1c and reference test OGGT at the level of cut off point 5.7 was possible to be pooled meta-analysis. The rest of meta-analyses are single analysis because the comparison was done only in one study. It should also be noted that the studies showed clinical heterogeneity, as described above.

Figure 10: Summary ROC Plot of tests: 1) HOMA-IR 3.4 vs OGTT; 2) HbA1c 5.7 vs OGTT; 3) HbA1c 5.8 vs OGTT; 4) HbA1c 6.5 vs OGTT; 9) FPG 7.0 or 2h glucose vs OGTT.

Figure 10: Summary ROC Plot of tests (1)



In this sub group summary ROC plot of tests, we can see that the optimal sensitivity and specificity for and index test HbA1c were 84.00% sensitivity and 99.00% specificity, a critical value of 6.5 versus a reference test OGTT. As the second most accurate was test was evaluated an index test FPG at the level of cut off point 7.0 with sensitivity 44.00% and specificity 99.60% or 2-h glucose versus OGTT as a reference test. All tests were shown above the AUC curve.

The results of the studies show that in Brar´s study (2014), HOMA-IR as an index test and OGTT as a reference test was performed at the level of cut off point 3.4. The sensitivity was 72.00% and specificity was 61.00% with TP 13, FP 51, FN 5, and TN 80. This test is shown as a grey circle in Figure 10. The curve of this test starts on the TPF axis around 0.05 and most closely approaches the diagonal curve called the "useless test" curve.

Three studies of Brar (2014), Ehehalt (2017) and Nam (2018) searched an index test HbA1c on the level of cut off point 5.7 versus reference test OGTT. The Brar´s study (2014) had sensitivity 75.00% and specificity 58.00% with TP 18, FP 53, FN 6 and TN 72. Ehehalt´s study (2017) had sensitivity 96.00% and specificity 76.00% with TP 48, FP 1168, FN 2 and TN 3619. Nam´s study (2018) had sensitivity 67.00% and specificity 74.00% with TP 123, FP 54, FN 61 and TN 151. The values of this test are plotted as a red diamond. Its curve starts at a value of approximately 0.11. However, if we look at the individual diamonds representing the results of the 3 studies used (Brar, 2014, Ehehalt, 2017 and Nam, 2018), we find that each of the diamonds is located in a different place in the space above the diagonal curve. The HbA1c index curve with cut off point 5.7 vs OGTT does not show the most accurate result in this meta-analysis.

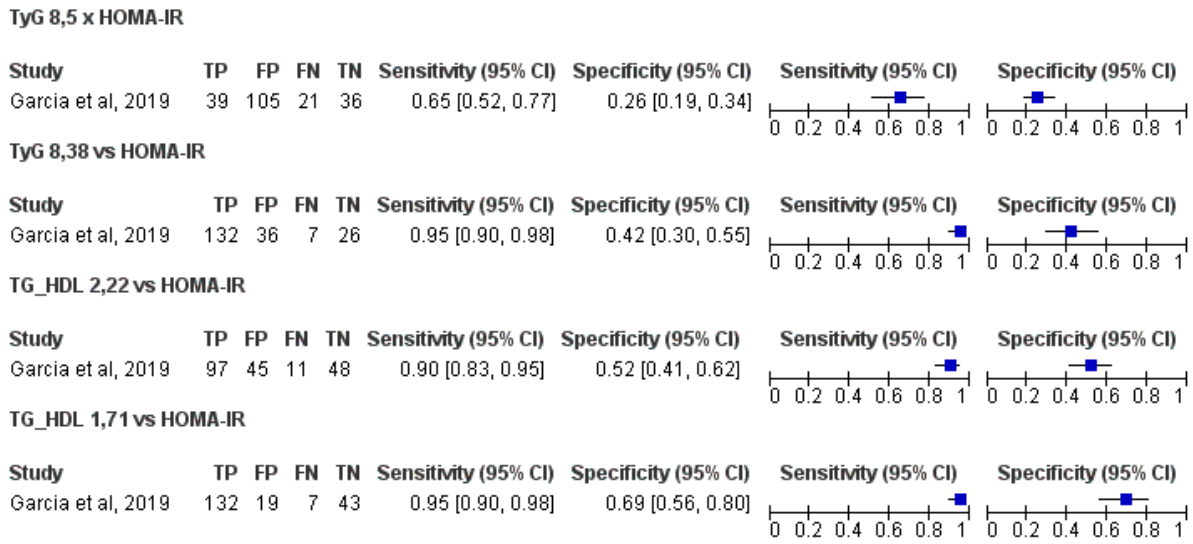
In Nam´s study (2018), cut of point of 5.8 was used for index test HbA1c versus reference test OGTT. The sensitivity was 64.00% and specificity was 84.00% with TP 99, FP 38, FN 56 and TN 196. Its representation in Figure 10 is like a light green square. It starts at a value of about 0.09 and copies the curve of the red diamond almost throughout the display (HbA1c cut off point 5.7 x OGTT).

Ehehalt´s study (2017) searched for the sensitivity and specificity of FPG or 2h glucose as an index test and OGTT as a reference test. The cut off point was determined at the level of 7.0. The sensitivity was 44.00% and specificity was 99.60% with TP 22, FP 21, FN 28 and TN 5229. In figure 10 it is shown as a yellow cross and starts on the TPF axis at a value of approximately 0.68 and from the result of the display we can see that this is the second most accurate result of these polled meta-analyses.

Study of Ehehalt (2017), was used an index test HbA1c and reference test OGTT at the level of cut off point 6.5. The sensitivity was 84.00% and specificity was 99.00% with TP 43, FP 33, FN 8 and TN 4681. The curve for this test is displayed in dark blue tringle and it is seen that that it is at a value reaching the axis in point 1. Therefore, it can be stated that in the SROC plot of listed tests with their cut off points, this test shows the most accurate value of all the tests used.

Figure 11: Forest plot of tests: 5) TyG 8.5 x HOMA-IR; 6) TyG 8.38 vs HOMA-IR; 7) TG_HDL 2.22 vs HOMA-IR; 8) TG_HDL 1.71 vs HOMA-IR

Figure 11: Sub meta-analysis (2) - forest plot of tests

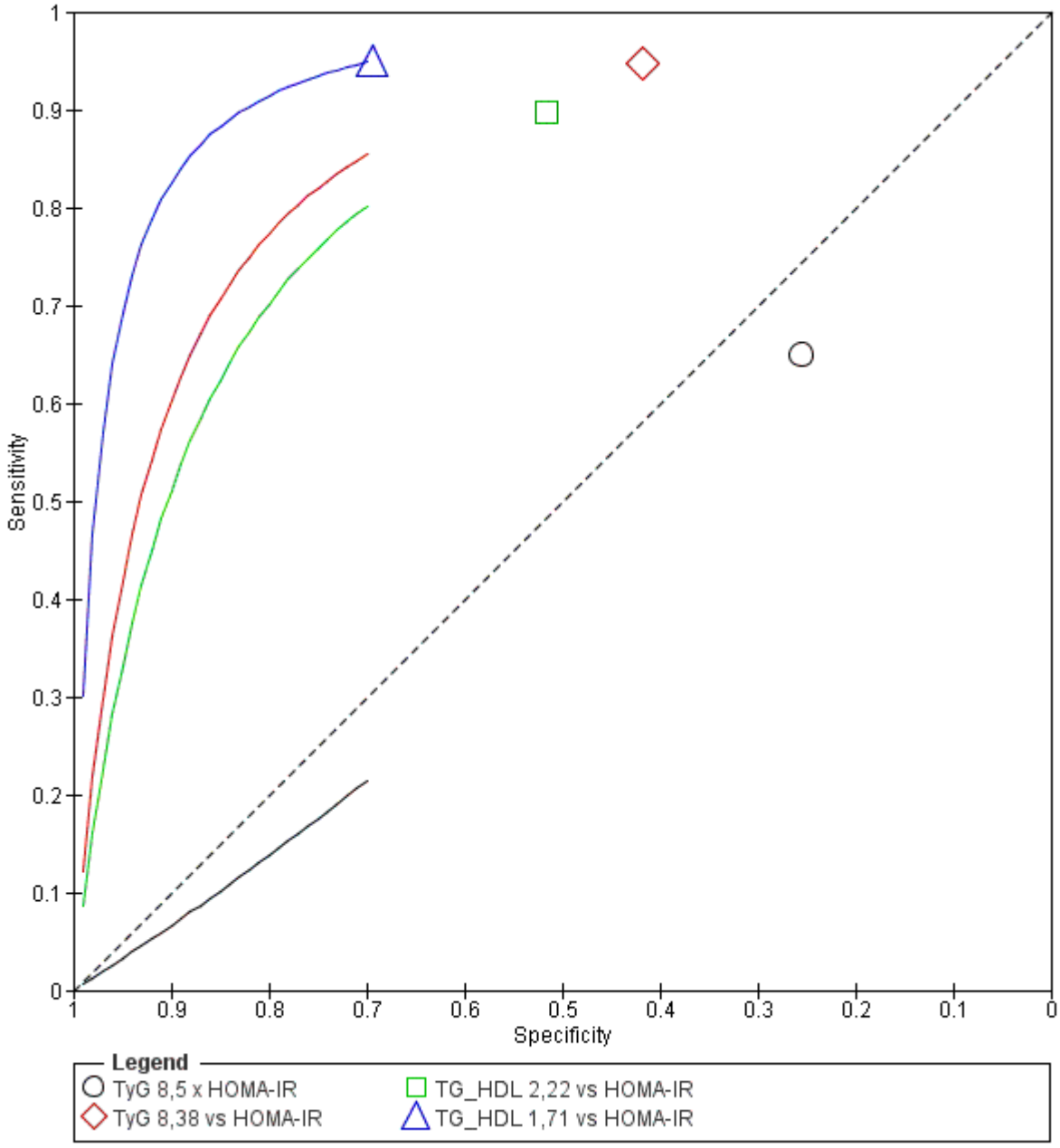


In the sub meta-analysis (2) 4 different tests with four different cut off points were used. All meta-analyses are single analysis because the comparison was done only in one study (Garcia, 2019).

Study from Garcia (2019) used TyG and TG_HDL as an index tests and HOMA-IR as a reference test at different levels of cut off points. The cut off point of 8.5 with an index test TyG and reference test had 65.00% sensitivity and 26.00% specificity with TP 39, FP 105, FN 21 and TN 36. The same tests (TyG and HOMA-IR) was used at the level of cut off point 8.38 with 95.00% sensitivity and 42.00% specificity with TP 132, FP 36, FN 7 and TN 26. In TG_HDL as an index test and HOMA-IR as a reference test the cut off point 2.22 was used. The sensitivity was 90.00% and specificity 52.00% with TP 97, FP 45, FN 11 and TN 48. The same tests (TG_HDL and HOMA-IR) was used at the level of cut off point 1.71 with 95.00% sensitivity and 69.00% specificity with TP 132, FP 19, FN 7 and TN 43.

Figure 12: Summary ROC Plot of tests: 5) TyG 8.5 x HOMA-IR; 6) TyG 8.38 vs HOMA-IR; 7) TG_HDL 2.22 vs HOMA-IR; 8) TG_HDL 1.71 vs HOMA-IR.

Figure 12: Summary of ROC Plot of tests (2)



In this summary ROC plot of tests, we can see that the optimal sensitivity and specificity for and index test TG_HDL were 95.00% sensitivity and 69.00% specificity, a critical value of 1.71 versus a reference test HOMA-IR. The curve is shown as a dark blue triangle. It starts at 0.3 of the TPF axis and most closely of all curves approaches 1.0. As the second most accurate index test was evaluated TyG at the level of cut off point 8.38 with sensitivity 95.00% and specificity 42.00% versus HOMA-IR as a reference test. The curve is in the form of a red diamond. Starts at roughly 0.11 at the TPF axis. The third light green curve represents the TG_HDL test index with a cut off point of 2.22 versus HOMA-IR. It starts slightly below 0.1 and has almost the

same course as the TyG 8.38 vs HOMA-IR curve. TyG as an index test with the level of cut off point 8.5 with sensitivity 65.00% and specificity 26.00% versus HOMA-IR is displayed as a grey circle and it can be seen that the result curve of this test was shown under the AUC curve.

2.6. Narrative description of the results of the included studies

In this chapter, we will describe narratively studies that were included in the SR DTA but in which meta-analysis could not be performed. The most common reason for avoiding quantitative synthesis are typically connected with heterogeneity. The included studies are usually too different, either statistically, methodologically or clinically (including methodological differences in interventions, metrics, results, participants and/or settings).

The biggest issues in existing studies is variability of index and reference tests across the studies and variability of thresholds. We were able to manually calculate data from (Brar et al, 2014), (Ehehalt, 2017), (Garcia, 2019) and (Nam, 2018). The rest of the results from the studies, in which the meta-analysis was not possible to be pooled, are presented using the narrative synthesis.

Therefore, the results of each study are going to be presented in individual subchapters in which brief summary of PIRD (Population, Index test, Reference test, Diagnosis of interest). The study results will be resumed. We will show the basic information about the included studies in tabular form which will be followed by the narrative description of the study results.

2.6.1. Overview of the studies included to SR DTA and significance of results

Only 5 studies (in total) provided two pairs of results in the group of **HbA1c (index test)** and **OGTT (reference test)**: Brar (2014) & Chan (2015), Brar (2014) & Nam (2018); and two pairs of results in the group of **HOMA-IR (index test)** and **OGTT (reference test)**: Atabek (2007) & Brar (2014) and Brar (2014) & Galhardo. These study results could be used for the meta-analysis because of the identical cut off points and identical index and reference tests.

The results from the other studies were difficult to compare as they had very heterogeneous combination of tests, results and determined cut off points. All of the studies were included because they represented a good source of information evidence about the diagnostic accuracy of used tests for the established diagnosis of interest of the research project.

2.6.1.1. **Atabek (2017):**

Title: Assessment of insulin sensitivity from measurements in fasting state and during an oral glucose tolerance test in obese children

Number of participants	Age groups	Tanner scale (I – V)	BMI	Index test	Cut off point (%)	Sensitivity (%)	Specificity (%)	PVV (%)	NPV (%)	Reference test
148 participants (86 girls and 62 boys)	mean age: 10.86 ± 3.08	Not known	Not known							OGTT
				HOMA-IR	2.7	80.00	59.10	-	-	
				FGIR	5.6	61.80	76.30	-	-	
				QUICKI	0.328	80.00	60.20	-	-	

Main findings:

P - 8-18 years old with BMI greater than or equal to the 95th percentile for age and gender;

I – HOMA-IR, FGIR, QUICKI;

R – OGTT;

D – pre-diabetes.

In this study, non-insulin resistant group and insulin resistant group were compared. The information about results of insulin resistant group were used in this SR DTA. The prevalence of insulin resistance, glucose intolerance and dyslipidaemia were 37, 1 %, 24, 3 % and 54 % respectively. No significant differences were detected between males and females with respect to mean age, BMI, waist-hip ratio, triglycerides, total cholesterol, HDL-cholesterol, LDL-cholesterol, systolic blood pressure and diastolic blood pressure. Hypertension was found in 21.6% (n = 32) with a significantly higher rate among the IR obese children. The mean systolic blood pressure was 118.5 ± 14.8 mm Hg and the mean diastolic blood pressure was 78.9 ± 11.0 mm Hg. The insulin sensitivity in IR obese children (FGIR 5.6 +/- 2.8, $p < 0.0001$, HOMA-IR 4.9 +/- 2.3, $p < 0.0001$, QUICKI 0.30 +/- 0.02, $p < 0.000$) and IGT (FGIR 8.2 +/- 9.1, $p = 0.834$, HOMA-IR 4.9 +/-3.3, $p = 0.003$, QUICKI 0.31 +/- 0.03, $p = 0.07$) and normo- and dislipidaemic obese children (FGIR 7.5 +/- 6.6, $p = 0.097$, HOMA-IR 4.2 +/- 2.7, $p = 0.028$, QUICKI 0.31 +/- 0.03, $p = 0.02$). Insulin resistance was positively correlated with BMI, systolic and diastolic blood pressure, triglycerides, fasting insulin, 120 min insulin, $AUC_{insulin}$ and HOMA-IR, and negatively with FGIR, ISI and QUICKI. Impaired glucose tolerance was positively correlated with fasting glucose, fasting insulin, HOMA-IR and $AUC_{glucose}$. Dyslipidaemia was positively correlated with sex, age, BMI, systolic blood pressure, fasting insulin, $AUC_{jnsulin}$ and HOMA-IR, and negatively with QUICKI. The cut-off points for diagnosis of insulin resistance were <5.6 for FGIR (sensitivity 61.8, specificity 76.3), >2.7 for HOMA-IR (sensitivity 80, specificity 59.1), and <0.328 for QUICKI (sensitivity 80, specificity 60.2) (Atabek & Pirgon, 2007).

2.6.1.2. **Brar (2014):**

Title: Screening Obese Children and Adolescents for Prediabetes and/or Type 2 Diabetes in Paediatric Practices: A Validation Study

Number of participants	Age groups	Tanner scale (I – V)	BMI (Z score)	Index test	Cut off point (%)	Sensitivity (%)	Specificity (%)	PVV (%)	NPV (%)	Reference test
149 obese patients: normal (n = 125), prediabetes (n = 2l), diabetes (n = 3)	13.8+/-3.1	Not known	BMI Z score: Normal: 2,3 +/-0,5, Pre-diabetes: 2,1 +/-0,7 , Diabetes: 2,1 +/-0,5							OGTT
				HbA1c	5.6	83.30	47.20	23.30	93.70	
					5.7	75.00	57.60	25.40	92.30	
					5.8	66.70	65.50	27.10	91.10	
					5.9	66.70	77.60	36.40	92.40	
				HOMA-IR	2.7	77.80	45.80	19.40	92.50	
					3.1	72.20	56.10	21.70	92.30	
					3.4	72.20	60.70	23.60	92.90	
					4	61.10	68.20	24.40	91.30	

Main findings:

P - Patients with a suspicion of diabetes, and/or related morbidities such as abnormal values of glucose, insulin, HbA1c, polycystic ovary syndrome, dyslipidaemia, hypertension, acanthosis nigricans, and metabolic syndrome;

I – HbA1c, HOMA-IR;

R – OGTT;

D - pre-diabetes.

Test performance of HOMA-IR for detecting prediabetes/T2DM at varying thresholds showed that a cut off point of 3.4 maintained the highest sensitivity without reducing specificity below 60%; overall test performance with AUC = 0.71 (95% CI = 0.57-0.84) was similar to HbAlc

alone, with AUC = 0.74 (95% CI = 0.61-0.87). Using the ADA-defined cut-point of 100 mg/dl, FPG had a sensitivity of 75%, excellent specificity of 100%, and the highest AUC (0.904; 95% CI = 0.81-0.99) when compared with HbAlc and HOMA-IR. The combination of HbAlc (>5.7%) and HOMAJR (≥3.4) results in a substantially higher sensitivity than either test alone, but with resulting poor specificity. Combining HbAlc (>5.7%) with FPG (>100 mg/dl) results in similarly high sensitivity while pre' serving the specificity seen with HbAlc alone. The combination of HbAlc and FPG was superior to the combination of HbAlc and HOMA-IR in terms of ability to rule out prediabetes/T2DM (LR negative 0.07 vs 0.14) and in terms of overall accuracy (AUC = 0.77 [95% CI = 0.68-0.85] vs 0.64 [95% CI: 0.53-0.75] (Brar et al., 2014).

2.6.1.3. **Bridges (2016)**

Title: Use of the triglyceride to HDL cholesterol ratio for assessing insulin sensitivity in overweight and obese children in rural Appalachia

Number of participants	Age groups	Tanner scale (I – V)	BMI (Z score)	Index test	Cut off point (%)	Sensitivity (%)	Specificity (%)	PVV (%)	NPV (%)	Reference test
223 (124 female, 99 male);	13.4 years (range, 10–17).	Not known	96,20+/-5,71							HOMA top quartile
				TRG/HDL	2.27	14.80	97.60	-	-	

Main findings:

P - 223 (124 female, 99 male); The average age of the population was 13.4 years (range, 10–17).;

I – TRG/HDL;

R – HOMA top quartile;

D - pre-diabetes.

TRG/HDL ratio correlated significantly with BMI percentile (r = 0.192, p = 0.004); insulin levels (r = 0.358, p < 0.001); and HOMA (r = 0.376, p < 0.001). There was no correlation between any of the metabolic parameters and age. The results of the regression models indicated that, although TRG/HDL ratio significantly predicted hyperinsulinemia (OR = 1.42, CI 1.18–1.70) and IR as defined by the top quartile of HOMA (OR = 1.47, CI 1.22– 1.79), the postestimation indicated only adequate prediction of the outcome variables. Model fit was improved when TRG/HDL ratio was added to a null model which contained the control variables of age, gender and BMI percentile. Likelihood ratio $\chi 2$ of the null model was 14.32 (p = 0.003) for hyperinsulinemia and 6.98 (p = 0.073) for top quartile of HOMA. Addition of TRG/HDL ratio improved these values to 30.37 (p < 0.001) and 30.36 (p < 0.001), respectively (Bridges, Jarrett, Thorpe, Baus, & Cochran, 2016).

2.6.1.4. **Ehehalt (2017):**

Title: Diabetes screening in overweight and obese children and adolescents: choosing the right test

Number of participants	Age groups	Tanner scale (I –V)	BMI (Z score)	Index test	Sensitivity (%)	Specificity (%)	PVV (%)	NPV (%)	Reference test
4848 (2668 girls)	Mean age: 13.1 ± 2.4,	not known	30.6 ± 5.4 kg/m2						OGTT
				FPG ≥ 126 mg/dl (≥7.0 mmol/l	18.00	99.80	52.90	99.10	
				FPG ≥ 126 mg/dl (≥7.0 mmol/l) and/ or 2-h glucose ≥ 200 mg/dl (≥11.1 mmol/l)	44.00	99.60	51.20	99.40	
				HbA1c ≥ 48 mmol/ mol (≥6.5%)	84.00	99.30	56.00	99.80	
				HbA1c ≥ 39 mmol/ mol (≥5.7%)	96.00	75.60	4.00	99.90	
				HbA1c ≥ 39 mmol/mol (≥5.7%) and/or FPG ≥ 100 mg/dl (≥5.6 mmol/l)	98.00	70.00	3.30	99.97	

Main findings:

P - Overweight, obese, and extremely obese children and adolescents aged 7 to 17 years;

I - FPG ≥ 126 mg/dl (≥7.0 mmol/l), FPG ≥ 126 mg/dl (≥7.0 mmol/l) and/ or 2-h glucose ≥ 200 mg/dl (≥11.1 mmol/l), HbA1c ≥ 48 mmol/ mol (≥6.5%), HbA1c ≥ 39 mmol/ mol (≥5.7%), HbA1c ≥ 39 mmol/mol (≥5.7%) and/or FPG ≥ 100 mg/dl (≥5.6 mmol/l);

R – OGTT;

D – pre-diabetes.

Dr. Wiegand was contacted twice via e-mail as a corresponding author to provide us additional data so we could pool the study results to the meta-analysis because subgroup of participants at risk for diabetes (based on the information on p 92, Table 1 in the study) was determined in the study. But we did not get any feedback from the authors.

OGTT identified 21.5% of the patients as having diabetes by using the HbA1c criteria. HbA1c identified 32% of the patients as having diabetes by using the OGTT criteria. The comparison of the classification of glucose tolerance status between OGTT and HbA1c showed significant differences ($p < 0.001$). Using both HbA1c (≥48 mmol/mol, ≥6.5%) and OGTT (FPG ≥ 126 mg/dl, ≥7.0 mmol/l and/or 2-h glucose ≥200 mg/dl, ≥11.1 mmol/l) as diagnostic criteria, 2.4% of our patients (n = 115, 55 females, mean age 14.0 ± 2.3, age range 8.3–17.9 years) could be classified as having diabetes. Within this group of 115 patients, 22.6% (n = 26) had FPG levels ≥126 mg/dl (≥7.0 mmol/l) and 68.7% (n = 79) had HbA1c ≥48 mmol/mol (≥6.5%). FPG ≥126 mg/dl (≥7.0 mmol/l) and 2-h glucose levels ≥200 mg/dl (≥11.1 mmol/l) were found in 46.1% (n = 53) of the patients, while the combination of FPG ≥126 mg/dl (≥7.0 mmol/l) and HbA1c ≥48 mmol/mol (≥6.5%) was found in 81.7% (n =94) of the patients. Based on these observations, HbA1c measurement seems to be a more promising screening method than FPG/OGTT. In this study group, however, the sensitivity for HbA1c ≥48 mmol/mol (≥6.5%) was also rather low (68.7%). Out of the 115 patients, 101 patients had HbA1c values ≥39 mmol/mol (≥5.7%) corresponding to a sensitivity of 87.8% (95% CI 80.4–93.2, n = 101). Specificity was found to be 76.3% (95% CI 75.1–77.5, n = 3612). Further analysis revealed a positive predictive value of 8.3% (6.8–10.0%, n = 101) and a negative predictive value of 99.6% (95% CI 99.4–99.8%, n = 3612). False-positive test results were found in 1121 patients, and false-negative test results were found in 14 patients. In the subgroup of IFG and IGT patients, an HbA1c cut-off level ≥39 mmol/mol (≥5.7%) detected 39.0% (95% CI 34.4– 43.7, n = 170) of IFG levels and 33.1% (95% CI 29.3–37.0, n = 198) of all IGT cases. Lowering the limit ofHbA1c from 39 mmol/mol (5.7%) to 31 mmol/mol (5.0%), 95% of all FPG ≥100 mg/dl (≥5.6 mmol/l) and of all 2-h glucose levels ≥140 mg/dl (≥7.8 mmol/l) would have been identified.

Based on the 50 patients with confirmed diabetes, the ROC analysis revealed for FPG an optimal threshold of 98 mg/dl (5.4 mmol/l) and for HbA1c a best cut-off value of42 mmol/ mol (6.0%) (Ehehalt et al., 2017).

2.6.1.5. **Galhardo (2015):**

Title: The Role of Haemoglobin A1c in Screening Obese Children and Adolescents for Glucose Intolerance and Type 2 Diabetes

Number of participants	Age groups	Tanner scale (I-V)	BMI (Z score)	Index test	Cut off (%)	Sensitivity (%)	Specificity (%)	PVV (%)	NPV (%)	Positive Likelihood Ratio	Reference test
266 patients (55,3% female)	12.3 median age (range: 8.9 to 17.6)	106 (39.9%) – pre-pubertal, 108 (40.6%) - pubertal	BMI z-score:3.35 ± 0.59								OGTT
				HbA1c	3.1	100.00	0.00	5.00	-	1.00	
					4.4	100.00	1.00	5.00	100	1.01	
					4.8	92.00	4.00	5.00	90.00	0.96	
					5.0	85.00	15.00	5.00	95.00	1.00	
					5.3 *	62.00	53.00	6.00	96.00	1.32	
					5.7[a]	23.00	89.00	9.00	86.00	2.01	
					5.9	23.00	96.00	23.00	96.00	5.75	
					6.1	8.00	99.00	30.00	95.00	8.00	
					6.3	8.00	100.00	100.00	95.00	-	
				Fasting blood glucose (mmol/l)	2.4	100.00	0.00	5.00	-	1.00	
					3.7	100.00	1.00	5.00	100.00	1.01	
					4.0	100.00	4.00	5.00	100.00	1.04	
					4.5	100.00	39.00	8.00	100.00	1.64	
					4.7 *	77.00	61.00	9.00	98.00	1.97	
					5.0	46.00	84.00	13.00	97.00	2.88	
					5.3	31.00	93.00	19.00	96.000	4.43	
					5.6[a]	8.00	98.00	17.00	95.00	4.00	
					5.7	8.00	99.00	30.00	95.00	8.00	
					7.6	0.00	100.00	-	95.00	-	
				HOMA-IR	0.1	100.00	0.00	5.00	-	1.00	
					1.1	100.00	11.00	6.00	100.00	1.12	
					3.5	92.00	53.00	9.00	99.00	1.96	
					4.0	85.00	61.00	10.00	99.00	2.18	
					4.5 *[a]	77.00	67.00	11.00	98.00	2.33	
					5.0	69.00	75.00	13.00	98.00	2.76	
					5.8	54.00	80.00	12.00	97.00	2.70	
					7.6	46.00	88.00	17.00	97.00	3.83	

		8.2	39.00	90.00	17.00	97.00	3.90	
		9.5	31.00	96.00	29.00	96.00	7.75	
		11.0	0.00	100.00	-	95.00	-	
	TG:HDL-C ratio	0.2	100.00	0.00	5.00	-	1.00	
		0.5	91.00	11.00	5.00	96.00	1.02	
		1.0	82.00	61.00	10.00	98.00	2.10	
		1.3	73.00	74.00	13.00	98.00	2.81	
		2.0	64.00	92.00	30.00	98.00	8.00	
		2.3*	27.00	96.00	26.00	96.00	6.75	
		3.0[a]	18.00	98.00	32.00	96.00	9.00	
		3.4	9.00	98.00	19.00	95.00	4.50	
		3.7	9.00	99.00	32.00	95.00	9.00	
		4.2	0.00	100.00	-	95.00	-	
	[a] Recommended cut off; * Optimized cut off							

Main findings:

P - 266 patients with 12.3 median age (range: 8.9 to 17.6 years of age);

I - HbA1c, Fasting blood glucose (mmol/l), HOMA-IR, TG:HLD_C ratio;

R - OGTT;

D – pre-diabetes.

According to the OGTT result, 253 (95.1%) patients were normoglycemic, 13 (4.9%) had pre-diabetes and no patient was diagnosed with DM2. Levels of glycated haemoglobin correlated positively with the AUC for glucose ($R^2 = 0.158$, $p < 0.001$, 95% CI 0.046 - 0.081), with OGTT ($R^2 = 0.064$, $p < 0.001$, 95% CI 0.003 - 0.010) and with fasting blood glucose levels ($R^2 = 0.021$, $p = 0.017$,95% CI 0.002-0.017). Nevertheless, the study result did not prove statistically significant difference between HbA1c geometric means both in normoglycemic or pre-diabetic patients ($p = 0.06$, 95% CI – 0.03 - 0.01). In addition, when the HbA1c level was used for pre-diabetes classification, 29 false positive and 10 false negative cases were found (and one patient with prediabetes incorrectly classified as diabetes). HbA1c's sensitivity, specificity, positive and negative predictive values and positive likelihood ratio for diagnosis of pre-diabetes, with a 5.7% cut-off value, were respectively 23.08%, 88.54%, 9.38%, 95.73% and 2.01. For this test, the area under the ROC curve was 0.59 (95% CI 0.40 - 0.78), showing its lack of discrimination power. In addition, fasting blood glucose level ($R^2 = 0.192$ $p < 0.001$, 95% CI 0.068-0.112), HOMA-IR ($R^2 = 0.042$, $p = 0.001$, $p = 0.001$, 95% CI 0.016-0.060) and TG: HDL-C ratio ($R^2 = 0.024$, $p = 0.017$, 95% CI 0.001-0.013) also correlated positively with glucose's AUC. Finally, unlike what was found regarding glycated haemoglobin, a statistically significant difference was found between the mean values of these three parameters in normoglycemic vs. pre-diabetic group of patients as well as a higher power of diagnostic discrimination shown by their ROC curves (Galhardo & Shield, 2015).

2.6.1.6. **Garcia (2019):**

Title: Diagnostic accuracy of triglyceride/glucose and triglyceride/HDL index as predictors for insulin resistance in children with and without obesity

Number of participants	Age groups	Tanner scale (I – V)	BMI (Z score)	Index test	Cut off (%)	Sensitivity (%)	Specificity (%)	PV V (%)	NP V (%)	Reference test
201 participants (42.78% male)	Median age: 8 years (range 5-9).	Not known	Not known							HOMA-IR
				TyG	8.5	65.00	25.70	-	-	
				Proposed TyG	8.38	95.00	42.30	-	-	
				TG/HDL	8.18	77.40	64.80	-	-	
				Proposed TG/HDL	1.17	95.00	68.60	-	-	

Main findings:

P - 5 and 9 years old; According to the percentile tables of the CDC corresponding to BMI and age, two groups were constituted: group with obesity-overweight (OO Group): 85th percentile (n = 97) and group with normal weight (NW Group): <percentile 85 (n = 104)

I – TyG, Proposed TyG, TG/HDL, Proposed TG/HDL;

R - HOMA-IR;

D – pre-diabetes.

The median of HOMA-IR was 1.51 (range 0.21-38.45), median of TyG: 8.32 (range 7.20-9.92) and median of TG/HDL 2.17 (range 0.43-12.29). These indexes were compared between Obese-Overweight group (OO Group) and Normal Weight Group (NW Group), and they were statistically significant ($P < 0.05$). Receiver Operating Curves (ROC) was performed and the sensitivity, specificity and of the cut off point of the established TyG and TG/HDL were recorded. Positive Predictive Value (PPV) of TyG =8.5: 21.66%; Negative predictive Value (NPV) TyG = 8.5: 4.96%; accuracy diagnostic TyG = 8.5: 73.13%. PPV TyG = 8.38: 13.66%; NPV TyG = 8.38: 1.61%; accuracy diagnostic TyG = 8.38: 61.19%. PPV TG/HDL = 2.22: 16.66% and NPV TG/THDL = 2.22: 2.15%; accuracy diagnostic TG/HDL = 2.22: 54.22%; PPV TG/HDL = 1.71: 13.66% and NPV TG/THDL =1.71: 1.61%, accuracy diagnostic TG/HDL =1.71: 39.80%. To evaluate the magnitude of each cardiometabolic risk factor with TyG and TG/HDL in comparison with HOMA was realized Odds Ratio and the results. ROC was analysed for each group (OO and NW) but there was no change in sensitivity and specificity values. No significant adverse events occurred while blood sample collected. Only two children

had a mild hematoma that disappeared in three days (Garcia, Urbina Trevino, Villalpando Sanchez, & Aguilar, 2019).

2.6.1.7. **Chan (2016):**

Title: Screening for type 2 diabetes and prediabetes in obese youth: evaluating alternate markers of glycemia-1,5-anhydroglucitol, fructosamine, and glycated albumin

Number of participants	Age groups	Tanner scale (I – V)	BMI (Z score)	Index test						Reference test
119 participants (62% female)	Median age: 14.1 (range: 10-18)	Not known	BMI z-score: 2.3 (range: 1.1-3.0)							
					2hG<140 mg/dl^{-1}	**2hG 140-199 mg/dl^{-1}**	**2hG≥200 mg/dl^{-1}**	**P - value**	**Adjusted p-value**	**2h-glucose category**
				FA*	209.0	208.0	226.0	0.0374	0.0051	
				GA*	11.0	11.0	14.0	0.0092	<0.0001	
				1.5 AG*	24.5	23.0	7.0	0.0063	0.0006	
					<5.7	**5.7-6.4**	**>6.4**	**P - value**	**Adjusted p-value**	**HbA1c**
				FA*	207	210	234	0.0024	0.0001	
				GA*	11	12	15	<0.0001	<0.0001	
				1.5 AG*	25.3	22.6	5.7	0.0034	<0.0001	
*FA = Fructosamin; GA = Glycated albumin; 1.5 AG = 1.5-anhydroglucitol										

Main findings:

P - Eligible participants included youth 10–18 yr of age with a body mass index (BMI) ≥85th‰;

I – Fructosamin, Glycated albumin, 1.5-anhydroglucitol;

R - 2h-glucose category, HbA1c;

D – pre-diabetes.

Dr. Chan was contacted twice via e-mail as an author to provide us additional data in 2x2 table so we could pool the study results to the meta-analysis because the data about sensitivity and specificity in the study were missing. But we did not get any feedback from the authors.

Approximately half of the participants were dysglycemic based on either 2hG ≥140mgdL−1 (40.2%) or HbA1c ≥5.7% (51.3%), whereas only 9% were dysglycemic by FPG ≥100mgdL−1. Median (min.–max.) values for all glycemic measures were as follows: FPG 86mgdL−1 (87–130mgdL−1), 2hG 131mgdL−1 (81–239mgdL−1), HbA1c 5.7% (4.9–7.7%), FA 209 μmolL−1

(169–270 μmolL−1), GA 11% (9–17%), and 1,5-AG 24.1mcgmL−1 (2.6–41mcgmL−1). ROC curves were generated to determine cut points for the alternate markers that optimized sensitivity and specificity for detecting 2hG ≥200 and ≥140mgdL−1, as well as HbA1c ≥6.5 and ≥5.7%. The alternate markers had similarly low ROC-AUCs for identifying prediabetes by 2hG and HbA1c. The ROC AUCs of the alternate markers were higher, however, for identifying diabetes by 2hG at the following cut points: FA≥219 μmolL−1, GA≥12%, 1,5-AG≤19.8 mcgmL−1; and for identifying diabetes by HbA1c (Fig. 1C) at the following cut points: FA≥219 μmolL−1,GA≥14%, and 1,5AG≤7mcgmL−1. For predicting 2hG ≥200mgdL−1, HbA1c had the highest ROC-AUC (0.97) at a cut off point of 6.1%, FA had the lowest (0.85), and these differences were statistically significant (p=0.047). The ROC-AUCs for predicting diabetes by 2hG among HbA1c (0.97), 1,5-AG (0.87) and GA (0.92), were not significantly different (Chan et al., 2016).

2.6.1.8. **Chan (2015):**

Title: Continuous glucose monitoring and its relationship to haemoglobin A1c and oral glucose tolerance testing in obese and prediabetic youth

Number of participants	Age groups	Tanner scale (I – V)	BMI (Z score)	Index test	Cut off	Sensitivity (%)	Specificity (%)	PVV (%)	NPV (%)	Reference test
98 participants (35.7% male)	Median age: 14.1 (range: 10.6-14.4)	I.-6.1%; II.-11.2%; III.-13.3; IV.-14.3; V.-55.1%	BMI z-score 2.3 (1.1-3.0)							N/A
				HbA1c	5.9 %	80.00	64.00			
				FPG	92 mg/dL	80.00	50.00			
				2-h glucose	153 mg/dL	69.00	79.00			

Main findings:

P - Males and females 10–18 years of age with a body mass index (BMI) in the 85th percentile or greater;

I – HbA1c, FPG, 2-h glucose;

R – N/A;

D – pre-diabetes.

Dr. Chan was contacted twice via e-mail as an author to provide us additional data in 2x2 table so we could pool the study results to the meta-analysis because the data about sensitivity and specificity in the study were missing. But we did not get any feedback from the authors.

CGM data were successfully collected on 98 obese youth. Median FPGs in normal and prediabetes HbA1c categories were normal (83 mg/dL and 91 mg/dL, respectively). Only 9 individuals had elevated FPG of 100 mg/dL or greater, with only one greater than 125 mg/dL; thus, CGM comparisons were not made by FPG category. 36 individuals had FPG of at least 90 mg/dL. 35 youth had 2-hour glucose in the prediabetes range (140–199 mg/dL) and 5 had 2-hour glucose of 200 mg/dL or greater. The subgroups of obese adolescents with normal HbA1c and with normal 2-hour glucose, spent 17% and 20% of the time greater than 120 mg/dL, respectively. Time spent greater than 140 mg/dL for those with normal HbA1c and those with normal 2-hour glucose, however, were only 1.2% and 1.3%, respectively. However, when categorized by HbA1c or 2-hour glucose, there were significant differences between normal glycaemic youth and youth with prediabetes for CGM outcomes. When categorized by HbA1c, differences were highly significant ($P < .0001$) for night average, night peak glucose, and night AUC, but not for CGM SD or excursions above200mg/dL. When categorized by2-hour glucose, differences were highly significant ($P< .0001$) for percentage of time at 140 mg/dL or greater, but not for night-average glucose, minimum-sensor glucose, and night AUC.

The magnitudes of correlation for HbA1c with average-sensor glucose, night-average sensor glucose, minimum sensor glucose, and AUC were greater than for 2-hour glucose. But the magnitudes of correlation between 2-hour glucose and peak-sensor glucose, SD, excursions greater than 140 mg/dL and greater than 200 mg/dL, and percentage of time spent greater than 140 mg/dL and greater than 200 mg/dL were greater than for HbA1c.

Logistic regression models were used to assess the ability of HbA1c, FPG, and 2-hour glucose to predict an abnormal CGM AUC by comparing ROC curves in the remaining (dysglycemic) patients. Two-hour glucose had the highest area under the ROC curve at 0.78, although there was no statistically significant difference between 2-hour glucose and the other variables for predicting abnormal CGM AUC.A cut off of 153 mg/dL maximized sensitivity (69%) and specificity (79%) for 2-hour glucose in predicting abnormal CGM AUC. A cut off of 5.9% for HbA1c maximized sensitivity (80%) and specificity (64%) for predicting abnormal CGM AUC. An FPG cut off of 92 mg/dL maximized sensitivity (80%) and specificity (50%) for predicting abnormal CGM AUC (Chan et al., 2015).

2.6.1.9. **Kang (2017)**

Title: Triglycerides/glucose index is a useful surrogate marker of insulin resistance among adolescents

Number of participants	Age groups	Tanner scale (I – V)	BMI (Z score)	Index test	Cut off (%)	Sensitivity (%)	Specificity (%)	PVV (%)	NPV (%)	Reference test
221 participants (168 males, 53 females)	Mean age: 11.1 +/- 1.5 yrs (range: 9-13)	Not known	BMI (with IR): 24.0 +/- 4.5; without IR: 19.9 +/- 3.6							HOMA-IR
				TyG	8.18	77.40	64.80	-	-	
				Tg/HDL	1.41	72.70	61.80	-	-	

Main findings:

P - Non-diabetic subjects aged 9–13 years from one middle and two elementary schools; 168 males and 53 females with a mean age of 11.1 ± 1.5 years, their BMI classified 16;

I – TyG, Tg/HDL;

R – HOMA-IR;

D – pre-diabetes.

HOMA-IR had a significant positive correlation with the TyG index ($r = 0.41$, $P < 0.001$), and TG/HDL-C ($r = 0.40$, $P < 0.001$) showed a similar correlation. The TyG index and TG/HDL-C showed a strong positive correlation ($r = 0.84$, $P < 0.001$). The best cut-offs of the TyG index and TG/HDL-C for insulin resistance diagnosis were 8.18 and 1.16, respectively. The area under the ROC curve for the TyG index was 0.734 (95% CI: 0.671 – 0.791) and showed no difference compared with the AUC of 214 TG/HDL-C (0.736, 95% CI: 0.673 – 0.793; P= 0.944). The ROC curve for the TyG index represented good sensitivity (77.3%) and specificity (68.3%) (Kang et al., 2017).

2.6.1.10. **Kasturi (2019)**

Title: Two- vs one-hour glucose tolerance testing: Predicting prediabetes in adolescent girls with obesity

Number of participants	Age groups	Tanner scale (I – V)	BMI (Z score)					
93 female participants	Age: 14.8 ± 1.6 yrs, range: 12-17 years	Not known	BMI: 32,6 +/-6,5					
	Baseline			6-weeks			Reproducibility	
OGTT Feature	Sensitivity	Specificity	ROC-AUC (95% CI)	Sensitivity	Specificity	ROC-AUC (95% CI)	Kappa	95 CI
Morphological features								
Glucose peak > 30 minutes	0.70	0.70	0.71 (0.55-0.86)	0.58	0.62	0.60 (0.45-0.76)	0.23	0.02-0.44
Monophasic curve	0.60	0.53	0.57 (0.40-0.74	0.58	0.48	0.53 (0.37-0.69)	0.23	0.02-0.43
1-hr glucose 155 mg/dL	0.40	0.95	0.67 (0.51-0.83)	0.33	0.98	0.66 (0.52-0.79)	0.42	0.07-0.77
COMBO	0.66	0.75	0.71 (0.56-0.85)	0.41	0.73	0.57 (0.42-0.72)	0.46	0.30-0.61
Gold standard criteria								
Fasting glucose 100 mg/dL							0.38	0.04-0.72
2-hr glucose 140 mg/dL							0.28	-0.08-0.64

Main findings:

P - Youth females with a first or second degree relative with type 2 diabetes and mild or moderate depressive symptoms. Youth females had overweight/obesity; age 14.8 ± 1.6 years, range: 12-17 years) who had overweight/obesity (body mass index [BMI] ≥ 85th percentile;

I – Glucose peak > 30 minutes, Monophasic curve, 1-hr glucose 155 mg/dL, COMBO;

R – Fasting glucose 100 mg/dL, 2-hr glucose 140 mg/dL ;

D – pre-diabetes.

This study was focused on a reproducibility and predictive ability of a morphological feature of the glucose curve (monophasic curve, glucose peak >30mins and 1-hr glucose 155mg/dL). That is the reason why it was necessary to change the tabular form of the results when describing the data obtained from the study.

The reproducibility and diagnostic accuracy between baseline and 6-weeks, κ coefficient was 0.48 for the morphological features of the OGTT. The percentage of youth with prediabetes

(12%) was the same at baseline and 6-weeks, (P=0.76). Six girls diagnosed with prediabetes at baseline were reclassified as NGT at 6-weeks, while 8 girls who were NGT at baseline were reclassified as prediabetes at 6-weeks. The ROC-AUCs of OGTT morphological features were not significantly different when compared at baseline or at 6-weeks (P 0.21). The predictive ability of baseline OGTT parameters at 1-year follow up showed that among the 72 girls examined, 11 had prediabetes at baseline (15%) and 7 (10%) had prediabetes at 1-year. From baseline to 1-year, 4 girls maintained a prediabetes diagnosis; the diagnosis of prediabetes resolved in 7 girls; and 3 girls were newly diagnosed with prediabetes at 1-year. Compared to the gold-standard OGTT criteria the area under the ROC curve (ROC-AUC: 0.73, 95% CI: 0.53–0.93) was significantly lower for monophasic curve variable (0.42, 95% CI: 0.22–0.63, P<0.001) but not different for 1-hr glucose 155mg/dL (0.67, 0.48–0.88), glucose peak >30mins (0.68, 0.49–0.87) or COMBO (0.77, 0.62–0.93). There was no difference between ROC-AUC for glucose peak, 1-hr glucose, and COMBO parameters (P=0.39) (Kasturi et al., 2019).

2.6.1.11. **Keskin (2005):**

Title: Homeostasis Model Assessment Is More Reliable Than the Fasting Glucose/Insulin Ratio and Quantitative Insulin Sensitivity Check Index for Assessing Insulin Resistance Among Obese Children and Adolescents

Number of participants	Age groups	Tanner scale (I – V)	BMI (Z score)	Index test	Cut off (%)	Sensitivity (%)	Specificity (%)	PVV (%)	NPV (%)	Reference test
57 (30 girls and 27 boys)	12.04 +/- 2.90 years	not known	BMI: with IR: 31.29 +/- 5.86; without IR: 28,23 +/- 4.94							OGTT
				HOMA-IR	3.16	76.00	66.00	-	-	
				QUICKI	-	-	-	-	-	
				FGIR	-	-	-	-	-	

Main findings:

P - 57 (30 girls and 27 boys) mean age: 12.04 +/- 2.90 years; mean BMI: 29.57 +/- 5.53);

I – HOMA-IR, QUICKI, FGIR;

R – OGTT;

D – pre-diabetes.

Prof. Keskin was contacted twice via e-mail as an author to provide us missing data about the tests included in the study (HOMA-IR, QUICKI and FGIR) so we could pool the study results to the meta-analysis. But we did not get any feedback from the authors.

The mean fasting glucose level was 82.67 +/- 9.23 mg/dL (range: 65-106 mg/dL), the mean fasting insulin level was 26.98 +/- 22.49 μU/mL (range: 1.45-109.72 μU/mL), and the mean sum of insulin levels was 447.32 +/- 145.22 μU/mL (range: 300.24-744.39 μU/mL) for the group with insulin resistance; the mean fasting glucose level was 80.44 +/- 10.51 mg/dL (range: 61-105 mg/ dL), the mean fasting insulin level was 16.65 +/- 13.85 μU/mL (range: 1.40-51.47 μU/mL), and the mean sum of insulin levels was 154.08 +/- 77.78 μU/mL (range: 24.86-275.00 μU/mL) for the group without insulin resistance. There were significant differences in the mean HOMA-IR (6.06 +/- 4.98 and 3.42 +/- 3.14, P < .05) and QUICKI (0.313 +/- 0.004 and 0.339 +/- 0.004, P < .05), but not FGIR, values between the 2 groups. Sensitivity and specificity calculations were based on insulin resistance with ROC analysis. HOMA had high sensitivity and specificity for measuring insulin resistance. The present HOMA cut off point for diagnosis of insulin resistance of 3.16 yielded a sensitivity of 76% and a specificity of 66% (Keskin, Kurtoglu, Kendirci, Atabek, & Yazici, 2005).

2.6.1.12. **Kim (2019):**

Title: Comparison of HbA1c and OGTT for the diagnosis of type 2 diabetes in children at risk of diabetes

Number of participants	Age groups	Tanner scale (I – V)	BMI (Z score)	Index test	Cut off	Non DM	DM	Total	Sensitivity (%)	Specificity (%)	PVV (%)	NPV (%)
190 (52.1 % female)	Age (yrs): 12.56+/- 3.44	Not known										
				HbA1c	≤6.5%	143	5	148	89.4	100.00	100.00	96.6
					≥6.5%	0	42	42	-	-	-	-
					Total	143	47	190	-	-	-	-
				OGTT	<200 mg/dL	143	17	160	63.8	100.00	100.00	89.4
					≥200 mg/dL	0	30	30	-	-	-	-
					Total	143	47	190	-	-	-	-
				FPG	<126 mg/dL	143	7	150	85.1	100.00	100.00	95.3
				FPG	≥126 mg/dL	0	40	40	-	-	-	-
					Total	143	47	190	-	-	-	-

Main findings:

P – 190 participants divided into 3 groups: normal glucose tolerance (NGT; n = 117), impaired glucose tolerance (IGT; n = 33), and diabetes (DM; n = 40) according to the OGTT.

I – HbA1c, OGTT, FPG;

R – N/A;

D – pre-diabetes.

In this study, the participants were divided into 3 groups based on OGGT performed at the beginning. The study is mainly focused on DM predictors of stated diagnostic tests but the descriptive data are available for the group of non-diabetic population.

Based on the OGTT performed on 190 subjects, 33 (17.4%) were diagnosed with IGT, and 40 (21.1%) were diagnosed with DM. The remaining 117 students (61.6%) were diagnosed with NGT. The mean age and BMI for all subjects were 12.6 years and 24.5 kg/m2, respectively. The BMI values were comparable across the 3 groups. There was an increased prevalence of female and elderly subjects among the IGT and DM groups compared to the NGT group. The mean FPG and 2-h OGTT levels in these groups were significantly higher than among subjects with NGT ($p < 0.001$). The average HbA1c level among all subjects was 6.3 +/- 1.8%. As expected, HbA1c levels were significantly higher in the DM group (9.0%) than in the IGT (6.1%) or NGT group (5.5%). Furthermore, c-peptide and HOMA-IR levels in the DM group were significantly higher than in the NGT and IGT groups.

Based on HbA1c, patients were categorized into 3 groups by the ADA criterion of HbA1c as follows: NGT group, 107 (55.3%) subjects; at risk for DM group, 41 (21.6%) subjects; DM group, 42 (22.1%) subjects. Although the majority (83.2%) of subjects with an HbA1c < 5.7% were classified as having NGT according to the OGTT, 16.8% of subjects were classified as having IGT or DM. Of 41 subjects in at the risk for DM group (HbA1c 5.7e6.4%), 27 (65.9%) were categorized as having NGT and 3 (7.3%) as having DM according to the OGTT. The majority (83.3%) of subjects with an HbA1c ≥6.5% were classified as having DM by the OGTT, while only 1 subject (2.4%) was classified as having NGT. Therefore, of 42 subjects with DM according to HbA1c criterion, 7 (16.7%) did not meet the standard diagnostic criteria for DM. The subjects with HbA1c ≥ 6.5% are evaluated to identify any differences in clinical parameters between subjects with DM (n = 35) and without DM (n = 7) according to the OGTT. However, there were no differences in age, sex difference, HbA1c level, serum c-peptide, HOMA-IR and cholesterol profile between the 2 groups except the FPG and 2-h OGTT (no data).

The diagnostic accuracies of HbA1c, 2-h OGTT and FPG criteria were evaluated by calculating a ROC curve with the 95% confidence intervals. The AUCs for each diagnostic criterion were 0.970 for HbA1c, 0.939 for FPG, and 0.977 for 2-h OGTT. These results indicate that HbA1c levels are a good screening tool in children at risk of developing pre-diabetes.

2.6.1.13. **Kurtoğlu (2010):**

Title: Insulin resistance in obese children and adolescents: HOMA-IR cut-off levels in the prepubertal and pubertal periods

Number of participants	Age groups	Tanner scale (I – V)	BMI (Z score)	Index test	Cut off (%)	Sensitivity (%)	Specificity (%)	PVV (%)	NPV (%)	Reference test
268 participants (141 girls, 127 boys)	Age: 5-18; Mean age: *pPB: 8.9+/-1.8; *pPG: 8.3+/-1.4; *PB: 13.6+/-1.6; *PG: 13.2+/-2.0.	*pPB: 46.2%; *pPG: 25.5%; *PB: 63.8%; *PG: 74.5%	BMI *pPB: 28.2±5.4; *pPG: 26.2±5.8; *PB: 30.9+/-4.9; *PG: 30.4+/-5.0.							OGTT
				HOMA-IR (*pPB)	2.67	88.20	65.50	-	-	
				HOMA-IR (*pPG)	2.22	100.00	42.30	-	-	
				HOMA-IR (*PB)	5.22	56.00	93.30	-	-	
				HOMA-IR (*PG)	3.82	77.10	71.40	-	-	
*pPB = prepubertal boys, pPG = prepubertal girls, PB = pubertal boys, PG = pubertal girls										

Main findings:

P - Children with obesity aged between 5 and 18 years (Age: 5-18; Mean age: *pPB: 8.9+/-1.8; *pPG: 8.3+/-1.4; *PB: 13.6+/-1.6; *PG: 13.2+/-2.0.);

I – HOMA-IR ;

R – OGTT;

D – pre-diabetes.

Chronological ages, BMI values, fasting blood sugar and insulin values, blood sugar and insulin values at 120 minutes, total insulin values measured during OGTT, FGIR and HOMA-IR values were calculated according to gender and pubertal status. Following OGTT, the rate of insulin resistance in the prepubertal period was 37% (n=17) in boys and 27.8% (n=10) in girls. In the pubertal children, these rates were 61.7% (n=50) in boys and 66.7% (n=70) in girls. There was not any difference in pre- and post-prandial blood glucose level in boys neither in the prepubertal nor in the pubertal groups when comparing the hyperinsulinemic and nonhyperinsulinemic groups for blood glucose levels at 0 and 120th minutes of OGTT; the same at girls´ groups. In pubertal girls, there was a significant difference between 0- and 120-

minute blood glucose levels between insulin resistant and non-resistant groups. HOMA-IR, fasting and 120-minute insulin values, FGIR and total insulin values were significantly different between the subjects of both sexes with and without insulin resistance both in the prepubertal and pubertal groups. HOMA-IR cut-off values for insulin resistance were calculated to be 2.67 (sensitivity 88.2%, specificity 65.5%) in boys and 2.22 (sensitivity 100%, specificity 42.3%) in girls in the prepubertal period, and 5.22 (sensitivity 56%, specificity 93.3%) in boys and 3.82 (sensitivity 77.1%, specificity 71.4%) in girls in the pubertal period. Fasting insulin levels above 15 μU/mL in the prepubertal period, 30 μU/mL in the pubertal period and 20 μU/mL in the post pubertal period, FGIR above 6, 120-minute insulin >75 μU/mL during OGTT and peak insulin above 150 μIU/mL are recommended as cut-off levels for hyperinsulinism and consequently as parameters showing insulin resistance. Fasting insulin, insulin and blood sugar at 120th minute, FGIR and HOMA-IR values were compared using the ROC analysis regarding their importance in determination of insulin resistance according to pubertal status and gender. With the exception of prepubertal girls, HOMAIR index was found to be the best determinant of insulin resistance in sub groups, insulin level at 120th minute was the best indicator of insulin resistance in the prepubertal girls (Kurtoglu et al., 2010).

2.6.1.14. **Lee (2019):**

Title: Discrepancies between Glycosylated Haemoglobin and Fasting Plasma Glucose for Diagnosing Impaired Fasting Glucose and Diabetes Mellitus in Korean Youth and Young Adults

Number of participants	Age groups	Tanner scale (I – V)	BMI (Z score)	Index test	Cut off (%)	Sensitivity (%)	Specificity (%)	PVV (%)	NPV (%)	Reference test
4129 (54.6% male)	10-19 yrs	Not known	Not known							FPG
				HbA1c	5.5	49.90	73.30	14.20	94.30	
					5.7	35.00	83.90	16.10	93.60	
					5.9	100.00	95.80	5.34	100.00	
					6.5	72.20	99.90	72.20	99.90	

Main findings:

P - 4,129 (45.1%) in the youth group (10 to 19 years of age);

I – HbA1c;

R – FPG;

D – pre-diabetes.

In this study, two research groups were assessed: the group 10-19y and 20-29y. Based on the inclusion criteria of the SR DTA population we describe only data for the age group 10-19y. This study represented 45.1 % of the total number of participants.

In the ROC curve analysis, the AUC (95% CI) for detecting IFG based on HbA1c level was 0.649 (95% CI, 0.648 to 0.650) for the youth group. The optimal HbA1c cut off point for diagnosing IFG was 5.6% (sensitivity 49.9%, specificity 73.3%) in the youth group. The AUC

(95% CI) for detecting DM_{FPG} based on HbA1c level was 0.996 (95% CI, 0.996 to 0.996) for the youth group. The optimal HbA1c cut off point for diagnosing DM_{FPG} was 5.9% (sensitivity 100%, specificity 95.8%) in the youth group. By using nationally representative survey data, we assessed the diagnostic performance of HbA1c cut off values recommended for the diagnosis of IFG and DM_{FPG} ≥6.5% for DM_{FPG} resulted in a sensitivity of 72.2% and a specificity of 99.9% in the youth group. However, the HbA1c cut off of ≥5.7% for IFG had a lower sensitivity and specificity than did the HbA1c cut off of 6.5% for diabetes in both groups. In the present study, the HbA1c cut off values that best coincided with the DM_{FPG} were 5.9% in the young group by using ROC curve analysis. The optimal HbA1c cut off levels for detecting IFG in the youth was 5.6% (J. Lee et al., 2019).

2.6.1.15. **Liang (2015):**

Title: Triglycerides and high-density lipoprotein cholesterol ratio compared with homeostasis model assessment insulin resistance indexes in screening for metabolic syndrome in the Chinese obese children: a cross section study

Number of participants	Age groups	Tanner scale (I – V)	BMI (Z score)	Index test	Cut off (%)	Sensitivity (%)	Specificity (%)	PVV (%)	NPV (%)	Reference test
976 participants (female: 286, male 690)	Age group <10 years 349, >= 10 years: 627	Pubertal stage: Prepubertal: 458, pubertal: 518	Not known							OGTT
				HOMA1-IR	>4.59	58.70	65.50	-	-	
				HOMA2-IR	>2.76	53.20	69.50	-	-	
				TG/HDL-C	>1.25	80.00	75.00	-	-	

Main findings:

P - 976 participants (female: 286, male 690), Total: sex (F/M): 286/690; age group: <10 years 349, >= 10 years: 627; Prepubertal: 458, pubertal: 518;

I – HOMA1-IR, HOMA2-IR, TG/HDL-C;

R – OGTT;

D – pre-diabetes.

The participants were divided into two groups – non-metabolic syndrome strata and metabolic syndrome strata.

The TG/HDL-C ratio was a better predictor of MS (acceptable sensitivity and specificity and higher AUC-ROC) than either HOMA1-IR or HOMA2-IR. The cut off values for MS were: TG/HDL-C ratio > 1.25 (sensitivity: 80 %; specificity: 75 %), HOMA1-IR > 4.59 (sensitivity: 58.7 %; specificity: 65.5 %) and HOMA2-IR > 2.76 (sensitivity: 53.2 %; specificity: 69.5 %).

After stratified by age group, puberty stage and sex, the cut offs of HOMA1-IR changed from 3.58–5.74 while the cut offs of HOMA2-IR fluctuated from 1.92–2.99. However, the cut offs of TG/HDL-C varied slightly from 1.21–1.53. The Overall AUC-ROC values for the prediction of MS were 0.640, 0.625, and 0.843 by HOMA1-IR, HOMA2-IR and TG/HDL-C respectively. Significant difference of the AUC-ROC values between HOMA-IR and TG/HDL-C was found with a higher sensitivity and specificity. When stratified by age group, gender and puberty stage the AUC-ROC values for the prediction by HOMA-IR were still lower than those by TG/HDL-C (Liang et al., 2015).

2.6.1.16. **Maffeis (2010):**

Title: Fasting Plasma Glucose (FPG) and the Risk of Impaired Glucose Tolerance in Obese Children and Adolescents

Number of participants	Age groups	Tanner scale (I – V)	BMI (Z score)	Index test	Cut off (%)	Sensitivity (%)	Specificity (%)	PVV (%)	NPV (%)	Reference test
563 (315 males, 248 females)	Girls: 11.1 (2.7); boys: 11.4 (2.5)	Both gender stage I group and stage II group	Not known							OGTT
				FPG (*pPG)	4.8	80.00	58.00	10.00	98.00	
				FPG (*PG)	4.8	75.00	65.00	22.00	95.00	
				FPG (*pPB)	4.8	66.00	61.00	7.00	98.00	
				FPG (*PB)	4.8	87.00	53.00	8.00	99.00	
				FPG (pooled sample)	4.8	77.00	58.00	12.00	97.00	
				FSI (*pPG)	13	100.00	69.00	17.00	100.00	
				FSI (*PG)	16	67.00	57.00	16.00	93.00	
				FSI (*pPB)	11	66.00	54.00	6.00	97.00	
				FSI (*PB)	14	75.00	59.00	8.00	98.00	
				FSI (pooled sample)	Specific group	71.00	59.00	11.50	96.00	
				HOMA-IR (*pPG)	2.85	100.00	73.00	18.00	100.00	
				HOMA-IR (*PG)	3.44	65.00	59.00	18.00	92.00	
				HOMA-IR (*pPB)	2.65	66.00	72.00	10.00	98.00	
				HOMA-IR (*PB)	3.25	75.00	67.00	9.00	98.00	
				HOMA-IR	Specific group	71.00	66.00	14.00	97.00	
*pPB = prepubertal boys, pPG = prepubertal girls, PB = pubertal boys, PG = pubertal girls										

Main findings:

P - 563 (315 males, 248 females), white ethnicity, age (4–17 years), and obesity;

I – FPG, FSI, HOMA-IR;

R – OGTT;

D – pre-diabetes.

FPG and FSI, but not HOMA-IR varied significantly according to gender. FSI and HOMA-IR not FPG, varied significantly according to puberty. Two-hour plasma glucose (after glucose load) was significantly higher in pubertal than in prepubertal children in both genders. Frequency of IFG was not significantly different between males and females. In females, a gender × puberty interaction was found, the IFG frequency being higher in pubertal than in prepubertal girls. Children with IGT had significantly higher FPG, higher FSI, and higher HOMA-IR than children without IGT. IGT frequency was higher in girls than in boys and it was affected by puberty in girls only, being higher in pubertal than in prepubertal girls. ROC curve analyses run for gender and puberty-adjusted biochemical parameters (FPG, FSI, HOMA-IR), were all significant and were not statistically different from each other, as demonstrated by their widely overlapping 95% confidence interval: area under the curve = 0.68 (0.59–0.76), P = 0.0002; area under the curve = 0.66 (0.56–0.76), P = 0.001; area under the curve = 0.68 (0.59–0.78), P = 0.0001, respectively. In the pooled population, FPG, FSI, and HOMA-IR did not show statistically different sensitivity, specificity, or predictive values. This result was also confirmed in each gender/puberty subgroup. Sensitivity and specificity of FPG, FSI, and HOMA-IR were not significantly different from gender/puberty subgroup. Threshold values varied among gender/puberty subgroups for FSI and HOMA-IR, but not for FPG, due to the minimal variation of this parameter according to gender and to the absence of variation according to puberty (Maffeis et al., 2010).

2.6.1.17. **Maldonado-Hernández (2016):**

Title: The 13C-Glucose Breath Test for Insulin Resistance Assessment in Adolescents: Comparison with Fasting and Post-Glucose Stimulus Surrogate Markers of Insulin Resistance

Number of participants	Age groups	Tanner scale (I – V)	BMI (Z score)	Index test	Sensitivity (%)	Specificity (%)	PVV (%)	NPV (%)	Accuracy
133 (62 females and 71 males)	Mean age: 13 (range: 9-16)	pubescent (53.4%; stages 2 and 3) and postpubescent (46.6%; stages 4 and 5)	BMI 23 (15.6-37.8)	*A% OD					
Pubescents Reference tests			HOMA-IR ≥p95	≤16.0%	78.90	62.10	70.30	72.00	71.00
			FPI≥p90	≤16.3%	82.80	60.60	64.90	80.00	76.60
			2-h OGTT PI≥65 μU/ml	≤14.6%	75.00	69.00	53.60	85.30	71.00
Post-pubescent Reference tests			HOMA-IR ≥p95	≤13.0%	77.80	70.50	61.80	83.80	73.20
			FPI≥p90	≤13.0%	87.50	63.60	41.10	94.60	73.60
			2-h OGTT PI≥65 μU/ml	≤12.6%	77.80	67.90	42.50	90.00	70.40
*A% OD: adjusted percentage of oxidized ^{13}C-glucose dose at 180 minutes									

Main findings:

P – 133 (62 females, 71 males), healthy adolescents aged between 10 and 16 years (mean age:13 years);

I – A% OD: adjusted percentage of oxidized ^{13}C-glucose dose at 180 minutes;

R – HOMA-IR, FPI≥p90, 2-h OGTT PI≥65 μU/ml;

D – pre-diabetes.

The tabular form of the results from this study is different because one index test (^{13}C-Glucose Breath Test) was compared to 3 different reference tests (HOMA-IR, FPI, and 2-h OGTT).

The parameters of weight, BMI, abdominal circumference, fasting plasma insulin, and HOMA-IR had statistically significant differences. When contrasting lean versus obese and overweight versus obese individuals, 2-h OGTT insulin and A% OD at 180 minutes differed significantly. The comparison of lean versus overweight and lean versus obese subjects revealed that the 2-h OGTT glucose was substantially different. Finally, fasting plasma glucose achieved a statistically relevant difference only between lean and obese individuals. Three multiple regression models with three different IR surrogates were used to determine the influence of Tanner stage and gender on ^{13}C-GBT; IR was defined as HOMA-IR ≥p95 reference score adjusted by gender and age fasting plasma insulin ≥p90 reference score adjusted by gender and Tanner stage, and 2-h OGTT insulin ≥65 μU/mL. Gender does not substantially alter ^{13}C-GBT when co-analysed with HOMA-IR (β=0.8; p=0.361), fasting plasma insulin (β=1.0; p=0.239),

and 2-h OGTT insulin (β=1.4; p=0.131). In contrast, it was established that Tanner stage modifies ^{13}C-GBT when co-evaluated with HOMA (β=-2.1; p=0.017), fasting plasma insulin (β=-1.9; p=0.034), and 2-h OGTT insulin (β=-2.1; p=0.017). In pubescent and post-pubescent individuals, the ^{13}C-GBT rendered the highest accuracy when compared to fasting plasma insulin. With said parameter, in pubescent individuals, an A% OD at 180 minutes ≤16.3% diagnoses IR with a sensitivity of 82.8%, a specificity of 60.6%, a positive predictive value (PPV) of 64.9%, and a negative predictive value (NPV) of 80.0%. In post-pubescent subjects, an A% OD at 180 minutes ≤13.0% indicates IR with a sensitivity of 87.5%, a specificity of 63.6%, a PPV of 41.1%, and an NPV of 94.6% (Maldonado-Hernández et al., 2016).

2.6.1.18. **Mutlu (2013):**

Title: Can HbA1c and One-Hour Glucose Concentration in Standard OGTT Be Used for Evaluation of Glucose Homeostasis in Childhood?

Number of participants	Age groups	Tanner scale (I –V)	BMI (Z score)	Index test	Cut off (%)	Sensitivity (%)	Specificity (%)	PVV (%)	NPV (%)	Reference test
106 participants (71 female, 35 male)	13.4+/-2.6 (7-18)	Not known	BMI: 31.5+/-(20.7-7.46)							OGTT
				HbA1c	5.5	63.00	70.00	-	-	
					5.2	78.00	37.00	-	-	
					5.3	72.00	49.00	-	-	

Main findings:

P – 106 obese/overweight children aged from 7 to 18 years: 13.4+/-2.6 (median: 13.5); Gender (F/M): 71 (67%), 35(33%);

I – HbA1c;

R – OGTT;

D – pre-diabetes.

Mean FG was 78.7±10 mg/dL (54-104 mg/dL), mean 2-hour glucose concentration was 119.6±27.8 mg/dL (50238 mg/dL), and mean HbA1c level was 5.3±0.5% (4-7.5%). Three subjects (3%) had IFG, 18 subjects (17%) had IGT, and 1 subject (1%) had diabetes according to their 2-hour glucose concentrations. Only one of the 18 subjects who had an IFG had IGT. Mean 30-minute insulin concentration of the group was 102.3±83 uU/mL. Their mean plasma triglyceride level was 118.2±62.7 mg/dL, total cholesterol level 163.1±52.4 mg/dL, HDL cholesterol 43.5±11.8 mg/dL, LDL cholesterol level was 92.9±27.1 mg/dL, and VLDL cholesterol level 22.9±13.8 mg/dL. There was a negative correlation between the 2-hour glucose and the 30-minute insulin concentrations ($p<0.01$) and positive correlations between the 2-hour glucose concentration and the FG level and between the 1-hour glucose and the HbA1c levels. However, the 2-hour glucose concentration was not correlated with age, pubertal stage, BMI, BMI-SDS, or WC (p=0.7, 0.6, 0.8, 0.9 and 0.7, respectively).

If a 5.5% cut-off value for HbA1c was accepted to be a predictor of IGT, the sensitivity was 63% and specificity was 70%. Although the cut-off values of 5.2 and 5.3% had higher sensitivity (78 and 72%, respectively), they had lower specificity (37 and 49%, respectively). 31% of the subjects who had HbA1c levels at or above 5.5% had IGT, however, this rate was significantly lower in the subjects who had HbA1c levels below 5.5% (10%) ($p<0.05$). Although only one (5.5%) of the 18 subjects with IGT had IFG, 12 (66.6%) of them had HbA1c at or above 5.5% (Mutlu, Ozsu, Cizmecioglu, & Hatun, 2013).

2.6.1.19. **Nam (2018):**

Title: HbA1c Cut off for Prediabetes and Diabetes Based on Oral Glucose Tolerance Test in Obese Children and Adolescents

Number of participants	Age groups	Tanner scale (I – V)	BMI (Z score)	Index test	Cut off (%)	Sensitivity (%)	Specificity (%)	PVV (%)	NPV (%)	Reference test
389 (male: 217 (55.8%); normoglycemia (n= 197), prediabetes (n=121), type 2 DM (n = 71)	Mean age was 13.0 ± 2.5 years	Not known	normoglycemia (2.3+/-0.8), prediabetes (2.2+/-0.7), type 2 DM (2.0+/-0.5)							OGTT
				HbA1c - prediabetes	5.8	64.10	83.80	79.40	70.50	

Main findings:

P – 389 children (48 overweight and 341 obese) and there were more boys (217, 55.8%) than girls. The mean age was 13.0 ± 2.5 years. The mean height SDS, body weight SDS, and BMI SDS were 0.9 ± 1.2, 2.2 ± 0.8, and 2.2 ± 0.6, respectively;

I – HbA1c;

R – OGTT;

D – pre-diabetes.

Based on the results of the OGTT, 197 (50.6%) subjects had normoglycemia, 121 (31.1%) had prediabetes, and 71 (18.3%) had type 2 DM. Due to our focus SR DTA we present only the results for the prediabetes group, which is our diagnosis of interest. In prediabetes group AUC was used to determine the diagnostic performance of HbA1c for prediabetes. The statistically optimal HbA1c cut off point for prediabetes was 5.8% (AUC, 0.795; 95% CI, 0.750–0.840), with a sensitivity of 64.1% and a specificity of 83.8%. The sensitivity of this study was lower and the specificity was higher than that of ADA criteria at the prediabetic cut off (≥ 5.7) (64.1% vs. 68.8% and 83.8% vs. 73.6%, respectively). Based on the ADA cut off point for HbA1c of 5.7%–6.4%, 17 (9.4%) of 180 children with prediabetes satisfied all three diagnostic criteria. Twenty-nine (16.1%) were omitted without 2-hr PG. Based on the cut off point for HbA1c of

5.8%–6.1% in the present study, 12 (7.7%) of 156 children with prediabetes satisfied all three diagnostic criteria; 40 (25.6%) were omitted without 2-hr PG (Nam et al., 2018).

2.6.1.20. **Nor (2015):**

Title: Triglyceride glucose (TYG) index as a surrogate measure of insulin sensitivity in obese adolescents

Number of participants	Age groups	Tanner scale (I – V)	BMI (Z score)	Index test	Cut off (%)	Sensitivity (%)	Specificity (%)	PVV (%)	NPV (%)	Reference test
225 obese adolescents (114 male and 111 female)	10-20 yr; mean age 14.2+1.9 years	Tanner stages II–V	Not known							Insulin-stimulated glucose disposal (Rd)
				TyG index	8.52	69.10	71.10	-	-	
					8.93	-	-	-	-	
					8.43	-	-	-	-	
				TyG/HDL	-	-	-	-	-	
				1/IF*	-	-	-	-	-	
*1/IF = 1/fasting insulin										

Main findings:

P – Participants' mean age 14.2+1.9 years (122 black and 103 white, 114 male and 111 female); They were between 10 and 20 yrs. old;

I – TyG index, TyG/HDL, 1/IF;

R – Insulin-stimulated glucose disposal (Rd);

D – pre-diabetes.

Prof. Noor was contacted via e-mail as an author to provide us additional data in 2x2 table so we could pool the study results to the meta-analysis because the data about sensitivity and specificity in the study were missing. She answered with apology that the set of original data are missing hence they are not able to provide original 2x2 table data that was needed.

The study participants were divided into 3 groups (normal glucose tolerance – OB-NGT, prediabetes – OB-preDM, type 2 diabetes mellitus – OB-T2DM). In the next paragraph, we stated the input evaluation of individual groups. Description of the results is based on data that were not evaluated for each group separately, but for the overall population included in the study. More detailed information about sensitivity, specificity and cut off points of the index tests are missing in the study too.

The overall mean for Rd and TyG index in our study population were 6.1±2.4mg/kg/min and 8.5±0.5, respectively. There were no significant differences among the groups with respect to BMI, sex, race, and waist circumference. Fasting glucose and triglycerides were significantly higher in the OB-T2DM compared with OB-preDM and OB-NGT. Rd significantly declined across the glycemic groups from OB-NGT to OB-preDM to OBT2DM, with a corresponding significant increase in TyG index, higher in OB-T2DM and OB-preDM vs. OB-NGT. With

regards to TG/HDL ratio and 1/IF, OB-NGT group had the lowest TG/HDL ratio and the highest 1/IF.

The AUC was lowest for TG/HDL and highest for 1/IF. The AUC for TyG index was 0.750 ($p<0.0001$) in the total population. A TyG index of 8.52, 69.1% sensitivity and 71.7% specificity, best predicted insulin resistance in the total population.

In multiple regression analyses with Rd as the dependent variable and age, sex, race, Tanner stage, BMI z-score, TyG index, and glycemic group as the independent variables, 51.4% of the variance in Rd ($p<0.0001$) was explained by TyG index (partial $r=-0.412$, $p<0.0001$), BMI z-score (partial $r=-0.514$, $p<0.0001$), sex (partial $r=-0.355$, $p<0.0001$), glycemic group (partial $r=-0.293$, $p<0.0001$), and race (partial $r=0.155$, p 0.048). Replacing the TyG index with 1/IF in the model increased the prediction of the variance in Rd to 57.7% ($p<0.0001$). Inclusion of both the TyG index and 1/IF in the regression model further improved the estimate of the variance in Rd to 64.8%. The addition of HDL did not contribute significantly to the variance in Rd (Nor, Lee, & Arslanian, 2015).

2.6.1.21. **Pandey (2017):**

Title: Anthropometric indicators as predictor of pre-diabetes in Indian adolescents

Number of participants	Age groups	Tanner scale (I – V)	BMI (Z score)	Index test	Cut off	Sensitivity (%)	Specificity (%)	PVV (%)	NPV (%)
526 (277 boys and 249 Girls)	Mean age of boys was 18.5+/-1.5yrs; the mean age of girls was 17.9+/-1.8yrs	Not known	BMI: boys: 22+/-3.5; girls: 20.8+/-4.1						
				BMI (boys)	≥22.8 kg/m^2	73.10	95.30	-	-
				BMI (girls)	≥20.5 kg/m^2	70.80	94.60	-	-
				Waist circumference (boys)	≥82.5 cm	75.30	92.40	-	-
				waist circumference (girls)	≥80.3 cm	72.70	94.80	-	-

Main findings:

P – 277 boys and 249 girls. The mean age of boys was 18.5+/-1.5 years and the mean age of girls was 17.9+/-1.8years;

I – BMI, waist circumference;

R – N/A;

D – pre-diabetes.

Prevalence of prediabetes among the study subjects was 32.1%. The ROC analysis for BMI showed good predictive power for pre-diabetes for both boys and girls. Area under the curve was 0.828 for boys and 0.838 for girls, respectively. he cut-offs of BMI to predict prediabetes were calculated as ≥22.8 kg/ m^2 in boys and ≥20.5 kg/m^2 in girls. ROC analysis for waist circumference also revealed that it was a good discriminator of prediabetes both for boys (area under the curve 0.804) and girls (area under the curve 0.795). The cut-offs for waist circumference to predict prediabetes were calculated as ≥ 82.5cm for boys and ≥80.3cm for girls. The sensitivity and specificity of the cut off for BMI in boys was 73.1% and 95.3% respectively, and the same in girls was 70.8% and 94.6%, respectively. The sensitivity and specificity of the cut off for waist circumference in boys was 75.3% and 92.4% respectively, and the same in girls was 72.7% and 94.8% respectively. Adolescents with raised BMI or increased waist circumference had a greater prevalence of prediabetes.

Multiple logistic regression analysis of the determinants of prediabetes showed that BMI, waist circumference and physical activity were significantly associated with pre-diabetes in adolescents. For every 1kg/m^2 increase in BMI, there was a 1.067 times increased risk of pre-diabetes and for every 1 cm increase in waist circumference, there was a 1.028 times higher risk of pre-diabetes (Pandey et al., 2017).

2.6.1.22. **Puri (2007):**

Title: Criteria for oral glucose tolerance testing of obese minority youth

Number of participants	Age groups	Tanner scale (I – V)	BMI (Z score)	Index test	Cut off (%)	Sensitivity (%)	Specificity (%)	PVV (%)	NPV (%)	Reference test
167 participants (75 males, 92 females)	10-18 years, mean age 14 +/- 2.3	Not known	BMI: Normal OGGT: 37.7+/- 7.4; IGT/DM2 : 41.8+/- 8.4							OGTT
				HOMA-IR (girls)	>4.5	100.00	55.10	-	-	
				HOMA-IR (boys)	>13	100.00	76.60	-	-	
				HbA1c (boys)	>5.8	100.00	76.60	-	-	
				Cholesterol (boys)	>200 mg/dL	100.00	76.60	-	-	

Main findings:

P – 167 participants ages 10-18 years or in puberty, BMI >85th percentile and family history of DM2, race/ ethnicity (African American, Caribbean Hispanic) with signs of insulin resistance, such as acanthosis nigricans; with a mean age 14 ± 2.3 years;

I – HOMA-IR, cholesterol, HbA1c;

R – OGTT;

D – pre-diabetes.

The study participants were divided into 2 groups (Obese: normal OGTT, Obese: IGT/DM2). A total of 21/167 (12.5%) of children screened had an abnormal OGTT (IGT/DM2), 11/75 (14.7%) boys, and 10/92 (10.9%) girls. In girls, HOMA >4.5 identified those with an abnormal OGTT (IGT/ DM2) with a sensitivity of 100% (95% CI 51.7, 97.1) and specificity 55.1% (95% CI 44.09, 66.17). In boys, HOMA-IR >13, HbA1c >5.8% or cholesterol >200 mg/dl identified those with an abnormal OGTT with a sensitivity of 100% (95% CI 71.5, 100) and specificity 76.6% (95% CI 66.2, 86.9). In this cohort, 29/61 (48%) girls met the screening criteria (HOMA >4.5), and 26/75 (35%) boys met the screening criteria (HOMA >13, HbA1c >5.8%, or total cholesterol >200 mg/dl).

The model was validated using an independent sample of patients followed in the Paediatric Endocrine Clinic at the Children's Hospital at Montefiore, who met the ADA diabetes screening criteria. In this independent sample, 8% of the 198 patients demonstrated IGT. The screening criteria identified these girls with 100% sensitivity (95% CI 39.8, 100) and specificity 39.3% (95% CI 28.8, 49.7), and these boys with 100% sensitivity (95% CI 2.5, 100) and specificity 70% (95% CI 58.4, 81.6) (Puri, Freeman, Garcia, Nussbaum, & Dimartino-Nardi, 2007).

2.6.1.23. **Sharma (2012):**

Title: Use of HbA1C testing to diagnose pre-diabetes in high risk African American children: A comparison with fasting glucose and HOMA-IR

Number of participants	Age groups	Tanner scale (I – V)	BMI (Z score)	Index test	Cut off (%)	Sensitivity (%)	Specificity (%)	PVV (%)	NPV (%)	Reference test
172 (70 boys, 102 girls	9–11 years (boys: 9.96; girls: 9.80)	I-V	BMI, z-score: boys: 1.85 (0.07); girls: 2.05 (0.06)							HbA1c
				HOMA-IR	2.5	93.00	21.00	-	-	
				Glucose	100 mg/dL	88.00	0.00	-	-	

Main findings:

P – 172 children (70 boys and 102 girls) aged 9–11 years with BMI's above the 85th percentile;

I – HOMA-IR, Glucose;

R – HbA1c;

D – pre-diabetes.

Glucose, insulin and HOMA-IR were significantly interrelated, but HbA1C was not significantly correlated with these biochemical prediabetes assessment variables, nor with anthropometric (BMI z score, WC) risk factors. Of the 172 participants included in this analysis, 21 (12.2%) had HbA1C concentrations above the cut off of 5.7 used to identify prediabetes. None (0%) of these 21 participants, however, were observed to have a glucose concentration above the pre-diabetes cut off of 110 mg/dl and only 61.9% of the participants had HOMA-IR above the pre-diabetes cut off of 2.5. If there is a dual role for fasting glucose and HbA1C for prediction of diabetes, 12.2% of this sample would be classified as pre-diabetic, a proportion identical to that determined using HbA1C alone. Applying the dual role concept, but including participants with values for HOMA-IR (instead of values for glucose) and/or HbA1C above their cut offs, the proportion of participants with pre-diabetes increased from 12.2% with HbA1C alone to 41.3% with these two markers. When compared to the previously identified glucose cut off of 110 mg/dl and HOMA-IR cut off of 2.5 for pre-diabetes, HbA1C showed high specificity (88 and 93%, respectively) but very low sensitivity (0 and 21%, respectively) (S. Sharma & Fleming, 2012).

2.6.1.24. **Tirabanchasak (2015):**

Title: Insulin dynamics and biochemical markers for predicting impaired glucose tolerance in obese Thai youth

Number of participants	Age groups	Tanner scale (I – V)	BMI (Z score)	Index test	Cut off (%)	Sensitivity (%)	Specificity (%)	PVV (%)	NPV (%)	Reference test
115 participants (males: 76, females: 39)	median age 12.6 years (range: 8.4–17.5)	45 (II-III), 40 (IV-V)	BMI 32.9 (24.0-57.5)							OGTT
				FG	N/A	-	-	-	-	
				HOMA-IR	N/A	-	-	-	-	

Main findings:

P – 115 obese subjects (76 males and 39 females, age ranging from 8 to 18 years) with obesity; median age of the patients was 12.6 years (range: 8.4–17.5)

I – FG, HOMA-IR;

R- OGTT;

D – pre-diabetes.

As it is stated in the study results part, the attempt to identify optimal predictive cut off values of fasting biochemical indices for predicting IGT was unsuccessful. ROC curve in the study provided analyses for FBG, HbA1c and 1-h postload glucose for predicting IGT. The area under the ROC curve (AUC) of HbA1c was 0.555 (0.410–0.700, p = 0.402). The AUC of FBG was 0.631 (0.508–0.754, p = 0.035). However, FBG had a low sensitivity and specificity for predicting IGT. Among FBG, HbA1c and 1-h postload glucose values, we found that 1-h postload glucose was the best predictor of IGT with an AUC of 0.712 (0.600–0.824, p < 0.001). The cut-off levels of 1-h postload glucose at 155 mg/dL had sensitivity 53.3%, specificity 79.5%, positive predictive value 50%, negative predictive value 79.5%, and accuracy 72.2%. The 1-h postload glucose values of 140 mg/dL gave a better sensitivity of 76.7%, but specificity decreased to 64.1% (Tirabanchasak, Siripunthana, Supornsilchai, Wacharasindhu, & Sahakitrungruang, 2015)

2.6.2. Summary of the narrative description of included studies

In the previous chapters, 24 studies were described as they were difficult to compare. Within these 24 studies, 20 different tests were used as an index test. HbA1c and FPG were used in some studies with different cut off points in order to find the one that would be most accurate. Within these 24 studies, 12 different tests were used as a reference test.

2.6.2.1. Summary of the narrative description of included studies containing index/reference tests that could not be compared

Although a total of 20 different tests were used in the 24 included studies, which were designated as index tests, they could not be compared in other studies because these tests were used individually in one included study.

Tests that were used in studies included in the SR DTA as an index test that could not be compared with the results of other studies are: FGIR, QUICKY, 2-h glucose, fructosamine, glycated albumin, 1.5-anhydroglucitol, glucose peak>30 minutes, monophasic curve, 1-h glucose 155 mg/dL. COMBO, FSI, % OD adjusted percentage of oxidized ^{13}C-glucose dose at 180 minutes, 1/IF, BMI, waist circumference. The reasons why the results of these index tests were not assessed are two: 1) the was no other study used the same index/reference test; 2) the results of the index test were completely missing for the individual index test.

Tests that were used in studies included in the SR DTA as a reference test that could not be compared with the results of other studies are: HOMA top quartile, fasting glucose 100 mg/dL, 2-hr glucose 140 mg/dL, FPI≥p90, PI≥65 μU/ml, insulin-stimulated glucose disposal (Rd). The reasons why the results of these reference tests were not assessed are two: 1) the was no other study used the same index/reference test; 2) the results of the index test were completely missing for the individual index test.

2.6.2.2. Description of the index/reference test of included studies that could be compared

In this chapter, the description of the index/reference test of included studies that could be compared will be described. The reason for this step is a large difference in the use of individual tests in the included studies, which, based on the analysis of the results, does not lead to a clear conclusion. This finding leads to a narrative synthesis of the results of the SR DTA, so we consider the description of individual tests in studies that could be compared to be one of the starting points that will lead to the substantiation of the results for implementation in practice.

HOMA-IR

Homeostatic model assessment (HOMA) is a method for assessing β-cell function and insulin resistance (IR) from basal (fasting) glucose and insulin or C-peptide concentrations. We can distinguish between HOMA-IR 1 (the original HOMA model) and HOMA-IR 2 (the uploaded HOMA model). HOMA1, the original model from Matthews et al. (Matthews et al., 1985) contained a simple mathematical approximation of the original nonlinear solution to the iterative equations (this is the explanation for the exponential functions, which are cancelled out, in that article). The equations are widely used and simplify to: HOMA1-IR = (FPI × FPG)/22.5 HOMA1-%B = (20 × FPI)/(FPG − 3.5) for IR and β-cell function, respectively, where FPI is fasting plasma insulin concentration (mU/l) and FPG is fasting plasma glucose (mmol/l). HOMA2 is the correctly solved computer model (Levy, Matthews, & Hermans, 1998)

and has nonlinear solutions. In this updated version it is possible to determine insulin sensitivity and β-cell function from paired fasting plasma glucose and radioimmunoassay insulin, specific insulin, or C-peptide concentrations. The authors recommend the computer software be used wherever possible (Levy et al., 1998). In 2004, the HOMA2 Calculator was released. This provides quick and easy access to the HOMA2 model for researchers who wish to use model-derived estimates of %B and %S, rather than linear approximations. It runs on a variety of computer platforms and can be downloaded on line which makes its approach to paediatricians possible.

HbA1c

Blood glycated haemoglobin (HbA1c) is considered to be a more accurate tool for determining blood glucose and is currently the routine and most effective tool for monitoring the course of diabetes. This is an indicator of so-called "long-term blood glucose", as it provides information on blood glucose for a period of 2-3 months. Glycated haemoglobin levels can be used in the screening for glucose homeostasis disorders, especially in relation to prediabetes. HbAlc is usually taken during sampling in a dialectological office. The patient does not have to be fasting. HbA1c values are expressed as a percentage (%) as the percentage of glycated haemoglobin of the total haemoglobin in the blood. The conversion of units is simple, mmol / mol is ten times the original values (Esteghamati et al., 2010).

FPG

Fasting plasma glucose is the level of sugar in the blood after someone has not eaten for a long time, usually overnight. It is often used as a measure of how well people with diabetes control their blood sugar. These levels may be too high - perhaps an indication that the person is suffering from diabetes or pre-diabetes. Fasting plasma glucose tests are important for individuals with pre-diabetes. This is because they are at increased risk of developing type 2 diabetes. They usually need to monitor their blood sugar frequently to make sure that the disorder has not developed. The pre-analytic requirements are fasting blood collection (min. 8 hours of fasting), determination in venous blood plasma (EDTA + NaF), separation of plasma from blood elements within 60 min after consumption, and when collecting urine, store urine until analysis at 4-8^{0}C, is bacterial contamination should be avoided (Gurung & Jialal, 2019).

OGTT

Oral glucose tolerance test is a series of blood tests that are used to evaluate an individual's response to drinking a standard quantity of a specific glucose-containing solution. The patient drinks the solution, and then the blood is drawn at specific intervals over the next several hours. Each blood test measures the amount of glucose (a particular form of simple sugar) in the blood. The tests are used to evaluate patients for the possibility that they have diabetes. An oral glucose tolerance test is usually performed when there is a suspicion that an individual has Type II diabetes, for example, when a serum glucose level has revealed an abnormality, when there is a strong family history of diabetes when an individual has specific risk factors for diabetes (such as being overweight), or when an individual is experiencing symptoms suggestive of diabetes (excessive thirst and/or hunger, urinary frequency, unintentional weight loss, severe fatigue and

weakness, and poor healing). Additionally, an oral glucose tolerance test is almost always ordered as a routine part of prenatal care during the second trimester of pregnancy, usually between 24 and 28 weeks of pregnancy. The oral glucose tolerance test should only be performed when the individual is in perfectly good health and normally ambulatory/active. For the 72 hours prior to undergoing the OGTT, the individual should be instructed to eat a high-carbohydrate diet (150-200 grams of carbohydrate per day). The test is done on a fasting basis, meaning that nothing should be eaten or drunk after midnight prior to the test (Cefalu et al., 2019).

TyG

Triglyceride-glucose index (TyG index) was made for a marker of insulin resistance, and calculated with fasting plasma glucose and triglycerides (Simental-Mendía, Rodríguez-Morán, & Guerrero-Romero, 2008). We can conclude that the correct formula for TyG index is Ln [fasting triglycerides (mg/dL) × fasting glucose (mg/dL)/2] which has been proposed by Simental et al. (Simental-Mendía et al., 2008).

TG_HDL

Lipoproteins are divided into 5 subgroups, by density/size (an inverse relationship), which also correlates with function and incidence of cardiovascular events. Unlike the larger lipoprotein particles, which deliver fat molecules to cells, HDL particles remove fat molecules from cells. The lipids carried include cholesterol, phospholipids, and triglycerides, amounts of each are variable (März et al., 2017). HDL particles remove fats and cholesterol from cells, including within artery wall atheroma, and transport it back to the liver for excretion or re-utilization; thus the cholesterol carried within HDL particles (HDL-C) is sometimes called "good cholesterol" (despite being the same as cholesterol in LDL particles). Those with higher levels of HDL-C tend to have fewer problems with cardiovascular diseases, while those with low HDL-C cholesterol levels (especially less than 40 mg/dL or about 1 mmol/L) have increased rates for heart disease (Toth, 2005). TGs (triglycerides): TGs are formed by combining glycerol with three molecules of fatty acid. TGs, as major components of VLDL and chylomicrons, play an important role in metabolism. When the body requires fatty acids as an energy source, the hormone glucagon signals the breakdown of the TGs by lipase to release free fatty acids. TGs are water-insoluble, non-polar neutral fats. These are not the structural components of biological membranes. TGs synthesis and storage mostly occur in liver and adipose tissue. Free fatty acids and glycerol must be activated prior to the synthesis of TGs into acyl-coA and glycerol-3-phosphate respectively.

2.6.3. General description of results

Only in three pairs of studies with HbA1c as an index test and HOMA-IR, and two pairs of studies with HOMA-IR as an index test and OGTT as a reference test, it was possible to make meta-analysis, because these pairs of studies had the same cut off point in range of 5.7 – 5.9 for HbA1c as an index test, and values 2.7 and 4 for HOMA-IR as an index test.

A total of 14 382 participants were included in these 24 studies. The age range was from 5 to 19 years and the mean age was approximately 12.64 years (counted from 23 studies; the mean age was not stated in Liang (2015). Information on participants´ Tanner scale was included in 9 studies. Tanner scale was defined in studies to determine the regimen to follow for paediatric or adolescent patients. Roman numerals I. - V were processed to define the degree of maturity.

Pubic Hair Scale (both males and females)

- Stage I.: No hair
- Stage II.: Downy hair
- Stage III.: Scant terminal hair
- Stage IV.: Terminal hair that fills the entire triangle overlying the pubic region
- Stage V.: Terminal hair that extends beyond the inguinal crease onto the thigh

Female Breast Development Scale

- Stage I.: No glandular breast tissue palpable
- Stage II.: Breast bud palpable under the areola (1st pubertal sign in females)
- Stage II.: Breast tissue palpable outside areola; no areolar development
- Stage IV.: Areola elevated above the contour of the breast, forming a "double scoop" appearance
- Stage V.: Areolar mound recedes into single breast contour with areolar hyperpigmentation, papillae development, and nipple protrusion

Male External Genitalia Scale

- Stage I.: Testicular volume < 4 ml or long axis < 2.5 cm
- Stage II.: 4 ml-8 ml (or 2.5 to 3.3 cm long), 1st pubertal sign in males
- Stage III.: 9 ml-12 ml (or 3.4 to 4.0 cm long)
- Stage IV.: 15-20 ml (or 4.1 to 4.5 cm long)
- Stage V: > 20 ml (or > 4.5 cm long) (Emmanuel & Bokor, 2017)

The 9 studies were from: (Galhardo (2015) – prepubertal and pubertal; Chan (2015) – I-V; Kurtoğlu (2010) – prepubertal and pubertal; Liang (2015) – prepubertal and pubertal; Maffeis (2010) – I-II; Maldonado-Hernández (2016) – pubescent and postpubescent; Nor (2015) – II.-V; Sharma (2012) – I-V; Tirabanchasak (2015) – II.-V.). Information on BMI or BMI z core was provided in 17 of 24 studies.

A reference test, which was designated as the "gold standard" for comparing diagnostic accuracy results with index test, was used in 21 studies. Chan (2015), Kim (20198) and Pandey (2017) in their studies did not mention a reference test.

Chan (2015) assessed relationships among CMG outcomes, HbA1c, and OGTT results (FPG and 2-h glucose).

Kim (2019) evaluated the correlation between plasma glucose (FPG and 2-h OGTT) and HbA1c and examined whether HbA1c could be used in place of the FPG and 2-h OGTT.

Pandey (2017) intended to find out the cut off values of BMI and waist circumference for predicting pre-diabetes.

In four studies, Brar (2014); Kim (2019); Lee (2019); Nam (2017); the participants were divided into subgroups.

From 149 obese participants in Brar (2014) study, normal (n=125), pre-diabetes (n=21) and diabetes (n=3) children were included. The numbers were described separately for pre-diabetes and diabetes group in result part of the study. Therefore, it was possible to include the study in our SR DTA.

From 190 participants in Kim (2019) study, normal glucose tolerance (n=117), impaired glucose tolerance (n= 33) and diabetes (n=40) children divided into 3 subgroups based on entrance OGTT measurement were included. Although the study is focused on mainly DM predictor of stated diagnostic tests, the descriptive data for non-diabetic population are available. Therefore, it was possible to include the study in our SR DTA.

From 7332 participants in Lee (2019) study, 4129 children were enrolled to the group aged 10-19 (the rest of 3203 participants were 20 – 29 years old). The numbers were described separately for the group aged 10-19 in result part of the study. Therefore, it was possible to include the study in our SR DTA.

From 389 participants in Nam (2018) study, normoglycemic (n=197), pre-diabetes (n=121) and diabetes (n=71) children were included. The numbers were described separately for pre-diabetes and diabetes group in result part of the study. Therefore, it was possible to include the study in our SR DTA.

In these 23 studies, 27 index tests and 12 reference tests were used. In connection with different types of diagnostic tests, the study was focused on which of the tests identifying pre-diabetes is more reliable and how to compare their quality. Therefore, the most frequently reported characteristics were sensitivity and specificity. The degree of sensitivity expresses the probability of correct diagnosis of positive cases; the degree of specificity expresses the probability of correct diagnosis of negative cases. A test with a high degree of sensitivity reveals a high proportion of real patients with a diagnosis of interest. However, with low degree of specificity, there is a risk that that a false positive result will be demonstrated. On the contrary, a test with high specificity gives only an exceptionally false positive result. However, with low test sensitivity, there is a risk of false negativity. Ideally, both sensitivity and specificity should be as high as possible. In practice, it is always necessary to proceed from real possibilities and the state of knowledge. In many situations, a diagnostic tool with these indicators around 60% can be a great benefit, in others the values are close to 100%.

In the 24 studies which were included in the SR DTA, we could find studies in which both sensitivity and specificity degrees was higher than 80 %.

Ehehalt (2017) reports the results comparing HbA1c (index test) with OGTT (reference test) on the level of HbA1c ≥48 mmol/mol (≥6.5%) with sensitivity 84 % and specificity 99.3 %.

Kim (2019) reports the results of HbA1c (cut off point ≤6.5%) with sensitivity 89.4% and specificity 100%; and FPG <126mg/dL with sensitivity 85.1 % and specificity 100 %.

Lee (2019) reports the results of HbA1c (index test) comparing with FPG (reference test) on the level of cut off point 5.9 % with sensitivity 100% and specificity 95.8 %.

In other studies, there were big difference between sensitivity and specificity, e.g. Galhardo (2015) reports the results of HbA1c (index test) comparing with OGTT (reference test) on the level of cut off point 3.1 with sensitivity 100% and specificity 0%; and on the level of cut off point 4.4 with sensitivity 100% and specificity 1%; the same results are stated for fasting blood glucose (mmol/l) on the level of cut off point 3.7% with sensitivity 100% and specificity 0%.

3. Discussion

3.1. General discussion

The increase in the incidence of T2DM in early age correlates with the global obesity pandemic, which is related to the way of life of the late 20th and early 21st centuries. 95% of children with T2DM have a body mass index (BMI) higher than the 85th percentile of the population. The lifestyle of families of children suffering from metabolic changes is characterized by overeating, inappropriate diet and minimal physical activity. The situation in this area will not improve even after the coronavirus pandemic, on the contrary. Due to anti-pandemic measures, there was a deepening of children's inactivity, insufficient physical movement and an increased incidence of sedentary lifestyle. This fact was proven by HBSC study (The Healthy Behaviour in School-aged children) (Ng, Cosma, Svacina, Boniel-Nissim, & Badura, 2021). For this reason, it can be assumed that the incidence of metabolic diseases due to a coronavirus pandemic will result in a higher incidence of children with this type of disease in the future.

In this systematic review of diagnostic test accuracy, we summarized the results of 25 studies reporting the accuracy of tests identifying pre-diabetes in children that met the inclusion criteria of this review. The identification of pre-diabetes relies on tests which were defined as index test, that could bring a new perspective in this field of diagnosing, and reference test being considered as so-called gold standard in diagnosing pre-diabetes. According to ADA, there are tests which are considered as a gold standard identifying pre-diabetes in adults but having these tests used in children are missing.

When comparing the 24 studies, we had to consider a few facts that accompanied the development of SR DTA.

From a total of 24 studies, 9 of them was published before 2015 (the oldest one included was from 2005), the rest was published in/or after 2015. We assume that one of the reasons why there is no more relevant literature before 2015, is this fact that pre-diabetes is a disease that has been developing predominantly in the last thirty years. This also gives us the reason why most studies were excluded for unsatisfactory design – 47 studies that did not meet study design criteria (mostly prevalence study). We see some interesting data in conference abstracts. Unfortunately, it was not possible to find the full text for them. Therefore, some of them had to be excluded.

Another observation can be seen in ontogenetic development and how it has been considered in individual studies. The IDF had set criteria for how the ontogenetic development conditions should be defined for different stages of ontogenetic development. These age periods can be divided into 6 to 10, 10–16, and >16 years. However, this division was not considered in any study. The question is, to what extent is this problem in identifying pre-diabetes and determining the most accurate test that reveals it? We see another problem in the division of the studied population into girls and boys. Some studies divided and compared the results in

girls and boys. We hypothesize that this is also one of the factors that can affect the detection of pre-diabetes, especially in the pubertal and adolescent population in terms of hormonal changes, fat storage and other signs of the stage of development 10-16, and >16 years, which is different in girls and boys.

The definition of index tests and test references were in all studies described clearly. What seems to be a problem was index/reference time interval because this information was missing in almost all included studies as well as the period that studies were carried out (beginning and end date of the individual study). In some studies, the cut-off point of the tests were missing. This issue mainly concerned studies that were published before 2015.

We also observed differences in the presentation of individual studies. Only two studies were developed according to STARD guideline. Other studies were designed according to the classical order "introduction, methods/materials, results, discussion, conclusion".

Only in one study from total 24, the authors published data containing false positive/negative numbers. Although we tried to contact the authors of the studies were the data needed for meta-analysis were missing we were not successful and we had to exclude these studies from the meta-analysis process.

3.2. Discussion to the results of included studies

The original review objective was to **identify all alternative tests currently in use for the diagnosis of type 2 pre-diabetes mellitus in children and establish their accuracy relative to this gold standard. The gold standard for the diagnosis of pre-diabetes was the measurement of fasting plasma glucose and the oral glucose tolerance test**.

All studies included in the SR DTA were discussing the diagnostic accuracy of tests detecting pre-diabetes. To put the results from the studies in a context, we divided them according to the pairs of tests which were used in the individual studies. We can determine seven pairs of the same tests that were supposed to be discussed in this part of the systematic review. The tests are listed the first index test, second reference test: HOMA-IR and OGTT (Atabek, 2017), (Brar, 2014), (Galhardo, 2015), (Keskin, 2005), (Kurtoğlu, 2010), (Liang, 2015), (Maffeis, 2010), (Puri, 2007), (Tirabanchasak, 2015); HbA1c and OGTT (Brar, 2014), (Ehehalt, 2017), (Galhardo, 2015), (Kim, 2019), (Mutlu, 2013), (Nam, 2018), (Puri, 2007); TyG and HOMA-IR (Garcia, 2019), Kang (2017); TG_HDL and HOMA-IR (Garcia, 2019), (Kang, 2017); FGIR and OGTT (Atabek, 2017), (Keskin, 2005); FPG and OGTT (Ehehalt, 2017), (Kim, 2019), Maffeis (2010); TrG_HDL and OGTT (Galhardo, 2015), (Bridges, 2016).

A total number of four studies were possible to be pooled in the meta-analyses. The studies from which meta-analyses was possible to be pooled, had two reference tests (HOMA-IR and OGTT) and five index tests (HbA1c, HOMA-IR, TyG, TG_HDL and FPG). The HbA1c had three different cut off points used: 5.7, 5.8 and 6.5. The TyG had two different cut off points

used: 8.5 and 8.38. The TG_HDL had two different cut off points used: 2.22 and 1.71. Separate meta-analyses were performed for both different test pairs and different cut off values.

The only meta-analysis which was possible to be pooled, had three studies from Brar (2014), Ehehalt (2017) and Nam (2018). The index test of these studies was HbA1c and reference test was OGTT. The cut off point was 5.7. In Figure 8 Summary ROC Plot of results, the curve of this meta-analysis is plotted with a red diamond. In the Brar´s study (2014), sensitivity was 75.00% and specificity was 58.00% with TP 18, FP 53, FN 6 and TN 72. In Ehehalt´s study (2017) was sensitivity 96.00% and specificity was 76.00% with TP 48, FP 1168, FN 2, TN 3619. In Nam´s study (2018) was sensitivity 67.00% and specificity was 74.00% with TP 123, FP 54, FN 61 and TN 151. From these results, it can be seen, that the Ehehalt´s study had the best results with the highest sensitivity and specificity. Curve of the meta-analysis starts at a value of approximately 0.11. However, if we look at the individual diamonds representing the results of the three studies used (Brar, 2014, Ehehalt, 2017 and Nam, 2018), we find that each of the diamonds is located in a different place in the space above the diagonal curve. The important fact which need to be mentioned here is, that all of these three studies had very different baseline characteristic of population and the tests were also not performed according to the same procedures, which could lead to different results. As it was mentioned in Chapter 2.5, in the Ehehalt (2017) study 4848 overweight, obese, and extremely obese children and adolescents from Germany aged 7 to 17 years were included. In Brar´s study (2014), the number of participants was 149 aged 13.8+/-3.1. It was conducted in the USA, and it included five different ethnicities. Population in Nam´s study (2018) was ten years and above with a body mass index ≥ 85th percentile for age and gender and having two or more additional risk factors for diabetes, consistent with American Diabetes Association (ADA). So, if we assess the Ehehalt´s study (2017), Brar´s study (2014), Nam´s study (2018), the biggest difference is in ethnicity. Nam vymazat, že testy dělali jen jednou – detailně se zaměřit na abstract

Other meta-analyses were performed on individual studies, as there were no other studies with which they could be compared. These are, therefore single meta-analyzes, the results of which, however, can be discussed.

In Nam´s study (2018), except of the cut off point 5.7 was used the cut off point 5.8 (HbA1c as an index test and OGTT as a reference test. The sensitivity and specificity are different. Sensitivity at the cut off point 5.8 is 64.00% and specificity was 84.00%, with TP 99, FP 38, FN 56 and TN 196 (at the cut off point 5.7 sensitivity was 67.00% and specificity was 74.00%). Therefore, we can state that HbA1c as an index test at the level of cut off point 5.7 versus OGTT as a reference test is more accurate. In Figure 8 Summary ROC Plot of results, the curve of this meta-analysis is plotted with a light green square. This curve, which starts on the TPF axis at about 0.09, essentially copies the meta-analysis curve for the TG_HDL test index with cut off point 2.22 vs HOMA-IR In Nam´s study (2018), the population from South Korea was 10 years and above with body mass index ≥ 85th percentile for age and gender and having two or more additional risk factors for diabetes, consistent with American Diabetes Association (ADA). We can hypothesize that in different population could be more accurate different cut off point.

When assessing an index test HbA1c and reference test OGGT, the study from Ehehalt (2017) used the third different cut off point included in single meta-analysis. The cut off point was 6.5 and the sensitivity was 84.00% and specificity was 99.00%, with TP 22, FP 21, FN 28 and TN 5229. In Figure 8 Summary ROC Plot of results, the curve of this meta-analysis is plotted with a dark blue triangle. It is seen that that it is at a value reaching the axis in point 1. This result is more accurate than Brar´s (2014) results but we must to take into account the difference in the population included in the Ehehalt´s study (2017). This study is defined as an observational multicenter analysis with 4848 German participants.

Another single analysis was focused on using TyG and TG_HDL versus HOMA-IR from Garcia´s study (2017). The cut off point for TyG was 8.5 (sensitivity 65.00%, specificity 25.70%, with TP 39, FP 105, FN 21, TN 36) and for proposed TyG 8.83 (sensitivity 95.00%, specificity 42.30%, with TP 132, FP 36, FN 7 and TN 26). TG/HDL as an index test and HOMA-IR as a reference test, the cut off point 2.22 was used and the sensitivity was 90.00%% and specificity was 52.00%%, with TP 97, FP 45, FN 11 and TN 48. The proposed value of threshold was 1.71 with sensitivity 95% and specificity 68.60%, with TP 132, FP 19, FN 7 and TN 43. In Figure 8 Summary ROC Plot of results, the curve of this meta-analysis (TG_HDL 1.71 x HOMA-IR) is plotted with dark yellow circle. And we can see, that the most accurate is TG_HDL with cut off point 1.71 versus HOMA-IR. The second accurate test in this analysis is TyG with cut off point 8.38 versus HOMA-IR. In Figure 8 Summary ROC Plot of results, the curve of this meta-analysis (TyG 8.38 x HOMA-IR) is plotted with pink plus sign. If we compare these two tests within the performed meta-analysis, we must state that the index TG_HDL test with cut off point 1.71 versus HOMA-IR as a reference test is the third most accurate test and in the sub-group analyses 2 (figure 12) comparing all tests to HOMA-IR the most accurate test.

In study, which researched HOMA-IR with cut off point 3.4 versus OGTT (Brar, 2014), sensitivity was 72.00% and specificity was 61.00%, with TP 13, FP 51, FN 5 and TN 80. In Figure 8 Summary ROC Plot of results, the curve of this meta-analysis is shown as a grey circle. The curve of this test starts on the TPF axis around 0.05 and most closely approaches the diagonal curve called the "useless test" curve.

The last single meta-analysis was made from the results of Ehehalt´s study (2017) with an index test FPG with cut off point 7.0 or 2h glucose versus OGTT. The sensitivity was 44.00% and specificity was 99.60%, with TP 22, FP 21, FN 28 and TN 5229. In Figure 8 Summary ROC Plot of results, it is shown as a yellow cross and starts on the TPF axis at a value of approximately 0.68 and from the result of the display we can see that this is the second most accurate result of these polled meta-analyses.

Based on the results of the meta-analyses and SROC, we can state that the most accurate results was found in Ehehalt´s study with an index test HbA1c at the level of cut off point 6.5 versus OGTT as a reference test. In Figure 8 Summary ROC Plot of results, the curve for this test is displayed in dark blue tringle and it is seen that it is at a value reaching the axis in point 1.

Therefore, it can be stated that in the SROC plot of listed tests with their cut off points, this test shows the most accurate value of all the tests used.

FPG as an index test and OGTT as a reference test were assessed in 4 studies. In Ehehalt´s study (2017), FPG was on levels of ≥126 mg/dL (which stands for ≥7.0 mmol/L) with 18% of sensitivity and 99.80% of specificity. The results of this study revealed that an optimal threshold for FPG is ≥98 mg/dL with sensitivity 70.00% and specificity 88.00% (which stands for ≥5.4 mmol/L) (Ehehalt et al., 2017).

In Chan´s study (2015), the threshold for FPG was 92 mg/dL (which stands for 5.1 mmol/L). The sensitivity and specificity were 80.00% and 50.00%. That shows a better result of the diagnostic accuracy of the FPG as a test index for the diagnosis of interest compared to the results of the study by Ehehalt (2017). But in Chan´s study, the reference test was determined as CGM data. CGM is a measurement of free-living glucose by continuous glucose monitoring using software accompanying iPro recorder, sensor readings were converted into excel format for each subject (Chan et al., 2015). It seems to be a new method in pre-diabetes identification. But more studies comparable in using of an index test FPG and reference test CGM are missing.

Another study which included FPG as an index test was Kim´s study (2019). There were used two thresholds: <126 mg/dL (which stands for <7.0 mmol/L) with sensitivity 85.1% and. mmol/L specificity 100.00%. The second cut off point was determined at ≥126 mg/dL. The problem is that data about sensitivity and specificity on the level at ≥126 mg/dL are not provided at this level of threshold. However, if we compare this threshold <126 mg/dL with sensitivity 85.1% and specificity 100.00%, the problem can be seen in the fact that the population from Kim´s study is only Korean (which is discussed below). The same threshold was used in Ehehalt´s study (2017) but we assume that the difference in sensitivity and specificity is because the population was much different. In studies of Ehehalt (2017) the optimal cut off point was 98 mg/dL with sensitivity 70.00% and specificity 88.00%. The methodology of Ehehalt´s study (2017) was observational multicentre study of youth in Germany with 4848 participants. This could be the reason, why two different cut off point in two different studies are determined. We can hypothesize that in Korean population, the most accurate cut off point could be <126 mg/dL with sensitivity 85.1% and specificity 100.00%. But this cut off point seems to be less accurate for the Western European population because the results from Ehehalt´s multicentre study stated cut off point on the level at 98 mg/dL. The same was proven in meta-analysis when the cut off point 7.0 was used, and the sensitivity was 44.00% and specificity was 99.60% The Chan´s study (2015), the cut off point for FPG was stated at level of 92 mg/dL. But this was determined based on the data obtain from CGM. There was no reference standard determined because this study was based more on the data collection (Chan, 2015). Therefore, there is the assumption that there are two thresholds with different accuracy based on the ethnicity and geographical-demographic factors.

The last study dealing with FPG as an index test was Maffeis´s study (2010). Population of this study was divided into four groups based on the pubertal stage according to Tanner scale. The cut off point for all groups was determined at the level 4.8. Sensitivity was the in pubertal boys (87.00%) and prepubertal girls (80.00%). Sensitivity in pubertal girls and prepubertal boys was

lower, for pubertal girls it was 75.00%, for prepubertal boys 66.00 Specificity was the highest in pubertal girls (58.00%) and prepubertal boys (61.00%). In prepubertal girls it was 58.00%, for pubertal boys it was 53.00%. The pooled sample was based on values 77.00% sensitivity and 58.00 of specificity. It can be said that the cut off point 4.8 is the best for pubertal girl with 75.00% of sensitivity and 65.00% of specificity.

We assessed 4 studies in the group of studies where index test was FPG and reference test was OGTT. Only in Maffeis´s study (2010) the population was divided into groups according to the (pre)pubertal stage. The population from the rest of three studies was in age range from 7 to 18 years. The most accurate threshold was proven in Kim´s study (2019). The threshold was on the level <7.0 mmol/L with sensitivity 85.1% and specificity 100.00%. We hypothesize that this may have been influenced by the population included in the study. The sample in Kim´s study (2019) was from Korea, the same ethnicity. The children were 12.56 +/- 3.44 years old and 45.3% of them was obese. In Chan´s study (2015), three different ethnicities (non-Hispanic white, black, and Hispanic) were enrolled. This could be a reason why the results with regard to sensitivity and specificity of the tests they are not as accurate as in Kim's study (2015), in which the population was of the same ethnicity. This assumption can also be confirmed by the fact that in the Maffeis´s study (2010), the population included in the research was only Italian and the age range was quite wide (from 4 – 17 years). The results of the Maffeis study (2010) showed higher sensitivity and specificity than the Chan´s study (2015). If we want to compare the results with the study by Ehehalt (2017), it is necessary to mention the fact that this study was defined as an observational multicentre analysis. The number of participants in this study was 4848. This is disproportionately more than in the studies by Kim (2019), Chan (2015) and Maffeis (2010). It can be said that the most accurate test for the detection of pre-diabetes was determined in Kim´s study (2019) on the level <7.0 mmol/L with 85.10% sensitivity and 100.00% specificity for Asian population. For pubertal girls, it seems to be the most accurate HOMA-IR with the level of cut off point 4.8 with 75.00% sensitivity and 65.00% specificity how it was proven in Maffeis´s study (2010) based on division the population into 4 groups according to Tanner scale.

The most commonly used reference test was OGTT and HOMA-IR. The interesting fact when evaluating these two tests in the studies was that in case of HOMA-IR the authors of some studies assessed the population included by gender or sexual maturity, which showed different results in assessing the diagnostic accuracy of HOMA-IR (as an index test) and OGTT (as a reference test).

In Atabek´s study (2007) and Brar´s study (2014), the same cut of point 2.7 was used. Another two studies with the same cut off point were Brar (2014) and Galhardo (2015). These studies (Brar, 2014) and (Galhardo, 2015) used cut off point 4.

The cut off point 2.7 was used in two studies with the time difference of 7 years from publication: Atabek (2007) and Brar (2014). The sensitivity and specificity in Atabek´s study were 80.00 % and 59.10 %. In Brar´s study, the sensitivity and specificity were 77.80 % and 45.80%. Atabek's study (2007) was therefore more accurate than the Brar´s study (2014) at the level of cut off point 2.7. When we consider the difference between the sensitivity and

specificity of these two studies at the cut off point 2.7 we can hypothesize that the population was not comparable. In Atabek´s study (2007), the Turkish population was 8-18 years old with BMI greater than or equal to the 95th percentile for age and gender; in Brar´s study (2014), the population included the patients with a suspicion of diabetes, and/or related comorbidities such as abnormal values of glucose, insulin, HbA1c, polycystic ovary syndrome, dyslipidaemia, hypertension, acanthosis nigricans and so on. Plus, in Brar´s study (2014), the population contained five different ethnicities living in NYC, USA. This can be the reason why there are differences in diagnostic accuracy of the test at the cut off point 2.7. Therefore, it can be stated that Atabek´s study (2007) brought more accurate results at the level of 2.7 with sensitivity 80.00% and specificity 59.10%.

In Brar´s study cut off points on the level of 3.1 (sensitivity 72.20% and specificity 56.10%), 3.4 (sensitivity 72.20% and specificity 60.70%) and 4 (sensitivity 61.10% and specificity 68.20%) was used. When pooled the cut off point 3.4 (Brar, 2014) in meta-analysis, the sensitivity was 72.00% and specificity was 61.00% It is necessary to mention that Atabek´s study included participants with mean age 10.86. Although another three cut off points (3.1, 3.4 and 4) are used in Brar´s study (2014), the most accurate stays the cut off point 2.7 from Atabek´s study (2007). When discussing the results from these two studies we must state that the population in both studies was the same in the number of participants included in the studies (Atabek´s – n = 148; Brar´s – n = 149) but differed significantly in the characteristics, which may affect the final result. Population in Atabek´s study (2007) was only from Turkey which indicates that they were probably the same ethnicity. According to the description in the study the participants were in a good health with normal thyroid function. In contrast, the population from the Brar´s study (2014) were included population represented Hispanic, white, Black, Asian and other ethnicities and the population were predisposed to pre-diabetes (e.g. abnormal values of glucose, hypertension, dyslipidaemia etc.). From this it can be concluded that if we have a uniform ethnic group that does not have associated comorbidities related to the occurrence of the diagnosis of interest, HOMA-IR with a cut off point 2.7 with 80.00% sensitivity and 59.10% specificity can be used. However, if we start from the fact that the population at Brar´s was diverse, then it appears to be the most accurate threshold 3.4 with 72.20% sensitivity and 60.70% specificity, as the Brar´s study (2014) showed.

Another study which included HOMA-IR as an index test, was study from Galhardo (2015). The range of cut off points was from 0.1 to 11.0. The interesting results were at the edges of this range. While in the 0.1 threshold the sensitivity was 100% and the specificity 0%, in the 11.0 threshold it was exactly the opposite (sensitivity 0%, specificity 100%). The recommended cut off point by ADA is 4.5. The same cut off point (4.5) came out as optimal in this study with sensitivity 77.00% and specificity 67.00%. Therefore, it is suggested by the authors of the study to be used on a first approach for the exclusion of patients with an adequate blood glucose level (Galhardo & Shield, 2015).

Galhardo´s study (2015) as well as in Brar´s study (2014), cut off point 4 was used. In Galhardo´s study, there was a higher sensitivity (80.00%) than in Brar´s study (61.10%). The specificity was opposite. In Galhardo study was lower (61.00%) than in Brar´s study (68.20%).

When we compare the results, we must state that the best results if it comes about sensitivity was in prepubertal girls because in both studies the sensitivity was 100.00%. The interesting fact is that in Kurtoğlu´s study (2010) the cut off point was for this group 2.22; in Maffeis´s study (2010) the cut off point was 2.85. The difference here is in specificity, In Kurtoğlu´s study (2010), the specificity in a group of prepubertal girls (cut off point 2.22) was 42.30%; in Maffeis´s study (2010) the specificity in the same group (cut off point 2.85) was 73.00%. An interesting finding would be whether the specificity in the population examined by Kurtoğlu (2010) would not increase if the cut off point of the group of prepubertal girls were changed. However, this is the most accurate test result for all 4 groups from both studies that were included. Thus, it can be argued that in prepubertal girls, based on a study by Maffeis (2010), the most accurate HOMA-IR is at the 2.85 level with 100% sensitivity and 73% Specificity. In other three groups there are big differences in results of sensitivity and specificity. Except of pubertal boys' group in Kurtoğlu´s study (2010) all groups have higher sensitivity than specificity. The sensitivity and specificity in pubertal boys ´group (Kurtoğlu, 2010) is 56.00% and 93.30%. The level of cut off point for this group was determined at 5.22. The cut off point for pubertal boys in Maffeis ´s study (2010) was lower – 3.22. Sensitivity was 75.00% and specificity was 67.00%. Another interesting fact is with groups of prepubertal boys. In both studies, almost the same cut off point was determined – Kurtoğlu (2010) – 2.67; Maffeis (2010) – 2.65. Although there is a difference in two hundredths, the difference in sensitivity and specificity is bigger. In Kurtoğlu´s study (2010), sensitivity was 88.20%, specificity was 65.50%; in Maffeis´s study (2010), sensitivity was 66.00%, specificity was 72.20%. We can therefore assume that the prepubertal boys have a more accurate test with a level of 2.67 (Kurtoğlu, 2010), although it differs by only two tenths from the cut off point used in the Maffeis´s study (2010). In group of pubertal girls were used cut off points 3.82 (Kurtoğlu) and 3.44 (Maffeis). It is not as big difference as in cut off points used for groups of pubertal boys in both studies. However, Kurtoğlu´s study (2010) showed more accurate results for this group in terms of sensitivity (77.10%) and specificity (71.40%); Maffeis´s study (2010) for this research group – sensitivity 65.00% and specificity 59.00%. Looking at the overall comparison of all the groups included in both studies, it can be stated that HOMA-IR with a cut off point of 3.82 can be used for the group of adolescent girls.

The third study divided the population to boys and girls was from Puri (2007). Unfortunately, there are information about mean age of each group missing, the same as Tanner scale of participants so it is not possible to compare this study with the previous one fully. The cut off point for girls was >4.5 – the sensitivity was 100.00% and specificity was 55.10%; the cut off point for boys was >13 – the sensitivity was 100.00% and specificity was 76.60% which can be considered as very good. The study results showed that a girl with HOMA >4.5 had an 18.6% chance of having an abnormal OGTT, and a boy with HOMA >13 had a 57.1% chance; cholesterol >200 mg/dl, 36.4% chance (Puri et al., 2007). When we compare the Puri´s study (2007) with studies of Kurtoğlu (2010) and Maffeis´s study (2010), we can say that the results need to be discussed with respect to the division into girls and boys. In Masfeeis´s study (2010) and Kurtoğlu´s study (2010), the participants were divided into four groups according to pubertal maturity and gender. However, in Puri´s study (2007), the population was divided only according to a gender. The question stays still how the division in Puri´s study was done because

it is not stated in the study. Therefore, we can only state that the results differ from the results of the Maffeis (2010) and Kurtoğlu (2010) studies. The reason is the different division of the population into individual groups.

If we were to compare this study with two others (Kurtoğlu, 2010; Maffies, 2010) in which groups of girls and boys were formed and divided according to puberty, we must state that the results of the Puri´s study (2007) do not correspond to the results reported by the other two studies. One reason may be that Puri's population is older (10-17 years); second reason can be that the study was focused on minority pubertal youth (African American, Caribbean Hispanic) in the USA. Another reason why the results cannot be discussed in more depth is the fact that Puri's study (2007) did not sufficiently explain the division into groups of boys and girls and their division according to puberty. Based on all the studies that examined HOMA-IR as an index test and OGTT as a reference test, we leaned towards results that are consistent with Galhard's study (2015). In Galhardo´s study (2015), the threshold was determined at level of 4.5 which is also in line with the ADA recommendation. In Puri´s study, cut off point for girls was 4.5 as well with 100.00% sensitivity and 55.10% specificity. Therefore, we can state that this result is confirmed. The interesting fact seems to be that the cut off point for pubertal boys was higher than the others in Puri´s study (2007) and in Kurtoğlu study (2010). In Puri´s study (2007), the cut off point for boys was 13; in Kurtoğlu´s study (2010), the cut off point was 5.22. Although there is a big difference in these two cut off points, the fact, that both groups of pubertal boys have higher cut off point than other groups can indicate the importance of different cut off points for individual groups of girls and boys also with regard to their pubertal maturity. However, the result for boy group from Puri´s study (2007) is very different from other results in studies dealing with identification of pre-diabetes in children. Although the number of participants in Puri´s study (2007) is similar to Maffeis (2010) and Kurtoğlu (2010), the difference is in the occurrence of symptoms indicating the presence of a diagnosis of interest because in Puri´s study (2007), participants with signs of insulin resistance (such as acanthosis nigricans) were included. This could be a reason why the results are different from studies from Kurtoğlu (2010) and Maffeis (2010).

If we take the results we obtained, we can state that for the population regardless of gender, age, ethnicity, the cut off point for HOMA-IR vs OGTT corresponds to the ADA recommendation, namely 4.5. The same follows from the findings of a study by Galhardo (2015) with sensitivity 77.00% and specificity 67.00%. For prepubertal girls, the 2.85 threshold appears to be the most favourable, as revealed by the Maffeis study (2010). In prepubertal boys, the most accurate cut off point appears to be 2.67, which emerged from the Kurtoğlu study (2010). In pubertal boys, cut off point 5.22 with sensitivity 56.00% and specificity 93.30% appears to be the most accurate, as revealed by the Kurtogğu study (2010). In pubertal girls, cut off point 3.82 appears to be the most accurate, as revealed by the Kurtoğlu study (2010).

Another group of studies used HbA1c as an index test and OGTT as a reference test. Blood glycated haemoglobin (HbA1c) is considered to be a more accurate tool for determining blood glucose and is currently the routine and most effective tool for monitoring the course of diabetes. This is an indicator of so-called "long-term blood glucose", as it provides information

on blood glucose for a period of 2-3 months. The value of glycated haemoglobin can be used in the screening of glucose homeostasis disorders, especially in relation to prediabetes (Čapková & Lustigová, 2017). The cut off point recommended by ADA is 5.7 for adult population.

In the Brar´s study (2014) the cut off points for HbA1c were 5.6 and 5.7 and 5.8 and 5.9. The highest sensitivity was demonstrated at cut off point 5.6 (83.30 %) but showed the lowest specificity (47.20). The highest specificity was demonstrated at the cut off point 5.9 (77.60 %) where the sensitivity was 66.70%. For the cut off point 5.7, the sensitivity was 75.00% and the specificity 57.60%. For the cut off point 5.8, the sensitivity was 66.70% and the specificity 65.50%. Based on the results of this study, a HbA1c threshold of 5.6 % may offer the best combination of sensitivity and specificity if the HbA1c test is used alone (Brar et al., 2014).

In Brar´s study (2014), there were participants who could be burdened with abnormal values glucose, insulin, polycystic ovary syndrome, dyslipidaemia, hypertension, or metabolic syndrome. As it was stated in Brar´s study (2014), the ethnic discrepancies in HbA1c in children can occur that are not explained by glycaemic status (Brar et al., 2014).

In the Nam´s study (2018), cut off point 5.8 was used for HbA1c. The sensitivity was 64.10 % and specificity was 83.80 %. (Nam et al., 2018). Comparing to Brar (2014) who used the same cut off point 5.8, we can state that the sensitivity is similar 64.10 %, respectively 66.70 %, but specificity is different: 83.80 %, respectively 65.60 %. However, as stated in the Nam´s study, the study discussed was conducted in a Korean population; hence, HbA1c cut off point may not be generalizable to other population (Nam et al., 2018). Further population in Nam´s study (2018) was 10 years and above with body mass index ≥ 85th percentile for age and gender and having two or more additional risk factors for diabetes, consistent with American Diabetes Association (ADA). So, if we assess the Brar´s study (2014) and Nam´s study (2018), the biggest difference is in ethnicity. The question stays whether this factor (ethnicity) can lead to different results.

In Chan´s study (2015) the same cut off point was used as in Brar´s study (2014) – at the level of 5.9. The sensitivity and specificity in this study was 80.00% and 64.00% (Chan et al., 2015). Comparing to Brar´s study, sensitivity was similar (80.00% - Chan, and 83.30 % - Brar) but specificity was different (64.00% - Chan, and 47.20 % Brar). When we compare the population from Brar´s study (2014) and Chan´s study (2015), we can state that there was a difference in them. While in the Chan´s study (2015) the population was without much burden with regard to the diagnosis of interest, in the population in the Brar´s study (2014) the population called for signs of suspicion of interest of interest such as abnormal values of glucose, insulin, HbA1c, dyslipidaemia, hypertension, acanthosis nigricans (Brar, 2014). This fact could lead to differences in test results at level 5.9. The most important finding of Chan´s study is that HbA1c is a measure of average glycemia whereas 2-h glucose reflects response to a glucose challenge. Data demonstrate that, despite the failure to identify one or the other test to be a stronger predictor of abnormal free-living glycemia, there are important and potentially clinically relevant differences in the pattern of relationships between HbA1c and 2-hour glucose (Chan

et al., 2015). Therefore, we can assume, that in Nam´s study (2015), the test with sensitivity 80.00% and specificity 64.00% at the level of 5.9 is more accurate for the population without much burden with regard to the diagnosis of interest.

In these three studies in the one from Brar´s study (2014), there are contained four cut off points which can be compared with cut off points from other three studies. The only cut off point which is not comparable within these studies is 5.6, which in Brar´s study showed the sensitivity 83.30% and specificity 47.20%. Unfortunately, there is no other study within all included studies with an index test HbA1c and reference test OGGT which bypassed the threshold level 5.6. Therefore, we can not compare other aspects that could led to one of the highest sensitivities within tests used in this group of diagnostic tests.

Other studies also worked with cut off points identical to those from Brar (2014), Nam (2018) and Chan (2015).

Another study in which the HbA1c as an index test was used was in Ehehalt´s study (2017). The threshold was determined at HbA1c ≥48 mmol/mol which corresponds with HbA1c≥6.5 mmol/mol. At this level, sensitivity and specificity was 84.00 % and 99.30 % which represents one of the best results when comparing with other studies results. When comparing the threshold HbA1c 39 mmol/mol which states ≥5.7, the sensitivity and specificity show differences (sensitivity 96.00% and specificity 75.60%). According to the ADA thresholds, the sensitivity for detecting pre-diabetes was better for HbA1c than that for OGTT. In this study is indicated that this could be due to a lower sensitivity of OGTT in obese children (Ehehalt et al., 2017). The reason why the cut off point ≥6.5 has higher sensitivity and specificity than cut off point ≥5.7 can be explained with the population included in the study. Unlike other studies included in the assessment, which had a maximum number of participants up to 500, this study is defined as an observational multicentre study. This study numbers 4848 participants from Germany. So, there is a huge number of participants who live in the territory of one state. However, the composition is not described in terms of nationality, but all participants included in the study were defined as obese. The mean age of participants was 13.1+/-2.4. Thus, it can be assumed that this population of children from Germany is more advanced on average, so cut off point 6.5, which is set by the ADA for the detection of pre-diabetes in the adult population, had a higher sensitivity but lower specificity than cut off point 5.7. If we compare the cut off point 5.7 with the studies in which it was also used (Brar, 2014) it can be argued that the results regarding sensitivity and specificity are very good. Cut off point 5.7 in the studies: Brar´s study (2014): sensitivity 75.00% and specificity 57.60; Ehehalt (2017) sensitivity 96.00% and specificity 75.60%. In Brar´s study (2014) is the result of both sensitivity and specificity the lowest. The reason can be seen in the fact that the population in this study was burdened by pre-existing predispositions to pre-diabetes. In the Ehehalt (2017) study, the inclusion criteria were (1) overweight, obese, and extremely obese children and adolescents aged 7 to 17 years; and (2) oral glucose tolerance test and HbA1c measurement on the same day. In Brar´s study (2014), the number of participants was 149. It was conducted in the USA and it included 5 different ethnicities. The inclusion criteria were suspicion of diabetes, and/or related comorbidities such as abnormal values of glucose, insulin, HbA1c, polycystic ovary syndrome, dyslipidemia,

hypertension, acanthosis nigricans, and metabolic syndrome and who had both OGTT and HbAlc tests performed within 3 months of one another. This could make a difference in the results in sensitivity on the level of cut off point at 5.7 with sensitivity 96.00 % and specificity 75.00 % (Ehehalt, 2017) and sensitivity 75.00% and specificity 57.60% (Brar, 2014). Therefore, we can hypothesize that in larger sample of participants where they can be/or can not be the burden of predispositions to the disease is the cut off point 5.7 more accurate than in a smaller participated studies with burden of predispositions to the disease of interest.

In Galhardo´s study (2015), nine different cut off point for HbA1c level is used. The recommended cut off which is used in this study is 5.7. This corresponds with ADA recommendation and with the results from the study of Brar (2014) and Ehehalt (2017) although in Galhardo´s study (2015), the sensitivity and specificity show the result 23.00 %, respectively 89.00%. Given that 95.1% of the population were free from symptoms leading to pre-diabetes (in Galhardo's study (2015) are marked as normal with normal glycemic), it can be assumed that for this reason the specificity is so high. The optimized cut off based on this study is 5.3 with sensitivity 62.00% and specificity 53.00 (Galhardo & Shield, 2015). As it is explained in the discussion part of this study, the reduction of the cut off point to 5.3 may correspond with (1) higher risk of complications associated to high blood glucose in young age (shown by its earlier onset compared to the adults) and (2) to higher physiological variability in this age group (for instance, according to the levels of haemoglobin, the glycosylation rate or the pubertal status) (Galhardo & Shield, 2015). Because no other study from all the included did not work with the level of cut off point 5.3 we can not compare the results. Although the reason, why this cut off point is optimized for this study, is explained in the study of Galhardo (2015) we can assume that the main reason, why the cut off point is optimized on this level is, that 95.1% of included participants are specific symptoms free.

Except of the cut of points 5.3 (which is optimized) and 5.7 (which is recommended by ADA) from Galhardo´s study (2015), another threshold which can be compared with other included studies is 5.9. The same cut off point is in Brar´s study (2014) and Chan´s study (2015). The sensitivity and specificity from these three studies are as follows: Brar´s study (2014): sensitivity 66.70% and specificity 77.60%; Chan´s study (2015): sensitivity 80.00% and specificity 64.00%; Galhardo´s study (2015): sensitivity 23.00% and specificity 96.00%. We can assume that the big difference between sensitivity and specificity can miss more participants suffering from disease of interest. Therefore, cut off point 5.7 still seems best to us, which is also supported by the results of other studies.

Mutlu (2013) stated three levels of threshold for HbA1c (5.5, 5.2, 5.3). The sensitivity and specificity at the level of 5.5 was 63.00%, respectively 70.00%; at the level of 5.2 it was 78.00%, respectively 37.00%, and at the level of 5.3 it was 72.00%, respectively 49.00. In this study was found a significant correlation between 2-hour glucose levels and FG, HbA1c, and one-hour glucose level. Based on that, the suggestion for cut off level of HbA1c is 5.5 (sensitivity 63.00% and specificity 70.00%) (Mutlu et al., 2013). When comparing this study with study of Galhardo (2015), who determined the optimized cut off point at 5.3, we can see the difference because in Mutlu´s study (2013) this cut off point is not that accurate as cut off

point 5.5. In Galhardo´s study (2015, the participants were from 90 % Caucasians, in Mutlu´s study (2013), the population was Turkish. The difference can be seen in the fact that in Multu´s study (2013) the participants underwent OGGT before they started participate in the study and they were defined as obese or overweigt. The BMI SDS in these children was 2.6+/-0.6. Based on the information from Galhardo´s study (2015), the participants had BMI z score 3.53+/-0.59. The presence of OGGT testing before participation in the study could cause different results in sensitivity and specificity at the level of cut off point 5.3. The recommended cut off point from Multu´s study (2013) is 5.5 with sensitivity 63.00% and specificity 70.00%. This cut off point was not used in any other study so we can not compare the results. Based on the heights of sensitivity and specificity at the level we can assume that this cut off point is accurate for Turkish population which shows signs of obesity or overweight. The question arises as to whether it is not possible to find a more accurate cut off point for children who are obese and for those who are morbidly obese.

The present studies revealed moderate agreement between HbA1c (as an index test) and OGGT (as a reference test) results. Results of the studies showed that the best combination for identification of pre-diabetes in children when using HbA1c and fasting glucose is potentially promising as a useful diagnostic method with the following refinement by OGTT to get more accurate identification in children with symptoms pre-diabetes. Sensitivity and specificity values differed from study to study. It should be noted here that the recommended cut off point for underestimation of prediabetes prevalence by ADA in the adult population is 6.5. As it was stated above, the ADA recommendation for identification of pre-diabetes in children is 5.7. This cut off point was determined in 4 studies which assessed HbA1c as an index test and OGTT as reference test. In study of Galhardo (2015), the optimal cut off point was determined as 5.3. Like it is explained in the discussion in Galhardo´s study, the reduction of cut off point to 5.3 may correspond with (1) higher risk of complications associated to high blood glucose in young age (shown by its earlier onset compared to the adults) and (2) to higher physiological variability in this age group (for instance, according to the levels of haemoglobin, the glycosylation rate or the pubertal status). However, they did not work with this fact in other studies, as other studies included participants who had symptoms that indicated metabolic changes, even if they were "only" overweight. Therefore, we cannot completely agree with the result achieved by Galhardo (2015) in his study. Among the other cut off points that showed high diagnostic accuracy, it was especially the cut off point 5.8, which turned out very well in the Brar (2014) and Nam (2018) studies. Sensitivity was 66.70% (Brar, 2014) and 64.10% (Nam, 2018); specificity was 65.50% (Brar, 2014) and 83.30% (Nam, 2018). However, it is important to reiterate that the population in these studies was somewhat different. In Brar´s study (2014), the population was Hispanic from 71 %, the study was conducted in the USA. In Nam´s study (2018), the population was Korean. The number of participants´ was in Nam´s study (2018) two and a half times higher than in Brar´s study (2014). On the other hand, the baseline characteristics were similar on both sides: 1) age 10 years and above or at the onset of puberty, 2) overweight or obese (body mass index [BMI] ≥ 85th percentile for age and gender), and 3) two or more additional risk factors for diabetes, consistent with American Diabetes Association (ADA) recommendations for type 2 DM screening, such as family history of type 2 DM, race or ethnicity, signs of insulin resistance or its associated conditions, maternal history

of DM or gestational DM. From this point of view, the level of HbA1c at level 5.8 seems to be a promising indicator with sensitivities and specificities confirmed in these studies: Sensitivity was 66.70% (Brar, 2014) and 64.10 (Nam, 2018); specificity was 65.50 (Brar, 2014) and 83.30% (Nam, 2018).

The best predictor of pre-diabetes from the included studies, the cut off point remains 5.7. It was recommended in studies from Brar (2014) – sensitivity 75.00%, specificity 57.60%; and Ehehalt (2017) – sensitivity 96.00% and specificity 75.60%. But the interesting fact is, that when the number were pooled in meta-analysis, the most accurate result was for cut off point 6.5 (Ehehalt, 2017) with sensitivity 84.00% and specificity 99.00%. The problem here is that unfortunately we do not have any other study assessing this level of cut off point to prove the hypothesis. One of the reason why the result in pooled meta-analysis was the most accurate is specific German population and high number of participants included in the study (n = 4848) characterized as an observational multicentre analysis. Although each study had a different number of participants and baseline characteristics at the beginning of the study, and each had a different enrolment of participants in the study, we can state that despite the difference, a cut off point of 5.7 was reached. Thus, it can not be said that HbA1c at the level of the cut off point 5.7 is indeed the "gold standard" for HbA1c although this cut off point is recommended by ADA. From the results of meta-analysis of the SR DTA, it was shown that index test of HbA1c and a reference test OGTT with the cut off point was 6.5 seems to be the most accurate in German population. However, it has been shown that there are differences between different participants groups for which a different cut off point might be more accurate with respect to glycemic index (Galhardo, 2015), obesity or overweight Nam, 2018, Brar, 2014), and the occurrence of pre-detected predispositions to pre-diabetes (Chan, 2015). Despite the fact that the average age is similar, we do not have accurate data on the exact age distribution and the data have not been evaluated - ie. when the average age is 12 years there can be a charge of 40 children 8 years old and vice versa 40 children 16 years old. In another study, on the other hand, there are actually children in mean age 12 years +/- 2.0 years, which may affect the results.

Within these 24 studies, there was one, where HOMA-IR was used as an index test and HbA1c was used as a reference test. In study from Sharma (2012) HOMA-IR with cut off point 2.5 with sensitivity 93.00% and specificity 21.00% was used. As it is stated in the study. The cut off point was determined on the previous research of Sharma et al from 2011 which was focused on identification of metabolic syndrome in African American Children Using HOMA-IR (Sushma Sharma, Lustig, & Fleming, 2011). The specificity of HOMA-IR is this study is very low. It can be cause by the fact that HOMA-IR as an index test and HbA1c as a reference test has together do not have as accurate diagnostic validity in terms of accuracy as other tests, e.g. OGTT as a reference test. Another reason can be seen in the fact that only African Americans with mean age 9.08 were included in this study therefore the specificity at the cut off point 2.5 is much lower then sensitivity. Within the studies, there is no other one which would compare HOMA-IR as an index test and HbA1c as reference test.

Two studies were assessing TyG as an index test and HOMA-IR as a reference test. In study from Kang (2017), TyG with the cut off point 8.18 was used. The sensitivity was 77.40% and

specificity was 64.80%. The study showed that the TyG index is a useful prediction marker of the development of type 2 diabetes for normal fasting glucose, whereas fasting plasma glucose had a higher hazard ratio than the TyG index for impaired fasting glucose. The TyG index might be a useful marker for identifying individuals with a high risk of developing diabetes. In addition, TyG index was significantly associated with an increased risk of developing CVD (Kang et al., 2017).

Another study using TyG and proposed TyG was from Garcia (2019). The cut off point for TyG was 8.5 (sensitivity 65.00%, specificity 25.70%) and for proposed TyG 8.83 (sensitivity 95.00%, specificity 42.30%).

When comparing studies of Kang (2017) and Garcia (2019) in which TyG was used as an index test and HOMA-IR was used as a reference test, we have to mention that in Garcia´s study (2019) there was comparison of TyG based on the formula used, which was calculated using the equation: Ln[fasting triglycerides in mg/dL x fasting glucose in mg/dL]/2. Proposed TyG on the level of 8.83 is the result of the determination obtained on the basis of the diagnostic accuracy of this index test. Although Garcia (2019) used the calculation to determine the proposed cut off point, the diagnostic transfer in the Kang study (2017) is more accurate. The reason may be the diversity of populations, because in the Kanga study (2017) it is a population that has a mean age of 11.1. The research was conducted in South Korea. In Gardia´s study (2019), mean age 8.0 participants who were obese and overweight were included. The study was conducted in Mexico. This can be the reason why there are difference in cut off point and sensitivity and specificity in Kang´s study (2017) is higher than in Garcia´s study (2019).

Another pair of tests used in the studies of Garcia (2019) and Kang (2017) was TG/HDL as an index test and HOMA-IR as a reference test. In Garcia (2017), the cut off point 8.18 was used and the sensitivity was 77.40% and specificity was 64.80%. The proposed value of threshold was 1.17 with sensitivity 95% and specificity 68.60%.

In Kang (2017), the cut off point used was 1.41 with sensitivity 72.20 and specificity 61.80.

When comparing studies of Kang (2017) and Garcia (2019) in which Tg/HDL was used as an index test and HOMA-IR was used as a reference test, we have to mention that in Garcia´s study (2019), there was comparison of Tg/HDL which was calculated with fasting triglycerides/fasting HDL. Proposed Tg/HDL on the level of 1.17 is the result of the determination obtained on the basis of the diagnostic accuracy of this index test. When comparing the results of the test, in this group of tests, proposed cut off point from Garcia´s study is more accurate (sensitivity 95.00% and specificity 68.60%) than in Kang´s study (sensitivity 72.70% and specificity 61.80%). It is necessary to mention that there is the diversity of populations, because in the Kang´s study (2017) it is a population that has a mean age of 11.1 and included non-diabetic subjects. The research was conducted in South Korea. In Gardia´s study (2019), mean age 8.0 participants who were obese and overweight were included. The study was conducted in Mexico. The question here arises if it is possible to compare two such a different study with the different baseline population.

TG_HDL-C ratio as an index test and OGTT as a reference test was used in study of Galhardo. Nine different cut off points in range 0.5 to 4.2. The cut off point 4.2 had the sensitivity 0%, but specificity 100.00%. From these 9 thresholds, 3.0 is recommended – the sensitivity is 18.00% and specificity is 98.00%. Based on the results of this study, cut off point 2.3 with sensitivity 27.00% and specificity 96.00% is optimized – the sensitivity is 27.00% and specificity is 96.00%. (Galhardo & Shield, 2015).

We compare using TG_HDL as an index test in studies of Garcia (2019), Kang (2017) and Galhardo (2015) with the reference test which may differ – HOMA-IR (for Garcia, 2019, and Kang, 2017) and OGTT (for Galhardo (2015). From the results it is seen that TG_HDL with HOMA-IR has more promising results than TG_HDL with OGTT if we consider the diagnostic accuracy of the tests. It can be stated that OGTT is more accurate, therefore we can hypothesize that in this comparison TG_HDL has a better result with a less accurate reference test. However, it is important to emphasize that the population of these studies is very divergent, so it would be useful to further investigate the possibilities for these diagnostic tests on a larger population, respectively different populations. If we compare the results described above with the results of HOMA-IR as an index of the test and OGTT as a reference of the test, we can state that the most accurate result for the population regardless of age, gender, ethnicity corresponds to the most accurate cut off point value 4.5 with sensitivity 77.010% and specificity 67.00% (Galhardo, 2015). The problem here is, that we can not compare the reference tests OGTT and HOMA-IR and their diagnostic accuracy because the OGGT was not used as an index test in these 24 included studies. Based on that we can hypothesize that OGTT as a reference test is more accurate.

In the study of Maffeis´s (2010), FSI as an index test and OGTT as a reference test were used. Although FSI as the index test was used only in this study, the results obtained are interesting. The population in the study was divided according to gender and pubertal development to four groups. The cut off points were determined as follows: for prepubertal girls 13 μU/ml with sensitivity 100.00% and specificity 69.00%; for pubertal girls 16 μU/ml with sensitivity 67.00% and specificity 57.00%; prepubertal boys 11 μU/ml with sensitivity 66.00% and specificity 54.00%; pubertal boys 14 μU/ml with sensitivity 75.00% and specificity 59.00%. It is seen that FSI showed homogenous sensitivity and specificity according to gender and puberty. The groups of participants included to the study were similar at the baseline characteristic. All participants were recruited to the study in Verona, Italy, which predicts the same ethnicity included to the study. However, from the results raised that it is necessary to determine specific threshold for each group to get as more accurate results as possible. Because within the included studies, there no other which would be focused on the assessment of FSI as an index test but the results from this study are promising, it could be recommended to suggest this test as an index to further testing to see if we can determine for different groups better matched cut off points, or which test to choose as a reference. The limitation here for this test used for identification of pre-diabetes (based on the results of Maffei´s study, 2010) would be the requirement for homogenous population when enrolling to the study.

OGGT is the gold standard as a reference test in most diagnostic test comparisons. SR DTA aimed to determine if a test has the ability to become a reference test. When we compare the tests and their diagnostic accuracy we will get the results. When we used FPG as an index test and OGTT as a reference test, we got the results which determine the cut off point ≥5.4mmol/L with sensitivity 70.00% and specificity 88.00% the most promising (Ehehalt, 2017). The limit for clear statement of this result was the methodological design of Ehehalt´s study (2017) because it was defined as observational multicentre analysis with 4848 participants from Germany. More accurate seemed to be study from Kim (2019) with cut off point - ≥7.0mmol/L with sensitivity 85.10% and specificity 100.00%. The limit for this study is that only Korean population was include in this study. So, the question here is whether the cut off point would indicate the same level of sensitivity and specificity in more diverse population or if we can hypothesize that FPG can be use only on homogenous population to get the most accurate results. Another question is about ethnicity because in Ehehalt´s study (2017), the population was European but in Kim´s study, the population was Asian. FPG as an index test and OGTT as a reference test was used in Maffeis´s study where the population was divided into 4 group according to gender and pubertal maturity. The best result came out at pubertal girls at the cut off point 4.8 with sensitivity 75.00% and specificity 65.00%.

However, the results are not the best when we compare them with HOMA-IR and HbA1c as an index test and OGTT as a reference test. When we should sum up the results of HOMA-IR as an index test and OGTT as a reference test we must state that the best results of determined cut off point was in Atabek´s study (2007). In this study, the cut off point used at levet 2.7 showed the sensitivity 80.00% and specificity 59.10%. The important information here is to say that population in Atabek´s study was Turkish – which means homogenous ethnicity, and without any comorbidities or predispositions to diagnosis of interest. This seemed to be important factor when determining the most accurate cut off point in pair of tests HOMA-IR (as an index test) and OGTT (as a reference test). Assumption for this statement are results from other studies with more diverse population or population with predispositions to the diagnosis of interest. If we had a study that involved different representations of ethnicities or races, it was reflected in the results of that study. We can hypothesize that the cut off point 3.4 with sensitivity 72.20% and specificity 60.70% (Brar´s study, 2014) was the most accurate. In this study 5 different ethnicities were included which may definitely influence the result and prove the fact that the cut off point can not be the same as for homogenous population in the study. In population of Caucasian with pre-existing predispositions to the diagnosis of interest, we can assume that cut off point 4.5 with sensitivity 77.00 and specificity 67.00% (Galhardo´s study, 2015) is the most accurate test.

When comparing HOMA-IR and OGTT vs FPG and OGTT as an identification tests for groups of youth divided into groups according to gender and Tanner scale, we must state that HOMA-IR in pubertal boys at the level 3.25 with sensitivity 75.00% and specificity 67.00% (Maffeis, 2010) is the most accurate cut off point. In prepubertal boys, HOMA-IR the cut off point at level of 2.67 with sensitivity 88.20% and specificity 65.50% (Kurtoğlu, 2010) is the most accurate. In pubertal girls, HOMA-IR cut off point at level 3.82 with sensitivity 77.10% and specificity 71.40% (Kurtoğlu, 2010) is the most accurate. When compared with the most

accurate result of FPG for the pubertal girls on the level of cut off point 4.8 with sensitivity 75.00% and specificity 65.00%, we can say that HOMA-IR in pubertal girls on the level of 3.82 with sensitivity 77.10% and specificity 71.40% is more accurate (Kurtoğlu, 2010). In prepubertal girls, HOMA-IR cut off point 2.85 with sensitivity 100.00% and specificity 73.00% (Maffeis, 2010) is the most accurate.

When we compare the results of HbA1c as an index test and OGTT as s reference test we must stay that Nam´s study (2015) the same as Kim´s study (2019) assessed the cut off point for prediabetes in Korean population. While Kim´s study considered the level of FPG ≥7.0 mmol/L (as an index test) with sensitivity 85.10% and specificity 100.00%, in Nam´s study (2015) with an index test HbA1c the most accurate cut off point at level 5.9 for this population, it showed 80.00% sensitivity and 64.00% specificity. Both of the populations from these two studies were similar at the baseline, but in Nam´s population there was 1) age 10 years and above or at the onset of puberty, 2) overweight or obese (body mass index [BMI] ≥ 85th percentile for age and gender), and 3) two or more additional risk factors for diabetes, consistent with American Diabetes Association (ADA) recommendations for type 2 DM screening, such as family history of type 2 DM, race or ethnicity, signs of insulin resistance or its associated conditions, maternal history of DM or gestational DM unburdened by predispositions to the disease. This reason could make the difference in sensitivity and specificity. We can assume that for Korean population which is burdened by predispositions to the prediabetes, the HbA1c test at the level of cut off point 5.9 with sensitivity 80.00% and specificity 64.00% is the most accurate (Nam, 2015). For Korean population unburden by predispositions to the disease the FPG test at the level of ≥126 mg/dL with sensitivity 85.10% and specificity 100.00% is the most accurate.

When considering HbA1c as an index test and OGTT as a reference test two cut off points were the most accurate – 5.7 and 5.8. Cut off point 5.7 is recommended by ADA. These two cut off points were included altogether in 4 studies (Brar, 2014; Ehehalt, 2017; Galhardo, 2015; Nam, 2017). When we compare the sensitivities and specificities of these two cut off points in listed studies, the highest accuracy is shown at the level of cut off point 5.7 – 96.00% sensitivity and 75.60% specificity (Ehehalt, 2017). In Galhardo´s study (2015), sensitivity was 23.00% and specificity was 89.00%. But like we discussed before, the population contained in this study was free from symptoms of prediabetes. Brar´s study (2014) was the only one which compare these two cut off point together. The results at the level of cut off point 5.7 is 75.00% sensitivity and 57.60% specificity, contrary at the level of cut off point 5.8, the sensitivity was 66.70% and specificity was 65.50%. In Nam´s study (2017), at the level of cut off point 5.8, the sensitivity was 64.10% and specificity was 83.30%.

Galhardo (2015) used in the study the index test TG_HDL as an index test and OGTT as a reference test. The recommended cut off point is 3.0 with sensitivity 18.00% and specificity 98.00%. From the Galhardo´s study raised that optimized cut off point is 2.3 with sensitivity 27.00% and specificity 96.00%. But these values do not show as accurate as the test values recommended by the ADA and IDF.

In addition to OGTT as a reference test, HOMA-IR was also used as a reference test. TyG and TG_HDL were used as an index test.

When comparing the results of TG-HDL as an index test and HOMA-IR as a reference test, we can use two studies from Kang (2017) and Garcia (2019). In Kang´s study (2017), the cut off point was 1.14 with sensitivity 72.20% and specificity 61.80%. In Garcia ´s study (2019), the cut off point was 8.18 with sensitivity 77.40% and specificity 64.80%. The proposed cut off point in this study was 1.17 with sensitivity 95.00% and specificity 68.60%. The same results rised from the meta-analysis where TH_HDL at the level of cut off point was 1.71 with sensitivity 95.00% and specificity 69.00%. When comparing these results with sensitivity and specificity from Galhardo´s study (2015), we hypothesize that OGGT still is more accurate index standard to TG_HDL test than HOMA-IR.

The number of studies identified (24) and patients enrolled 14 382 were sufficient to answer the review question about diagnostic accuracy identifying pre-diabetes in children. But we have to state that patient enrolment, using of both index test and reference standard, and clinical setting we not homogenously suitable for the analysis across all studies. As expected in diagnostic test accuracy meta-analysis and heterogeneity was a problem. Therefore, the main part on which we can make a conclusion for this SR DTA is based mainly on narrative description of the studies.

Based on the results of all 24 included studies, we can state that the reference test certainly seems to be the OGTT. However, it is not possible to determine exactly which of the tests used in each study could be designated as the new reference test. However, what is clear from the results of these 24 included studies is the finding that the use of index tests and their exact cut off points may vary in terms of population, age, gender, sexual maturity, body physiognomy, or geographical-demographic factors. on the composition of the size of the selected population sample, as well as on the methodology of the study, which should meet the requirements for the creation of a study dealing with the diagnostic accuracy of tests-

In the end, there is one important question whether we can determine only one test to say “This is “gold standard”. Based on the results of this review it seems to be proven that although there are gold standards of two diagnostic test existing (HbA1C – cut off 5.7, HOMA-IR cut off 4.5 – both recommended by ADA), they are not always usable and suitable for individual population. The need for more test variability is one of the outcomes of this SR DTA. In conclusion, it must be said that HbA1c at the level 5.7 has not been shown to be the most accurate test in diagnosing pre-diabetes in children. This also needs to be further verified in future research. Based on the meta-analysis, it was found out that index test of HbA1c and a reference test OGTT with the cut off point was 6.5 seems to be the most accurate in German population.

3.2.1. Limitations of included studies

The biggest limitation of included studies is seen in the fact that almost all of them did not follow STARD guideline. This was particularly noticeable in the description of the individual parts of these studies, in which some parts were insufficiently described or completely missing.

Only three studies stated the final numbers of children who suffered from pre-diabetes after ending of testing (Brar et al., 2014), (Garcia et al., 2019), (Nam et al., 2017).

Another limit was the selection of participants. Some studies were based solely on medical documentation. Other studies included participants who volunteered for the study.

Also, the fact that the participants were not divided into subgroups, even in cases where it was a group in the age range of 4-17 and 5-18 years.

Some of the study did not have a suitable study design for diagnostic accuracy. Therefore, it was very hard or impossible to obtain the data necessary for more reliable results of this SR DTA.

3.2.2. Limitations of the review

This review had some limitations. The main limitation of this review is that, despite the high number of patients enrolled, heterogeneity is remarkably high, and we were not able to perform the meta-analysis to investigate the accuracy of included diagnostic tests for all studies statistically.

Further, it must be stated to the results of meta-analyses and SROC that this result has its limits, mainly due to the different population that was included in the studies, different ages of different populations, different ethnicities, but also in terms of methodologies of individual studies and procedures for collecting data from individual diagnostic tests. It is also important to mention the fact that only four studies could be calculated TP, TN, FN a FP. Therefore, the following findings from the 24 included studies were described by narrative synthesis.

Despite the great efforts of the authors to perform a reliable and exhaustive systematic literature search including hand search of databases and literature reference lists, it is a small chance that some studies that could be included in the review were omitted. Except of this, the population from included studies was very variable and there are differences in the detection of metabolic changes in children in different countries and based on different national standards in the countries in which the studies took place.

3.2.3. Strengths of the review

Strengths of this systematic review of diagnostic test accuracy include extensive literature search, assessment of risk of bias using and data extraction using standardized System for the Unified Management of the Assessment and Review of Information (SUMARI) from the JBI, and detailed narrative description of the results of the included studies. The manual calculation was performed, so meta-analyses and SROC analyses were possible for four studies and nine comparisons of different tests and their cut off points respectively. The results of this review have revealed that there is a need for further research in this area. Many studies assess the prevalence of overweight and obesity as well as various blood parameters of people with T2DM, especially in the adult population. However, only in a few studies they are all analysed together and considered in terms of pre-diabetes.

3.3. Implications for practice

In the SR DTA, 24 studies were examined, and 14 382 patients were enrolled. It was possible to pool the meta-analyses from total of four studies. Following implications for practice might be formulated based on the results.

- **The most accurate cut off point for FPG as an index test and OGTT as a reference test for homogenous Korean population (aged 12.5+/-3.44, 52.1% girls, BMI not known) is cut off point ≥7.0 mmol/L with sensitivity 85.10% and specificity 100.00% (Kim, 2019).**

Index test	Reference test	Cut off point	Sensitivity (%)	Specificity (%)	Population	Author, year
FPG	OGTT	≥7.0 mmol/L	85.10	100.00	N = 190, South Korea	Kim, 2019

- **The most accurate cut off point for FPG as an index test and OGTT as a reference test for homogenous German population (aged 13.1+/-2.4, 55% girls, BMI 30.6+/-5.4 kg/m^2) is cut off point ≥7.0 mmol/L with sensitivity 44.00% and specificity 99.60% (Ehehalt, 2017).**

Index test	Reference test	Cut off point	Sensitivity (%)	Specificity (%)	Population	Author, year
FPG	OGTT	≥7.0 mmol/L	44.00	99.60	N = 4848, Germany	Ehehalt, 2017

This test is based on the overall synthesis in the Kim study (2019) at 85.10% sensitivity and 100.00% specificity, but the result could only be reported by narrative synthesis. However, from a meta-analysis performed from the results of the Ehehalt study (2017), FPG emerged as the most accurate test with a sensitivity of 44.00% and a specificity of 99.60% as the most accurate for the German population. All the most accurate tests are for specific populations, but

they are quite similar. Although we do not know the BMI of the Korean population, it can be assumed that since the results are similarly accurate, the Korean population was also obese.

- **The most accurate cut off point for HbA1c as an index test and OGTT as a reference test for young German population (aged 13.1+/-2.4, 55% girls, BMI 30.6+/-5.4 kg/m^2) included in the study is cut off point 6.5 with sensitivity 84.00% and specificity 99.00% (Ehehalt, 2017).**

Index test	Reference test	Cut off point	Sensitivity (%)	Specificity (%)	Population	Author, year
HbA1c	**OGTT**	**6.5**	**84.00**	**99.00**	**N = 4848, Germany**	**Ehehalt, 2017**

- **The most accurate cut off point for HbA1c as an index test and OGTT as a reference test for homogenous Korean population (aged 13.0+/-2.5, 54% girls, BMI Z score 2.3+/-0.8) is cut off point 5.9 with sensitivity 80.00% and specificity 64.00% (Nam, 2017).**

Index test	Reference test	Cut off point	Sensitivity (%)	Specificity (%)	Population	Author, year
HbA1c	**OGTT**	**5.9**	**80.00**	**64.00**	**N = 389, S. Korea**	**Nam, 2017**

For the studies from Nam (2018) and Ehehalt (2017), 2 different cut off points were used. It can be assumed that this is due to the use of two different study designs. In Nam´s study (2018) was used restrospective chart review, in Ehehalt´s study (2017) desing was characterized as an observational multicenter analysis.

- The most accurate cut off point for HOMA-IR as an index test and OGTT as a reference test for homogenous Turkish population (aged 10.86+/-3.08, 58% girls, BMI not known) is cut off point 2.7 with sensitivity 80.00% and specificity 59.10% (Atabek, 2007).

Index test	Reference test	Cut off point	Sensitivity (%)	Specificity (%)	Population	Author, year
HOMA-IR	OGTT	2.7	80.00	59.10	N = 148, Turkey	Atabek, 2007

The most accurate cut off point for HOMA-IR as an index test and OGTT as a reference test for inhomogeneous population (aged 13.8+/-3.1, total 149 participants, BMI Z score 2.3+/-0.5) representing the ethnicities of Hispanic/White/Black/Asian/others included in the study is cut off point 3.4 with sensitivity 72.00% and specificity 60.70% Brar, 2014). This was confirned by the meta-analysis where an index test HOMA-IR at the level of cut off point 3.4 versus OGTT was evaluated as the eighth most accurate test.

Index test	Reference test	Cut off point	Sensitivity (%)	Specificity (%)	Population	Author, year
HOMA-IR	OGTT	3.4	72.00	60.70	N = 149, USA	Brar, 2014

- The most accurate cut off point for HOMA-IR as an index test and OGTT as a reference test for homogeneous population (aged 12.3, 55% girls, BMI Z score 3.35+/-0.59) of Caucasian included in the study is cut off point 4.5 with sensitivity 77.00% and specificity 67.00% (Galhardo, 2015).

Index test	Reference test	Cut off point	Sensitivity (%)	Specificity (%)	Population	Author, year
HOMA-IR	OGTT	4.5	77.00	67.00	N = 266, UK	Galhardo, 2015

- The most accurate cut off point for HOMA-IR as an index test and OGTT as a reference test for pubertal boys (aged 11.4+/-2.5, BMI not known) according to Tanner scale included in the study is cut off point 3.25 with sensitivity 75.00% and specificity 67.00% (Maffeis, 2010).

Index test	Reference test	Cut off point	Sensitivity (%)	Specificity (%)	Population	Author, year
HOMA-IR	OGTT	3.25	75.00	67.00	N = 315, Italy	Maffeis, 2010

- The most accurate cut off point for HOMA-IR as an index test and OGTT as a reference test for prepubertal boys (aged 11.4+/-2.5, BMI not known) according to Tanner scale included in the study is cut off point 2.67 with sensitivity 88.20% and specificity 65.50% (Kurtoğlu, 2010).

Index test	Reference test	Cut off point	Sensitivity (%)	Specificity (%)	Population	Author, year
HOMA-IR	OGTT	2.67	88.20	65.50	N = 127, Turkey	Kurtoğlu, 2010

- The most accurate cut off point for HOMA-IR as an index test and OGTT as a reference test for pubertal girls (aged 11.1+/-2.5, BMI not known) according to Tanner scale included in the study is cut off point 3.82 with sensitivity 77.10% and specificity 71.40% (Kurtoğlu, 2010).

Index test	Reference test	Cut off point	Sensitivity (%)	Specificity (%)	Population	Author, year
HOMA-IR	OGTT	3.82	77.10	71.40	N = 141, Turkey	Kurtoğlu, 2010

- The most accurate cut off point for HOMA-IR as an index test and OGTT as a reference test for prepubertal girls (aged 11.1+/-2.5, BMI not known) according to Tanner scale included in the study is cut off point 2.85 with sensitivity 100.00% and specificity 73.00% (Maffeis, 2010).

Index test	Reference test	Cut off point	Sensitivity (%)	Specificity (%)	Population	Author, year
HOMA-IR	OGTT	3.85	100.00	73.00	N = 248, Turkey	Kurtoğlu, 2010

- **The most accurate cut off point for TG_HDL as an index test and HOMA-IR as a reference test for Mexican population (aged 8.0, 58% girls, BMI not known) included in the study is cut off point 1.71 with sensitivity 95.00% and specificity 68.60% (Garcia, 2019). This was confirned by the meta-analysis where an index test TG_HDL at the level of cut off point 1.71 versus HOMA-IR was evaluated as the third most accurate test.**

Index test	**Reference test**	**Cut off point**	**Sensitivity (%)**	**Specificity (%)**	**Population**	**Author, year**
TG_HDL	**HOMA-IR**	**1.71**	**95.00**	**68.60**	**N = 201, Mexico**	**Garcia, 2019**

Based on the results, we got from the meta-analyses, we can say that the most accurate cut off point is 6.5 in HbA1c as an index test with sensitivity 84.00% and specificity 99.00% versus OGTT as a reference test (Ehehalt, 2017). The second most accurate cut off point is 7.0 in FPG as an index test with 85.10% sensitivity and 100.00% specificity versus OGTT as a reference test (Kim, 2019). And the third most accurate cut off point is 1.71 in TG_HDL as an index test with 95.00% sensitivity and 68.60% specificity versus HOMA-IR (Garcia, 2019). All of these three results were confirmed by the meta-analyses. When we compare the baseline characteristic of the participants included in these three studies, we can state that the sex representation of the children included in each study was similar, around 50% and 50%. The population in Ehehelt´s study (2017) and Kim´s study (2019) was approximately the same (Ehehalt´s study – 13.1+/-2.4; Kim´s study – 12.5+/-3.44). In Garcia´s study (2019), the population was younger – 8.0 year which could play important role if we are considering selecting the most accurate test for a given age group. However, this hypothesis must be confirmed by further research (please, see Chapter 3.4). Unfortunately, we do not have

complete data about the obesity or BMI from all three studies. Only in Ehehalt´s study (2017) was stated that the BMI of participants was 30.6+/-5.4 kg/m^2. But we did not get this information from either Kim´s or Garcia's study. Another information we lack is the effect of risk factors on the incidence of pre-diabetes in children. This information, e.g. on eating habits, sedentary lifestyle, physical activity, was not included in the description of the population in the individual studies. However, this information we consider as a useful for practice.

3.4. Implications and recommendations for research

The following part will describe the results of the narrative synthesis of this SR DTA.

Although SR DTA had no time-limit on published studies, more relevant studies were found after 2010. Nevertheless, it is necessary to be able to distinguish and predict changes between the effect of confounders, which relate to ontological, geographical or demographic factors. There is growing evidence that OGTT as the gold standard in the diagnosis of pre-diabetes in adults is not the gold standard in the diagnosis of pre-diabetes in children. In this SR DTA, we did come across interesting and new research that included not only standard index / reference tests, but also attempts to uncover new diagnostic tools.

1. **Defining the enrolment** – the studies should have a cross-sectional design and the enrolment should be random or at least consecutive; the enrolment should be clearly described including recruitment centres.
2. **Defining the population** – the studies should clearly state what is their population with regard to age, gender, ethnicity, pubertal maturity, genetic predisposition to the disease of interest, or already established risks of the disease.
3. **Defining the ontogenetic stage of population** – the ontogenetic development should be respected in the studies dealing with metabolic diseases because the development of the metabolic system of each individual is individual and is influenced not only by age, but also by gender, ethnicity and other geographical and demographic factors.
4. **Defining cut off point of the tests, time frame, and time interval** – the studies should clearly state the period when study was carried out (the beginning and the end date), index/reference time interval should be described, and the cut off point should be defined clearly.
5. **Publishing results with transparency** – primary, raw data provided standardly in 2x2 tables in almost all studies were missing. This fact limits researchers who want to perform an extensive analysis of the results of their research in terms of comparisons across studies, reducing transparency, increasing the risk of systematic bias, and overall contributes to the difficulty of transmitting scientific evidence.
6. **STARD** – all studies should be strictly developed using STARD guideline.
7. **There are two suggestions about the further research**:
    a) To answer the question about diagnostic test accuracy identifying pre-diabetes it is necessary to make more studies dealing with this topic. However, it is important to monitor the participants involved with regard to their ontogenetic development, gender, pubertal maturity and ethnicity. The studies, that have

been carried out to March 2020 focused on the topic of diagnostic tests revealing pre-diabetes in children, have very different characteristics in terms of the baseline characteristics of the population (ontogenetic development, gender, pubertal maturity and ethnicity). In the case of ontogenetic development, which according to the IDF is divided into 3 basic ontogenetic stages, the recommendation in this point (1) is to divide the study population according to these criteria a) either examine different index / reference tests in populations representing one stage of ontogenetic development or b) use the same index / reference test on a population representing all 3 ontogenetic stages and compare diagnostic accuracy across these populations. Both of these research strategies can also be recommended for baseline characteristics related to puberty or gender. Ethnicity depends on the geographical location of the study where it is conducted. Nevertheless, in studies carried out up to March 2020, the population was not evenly represented in terms of the representation of individual ethnicities. The recommendation in this point (2) requires an equally by numbers representation of individual ethnic groups in future research, which would be a representative sample of the research.

b) To answer the question about diagnostic test accuracy identifying pre-diabetes it is necessary to evaluate not only one test, but more tests which can be used in the initial phase of identification followed by other test which can confirm the diagnosis of interest. This SR DTA revealed index tests (FGIR, QUICKY, 2-h glucose, fructosamine, glycated albumin, 1,5-anhydroglucitol, glucose peak> 30 minutes, monophasic curve, 1-h glucose 155 mg / dL. COMBO, FSI,% OD adjusted percentage of oxidized 13C-glucose dose at 180 minutes, 1 / IF, BMI, waist circumference) and reference tests (HOMA top quartile, fasting glucose 100 mg / dL, 2-hr glucose 140 mg / dL, FPI≥p90, PI≥ 65 μU / ml, insulin-stimulated glucose disposal (Rd), which could not be further compared with other tests for two main reasons: 1) the was no other study used the same index / reference test; 2) the results of the index test were completely missing for the individual index test. Furthermore, six pairs of tests were revealed in this SR DTA that could be compared with each other (index tests: HOMA-IR, HbA1c, TyG, TG_HDL, FPG; reference test: OGTT, HOMA-IR) We hypothesize that with a better managed methodology of studies that used index / reference tests that were impossible be compared, some of them have the potential to be further investigated for their accuracy in detecting pre-diabetes in children when used in the light of the recommendations contained therein. in paragraph 7a) on ontogenetic development, gender, pubertal maturity and ethnicity. However, as it was not possible to obtain more relevant results in the given studies, it is necessary to add as one of the recommendations to fill the gap in research concerning the diagnosis of pre-diabetes in children by methodologically well-conducted cross-sectional studies.

c) To answer the question about recommendations of reference tests for the research, we can state that based on the results of this SR DTA, OGTT was proven as a “gold standard” which could be used as a reference test in the future

research in diagnosing of pre-diabetes. Another test which was tested as a reference test was HOMA-IR. Based on our results, it was shown that FPG at the cut-off level of 7.0 and HbA1c at the cut-off level of 6.5 can also be considered. These tests also showed promising results in some studies. Therefore, our recommendation for research is to use HOMA-IR as a reference test and try to find the most accurate threshold in an enrolled population of each future study. As written in Chapter 2.5.2.1, some studies have also examined the accuracy of new diagnostic tests that could be used as a reference test: HOMA top quartile, fasting glucose 100 mg/dL, 2-hr glucose 140 mg/dL, FPI≥p90, PI≥65 μU/ml, insulin-stimulated glucose disposal (Rd). In the future research, attention should also be paid to these tests, which could be used as a reference test.

d) Very promising seems to be FPG as an index test especially in young Korean population. This test has the most accurate values, so there is a recommendation to use this test not only in the young Korean population, but also in other types of populations for which it would be appropriate to find the most accurate cut off point. Furthermore, HbA1c at level 5.7 (which is recommended as a cut off point by ADA) has not been shown to be the most accurate test for the generalized population of children. This is one of the main results raised from the SR DTA. This also needs to be further verified in future research. Based on the meta-analysis, it was found out that index test of HbA1c and a reference test OGTT with the cut off point was 6.5 seems to be the most accurate in German population. This needs to be tested in other types of population as well. Another promising test seems to be HOMA-IR. In our SR DTA, the value of 4.5 was proved to be the most accurate, which corresponds to the ADA recommendation. However, even this result should be confirmed by more research. As written in Chapter 2.5.2.1, some studies have also examined the accuracy of new diagnostic tests that could be used as an index test: FGIR, QUICKY, 2-h glucose, fructosamine, glycated albumin, 1.5-anhydroglucitol, glucose peak>30 minutes, monophasic curve, 1-h glucose 155 mg/dL. COMBO, FSI, % OD adjusted percentage of oxidized ^{13}C-glucose dose at 180 minutes, 1/IF, BMI, waist circumference. In the future research, attention should also be paid to these tests, which could be used as an index test.

3.5. Conclusion

Based on the results of the studies included in this SR DTA, we can state that meta-analysis was possible to conduct. Although clinical and methodological aspects of the studies were very different, we were able to manually calculate data from (Brar et al, 2014), (Ehehalt, 2017), (Garcia, 2019) and (Nam, 2018). The most accurate result was found in Ehehalt´s study. An index test of this study was HbA1c and a reference test was OGTT, the cut off point was 6.5. In summary ROC plot of all 9 tests, the analysis showed this cut off point to be the most accurate

in German population, with sensitivity 84.00% and specificity 99.00% The level of confidence in the findings of the included studies is very low certainty, because in particular the definition of the population and its distribution in individual studies can be taken as a counfounding factor. Following an initial scoping of the literature, it was hypothesised that a new "gold standard" could be found in the diagnosis of pre-diabetes in children, as foreshadowed in several articles on the diagnostic accuracy of diagnostic tests for this disease of interest.

A series of protocols for each study following STARD guideline need to be developed to ensure provision of clarity because clinical researchers need to use similar standard practices and methods to aid comparison of different cut off points for different diagnostic tests detecting pre-diabetes in children in a different ontogenetic stage of development. In addition, study authors should always publish all details. Plus, using of the preferred study design for synthesizing evidence of diagnostic accuracy (cross-sectional study) would provide more appropriate and accurate data for this type of systematic review.

The narrative synthesis of this SR DTA included 24 studies that met the criteria for inclusion according to the acronym PIRD. All 24 studies were analysed for index / reference test and their sensitivity and specificity for the cut off point determined in each study.

The most accurate index test from all of 24 included studies was detected in Kim´s study (2019). This study examined the diagnostic accuracy of the index / reference tests of FPG / OGTT. The results of the study are as follows: for FPG as an index test and OGTT as a reference test for homogenous Korean population is cut off point ≥7.0 mmol/L with sensitivity 85.10% and specificity 100.00%. The total number of participants was n = 190; mean age was 12.56+/-3.44; sex (%) was n = 99 (52.1) females, and n = 91 (47.9) males. All participants were consecutively enrolled in Chonbuk National University Children´s Hospital between 2010 and 2017. All participants were of the same ethnic group (Kim, 2019).

As discussed earlier, each diagnostic test has its own specifics and method of execution. FPG is collected from patients who have been fasting for 8 hours. It is necessary to use a container with an antiglycolytic mixture (EDTA + NaF) and it is necessary to collect non-coagulating blood, because FPG cannot be examined from serum. The OGTT as „gold standard“ has more demanding requirements for the method of execution (e.g. 12 hours fasting), the number of samples (affected by the result of the first sampling) and the time needed to complete the results (more than 2 hours), determine the values and draw conclusions. Both tests require the collection of non-coagulated venous blood, rapid transport to the laboratory and separation of blood cells and plasma. For FPG, it is necessary to separate the plasma from other elements within 60 minutes after collection. If the glucose value does not exceed a defined amount at this point, the test should be supplemented by OGGT testing. From this point of view, we can conclude that FPG is less invasive, less demanding and easier to perform, both for the patient undergoing the test and for the physician performing the test, especially in the first phase of the detecting of pre-diabetes. This conclusion based on the preference of FPG as a test index was identified on a specific population consisting of a homogeneous Korean paediatric population, but nevertheless shows the most accurate diagnostic accuracy among the tests from all 24 studies included in that SR DTA.

The second most accurate index test from all of 24 included studies was detected in Ehehalt´s study (2017). This study examined the diagnostic accuracy of the index / reference tests of HbA1c versus OGTT. The results of the study are as follows: for HbA1c as an index test and OGTT as a reference test for homogenous German population is cut off point 6.5 with sensitivity 84.00% and specificity 99.00%. The total number of participants was n = 4848; mean age was 13.1+/-2.4; sex (%) was n = 2668 (55%) females. The study was characterized as an observational multicentre analysis. A total of 6 medical facilities in Germany were involved in the study.

If we consider the use of the Glycated Haemoglobin (HbA1c) Test, we will determine the average blood glucose level usually in the last 2-3 months. In some cases, the measurement period may be shortened or even extended to 6 months. Haemoglobin is a high-iron protein found inside red blood cells. After the patient receives a diet, glucose is released into the blood and begins to bind to haemoglobin in the red blood cells. Thus, glycated haemoglobin expresses what portion of glucose is associated with haemoglobin, and this value is expressed as a percentage. Red blood cells live in the body for about 3 months and then disappear in the spleen and liver and re-form in the bone marrow. Therefore, the glycated haemoglobin test reflects the amount of bound glucose in the last 3 months or so. At the doctor's visit, blood will be taken from the patient's hand and then sent for examination. The amount of glucose bound to haemoglobin does not affect the type of diet in the short term and therefore fasting is not required before the test itself. As the test results do not affect the diet before the test itself, no preparation is necessary. This test can also be performed at home using a purchased device for measuring A1c values. However, a home test should not be considered a substitute for a physician test, but can be used to get an idea of glycated haemoglobin levels. Although the test is used to determine glucose levels in the last 2-3 months, the specific frequency of measurements will ultimately depend on the doctor, who may recommend measurements more often based on the health, regimen, diet and many other factors. A more significant change in glycated haemoglobin can be seen two weeks after the first measurement, so in the event of large changes in diet, exercise or in the amounts or types of medication patients are taking, the frequency of the measurement may increase. If we consider the use of the HbA1c test, which was the most accurate test in the German population at the level of cut off point 6.5, we must state that the great advantage is the fact that the patient is not limited in intake of carbohydrates or other diets as in other tests. We consider the longer time-consuming to be a disadvantage, because repeated measurements are performed. However, as already shown from the information above, HbA1c measurements can also be performed at home.

The HOMA-IR (Homeostasis Model Assessment of Insulin Resistance) index can be used to assess insulin resistance. It is calculated according to the formula: HOMA-IR = fasting insulin (μU / ml) x fasting glucose (mmol / l) / 22.5. An increase in HOMA-IR values is seen with an increase in fasting glucose or insulin levels. This is consistent with increased cell and tissue insulin resistance and an increased risk of T2DM and cardiovascular disease. The calculation of this indicator can also be used if there is a suspicion of the development of insulin resistance in polycystic ovary syndrome in women, gestational diabetes mellitus, chronic renal failure, chronic hepatitis B and C, non-alcoholic liver steatosis, a number of infectious, oncological,

autoimmune diseases and treatment with certain drugs. (glucocorticoids, oral contraceptives and others). If it is necessary to perform tests to determine the HOMA index, a number of mandatory rules should be followed: Blood sampling for analysis must be performed in the morning between 8 and 11 o'clock. The patient must not eat for 8 to 14 hours before taking blood. Only water is allowed. It is necessary to reduce food intake before the day of testing. Blood is taken from a vein to detect metabolic disorders. The HOMA index is calculated according to the following formula: IRI - content of immunoreactive insulin contained in the blood; FPG - glucose in blood plasma. Calculating HOMA-IR is a more complex process in the sense that it is a mathematical model that presents more data that it calculates. For the patient, fasting may be somewhat limiting, which should last 8 to 14 hours before collection. At the same time, it is required to reduce nutrient intake one day before testing. A certain advantage of HOMA-IR is that it can be calculated using software (provided we have a laboratory in which we process the samples taken). This software is also available online, so it can be equipped in a doctor's office quite easily.

Although DTA SR did not provide strong evidence of the accuracy of tests diagnosing pre-diabetes in children, it emphasized the need for objective, empirical scientific evidence and reduced heterogeneity to investigate this issue in paediatric, endocrinological and dialectological societies. This dissertation thesis presents the difficulties that will continue to appear in the synthesis of scientific knowledge in this area, unless consistent protocols are developed for future research and publication following existing guidelines. Only on the basis of well-performed primary studies with provided raw data, it will be possible to perform a synthesis of evidence that can be implied in practice. This may be a driving force for change in children and adolescents with metabolic diseases, which may have an impact on public health both in terms of the presence and future of the current paediatric population.

REFERENCE LIST

ADA. (2010). Diagnosis and classification of diabetes mellitus. *Diabetes Care, 33*(Supplement 1), S62-S69.

ADA. (2018, 2021). Statistics About Diabetes. Retrieved from https://www.diabetes.org/resources/statistics/statistics-about-diabetes

Aggarwal, R., & Ranganathan, P. (2019). Study designs: Part 2–Descriptive studies. *Perspectives in clinical research, 10*(1), 34. Retrieved from http://www.picronline.org/article.asp?issn=2229-3485;year=2019;volume=10;issue=1;spage=34;epage=36;aulast=Aggarwal

Akobeng, A. K. (2007). Understanding diagnostic tests 3: receiver operating characteristic curves. *Acta paediatrica, 96*(5), 644-647. Retrieved from https://onlinelibrary.wiley.com/doi/pdfdirect/10.1111/j.1651-2227.2006.00178.x?download=true

Altman, D. G., & Bland, J. M. (1994). Diagnostic tests. 1: Sensitivity and specificity. *BMJ: British Medical Journal, 308*(6943), 1552.

Aromataris, E., & Pearson, A. (2014). The Systematic Review: An Overview. *AJN The American Journal of Nursing, 114*(3), 53-58. doi:10.1097/01.NAJ.0000444496.24228.2c

Arslanian, S. (2002). Type 2 diabetes in children: clinical aspects and risk factors. *Hormone Research in Paediatrics, 57*(Suppl. 1), 19-28.

Atabek, M. E., & Pirgon, O. (2007). Assessment of insulin sensitivity from measurements in fasting state and during an oral glucose tolerance test in obese children. *J Pediatr Endocrinol Metab, 20*(2), 187-195. Retrieved from https://www.degruyter.com/view/j/jpem.2007.20.2/jpem.2007.20.2.187/jpem.2007.20.2.187.xml

Boon MH, T. H., Shaw B, Akl EA, Lhachimi SK, López-Alcalde J,, Klugar M, C. L., Saz-Parkinson Z, Mustafa RA, Langendam MW, Crane O, Morgan RL, Rehfuess, & E, J. B., Chong LY, Guyatt GH, Schünemann HJ, Katikireddi SV. (2021). Challenges in applying the GRADE approach in public health guidelines and systematic reviews:

A concept paper from the GRADE Public Health Group. *Journal of Clinical Epidemiology, 135*(1). doi:10.1016/j.jclinepi.2021.01.001

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., Lijmer, J. G. (2003). The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Annals of internal medicine, 138*(1), W1-12.

Bramer, W. M., Giustini, D., de Jonge, G. B., Holland, L., & Bekhuis, T. (2016). De-duplication of database search results for systematic reviews in EndNote. *Journal of the Medical Library Association: JMLA, 104*(3), 240.

Brar, P. C. M. D., Mengwall, L. B. A., Franklin, B. H., & Fierman, A. H. M. D. (2014). Screening Obese Children and Adolescents for Prediabetes and/or Type 2 Diabetes in Pediatric Practices: A Validation Study. *Clinical Pediatrics, 53*(8), 771. Retrieved from https://search.proquest.com/docview/1535385887?accountid=45351

Bridges, K. G., Jarrett, T., Thorpe, A., Baus, A., & Cochran, J. (2016). Use of the triglyceride to HDL cholesterol ratio for assessing insulin sensitivity in overweight and obese children in rural Appalachia. *J Pediatr Endocrinol Metab, 29*(2), 153-156. doi:10.1515/jpem-2015-0158

Brożek, J., Akl, E. A., Jaeschke, R., Lang, D. M., Bossuyt, P., Glasziou, P., Meerpohl, J. (2009). Grading quality of evidence and strength of recommendations in clinical practice guidelines: Part 2 of 3. The GRADE approach to grading quality of evidence about diagnostic tests and strategies. *Allergy, 64*(8), 1109-1116.

Campbell, Klugar, M., Ding, S., Carmody, D., Hakonsen, S., & Jadotte, Y. (2015). The systematic review of studies of diagnostic test accuracy. *Joanna Briggs Institute reviewers' manual*, 1-46.

Campbell, Klugar, M., Ding, S., Carmody, D., Hakonsen, S., Jadotte, Y., Munn, Z. (2015). Diagnostic test accuracy: methods for systematic review and meta-analysis. *International Journal of Evidence-Based Healthcare, 13*(3), 154-162.

Cefalu, W. T., Berg, E. G., Saraco, M., Petersen, M. P., Uelmen, S., & Robinson, S. (2019). Classification and diagnosis of diabetes: standards of medical care in diabetes-2019. *Diabetes Care, 42*, S13-S28.

Cochrane, C. (2020). Review Manager (RevMan)[Computer program]. Version 5.4, .

Cooper, H., Hedges, L. V., & Valentine, J. C. (2019). *The handbook of research synthesis and meta-analysis*: Russell Sage Foundation.

Čapková, N., & Lustigová, M. (2017). Vybrané ukazatele zdravotního stavu české populace—studie EHES 2014. *Hygiena, 62*(1), 35-37.

D'Adamo, E., & Caprio, S. (2011). Type 2 Diabetes in Youth: Epidemiology and Pathophysiology. *Diabetes Care, 34*(Supplement 2), S161-S165. doi:10.2337/dc11-s212

Dabelea, D., Bell, R. A., D'Agostino Jr, R. B., Imperatore, G., Johansen, J. M., Linder, B., Mayer-Davis, E. J. (2007). Incidence of diabetes in youth in the United States. *Jama, 297*(24), 2716-2724.

Danaei, G., Finucane, M. M., Y, L., Singh, G. M., Cowan, G. M., Paciorek, M. J., & Stevens, C. J. (2011). National, regional and global trends in fasting plasma glucose and diabetes prevalence since 1980: systematic analysis of health examination surveys and epidemilogical studies with 370 country-years and 2.7 participants. *The Lancet, 378*(9785), 31-40. doi:10.1016/S0140-6736-11-60679-X

Egger, M., Smith, G. D., & Altman, D. G. (2007). *Systematic Reviews in Health Care: Meta-analysis in context* (2nd ed.). London: BMJ Publishing Group.

Ehehalt, S., Wiegand, S., Körner, A., Schweizer, R., Liesenkötter, K. P., Partsch, C. J., Reinehr, T. (2017). Diabetes screening in overweight and obese children and adolescents: choosing the right test. *European Journal of Pediatrics, 176*(1), 89-97. doi:10.1007/s00431-016-2807-6

Emmanuel, M., & Bokor, B. R. (2017). Tanner stages.

Esteghamati, A., Ashraf, H., Khalilzadeh, O., Zandieh, A., Nakhjavani, M., Rashidi, A., . . . Asgari, F. (2010). Optimal cut-off of homeostasis model assessment of insulin resistance (HOMA-IR) for the diagnosis of metabolic syndrome: third national surveillance of risk factors of non-communicable diseases in Iran (SuRFNCD-2007). *Nutrition & metabolism, 7*(1), 1-8.

Evans, D., & Pearson, A. (2001). Systematic reviews: gatekeepers of nursing knowledge. *Journal of Clinical Nursing, 10*(5), 593-599. doi:10.1046/j.1365-2702.2001.00517.x

Fegan, D. F. (2000). Evaluation of diagnostic tests: the epidemiological approach. In *DNA-based molecular diagnostic techniques* (pp. 30-37): FAO, Rome, Italy.

Galhardo, J., & Shield, J. (2015). The Role of Haemoglobin A1c in Screening Obese Children and Adolescents for Glucose Intolerance and Type 2 Diabetes. *Acta Med Port, 28*(3), 307-315. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/26421782

Garcia, A. G., Urbina Trevino, M. V., Villalpando Sanchez, D. C., & Aguilar, C. A. (2019). Diagnostic accuracy of triglyceride/glucose and triglyceride/HDL index as predictors for insulin resistance in children with and without obesity. *Diabetes Metab Syndr, 13*(4), 2329-2334. doi:10.1016/j.dsx.2019.05.020

Gayoso-Diz, P., Otero-González, A., Rodriguez-Alvarez, M. X., Gude, F., García, F., De Francisco, A., & Quintela, A. G. (2013). Insulin resistance (HOMA-IR) cut-off values and the metabolic syndrome in a general adult population: effect of gender and age: EPIRCE cross-sectional study. *BMC endocrine disorders, 13*(1), 1-10.

Glaszious, P., Irwing, L., Bain, C., & Colditz, G. (2001). *Systematic reviews in health care: A practical guide*. Cambridge: Press Syndicate of the University of Cambridge.

Godin, K., Stapleton, J., Kirkpatrick, S. I., Hanning, R. M., & Leatherdale, S. T. (2015). Applying systematic review search methods to the grey literature: a case study examining guidelines for school-based breakfast programs in Canada. *Systematic Reviews, 4*(1), 138.

Gopalakrishna, G., Mustafa, R. A., Davenport, C., Scholten, R. J., Hyde, C., Brozek, J., . . . Langendam, M. W. (2014). Applying Grading of Recommendations Assessment, Development and Evaluation (GRADE) to diagnostic tests was challenging but doable. *Journal of Clinical Epidemiology, 67*(7), 760-768.

Goran, M. I. (2002). Impaired glucose tolerance in obese children and adolescents. *The New England journal of medicine, 347*(4), 290-292; author reply 290.

Gough, D., Oliver, S., & Thomas, J. (2017). *An introduction to systematic reviews*: Sage.

Grundy, S. M. (2012). Pre-diabetes, metabolic syndrome, and cardiovascular risk. *Journal of the American College of Cardiology, 59*(7), 635-643.

Gurung, P., & Jialal, I. (2019). Plasma Glucose.

Guyatt, G., Oxman, A. D., Akl, E. A., Kunz, R., Vist, G., Brozek, J., Debeer, H. (2011). GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology, 64*(4), 383-394.

Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., & Schünemann, H. J. (2008). GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ, 336*(7650), 924-926. doi:10.1136/bmj.39489.470347.AD

Haines, L., Wan, K. C., Lynn, R., Barrett, T. G., & Shield, J. P. H. (2007). Rising Incidence of Type 2 Diabetes in Children in the U.K. *Diabetes Care, 30*(5), 1097-1101. doi:10.2337/dc06-1813

Harris, A. D., McGregor, J. C., Perencevich, E. N., Furuno, J. P., Zhu, J., Peterson, D. E., & Finkelstein, J. (2006). The use and interpretation of quasi-experimental studies in medical informatics. *Journal of the American Medical Informatics Association, 13*(1), 16-23. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1380192/pdf/16.pdf

Haynes, R. B. (2006). Of studies, syntheses, synopses, summaries, and systems: the "5S" evolution of information services for evidence-based healthcare decisions. *BMJ Evidence-Based Medicine, 11*(6), 162-164.

Higgins, J. P. T., & Green, S. (2008). *Cochrane Handbook of Systematic Reviews of Interventions*. Chichester: John Wiley & Sons Ltd.

Hopewell, S., McDonald, S., Clarke, M. J., & Egger, M. (2007). Grey literature in meta-analyses of randomized trials of health care interventions. *Cochrane Database of Systematic Reviews*(2).

Chan, C. L., Pyle, L., Kelsey, M., Newnes, L., Zeitler, P. S., & Nadeau, K. J. (2016). Screening for type 2 diabetes and prediabetes in obese youth: evaluating alternate markers of glycemia-1,5-anhydroglucitol, fructosamine, and glycated albumin. *Pediatric Diabetes, 17*(3), 206-211. doi:10.1111/pedi.12258

Chan, C. L., Pyle, L., Newnes, L., Nadeau, K. J., Zeitler, P. S., & Kelsey, M. M. (2015). Continuous glucose monitoring and its relationship to hemoglobin A1c and oral glucose tolerance testing in obese and prediabetic youth. *J Clin Endocrinol Metab, 100*(3), 902-910. doi:10.1210/jc.2014-3612

Chappell, F., Raab, G., & Wardlaw, J. (2009). When are summary ROC curves appropriate for diagnostic meta-analyses? *Statistics in medicine, 28*(21), 2653-2668.

IDF. (2006). *The IDF consensus worldwide definition of the metabolic syndrome*. Retrieved from Brussels: https://www.idf.org/e-library/consensus-statements/60-idfconsensus-worldwide-definitionof-the-metabolic-syndrome.html

JBI. (2019, 2020). JBI reviewer's manual 2019. *URL: https://wiki. joannabriggs. org/display/MANUAL/JBI+ Reviewer% 27s+ Manual (accessed 12 August 2019).*

Jordan, Z., Lockwood, C., Munn, Z., & Aromataris, E. (2019). The updated Joanna Briggs Institute model of evidence-based healthcare. *JBI Evidence Implementation, 17*(1), 58-71.

Kang, B., Yang, Y., Lee, E. Y., Yang, H. K., Kim, H. s., Lim, S. y., . . . Yoon, K. h. (2017). Triglycerides/glucose index is a useful surrogate marker of insulin resistance among adolescents. *International Journal of Obesity, 41*(5), 789-792. doi:http://dx.doi.org/10.1038/ijo.2017.14

Kasturi, K., Onuzuruike, A. U., Kunnam, S., Shomaker, L. B., Yanovski, J. A., & Chung, S. T. (2019). Two- vs one-hour glucose tolerance testing: Predicting prediabetes in adolescent girls with obesity. *Pediatr Diabetes, 20*(2), 154-159. doi:10.1111/pedi.12803

Keskin, M., Kurtoglu, S., Kendirci, M., Atabek, M. E., & Yazici, C. (2005). Homeostasis Model Assessment Is More Reliable Than the Fasting Glucose/Insulin Ratio and Quantitative Insulin Sensitivity Check Index for Assessing Insulin Resistance Among Obese Children and Adolescents. *Pediatrics, 115*(4), 1066-1063. Retrieved from https://search.proquest.com/docview/228347793?accountid=45351

Khan, K., Kunz R., Kleijnen J., & Antes, G. (2011). *Systematic Reviews to Support Evidence-based medicine*. London: Hodder & Stoughton Ltd.

Kim, S. Y. (2013). Efficacy versus effectiveness. *Korean journal of family medicine, 34*(4), 227-227. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3726789/pdf/kjfm-34-227.pdf

Klugar, M. (2015). *Systematická review ve zdravotnictví* (1st ed.). Olomouc: Palacký University Olomouc.

Klugar, M. (2016). A protocol is essential for a systematic review as randomization is for randomized controlled trials. *JBI Evidence Synthesis, 14*(7), 1-2.

Kontopantelis, E., Doran, T., Springate, D. A., Buchan, I., & Reeves, D. (2015). Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis. *BMJ : British Medical Journal, 350*, h2750. doi:10.1136/bmj.h2750

Korevaar, D. A., Cohen, J. F., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., . . . Altman, D. G. (2016). Updating standards for reporting diagnostic accuracy: the development of STARD 2015. *Research integrity and peer review, 1*(1), 1-10.

Kurtoglu, S., Hatipoglu, N., Mazicioglu, M., Kendirici, M., Keskin, M., & Kondolot, M. (2010). Insulin resistance in obese children and adolescents: HOMA-IR cut-off levels in the prepubertal and pubertal periods. *J Clin Res Pediatr Endocrinol, 2*(3), 100-106. doi:10.4274/jcrpe.v2i3.100

Lam, D. W., & LeRoith, D. (2012). The worldwide diabetes epidemic. *Current Opinion in Endocrinology, Diabetes and Obesity, 19*(2), 93-96.

Langendam, M. W., Akl, E. A., Dahm, P., Glasziou, P., Guyatt, G., & Schünemann, H. J. (2013a). Assessing and presenting summaries of evidence in Cochrane Reviews. *Systematic Reviews, 2*(1), 1-9.

Langendam, M. W., Akl, E. A., Dahm, P., Glasziou, P., Guyatt, G., & Schünemann, H. J. (2013b). *Assessing and presenting summaries of evidence in Cochrane Reviews* (Vol. 2).

Lee, J., Kim, K. W., Choi, S. H., Huh, J., & Park, S. H. (2015). Systematic review and meta-analysis of studies evaluating diagnostic test accuracy: a practical review for clinical researchers-part II. Statistical methods of meta-analysis. *Korean journal of radiology, 16*(6), 1188-1196.

Lee, J., Lee, Y. A., Kim, J. H., Lee, S. Y., Shin, C. H., & Yang, S. W. (2019). Discrepancies between Glycosylated Hemoglobin and Fasting Plasma Glucose for Diagnosing Impaired Fasting Glucose and Diabetes Mellitus in Korean Youth and Young Adults. *Diabetes Metab J, 43*(2), 174-182. doi:10.4093/dmj.2018.0046

Leeflang Mariska MG, Deeks, J. J., Takwoing Yemisi, & Petra, M. (2013). Cochrane diagnostic test accuarcy reviews. *Systematic Reviews, 2*(82). doi:10.1186/2046-4053-2-82

Levin, K. A. (2006). Study design VI-ecological studies. *Evidence-based dentistry, 7*(4), 108-108. Retrieved from https://www.nature.com/articles/6400454.pdf

Levy, J. C., Matthews, D. R., & Hermans, M. P. (1998). Correct homeostasis model assessment (HOMA) evaluation uses the computer program. *Diabetes Care, 21*(12), 2191-2192.

Liang, J., Fu, J., Jiang, Y., Dong, G., Wang, X., & Wu, W. (2015). TriGlycerides and high-density lipoprotein cholesterol ratio compared with homeostasis model assessment insulin resistance indexes in screening for metabolic syndrome in the chinese obese children: a cross section study. *BMC Pediatrics, 15*. Retrieved from https://search.proquest.com/docview/1780070068?accountid=45351

Lockwood, C., & White, S. (2012). *Synthesizing descriptive evidence*: Lippincott Williams & Wilkins.

Lord, S. J., Irwig, L., & Simes, R. J. (2006). When is measuring sensitivity and specificity sufficient to evaluate a diagnostic test, and when do we need randomized trials? *Annals of internal medicine, 144*(11), 850-855.

Lucas, P. J., Baird, J., Arai, L., Law, C., & Roberts, H. M. (2007). Worked examples of alternative methods for the synthesis of qualitative and quantitative research in systematic reviews. *BMC Medical Research Methodology, 7*(1), 1-7.

Macaskill, P., Gatsonis, C., Deeks, J., Harbord, R., & Takwoingi, Y. (2010). Cochrane handbook for systematic reviews of diagnostic test accuracy. *Version 0.9. 0. London: The Cochrane Collaboration*.

Maffeis, C., Pinelli, L., Brambilla, P., Banzato, C., Valzolgher, L., Ulmi, D., Morandi, A. (2010). Fasting Plasma Glucose (FPG) and the Risk of Impaired Glucose Tolerance in Obese Children and Adolescents. *Obesity, 18*(7), 1437-1442. doi:http://dx.doi.org/10.1038/oby.2009.355

Maldonado-Hernández, J., Martínez-Basila, A., Salas-Fernández, A., Navarro-Betancourt, J. R., Piña-Aguero, M. I., & Bernabe-García, M. (2016). The 13C-Glucose Breath Test for Insulin Resistance Assessment in Adolescents: Comparison with Fasting and Post-Glucose Stimulus Surrogate Markers of Insulin Resistance. *Journal of Clinical Research in Pediatric Endocrinology, 8*(4), 419-424. doi:http://dx.doi.org/10.4274/jcrpe.3260

März, W., Kleber, M. E., Scharnagl, H., Speer, T., Zewinger, S., Ritsch, A., Laufs, U. (2017). HDL cholesterol: reappraisal of its clinical relevance. *Clinical Research in Cardiology, 106*(9), 663-675.

Matthews, D. R., Hosker, J., Rudenski, A., Naylor, B., Treacher, D., & Turner, R. (1985). Homeostasis model assessment: insulin resistance and β-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia, 28*(7), 412-419.

Melnyk, B. M., & Fineout-Overholt, E. (2011). *Evidence-based practice in nursing & healthcare: A guide to best practice*: Lippincott Williams & Wilkins.

MG, L. M., Deeks, J. J., Gatsonis, C., & Bossuyt, P. M. (2008). Systematic Reviews of Diagnostic Test Accuracy. *Ann Intern Med., 149*(12), 889-897.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews, 4*(1), 1.

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic review or scoping approach. *BMC Medical Research Methodology, 18*(143). doi:10.1186/s12874-018-0611-x

Munn, Z., Stern, C., Aromataris, E., Lockwood, C., & Jordan, Z. (2018). What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. *BMC Medical Research Methodology, 18*(1), 1-9.

Mutlu, G. Y., Ozsu, E., Cizmecioglu, F. M., & Hatun, S. (2013). Can HbA1c and One-Hour Glucose Concentration in Standard OGTT Be Used for Evaluation of Glucose Homeostasis in Childhood? *Journal of Clinical Research in Pediatric Endocrinology, 5*(2), 80-84. doi:10.4274/Jcrpe.889

Nam, H. K., Cho, W. K., Hyun, J. K., Rhie, Y. J., Chung, S., Lee, K. H., & Suh, B. K. (2017). Optimal cutoff of hemoglobin A1c for diagnosing diabetes and prediabetes in obese children and adolescents. *Endocrine Reviews, 38*(3). Retrieved from http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L617151822

Nam, H. K., Cho, W. K., Kim, J. H., Rhie, Y. J., Chung, S., Lee, K. H., & Suh, B. K. (2018). HbA1c Cutoff for Prediabetes and Diabetes Based on Oral Glucose Tolerance Test in Obese Children and Adolescents. *J Korean Med Sci, 33*(12), e93. doi:10.3346/jkms.2018.33.e93

Nasser, M., & Fedorowicz, Z. (2011). Guest Editorial-Grading the quality of evidence and strength of recommendations: the GRADE approach to improving dental Clinical Guidelines. *Journal of Applied Oral Science, 19*(1), 0-0.

Ng, K., Cosma, A., Svacina, K., Boniel-Nissim, M., & Badura, P. (2021). Czech adolescents' remote school and health experiences during the spring 2020 COVID-19 lockdown. *Preventive Medicine Reports, 22*, 101386.

Nor, N. S. M., Lee, S. J., & Arslanian, S. A. (2015). Triglyceride glucose (TYG) index as a surrogate measure of insulin sensitivity in obese adolescents. *Endocrine Reviews, 36*. Retrieved from http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L613816977

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., . . . Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ, 372*, n71. doi:10.1136/bmj.n71

Pandey, U., Midha, T., Rao, Y. K., Katiyar, P., Wal, P., Kaur, S., & Martolia, D. S. (2017). Anthropometric indicators as predictor of pre-diabetes in Indian adolescents. *Indian Heart J, 69*(4), 474-479. doi:10.1016/j.ihj.2017.06.006

Parikh, R., Mathai, A., Parikh, S., Sekhar, G. C., & Thomas, R. (2008). Understanding and using sensitivity, specificity and predictive values. *Indian journal of ophthalmology, 56*(1), 45.

Pearson, A., Jordan, Z., & Munn, Z. (2012). Translational Science and Evidence-Based Healthcare: A Clarification and Reconceptualization of How Knowledge Is Generated and Used in Healthcare. *Nursing Research and Practice, 2012*. doi:10.1155/2012/792519

Pearson, A., Loveday, H., & Holopainen, A. (2012). *Critically appraising evidence for healthcare*: Lippincott Williams & Wilkins.

Pearson, A., Stannard, D., & Hu, Y. (2012). *Clinical Wisdom and Evidence-Based Healthcare* (A. Pearson Ed.). Australia: Lippincott Williams & Wilkins, a Wolters KLuwer business.

Pearson, A., Weeks, S., & Stern, C. (2011). *Translation Science and The JBI Model of Evidence-Based Healthcare* (A. Pearson Ed.). Australia: Lippincoot Williams & Wilkins, a Wolters Kluwer business.

Pearson, A., Wiechula, R., Court, A., & Lockwood, C. (2005). The JBI model of evidence-based healthcare. *International Journal of Evidence-Based Healthcare, 3*(8), 207-215. Retrieved from https://journals.lww.com/ijebh/Fulltext/2005/09000/The_JBI_model_of_evidence_based_healthcare.1.aspx

https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1479-6988.2005.00026.x

Phillips, B., Ball, C., Sackett, D., Badenoch, D., Straus, S., Haynes, B., & Dawes, M. (1998, 2020). Oxford Centre for Evidence-based Medicine - Levels of Evidence. Retrieved from https://www.cebm.net/2009/06/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., . . . Duffy, S. (2006). Guidance on the conduct of narrative synthesis in systematic reviews. *A product from the ESRC methods programme Version, 1*, b92.

Puri, M., Freeman, K., Garcia, M., Nussbaum, H., & Dimartino-Nardi, J. R. (2007). Criteria for oral glucose tolerance testing of obese minority youth. *J Pediatr Endocrinol Metab, 20*(6), 703-709. Retrieved from https://www.degruyter.com/view/j/jpem.2007.20.6/jpem.2007.20.6.703/jpem.2007.20.6.703.xml

Rezigalla, A. A. (2020). Observational Study Designs: Synopsis for Selecting an Appropriate Study Design. *Cureus, 12*(1).

Robson, L., Shannon, H., Goldenhar, L., & Hale, A. (2001). Before-and-after design: A simple evaluation design. *Guide to Evaluating the Effectiveness of Strategies for Preventing Work Injuries: How to Show Whether a Safety Intervention Really Works. Druid Hills: Centers for Disease Control and Prevention*.

Rosenbloom, A. L., Silverstein, J. H., Amemiya, S., Zeitler, P., & Klingensmith, G. J. (2009). Type 2 diabetes in children and adolescents. *Pediatric Diabetes, 10*(s12), 17-32. doi:10.1111/j.1399-5448.2009.00584.x

Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: what it is and what it isn't. *BMJ, 312*(7023), 71-72. doi:10.1136/bmj.312.7023.71

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Tarantola, S. (2008). *Global sensitivity analysis: the primer*: John Wiley & Sons.

Sharma, S., & Fleming, S. E. (2012). Use of HbA1C testing to diagnose pre-diabetes in high risk African American children: A comparison with fasting glucose and HOMA-IR. *Diabetes and Metabolic Syndrome: Clinical Research and Reviews, 6*(3), 157-162. doi:10.1016/j.dsx.2012.09.004

Sharma, S., Lustig, R. H., & Fleming, S. E. (2011). ORIGINAL RESEARCH Identifying Metabolic Syndrome in African American Children Using Fasting HOMA-IR in Place of Glucose. *Preventing Chronic Disease*, 10_0036.

Shaw, J. (2007). Epidemiology of childhood type 2 diabetes and obesity. *Pediatric Diabetes, 8*(s9), 7-15. doi:10.1111/j.1399-5448.2007.00329.x

Schünemann, H. J. (2019). Chapter 14: Completing 'Summary of findings' tables and grading the certainty of the evidence: Higgins JPT. *Thomas J, Chandler J, et al., Cochrane Handbook for systematic reviews of interventions version, 6*.

Schünemann, H. J., Hill, S. R., Kakad, M., Vist, G. E., Bellamy, R., Stockman, L., . . . Uyeki, T. M. (2007). Transparent development of the WHO rapid advice guidelines. *PLoS medicine, 4*(5), e119.

Schünemann, H. J., Oxman, A. D., Brozek, J., Glasziou, P., Jaeschke, R., Vist, G. E., . . . Montori, V. M. (2008). Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ, 336*(7653), 1106-1110.

Schwartz, A. (2014). Diagnostic test calculator, 2006. *Available at: araw. mede. uic. edu/cgi-bin/testcalc. pl (accessed Feb 21, 2011)*.

Sibbald, B., & Roland, M. (1998). Understanding controlled trials. Why are randomised controlled trials important? *BMJ: British Medical Journal, 316*(7126), 201. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2665449/pdf/9468688.pdf

Simental-Mendía, L. E., Rodríguez-Morán, M., & Guerrero-Romero, F. (2008). The product of fasting glucose and triglycerides as surrogate for identifying insulin resistance in apparently healthy subjects. *Metabolic syndrome and related disorders, 6*(4), 299-304.

Tabak, A. G., Herder, C., Rathmann, W., Brunner, E. J., & Kivimaki, M. (2012). Prediabetes: a high-risk state for diabetes development. *The Lancet, 379*(9833), 2279-2290. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3891203/pdf/nihms535847.pdf

Tirabanchasak, S., Siripunthana, S., Supornsilchai, V., Wacharasindhu, S., & Sahakitrungruang, T. (2015). Insulin dynamics and biochemical markers for predicting impaired glucose tolerance in obese Thai youth. *Journal of Pediatric Endocrinology & Metabolism, 28*(9-10), 1039-1045. doi:http://dx.doi.org/10.1515/jpem-2014-0273

Toth, P. P. (2005). The "good cholesterol" high-density lipoprotein. *Circulation, 111*(5), e89-e91.

Tucková, D., Klugar, M., Lasota, T., Lasák, J., Campbell, J. M., & Munn, Z. (2017). Alternative tests versus measurement of fasting plasma glucose and oral glucose tolerance test for diagnosis of pediatric type 2 pre-diabetes: a systematic review protocol. *JBI Evidence Synthesis, 15*(8), 2028-2032.

Tučková, D., & Klugar, M. (2015). Sekundární výzkum a jeho charakteristika. In Š. Vévodová & K. a. k. Ivanová (Eds.), *Základy metodologie výzkumu pro nelékařské zdravotnické profese*

(pp. 167-192). Olomouc: Univerzita Palackého.

Vickers, A. J. (2008). Decision analysis for the evaluation of diagnostic tests, prediction models, and molecular markers. *The American Statistician, 62*(4), 314-320.

Walden University. (2020). Evidence-Based Practice Research: Levels of Evidence Pyramid. Retrieved from https://academicguides.waldenu.edu/library/healthevidence/evidencepyramid

Wang, J. J., & Attia, J. (2010). Study designs in epidemiology and levels of evidence. *American journal of ophthalmology, 149*(3), 367-370.

White, S., Schultz, T., & Y., E. (2011). *Synthetizing Evidence of Diagnostic Accuracy* (P. A. Pearson Ed.). Philadelphia: Lippincott Williams and Wilkins; a Wolters Kluwer.

Whiting, P. F., Guariguata, L., Weil, C., & Shaw, J. (2011). IDF Atlas: Global estimates of the prevalence of diabetes for 2011 and 2030. *Diabetes Research and Clinical Practice, 94*(3), 311-321. doi:10.1016/j.diabres.2011.10.029

Whiting, P. F., Rutjes, A. W., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., . . . Bossuyt, P. M. (2011). QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of internal medicine, 155*(8), 529-536.

WHO. (1999). *Definition, diagnosis and classification of diabetes mellitus and its complications : report of a WHO consultation. Part 1, Diagnosis and classification of diabetes mellitus.* (WHO/NCD/NCS/99.2). Retrieved from Geneva: https://apps.who.int/iris/handle/10665/66040

WHO. (2009). *Global health risks: mortality and burden of disease attributable to selected major risks*: World Health Organization.

WHO. (2016a). *Diabetes Fact Sheet*. Retrieved from Geneva: https://www.euro.who.int/__data/assets/pdf_file/0010/305389/Diabetes-Fact-Sheet-en.pdf

WHO. (2016b). *Global report on diabetes*. Retrieved from Geneva: https://www.who.int/publications/i/item/9789241565257

Willis, B. H., & Quigley, M. (2011). Uptake of newer methodological developments and the deployment of meta-analysis in diagnostic test research: a systematic review. *BMC Medical Research Methodology*, *11*(1), 1-8.

SUMMARY OF FIGURES

SUMMARY OF TABLES

SUMMARY OF APPENDIXES

Appendix 1

PRISMA-P (Preferred Reporting Items for Systematic review and Meta-Analysis Protocols) 2015 checklist: recommended items to address in a systematic review protocol*

Section and topic	Item No	Checklist item
ADMINISTRATIVE INFORMATION		
Title:		
Identification	1a	Identify the report as a protocol of a systematic review
Update	1b	If the protocol is for an update of a previous systematic review, identify as such
Registration	2	If registered, provide the name of the registry (such as PROSPERO) and registration number
Authors:		
Contact	3a	Provide name, institutional affiliation, e-mail address of all protocol authors; provide physical mailing address of corresponding author
Contributions	3b	Describe contributions of protocol authors and identify the guarantor of the review
Amendments	4	If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments
Support:		
Sources	5a	Indicate sources of financial or other support for the review
Sponsor	5b	Provide name for the review funder and/or sponsor
Role of sponsor or funder	5c	Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol
INTRODUCTION		
Rationale	6	Describe the rationale for the review in the context of what is already known
Objectives	7	Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO)
METHODS		
Eligibility criteria	8	Specify the study characteristics (such as PICO, study design, setting, time frame) and report characteristics (such as years considered, language, publication status) to be used as criteria for eligibility for the review
Information sources	9	Describe all intended information sources (such as electronic databases, contact with study authors, trial registers or other grey literature sources) with planned dates of coverage
Search strategy	10	Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated

Study records:		
Data management	11a	Describe the mechanism(s) that will be used to manage records and data throughout the review
Selection process	11b	State the process that will be used for selecting studies (such as two independent reviewers) through each phase of the review (that is, screening, eligibility and inclusion in meta-analysis)
Data collection process	11c	Describe planned method of extracting data from reports (such as piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators
Data items	12	List and define all variables for which data will be sought (such as PICO items, funding sources), any pre-planned data assumptions and simplifications
Outcomes and prioritization	13	List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale
Risk of bias in individual studies	14	Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis
Data synthesis	15a	Describe criteria under which study data will be quantitatively synthesised
	15b	If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data and methods of combining data from studies, including any planned exploration of consistency (such as I^2, Kendall's τ)
	15c	Describe any proposed additional analyses (such as sensitivity or subgroup analyses, meta-regression)
	15d	If quantitative synthesis is not appropriate, describe the type of summary planned
Meta-bias(es)	16	Specify any planned assessment of meta-bias(es) (such as publication bias across studies, selective reporting within studies)
Confidence in cumulative evidence	17	Describe how the strength of the body of evidence will be assessed (such as GRADE)

*** It is strongly recommended that this checklist be read in conjunction with the PRISMA-P Explanation and Elaboration (cite when available) for important clarification on the items. Amendments to a review protocol should be tracked and dated. The copyright for PRISMA-P (including checklist) is held by the PRISMA-P Group and is distributed under a Creative Commons Attribution Licence 4.0.**

From: Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart L, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015 Jan 2;349(jan02 1):g7647.

Appendix 2

Databases of published literature

(cited from The JBI Reviewers´Manual 2015: The systematic review of studies of diagnostic test accuracy)

Nursing and allied health

- Allied and Complementary Medicine (AMED):
(http://www.ebscohost.com/academic/AMED-The-Allied-and-Complementary-Medicine-Database)

- British Nursing Index (BNI):
(www.bniplus.co.uk/)

- Cumulative Index to Nursing and Allied Health (CINAHL):
(www.cinahl.com/)

Primary Care

- Essential Evidence Plus (formerly Patient Oriented Evidence that Matters (InfoPOEMs)):
(www.essentialevidenceplus.com/)

Social science psychology and psychiatry

- Applied Social Sciences Index and Abstracts (ASSIA):
(http://www.proquest.com/products-services/ASSIA-Applied-Social-Sciences-Index-and-Abstracts.html)

- PsycINFO:
(www.apa.org/psycinfo/)

- Sociological Abstracts:
(http://proquest.libguides.com/SocAbs)

Biology and chemistry

- Biological Abstracts / BIOSIS Previews:
(http://thomsonreuters.com/biosis-previews/)

- Chemical Abstracts:
(www.cas.org/)

- Database of the International Federation of Clinical Chemistry and Laboratory Medicine
Committee for Evidence-based Laboratrory Medicine (IFCC C-EBLM database)

International health

- Global Health
Available via: (www.cabi.org)

In addition to subject-specific databases, general search engines include:

- Google Scholar (free on the Internet)
(www.schilar.googlecom/advanced_scholar_search)

- Turning Research into Practice (TRIP) database (evidence-based healthcare resource) (free on the Internet): (www.tripdatabase.com/)

"Citation searching"

Citation searching is an important and effective adjunct to database searching and hand searching. Information about these citation indexes is available at: Cochrane handbook

- Science Citation Index:
(scientific.thomson.com/products/sci)

- Social Sciences Citation Index:
(scientific.thomson.com/products/ssci)

- Web of Science:
(scientific.thomson.com/products/wos)

- Web of Knowledge:
(isiwekofknowledge.com/)

- Scopus:
(http://www.elsevier.com/online-tools/scopus)

Theses specific database

- ProQuest Dissertations & Theses Database:
(www.proquest.co.uk/products_pq/descriptions/pqdt.shtml)

- Dissertation Abstracts Online (DIALOG):

- Index to Theses in Great Britain and Ireland
(www.thesis.com)

- DissOnline: indexes 50,000 German dissertations:
(www.dissonline.de/)

Grey literature database

- MedNar:

(www.mednar.com/mednar)

- OpenSIGLE: (http://www.greynet.org/opensiglerepository.html)
- National Technical Information Service (NTIS): (www.ntis.gov/)
- WorldWideScience.org: (worldwidescience.org/index)
- Open Grey: (http://www.opengrey.eu/)

Appendix 3

JBI CRITICAL APPRAISAL CHECKLIST FOR DIAGNOSTIC TEST ACCURACY STUDIES

Reviewer_____Date_____

Author_____Year_____ Record Number_____

	Yes	No	Unclear	Not applicable
1. Was a consecutive or random sample of patients enrolled?	☐	☐	☐	☐
2. Was a case control design avoided?	☐	☐	☐	☐
3. Did the study avoid inappropriate exclusions?	☐	☐	☐	☐
4. Were the index test results interpreted without knowledge of the results of the reference standard?	☐	☐	☐	☐
5. If a threshold was used, was it pre-specified?	☐	☐	☐	☐
6. Is the reference standard likely to correctly classify the target condition?	☐	☐	☐	☐
7. Were the reference standard results interpreted without knowledge of the results of the index test?	☐	☐	☐	☐
8. Was there an appropriate interval between index test and reference standard?	☐	☐	☐	☐
9. Did all patients receive the same reference standard?	☐	☐	☐	☐
10. Were all patients included in the analysis?	☐	☐	☐	☐

Overall appraisal: Include ☐ Exclude ☐ Seek further info ☐

Comments (Including reason for exclusion)

_____

_____

Appendix 4

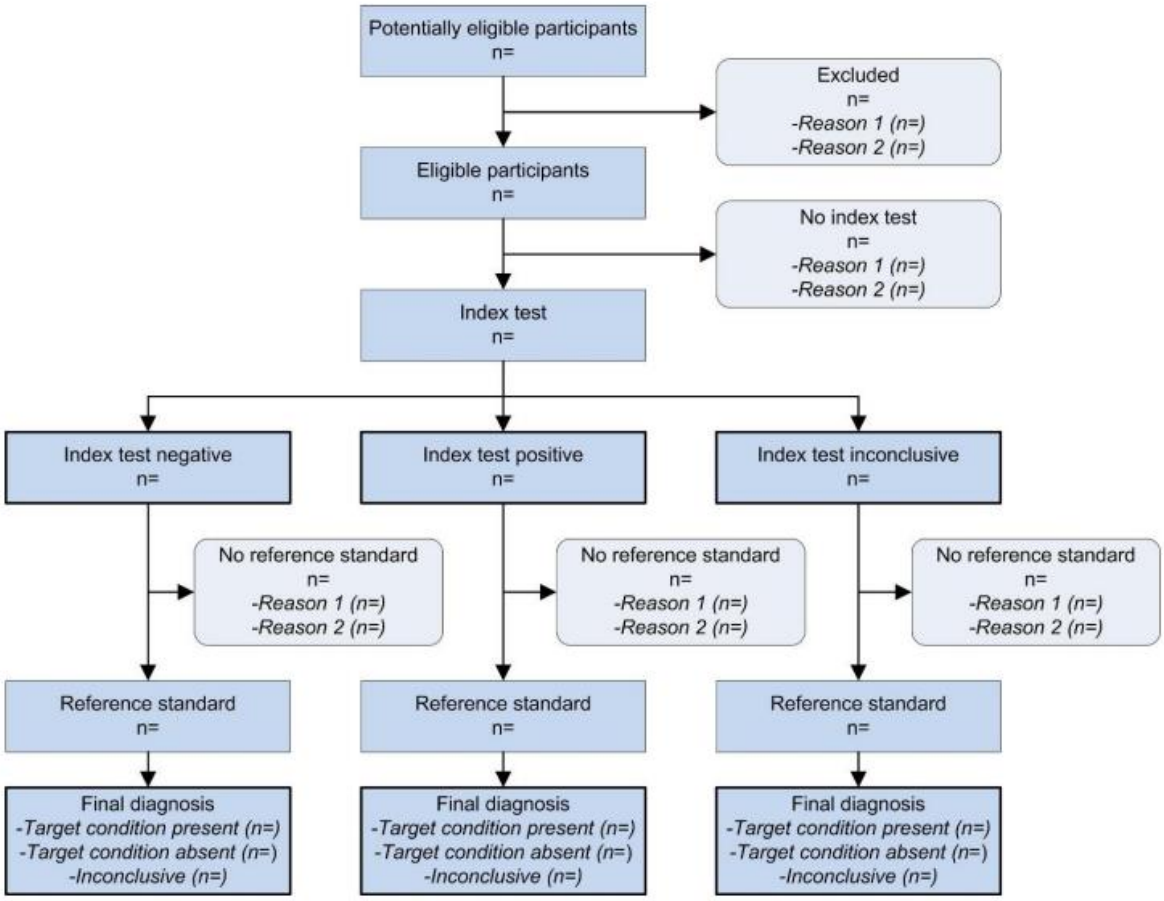
Author/Date	
Inclusion/exclusion criteria: i.e. presenting symptoms, results from previous tests	Inclusion: Exclusion:
Sample size	
Participant demographics (i.e. age, sex, spectrum of presenting symptoms, comorbidity, current treatments, recruitment centres)	
Study methodology (consecutive or random; retrospective or prospective)	
Period that study was carried out (beginning and end date)	
Index test description (including criteria for positive test)	
Reference test description (including criteria for positive test)	
Geographical location of data collection	
Setting of data collection	
Persons executing and interpreting index tests (numbers, training, and expertise)	
Persons executing and interpreting reference test	
Index/reference time interval (and treatments carried out in between)	
Distribution of severity of disease in those with target condition	
Other diagnoses in those without target condition	
Adverse events from index test	
Adverse events from reference test	

Index test results Threshold=	**Condition positive**	**Condition negative**	**Total**
Index test positive (T+)			
Index test negative (T-)			
Total			

Appendix 5

Section & Topic	No	Item
TITLE OR ABSTRACT		
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)
ABSTRACT		
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)
INTRODUCTION		
	3	Scientific and clinical background, including the intended use and clinical role of the index test
	4	Study objectives and hypotheses
METHODS		
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)
Participants	**6**	Eligibility criteria
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)
	8	Where and when potentially eligible participants were identified (setting, location and dates)
	9	Whether participants formed a consecutive, random or convenience series
Test methods	**10a**	Index test, in sufficient detail to allow replication
	10b	Reference standard, in sufficient detail to allow replication
	11	Rationale for choosing the reference standard (if alternatives exist)
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test
	13b	Whether clinical information and index test results were available to the assessors of the reference standard
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy
	15	How indeterminate index test or reference standard results were handled
	16	How missing data on the index test and reference standard were handled
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory
	18	Intended sample size and how it was determined
RESULTS		
Participants	**19**	Flow of participants, using a diagram
	20	Baseline demographic and clinical characteristics of participants
	21a	Distribution of severity of disease in those with the target condition
	21b	Distribution of alternative diagnoses in those without the target condition
	22	Time interval and any clinical interventions between index test and reference standard
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)
	25	Any adverse events from performing the index test or the reference standard
DISCUSSION		
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability
	27	Implications for practice, including the intended use and clinical role of the index test
OTHER INFORMATION		
	28	Registration number and name of registry
	29	Where the full study protocol can be accessed
	30	Sources of funding and other support; role of funders

Prototypical STARD diagram to report flow of participants through the study



Appendix 6

Studies that did not meet population criteria:

1. Abdul-Ghani, M. A., Abdul-Ghani, T., Müller, G., Bergmann, A., Fischer, S., Bornstein, S., DeFronzo, R. A., Schwarz, P. *Role of glycated hemoglobin in the prediction of future risk of T2DM.* (2011) – population age did not meet the criteria for inclusion

2. Alhalbouni, H., Kabalan, Y., Alquobaili, F. *Relation of plasma obestatin levels with BMI and HOMA-IR in syrian obese patients with type 2 diabetes.* (2017) – population age did not meet the criteria for inclusion

3. Alqahtani, N., Khan, W. A. G., Alhumaidi, M. H., Ahmed, Y. A. A. R. *Use of glycated hemoglobin in the diagnosis of diabetes mellitus and pre-diabetes and role of fasting plasma glucose, oral glucose tolerance test.* (2013) – population age did not meet the criteria for inclusion

4. Assyov, Y., Gateva, A., Tsakova, A., Kamenov, Z. *Irisin in the Glucose Continuum.* (2016) – population age did not meet the criteria for inclusion

5. Khokhar, A., Naraparaju, G., Friedman, M., Perez-Colon, S., Umpaichitra, V., Chin, Vivian L. *Comparison of A1C to Oral Glucose Tolerance Test for the Diagnosis of Prediabetes in Overweight and Obese Youth.* (2017) - population age did not meet the criteria for inclusion

6. Kumbhojkar, A., Saraff, V., Nightingale, P., Hogler, W. *Glycated haemoglobin as a screening test for abnormal glucose homeostasis in childhood obesity*. (2020) - population age did not meet the criteria for inclusion

7. Lee, J. M., Gebremariam, A., Wu, E. L., Larose, J., Gurney, J. G. *Evaluation of nonfasting tests to screen for childhood and adolescent dysglycemia.* (2011) - population with predisposition to the diagnosis of interest

8. Lee, H. S., Park, H. K., Hwang, J. S. *HbA1c and glucose intolerance in obese children and adolescents*. (2012) - population with predisposition to the diagnosis of interest

9. Mo, Y., Ma, X., Li, H., Ran, X., Yang, W., Li, Q., Peng, Y., Li, Y., Gao, X., Luan, X., Wang, W., Xie, Y., Zhou, J., Jia, W. *Relationship between glycated albumin and glycated hemoglobin according to glucose tolerance status: A multicenter study*. (2016) - population age did not meet the criteria for inclusion

10. Moadab, M. H., Kelishadi, R., Hashemipour, M., Amini, M., Poursafa, P. *The prevalence of impaired fasting glucose and type 2 diabetes in a population-based sample of overweight/obese children in the Middle East.* (2010) - population age did not meet the criteria for inclusion

11. Morandi, A., Maschio, M., Marigliano, M., Miraglia Del Giudice, E., Moro, B., Peverelli, P., Maffeis, C. *Screening for impaired glucose tolerance in obese children and adolescents: A validation and implementation study*. (2014) - population age did not meet the criteria for inclusion

12. Nowicka, P., Santoro, N., Liu, H. B., Lartaud, D., Shaw, M. M., Goldberg, R., Guandalini, C., Savoye, M., Rose, P., Caprio, S. *Utility of Hemoglobin A(1c) for Diagnosing Prediabetes and Diabetes in Obese Children and Adolescents*. (2011) - population age did not meet the criteria for inclusion

13. Okosun, I. S., Seale, J. P., Lyn, R., Davis-Smith, Y. M. *Improving Detection of Prediabetes in Children and Adults: Using Combinations of Blood Glucose Tests*. (2015) - population age did not meet the criteria for inclusion

14. Olson, B. P., Matter, N. I., Ediger, M. N., Hull, E. L., Maynard, J. D. *Noninvasive skin fluorescence spectroscopy is comparable to hemoglobin A1c and fasting plasma glucose for detection of abnormal glucose tolerance*. (2013) - population age did not meet the criteria for inclusion

15. Park, S. H., Yoon, J. S., Won, K. C., Lee, H. W. *Usefulness of Glycated Hemoglobin as Diagnostic Criteria for Metabolic Syndrome*. (2012) - population age did not meet the criteria for inclusion

16. Serdar, M. A., Serteser, M., Ucal, Y., Karpuzoglu, H. F., Aksungar, F. B., Coskun, A., Kilercik, M., Unsal, I., Ozpinar, A. *An Assessment of HbA1c in Diabetes Mellitus and Pre-diabetes Diagnosis: a Multi-centered Data Mining Study*. (2020) - population age did not meet the criteria for inclusion

17. Shalitin, S., Abrahami, M., Lilos, P., Phillip, M. *Insulin resistance and impaired glucose tolerance in obese children and adolescents referred to a tertiary-care center in Israel*. (2005) - population age did not meet the criteria for inclusion

18. Simental-Mendia, L. E., Gamboa-Gomez, C. I., Aradillas-Garcia, C., Rodriguez-Moran, M., Guerrero-Romero, F. *The triglyceride and glucose index is a useful biomarker to recognize glucose disorders in apparently healthy children and adolescents*. (2020) - population age did not meet the criteria for inclusion

19. Vijayakumar, P., Nelson, R. G., Hanson, R. L., Knowler, W. C., Sinha, M. *HbA1c and the Prediction of Type 2 Diabetes in Children and Adults*. (2017) - population age did not meet the criteria for inclusion

Studies that did not meet diagnosis of interest criteria

1. Acosta-Garcia, E., Concepcion-Paez, M. *Cardiometabolic index as a predictor of cardiovascular risk factors in adolescents*. (2018) - diagnosis of interest did not meet the criteria for inclusion (cardiovascular risk factors – dyslipidemia, hypertension, IFG)

2. Kruger, H. Salome, Faber, Mieke, Schutte, Aletta E., Ellis, Suria M. *A proposed cutoff point of waist-to-height ratio for metabolic risk in African township adolescents.* (2013) – diagnosis of interest did not meet the criteria for inclusion (MS)

3. Lantigua, H., Rubio, N., Yafi, M. *25th European Congress on Obesity, Vienna, Austria, May 23-26, 2018: Abstracts.* (2018) – diagnosis of interest did not meet the criteria for inclusion (T2DM)

4. Li, Y., Zou, Z., Luo, J., Ma, J., Ma, Y., Jing, J., Zhang, X., Luo, C., Wang, H., Zhao, H., Pan, D., Jia, P. *The predictive value of anthropometric indices for cardiometabolic risk factors in Chinese children and adolescents: A national multicenter school-based study.* (2020) - diagnosis of interest did not meet the criteria for inclusion (T2DM)

5. Liang, J., Fu, J., Jiang, Y., Dong, G., Wang, X., Wu, W. *TriGlycerides and high-density lipoprotein cholesterol ratio compared with homeostasis model assessment insulin resistance indexes in screening for metabolic syndrome in the chinese obese children: a cross section study.* (2015) - diagnosis of interest did not meet the criteria for inclusion (MS)

6. Ozer, S., Yilmaz, R., Ozlem Kazanci, N., Sonmezgoz, E., Karaaslan, E., Altuntas, B., Emre Kuyucu, Y. *Higher hdl levels are a preventive factor for metabolic syndrome in obese Turkish children*. (2014) - diagnosis of interest did not meet the criteria for inclusion (MS)

7. Rivera-Hernandez, A., Zurita-Cruz, J. N., Garrido-Magana, E., Fiorentini-Fayad, G. M., Nishimura-Meguro, E. *Glycosylated hemoglobin A1c as a diagnostic test for diabetes mellitus in adolescents with overweight and obesity*. (2015) - diagnosis of interest did not meet the criteria for inclusion (T2DM)

8. Shah, S., Kublaoui, B. M., Oden, J. D., White, P. C. *Screening for Type 2 Diabetes in Obese Young.* (2009) - diagnosis of interest did not meet the criteria for inclusion (T2DM)

9. Tapia Ceballos, L. et al. *Prevalence of metabolic syndrome and its components in obese children and adolescents.* (2007) - diagnosis of interest did not meet the criteria for inclusion (MS)

10. Toledo-Corral, C. M., Vargas, L. G., Goran, M. I., Weigensberg, M. J. *Hemoglobin A1c above Threshold Level is Associated with Decreased beta-Cell Function in Overweight Latino Youth.* (2012) - diagnosis of interest did not meet the criteria for inclusion (beta-cell function)

11. Vitariusova, E., Kostalova, L., Pribilincova, Z., Hlavata, A., Kovacs, L. *Problems of metabolic syndrome diagnostics in children.* (2010) - diagnosis of interest did not meet the criteria for inclusion (MS)

12. Yang, C. Y., Li, H. Y., Sung, F. C., Tan, E. C., Wei, J. N., Chuang, L. M. *Relationship between fasting plasma glucose and incidence of diabetes in children and adolescents*. (2019) - diagnosis of interest did not meet the criteria for inclusion (T2DM)

13. Yoon, J. S., So, C. H., Lee, H. S., Hwang, J. S. *Glycated hemoglobin A1c as a screening test for detecting type 2 diabetes mellitus in obese children and adolescents*. (2018) - diagnosis of interest did not meet the criteria for inclusion (T2DM)

14. Zhu, W. F., Liang, L., Wang, C. L. *Triglyceride and non-HDL-C are better predictors of cardiovascular disease risk factors in Chinese Han children and adolescents than LDL-C.* (2013) - diagnosis of interest did not meet the criteria for inclusion (cardiovascular disease risk factors)

Studies that did not meet study design criteria:

1. Abrams, P., Levitt Katz, L. E., Moore, R. H., Xanthopoulos, M. S., Bishop-Gilyard, C. T., Wadden, T. A., Berkowitz, R. I. *Threshold for improvement in insulin sensitivity with adolescent weight loss.* (2013) – not a DTA study design

2. Al Amiri, E., Abdullatif, M., Abdulle, A., Al Bitar, N., Afandi, E. Z., Parish, M., Darwiche, G. *The prevalence, risk factors, and screening measure for prediabetes and diabetes among Emirati overweight/obese children and adolescents.* (2015) – not a DTA study design

3. Aldhoon-Hainerova, I., Zamrazilova, H., Dusatkova, L., Sedlackova, B., Hlavaty, P., Hill, M., Hampl, R., Kunesova, M., Hainer, V. *Glucose homeostasis and insulin resistance: prevalence, gender differences and predictors in adolescents.* (2014) – not a DTA study design

4. Alikasifoglu, A., Gonc, N., Ozon, Z. A., Sen, Y., Kandemir, N. *The relationship between serum adiponectin, tumor necrosis factor-alpha, leptin levels and insulin sensitivity in childhood and adolescent obesity: adiponectin is a marker of metabolic syndrome.* (2009) – not a DTA study design

5. Babaoglu, K., Hatun, S., Arslanoglu, I., Isguven, P., Bas, F., Ercan, O., Darendeliler, F., Bundak, R., Saka, N., Gunoz, H., Bereket, A., Memioglu, N., Neyzi, O. *Evaluation of*

glucose intolerance in adolescents relative to adults with type 2 diabetes mellitus. (2006) - not a DTA study design

6. Bahíllo-Curieses, M. P., Hermoso-López, F., Martínez-Sopena, M. J., Cobreros-García, P., García-Saseta, P., Tríguez-García, M., Marugán-Miguelsanz, J. M. *Prevalence of insulin resistance and impaired glucose tolerance in a sample of obese Spanish children and adolescents.* (2012) - not a DTA study design

7. Bergman, M., Manco, M., Sesti, G., Dankner, R., Pareek, M., Jagannathan, R., Chetrit, A., Abdul-Ghani, M.,, Buysschaert, M., Olsen, M. H., Nilsson, P. M., Medina, J. L., Roth, J., Groop, L., Del Prato, S., Raz, I., Ceriello, A. *Petition to replace current OGTT criteria for diagnosing prediabetes with the 1-hour post-load plasma glucose>/=155mg/dl (8.6mmol/L).* (2018) - not a DTA study design

8. Berhan, Y. T., Möllsten, A., Carlsson, A., Högberg, L., Ivarsson, A., Dahlquist, G. *Five-region study finds no evidence of undiagnosed type 2 diabetes in Swedish 11- to 13-year-olds.* (2014) - not a DTA study design

9. Bobbert, T., Schwarz, F., Fischer-Rosinsky, A., Maurer, L., Möhlig, M., Pfeiffer, A. F. H., Mai, K., Spranger, J. *Chemerin and prediction of Diabetes mellitus type 2.* (2015) – not a DTA study design

10. Brar, P. C. *Update on the current modalities used to screen high risk youth for prediabetes and/or type 2 diabetes mellitus.* (2019) – not a diagnostic study design

11. Brown, R. J., Yanovski, J. A. *Estimation of insulin sensitivity in children: Methods, measures and controversies.* (2014) – not a diagnostic study design

12. Chan, C. L., Drews, K. L., Buse, J. B., Zeitler, P. S., Kelsey, M. M. *10th Individual Abstracts for International Meeting of Pediatric Endocrinology: Free Communication and Poster Sessions, Abstracts.* (2017) – not a diagnostic study design

13. Cockcroft, E. J., Williams, C. A., Jackman, S. R., Armstrong, N., Barker, A. R. *Agreement and Reliability of Fasted and Oral Glucose Tolerance Test-Derived Indices of Insulin Sensitivity and Beta Cell Function in Boys.* (2017) – not a diagnostic study design

14. d'Annunzio, G., Vanelli, M., Pistorio, A., Minuto, N., Bergamino, L., Iafusco, D., Lorini, R. *Insulin resistance and secretion indexes in healthy Italian children and adolescents: a multicentre study.* (2009) – not a diagnostic study design

15. da Silva, R. C. Q., Lopes, M. W., Dib, A. R., Atala, S. *Insulin resistance, [beta]-cell function, and glucose tolerance in Brazilian adolescents with obesity or risk factors for type 2 diabetes mellitus.* (2007) – not a diagnostic study design

16. Dilli, D., Bostanci, I., Dallar, Y., Gucuk, S. *Glycohemoglobin screening in adolescents attending to the Department of Paediatrics at a tertiary hospital in Turkey.* (2008) – not a diagnostic study design

17. Donoso, M. A., Muñoz-Calvo, M. T., Barrios, V., Martínez, G., Hawkins, F., Argente, J. *Increased leptin/adiponectin ratio and free leptin index are markers of insulin resistance in obese girls during pubertal development.* (2013) – not a diagnostic study design

18. Dubinina, I. A., Chistiakov, D. A., Eremina, I. A., Brovkin, A. N., Zilberman, L. I., Nikitin, A. G., Kuraeva, T. L., Nosikov, V. V., Peterkova, V. A., Dedov, I. I. *Studying progression from glucose intolerance to type 2 diabetes in obese children.* (2014) – not a diagnostic study design

19. El Awwa, A., Soliman, A., Al-Ali, M., Yassin, M., De Sanctis, V. *Continuous glucose monitoring, oral glucose tolerance, and insulin - glucose parameters in adolescents with simple obesity*. (2012) – not a diagnostic study design

20. Elst, M. A. J., van der Aa, M. P., van Mil, E. G. A. H., van der Vorst, M. M. J. *Screening for type 2 diabetes mellitus: The holy grail?* (2015) – not a diagnostic study design

21. Eyzaguirre, F., Mericq, V. *Insulin Resistance Markers in Children.* (2009) – not a diagnostic study design

22. Groot, C. J., Grond, J. V., Delgado, Y., Rings, E. H., Hannema, S. E., van den Akker, E. L. *High predictability of impaired glucose tolerance by combining cardiometabolic screening parameters in obese children.* (2017) – not a diagnostic study design

23. Gunczler, P., Lanes, R. *Relationship between different fasting-based insulin sensitivity indices in obese children and adolescents.* (2006) – not a diagnostic study design

24. Henderson, M., Baillargeon, J. P., Rabasa-Lhoret, R., Chiasson, J. L., Hanley, J., Lambert, M. *Estimating insulin secretion in youth using simple indices derived from the oral glucose tolerance test.* (2012) - not a diagnostic study design

25. Henderson, M., Rabasa-Lhoret, R., Bastard, J. P., Chiasson, J. L., Baillargeon, J. P., Hanley, J. A., Lambert, M. *Measuring insulin sensitivity in youth: How do the different indices compare with the gold-standard method?* (2012) – not a diagnostic study design

26. Kaya, A., Kocyigit, C., Catli, G., Ozkan, E. B., Dundar, B. N. *The Relationship Between Glycemic Variability and Inflammatory Markers in Obese Children with Insulin Resistance and Metabolic Syndrome.* (2017) – not a diagnostic study design

27. Kim, J. Y., Goran, M. I., Toledo-Corral, C. M., Weigensberg, M. J., Shaibi, G. Q. *Comparing glycemic indicators of prediabetes: a prospective study of obese Latino Youth* (2015) – not a diagnostic study design

28. Kim, J. Y., Tfayli, H., Bacha, F., Lee, S., Michaliszyn, S. F., Yousuf, S., Gebara, N., Arslanian, S. *Beta-cell function, incretin response, and insulin sensitivity of glucose and fat metabolism in obese youth: Relationship to OGTT-time-to-glucose-peak.* (2020) – not a diagnostic study design

29. Lee, J. A., Laurson, K. R. *Obesity and Insulin Resistance Screening Tools in American Adolescents: National Health and Nutrition Examination Survey (NHANES) 1999 to 2010.* (2016) – not a DTA study design

30. Lee, J. M., Eason, A., Nelson, C. *Use of HbA1c in the Diagnosis of Diabetes in Adolescents.* (2014) – not a DTA study design

31. Lentferink, Y. E., Elst, M. A. J., Knibbe, C. A. J., van der Vorst, M. M. J. *Predictors of Insulin Resistance in Children versus Adolescents with Obesity.* (2017) - not a diagnostic study design

32. Li, G., Han, L., Wang, Y., Zhao, Y., Li, Y., Fu, J., Li, M., Gao, S., Willi, S. M. *Evaluation of ADA HbA1c criteria in the diagnosis of pre-diabetes and diabetes in a population of Chinese adolescents and young adults at high risk for diabetes.* (2018) – not a diagnostic study design

33. Masuccio, F. G., Lattanzio, F. M., Matera, S., Giannini, C., Chiarelli, F., Mohn, A. *Insulin Sensitivity in Prepubertal Caucasian Normal Weight Children*. (2009) - not a diagnostic study design

34. Mazza, C. S., Ozuna, B., Krochik, A. G., Araujo, M. B. *Prevalence of type 2 diabetes mellitus and impaired glucose tolerance in obese Argentinean children and adolescents*. (2005) - not a diagnostic study design

35. Nogueira-de-Almeida, C. A., de Mello, E. D. *Different criteria for the definition of insulin resistance and its relation with Dyslipidemia in overweight and obese children and adolescents.* (2018) - not a diagnostic study design

36. Nsiah-Kumi, P. A., Lasley, S., Whiting, M., Brushbreaker, C., Erickson, J. M., Qiu, F., Yu, F., Larsen, J. L. *Diabetes, pre-diabetes and insulin resistance screening in Native American children and youth*. (2013) - not a diagnostic study design

37. Ogawa, E., Urakami, T., Suzuki, J., Yoshida, A., Takahashi, S., Mugishima, H. *Usefulness of HbA1c to diagnose diabetes among Japanese children detected by a urine glucose screening program in the Tokyo Metropolitan Area.* (2012) - not a diagnostic study design

38. Önal, Z. E., Atasayan, V., Gürbüz, T., Hepkaya, E., Nuhoglu, Ç. *Association of glycosylated hemoglobin (HbA1c) levels with Iinsulin resistance in obese children*. (2014) - not a diagnostic study design

39. Sahin, N. M., Kinik, S. T., Tekindal, M. A. *OGTT results in obese adolescents with normal HOMA-IR values*. (2013) - not a diagnostic study design

40. Shashaj, B., Luciano, R., Contoli, B., Morino, G. S., Spreghini, M. R., Rustico, C., Sforza, R. W., Dallapiccola, B., Manco, M. *Reference ranges of HOMA-IR in normal-weight and obese young Caucasians.* (2016) - not a diagnostic study design

41. Tresaco, B.. Bueno, G., Moreno, L. A., Garagorri, J. M., Bueno, M. *Insulin resistance and impaired glucose tolerance in obese children and adolescents*. (2003) - not a diagnostic study design

42. Urakami, T., Habu, M., Kuwabara, R., Komiya, K., Nagano, N., Suzuki, J., Mugishima, H. *Insulin resistance at diagnosis in Japanese children with type 2 diabetes mellitus.* (2012) - not a diagnostic study design

43. Valerio, G., Licenziati, M. R., Iannuzzi, A., Franzese, A., Siani, P., Riccardi, G., Rubba, P. *Insulin resistance and impaired glucose tolerance in obese children and adolescents from Southern Italy.* (2006) - not a diagnostic study design

44. Valery, P. C., Moloney, A., Cotterill, A., Harris, M., Sinha, A. K., Green, A. C. *Prevalence of obesity and metabolic syndrome in Indigenous Australian youths*. (2009) - not a diagnostic study design

45. Van Der Aa, M. P., Fazeli Farsani, S., Kromwijk, L. A. J., De Boer, A., Knibbe, C. A. J., Van Der Vorst, M. M. J. *How to screen obese children at risk for type 2 diabetes mellitus?* (2014) - not a diagnostic study design

46. Vijayadeva, V., Nichols, G. A. *Impact of implementing glycated hemoglobin testing for identification of dysglycemia in youth.* (2014) - not a diagnostic study design

47. Wang, C. L., Liang, L., Fu, J. F., Hong, F. *Comparison of methods to detect insulin resistance in obese children and adolescents*. (2005) - not a diagnostic study design

<u>Studies that were only abstracts available</u>

1. Gao, S., Li, M., Qu, X. X., Wang, Y. H., Zhang, X. J., Zhang, X. J., Willi, S. M. *ADA HbA1c Diagnostic Criteria Fail to Identify Prediabetes and Diabetes in a Population of Chinese Adolescents and Young Adults at High Risk for Diabetes.* (2015) – conference abstract

2. Hwang, J. W., Kim, S. Y., Lee, D. Y., Kim, M. S. *Hemoglobin A1c measurement for the diagnosis of type 2 diabetes in Korean children.* (2016) – conference paper

3. Kaya, A., Kocyigit, C., Catli, G., Can, P. S., Sutcu, R., Dundar, B. N. *55th Annual Meeting of the European Society for Paediatric Endocrinology (ESPE), Paris, France, September 10-12, 2016: Abstracts.* (2016) – conference paper, no data available

4. Kim, M. S., Lee, D. Y. *Poster Sessions/Hemoglobin A1c measurement for the diagnosis of type 2 diabetes in Korean children.* (2015) - conference paper, no data available

5. Kumbhojkar, A. P., Saraff, V., Hogler, W. *10th Individual Abstracts for International Meeting of Pediatric Endocrinology: Free Communication and Poster Sessions, Abstracts.* (2017) - conference paper, no data available

6. Lantigua, H., Rubio, N., Yafi, M. *25th European Congress on Obesity, Vienna, Austria, May 23-26, 2018: Abstracts.* (2018) – conference paper, no data available

7. Lee, J., Kim, J. H. *Optimal cutoff of hemoglobin A1c for detecting impaired fasting glucose in Korean Youth: The Korea National Health and Nutrition Survey 2011-2012.kim* (2016) - conference paper, no data available

8. Pellegrin, M. C., Radillo, L., Grillo, A., Tornese, G., Faleschini, E., Ventura, A. *Predictive value of glycated hemoglobin or OGTT in Italian overweight/obese children*. (2015) - conference paper, no data available

9. Vukovic, R., Milenkovic, T., Mitrovic, K., Todorovic, S., Plavsic, L., Soldatovic, I. *10th Individual Abstracts for International Meeting of Pediatric Endocrinology: Free Communication and Poster Sessions, Abstracts.* (2017) - conference paper, no data available

10. Yoon, J. S., So, C. H., Lee, H. S., Lim, J. S., Hwang, J. S. *Usefulness of combined use of HBALC and fasting plasma glucose for screening of glucose intolerance and diabetes in children and adolescents*. (2017) - conference paper, no data available

Appendix 7

P	1.	Children **OR** teenager **OR** kid **OR** non adults **OR** early ontogenetic stage **OR** youngster **OR** adolescent **OR** youth **OR** child **OR** young **OR** youth diabetes **OR** juvenile diabetes
I	2.	HOMA **OR** HOMA-IR **OR** homeostatic model assessment of insulin resistance **OR** postprandial glucose test
	3.	HbA1c **OR** H$moglobin A1c **OR** glycoh$moglobin **OR** h$moglobin A1c **OR** A1c **OR** A1c h$moglobin **OR** HGBA1C **OR** hemoglobin A1C
	4.	1.5 anhydroglucitol **OR** 1.5 – AG **OR** anhydro-d-glucitol **OR** glycoMark **OR** anhydroglucitol **OR** dianhydro-d-glucitol
R	5.	OGTT **OR** oral glucose tolerance test
	6.	FPG **OR** fasting plasma glucose
D.	7.	Type 2 pre-diabetes mellitus **OR** insulin resistance **OR** impaired glucose tolerance **OR** impaired glucose metabolism **OR** type 2 diabetes **OR** non-insulin dependent diabetes
	8.	Diag **OR** sensitivity **OR** specificity **OR** predictive value **OR** ROC **OR** receiver operating characteristic
	9.	1 AND (2 OR 3 OR 4) AND (5 OR 6) AND 7 AND 8

Appendix 8

MEDLINE(R) ALL <1946 to March 19, 2020> (Ovid)

Search was conducted on 20th March 2020 at 11:20 am (CET) in Keywords

#	search string	results
1	exp Child/	1884300
2	exp Adolescent/	1998137
3	child*.mp	2402528
4	adolescen*.mp	2072531
5	kid?.mp.	8593
6	youngster?.mp	2492
7	youth?.mp	77349
8	teen*.mp.	30262
9	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8	3485569
10	obesity.mp	316442
11	overweight.mp	72197
12	body weight.mp	339271
13	exp body mass index/	124043
14	body mass index.mp	233715
15	Quetelet? Index.mp	713
16	Quetelet's Index.mp	226
17	BMI.mp	140603
18	obesity/ or pediatric obesity/	184537
19	exp Waist-Hip Ratio/	4039
20	Waist Hip Ratio?.mp.	7075

21	Waist-Hip Ratio?.mp.	7075
22	Waist to Hip Ratio?.mp.	11259
23	Waist-to-Hip Ratio?.mp	11259
24	WHR.mp	4642
25	10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24	778827
26	exp Hypertension/	251488
27	hypertension.mp	481143
28	high blood pressure.mp.	14793
29	26 or 27 or 28	485902
30	anhydroglucitol?.mp.	505
31	1,5-AG.mp	203
32	anhydro-D-glucitol?.mp	162
33	Deoxy-D-glucopyranose?.mp	105
34	GlycoMark.mp.	14
35	30 or 31 or 32 or 33 or 34	683
36	exp Triglycerides/	75993
37	triglyceride?.mp	142898
38	triacylglycerol?.mp	16203
39	triacylglyceride?.mp.	980
40	TG.mp	52037
41	TAG.mp	34730
42	36 or 37 or 38 or 39 or 40 or 41	217444
43	exp Cholesterol, HDL/	27983
44	HDL? Cholesterol?.mp	28330

45	High Density Lipoprotein Cholesterol?.mp	25444
46	alpha-Lipoprotein Cholesterol?.mp	67
47	alpha lipoprotein cholesterol?.mp	67
48	43 or 44 or 45 or 46 or 47	62699
49	Dyslipidemias/	11206
50	Dyslipidemia?.mp	32840
51	Dyslipoproteinemia?.mp.	846
52	49 or 50 or 51	33571
53	homeostatic model assessment of insulin resistance.mp.	1723
54	HOMA.mp.	15702
55	HOMA-IR.mp.	11211
56	53 or 54 or 55	16356
57	postprandial glucose.mp	3390
58	PPG.mp	3328
59	57 or 58	6492
60	exp Glycated Hemoglobin A/	34262
61	Glycated H?emoglobin?.mp.	40539
62	Glycosylated H?emoglobin?.mp	9945
63	Glycoh?emoglobin?.mp	970
64	HbA1*.mp.	34642
65	A1c.mp.	21735
66	Hb A1*.mp.	610
67	HGBA1C.mp.	179
68	Hb1c.mp	15

69	60 or 61 or 62 or 63 or 64 or 65 or 66 or 67 or 68	65958
70	exp Glucose Tolerance Test/	34553
71	Glucose Tolerance.mp	58583
72	OGTT.mp	8481
73	GTT.mp	1718
74	70 or 71 or 72 or 73	60123
75	glucose test?.mp.	1122
76	fasting glucose.mp	16800
77	fasting blood glucose.mp	11639
78	fasting plasma glucose.mp	12352
79	FPG.mp	5810
80	fasting blood sugar.mp	2330
81	fasting plasma sugar.mp	24
82	fasting sugar.mp	48
83	FBS.mp	7707
84	75 or 76 or 77 or 78 or 79 or 80 or 81 or 82 or 83	50227
85	exp Prediabetic State/	6582
86	prediabet*.mp	10594
87	pre-diabet*.mp	2606
88	pre diabet*.mp	2606
89	borderline diabet*.mp.	119
90	chemical? diabet*.mp.	251
91	latent diabet*.mp.	338
92	T2P.mp	53

93	Impaired fasting glucose.mp	3831
94	Impaired fasting glyce?mia.mp	129
95	IFG.mp	3581
96	Impaired glucose tolerance?.mp.	10965
97	IGT.mp	4870
98	Impaired glucose metabolism.mp.	1334
99	exp Glucose Intolerance/	8517
100	Glucose Intolerance?.mp.	15947
101	85 or 86 or 87 or 88 or 89 or 90 or 91 or 92 or 93 or 94 or 95 or 96 or 97 or 98 or 99 or 100	38330
102	Diagnosis/	17305
103	diagnos*.mp	4907489
104	detect*.mp.	2323874
105	accura*.mp	784514
106	exp "Sensitivity and Specificity"/	575535
107	sensitiv*.mp.	1663848
108	specificit*.mp	1048949
109	Receiver Operating Characteristic?.mp	66295
110	ROC.mp	86399
111	Predictive Value?.mp	271480
112	102 or 103 or 104 or 105 or 106 or 107 or 108 or 109 or 110 or 111	8411683
113	25 or 29 or 35 or 42 or 48 or 52 or 56 or 59 or 69	1421989
114	69 or 74 or 84	151575
115	9 and 101 and 112 and 113 and 114	1325
116	2015 -Current"	399

Embase (Elsevier)

Search was conducted on 20th March 2020 at 11:45 am (CET) in Title, Abstract, Keywords

#	search string	results
1	'child'/exp	2840901
2	'adolescent'/exp	1636695
3	child*:ti,ab,kw	1807784
4	adolescen*:ti,ab,kw	368654
5	kid$:ti,ab,kw	12274
6	youngster$:ti,ab,kw	3473
7	youth$:ti,ab,kw	89145
8	teen*:ti,ab,kw	41335
9	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8	4158938
10	obesity:ti,ab,kw	369395
11	overweight:ti,ab,kw	102662
12	'body weight':ti,ab,kw	270747
13	'body mass'/exp	421532
14	'body mass index':ti,ab,kw	261015
15	whr:ti,ab,kw	6877
16	'quetelet* index':ti,ab,kw	567
17	bmi:ti,ab,kw	301483
18	'obesity'/de OR 'adolescent obesity'/de OR 'childhood obesity'/de OR 'diabetic obesity'/de	436087
19	'waist hip ratio'/exp	14013
20	'waist hip ratio$':ti,ab,kw	6413
21	'waist-hip ratio$':ti,ab,kw	6413

22	'waist to hip ratio$':ti,ab,kw	7812
23	'waist-to-hip ratio$':ti,ab,kw	7812
24	#10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23	1105798
25	'hypertension'/de OR 'diabetic hypertension'/de	593001
26	hypertension:ti,ab,kw	597381
27	'high blood pressure':ti,ab,kw	22178
28	#25 OR #26 OR #27	839091
29	anhydroglucitol$:ti,ab,kw	583
30	'1,5 ag':ti,ab,kw	361
31	'anhydro-d-glucitol$':ti,ab,kw	211
32	'deoxy-d-glucopyranose$':ti,ab,kw	128
33	glycomark:ti,ab,kw	34
34	#29 OR #30 OR #31 OR #32 OR #33	703
35	'triacylglycerol'/exp	198268
36	triglyceride$:ti,ab,kw	157061
37	triacylglycerol$:ti,ab,kw	18564
38	triacylglyceride$:ti,ab,kw	1219
39	tg:ti,ab,kw	78785
40	tag:ti,ab,kw	42510
41	#35 OR #36 OR #37 OR #38 OR #39 OR #40	333089
42	'high density lipoprotein cholesterol'/exp	102679
43	'hdl$ cholesterol$':ti,ab,kw	41031
44	'high density lipoprotein cholesterol$':ti,ab,kw	30712
45	'alpha-lipoprotein cholesterol$':ti,ab,kw	23

46	'alpha lipoprotein cholesterol$':ti,ab,kw	23
47	#42 OR #43 OR #44 OR #45 OR #46	120269
48	'dyslipidemia'/de	69304
49	dyslipidemia$:ti,ab,kw	50495
50	dyslipoproteinemia$:ti,ab,kw	1098
51	#48 OR #49 OR #50	82171
52	'homeostatic model assessment of insulin resistance':ti,ab,kw	1218
53	homa:ti,ab,kw	28522
54	'homa-ir':ti,ab,kw	20154
55	#52 OR #53 OR #54	28923
56	'postprandial glucose':ti,ab,kw	5103
57	ppg:ti,ab,kw	4912
58	#56 OR #57	9559
59	'glycosylated hemoglobin'/exp	122129
60	'glycated h$emoglobin$':ti,ab,kw	15744
61	'glycosylated h$emoglobin$':ti,ab,kw	13289
62	glycoh$emoglobin$:ti,ab,kw	1348
63	hba1*:ti,ab,kw	59656
64	a1c:ti,ab,kw	22623
65	'hb a1*':ti,ab,kw	759
66	hgba1c:ti,ab,kw	677
67	hb1c:ti,ab,kw	47
68	#59 OR #60 OR #61 OR #62 OR #63 OR #64 OR #65 OR #66 OR #67	138111
69	'glucose tolerance test'/exp	63110

70	'glucose tolerance':ti,ab,kw	64886
71	ogtt:ti,ab,kw	16136
72	gtt:ti,ab,kw	3317
73	#69 OR #70 OR #71 OR #72	94050
74	'glucose test$':ti,ab,kw	1828
75	'fasting glucose':ti,ab,kw	29082
76	'fasting blood glucose':ti,ab,kw	18966
77	'fasting plasma glucose':ti,ab,kw	18826
78	fpg:ti,ab,kw	10082
79	'fasting blood sugar':ti,ab,kw	4141
80	'fasting plasma sugar':ti,ab,kw	35
81	'fasting sugar':ti,ab,kw	115
82	fbs:ti,ab,kw	13855
83	#74 OR #75 OR #76 OR #77 OR #78 OR #79 OR #80 OR #81 OR #82	83882
84	'impaired glucose tolerance'/exp	30068
85	prediabet*:ti,ab,kw	15400
86	'pre-diabet*':ti,ab,kw	4971
87	'pre diabet*':ti,ab,kw	4971
88	'borderline diabet*'	182
89	'chemical$ diabet*':ti,ab,kw	362
90	'latent diabet*':ti,ab,kw	443
91	t2p:ti,ab,kw	66
92	'impaired fasting glucose':ti,ab,kw	6155
93	'impaired fasting glyce$mia':ti,ab,kw	186

94	ifg:ti,ab,kw	5760
95	'impaired glucose tolerance$':ti,ab,kw	16615
96	igt:ti,ab,kw	7960
97	'impaired glucose metabolism':ti,ab,kw	2038
98	'glucose intolerance'/exp	17967
99	'glucose intolerance$':ti,ab,kw	14964
100	#84 OR #85 OR #86 OR #87 OR #88 OR #89 OR #90 OR #91 OR #92 OR #93 OR #94 OR #95 OR #96 OR #97 OR #98 OR #99	67515
101	'diagnosis'/de OR 'diagnostic test'/de OR 'diagnostic accuracy'/exp OR 'diagnostic test accuracy study'/exp	1727079
102	diagnos*:ti,ab,kw	3541990
103	detect*:ti,ab,kw	2923534
104	accura*:ti,ab,kw	996216
105	'sensitivity and specificity'/exp OR 'receiver operating characteristic'/exp OR 'predictive value'/exp	520990
106	sensitiv*:ti,ab,kw	1719023
107	specificit*:ti,ab,kw	622087
108	'receiver operating characteristic$':ti,ab,kw	86175
109	roc:ti,ab,kw	90599
110	'predictive value$':ti,ab,kw	158559
111	#101 OR #102 OR #103 OR #104 OR #105 OR #106 OR #107 OR #108 OR #109 OR #110	8384528
112	#24 OR #28 OR #34 OR #41 OR #47 OR #51 OR #55 OR #58 OR #68	2142760
113	#68 OR #73 OR #83	269725
114	#9 AND #100 AND #111 AND #112 AND #113	1844
115	limitation 2. 4. 2015 – -20. 3. 2020	701

CINAHL (EBSCO)

Search was conducted on 19th March 2020 at 12:55 pm (CET) in Title and Abstract

#	search string	results
1	(MH "Child+")	692806
2	(MH "Adolescence+")	552742
3	TI child* OR AB child*	508069
4	TI Adolescen* OR AB Adolescen*	142099
5	TI kid# OR AB kid#	10205
6	TI youngster# OR AB youngster#	899
7	TI youth# OR AB youth#	52674
8	TI teen* OR AB teen*	19706
9	S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8	1192084
10	TI obesity OR AB obesity	82439
11	TI overweight OR AB overweight	33263
12	TI "body weight" OR AB "body weight"	30706
13	(MH "Body Mass Index")	90919
14	TI "body mass index" OR AB "body mass index"	66481
15	TI "Quetelet* Index" OR AB "Quetelet* Index"	41
16	TI WHR OR AB WHR	1154
17	TI BMI OR AB BMI	52524
18	(MH "Obesity") OR (MH "Pediatric Obesity")	99089
19	(MH "Waist-Hip Ratio")	3162
20	TI "Waist Hip Ratio#" OR AB "Waist Hip Ratio#"	1118
21	TI "Waist-Hip Ratio#" OR AB "Waist-Hip Ratio#"	1118

22	TI "Waist to Hip Ratio#" OR AB "Waist to Hip Ratio#"	1846
23	TI "Waist-to-Hip Ratio#" OR AB "Waist-to-Hip Ratio#"	1846
24	S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23	241923
25	(MH "Hypertension+")	80157
26	TI hypertension OR AB hypertension	84223
27	TI "high blood pressure" OR AB "high blood pressure"	4969
28	S25 OR S26 OR S27	120629
29	TI anhydroglucitol# OR AB anhydroglucitol#	9
30	"1,5-AG" OR AB "1,5-AG"	54
31	TI "anhydro-D-glucitol#" OR AB "anhydro-D-glucitol#"	2
32	TI "Deoxy-D-glucopyranose#" OR AB "Deoxy-D-glucopyranose#"	37
33	TI GlycoMark OR AB GlycoMark	9
34	S29 OR S30 OR S31 OR S32 OR S33	68
35	(MH "Triglycerides")	14873
36	TI triglyceride# OR AB triglyceride#	18858
37	TI triacylglycerol# OR AB triacylglycerol#	1728
38	TI triacylglyceride# OR AB triacylglyceride#	68
39	TI TG OR AB TG	5501
40	TI TAG OR AB TAG	3008
41	S35 OR S36 OR S37 OR S38 OR S39 OR S40	31845
42	(MH "Lipoproteins, HDL Cholesterol")	9151
43	TI "HDL# Cholesterol#" OR AB "HDL# Cholesterol#"	5631
44	TI "High Density Lipoprotein Cholesterol#" OR AB "High Density Lipoprotein Cholesterol#"	6612
45	TI "alpha-Lipoprotein Cholesterol#" OR AB "alpha-Lipoprotein Cholesterol#"	16578

46	TI "alpha lipoprotein cholesterol#" OR AB "alpha lipoprotein cholesterol#"	16578
47	S42 OR S43 OR S44 OR S45 OR S46	16921
48	(MH "Hyperlipidemia+")	21315
49	TI Dyslipidemia# OR AB Dyslipidemia#	6824
50	TI Dyslipoproteinemia# OR AB Dyslipoproteinemia#	53
51	S48 OR S49 OR S50	25158
52	TI "homeostatic model assessment of insulin resistance" OR AB "homeostatic model assessment of insulin resistance	591
53	TI HOMA OR AB HOMA	3856
54	TI "HOMA-IR" OR AB "HOMA-IR"	2909
55	S52 OR S53 OR S54	4127
56	TI "postprandial glucose" OR AB "postprandial glucose"	1281
57	TI PPG OR AB PPG	559
58	S56 OR S57	1725
59	(MH "Hemoglobin A, Glycosylated")	17742
60	TI "Glycated H#emoglobin#" OR AB "Glycated H#emoglobin#"	3824
61	TI "Glycosylated H#emoglobin#" OR AB "Glycosylated H#emoglobin#"	3061
62	TI Glycoh#emoglobin# OR AB Glycoh#emoglobin#	218
63	TI HbA1* OR AB HbA1*	12529
64	TI A1c OR AB A1c	6833
65	TI "Hb A1*" OR AB "Hb A1*"	156
66	TI HGBA1C OR AB HGBA1C	89
67	TI Hb1c OR AB Hb1c	4
68	S59 OR S60 OR S61 OR S62 OR S63 OR S64 OR S65 OR S66 OR S67	29301
69	(MH "Glucose Tolerance Test")	7576

70	TI "Glucose Tolerance" OR AB "Glucose Tolerance"	9469
71	TI OGTT OR AB OGTT	2025
72	TI GTT OR AB GTT	255
73	S69 OR S70 OR S71 OR S72	12832
74	TI "glucose test#" OR AB "glucose test#"	376
75	TI "fasting glucose" OR AB "fasting glucose"	5564
76	TI "fasting blood glucose" OR AB "fasting blood glucose"	3119
77	TI "fasting plasma glucose" OR AB "fasting plasma glucose"	3782
78	TI FPG OR AB FPG	1477
79	TI "fasting blood sugar" OR AB "fasting blood sugar"	665
80	TI "fasting plasma sugar" OR AB fasting plasma sugar"	7
81	TI "fasting sugar" OR AB "fasting sugar"	8
82	TI FBS OR AB FBS	883
83	S74 OR S75 OR S76 OR S77 OR S78 OR S79 OR S80 OR S81 OR S82	13472
84	(MH "Prediabetic State")	3205
85	TI Prediabet* OR AB Prediabet*	2910
86	TI "pre-diabet*" OR AB "pre-diabet*"	984
87	TI "pre diabet*" OR AB "pre diabet*"	984
88	TI "borderline diabet*" OR AB "borderline diabet*"	16
89	TI "chemical# diabet*" OR AB "chemical# diabet*"	2
90	TI "latent diabet*" OR AB "latent diabet*"	6
91	TI T2P OR AB T2P	4
92	TI "Impaired fasting glucose" OR AB "Impaired fasting glucose"	1490
93	TI "Impaired fasting glyce#mia" OR AB "Impaired fasting glyce#mia"	41

94	TI IFG OR AB IFG	955
95	TI "Impaired glucose tolerance#" OR AB "Impaired glucose tolerance#"	2930
96	TI IGT OR AB IGT	1350
97	TI "Impaired glucose metabolism" OR AB "Impaired glucose metabolism"	340
98	(MH "Glucose Intolerance")	3328
99	TI "Glucose Intolerance#" OR AB "Glucose Intolerance#"	1890
100	S84 OR S85 OR S86 OR S87 OR S88 OR S89 OR S90 OR S91 OR S92 OR S93 OR S94 OR S95 OR S96 OR S97 OR S98 OR S99	11610
101	(MH "Diagnosis")	8509
102	TI Diagnos* OR AB Diagnos*	531571
103	TI Detect* OR AB Detect*	228502
104	TI Accura* OR AB Accura*	140337
105	(MH "Sensitivity and Specificity") OR (MH "ROC Curve") OR (MH "Predictive Value of Tests")	140726
106	TI Sensitiv* OR AB Sensitiv*	164557
107	TI Specificit* OR AB Specificit*	58465
108	TI "Receiver Operating Characteristic#" OR AB "Receiver Operating Characteristic"	17642
109	TI ROC OR AB ROC	12025
110	TI "Predictive Value#" OR AB "Predictive Value#"	26663
111	S101 OR S102 OR S103 OR S104 OR S105 OR S106 OR S107 OR S108 OR S109 OR S110	958846
112	S24 OR S28 OR S34 OR S41 OR S47 OR S51 OR S55 OR S58 OR S68	393741
113	S68 OR S73 OR S83	47775
114	S9 AND S100 AND S111 AND S112 AND S113	404
115	limitation 1. 4. 2015 – 20. 3. 2020	137

Web of Science Core Collection

Search was conducted on 19th March 2020 at 10:40 am (CET) in Topic

#	search string	results
1	TOPIC: (child*)	1813250
2	TOPIC: (Adolescen*)	447269
3	TOPIC: (kid$)	19456
4	TOPIC: (youngster$)	3644
5	TOPIC: (youth$)	152972
6	TOPIC: (teen*)	41657
7	#6 OR #5 OR #4 OR #3 OR #2 OR #1	2117532
8	TOPIC: ("obesity")	333796
9	TOPIC: ("overweight")	93847
10	TOPIC: ("body weight")	205582
11	TOPIC: ("body mass index")	202401
12	TS=("Quetelet* Index")	276
13	TS=("WHR")	4363
14	TOPIC: ("BMI")	139003
15	TOPIC: ("Waist Hip Ratio$")	4041
16	TOPIC: ("Waist-Hip Ratio$")	4041
17	TOPIC: ("Waist to Hip Ratio$")	5587
18	TOPIC: ("Waist-to-Hip Ratio$")	5587
19	#18 OR #17 OR #16 OR #15 OR #14 OR #13 OR #12 OR #11 OR #10 OR #9 OR #8	67074
20	TOPIC: ("hypertension")	437377
21	TOPIC: ("high blood pressure")	15221

22	#21 OR #20	443629
23	TOPIC: (anhydroglucitol$)	535
24	TOPIC: ("1,5-AG")	252
25	TOPIC: ("anhydro-D-glucitol$")	329
26	TOPIC: ("Deoxy-D-glucopyranose$")	204
27	TOPIC: ("GlycoMark")	34
28	#27 OR #26 OR #25 OR #24 OR #23	965
29	TOPIC: (triglyceride$)	113935
30	TOPIC: (triacylglycerol$)	23194
31	TOPIC: (triacylglyceride$)	1375
32	TOPIC: ("TG")	98054
33	TOPIC: ("TAG")	62172
34	#33 OR #32 OR #31 OR #30 OR #29	273354
35	TOPIC: ("HDL* Cholesterol$")	28264
36	TOPIC: ("High Density Lipoprotein Cholesterol$")	23082
37	TOPIC: ("alpha-Lipoprotein Cholesterol$")	18
38	TOPIC: ("alpha lipoprotein cholesterol$")	18
39	#38 OR #37 OR #36 OR #35	48625
40	TOPIC: (Dyslipidemia$)	30074
41	TOPIC: (Dyslipoproteinemia$)	805
42	#41 OR #40	30787
43	TOPIC: ("homeostatic model assessment of insulin resistance")	886
44	TOPIC: ("HOMA")	15364
45	TOPIC: ("HOMA-IR")	10564

46	#45 OR #44 OR #43	15714
47	TOPIC: ("postprandial glucose")	3865
48	TOPIC: ("PPG")	642
49	#48 OR #47	10049
50	TOPIC: ("Glycated Hemoglobin$" OR "Glycated Haemoglobin$")	12143
51	TOPIC: ("Glycosylated Hemoglobin$" OR "Glycosylated Haemoglobin$")	9382
52	TOPIC: ("Glycohemoglobin$" OR "Glycohaemoglobin$")	1117
53	TOPIC: ("HbA1*")	24615
54	TOPIC: ("A1c")	11824
55	TOPIC: ("Hb A1*")	353
56	TOPIC: ("HGBA1C")	155
57	TOPIC: ("Hb1c")	14
58	#57 OR #56 OR #55 OR #54 OR #53 OR #52 OR #51 OR #50	45932
59	TOPIC: ("Glucose Tolerance")	56429
60	TOPIC: ("OGTT")	7632
61	TOPIC: ("GTT")	1568
62	#61 OR #60 OR #59	58491
63	TOPIC: ("glucose test$")	974
64	TOPIC: ("fasting glucose")	18106
65	TOPIC: ("fasting blood glucose")	10239
66	TOPIC: ("fasting plasma glucose")	11745
67	TOPIC: ("FPG")	5566
68	TOPIC: ("fasting blood sugar")	19
69	TOPIC: ("fasting plasma sugar")	14

70	TOPIC: ("fasting sugar")	40
71	TOPIC: ("FBS")	7896
72	#71 OR #70 OR #69 OR #68 OR #67 OR #66 OR #65 OR #64 OR #63	49281
73	TOPIC: (Prediabet*)	8096
74	TOPIC: ("pre-diabet*")	2922
75	TOPIC: ("pre diabet*")	2922
76	TOPIC: ("borderline diabet*")	91
77	TOPIC: ("chemical$ diabet*")	178
78	TOPIC: ("latent diabet*")	145
79	TOPIC: ("T2P")	55
80	TOPIC: ("Impaired fasting glucose")	4734
81	TOPIC: ("Impaired fasting glyce$mia")	143
82	TOPIC: ("IFG")	35
83	TOPIC: ("Impaired glucose tolerance$")	17308
84	TOPIC: ("IGT")	4973
85	TOPIC: ("Impaired glucose metabolism")	142
86	TOPIC: ("Glucose Intolerance$")	1135
87	#86 OR #85 OR #84 OR #83 OR #82 OR #81 OR #80 OR #79 OR #78 OR #77 OR #76 OR #75 OR #74 OR #73	4287
88	TOPIC: (Diagnos*)	2456779
89	TOPIC: (Detect*)	3631189
90	TOPIC: (Accura*)	2056559
91	TOPIC: (Sensitiv*)	2167097
92	TOPIC: (Specificit*)	53177
93	TOPIC: ("Receiver Operating Characteristic$")	67389

94	TOPIC: ("ROC")	55836
95	TOPIC: ("Predictive Value$")	105254
96	#95 OR #94 OR #93 OR #92 OR #91 OR #90 OR #89 OR #88	8873906
97	#58 OR #49 OR #46 OR #42 OR #39 OR #34 OR #28 OR #22 OR #19	1339420
98	#72 OR #62 OR #58	132291
99	#98 AND #97 AND #96 AND #87 AND #7	972
100	Limitation 2015–2020	310

Scopus

Search was conducted on 19th March 2020 at 3:00 pm–6:00 pm (CET) in Article title, Abstract, Keywords.

#	search string	results
1	TITLE-ABS-KEY (child*)	3206117
2	TITLE-ABS-KEY (adolescen*)	2305498
3	TITLE-ABS-KEY (kid OR kids)	18237
4	TITLE-ABS-KEY (youngster*)	6466
5	TITLE-ABS-KEY (youth*)	163595
6	TITLE-ABS-KEY (teen*)	53100
7	#1 OR #2 OR #3 OR #4 OR #5 OR #6	4429259
8	TITLE-ABS-KEY (obesity)	454598
9	TITLE-ABS-KEY (overweight)	85932
10	TITLE-ABS-KEY ("body weight")	531773
11	TITLE-ABS-KEY ("body mass index")	245370
12	TITLE-ABS-KEY ("Quetelet* Index")	842
13	TITLE-ABS-KEY (bmi)	160777
14	TITLE-ABS-KEY ("Waist Hip Ratio*")	13838
15	TITLE-ABS-KEY ("Waist-Hip Ratio*")	13838
16	TITLE-ABS-KEY ("Waist to Hip Ratio*")	6609
17	TITLE-ABS-KEY ("Waist-to-Hip Ratio*")	6609
18	TITLE-ABS-KEY (whr)	6095
19	#8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18	1056537
20	TITLE-ABS-KEY (hypertension)	766460
21	TITLE-ABS-KEY ("high blood pressure")	22090

22	#20 OR #21	771972
23	TITLE-ABS-KEY (anhydroglucitol*)	564
24	TITLE-ABS-KEY ("1,5-AG")	249
25	TITLE-ABS-KEY ("anhydro-D-glucitol*")	259
26	TITLE-ABS-KEY ("Deoxy-D-glucopyranose*")	258
27	TITLE-ABS-KEY (glycomark)	16
28	#23 OR #24 OR #25 OR #26 OR #27	1007
29	TITLE-ABS-KEY (triglyceride*)	170771
30	TITLE-ABS-KEY (triacylglycerol*)	209273
31	TITLE-ABS-KEY (triacylglyceride*)	1578
32	TITLE-ABS-KEY (tg)	152094
33	TITLE-ABS-KEY (tag)	121090
34	#29 OR #30 OR #31 OR #32 OR #33	512101
35	TITLE-ABS-KEY ("HDL* Cholesterol*")	36512
36	TITLE-ABS-KEY ("High Density Lipoprotein Cholesterol*")	96486
37	TITLE-ABS-KEY ("alpha-Lipoprotein Cholesterol*")	78
38	TITLE-ABS-KEY ("alpha lipoprotein cholesterol*")	78
39	#35 OR #36 OR #37 OR #38	109781
40	TITLE-ABS-KEY (dyslipidemia*)	63617
41	TITLE-ABS-KEY (dyslipoproteinemia*)	1541
42	#40 OR #41	64909
43	TITLE-ABS-KEY ("homeostatic model assessment of insulin resistance")	1037
44	TITLE-ABS-KEY (homa)	17564
45	TITLE-ABS-KEY ("HOMA-IR")	12105

46	#43 OR #44 OR #45	17964
47	TITLE-ABS-KEY ("postprandial glucose")	3937
48	TITLE-ABS-KEY (ppg)	9025
49	#47 OR #48	12702
50	TITLE-ABS-KEY ("Glycated H*emoglobin*")	16856
51	TITLE-ABS-KEY ("Glycosylated H*emoglobin*")	35436
52	TITLE-ABS-KEY (glycoh*emoglobin*)	1207
53	TITLE-ABS-KEY (hba1*)	40850
54	TITLE-ABS-KEY (a1c)	71863
55	TITLE-ABS-KEY ("Hb A1*")	711
56	TITLE-ABS-KEY (hgba1c)	198
57	TITLE-ABS-KEY (hb1c)	23
58	#50 OR #51 OR #52 OR #53 OR #54 OR #55 OR #56 OR #57	99398
59	TITLE-ABS-KEY ("Glucose Tolerance")	90047
60	TITLE-ABS-KEY (ogtt)	10078
61	TITLE-ABS-KEY (gtt)	2614
62	#59 OR #60 OR #61	91918
63	TITLE-ABS-KEY ("glucose test*")	2514
64	TITLE-ABS-KEY ("fasting glucose")	18469
65	TITLE-ABS-KEY ("fasting blood glucose")	14800
66	TITLE-ABS-KEY ("fasting plasma glucose")	13800
67	TITLE-ABS-KEY (fpg)	7089
68	TITLE-ABS-KEY ("fasting blood sugar")	3619
69	TITLE-ABS-KEY ("fasting plasma sugar")	28

70	TITLE-ABS-KEY ("fasting sugar")	71
71	TITLE-ABS-KEY (fbs)	30777
72	#63 OR #64 OR #65 OR #66 OR #67 OR #68 OR #69 OR #70 OR #71	81928
73	TITLE-ABS-KEY (prediabet*)	11303
74	TITLE-ABS-KEY ("pre-diabet*")	3042
75	TITLE-ABS-KEY ("pre diabet*")	3042
76	TITLE-ABS-KEY ("borderline diabet*")	170
77	TITLE-ABS-KEY ("chemical* diabet*")	440
78	TITLE-ABS-KEY ("latent diabet*")	498
79	TITLE-ABS-KEY (t2p)	117
80	TITLE-ABS-KEY ("Impaired fasting glucose")	4347
81	TITLE-ABS-KEY ("Impaired fasting glyce*mia")	165
82	TITLE-ABS-KEY (ifg)	4063
83	TITLE-ABS-KEY ("Impaired glucose tolerance*")	26595
84	TITLE-ABS-KEY (igt)	6254
85	TITLE-ABS-KEY ("Impaired glucose metabolism")	1435
86	TITLE-ABS-KEY ("Glucose Intolerance*")	22446
87	#73 OR #74 OR #75 OR #76 OR #77 OR #78 OR #79 OR #80 OR #81 OR #82 OR #83 OR #84 OR #85 OR #86	54864
88	TITLE-ABS-KEY (diagnos*)	4950878
89	TITLE-ABS-KEY (detect*)	4963555
90	TITLE-ABS-KEY (accura*)	3159895
91	TITLE-ABS-KEY (sensitiv*)	3208639
92	TITLE-ABS-KEY (specificit*)	1264075
93	TITLE-ABS-KEY ("Receiver Operating Characteristic*")	119467

94	TITLE-ABS-KEY ("ROC")	100757
95	TITLE-ABS-KEY ("Predictive Value*")	324081
96	#88 OR #89 OR #90 OR #91 OR #92 OR #93 OR #94 OR #95	13985131
97	#19 OR #22 OR #28 OR #34 OR #39 OR #42 OR #46 OR #49 OR #58	2223927
98	#58 OR #62 OR #72	237248
99	#7 AND #87 AND #96 AND #97 AND #98	1908
100	Limitation 2015–2020	611

Emcare <1995 to 2020 week 11> (Ovid)

Search was conducted on 20th March 2020 at 12:00 am (CET) in Keywords.

#	search string	results
1	exp Child/	609729
2	exp Adolescent/	327336
3	child*.mp	653525
4	adolescen*.mp	358791
5	kid?.mp.	3660
6	youngster?.mp	1179
7	youth?.mp	49593
8	teen*.mp.	15201
9	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8	921217
10	obesity.mp	148814
11	overweight.mp	36626
12	body weight.mp	100544
13	exp body mass index/	132633
14	body mass index.mp	77502
15	Quetelet? Index.mp	97
16	Quetelet's Index.mp	22
17	BMI.mp	62668
18	obesity/ or pediatric obesity/	119594
19	exp Waist-Hip Ratio/	5232
20	Waist Hip Ratio?.mp.	5528
21	Waist-Hip Ratio?.mp.	5528

22	Waist to Hip Ratio?.mp.	5685
23	Waist-to-Hip Ratio?.mp	5685
24	WHR.mp	1832
25	10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24	298048
26	exp Hypertension/	151902
27	hypertension.mp	168451
28	high blood pressure.mp.	4645
29	26 or 27 or 28	183269
30	anhydroglucitol?.mp.	110
31	1,5-AG.mp	59
32	anhydro-D-glucitol?.mp	13
33	Deoxy-D-glucopyranose?.mp	0
34	GlycoMark.mp.	6
35	30 or 31 or 32 or 33 or 34	125
36	exp Triglycerides/	41457
37	triglyceride?.mp	25640
38	triacylglycerol?.mp	46666
39	triacylglyceride?.mp.	150
40	TG.mp	9325
41	TAG.mp	5600
42	36 or 37 or 38 or 39 or 40 or 41	57956
43	exp Cholesterol, HDL/	28500
44	HDL? Cholesterol?.mp	7915
45	High Density Lipoprotein Cholesterol?.mp	29251

46	alpha-Lipoprotein Cholesterol?.mp	0
47	alpha lipoprotein cholesterol?.mp	0
48	43 or 44 or 45 or 46 or 47	29879
49	Dyslipidemias/	5008
50	Dyslipidemia?.mp	17973
51	Dyslipoproteinemia?.mp.	198
52	49 or 50 or 51	18142
53	homeostatic model assessment of insulin resistance.mp.	620
54	HOMA.mp.	4597
55	HOMA-IR.mp.	3367
56	53 or 54 or 55	4883
57	postprandial glucose.mp	1467
58	PPG.mp	917
59	57 or 58	2274
60	exp Glycated Hemoglobin A/	31120
61	Glycated H?emoglobin?.mp.	4516
62	Glycosylated H?emoglobin?.mp	7185
63	Glycoh?emoglobin?.mp	264
64	HbA1*.mp.	14151
65	A1c.mp.	28393
66	Hb A1*.mp.	142
67	HGBA1C.mp.	73
68	Hb1c.mp	4
69	60 or 61 or 62 or 63 or 64 or 65 or 66 or 67 or 68	33551

70	exp Glucose Tolerance Test/	11818
71	Glucose Tolerance.mp	20495
72	OGTT.mp	2264
73	GTT.mp	297
74	70 or 71 or 72 or 73	21434
75	glucose test?.mp.	428
76	fasting glucose.mp	6342
77	fasting blood glucose.mp	3814
78	fasting plasma glucose.mp	4437
79	FPG.mp	1761
80	fasting blood sugar.mp	832
81	fasting plasma sugar.mp	8
82	fasting sugar.mp	9
83	FBS.mp	1437
84	75 or 76 or 77 or 78 or 79 or 80 or 81 or 82 or 83	16079
85	exp Prediabetic State/	7885
86	prediabet*.mp	2789
87	pre-diabet*.mp	1066
88	pre diabet*.mp	1066
89	borderline diabet*.mp.	21
90	chemical? diabet*.mp.	4
91	latent diabet*.mp.	9
92	T2P.mp	16
93	Impaired fasting glucose.mp	1610

94	Impaired fasting glyce?mia.mp	62
95	IFG.mp	1487
96	Impaired glucose tolerance?.mp.	8415
97	IGT.mp	1669
98	Impaired glucose metabolism.mp.	397
99	exp Glucose Intolerance/	3706
100	Glucose Intolerance?.mp.	4254
101	85 or 86 or 87 or 88 or 89 or 90 or 91 or 92 or 93 or 94 or 95 or 96 or 97 or 98 or 99 or 100	14567
102	Diagnosis/	206940
103	diagnos*.mp	840626
104	detect*.mp.	413413
105	accura*.mp	305542
106	exp "Sensitivity and Specificity"/	84435
107	sensitiv*.mp.	327437
108	specificit*.mp	137866
109	Receiver Operating Characteristic?.mp	40509
110	ROC.mp	18355
111	Predictive Value?.mp	62206
112	102 or 103 or 104 or 105 or 106 or 107 or 108 or 109 or 110 or 111	1475568
113	25 or 29 or 35 or 42 or 48 or 52 or 56 or 59 or 69	495330
114	69 or 74 or 84	59806
115	9 and 101 and 112 and 113 and 114	450
116	Limitation 2015 – 2020	154

ProQuest Dissertations & Theses Global

Search was conducted on 19th March 2020 at 6:50 pm– 9:00 pm (CET) in Title and Abstract

#	search string	results
1	AB,TI(child*)	230082
2	AB,TI(Adolescen*)	54032
3	AB,TI(kid?)	2498
4	AB,TI(youngster?)	1252
5	AB,TI(youth?)	36173
6	AB,TI(teen*)	8955
7	1 OR 2 OR 3 OR 4 OR 5 OR 6	289984
8	AB,TI(obesity)	14426
9	AB,TI(overweight)	5673
10	AB,TI("body weight")	10140
11	AB,TI("body mass index")	5465
12	AB,TI("Quetelet? Index")	44
13	AB,TI("Quetelet's Index")	10
14	AB,TI(BMI)	7796
15	AB,TI("Waist Hip Ratio?")	181
16	AB,TI("Waist-Hip Ratio?")	181
17	AB,TI("Waist to Hip Ratio?")	280
18	AB,TI("Waist-to-Hip Ratio?")	280
19	AB,TI(WHR)	301
20	8 OR 9 OR 10 OR 11 OR 12 OR 13 OR 14 OR 15 OR 16 OR 17 OR 18 OR 19	30969

21	AB,TI(hypertension)	10127
22	AB,TI("high blood pressure")	966
23	21 OR 22	10657
24	AB,TI(anhydroglucitol?)	11
25	AB,TI(1,5-AG)	36
26	AB,TI("anhydro-D-glucitol?")	4
27	AB,TI("Deoxy-D-glucopyranose?")	9
28	AB,TI("GlycoMark")	1
29	24 OR 25 OR 26 OR 27 OR 28	56
30	AB,TI(triglyceride?)	4633
31	AB,TI(triacylglycerol?)	1073
32	AB,TI(triacylglyceride?)	111
33	AB,TI(TG)	6049
34	AB,TI(TAG)	8979
35	30 OR 31 OR 32 OR 33 OR 34	19582
36	AB,TI("HDL? Cholesterol?")	904
37	AB,TI("High Density Lipoprotein Cholesterol?")	518
38	AB,TI("alpha-Lipoprotein Cholesterol?")	0
39	AB,TI("alpha lipoprotein cholesterol?")	0
40	36 OR 37 OR 38 OR 39	1390
41	AB,TI(Dyslipidemia?)	642
42	AB,TI(Dyslipoproteinemia?)	13
43	41 OR 42	655
44	AB,TI("homeostatic model assessment of insulin resistance")	20

45	AB,TI("HOMA")	388
46	AB,TI("HOMA-IR")	255
47	44 OR 45 OR 46	393
48	AB,TI("postprandial glucose")	120
49	AB,TI(PPG)	414
50	48 OR 49	531
51	(AB,TI("Glycated Hemoglobin?") OR AB,TI("Glycated Haemoglobin?"))	210
52	AB,TI("Glycosylated Hemoglobin?") OR AB,TI(Glycosylated Haemoglobin?)	358
53	AB,TI(Glycoh?emoglobin?)	40
54	AB,TI(HbA1*)	942
55	AB,TI(A1c)	484
56	AB,TI("Hb A1*")	8
57	AB,TI(HGBA1C)	41
58	AB,TI(Hb1c)	0
59	51 OR 52 OR 53 OR 54 OR 55 OR 56 OR 57 OR 58	1707
60	AB,TI("Glucose Tolerance")	1592
61	AB,TI(OGTT)	244
62	AB,TI(GTT)	95
63	60 OR 61 OR 62	1698
64	AB,TI("glucose test?")	40
65	AB,TI("fasting glucose")	490
66	AB,TI("fasting blood glucose")	400
67	AB,TI("fasting plasma glucose")	253
68	AB,TI(FPG)	223

69	AB,TI("fasting blood sugar")	50
70	AB,TI("fasting plasma sugar")	0
71	AB,TI("fasting sugar")	2
72	AB,TI(FBS)	677
73	64 OR 65 OR 66 OR 67 OR 68 OR 69 OR 70 OR 71 OR 72	1934
74	AB,TI(Prediabet*)	323
75	AB,TI("pre-diabet*")	240
76	AB,TI("pre diabet*")	240
77	AB,TI("borderline diabet*")	7
78	AB,TI("chemical? diabet*")	1
79	AB,TI("latent diabet*")	1
80	AB,TI(T2P)	4
81	AB,TI("Impaired fasting glucose")	59
82	AB,TI("Impaired fasting glyce?mia")	2
83	AB,TI(IFG)	126
84	AB,TI("Impaired glucose tolerance?")	370
85	AB,TI(IGT)	271
86	AB,TI("Impaired glucose metabolism")	39
87	AB,TI("Glucose Intolerance?")	431
88	74 OR 75 OR 76 OR 77 OR 78 OR 79 OR 80 OR 81 OR 82 OR 83 OR 84 OR 85 OR 86 OR 87	1545
89	AB,TI(Diagnos*)	110389
90	AB,TI(Detect*)	257870
91	AB,TI(Accura*)	235038
92	AB,TI(Sensitiv*)	193387

93	AB,TI(Specificit*)	40767
94	AB,TI("Receiver Operating Characteristic?")	1300
95	AB,TI(ROC)	2302
96	AB,TI("Predictive Value?")	3555
97	89 OR 90 OR 91 OR 92 OR 93 OR 94 OR 95 OR 96	693775
98	20 OR 23 OR 29 OR 35 OR 40 OR 43 OR 47 OR 50 OR 59	59153
99	59 OR 63 OR 73	4797
100	7 AND 88 AND 97 AND 98 AND 99	29
101	Limitation 2015 – 2020	9

Cochrane Library

Search was conducted on 19th March 2020 at 12:30 pm (CET) in Title, Abstract, Keywords

#	search string	results
1	MeSH descriptor: [Child] explode all trees	1238
2	MeSH descriptor: [Adolescent] explode all trees	100696
3	child*:ti,ab,kw	148002
4	Adolescen*:ti,ab,kw	130173
5	kid?:ti,ab,kw	1064
6	youngster?:ti,ab,kw	157
7	youth?:ti,ab,kw	6757
8	teen*:ti,ab,kw	2575
9	#1 or #2 or #3 or #4 or #5 or #6 or #7 or #8	234609
10	obesity:ti,ab,kw	34248
11	Overweight:ti,ab,kw	15697
12	body NEXT weight:ti,ab,kw	44733
13	MeSH descriptor: [Body Mass Index] explode all trees	9782
14	body NEXT mass NEXT index:ti,ab,kw	34928
15	Quetelet? NEXT Index:ti,ab,kw	55
16	Quetelet's NEXT Index:ti,ab,kw	50
17	BMI:ti,ab,kw	37016
18	MeSH descriptor: [Obesity] this term only	10938
19	MeSH descriptor: [Pediatric Obesity] this term only	1091
20	MeSH descriptor: [Waist-Hip Ratio] explode all trees	250
21	Waist NEXT Hip NEXT Ratio?:ti,ab,kw	1424

22	Waist-Hip NEXT Ratio?:ti,ab,kw	1424
23	Waist NEXT to NEXT Hip NEXT Ratio?:ti,ab,kw	839
24	Waist-to-Hip NEXT Ratio?:ti,ab,kw	839
25	WHR:ti,ab,kw	572
26	#10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21 or #22 or #23 or #24 or #25	105654
27	MeSH descriptor: [Hypertension] explode all trees	17585
28	hypertension:ti,ab,kw	58046
29	high NEXT blood NEXT pressure:ti,ab,kw	2380
30	#27 or #28 or #29	58894
31	anhydroglucitol?:ti,ab,kw	108
32	"1,5-AG":ti,ab,kw	110
33	anhydro-D-glucitol?:ti,ab,kw	14
34	Deoxy-D-glucopyranose?:ti,ab,kw	0
35	GlycoMark:ti,ab,kw	8
36	#31 or #32 or #33 or #34 or #35	208
37	MeSH descriptor: [Triglycerides] explode all trees	6217
38	triglyceride?:ti,ab,kw	21328
39	triacylglycerol?:ti,ab,kw	8181
40	triacylglyceride?:ti,ab,kw	64
41	TG:ti,ab,kw	6281
42	TAG:ti,ab,kw	977
43	#37 or #38 or #39 or #40 or #41 or #42	28233
44	MeSH descriptor: [Cholesterol, HDL] explode all trees	3649
45	HDL NEXT Cholesterol? OR HDLs NEXT Cholesterol?:ti,ab,kw	8120

46	High NEXT Density NEXT Lipoprotein NEXT Cholesterol?:ti,ab,kw	7489
47	alpha-Lipoprotein NEXT Cholesterol?:ti,ab,kw	2
48	alpha NEXT lipoprotein NEXT cholesterol?:ti,ab,kw	2
49	#44 or #45 or #46 or #47 or #48	13992
50	MeSH descriptor: [Dyslipidemias] this term only	1168
51	Dyslipidemia?:ti,ab,kw	5091
52	Dyslipoproteinemia?:ti,ab,kw	45
53	#50 or #51 or #52	5132
54	homeostatic NEXT model NEXT assessment NEXT of NEXT insulin NEXT resistance:ti,ab,kw	213
55	HOMA:ti,ab,kw	4503
56	"HOMA-IR":ti,ab,kw	3086
57	#54 or #55 or #56	4588
58	postprandial NEXT glucose:ti,ab,kw	2282
59	PPG:ti,ab,kw	733
60	#58 or #59	2763
61	MeSH descriptor: [Glycated Hemoglobin A] explode all trees	5487
62	Glycated NEXT Hemoglobin? OR Glycated NEXT haemoglobin?:ti,ab,kw	7992
63	Glycosylated NEXT H?emoglobin? OR Glycosylated NEXT haemoglobin?:ti,ab,kw	3395
64	Glycoh?emoglobin?:ti,ab,kw	113
65	HbA1*:ti,ab,kw	16894
66	A1c:ti,ab,kw	21484
67	Hb NEXT A1*:ti,ab,kw	16903
68	HGBA1C:ti,ab,kw	84
69	Hb1c:ti,ab,kw	8

70	#61 or #62 or #63 or #64 or #65 or #66 or #67 or #68 or #69	25947
71	MeSH descriptor: [Glucose Tolerance Test] explode all trees	1979
72	Glucose NEXT Tolerance:ti,ab,kw	8328
73	OGTT:ti,ab,kw	2177
74	GTT:ti,ab,kw	261
75	#71 or #72 or #73 or #74	8832
76	glucose NEXT test?:ti,ab,kw	241
77	fasting NEXT glucose:ti,ab,kw	4343
78	fasting NEXT blood NEXT glucose:ti,ab,kw	3431
79	fasting NEXT plasma NEXT glucose:ti,ab,kw	4821
80	FPG:ti,ab,kw	2626
81	fasting NEXT blood NEXT sugar:ti,ab,kw	1256
82	fasting NEXT plasma NEXT sugar:ti,ab,kw	10
83	fasting NEXT sugar:ti,ab,kw	20
84	FBS:ti,ab,kw	1131
85	#76 or #77 or #78 or #79 or #80 or #81 or #82 or #83 or #84	14658
86	MeSH descriptor: [Prediabetic State] explode all trees	912
87	Prediabet*:ti,ab,kw	2388
88	pre-diabet*:ti,ab,kw	831
89	pre NEXT diabet*:ti,ab,kw	831
90	borderline NEXT diabet*:ti,ab,kw	34
91	chemical NEXT diabet* OR chemicals NEXT diabet*:ti,ab,kw	22
92	latent NEXT diabet*:ti,ab,kw	7
93	T2P:ti,ab,kw	14

94	Impaired NEXT fasting NEXT glucose:ti,ab,kw	671
95	Impaired NEXT fasting NEXT glyce?mia:ti,ab,kw	30
96	IFG:ti,ab,kw	521
97	Impaired NEXT glucose NEXT tolerance?:ti,ab,kw	2825
98	IGT:ti,ab,kw	986
99	Impaired NEXT glucose NEXT metabolism:ti,ab,kw	147
100	MeSH descriptor: [Glucose Intolerance] explode all trees	1085
101	Glucose NEXT Intolerance?:ti,ab,kw	1830
102	#86 or #87 or #88 or #89 or #90 or #91 or #92 #93 or #94 or #95 or #96 or #97 or #98 or #99 or #100 or #101	5682
103	MeSH descriptor: [Diagnosis] this term only	61
104	Diagnos*:ti,ab,kw	214497
105	Detect*:ti,ab,kw	79491
106	Accura*:ti,ab,kw	30185
107	MeSH descriptor: [Sensitivity and Specificity] explode all trees	15075
108	Sensitiv*:ti,ab,kw	69330
109	Specificit*:ti,ab,kw	18862
110	Receiver NEXT Operating NEXT Characteristic?:ti,ab,kw	3411
111	ROC:ti,ab,kw	3057
112	Predictive NEXT Value?:ti,ab,kw	12862
113	#103 or #104 or #105 or #106 or #107 or #108 or #109 or #110 or #111 or #112	334624
114	#26 or #30 or #36 or #43 or #49 or #53 or #57 or #60 or #70	189446
115	#70 or #75 or #85	38455
116	#9 and #114 and #115 and #102 and #113	146
117	Limitation Apr 2015 – Mar 2020	94

Bibliographia medica Čechoslovaca

Search was conducted on 20th March 2020 at 9:35 am – 10:40 am (CET) in Kdekoliv

#	search string	results
1	Dítě (MeSH)	51078
2	Mladiství (MeSH)	23509
3	"Dítě*"	65228
4	"Dět*"	170882
5	"Adolescen*"	27433
6	"Mladistv*"	24753
7	"Dospívající*"	24217
8	"dospívání*"	23746
9	"Mládež*"	27398
10	"Teenage*"	23595
11	1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10	187393
12	"Obezit*"	8415
13	"Obesit*"	8041
14	"Obézn*"	1850
15	"Nadváh*"	844
16	"Otyl*"	6213
17	"Tělesn*" and "hmotnost*"	4499
18	"Index*" and tělesné and hmotnosti	1727
19	index tělesné hmotnosti (MeSH)	1642
20	"body mass index"	2134
21	"Quetelet*" and "Index*"	1645

22	"BMI"	2520
23	Obezita (MeSH)	6043
24	poměr pasu a boků (MeSH)	69
25	"Poměr*" and "pas*" and "bok*"	78
26	WHR	99
27	12 OR 13 OR 14 OR 15 OR 16 OR 17 OR 18 OR 19 OR 20 OR 21 OR 22 OR 23 OR 24 OR 25 OR 26	13884
28	Hypertenze (MeSH)	9829
29	"Hypertenz*"	16460
30	"Hypertens*"	17200
31	"Vysok*" and "krev*" and "tlak*"	10154
32	"Hyperton*"	10533
33	28 OR 29 OR 30 OR 31 OR 32	20728
34	"Anhydroglucitol*"	1
35	"anhydro-D-glucitol*"	0
36	"Deoxy-D-glucopyranose*"	0
37	"GlycoMark*"	2
38	34 OR 35 OR 36 OR 37	3
39	Triglyceridy (MeSH)	992
40	"Triglycerid*"	1730
41	"Triacylglycerol*"	1429
42	"Triacylglycerid*"	10
43	"TG"	372
44	"TAG"	231
45	39 OR 40 OR 41 OR 42 OR 43 OR 44 OR 45	2337

46	HDL-cholesterol (MeSH)	372
47	"Cholesterol*" and "HDL*"	1080
48	"HDL-cholesterol*"	824
49	"cholesterol-HDL*"	464
50	"High Density Lipoprotein Cholesterol*"	432
51	"alpha-Lipoprotein Cholesterol*"	372
52	"alpha lipoprotein cholesterol*"	372
53	46 OR 47 OR 48 OR 49 OR 50 OR 51 OR 52	1109
54	Dyslipidemie (MeSH)	1264
55	"Dyslipidemi*"	2135
56	"Dyslipoproteinemi*"	1468
57	"Hyperlipoproteinémi*"	1479
58	HLP	59
59	54 OR 55 OR 56 OR 57 OR 58	3575
60	"homeostatic model assessment of insulin resistance"	1
61	"HOMA"	114
62	"HOMA-IR"	58
63	60 OR 61 OR 62	114
64	postprandiální and "glykémi*"	1454
65	postprandiální and "gluko*"	866
66	64 OR 65	1571
67	hemoglobin A glykosylovaný (MeSH)	1055
68	"Glykovan*" and "hemoglobin*"	1228
69	"Glykosylovan*" and "hemoglobin*"	1075

70	"Glykozylovan*" and "hemoglobin*"	15
71	"Glykohemoglobin*"	13
72	"A1a*"	1064
73	"A1b*"	1061
74	"HbA1*"	1296
75	"A1c"	1075
76	"Hb A1*"	1057
77	"HGBA1C"	1
78	"Hb1c"	1
79	67 OR 68 OR 69 OR 70 OR 71 OR 72 OR 73 OR 74 OR 75 OR 76 OR 77 OR 78	1492
80	glukózový toleranční test (MeSH)	608
81	"Glukózov*" and "toleran*"	953
82	"Glukósov*" and "toleran*"	19
83	"OGTT"	647
84	"GTT"	32
85	80 OR 81 OR 82 OR 83 OR 84	1017
86	"Glukózov*" and "test*"	708
87	"Glukósov*" and "test*"	87
88	"Glykém*" and nalačno	637
89	"Glukóz*" and nalačno	447
90	"Glukos*" and nalačno	485
91	"Cukr*" and nalačno	593
92	FPG	25
93	"FBS"	23

94	86 OR 87 OR 88 OR 89 OR 90 OR 91 OR 92 OR 93	1701
95	Prediabetes (MeSH)	181
96	"Prediabet*"	319
97	"pre-diabet*"	862
98	"pre diabet*"	587
99	"Chemic*" and "diabet*"	480
100	"Latent*" and "diabet*"	73
101	"Hran*" and "diabet*"	119
102	"poru*" and "gluk*" and "toleranc*"	452
103	"zhorš*" and "gluk*" and "toleranc*"	25
104	PGT	6
105	"Poru*" and "glykem*" and nalačno	88
106	"Hran*" and "glykem*" and nalačno	15
107	"zhorš*" and "glykem*" and nalačno	23
108	HGL	1
109	"IFG"	16
110	"IGT"	27
111	"Glukos*" and "intoleranc*"	262
112	"Glukóz*" and "intoleranc*"	278
113	porucha glukózové tolerance (MeSH)	243
114	95 OR 96 OR 97 OR 98 OR 99 OR 100 OR 101 OR 102 OR 103 OR 104 OR 105 OR 106 OR 107 OR 108 OR 109 OR 110 OR 111 OR 112 OR 113	2028
115	Diagnóza (MeSH)	1529
116	"Diagnos*"	108832

117	"Diagnoz*"	43920
118	"Vyšetř*"	40751
119	"Detek*"	6126
120	"Přesn*"	2917
121	senzitivita a specificita (MeSH)	2160
122	"Sensitiv*"	13980
123	"Senzitiv*"	6813
124	"Specificit*"	6301
125	"Receiver Operating Characteristic"	290
126	"ROC"	2489
127	"Prediktiv*" and "hodnot*"	1993
128	115 OR 116 OR 117 OR 118 OR 119 OR 120 OR 121 OR 122 OR 123 OR 124 OR 125 OR 126 OR 127	152888
129	27 OR 33 OR 38 OR 45 OR 53 OR 59 OR 63 OR 66 OR 79	39269
130	79 OR 85 OR 94	3250
131	11 AND 114 AND 128 AND 129 AND 130	90
132	limitace 2015 – 2020	33

PEDro (Physiotherapy Evidence Database)

Search was conducted on 20th March 2020 at 2:50 –– 3:00 pm (CET).

#	search string	results
1	prediabetes OR "pre-diabetes" OR "pre diabetes" OR prediabetic OR "impaired fasting" OR IFG OR "Impaired glucose" OR IGT OR "Glucose Intolerance" OR "borderline diabetes" OR "chemical diabetes" OR "chemicals diabetes“ OR "latent diabetes" (Abstract/Title)	
2	paediatrics (Subdiscipline)	
3	1 AND 2	4
4	2015 – 2020	1

MedNar search strategy

Search was conducted on 20th March 2020 at 14:15 am (CET) in these sources All Annual Reviews, Centerwatch, ClinicalTrials.gov, , Drugs.com, Fierce Pharma , Mayo Clinic, Merck Manual, WebMD, American College of Physicians, American Diabetes Association, Journal of the American Medical Association, Eunice Kennedy Shriver National Institute of Child Health and Human Development, John E. Fogarty International Center for Advanced Study in the Health Sciences, , National Heart Lung and Blood Institute, National Inst. Of Aging, National Institute of Allergy, National Institute on Deafness and Other Communication Disorders, National Institute on Drug Abuse, National Library of Medicine, NIH MedlinePlus, Office of Dietary Supplements, Administration on Aging, CDC, Drug Information Portal, EPA Pesticide Factsheets, Fedstats, HealthFinder, PILOTS Database, Substance Abuse & Mental Health Services Administration, U.S. Department of Health and Human Services, U.S. Food and Drug Administration, World Health Organization

#	search string	results
1	child* OR adolescen* OR kid OR kids OR youngster* OR youth* OR teen* (Full Record)	
2	prediabet* OR "pre-diabet*" OR "pre diabet*" OR "impaired fasting" OR IFG OR "Impaired glucose" OR IGT OR "Glucose Intolerance" OR "borderline diabet*" OR "chemical* diabet*" OR "latent diabet*" (Title)	
4	1 AND 2	218

5	2015 – 2020	40

OpenGrey

Search was conducted on 20th March 2020 at 2:35 pm (CET).

#	search string	results
1	(prediabetes OR "pre-diabetes" OR "pre diabetes" OR prediabetic OR "impaired fasting" OR IFG OR "Impaired glucose" OR IGT OR "Glucose Intolerance" OR "borderline diabetes" OR "chemical diabetes" OR "chemicals diabetes“ OR "latent diabetes") AND (child* OR adolescen* OR kid OR kids OR youngster* OR youth* OR teen*)	7
2	2015 – 2020	0

Clinical trials.gov

Search was conducted on 20th March 2020 at 1:25 pm (CET).

#	search string	results
1	prediabetes (Condition or Disease; Also searched for Glucose Intolerance, Pre diabetics, and Impaired glucose tolerance.)	
2	Child - birth–17 (Age Group)	
3	1 AND 2	91
4	2015 – 2020 (First posted)	39

WHO International Clinical Trials Registry Platform (WHO ICTRP)

Search was conducted on 20th March 2020 at 3:20 pm (CET).

#	search string	results
1	prediabetes OR "pre-diabetes" OR "pre diabetes" OR prediabetic OR "impaired fasting" OR IFG OR "Impaired glucose" OR IGT OR "Glucose Intolerance" OR	69

	"bordeline diabetes" OR "chemical diabetes" OR "chemicals diabetes“ OR "latent diabetes" (Condition, without synonyms, in clinical trials in children, recruitment status: all)	
5	2015 – 2020	29

Current control trials (ISRCTN registry)

Search was conducted on 20th March 2020 at 1:30 -2:00 pm (CET).

#	search string	results
1	Prediabetes OR pre-diabetes OR "pre diabetes" OR prediabetic OR impaired fasting OR IFG OR "Impaired glucose" OR IGT OR "Glucose Intolerance" OR "borderline diabetes" OR "chemical diabetes" OR "chemicals diabetes“ OR "latent diabetes" (Condition)	
2	Age range: Child, Mixed, All, Not Specified	
3	1 AND 2	2
4	2015-2020	2

Appendix 9

Author	Atabek, Pirgon
Year of publication	2007
Inclusion criteria: i.e. presenting symptoms, results from previous tests	8-18 years old with BMI greater than or equal to the 95th percentile for age and gender
Exclusion criteria: i.e. presenting symptoms, results from previous tests	prior major illness, including type 1 or type 2 diabetes mellitus, took medications, or had a condition known to influence body composition, insulin action, or insulin secretion (e.g. glucocorticoid therapy, hypothyroidism, Cushing's disease).
Sample size	148 participants (86 girls, 62 boys)
Participant demographics (i.e. age, sex, spectrum of presenting symptoms, comorbidity, current treatments)	86 girls and 62 boys, mean age: 10.86 ± 3.08 years, mean body mass index [BMI]: 27.7 ± 4.2, all participants were in good health and had normal thyroid function.
Recruitment centres	Department of Paediatric Endocrinology Unit at Selcuk University Hospital in Konya, Turkey.
Study methodology (consecutive or random; retrospective or prospective)	Consecutive enrolment
Period that study was carried out (beginning and end date)	N/A
Index test description (including criteria for positive test)	FGIR (Fasting glucose to insulin ration - HOMA-IR was calculated as fasting insulin concentration (μυ/ml) χ fasting glucose concentration (mmol/l)/22.5; QUICKI was calculated as l/[(log fasting insulin concentration (μυ/ ml) + log fasting glucose concentration (mg/dl)]
Reference test description (including criteria for positive test)	An OGTT was conducted using a dose of 1.75 g glucose/kg body weight (to a maximum of 75 g). Venous blood samples were obtained at 0, 30, 60, 90 and 120 min to measure plasma glucose and insulin levels in the morning by venepuncture after an overnight fast. The OGTT was used - the following: 2-hour post-load glucose (2h.PG) <140 mg/dl (<7.8 mmol/1) = normal glucose tolerance; 2h.PG >140 mg/dl (>7.8 mmol/1) and <200 mg/dl (<11.1 mmol/1) = impaired glucose tolerance (IGT).
Geographical location of data collection	Konya, Turkey
Setting of data collection	Department of Pediatric Endocrinology Unit at Selcuk University Hospital in Konya, Turkey.
Persons executing and interpreting index tests (numbers, training, and expertise)	Not known
Persons executing and interpreting reference test	Not known
Index/reference time interval (and treatments carried out in between)	Not known
Distribution of severity of disease in those with target condition	N/A
Other diagnoses in those without target condition	No other diagnosis was reported.
Adverse events from index test	N/A
Adverse events from reference test	N/A

Author:	Brar, Mengwall, Franklin, Fierman
Year of publication	2014
Inclusion criteria: i.e. presenting symptoms, results from previous tests	Patients with a suspicion of diabetes, and/or related morbidities such as abnormal values of glucose, insulin, HbA1c, polycystic ovary syndrome, dyslipidaemia, hypertension, acanthosis nigricans, and metabolic syndrome and who had both OGTT and HbAlc tests performed within 3 months of one another.
Exclusion criteria: i.e. presenting symptoms, results from previous tests	Not specified.
Sample size	149
Participant demographics (i.e. age, sex, spectrum of presenting symptoms, comorbidity, current treatments)	149 obese patients: normal (n = 125), prediabetes (n = 2l), diabetes (n = 3). The majority of the patients (71.1%) were Hispanic, and 62.40% of the patients were female. **For normal** (n=125), age (years) 13,8+/-3,1; Sex (%) M/F 38,4/61,6; Race/Ethnicity: H/W/B/A/O 74/2/8/7/9; BMI Z score 2,3 +/-0,5; Fasting glucose (mg/dL) 85,4+/-7; 2-hour glucose (mg/dL) 98,0+/-20; fasting insulin 16,8+/-12; 2-hour insulin 74,1+/-61; HbA1c 5,6+/-0,3; HOMA-IR 3,6+/-2,6/ **For prediabetes** (n=21): age (years) 13,0+/-3,7; Sex (%) M/F 33,3/66,7; Race/Ethnicity: H/W/B/A/O 57/5/5/14/19; BMI Z score 2,1 +/-0,7; Fasting glucose (mg/dL) 100,4+/-10; 2-hour glucose (mg/dL) 131,3+/-29; fasting insulin 23,2,+/-17; 2-hour insulin 127,1+/-108; HbA1c 5,9+/-0,5; HOMA-IR 5,8+/-4,5, **For Diabetes** (n=3): age (years) 13,5+/-0,1; Sex (%) M/F 33,3/66,7; Race/Ethnicity: H/W/B/A/O 33/0/0/67/0; BMI Z score 2,1 +/-0,5; Fasting glucose (mg/dL) 143,3+/-60; 2-hour glucose (mg/dL) 266,3+/-84; fasting insulin 51,5+/-40; 2-hour insulin 290+/-1; HbA1c 7,3+/-0,9; HOMA-IR 13,1+/-8,8
Recruitment centres	Bellevue Hospital NYC USA
Study methodology (consecutive or random; retrospective or prospective)	a retrospective chart review of patients (endocrine clinic 2005-2010)
Period that study was carried out (beginning and end date)	2005-2010
Index test description (including criteria for positive test)	HbAlc assays were done between 2005 and 2010 using borate affinity chromatography (Belleveu Hospital) and then by immune turbidirretric calorimetry (Quest Diagnostics, Teterboro, NJ). Both these methods met the NGSP (National Glycohemoglobin Standardization Pro gram) certification. HOMA-IR, a validated measure of insulin sensitivity,r3 was calculated using the following values for fasting glucose and insulin: HOMA-IR = fasting plasma insulin (FPI; in plU/mo1) x FPG (in mmollL)1122.5. BMI was calculated as weight (in kilograms) divided by height (in meters) squared. BMI percentiles and Z scores were obtained using age and gender-specific reference data

Reference test description (including criteria for positive test)	The OGTT outcome was considered positive if fasting glucose was >100 mg/dl aniVor 2-hour glucose was >140.
Geographical location of data collection	Bellevue Hospital NYC USA
Setting of data collection	Endocrine clinic at Bellevue Hospital
Persons executing and interpreting index tests (numbers, training, and expertise)	Not known
Persons executing and interpreting reference test	Not known
Index/reference time interval (and treatments carried out in between)	HbA1c and OGTT were measured on the same day in 55% of patients and within 1 month in 75%.
Distribution of severity of disease in those with target condition	N/A
Other diagnoses in those without target condition	No other diagnosis was reported.
Adverse events from index test	N/A
Adverse events from reference test	N/A

Author	Bridges, Thorpe, Baus, Cochran
Year of publication	2016
Inclusion criteria: i.e. presenting symptoms, results from previous tests	Children - BMI at or above the 85th percentile and lower than the 95th percentile, and above 95th percentile
Exclusion criteria: i.e. presenting symptoms, results from previous tests	No exclusion criteria
Sample size	223 (124 female, 99 male); The average age of the population was 13.4 years (range, 10–17).
Participant demographics (i.e. age, sex, spectrum of presenting symptoms, comorbidity, current treatments)	Female (n= 124): age, years: 13,40+/-2.26; BMI percentile 95,94+/-6,01; glucose (mg/dL) 90,11+/-8,90; HOMA-IR 5,28+/-3,58; Insulin 23,48+/-15,04; LDL-C (mg/dL) 89,50+/-23,55; HDL-C (mg/dL) 44,42+/-10,44; TRG (mg/dL) 101,06+/-53,16; Total C (mg/dL) 154,07+/-29,59; TRG/HDL ratio 2,50+/-1,84/ Male (n=99): age, years: 13,39+/-2.08; BMI percentile 96,20+/-5,71; glucose (mg/dL) 95,75+/-8,01; HOMA-IR 4,94+/-3,79; Insulin 20,63+/-14,70; LDL-C (mg/dL) 89,99+/-24,51; HDL-C (mg/dL) 39,77+/-7,93; TRG (mg/dL) 117,71+/-72,19; Total C (mg/dL) 153,28+/-28,16; TRG/HDL ratio 3,30+/-1,90
Recruitment centres	N/A
Study methodology (consecutive or random; retrospective or prospective)	random sampling, paediatric electronic medical records (data collected from a chart review using a standardized data collection tool)
Period that study was carried out (beginning and end date)	a two-year period (2012– 2014)
Index test description (including criteria for positive test)	HOMA was defined as (fasting insulin [μIU/mL] x fasting glucose [mg/dL])/405.
Reference test description (including criteria for positive test)	Hyperinsulinemia was defined as a fasting insulin level > 25 μIU/mL and impaired fasting glucose was defined as a fasting blood glucose level between 100 and 125 mg/dL
Geographical location of data collection	Lewisburg, West Virginia, USA
Setting of data collection	Robert C. Byrd Clinic in Lewisburg, WV, USA
Persons executing and interpreting index tests (numbers, training, and expertise)	Not known
Persons executing and interpreting reference test	Not known
Index/reference time interval (and treatments carried out in between)	N/A
Distribution of severity of disease in those with target condition	N/A
Other diagnoses in those without target condition	No other diagnosis was reported.
Adverse events from index test	N/A
Adverse events from reference test	N/A

Author	Ehehalt, Wiegand, Korner, Schweizer, et al
Year of publication	2017
Inclusion criteria: i.e. presenting symptoms, results from previous tests	(1) overweight, obese, and extremely obese children and adolescents aged 7 to 17 years; and (2) oral glucose tolerance test and HbA1c measurement on the same day.
Exclusion criteria: i.e. presenting symptoms, results from previous tests	(1) previously known diabetes; (2) tumour or severe systemic disease; (3) blood transfusion and significant blood loss within the last year; (4) hematologic diseases, hemoglobinopathies, renal insufficiency, chronic lead poisoning, and galactosemia; (5) syndromes associated with obesity; and (6) drugs affecting glucose and HbA1c levels, respectively.
Sample size	4848
Participant demographics (i.e. age, sex, spectrum of presenting symptoms, comorbidity, current treatments)	children and adolescents (2668 girls) with a mean BMI of 30.6 ± 5.4 kg/m2 (BMI-SDS 2.8 ± 0.6). Mean age was 13.1 ± 2.4 years. Within the study group, 15.7% (n = 759) were overweight, 46.4% (n =2251) were obese, and 37.9% (n = 1838) were extremely obese.
Recruitment centres	Vestische Hospital for Children and Adolescents Datteln, University of Witten/Herdecke, Germany (n = 2934); University Hospital for Children and Adolescents, University of Leipzig, Germany (n = 889); Endokrinologikum Berlin, Germany (n = 415); University Hospital for Children and Adolescents, Charité University Medicine, Berlin, Germany (n = 212); University Hospital for Children and Adolescents, University of Tübingen, Germany (n = 208); Endokrinologikum, Hamburg, Germany (n = 190)].
Study methodology (consecutive or random; retrospective or prospective)	an observational multicentre analysis
Period that study was carried out (beginning and end date)	N/A
Index test description (including criteria for positive test)	HbA1c levels were measured by using immunoassay and high-performance liquid chromatography (HPLC) methods, respectively, which are certified and standardized to the DCCT assay. It was used HbA1c ≥39 mmol/mol (≥5.7%) as cut-off level, too.
Reference test description (including criteria for positive test)	Blood samples taken for fasting glucose and glucose at 120 min (depending on the clinical or scientific question together with insulin at 0, 30, 60, 90, 120 min).
Geographical location of data collection	Germany
Setting of data collection	Vestische Hospital for Children and Adolescents Datteln, University of Witten/Herdecke, University Hospital for Children and Adolescents, University of Leipzig, Endokrinologikum Berlin, University Hospital for Children and Adolescents, Charité University Medicine, University Hospital for Children and Adolescents, University of Tübingen, Endokrinologikum, Hamburg
Persons executing and interpreting index tests (numbers, training, and expertise)	Not known
Persons executing and interpreting reference test	Not known

Index/reference time interval (and treatments carried out in between)	N/A
Distribution of severity of disease in those with target condition	N/A
Other diagnoses in those without target condition	No other diagnosis was reported.
Adverse events from index test	N/A
Adverse events from reference test	N/A

Author	Galhardo, Shield
Year of publication	2015
Inclusion criteria: i.e. presenting symptoms, results from previous tests	Recommended criteria were recruited for DM2 screening; body mass index z-score 3.35 ± 0.59.
Exclusion criteria: i.e. presenting symptoms, results from previous tests	Pregnancy; known intolerance to glucose or diabetes; chronic medication, namely hypoglycaemic; haemoglobinopathy or other condition associated to any change in erythrocyte survival.
Sample size	266 (range: 8.9 to 17.6 years of age)
Participant demographics (i.e. age, sex, spectrum of presenting symptoms, comorbidity, current treatments)	266 patients with 12.3 median age (range: 8.9 to 17.6 years of age) were assessed, from which 147 (55.3%) were female. Regarding patient's ethnicity, 240 (90.2%) were Caucasian, 22 (8.3%) Black and the remaining were of a mixed ethnicity: As regards family history, 215 (80.8%) patients had obese members in the family and 74 (27.8%) had 1st or 2nd -degree family members with DM2. The mothers of 15 patients (5.6%) developed gestational diabetes and 11 (4.1%) patients were born small for their gestational age. BMI average z-scores were 3.35 ± 0.59 while average z-scores of body fat percentage were 2.84 ± 0.61. A 36.51 ± 8.06% average central fat percentage was found. According to Tanner's classification, 106 (39.9%) patients were classified as pre-pubertal, 108 (40.6%) as pubertal and the remaining as post-pubertal.
Recruitment centres	A tertiary level British Paediatric Hospital
Study methodology (consecutive or random; retrospective or prospective)	A cross-sectional study
Period that study was carried out (beginning and end date)	first semester 2012
Index test description (including criteria for positive test)	Pre-diabetes was defined with the use of HbA1c level as diagnostic test at levels between 5.7% and 6.4% and diabetes at levels ≥ 6.5%. HbA1c levels were obtained through a NGSP assay, using human antiHbA1c monoclonal antibody.
Reference test description (including criteria for positive test)	OGTT test was adopted as the gold standard: prediabetes was diagnosed with a 2-hour blood glucose level between 7.8 mmol/L (140 mg/dL) and 11.0 mmol/L (199 mg/dL), while DM2 was defined with values ≥ 11.1 mmol/L (200 mg/dL). Pre-diabetes was defined with the use of HbA1c level as diagnostic test at levels between 5.7% and 6.4% and diabetes at levels ≥ 6.5%. As regards fasting blood glucose, pre-diabetes was established at levels of 5.6 mmol/L (100 mg/dL) to 6.9 mmol/L (125 mg/dL) and diabetes at levels ≥ 7.0 mmol/L (126 mg/dL).13 The blood glucose area under the curve (AUC) was calculated using the trapezoidal rule: AUC = 0.25 × [fasting blood glucose + 2 × (blood glucose at 30 min) + 2 × (blood glucose at 60 min) + 2 × (blood glucose at 90 min) + blood glucose at 120 min]. The insulin resistance index was calculated as HOMA-IR = [fasting insulin level (μIU/mL) x fasting blood glucose (mg/ dL)] / 405 and was

	considered as significant when ≥ 4.5. The TG: HDL-C ratio was considered elevated when ≥ 3.0.
Geographical location of data collection	Bristol, UK
Setting of data collection	Paediatric Hospital
Persons executing and interpreting index tests (numbers, training, and expertise)	Not known
Persons executing and interpreting reference test	Not known
Index/reference time interval (and treatments carried out in between)	N/A
Distribution of severity of disease in those with target condition	N/A
Other diagnoses in those without target condition	No other diagnosis was reported.
Adverse events from index test	N/A
Adverse events from reference test	N/A

Author	García, Trevino, Schanchez, Aguilar
Year of publication	2019
Inclusion criteria: i.e. presenting symptoms, results from previous tests	5 and 9 years old; According to the percentile tables of the CDC corresponding to BMI and age, two groups were constituted: group with obesity-overweight (OO Group): 85th percentile (n = 97) and group with normal weight (NW Group): <percentile 85 (n = 104)
Exclusion criteria: i.e. presenting symptoms, results from previous tests	Children with diabetes, hypertension, hypothyroidism or chronic illnesses
Sample size	201
Participant demographics (i.e. age, sex, spectrum of presenting symptoms, comorbidity, current treatments)	The median age of children was 8 years (range 5-9), and 42.78% were male.
Recruitment centres	Family Medicine Unit (FMU) No. 80 of IMSS of Morelia, Michoacán, Mexico
Study methodology (consecutive or random; retrospective or prospective)	a prospective, comparative cross-sectional study
Period that study was carried out (beginning and end date)	1 March 2016 to 28 February 2017
Index test description (including criteria for positive test)	TyG was calculated as Ln[fasting triglycerides in mg/dL x fasting glucose in mg/dL]/2. TG/HDL was calculated with fasting triglycerides/fasting HDL. The cut off point of TyG [TyG = Ln(99.9 x 99.9/2)] = 8.5 and TG/HDL [TG/ HDL = 99.9/44.9 = 2.22]
Reference test description (including criteria for positive test)	HOMA-IR = [Fasting glucose (mmol/L) x fasting insulin (μU/mL)]/22.5 and the cut point used was 8.23 that corresponding to ≥90th percentile.
Geographical location of data collection	Mexico
Setting of data collection	Family Medicine Unit
Persons executing and interpreting index tests (numbers, training, and expertise)	Not known
Persons executing and interpreting reference test	Not known
Index/reference time interval (and treatments carried out in between)	N/A
Distribution of severity of disease in those with target condition	N/A
Other diagnoses in those without target condition	No other diagnosis was reported.
Adverse events from index test	N/A
Adverse events from reference test	N/A

Author	Chan, Pyle, Kelsey, Newnes, Zeitler et al
Year of publication	2016
Inclusion criteria: i.e. presenting symptoms, results from previous tests	Eligible participants included youth 10–18 yr of age with a body mass index (BMI) ≥85th‰
Exclusion criteria: i.e. presenting symptoms, results from previous tests	Exclusion criteria included medications known to affect glycemia (insulin, other diabetes medications, atypical antipsychotics, glucocorticoids), BMI <85th‰, anaemia, hemoglobinopathy, chronic illness likely to affect red cell life span, pregnancy, and outside HbA1c >7.5% (i.e. requiring immediate diabetes treatment),
Sample size	117
Participant demographics (i.e. age, sex, spectrum of presenting symptoms, comorbidity, current treatments)	Participants were a median of 14.1 yr of age (range: 10–18 yr), median BMI z-score of 2.3 (range: 1.1–3.0), 62% female, 59% Hispanic, 22% White, and 17% Black. Approximately half of the participants were dysglycemic based on either 2hG ≥140mgdL−1 (40.2%) or HbA1c ≥5.7% (51.3%), whereas only 9% were dysglycemic by FPG ≥100mgdL−1. Median (min.–max.) values for all glycemic measures were as follows: FPG 86mgdL−1 (87–130mgdL−1), 2hG 131mgdL−1 (81–239mgdL−1), HbA1c 5.7% (4.9–7.7%), FA 209 μmolL−1 (169–270 μmolL−1), GA 11% (9–17%), and 1,5-AG 24.1mcgmL−1 (2.6–41mcgmL−1)
Recruitment centres	Primary care, weight management, and endocrine clinics in Denver, Colorado
Study methodology (consecutive or random; retrospective or prospective)	Consecutive
Period that study was carried out (beginning and end date)	N/A
Index test description (including criteria for positive test)	Alternate nonfasting glycemic markers – 1,5-anhydroglucitol (1,5AG), fructosamine (FA), and glycated albumin (GA)
Reference test description (including criteria for positive test)	OGTT, HbAc
Geographical location of data collection	Denver, Colorado, USA
Setting of data collection	Primary care, weight management, and endocrine clinics
Persons executing and interpreting index tests (numbers, training, and expertise)	Not known
Persons executing and interpreting reference test	Not known
Index/reference time interval (and treatments carried out in between)	N/A
Distribution of severity of disease in those with target condition	N/A
Other diagnoses in those without target condition	No other diagnosis was reported.
Adverse events from index test	N/A
Adverse events from reference test	N/A

Author	Chan, Pyle, Newnes, Nadeau, et al
Year of publication	2015
Inclusion criteria: i.e. presenting symptoms, results from previous tests	Males and females 10–18 years of age with a body mass index (BMI) in the 85th percentile or greater
Exclusion criteria: i.e. presenting symptoms, results from previous tests	Exclusion criteria included HbA1c greater than 7.5% because these individuals required immediate initiation of glucose-lowering therapy, medications known to affect blood glucose, known anemia, hemoglobinopathy, chronic illness likely to affect red cell life span, and pregnancy.
Sample size	118 (Eight were excluded due to missing FPG, HbA1c, or 2-hour glucose. Another 12 patients were excluded due to incomplete CGM data, leaving 98 participants with 48 hours of CGM data)
Participant demographics (i.e. age, sex, spectrum of presenting symptoms, comorbidity, current treatments)	N=98; age: 14.1 (10.0-18.0), M, N (%): 35 (35,7); Ethnicity, N (%): Non-hispanic white 24 (24.5), Black: 11 (11.2), Hispanic: 61 (62.2), Other 2 (2.0); Family history of diabetes: No, N (%) 19 (19.4), Yes, N (%) 78 (79.6); Weight, kg: 87.7 (41.2-171.4); height, cm: 161.2 (140.0-189.0); BMI, kg/m2: 32.5 (21.0-55.5); BMI z score: 2.3 (1.1-3.0); Waist circumference, cm: 101.0 (39.5-158.9); Hip circumference, cm: 108.2 (77.0-167.5); Waist-to hip ratio: 0.9 (0.4-1.7); SBP, mm Hg: 120.5 (95-161); SBP percentile: 87.0 (17.4-100.0); DBP, mm Hg: 69.5 (48-100); DBP percentile: 65.2 (10.2-99.7); Tanner stage, N %: I. 6 (6.1), II. 11 (11.2), III. 13 (13.3), IV. 14 (14.3), V. 54 (55.1); ALT, U/L_ 30.0 (6.0-182.0); AST, U/L 38.0 (16.0-116.0); Fasting plasma glucose, mg/dL: 86 (65-130); 2-h OGTT, mg/dL: 131 (84-289); Total cholesterol, mg/dL: 166 (81-292); TG, mg/dL: 137 (31-596); LDL, mg/dL: 99 (25-218); HDL, mg/dL: 37 (24-68); non-HDL, mg/dL: 128 (42-242); TG/HDL: 3.6 (0.7-20.1)
Recruitment centres	weight management and endocrine clinics at Children's Hospital, and primary care, school-based, and community health clinics
Study methodology (consecutive or random; retrospective or prospective)	a cross-sectional study
Period that study was carried out (beginning and end date)	N/A
Index test description (including criteria for positive test)	Descriptive statistics for demographic and clinical variables were calculated by category of HbA1c (<5.7%, 5.7–6.4%, >6.4%). The cut off value for HbA1c and 2-hour glucose that maximized sensitivity and specificity for identifying abnormal CGM AUC was determined. 2-hour glucose (<140 mg/dL, 140–199 mg/ dL, ≥200 mg/dL). The cut off value for HbA1c and 2-hour glucose that maximized sensitivity and specificity for identifying abnormal CGM AUC was determined
Reference test description (including criteria for positive test)	N/A
Geographical location of data collection	Denver, Colorado, USA

Setting of data collection	weight management and endocrine clinics and hospital, and primary care, school-based, and community health clinics
Persons executing and interpreting index tests (numbers, training, and expertise)	Not known
Persons executing and interpreting reference test	Not known
Index/reference time interval (and treatments carried out in between)	N/A
Distribution of severity of disease in those with target condition	N/A
Other diagnoses in those without target condition	No other diagnosis was reported.
Adverse events from index test	N/A
Adverse events from reference test	N/A

Author	Kang, Yang, Lee, Yang, Kim, Lim et al
Year of publication	2017
Inclusion criteria: i.e. presenting symptoms, results from previous tests	Not described
Exclusion criteria: i.e. presenting symptoms, results from previous tests	Not described
Sample size	231
Participant demographics (i.e. age, sex, spectrum of presenting symptoms, comorbidity, current treatments)	Non-diabetic subjects aged 9–13 years from one middle and two elementary schools; 168 males and 53 females with a mean age of 11.1 ± 1.5 years, their BMI classified 16
Recruitment centres	one middle and two elementary schools
Study methodology (consecutive or random; retrospective or prospective)	a cross-sectional study
Period that study was carried out (beginning and end date)	May to June 2014
Index test description (including criteria for positive test)	The TyG index was calculated as the natural logarithm (ln) of the product of plasma glucose and TG using the formula: ln (TG [mg dl− 1] × fasting glucose [mg dl − 1]/2).
Reference test description (including criteria for positive test)	The HOMA-IR index was calculated with the formula: fasting insulin (U ml− 1) × fasting glucose (mmol l − 1)/22.5. Insulin resistance in this study was defined as the value equal to or greater than the 95th percentile of the age- and sex-specific HOMA-IR of Korean adolescents.
Geographical location of data collection	Chung-ju city, North Chungcheong Province in South Korea
Setting of data collection	middle and elementary schools
Persons executing and interpreting index tests (numbers, training, and expertise)	Not known
Persons executing and interpreting reference test	Not known
Index/reference time interval (and treatments carried out in between)	Blood sampling for biochemical assays were drawn in the morning after 10–12 h of overnight fasting from an antecubital vein.
Distribution of severity of disease in those with target condition	N/A
Other diagnoses in those without target condition	No other diagnosis was reported.
Adverse events from index test	N/A
Adverse events from reference test	N/A

Author	Kasturi, Onuzuruike, Kunnam, Shomaker, Yanovski, Chung
Year of publication	2016
Inclusion criteria: i.e. presenting symptoms, results from previous tests	Adolescent girls who had overweight/obesity (BMI≥85th percentile). Youth with a first-or-second degree relative with type 2 diabetes and mild or moderate depressive symptoms.
Exclusion criteria: i.e. presenting symptoms, results from previous tests	Major depressive disorder (MDD), and psychiatric disorders
Sample size	93
Participant demographics (i.e. age, sex, spectrum of presenting symptoms, comorbidity, current treatments)	Youth females with a first or second degree relative with type 2 diabetes and mild or moderate depressive symptoms. Youth females had overweight/obesity; age 14.8 ± 1.6 years, range: 12-17 years) who had overweight/obesity (body mass index [BMI] ≥ 85th percentile. Youth females were black, white, Asian and mixed race and at the baseline with: BMI (kg/m2) 32,6 +/-6,5 Systolic BP (mmHg) 118,5+/-9,5 Diastolic BP (mmHg) 64,6+/-8,3 Fasting glucose (mg/dL) 89,0+/-7,5 2-hr glucose (mg/dL) 103,3+/-21,4 Fasting insulin (µU/mL)# 20,3 (13,8-27,7) Prediabetes, n (%) 12 (13) HbA1c (%) 5,3+/-0,4 Insulinogenic index 4,2 (2,5-6,6) Matsuda index 2,3 (1,5-3,2)
Recruitment centres	N/A
Study methodology (consecutive or random; retrospective or prospective)	a secondary analysis of a randomized controlled trial of cognitive behavioral therapy (CBT) vs health education in adolescent girls
Period that study was carried out (beginning and end date)	Baseline accuracy and reproducibility OGTT, 6-wk OGGT, 1 y follow up (September 2011 – July 2014)
Index test description (including criteria for positive test)	???
Reference test description (including criteria for positive test)	OGTT fasting and 2-hr glucose criteria, was designated the reference variable or gold-standard at each time point (baseline, 6-weeks and 1-year). Using standard OGTT criteria, prediabetes was defined as fasting glucose 100 mg/dL and <126 mg/dL, insulinogenic index (n=6 at screening and 6-week and n=3 at 1-year follow-up), Matsuda index (n=10 at screening, n=18 at 6-weeks and n=5 at 1-year follow-up) Simply 1-hr OGTT
Geographical location of data collection	USA
Setting of data collection	NIH Hatfield Clinical Research Center
Persons executing and interpreting index tests (numbers, training, and expertise)	Not known
Persons executing and interpreting reference test	Not known
Index/reference time interval (and treatments carried out in between)	Baseline, 6 weeks and 1 year follow up. Both tests were measured at the same time.
Distribution of severity of disease in those with target condition	N/A
Other diagnoses in those without target condition	No other diagnosis was reported.
Adverse events from index test	N/A
Adverse events from reference test	N/A

Author	Keskin, Kurtogğu, Kendirci, Atabek, Yazici
Year of publication	2005
Inclusion criteria: i.e. presenting symptoms, results from previous tests	All subjects were healthy and had normal thyroid function. All subjects had a BMI above the 95th percentile for age and gender and thus were classified as obese.
Exclusion criteria: i.e. presenting symptoms, results from previous tests	Not described
Sample size	57 (30 girls and 27 boys)
Participant demographics (i.e. age, sex, spectrum of presenting symptoms, comorbidity, current treatments)	mean age: 12.04 +/- 2.90 years; mean BMI: 29.57 +/- 5.53), **Obese subjects with IR** (n=25); age (y):12.88 +/- 2.88; Gender, M/F: 11/14; BMI, kg/m2: 31.29 +/- 5.86; Fasting glucose level, mg/dL: 82.67 +/- 9.23 (65-106), Fasting insulin level uU/mL: 26.98 +/- 22.49 (1.45–109.72); Sum of insulin levels: uU/mL 447.32 +/- 145.22 (300.24–744.39); **Obese subject without IR** (n= 32): age (y): 11.38 +/- 2.79; Gender, M/F: 16/16; BMI, kg/m2: 28,23 +/- 4.94; Fasting glucose level, mg/dL: 80.44 +/- 10,51 (61-105), Fasting insulin level uU/mL: 16,65 +/- 13,85 (1.40–51,47); Sum of insulin levels: uU/mL 154,08 +/- 77,78 (24,86–275,00)
Recruitment centres	Department of Pediatric Endocrinology of Erciyes University, Faculty of Medicine
Study methodology (consecutive or random; retrospective or prospective)	Consecutive enrolment
Period that study was carried out (beginning and end date)	Not known
Index test description (including criteria for positive test)	The HOMA index, QUICKI, and FGIR were derived as estimates of insulin resistance. The HOMA index was calculated as fasting insulin concentration (□U/mL) /// fasting glucose concentration (mmol/L)/22.5, assuming that normal young subjects have an insulin resistance of 1. The QUICKI was calculated as 1/[log fasting insulin concentration (□U/mL) ⩽ log glucose concentration (mg/dL)].
Reference test description (including criteria for positive test)	After a 3-day, performance of high-carbohydrate diet (300 g/day) and an overnight fast, a standard OGTT (1.75 g/kg or a maximum of 75 g of glucose). Blood samples were obtained 0, 30, 60, 90, and 120 minutes after glucose administration, for glucose and insulin measurements. Plasma glucose levels were measured with the glucose oxidase method and a modified Trinder colour reaction, catalysed by the peroxidase enzyme, and insulin levels were measured with an immunoradiometric assay kit.
Geographical location of data collection	Kayseri, Turkey
Setting of data collection	Department of Pediatric Endocrinology of Erciyes University Faculty of Medicine
Persons executing and interpreting index tests (numbers, training, and expertise)	Not known
Persons executing and interpreting reference test	Not known
Index/reference time interval (and treatments carried out in between)	Blood samples were obtained 0, 30, 60, 90, and 120 minutes after glucose administration, for glucose and insulin measurements (reference test)

Distribution of severity of disease in those with target condition	N/A
Other diagnoses in those without target condition	No other diagnosis was reported.
Adverse events from index test	N/A
Adverse events from reference test	N/A

Author(s):	Kim, Jo, Lee
Year of publication	2018
Inclusion criteria: i.e. presenting symptoms, results from previous tests	Not specified
Exclusion criteria: i.e. presenting symptoms, results from previous tests	Subjects were excluded if they were known to have diabetes or renal glucosuria prior to the study and if their hemoglobin levels were less than 10 g/dL.
Sample size	190 children
Participant demographics (i.e. age, sex, spectrum of presenting symptoms, comorbidity, current treatments)	N=190; age (yrs): 12.56+/-3.44; Sex (%): F 99 (52.1), M 91 (47.9); BMI (kg/m2): 24.50+/-5.12; Obesity (%): 86 (45.3); Family history of DM (%): 52 (39.1); Hb (g/dL): 14.09+/-1.17; HbA1c (%): <5.7% (NGT): 117 (61.6), 5.7-6.4% (at the risk for DM): 41 (21.6), ≥6.5% (DM): 42 (22.1); FPG (mg/dL): 110.74+/-52.92; 2-h OGTT (mg/dL): 177.91+/-127.24; Serum c-peptide (ng/mL): 2.67+/-1.54; HOMA-IR: 4.96+/-5.76; Cholesterol (mg/dL): 170.07+/-34.67; HDL (mg/dL): 47.36+/-12.09; LDL (mg/dL): 100.26+/-33.29; TG (mg/dL): 177.84+/-67.24
);R.ecruitment centres	+/-Chonbuk .National University Children's ; HOMA-IR:. .Hospital
Study methodology (consecutive or random; retrospective or prospective)	Consecutive - children attending Chonbuk National University Children's Hospital for an OGTT to confirm the diagnosis of diabetes.
Period that study was carried out (beginning and end date)	2010-2017
Index test description (including criteria for positive test)	HbA1c was measured via high-performance liquid chromatography. Serum insulin and c-peptide were measured by immunoradiometric assay (IRIMA) using commercial kits (DIAsource ImmunoAssay S.A., Belgium for insulin; Institute of Isotopes Co., Ltd., Budapest, Hungary for c-peptide). Insulin sensitivity was estimated using the previously validated homeostasis model assessment of insulin resistance (HOMA-IR) index
Reference test description (including criteria for positive test)	A 2-hour plasma glucose level following OGTT (2-h OGTT) > 200 mg/dL (11.1 mmol/L) or a fasting plasma glucose (FPG) ≥ 126 mg/dL (7.0 mmol/L), impaired glucose tolerance (IGT) as a 2-h OGTT 140-199 mg/dL (7.8-11.0 mmol/L), and normal glucose tolerance (NGT) as a 2-h OGTT < 140 mg/dL (7.8 mmol/L) according to the WHO criteria based on OGTT.
Geographical location of data collection	Seoul, Republic of Korea
Setting of data collection	National University Children's Hospital
Persons executing and interpreting index tests (numbers, training, and expertise)	Not known
Persons executing and interpreting reference test	Not known
Index/reference time interval (and treatments carried out in between)	
Distribution of severity of disease in those with target condition	N/A
Other diagnoses in those without target condition	No other diagnosis was reported.
Adverse events from index test	N/A
Adverse events from reference test	N/A

Author:	Kurtoğlu, Hatipoglu, Mazicioglu, Kendirci, Keskin, Kondolot
Year of publication	2010
Inclusion criteria: i.e. presenting symptoms, results from previous tests	Children had presented with obesity aged between 5 and 18 years.
Exclusion criteria: i.e. presenting symptoms, results from previous tests	Those with an underlying endocrinologic disease or/and those under medication were excluded from the study.
Sample size	268 participants
Participant demographics (i.e. age, sex, spectrum of presenting symptoms, comorbidity, current treatments)	Forty-six boys (46.2%) were evaluated as prepubertal and 81 (63.8%) as pubertal. Of the girls, 36 (25.5%) were evaluated as prepubertal and 105 (74.5%) as pubertal. In the prepubertal groups, the mean age was 8.9±1.8 years in boys and 8.3±1.4 years in girls, and the mean BMI was 28.2±5.4 kg/m2 in boys and 26.2±5.8 kg/m2 in girls. In the pubertal groups, the mean age was 13.6±1.6 years in boys and 13.2±2.0 years in girls, and the mean BMI was 30.9±4.9 kg/m2 in boys and 30.4±5.0 kg/m2 in girls.
Recruitment centres	Pediatric Endocrinology Clinic at Erciyes University, Faculty of Medicine
Study methodology (consecutive or random; retrospective or prospective)	consecutive
Period that study was carried out (beginning and end date)	Not known
Index test description (including criteria for positive test)	HOMA-IR was calculated using the equation: HOMA-IR=Fasting insulin (μU/mL) x Fasting glucose (mg/dL) /405.
Reference test description (including criteria for positive test)	OGTT was carried out in order to determine insulin resistance. Following a 3-day high carbohydrate diet (300 g /day) and overnight fasting, an oral dose of 1.75 g/kg (maximum 75 g) glucose was given, and blood samples were taken at 0, 30, 60, 90 and 120 minutes from a venous catheter for glucose and insulin assessments. A total (the sum of insulin levels at 0, 30, 60, 90, and 120 minutes during the OGTT) insulin level exceeding 300 μU/mL was taken as hyperinsulinemia. At each point in OGTT, both glucose and insulin levels were measured, and then, total insulin levels exceeding 300 μU/mL were recorded as hyperinsulinemia. Glucose levels at 120th minute were taken as a criterion for impaired glucose tolerance or DM
Geographical location of data collection	Kayseri, Turkey
Setting of data collection	Pediatric Endocrinology Clinic at Erciyes University, Faculty of Medicine
Persons executing and interpreting index tests (numbers, training, and expertise)	Not known
Persons executing and interpreting reference test	Not known
Index/reference time interval (and treatments carried out in between)	Not known.
Distribution of severity of disease in those with target condition	N/A
Other diagnoses in those without target condition	No other diagnosis was reported.
Adverse events from index test	N/A
Adverse events from reference test	N/A

Author:	Lee, J., Lee, Ya, Kim, Lee, SY, Choong, Yang
Year of publication	2019
Inclusion criteria: i.e. presenting symptoms, results from previous tests	9,502 subjects aged 10 to 29 years were considered as potential participants
Exclusion criteria: i.e. presenting symptoms, results from previous tests	Participants who had fasted <8 hours before sample collection (n=310), had no glucose or HbA1c data (n=1,647), were previously diagnosed with DM (n=11), were pregnant at the time of the survey (n=53), or had anemia with a haemoglobin level <11.5 g/dL in children aged 10 to 11 years, <12 g/dL in children aged 12 to 14 years and females aged ≥15 years, or <13 g/dL for males aged ≥15 years (n=343). Subjects with hemoglobinopathy were not considered for the present study, because the condition is extremely rare in the Korean population.
Sample size	4129
Participant demographics (i.e. age, sex, spectrum of presenting symptoms, comorbidity, current treatments)	4,129 (45.1%) were in the youth group (10 to 19 years of age) and 3,203 (54.9%) were in the young adult group (20 to 29 years of age)
Recruitment centres	
Study methodology (consecutive or random; retrospective or prospective)	a nationally representative cross-sectional examination of non-institutionalized Korean citizens with a multi-stage clustered probability design conducted by the Korea Centres for Disease Control and Prevention
Period that study was carried out (beginning and end date)	2011-2016
Index test description (including criteria for positive test)	The HbA1c cut off criteria recommended by the ADA and KDA, namely, ≥6.5% for DM by HbA1c (DM_{A1C}) and 5.7% to 6.4% for prediabetes by HbA1c ($PreDM_{A1C}$).
Reference test description (including criteria for positive test)	DM by FPG (DMFPG) was defined as an FPG level ≥126 mg/dL, and IFG was defined as an FPG level between 100 and 125 mg/dL.
Geographical location of data collection	Seoul, Republic of Korea
Setting of data collection	the Korea National Health and Nutrition Examination Survey
Persons executing and interpreting index tests (numbers, training, and expertise)	Not known
Persons executing and interpreting reference test	Not known
Index/reference time interval (and treatments carried out in between)	Not known.
Distribution of severity of disease in those with target condition	N/A
Other diagnoses in those without target condition	No other diagnosis was reported.
Adverse events from index test	N/A
Adverse events from reference test	N/A

Author:	Liang, Fu, Jiang, Dong, Wang, Wu
Year of publication	2015
Inclusion criteria: i.e. presenting symptoms, results from previous tests	Obese schoolchildren with complete record were eligibly included in the current study. Age- and sex-specific Body Mass Index (BMI) percentiles Age- and sex-specific Body Mass Index (BMI) percentiles.
Exclusion criteria: i.e. presenting symptoms, results from previous tests	the known presence of diabetes or high blood pressure, the use of drugs which influence glucose or lipid metabolism (glucocorticoid), specific causes of endocrine or genetic obesity, low birth weight, distress during blood sampling or a difficult phlebotomy (more than 5 min) as well as menstrual cycle changes that indicate the presence of Polycystic Ovary Syndrome in female participants.
Sample size	976 participants (female: 286, male 690)
Participant demographics (i.e. age, sex, spectrum of presenting symptoms, comorbidity, current treatments)	Total: sex (F/M): 286/690; age group: <10 years 349, >= 10 years: 627; Pubertal stage: Prepubertal: 458, pubertal: 518 Strata Non-MS: sex (F/M): 231 (74.5%)/511 (74.1%); age group: <10 years 281 (80.5%), >= 10 years 443 (70.7%); Prepubertal: 372 (81.2%), pubertal: 352 (68.0%) Strata MS: sex (F/M): 73 (25.2%)/197 (25.9%); age group: <10 years 68 (19.5%), >= 10 years 184 (29.3%); Prepubertal: 86 (18.8%), pubertal: 166 (32.0%)
Recruitment centres	Endocrinology Department of the Children's Hospital, Zhejiang University,
Study methodology (consecutive or random; retrospective or prospective)	a cross-sectional study
Period that study was carried out (beginning and end date)	Between May 2007 and June 2013
Index test description (including criteria for positive test)	Insulin resistance index was calculated by homeostasis model assessment of insulin resistance (HOMA1-IR) as (fasting insulin mU/L) × (fasting glucose mmol/L)/22.5 and the HOMA2-IR index was obtained by the program HOMA Calculator v2.2.2 at http://www.dtu.ox.ac.uk/homacalculator/index.php.; TG/HDL-C > 1.25.
Reference test description (including criteria for positive test)	OGTT
Geographical location of data collection	China
Setting of data collection	Endocrinology Department of the Children's Hospital
Persons executing and interpreting index tests (numbers, training, and expertise)	Not known
Persons executing and interpreting reference test	Not known
Index/reference time interval (and treatments carried out in between)	Not known.
Distribution of severity of disease in those with target condition	N/A
Other diagnoses in those without target condition	No other diagnosis was reported.

Adverse events from index test	N/A
Adverse events from reference test	N/A

Author:	Maffeis, Pinelli, Brambilla, Banzato, Valzogher, Ulmi, di Candia et al
Year of publication	2010
Inclusion criteria: i.e. presenting symptoms, results from previous tests	Inclusion criteria: white ethnicity, age (4–17 years), and obesity
Exclusion criteria: i.e. presenting symptoms, results from previous tests	Exclusion criteria were obesity associated with endocrine disorders, chronic diseases, malformations, and chronic use of drugs.
Sample size	563 (315 males, 248 females)
Participant demographics (i.e. age, sex, spectrum of presenting symptoms, comorbidity, current treatments)	Girls (total): Age (years): 11.2 (2.7); Height (m): 148.1 (13.0); Weight (kg): 67.3 (21.5); BMI: 30.0 (6.0); Z-BMI: 2.23 (0.61); FPG (mmol/l): 4.76 (0.53); FSI (μU/ml): 14.8 (10.2); SBP (mm Hg): 119 (16.5); DBP (mm Hg): 68.75 (13.25); High BP (%): 52; HOMA-IR: 2.49 (1.65); 2hPG (mmol/l): 5.81 (1.21); IFG (%): 5.2; IGT (%): 10.1 Boys (total): Age (years): 11.4 (2.5); Height (m): 152.0 (15.0); Weight (kg): 69.1 (21.0); BMI: 29.2 (4.8); Z-BMI: 2.11 (0.52); FPG (mmol/l): 4.86 (0.49); FSI (μU/ml): 12.5 (7.5); SBP (mm Hg): 118.5 (15.5); DBP (mm Hg): 71.0 (13); High BP (%): 41; HOMA-IR: 2.74 (1.77); 2hPG (mmol/l): 5.80 (1.03); IFG (%): 6.3; IGT (%): 4.5
Recruitment centres	University Hospitals
Study methodology (consecutive or random; retrospective or prospective)	Consecutive sampling
Period that study was carried out (beginning and end date)	N/A
Index test description (including criteria for positive test)	Homeostasis model assessment of insulin resistance (HOMAIR) was calculated as fasting serum insulin (FSI) (μU/ml) × fasting plasma glucose (FPG) (mmol/l)/22.5
Reference test description (including criteria for positive test)	OGTT (1.75-g/kg oral glucose, maximum 75 g), after a 12-h overnight fast; blood samples were obtained at 0, 30, 60, 90, and 120 min for determination of glucose and insulin
Geographical location of data collection	Italy
Setting of data collection	Verona and Milan University Hospitals
Persons executing and interpreting index tests (numbers, training, and expertise)	Not known
Persons executing and interpreting reference test	Not known
Index/reference time interval (and treatments carried out in between)	Not known.
Distribution of severity of disease in those with target condition	N/A
Other diagnoses in those without target condition	No other diagnosis was reported.
Adverse events from index test	N/A
Adverse events from reference test	N/A

Author:	Maldonado-Hernández, Martínez-Basila, Salas-Fernández, Navarro-Betancourt, Piña-Aguero, Bernabe-García
Year of publication	2016
Inclusion criteria: i.e. presenting symptoms, results from previous tests	Apparently healthy adolescents aged between 10 and 16 years assented to participate in the study.
Exclusion criteria: i.e. presenting symptoms, results from previous tests	Current chronic disease, diagnosed T2DM or presence of a capillary blood glucose level of ≥126 mg/dL, the use of medications that affect glucose metabolism, and fever in the last 48 hours.
Sample size	133 (62 females and 71 males)
Participant demographics (i.e. age, sex, spectrum of presenting symptoms, comorbidity, current treatments)	Mean age was 13 years, weight and abdominal circumference values ranged from 34 to 113 kg and from 63 to 129 cm, respectively. Body mass index (BMI) presented a median of 23 (15.6 to 37.8 kg/m2), and this parameter was used to classify individuals into three groups according to the child growth standards established by the World Health Organization (18), namely, lean (BMI between p3 and p85, 42.1%), overweight (BMI >p85, 14.3%) and obese (BMI >p97, 43.6%)
Recruitment centres	Medical Nutrition Research Unit of the Mexican Social Security Institute
Study methodology (consecutive or random; retrospective or prospective)	Cross-sectional study
Period that study was carried out (beginning and end date)	Enrolling participants during 2011
Index test description (including criteria for positive test)	the ^{13}C-glucose breath test (13C-GBT) - The ^{13}C-GBT consists of ingestion of a ^{13}C-glucose dose used as a tracer to label exhaled CO_2 together with an oral load of non-labelled glucose to challenge insulin-dependent tissues.
Reference test description (including criteria for positive test)	homeostasis model assessment (HOMA) IR (≥p95 adjusted by gender and age), fasting plasma insulin (≥p90 adjusted by gender and Tanner stage), results of 2-h oral glucose tolerance test (OGTT), insulin levels (≥65 μU/mL)
Geographical location of data collection	Mexico City, Mexico
Setting of data collection	Medical Nutrition Research Unit
Persons executing and interpreting index tests (numbers, training, and expertise)	Not known
Persons executing and interpreting reference test	Not known
Index/reference time interval (and treatments carried out in between)	Not known.
Distribution of severity of disease in those with target condition	N/A
Other diagnoses in those without target condition	No other diagnosis was reported.
Adverse events from index test	N/A
Adverse events from reference test	N/A

Author:	Mutlu, Özsu, Cizmecioglu, Hatun
Year of publication	2013
Inclusion criteria: i.e. presenting symptoms, results from previous tests	obese/overweight children aged from 7 to 18 years who underwent OGTT
Exclusion criteria: i.e. presenting symptoms, results from previous tests	N/A
Sample size	106 obese/overweight children aged from 7 to 18 years
Participant demographics (i.e. age, sex, spectrum of presenting symptoms, comorbidity, current treatments)	Age (years): 13.4+/-2.6 (median: 13.5); Gender (F/M): 71 (67%)/35(33%); Pubertal status (prepubertal, midpubertal, pubertal) (n): 10 (9%)/45 (43%)/51 (48%); BMI (kg/m2): 31.5+/-5.1 (20.7-46) (median: 31.8); BMI-SDS: 2.6+/-0.6 (0.9-4.5) (median: 2.6); Waist circumference (cm): 94.6+/-11.6 (65-123) (median: 94); HbA1c (%):: 5.3+/-0.7 (4-7.5); Impaired fasting glucose [n (%)] 3 (3%); 30 minute insulin level (µU/mL): 102.3+/-83 (1.7-476); Impaired glucose tolerance (n, %): 18 (17%); Triglyceride level (mg/dL): 118.2+/-62.7 (41-337); Total cholesterol level (mg/dL): 163.1+/-52.4 (40-575); HDL level (mg/dL): 43.5+/-11.8 (4-71); LDL level (mg/dL): 92.9+/-27.1 (13.6-176); VLDL level (mg/dL): 22.9+/-13.9 (10-94)
Recruitment centres	N/A
Study methodology (consecutive or random; retrospective or prospective)	Medical records evaluated retrospectively
Period that study was carried out (beginning and end date)	between February 2010 and February 2011
Index test description (including criteria for positive test)	HbA1c; Homeostasis model assessment of insulin resistance (HOMA-IR) was calculated with the formula [(FG (nmol/L) x fasting insulin (µU/mL)/22.5]. A level above 3.16 was accepted as a marker of insulin resistance,
Reference test description (including criteria for positive test)	OGTT - 1-hour glucose level in oral glucose tolerance test (OGTT); Impaired FG (IFG) was defined as a FG concentration between 100-125 mg/dL and impaired glucose tolerance (IGT) as a 2-hour glucose concentration between 140-199 mg/dL. Diabetes was defined as either a FG at or above 126 mg/dL or a 2-hour glucose concentration in OGTT at or above 200 mg/dL, as per the criteria of the American Diabetes Association (ADA)
Geographical location of data collection	Turkey
Setting of data collection	N/A
Persons executing and interpreting index tests (numbers, training, and expertise)	Not known
Persons executing and interpreting reference test	Not known
Index/reference time interval (and treatments carried out in between)	N/A
Distribution of severity of disease in those with target condition	N/A
Other diagnoses in those without target condition	No other diagnosis was reported.
Adverse events from index test	N/A
Adverse events from reference test	N/A

Author:	Nam, Cho, Kim, Rhie, Chung, Lee, Suh
Year of publication	2017
Inclusion criteria: i.e. presenting symptoms, results from previous tests	1) age 10 years and above or at the onset of puberty, 2) overweight or obese (body mass index [BMI] ≥ 85th percentile for age and gender), and 3) two or more additional risk factors for diabetes, consistent with American Diabetes Association (ADA) recommendations for type 2 DM screening, such as family history of type 2 DM, race or ethnicity, signs of insulin resistance or its associated conditions, maternal history of DM or gestational DM
Exclusion criteria: i.e. presenting symptoms, results from previous tests	Children and adolescents with known diabetes or newly diagnosed type 1 DM (low C-peptide levels and the presence of beta-cell autoantibodies) or anemia (hemoglobin [Hb] < 11.5 g/dL in subjects under the age of 12 years; Hb < 13.0 g/dL and Hb < 12.0 g/dL in boys and girls aged 12 years and over, respectively)
Sample size	389
Participant demographics (i.e. age, sex, spectrum of presenting symptoms, comorbidity, current treatments)	389 children and adolescents less than 20 years (217 boys, 55.8%); normoglycemia (FPG < 5.6 mmol/L and 2-hr PG < 7.8 mmol/L), prediabetes (FPG 5.6–6.9 mmol/L or 2-hr PG 7.8–11.0 mmol/L) or type 2 DM (FPG ≥ 7.0 mmol/L or 2-hr PG ≥ 11.1 mmol/L); 48 overweight and 341 obese and there were more boys (217, 55.8%) than girls. The mean age was 13.0 ± 2.5 years. The mean height SDS, body weight SDS, and BMI SDS were 0.9 ± 1.2, 2.2 ± 0.8, and 2.2 ± 0.6, respectively. About half of the children (203, 52.2%) had a family history of DM in first- and second degree relatives. Their mean FPG, 2-hr PG and HbA1c levels were 6.1 ± 2.6 mmol/L, 9.0 ± 5.2 mmol/L, and 6.3% ± 2.1%, respectively
Recruitment centres	Pediatric Endocrinology Clinic
Study methodology (consecutive or random; retrospective or prospective)	retrospectively reviewed the medical records
Period that study was carried out (beginning and end date)	between January 2010 and June 2016
Index test description (including criteria for positive test)	HbA1c: The diagnostic performance of HbA1c was investigated using sensitivity, specificity, positive predictive value, and negative predictive value at thresholds of 5.7% for prediabetes and 6.5% for diabetes, as recommended by the ADA. The area under the receiver operating characteristic curve (AUC) was generated to assess the predictive capability of HbA1c for prediabetes and diabetes. The optimal cutoff points were determined as the points at which the distance between the AUC curve and the point with a sensitivity of 1 and a specificity of 0 was minimized
Reference test description (including criteria for positive test)	OGTT
Geographical location of data collection	Korea
Setting of data collection	6 University Hospitals
Persons executing and interpreting index tests (numbers, training, and expertise)	Not known

Persons executing and interpreting reference test	Not known
Index/reference time interval (and treatments carried out in between)	Not known.
Distribution of severity of disease in those with target condition	N/A
Other diagnoses in those without target condition	No other diagnosis was reported.
Adverse events from index test	N/A
Adverse events from reference test	N/A

Author:	Nor, Lee, Bacha, Tfayli, Arslanian
Year of publication	2015
Inclusion criteria: i.e. presenting symptoms, results from previous tests	N/A
Exclusion criteria: i.e. presenting symptoms, results from previous tests	N/A
Sample size	225
Participant demographics (i.e. age, sex, spectrum of presenting symptoms, comorbidity, current treatments)	Participants' mean age 14.2+1.9 years (122 black and 103 white, 114 male and 111 female); They were between 10 and 20 yr old with Tanner stages II–V. Among them, 156 had normal glucose tolerance (OBNGT), 37 prediabetes (OB-preDM) [which included 4 impaired fasting glucose (IFG), 30 impaired glucose tolerance (IGT), and 3 with both IFG and IGT], and 32 type 2 diabetes mellitus (OB-T2DM) with negative glutamic acid decarboxylase (GAD) and tyrosine phosphatase-related islet antigen 2 (IA2) autoantibodies. Among the OB-T2DM patients, 7 were on lifestyle modification alone, 15 on Metformin alone, 3 on insulin alone, and 7 on Metformin and insulin together.
Recruitment centres	outpatient obesity and diabetes clinics in the Weight Management and Wellness Center and the Division of Pediatric Endocrinology at the Children's Hospital of Pittsburgh of University of Pittsburgh Medical Center (UPMC)
Study methodology (consecutive or random; retrospective or prospective)	Cross-sectional data
Period that study was carried out (beginning and end date)	N/A
Index test description (including criteria for positive test)	Tyg Index, TyG/HDL, 1/IF TyG index was calculated as the Ln [fasting triglycerides(mg/dl)×fasting glucose(mg/dL)/2]; The calculation for the fasting insulin (IF) was made by using the mean of four determinations obtained before the start of the clamp (times −30, −20, −10, and 0 min).
Reference test description (including criteria for positive test)	Insulin simulated glucose disposal (Rd) - Insulin-stimulated Rd, expressed in mg/kg/min was calculated during the last 30 min of the clamp to reflect in vivo insulin sensitivity
Geographical location of data collection	Pittsburgh, USA
Setting of data collection	Outpatients obesity and diabetes clinic and children´s hospital
Persons executing and interpreting index tests (numbers, training, and expertise)	Not known
Persons executing and interpreting reference test	Not known
Index/reference time interval (and treatments carried out in between)	Not known.
Distribution of severity of disease in those with target condition	N/A
Other diagnoses in those without target condition	No other diagnosis was reported.
Adverse events from index test	N/A
Adverse events from reference test	N/A

Author:	Pandey, Midha, Rao, Katiyar, Wal, Kaur, Martolia
Year of publication	2017
Inclusion criteria: i.e. presenting symptoms, results from previous tests	N/A
Exclusion criteria: i.e. presenting symptoms, results from previous tests	N/A
Sample size	526
Participant demographics (i.e. age, sex, spectrum of presenting symptoms, comorbidity, current treatments)	277 boys and 249 girls. The mean age of boys was 18.5+/-1.5 years and the mean age of girls was 17.9+/-1.8years. Average BMI of the study subjects was 22.0+/-3.5 kg/m2 in boys and 20.8+/-4.1kg/m2 in girls. The mean waist circumference of boys was 80.2+/-15.3cm and that of girls was 72.9+/-17.5 cm. Average fasting blood glucose level of the study population was 92.9+/-12.4 mg/dl and 87.9+/-14.8 mg/dl among boys and girls respectively. Prevalence of prediabetes among the study subjects was 32.1% and that of diabetes was 0.8%. None of the subjects had previously diagnosed Type 1 or Type 2 diabetes mellitus.
Recruitment centres	Institute of Paramedical Science
Study methodology (consecutive or random; retrospective or prospective)	Cross-sectional study
Period that study was carried out (beginning and end date)	N/A
Index test description (including criteria for positive test)	BMI boys ≥22.8 kg/m2; BMI girls ≥20.5 kg/m2 Waist circumference boys ≥82.5 cm; Waist circumference girls ≥80.3 cm
Reference test description (including criteria for positive test)	N/A
Geographical location of data collection	India
Setting of data collection	the Institute of Paramedical Sciences, affiliated to Chatrapati Shahuji Maharaj University, Kanpur
Persons executing and interpreting index tests (numbers, training, and expertise)	Not known
Persons executing and interpreting reference test	Not known
Index/reference time interval (and treatments carried out in between)	Not known.
Distribution of severity of disease in those with target condition	N/A
Other diagnoses in those without target condition	No other diagnosis was reported.
Adverse events from index test	N/A
Adverse events from reference test	N/A

Author:	Puri, Freeman, Garcia, Nussbaum, Nardi
Year of publication	2007
Inclusion criteria: i.e. presenting symptoms, results from previous tests	ages 10-18 years or in puberty, BMI >85th percentile and family history of DM2, race/ ethnicity (African American, Caribbean Hispanic) with signs of insulin resistance, such as acanthosis nigricans.
Exclusion criteria: i.e. presenting symptoms, results from previous tests	N/A
Sample size	167
Participant demographics (i.e. age, sex, spectrum of presenting symptoms, comorbidity, current treatments)	ages 10-18 years or in puberty, BMI >85th percentile and family history of DM2, race/ ethnicity (African American, Caribbean Hispanic) with signs of insulin resistance, such as acanthosis nigricans; with a mean age 14 ± 2.3 years, BMI 38.1 ± 7.5 kg/m2, and BMI z-score 2.47 ± 0.36, who met all the ADA criteria for DM2 screening, were enrolled to undergo an OGTT. (See Table 1 for descriptive statistics.) There were no ethnic differences between boys and girls.
Recruitment centres	Pediatric Diabetes Clinics and Pediatric Endocrine Clinics at the Children's Hospital at Montefiore in the Bronx
Study methodology (consecutive or random; retrospective or prospective)	Consecutive
Period that study was carried out (beginning and end date)	N/A
Index test description (including criteria for positive test)	HOMA - Homeostatic model assessment (HOMAIR) and glucose to insulin ratios were calculated for all subjects from glucose and insulin levels obtained in the fasting state.
Reference test description (including criteria for positive test)	An OGTT was performed on 167 minority youth after a 12 hour overnight fast. Blood samples for insulin and plasma glucose levels were obtained at 0 minutes and 30, 60, 90, and 120 minutes after an oral dose of glucose 1.75 g/kg with a maximum dose of 75 g. Insulin levels were determined using a solid-phase two-site chemiluminescent immunometric assay.
Geographical location of data collection	Bronx, New York, USA
Setting of data collection	Children´s Hospital
Persons executing and interpreting index tests (numbers, training, and expertise)	Not known
Persons executing and interpreting reference test	Not known
Index/reference time interval (and treatments carried out in between)	Not known.
Distribution of severity of disease in those with target condition	N/A
Other diagnoses in those without target condition	No other diagnosis was reported.
Adverse events from index test	N/A
Adverse events from reference test	N/A

Author:	Sharma, Fleming
Year of publication	2012
Inclusion criteria: i.e. presenting symptoms, results from previous tests	African American children with BMI´s at or above 85th percentile.
Exclusion criteria: i.e. presenting symptoms, results from previous tests	8 years of age or younger; 11 years of age or older; fasting glucose 120 mg/dl; any known metabolic disease; and taking medications known to affect the study outcomes.
Sample size	172
Participant demographics (i.e. age, sex, spectrum of presenting symptoms, comorbidity, current treatments)	Boys (n=70): age: 9.96 (0.09); Pubertal stage (1-5): 2.00 (0.17); BMI, z-score: 1.85 (0.07); Waist circumference (WC), cm: 80.1 (1.52); HbA1c: 5.14 (0.05); Fasting glucose, mg/dl: 87.7 (0.70); Insulin, μU/ml: 9.29 (0.75); HOMA-IR: 2.05 (0.18) Girls (n=102): age: 9.80 (0.08); Pubertal stage (1-5): 2.58 (0.13); BMI, z-score: 2.05 (0.06); Waist circumference (WC), cm: 86.0 (1.46); HbA1c: 5.17 (0.04); Fasting glucose, mg/dl: 86.9 (0.73); Insulin, μU/ml: 13.1 (0.81); HOMA-IR: 2.85 (0.19)
Recruitment centres	Participants were recruited by distributing pamphlets at local recreational sites and schools in inner-city Oakland
Study methodology (consecutive or random; retrospective or prospective)	A full set of data of cross-sectional analysis
Period that study was carried out (beginning and end date)	N/A
Index test description (including criteria for positive test)	Fasting glucose and insulin values were used to calculate homeostasis model assessment of insulin resistance (HOMA-IR), defined as fasting glucose (mmol/l) x insulin (μU/ml)/22.5, and used as an index of insulin resistance; Prediabetes was assessed using previously recommended cut offs for fasting plasma glucose of ≥110 mg/dl, HOMA-IR of ≥2.5
Reference test description (including criteria for positive test)	HbA1C was analysed by a commercial lab (Diabetes Technologies Inc.) using a dual HPLC method which improves accuracy of measurements by detecting and reducing error due to variants that interfere with interpretation of HbA1C data; Prediabetes was assessed using previously recommended cut offs HbA1c of ≥5.7%.
Geographical location of data collection	Oakland, CA, USA
Setting of data collection	Children's Hospital and Research Center Oakland, CA, Oakland
Persons executing and interpreting index tests (numbers, training, and expertise)	Not known
Persons executing and interpreting reference test	Not known
Index/reference time interval (and treatments carried out in between)	Not known.
Distribution of severity of disease in those with target condition	N/A
Other diagnoses in those without target condition	No other diagnosis was reported.
Adverse events from index test	N/A
Adverse events from reference test	N/A

Author:	Tirabanchasak, Siripunthana, Supornsilchai, Wacharasindhu, Sahakitrungruang
Year of publication	2015
Inclusion criteria: i.e. presenting symptoms, results from previous tests	pediatric patients who have risk factors according to the American Diabetes Association guidelines
Exclusion criteria: i.e. presenting symptoms, results from previous tests	known endocrine/metabolic disorders, and patients who have taken medications affecting glucose metabolism
Sample size	115 obese subjects (76 males and 39 females, age ranging from 8 to 18 years)
Participant demographics (i.e. age, sex, spectrum of presenting symptoms, comorbidity, current treatments)	All 115 participants met criteria of obesity. Median age of the patients was 12.6 years (range: 8.4–17.5), and median BMI-SDS was 3.3 (range: 2.0–4.6). Some 85 patients (~74%) already entered puberty, in which 45 subjects were in early puberty (Tanner II–III) and 40 children were in late puberty (Tanner IV–V). A family history of T2DM was found in 72 patients (62.5%) and a history of maternal GDM was found in five patients (4.3%). The family history of obesity was documented in 41 patients (35.7%). According to obesity-related complications, dyslipidemia was the most common one which was found in 57 patients (49.6%). Hypertension was found in 48 patients (41.7%). Sleep apnoea was found in 45 patients (39.1%). NASH was observed in 33 patients (28.7%). PCOS was found in five female patients (12.8%). We found 11 patients (9.6%) with behavioral problems such as attention-deficit/hyperactivity disorder, intellectual disability, adjustment disorder, school refusal, and anxiety disorder. There were no significant differences between males and females for any of these factors, except median BMI-SDS in boys were significantly higher, and NASH was more common in boys than in girls
Recruitment centres	King Chulalongkorn Memorial Hospital
Study methodology (consecutive or random; retrospective or prospective)	The study protocol; Data collected from the medical charts
Period that study was carried out (beginning and end date)	during 2007–2013
Index test description (including criteria for positive test)	FG, HOMA insulin sensitivity was assessed by the ratio of fasting glucose (FG) to fasting insulin (FI) (FG/FI), whole body insulin sensitivity index (WBISI) using the Matsuda method (17), and the quantitative insulin sensitivity check index (QUICKI) pancreatic β-cell function was assessed by HOMA-derived β-cell function (HOMA-β), and insulinogenic index (IGI)
Reference test description (including criteria for positive test)	An OGTT using 1.75 g/kg of glucose (maximum 75 g) was performed on each patient in the fasting state. Blood samples for glucose and insulin levels were collected at 0, 30, 60, 90, and 120 min. A patient was classified as having ‘IFG’ if FBG was between 100 and 125 mg/dL and ‘IGT’ if the 2-h post-OGTT glucose was 140–199 mg/dL. T2DM was diagnosed if FBG ≥ 126 mg/dL or the 2-h post-OGTT glucose ≥ 200 mg/dL
Geographical location of data collection	Bangkok, Thailand

Setting of data collection	Hospital
Persons executing and interpreting index tests (numbers, training, and expertise)	Not known
Persons executing and interpreting reference test	Not known
Index/reference time interval (and treatments carried out in between)	Not known.
Distribution of severity of disease in those with target condition	N/A
Other diagnoses in those without target condition	No other diagnosis was reported.
Adverse events from index test	N/A
Adverse events from reference test	N/A