A systematic review and mixed treatment comparison assessing the comparative effectiveness and safety of oral antihyperglycemic drugs as monotherapy in patients with type 2 diabetes mellitus

---

Dissertation

zum Erwerb des Doktorgrades der Humanbiologie
an der Medizinischen Fakultät
der Ludwig-Maximilians-Universität zu München

angefertigt am Institut für medizinische Informationsverarbeitung, Biometrie und
Epidemiologie der Medizinischen Fakultät der Ludwig-Maximilians-Universität zu München

vorgelegt von
Thomas Bartmus, MPH

eingereicht am Institut für medizinische Informationsverarbeitung, Biometrie und
Epidemiologie der Medizinischen Fakultät der Ludwig-Maximilians-Universität zu München

2014

Aus dem Institut für medizinische Informationsverarbeitung, Biometrie und Epidemiologie der Medizinischen Fakultät
der Ludwig-Maximilians-Universität zu München
Vorstand: Herr Professor Dr. rer. nat. Ulrich Mansmann

A systematic review and mixed treatment comparison assessing the comparative effectiveness and safety of oral antihyperglycemic drugs as monotherapy in patients with type 2 diabetes mellitus

Eine systematische Übersichtsarbeit und Netzwerk- Metaanalyse zur Beurteilung der Wirksamkeit und Sicherheit verschiedener oraler Antidiabetika als Monotherapie bei Patienten mit Diabetes Mellitus Typ 2

Dissertation
zum Erwerb des Doktorgrades der Humanbiologie
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München

vorgelegt von

Thomas Bartmus, MPH

aus

Schäßburg

Jahr

2014

Mit Genehmigung der Medizinischen Fakultät
der Universität München

Berichterstatter:

Prof. Dr. rer. nat. Ulrich Mansmann

Mitberichterstatter:

Prof. Dr. rer. nat. Konstantin Strauch

Prof. Dr. med. Dr. med. habil. Hans Uwe Janka

Prof. Dr. med. Eberhard Standl

Dekan:

Prof. Dr. med. Dr. h.c. Maximilian Reiser, FACR, FRCR

Tag der mündlichen Prüfung:

02.07.2014

Danksagung

Unsere Träume können sich nur dann erfüllen,
wenn wir unbeirrbar an ihre Erfüllbarkeit glauben;
denn gerade unser Glauben
gibt ihnen die Kraft, schließlich wahr zu werden.

unbekannt

An dieser Stelle möchte ich mich bei allen Personen bedanken, die mir bei der Erstellung der Doktorarbeit unterstützt haben. Herrn Professor Dr. rer. nat. Ulrich Mansmann danke ich für die Betreuung meiner Doktorarbeit, sowie für den anregenden Diskussionsaustausch, den fachlichen Rat und die zielführenden Hinweise. Herrn Sudip Jung Karki danke ich, dass er mir während seines Praktikums im Rahmen seines Masterstudiums zum Master of Science of Epidemiology als Zweitbegutachter für die Komponenten der systematischen Übersichtsarbeit zur Verfügung stand, die gemäß guter wissenschaftlicher Praxis nach dem Vier-Augen-Prinzip durchzuführen sind. Darüber hinaus möchte ich mich bei den wissenschaftlichen KollegInnen Frau Dr. Eva Annette Rehfuess, BA, MA, Herrn Tianan Yang, Frau Mariangela Salomon, Frau Olga Petrich und Herrn Dr. Shinji Takenaka bedanken, die mir als MuttersprachlerInnen bei der Bewertung, Übersetzung und Datenextraktion von Publikationen in italienischer, chinesischer, spanischer, russischer und japanischer Sprache geholfen haben.

Zuletzt gilt mein Dank meiner Familie für die große Unterstützung während dieser Zeit.

Contents

List of tables

List of figures

List of abbreviations

ADA	American Diabetes Association
AE	Adverse event
AGD	Aggregated level data
AIC	Akaike information criterion
AKDÄ	Drug Commission of the German Medical Association
AMNOG	Act on the Reform of the Market for Medicinal Products
BMI	Body-mass index
CDA	Canadian Diabetes Association
CENTRAL	Cochrane Central Register for Controlled Trials
CI	Frequentist confidence interval
CrI	Bayesian credible interval
DAG	Directed acyclic graph
DDG	German Diabetes Association
DEGAM	German College of General Practitioners and Family Physicians
DIC	Deviance information criterion
DM	Diabetes mellitus
DPP-4	Dipeptidylpeptidase four
EASD	European Association for the Study of Diabetes
EPAR	European public assessment report
EU	European Union
FDA	Food and Drug Administration
FE	Fixed effects model
FPG	Fasting plasma glucose
GLP-1	Glucagon-like peptide 1
HbA1c	Glycated Haemoglobin
HDL-C	High density lipoprotein cholesterol
HR	Hazard ratio
HRQoL	Health-related quality of life
HTA	Health technology assessment
ICF	Inconsistency factor
ICDF	Inconsistency degrees of freedom
IFG	Impaired fasting glucose
IPD	Individual patient data
IGT	Impaired glucose tolerance
ISPOR	International Society of Pharmacoeconomics and Outcomes Research
IQWiG	Institute for Quality and Efficiency in Health Care
ITC	Indirect treatment comparisons
ITT	Intention-to-treat analysis
KG	Kilogram
LDL-C	Low density lipoprotein cholesterol
LOCF	Last observation carried forward
MA	Meta-analysis
MAH	Market authorisation holder
MCMC	Markov chain Monte Carlo
MD	Difference in means
MedDRA	Medical dictionary for regulatory activities

MESH	Medical subject heading
MTC	Mixed treatment comparison
MTM	Multiple treatment meta-analysis
NICE	National Institute for Health and Care Excellence
NMA	Network meta-analysis
OAD	Oral antidiabetic drug
OR	Odds ratio
P	P-value
PPAR-gamma	Peroxisome proliferator-activated receptor-gamma
PPV	Positive predictive value
PRISMA	Preferred reporting items for systematic reviews and meta-analyses
PRO	Patient reported outcomes
RCT	Randomized controlled trial
RE	Random effects model
RR	Relative risk; risk ratio
SA	Sensitivity analysis
SAE	Severe adverse event
SD	Standard deviation
SEM	Standard error of the mean
SGB V	Social Code Book V
SHI	Statutory health insurance (GKV)
SmPC	Summary of Product Characteristics
T1DM	Type 1 diabetes mellitus
T2DM	Type 2 diabetes mellitus
TG	Triglyceride
U.S.	United States
WHO	World Health Organization

1. Introduction

1.1 Diabetes mellitus: classification, diagnosis and pathophysiology

Diabetes mellitus (DM) represents a group of metabolic disorders with the lead symptom hyperglycemia [1]; the loss of glycemic control is induced by a dysfunction in the metabolism of the hormone insulin which is responsible for blood glucose regulation; hyperglycemia is characterized either by a defective insulin secretion (insulin deficiency) of the beta cells of the pancreas or an impaired pharmacodynamic of insulin on its target organs, mainly muscle and liver cells (insulin resistance), or both. According to the etiological classification of the American Diabetes Association (ADA) diabetes mellitus is sub-divided into four main categories: type 1 diabetes mellitus (T1DM), type 2 diabetes mellitus (T2DM), other specific types and gestational diabetes mellitus [2] (see Table 1).

Table 1: Etiologic classification of diabetes mellitus [2]

Type 1 diabetes (beta-cell destruction, usually leading to absolute insulin deficiency) A. Immune mediated B. Idiopathic
Type 2 diabetes (may range from predominantly insulin resistance with relative insulin deficiency to a predominantly secretory defect with insulin resistance)
Other specific types A. Genetic defects of beta-cell function B. Genetic defects in insulin action C. Diseases of the exocrine pancreas D. Endocrinopathies E. Drug- or chemical-induced F. Infections G. Uncommon forms of immune-mediated diabetes H. Other genetic syndromes sometimes associated with diabetes
Gestational diabetes mellitus

T1DM, also called insulin dependent diabetes mellitus accounts for approximately 6-7% of diagnosed cases in Germany and is mediated by an autoimmune destruction of the ß-cells of the pancreas. Its acute manifestation is accompanied with symptoms of glucose induced diuresis, exsiccation and polydipsia [3].
T2DM, also referred to as non-insulin dependent diabetes mellitus or adult-onset diabetes mellitus is characterized by insulin resistance and a relative insulin deficiency; at the onset of the disease insulin secretion is predominantly insufficient and delayed in the early postprandial phase, whereas the basal blood levels are normal or can even be elevated in the late phase. However, during the course of disease insulin secretion can worsen and lead to an absolute insulin deficiency. Apart from lack of physical activity, hypertension and dyslipoproteinemia, adipositas, especially affecting the visceral fat tissue, are the main drivers of the insulin resistance [1]. Onset of T2DM is usually subtle and epidemiologic studies show that the metabolic dysfunction precedes the clinical diagnosis for years. Manifestation of T2DM occurs if the insulin resistance is no longer compensated by the impaired insulin secretion; individuals exhibit abnormal plasma glucose concentrations but are usually free of symptoms. According to the ADA recommendations diabetes mellitus can be diagnosed as described in Table 2 [2]; unless hyperglycemia is unambiguous, diagnosis must be confirmed on a subsequent day by any of the three other methods given.

Table 2: Criteria for the diagnosis of diabetes mellitus [2]

HbA1C ≥6.5%. The test should be performed in a laboratory using a method that is NGSP certified and standardized to the DCCT assay[a].
FPG ≥126 mg/dl (7.0 mmol/l). Fasting is defined as no caloric intake for at least 8 h[a].
2-h plasma glucose ≥200 mg/dl (11.1 mmol/l) during an OGTT. The test should be performed as described by the WHO, using a glucose load containing the equivalent of 75 g anhydrous glucose dissolved in water[a].
In a patient with classic symptoms of hyperglycemia or hyperglycemic crisis, a random plasma glucose ≥200 mg/dl (11.1 mmol/l).

[a] In the absence of unequivocal hyperglycemia, criteria 1–3 should be confirmed by repeat testing.
HbA1c, glycated haemoglobin. FPG, fasting plasma glucose. OGTT, oral glucose tolerance test. WHO, World Health Organization.

The prediabetic stages of impaired glucose tolerance (IGT) and impaired fasting glucose (IFG) are defined by a 2-hour postprandial glucose level of 7.8 to 11.1 mmol/l and a fasting blood glucose level of 5.6 to 7 mmol/l, respectively. Individuals suffering from these conditions are at high risk and should be monitored tightly and offered measures to eliminate factors that promote development of diabetes mellitus [1].

All forms of diabetes lead to microvascular and macrovascular complications. Hyperglycemia-induced injury of blood vessels is mediated by upregulated pathological sideways of glucose metabolism. The damaged endothelial cells and pericytes trigger atherosclerotic transformations which eventually lead to microangiopathies such as retinopathy, neuropathy and nephropathy and macroangiopathies of the heart, brain and extremities translating into coronary heart disease, heart failure, heart attack, stroke, (peripheral) artery disease, amputation of diabetic foot and more [3].

1.2 Epidemiology and burden of disease of type 2 diabetes mellitus

Growing levels of obesity and physical inactivity together with an ageing population and a greater longevity among patients with T2DM account for the increasing rise of this emerging pandemic worldwide. Globally the prevalence of T2DM rose from 171 million people in 2000 to 220 million people in 2004 and estimates predict an increase to 380 million for the year 2025 [4]. National surveillance data from the United States showed that from 1980 to 2011 the age-adjusted incidence for all forms of diabetes increased from 3.5 to7.6 cases per 1,000 population per year [5]. This menacing trend is reflected in numbers on a national level, too. In Germany the prevalence for DM lies between 7-8% of the general population, of which T2DM accounts for 80-90% of all cases [1,6]. According to data from a nationwide statutory health insurance (SHI) record the prevalence is expected to rise from 4.7% to 6.7% from 2008 to 2020; the one year incidence per 1,000 population is estimated to step up from 2.2 to 3.1 cases in that period [7]. These figures may even underestimate the real burden of disease since numbers refer only to diagnosed patients and definition of T2DM for identifying eligible patients in the database was rather restrictive.

The mean age of newly diagnosed patients equals approximately 60 years [1,7]. The main hazard for developing T2DM is summarized by the term metabolic syndrome which is defined by visceral adiposity plus two additional items from the following elements: abnormal trigylcerides, decreased high density lipoprotein cholesterol, hypertension or impaired fasting glucose [1]. Further risks are a positive familial anamnesis, higher age, lifestyle factors such as socioeconomic status, lack of physical activity, nutrition with a high fat content and smoking, drugs that interfere with glucose metabolism, gestational diabetes mellitus and other endocrine disorders [8].

About 80% of patients with T2DM develop cardiovascular complications which are by far the main drivers of morbidity and mortality [1]. Almost every second patient with T2DM dies from a cardiovascular cause; angina pectoris, myocardial infarction, stroke, peripheral artery disease, and congestive heart failure are associated with T2DM [4]. Evidence shows that the risk of myocardial infarction in people suffering from diabetes equals that of non-diabetic persons with a history of myocardial infarction [9]. Even in the context of the metabolic syndrome where patients exhibit multiple cardiovascular risk factors evidence suggests that type 2 diabetes acts as an independent predictor. Thus, the adjusted risk for the development of myocardial infarction and stroke of diabetics compared with healthy controls was increased 1.5- to 4.5- fold in women and 1.5- to 2- fold in men and 2- to 6.5- fold in women and 1.5- to 2- fold in men, respectively [10].

Regarding the microvascular complications T2DM is the main reason for end-stage renal disease and accounts for up to 50% of patients undergoing renal replacement therapy [4]. Incidence rates of renal failure average 6 events per 1000 person years; of these patients one third dies and the others require dialysis [11]. Moreover diabetic retinopathy is the most common cause of blindness among people aged 30-69 years. It is estimated to account for 5% of all cases of blindness globally [4]. Furthermore, regarding T2DM the prevalence of diabetic neuropathy is in the range of 13-46% [1]. Affected patients have a more than 25-fold increased risk of amputation compared to people without diabetes [4]. The prognosis of patients already having diabetic complications is fairly poor. For example the median survival of patients referred to a diabetes clinic was only 61 months; likewise, odds ratios for all cause and cardiovascular mortality in patients with diabetic retinopathy compared with non-diabetic controls were 5.1 and 5.6, respectively [12].

Astonishingly a substantial proportion of newly diagnosed patients already suffer diabetic complications. For instance, data from an evaluation of a disease management program in Germany revealed several comorbidities already present at the time of diagnosis such as hypertension (83.8%), dyslipidemia (65.2%), coronary heart disease (27.1%), peripheral artery disease (10%), heart failure (8%), heart attack (6.9%) and stroke (5.8%). Even the diabetes related complications neuropathy, retinopathy and nephropathy existed to a considerable degree at baseline affecting 20.4, 10.7 and 9.9% of patients, respectively [13].

T2DM also has an important impact on health-related quality of life (HRQoL). A pooled analysis of four population based surveys showed that HRQoL measured by the physical components of the 36-item Short Form Health Survey (SF-36) was significantly decreased in diabetics [14].

The impact of premature death in patients with T2DM is well documented. It is assumed that life expectancy on average is shortened by 5-10 years [15]. A German cohort study estimates that life years lost in diabetics depending on the socio-economic status range from 4.88-7.97 [16].

1.3 Health economic burden of disease of type 2 diabetes mellitus

Apart from the burden of disease the care and treatment of patients with DM has a substantial economic impact on health care expenditures. In the US alone, treatment of diabetic patients is estimated to exceed 100 billion dollars each year [17]. Regarding T2DM a cost of illness study focusing on the diabetes-related complications revealed that costs per year increased by USD ($) 1,087(50%), $ 7,352 (360%), and $15,675(771%) for patients with cardiovascular drug therapy, after a major cardiovascular event and for end stage renal disease, respectively, compared to patients with T2DM without any complications [18].
A Germany costing study estimated the direct medical excess costs of individuals with T2DM (diagnosed and undiagnosed) compared with individuals without the condition from a societal perspective including indirect cost such as productivity losses to rise from EUR (€) 11.8 billion in 2010 to €21.1 billion in 2040 [19]. Even if the indirect costs are ignored, another study indicates that patients with T2DM incur more than four times higher direct costs than patients without the disease mainly due to inpatient costs of diabetic-related complications and that these costs account for 14.2% of total health care costs in Germany [20].

1.4 Oral antidiabetic drugs as monotherapy for the treatment of type 2 diabetes mellitus

An comprehensive and effective diabetes care should encompass several aspects: reduction of morbidity and mortality including the cardiovascular risk, avoiding diabetes-related acute and chronic complications, elimination or relief of symptoms by optimizing metabolic control with respect to blood glucose, glycated haemoglobin, lipids, body-mass index (BMI) and blood pressure (see Table 3), treatment of comorbidities and improvement of HRQoL [6].

Table 3: Therapeutic targets of metabolic control for diabetics [6]

Indicator	Therapeutic range
Blood glucose - fasting - 2 hour postprandial	 90–120 mg/dl (5.0–6.7 mmol/l) 130–160 mg/dl (7.2–8.9 mmol/l)
HbA1c	6.5%
Lipids	LDL-C < 100 mg/dl (< 2.5 mmol/l) Secondary goals: HDL-C > 40 mg/dl (> 1.0 mmol/l) Fasting TG < 150 mg/dl (< 1.7 mmol/l)
BMI	< 25kg/m^2
Blood pressure	Systolic ≤130 mmHg Diastolic ≤ 80 mmHg

HbA1c, glycated haemoglobin. LDL-C, low density lipoprotein cholesterol. HDL-C, high density lipoprotein cholesterol. Tg, triglycerides. BMI, body-mass index. mmHg, unit of blood pressure.

To achieve these goals a multifaceted approach is necessary which goes far beyond the pharmacological management of hyperglycemia. However, since the scope of this work focuses on the relative efficacy and safety of oral antidiabetic drugs as monotherapy the other aspects of disease management and therapeutic regimes other than oral monotherapy will only be outlined shortly.
Non pharmacological interventions such as patient education, increased physical activity, low-caloric diet and weight decrease constitute always essential components of every treatment strategy.

In randomized controlled trials (RCT) such lifestyle interventions have proven to be able to reduce the incidence of T2DM in people with impaired glucose tolerance [21,22]. Moreover, smoking is a further risk factor for nephropathy an retinopathy [1].
Thus, diabetic patients should be motivated to give up smoking. Self-monitoring of blood glucose (SMBG) is another integral part of diabetes self-management. However it must be differentiated which subgroups of diabetics benefit from SMBG. There is consensus that patients with insulin therapy profit from self-measurement the frequency of it depending on the intensiveness of drug regimen [1,23]. For patients with T2DM on oral antidiabetic drugs the evidence is controversial. A benefit assessment of the Institute for Quality and Efficiency in Health Care (IQWiG) a German health technology assessment body which is commissioned by the Federal Joint Committee (GBA) which is the highest decision-making body of the joint self-government of physicians, dentists, hospitals and the statutory health insurance in Germany concluded that there is no proof of benefit [24]. Neither observational studies yielded an association between SMBG and decreased mortality or morbidity, nor did interventional studies show a considerable effect of the intervention. Although SMBG showed a statistically significant but clinically non-relevant difference in lowering glycated haemoglobin (HbA1c) (the difference in means (MD) was -0.23% in favour of SMBG; 0.95 confidence interval ($CI_{0.95}$):[-0.12;-0.34]), there were no statistically significant differences regarding hypoglycemia, body weight or HRQoL. As a consequence of this assessment the GBA decided that SMBG is no longer reimbursed by the SHI apart from some exceptions [25].

The control of existent risk factors predominantly hypertension and dyslipidemia as well as the treatment of comorbidities such as heart failure, chronic kidney disease, retinopathy, neuropathy and foot ulcers play a vital role in the medical management of diabetics. For hypertension and dyslipidemia it has been shown that they are independent predictors that increase the incidence of nephro- and neuropathy and coronary heart disease, respectively [26-28]. Thus it is advisable to treat diabetics with blood pressure lowering agents such as angiotensin converting enzyme- inhibitors or angiotensin II receptor blockers which show additional renal benefit and lipid lowering agents, especially statins to therapeutic targets [1]; nevertheless excessive lowering of blood pressure and lipids did not improve cardiovascular outcomes [29,30]. Furthermore, patients at high risk for cardiovascular events should receive antiplatelet therapy with low dose acetylsalicylic acid or Clopidogrel.
The incidence of heart failure is 2-4- fold higher in diabetics than people without; patients should be treated according to clinical guideline algorithms with special regard to the adequate selection of drugs due to diabetic-related complications, e.g. renal dysfunction.
For the prevention of microangiopathies like chronic kidney disease, retinopathy, neuropathy and food problems tight metabolic control (blood glucose, lipids and blood pressure) together with screening for complications like albuminuria, decrease of renal function, altered vision, peripheral loss of sensitivity and food examination are fundamental [23].

Based upon the symptoms and the level of hyperglycemia at diagnosis and after non-pharmacological interventions are not sufficient anymore pharmacological treatment of T2DM is usually initiated with an oral antidiabetic drug (OAD) as monotherapy (see chapter 1.5) if hyperglycemia is moderate; given that hyperglycemia is pronounced or patients present with symptoms and metabolic decompensation, oral combination therapy or initiating insulin therapy may be considered [6,23,31]. Due to the progressive nature of disease adaption of therapeutic strategies is necessary if glycemic targets are no longer met. Usually oral monotherapy is extended by a second and third antihyperglycemic agent; after glycemic failure of oral treatment the transition to injectable therapy takes place.

This is the domain of rapid-, short-, intermediate-, long acting and premixed insulin analogues as well as the newer generation of glucagon-like peptide 1(GLP-1) receptor agonists, e.g. Exenatide, Liraglutide and Lixisenatide. These drugs can be administered in a variety of therapeutic regimes such as monotherapy, combination therapy among each other or combination therapy with OAD which is displayed by the different therapeutic algorithms of the corresponding clinical guidelines [1,23,31].

Epidemiological studies have shown a correlation between the surrogate HbA1c and micro- and macrovascular risk. For example, in the UKPDS 35 study each 1% reduction in HbA1c was associated with a reduction of the risk for diabetes related death, myocardial infarction and microvascular complications of 21%, 14% and 37%, respectively [32].

Moreover, several interventional studies confirmed the positive effect of hyperglycemic control on the reduction of microvascular complications. In a large clinical trial intensive glucose control of Gliclazide plus other OAD to achieve HbA1c targets below 6.5% versus normal glucose control reduced the risk for microvascular events (hazard ratio (HR):0.86; $CI_{0.95}$:[0.77;0.97]; p-value (p)=0.01), a finding which was mainly driven by the reduced incidence of nephropathy (HR:0.79; $CI_{0.95}$:[0.66;0.93]; p=0.006); indeed the benefit was narrowed by weight gain (MD:0.7kg; p<0.001) and an increased incidence of severe hypoglycemia (HR:1.86; $CI_{0.95}$:[1.42;2.40]; p<0.001) in the intensive glucose control group [33]. In the UKPDS 33 study, after a follow-up of 10 years patients with T2DM who were administered an intensive antihyperglycemic treatment (sulfonylurea or insulin) displayed a reduced incidence of a combined microvascular endpoint, including the risk of retinal photocoagulation, compared to patients treated with diet only (relative risk (RR):0.75; $CI_{0.95}$: [0.6;0.93]) [34]. Evidence for a prevention of macrovascular events by drug intervention is scarce, only one trial in obese patients yielded a positive effect [35] (see Metformin).

On the other hand the approach to aim at very tight glucose control bears risks for patient safety, too. In a trial with a very ambitious glycemic target of HbA1c < 6% the higher mortality in the intensive-therapy group led to early discontinuation [36].

Thus, clinical guidelines are determined by the positive effect of drugs mainly on microvascular complications on the one hand and of reservation against a too restrictive lowering due to excess mortality on the other hand and recommendations suggest a range of 6.5-7.5% as target for HbA1c [1,23,31]. The decision required shared decision making involving the patient and should always balance potential benefits against the risks of therapy. A target close to an HbA1c of 6.5% should only be considered if it can be achieved by lifestyle modification or drugs with a low risk of severe adverse events, e.g. Metformin. If factors like age, life expectancy and comorbidities argue against an aggressive glucose control therapeutic goals should be eased and an HbA1c of 7.5% or above seems reasonable.

Alpha-glucosidase inhibitors: Acarbose, Miglitol

Acarbose and Miglitol exhibit their antihyperglycemic effect by inhibiting gastrointestinal enzymes that metabolize carbohydrates; thus digestion and absorption of carbohydrates is delayed and reduced; this cellular mechanism is unlikely to cause hypoglycemia and to affect body weight [3]. A systematic Cochrane review resumed that compared to placebo Acarbose and Miglitol reduce HbA1c levels by approximately 0.8% ($CI_{0.95}$:[0.6;0.9]) and 0.7% ($CI_{0.95}$:[0.4;0.9]), respectively [37]. Due to their pharmacodynamic profile this drug class is especially efficacious in reducing postprandial glucose; however, there is no proof that in RCT the reduction of the surrogate postprandial glucose is correlated with clinical relevant

endpoints [1].There is a lack of evidence regarding clinical endpoints due to micro- and macrovascular complications. The approved dose ranges from 150-300mg and 50-600mg per day for Miglitol and Acarbose, respectively. Gastrointestinal side effects such as diarrhea and flatulence occur frequently and impair compliance; they can be mitigated if therapy is initiated at 50mg and dose escalation is accomplished carefully. The drugs should not be used in people with severe gastrointestinal disorders, chronic inflammations, ulcers and malabsorption. Voglibose was not part of this investigation since it did not meet the eligibility criteria (see 3.1.1).

Dipeptidylpeptidase-4 (DPP-4) inhibitors (gliptins): Sitagliptin, Saxagliptin, Vildagliptin, Linagliptin

The DPP-4 inhibitors Sitagliptin, Saxagliptin, Vildagliptin and Linagliptin belong besides the GLP-1 receptor agonists to the group of incretin agents. Upon glucose stimulation incretins like GLP-1 act as potent insulinotropic hormones by suppressing glucagon secretion and delaying gastric emptying and secretion, thus reducing postprandial glucose. Since the biological half-life of GLP-1 is relatively short DPP-4 inhibitors can sustain its antihyperglycemic effects by inhibiting degradation. Compared to placebo therapy, Sitagliptin and Vildagliptin resulted in an HbA1c reduction of approximately 0.7% and 0.6%, respectively [38]. Similarly, clinical reviews of Linagliptin and Saxagliptin yielded comparable results for their antihyperglycemic efficacy; their placebo subtracted change in reduction of HbA1c levels ranged from 0.5%-0.7% and 0.5%-0.6%, respectively [39,40]. Long-term outcome trials are currently ongoing to determine whether gliptins reduce cardiovascular risk in T2DM [41]. The approved doses are 25-100mg, 50-100mg, 2.5-5mg and 5mg for Sitagliptin, Vildagliptin, Saxagliptin and Linagliptin, respectively. Due to the glucose dependent effect of gliptins they are unlikely to cause profound hypoglycemia. Apart from manageable side effects such as nausea and headache gliptins display fair tolerability; anyway gliptins may be associated with a higher risk for upper respiratory infections and pancreatitis which is not finally resolved [38]. Care must be taken if these drugs are given to diabetics with renal disease; renal elimination of the drugs may lead to accumulation and clinical experience is limited. However, Sitagliptin may be given at a dose of 25mg per day even if patients have end stage renal disease or a creatinin clearance less than 30ml/min; Linagliptin is hardly renally eliminated and can be given to patients for whom Metformin is contraindicated due to renal impairment.
Alogliptin was not part of this investigation since it did not meet the eligibility criteria (see 3.1.1).

Metformin

Metformin exerts its pharmacodynamic effect by inhibiting hepatic glucose production and improving peripheral insulin sensitivity. Moreover, it positively effects dyslipidemia and activated hemostasis two pathological processes often present in diabetics which may contribute to its protective cardiovascular effect [42]. Its mode of action does not seem to be associated with extensive hypoglycemia or weight gain [3]. In placebo controlled trials Metformin reduced mean HbA1c levels by 0.9% ($CI_{0.95}$:[0.6;1.1]) [43]; the glucose-lowering effect is maintained even after a follow-up of 10 years; in the UKPDS-34 study the difference in means in obese patients receiving Metformin compared to diet alone was -0.6%. For now, Metformin is the only OAD reducing the incidence of macrovascular complications [35].

Trialed against placebo, Metformin reduced the incidence of any diabetes-related endpoints (including macrovascular and microvascular complications), diabetes-related death, all-cause mortality and myocardial infarction by 0.32% (RR:0.68; $CI_{0.95}$:[0.58;0.87]), 0.42% (RR:0.58; $CI_{0.95}$:[0.37;0.91]), 0.36 (RR:0.64; $CI_{0.95}$:[0.45;0.91]) and 0.39% (RR:0.61; $CI_{0.95}$:[0.41;0.89]), respectively; compared with an active control consisting of Chlorpropamide, Glibenclamide or Insulin, Metformin significantly reduced the risk of any diabetes-related endpoints, all-cause mortality and stroke (RR:0.74 (per 1000 patient-years); p=0.0034, RR:0.72; p=0.021 and RR:0.53; p=0.032, respectively). The approved dose ranges from 500-3000mg per day. According to the Summary of Product Characteristic (SmPC) gastrointestinal side effects like nausea, vomiting and diarrhea are common (> 10%) but can be handled by careful dose escalation. A rare but often fatal complication is lactate acidosis (< 0.01%). Metformin is contraindicated in patients with renal dysfunction (creatinin clearance < 60ml/min) a complication which occurs frequently in diabetics, especially within the geriatric patients. Therefore, Metformin is not suited for a considerable share of patients; there is some evidence that Metformin can even be used in patients with a creatinin clearance > 30ml/min [44]. However, this approach is not supported by clinical trials with renal safety endpoints; thus, in this patient collective, Metformin should be administered with caution and renal function must be monitored.

Insulin secretagogues –sulfonylureas: Glibenclamide, Gliclazide, Glimepiride, Glipizide, Tolbutamide (Gliquidone, Chlorpropamide)

Sulfonylureas promote their glucose lowering effect by stimulating insulin secretion through inhibition of potassium dependent channels of the beta cells of the pancreas. As a consequence they tend to induce hypoglycemia and weight gain and are more prone to induce secondary therapeutic failure due to increasing insulin deficiency of the beta cells compared to other OAD [3,45]. All sulfonylureas reduce glycemia similarly; a meta-analysis yielded a placebo subtracted mean change in HbA1c of -1.51% ($CI_{0.95}$:[-1.78;-1.25]) [23,46]. As mentioned before, the treatment with Chlorpropamide and Glibenclamide seems to be associated with reduced incidence of microvascular complications and after 10 year follow-up the risk reduction became even significant for macrovascular complications such as myocardial infarction and death from any cause due to an increased number of events [34,47]; however the results should be interpreted with caution since the study has been criticized for several reasons. For example, patients were randomized to an active control group consisting of Chlorpropamide, Glibenclamide and Insulin and compared to therapy with diet alone; thus the assignment was rather adequate for comparing intensive vs. conventional therapy and not for inference about certain drugs. The validity of results may be challenged due to several amendments of power calculation and specification of endpoints and it cannot be ruled out definitely that reporting was driven by intermediate results. Moreover, co-medication and additional glucose lowering therapies beyond therapy failure have not been assessed rigorously enough and particularly after such a long follow-up it is difficult to draw conclusions about the initially assigned therapies [1]. The administered doses are 1.75-10.5mg, 30-120mg, 1-6mg, 2.5-40mg and 250-300mg per day for Glibenclamide, Gliclazide, Glimepiride, Glipizide and Tolbutamide, respectively. Frequent side effects are nausea, emesis, weight gain and hypoglycemia. However, hypoglycemia is often caused by medication errors (overdose, omitting a meal, alcohol consumption, etc.) and the incidence of clinically relevant severe hypoglycemias or major hypoglycemic events (nightly or sustained events) is elevated but baseline risk is rare [33,34]. The excess mortality of the combination of Glibenclamide and Metformin in subgroup analysis of the UKPS-34 raised concerns over the safety of this combination; epidemiological studies show conflicting results [6]; even if this question is not finally settled this combination should be avoided.

Gliquidone and Chlorpropamide were not part of this investigation since retrieved studies did not meet the eligibility criteria (see 3.1.1).

Insulin secretagogues – meglitinides (glinides): Repaglinide, Nateglinide

Like sulfonylureas meglitinides act as betazytotropic substances and increase insulin secretion. However, onset of action occurs earlier and half-life is shorter compared to sulfonylureas; this phenomenon leads to fair control of postprandial glucose and their flexible dosing makes them an attractive option for food uptake adapted therapy and for patients with irregular eating behaviour; on the other hand, due to their insulinotropic effect they are likely to cause hypoglycemia and weight gain, just as sulfonylureas [3,6]. Compared to placebo, Repaglinide reduced mean HbA1c levels by 1.08% ($CI_{0.95}$:[0.43;1.73]). Repaglinide may be more beneficial than Nateglinide; in a head to head trial the difference in means was -0.53% ($CI_{0.95}$:[-0.93;-0.13]) [48]. There is no evidence for a benefit of meglitinides on micro- or macrovascular outcomes. The approved doses are 0.5-16mg and 180-540mg per day for Repaglinide and Nateglinide, respectively. The side effect profile is similar to that of the sulfonylureas: main adverse events are gastrointestinal disorders, hypoglycemia and weight gain, the latter less pronounced compared to sulfonylureas [6].

Thiazolidindiones: Pioglitazone

Pioglitazone is a peroxisome proliferator-activated receptor-gamma (PPAR-gamma) agonist. It enhances insulin sensitivity („insulin-sensitizer“) of the target organs like muscle-, liver- and fat cells resulting in an inhibition of hepatic glucose production, a reduction of abdominal fat tissue and an increased glucose uptake and glycogen synthesis [3]. Its mode of action and effect on lowering HbA1c is promising whereas it clinical safety profile and recent pharmacovigilance data limit a widespread use. The reduction of HbA1c for Pioglitazone lies within the range of 0.5-1.4% [6]. Clinical data regarding cardiovascular outcomes raises doubts about a positive benefit-risk balance of Pioglitazone. In a randomized controlled trial Pioglitazone 45mg per day vs. placebo in addition to baseline antihyperglycemic medication did not affect the primary endpoint positively (composite of all-cause mortality, non-fatal myocardial infarction (including silent myocardial infarction), stroke, acute coronary syndrome, endovascular or surgical intervention in the coronary or leg arteries, and amputation above the ankle; HR:0.90; $CI_{0.95}$:[0.80;1.02]; p=0.095) [49]. A secondary main endpoint (composite of all-cause mortality, myocardial infarction and stroke), which was not specified in the study protocol was reported to yield a statistical significant benefit in favour of Pioglitazone vs. placebo (incidence: 11.6 vs. 13.6%; HR: 0,84; $CI_{0.95}$:[0.72;0.98]; p=0.027; number needed to treat (NNT) =50), whereas the risk of developing heart failure was elevated in patients taking Pioglitazone (incidence: 10.8 vs. 7.5%; p<0.0001; number needed to harm (NNH)= 30). In a subgroup analysis of patients with previous myocardial infarction Pioglitazone had a beneficial effect on the pre-specified endpoint of fatal and non-fatal myocardial infarction (incidence: 5.3 vs. 7.2%; HR:0.72; $CI_{0.95}$:[0.52;0.99]; p=0.045; NNT=53). Anew, Pioglitazone increased the risk of heart failure (incidence: 13.5 vs. 9.6%; HR: 1.43; $CI_{0.95}$: [1.13;1.81]; p=0.003; NNH= 26) [50]. The results must be scrutinized; as mentioned, in the core publication the main secondary endpoint was not pre-specified and analyzed post-hoc; thus a reporting bias driven by favourable results for some arbitrarily chosen endpoint cannot be excluded. The benefit-risk ratio seems to be unfavourable since the NNT exceeds the NNH in both studies. Moreover, in the second study randomization was not stratified by previous myocardial infarction to ensure that potential effect modifiers are distributed evenly among intervention groups. In addition, the authors do not state whether they adjusted for multiple testing; since the reported p-value displays marginal statistical

significance the inference may be misleading due to an uncorrected inflation of the alpha error. Based upon these results the Federal Joint Committee denied further reimbursement of Pioglitazone for SHI insurants. The approved dose range is 15-45mg per day. Side effects apart from heart failure are weight gain, fluid retention and edema and an increased incidence of peripheral fractures. Recently, pharmacovigilance data from the Food and Drug Administration (FDA) found a potential association between the exposure to Pioglitazone and bladder carcinoma; for the whole cohort there was a numerical trend towards a higher risk of incident bladder carcinoma in patients taking Pioglitazone compared to controls (HR: 1.2; $CI_{0.95}$: [0.9;1.5]) , which reached statistical significance in the stratum of patients being exposed to the drug for more than 2 years [51].
Rosiglitazone was not part of this investigation since it did not meet the eligibility criteria (see 3.1.1).

Other drugs: Colesevelam, Dapagliflozin

Colesevelam is is a bile acid sequestrant which binds bile acids in the intestinal tract. Its glucose lowering efficacy is modest with a reduction of HbA1c levels of approximately 0.5%; however it was only investigated as add-on therapy in patients who maintained their pre-existing antidiabetic regimen. Dapaglifozin is the first agent of a novel class of OAD known as sodium-glucose co-transporter-2 (SGLT2) inhibitors. Its mode of action is insulin independent; by inhibiting the SGLT2 protein in the kidneys, Dapagliflozin reduces renal glucose reabsorption, leading to urinary glucose excretion and a blood glucose reduction [52]. According to the pivotal studies submitted for approval its efficacy is moderate; mean HbA1c levels were reduced by 0.5%-0.7%; data on clinically relevant endpoints isn´t available yet [53]. In patients with moderate renal insufficiency Dapagliflozin loses its glucose lowering effect. Frequent side effects are infections of the urinary tract. Concerns about hepatotoxicity and a potential association with an increased incidence of bladder and prostate carcinoma led to rejection of approval by the FDA until these safety issues are answered by the manufacturer [54,55].
Both drugs were not part of this investigation. Colesevelam did not meet the eligibility criteria and Dapagliflozin was only approved in Europe after the updated literature research [53] (see 3.1.1).

1.5 Clinical guidelines for treatment of type 2 diabetes mellitus with oral antidiabetic drugs as monotherapy

The recommendations of the different clinical guidelines relating to the decision which OAD should be used initially to start monotherapy in patients with T2DM are reflected in the corresponding therapy algorithms. There is consensus that Metformin is the standard of care as initiating agent due to its beneficial effects regarding cardiovascular protection, especially in overweight diabetics [1,23,31,56,57]. Only the clinical guideline from the National Institute for Health and Care Excellence (NICE) suggests considering a sulfonylurea as an option for first-line therapy in not overweight patients or if a rapid response is required due to hyperglycemic symptoms. Though, there is debate about which OAD should be used as initial therapy if Metformin is contraindicated or not tolerated. For example, the Canadian Diabetes Association (CDA) doesn´t favour any specific drug but advises against Pioglitazone due to its disadvantageous safety profile [23]. Some guidelines recommend only a subset of candidates: the NICE guideline mentions only glinides and Acarbose whereas the joint position statement of the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD) names sulfonylureas, glinides, gliptins and Pioglitazone as alternatives to Metformin [31,56].

Interestingly, the German Disease Management Guideline for T2DM provides no consistent recommendation but provides instead to different therapy algorithms: the one issued by the Drug Commission of the German Medical Association (AKDÄ) and the German College of General Practitioners and
Family Physicians (DEGAM) discriminates between drug classes with proven clinical benefit regarding clinically relevant endpoints (sulfonylureas) and without benefit (alpha-glucosidase inhibitors, gliptins and glinides) advocating indirectly sulfonylureas as second-line therapy, whereas the one of the German Diabetes Association (DDG) has no explicit recommendations for second-line therapy and schedules all other OAD equal to another [1].

1.6 Indirect comparisons of health care interventions

In medicine today it is uncommon that there exists solely one therapeutic intervention for patients with a certain condition. In fact, health care professionals often have to make decisions between different competing treatments for the question at hand to select the best option and to ensure optimal care for their patients. Usually, systematic reviews and meta-analyses from randomized controlled trials focusing on pairwise direct comparisons are the most reliable sources of evidence. Ideally, there are one or several RCT including every relevant treatment option so that judgment can be based on direct evidence, which is an unrealistic scenario if the competing interventions are numerous [58]. For example, in clinical trials of new drugs they are often compared to placebo or standard care which is sufficient to gain approval of regulatory authorities. However, demonstrating effectiveness against other relevant competitors and cost-effectiveness by means of health economic evaluation becomes nowadays an important task for market access and successful reimbursement. Since a decision has to be made anyway, in the absence of direct evidence this is when indirect treatment comparisons (ITC) are of great value to enable decision makers to draw the required conclusions. Moreover, if direct evidence is of poor methodological quality or inconclusive indirect comparisons provide an additional piece of information so that the decision has not to rely solely on the questionable direct evidence. Even if direct evidence is conclusive the combination of direct and indirect evidence can strengthen the results of direct comparisons only by increasing the precision of estimates and by lending more external validity to the analysis due to a broadened population sample [59]. For a systematic approach it is even mandatory to encompass the whole body of evidence by considering indirect evidence, too. Anyway it is a pivotal issue to decide under which circumstances direct and indirect evidence should be combined (see chapter 3.2.3). Another disadvantage of classical meta-analysis is that results are limited to pairwise comparisons making it difficult to come to an overall conclusion which treatment is the best if several treatment options are available. The methods underlying indirect comparison are able to produce such rankings and probability statements which treatment is the best and can ease decision making across a range of interventions. Thus, indirect comparisons are an important tool in the field of evidence research synthesis and are becoming more popular a fact that is reflected by the increasing number of publications of systematic reviews using statistical methods of indirect comparisons [60].
In the simplest case intervention A has been compared to intervention B in a A-B trial and to intervention C in an A-C trial but there is no B-C trial; thus, an estimate of the relative effect of C vs. B d_{BC} can be derived from the indirect comparison of the effect of B vs. A (d_{AB}) and the effect of C vs. A (d_{AC}) via the common comparator A (see Figure 1, first network on the left)

$$d_{BC} = d_{AC} - d_{AB} \tag{1.1}$$

Since the relative treatment effects add together the appropriate scale of measurement must be chosen (e.g. the outcome measure must be normally distributed such as odds ratios, hazard ratios and relative risks on a log scale or difference in means) [58].

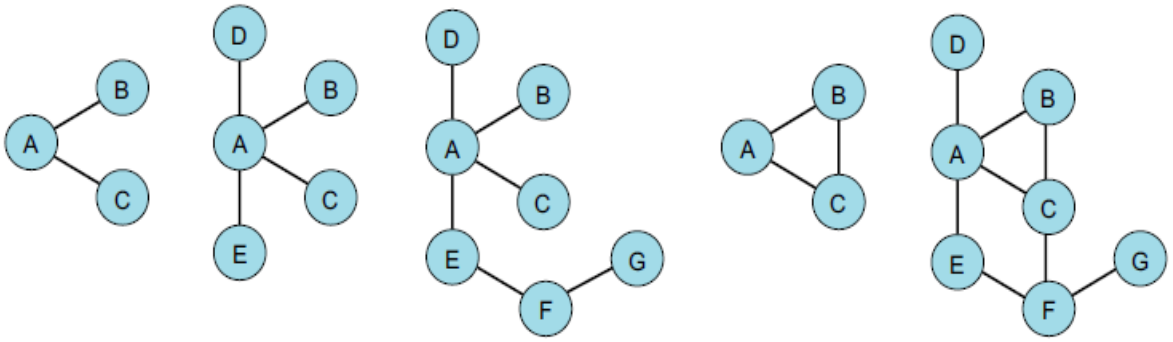


Figure 1: Structures of different evidence networks of network meta-analysis [59]

The indirect comparison is anchored on treatment A as bridging comparator, what is called an adjusted indirect comparison, too. It is fundamental that indirect comparisons do not break randomization. An unadjusted comparison of B vs. C which is only based on the results of the B arm of the A-B trial and the C arm of the A-C trial also called “naïve indirect comparison” ignores randomization and is no more valid than a comparison based on single arm from observational studies and should therefore be avoided [61]. A common critique about indirect comparisons is the assumption that they suffer from observational biases. As long as the indirect evidence is based on relative treatment effects and randomization is preserved they suffer from unobserved effect modifiers since patients are randomized within trials but not across trials, just as pairwise meta-analyses do. Both give unbiased estimates as long as the effect modifiers have similar distributions across pairwise comparisons or across the whole set of comparisons in the network (see chapter 3.2.3). Thus, both are superior compared to observational studies since they are based on RCT [62].
As long as the network is connected (e.g. each contrast has a path from one to the other) an indirect comparison of each intervention with any other is possible (see 2nd and 3rd network, Figure 1), although comparisons with longer paths will have less precision [59]. The fourth and fifth network in Figure 1 contains closed loops. In the fourth network the comparison B-C has direct evidence from the B-C trial and indirect evidence from the A-B and A-C trials; in the fifth network there are two closed loops, triangle ABC and diamond ACEF, both supported by direct and indirect evidence. Networks containing such loops are called mixed treatment comparisons (MTC) or multiple treatment meta-analysis (MTM). For such networks an important assumption for pooling all the evidence is that indirect comparisons are consistent with direct comparisons (see chapter 3.2.3). Sometimes there is confusion about the terminology regarding indirect comparisons. This work refers to the proposed definitions by Jansen et al., 2011 (see Figure 1): a network meta-analysis (NMA) is any synthesis where the

evidence base comprises two or more than two RCT connecting more than two interventions. An MTC or MTM consists of at least one closed loop in the network, for any other analysis of an open loop network the term ITC is appropriate [59].

1.7 Comparative effectiveness research within the framework of benefit assessment, reimbursement and health economic evaluation

Since pressure on efficient allocations of resources within health care systems is rising there is an demanding increase for comparative effectiveness research including the whole set of relevant treatments which in turn lays the foundation for health economic evaluation of these interventions. For instance, the Cochrane Collaboration introduced a new form of reviews assessing the relative effectiveness of several interventions for the same condition and established a methods group to enhance diffusion of this technology [63].
In the United Kingdom the NICE has a strong record of technical appraisal of new health care interventions; in its guide to the methods of technology appraisal it embraces MTC if it adds information that is not available from head-to-head comparisons or if direct evidence is lacking at all [64]. Moreover, the NICE maintains the Decision Support Unit an institution supporting the institute's technology appraisal program by providing useful technical support documents for evidence synthesis, utilities and other topics [65]. Other countries for instance the United States, Canada or Australia are undertaking similar efforts to improve efficiency of their health care delivery; thus their evidence bodies and health technology assessment (HTA) authorities appreciate methods of indirect comparisons in their methods guides, too [66-68].
In Germany, with the Act on the Reform of the Market for Medicinal Products (AMNOG) a benefit assessment of pharmaceuticals according to German Social Code V (SGB V) was introduced. Within three months after market authorisation the market authorisation holder (MAH) has to submit a dossier to the GBA who is in charge of conducting the benefit assessment - usually by commissioning the IQWiG. This assessment is the basis for the negotiations of reimbursement prices between the SHI and manufacturers. In case of disagreement a cost-benefit assessment can be initiated to derive adequate pricing levels of pharmaceuticals [69]. In this setting, both the GBA in its rules of procedure as well as the IQWiG in their methods papers on benefit assessment and on health economic evaluation acknowledge the need for indirect comparisons. For health economic evaluation the IQWiG admits that assessments are generally impractical without indirect comparisons. For benefit assessment the institute concedes that consideration of indirect comparisons may be appropriate but its use automatically downgrades the certainty of findings. Only adjusted indirect comparisons are considered and the homogeneity of pairwise comparisons as well as the consistency in the network has to be assessed. Furthermore the statistical model and code and the underlying assumptions have to be fully reported [69-71]. Already two oral antidiabetic drugs, Linagliptin and Dapagliflozin have undergone the process of benefit assessment in Germany. For both drugs the IQWiG concluded that due to a lack of evidence benefit with respect to the appropriate comparator could not be established. The MAH could have overcome this issue by making use of indirect comparisons to meet the requirements of the dossier, which wasn´t done. As a consequence Linagliptin will not be accessible for diabetic patients in Germany.

2. Research objective

T2DM is increasingly becoming a global pandemic and especially in developed countries increasing levels of obesity and sedentary lifestyle enhance its dissemination. The burden of disease in terms of micro- and macrovascular complications, loss of HRQoL and premature death as well as the socioeconomic burden is substantial. Therefore, identifying effective strategies including lifestyle interventions and pharmacological treatment that control diabetes and its complications are strongly needed to improve outcomes of affected patients and to keep health care expenditures manageable. Clinical guidelines tend to provide insufficient guidance on the sequential therapy when T2DM progresses. For example most guidelines don´t have clear preferences for combination therapy regarding the second or third antihyperglycemic agent [1,23,31]. However, there is evidence from MTC which double or triple combination therapy might be favoured [72,73]. For oral monotherapy clinical guidelines agree that Metformin is the gold standard but no statements are made which OAD should be used second line if Metformin is contraindicated or not well tolerated. This is an important question since gastrointestinal intolerance may limit its use in some patients and a substantial share of patients with T2DM already suffers from chronic kidney disease at time of diagnosis and can´t be treated with Metformin [13,74]. The lack of recommendation may be attributed to the fact that not all OAD have been trialed directly against each other and no indirect comparisons have been made yet. Moreover, it is unfortunate that patients don´t get access to new drugs due to the inability or reluctance to use indirect comparisons for benefit assessment providing the rationale for this work. Thus, the objective of this thesis is a systematic review and mixed treatment comparison assessing the comparative effectiveness and safety of oral antihyperglycemic drugs as monotherapy in patients with type 2 diabetes mellitus.

3. Methods

Although there exists more literature about conducting systematic reviews especially regarding different quality appraisal tools for assessing the risk of bias of studies, the single steps of this systematic review including the definition of the eligibility criteria, the search for records, the study selection and data collection and the assessment of the risk of bias were performed in close dependence on the Cochrane Handbook for Systematic Reviews of Interventions, since it represents one of the most elaborated and sophisticated guides for this kind of research. Unless otherwise stated, the section 3.1 of this work refers to the Cochrane Handbook Part 2 Chapter 5-8 [75]. Furthermore the conducting and reporting of the systematic review, meta-analysis and mixed treatment comparison will follow the methods of the PRISMA (preferred reporting items for systematic reviews and meta-analyses) statement and the guidance of the ISPOR (International Society of Pharmacoeconomics and Outcome Research) task force on indirect treatment comparisons, respectively [76,77].

Since reviews are by nature retrospective the used methods should be established and documented prior to the review process. The publication of a review protocol reduces the impact of review authors' biases and provides transparency of methods and processes. Every effort should be undertaken to comply with the predetermined protocol. However, changes in a review protocol are sometimes necessary due to unanticipated circumstances such as issues with data collection or reporting of data and justifiable, as long as they are not made on the basis of how they affect the outcome. Thus, amendments of the protocol and their underlying rationale should be reported in the review. Section three is written in the form of such a protocol pre-specifying the applied methods of systematic review and the statistical analysis plan.

3.1 Systematic Review

3.1.1 Defining the scope of the review by eligibility criteria

A systematic review in contrast to a narrative approach is characterized by a focused and clearly defined objective: the research question. This research question (see 2.) translates into pre-specified eligibility criteria. Based on them the inclusion and exclusion of studies for the review is accomplished. The PICOS (participants, interventions, comparisons, outcomes and study design) scheme was applied to cover all relevant aspects of the research question and develop the single eligibility criteria. A systematic review should contain all relevant outcomes focusing on 2-3 primary endpoints including beneficial effects and adverse events. However, the choice of endpoints determines how broad or narrow the scope of the review gets; the broader the eligibility criteria are defined the bigger is the evidence base to be included and the higher is the external validity in terms of generalizability of findings; on the other hand heterogeneity might be increased due to mixing „apples and oranges" which makes interpretation of results difficult (for narrowing the eligibility criteria, it is just the other way around). Thus, the analyst has to make a tradeoff between these opposites, has to balance their assets and drawbacks pragmatically and whatever approach is taken, must be transparent by pointing out how decisions are derived.
The eligibility criteria will be piloted against a sample of reports (including ones that are thought (not) eligible) to refine and clarify them if necessary.

3.1.1.1 Participants

Studies will be eligible for inclusion if they were conducted in adults aged 18 years or older with type 2 diabetes. To be consistent with changes in classification and diagnostic criteria of type 2 diabetes mellitus through the years, the diagnosis should have been established using the standard criteria valid at the time of the beginning of the trial (see 3.1.3.3).

3.1.1.2 Interventions/ Comparators

Studies will be valid for inclusion if patients were assigned to oral monotherapy (non-pharmacological lifestyle measures such as diet or physical activity were allowed; previous therapy itself was no criterion for exclusion as long as all other drugs than the interventions were discontinued at time of randomization) and one of the following drugs holding a market authorisation within the European Union or the United States alone were compared with each other or with placebo: Metformin, sulfonylureas (Gliclazide, Gliquidone, Glimepiride/Glyburide, Glibenclamide, Chlorpropamide, Glipizide, Tolbutamide or Tolazamide), Pioglitazone, meglitinides (Repaglinide or Nateglinide), alpha-glucosidase inhibitors (Miglitol or Acarbose) or DPP-4 inhibitors (Sitagliptin, Vildagliptin, Linagliptin, or Saxagliptin). Rosiglitazone will not be considered since it has been withdrawn from the market (association with an increased risk of myocardial infarction and cardiovascular death). Colesevelam was only investigated as combination therapy and is therefore not meeting the eligibility criteria. Dapagliflozin was only approved (2012/11/12) after the updated literature research (2012/07/30) and hence could not be considered.

One vital question regarding the definition of nodes in a network is whether treatments, especially drugs of the same class should be pooled or if each single drug should be analyzed separately. The lumping of treatments seems justified if the treatment effects of drugs within a drug class are comparable enough to make the assumption of a group effect. A benefit is that in a complex network with many nodes assessing drug classes rather than single drugs eases reading and conveying of key findings and results become more precise. On the other hand, if the treatment effects within a drug classes differ considerably and the assumption of a group effect may not hold, analyzing drug classes can introduce an amount of heterogeneity into the network which may complicate interpretation of results.
Based upon evidence from the existing literature in this thesis the assumption is made that the treatment effects of the single sulfonylureas, meglitinides, alpha-glucosidase inhibitors and DPP-4 inhibitors are comparable enough to analyze them as drug classes [23,37-40,46,48]; moreover, in the stage of conceptualization, either seven treatments (drug classes) or 20 treatments (single drugs) would yield an outcome matrix of T*(T-1)/2= 21 or 190 relative effect estimates, the latter making interpretation of results almost impractical. As a consequence, studies comparing drugs from the same class against each other are not eligible for inclusion.
Another method applied by the IQWiG in its first health economic evaluation of antidepressants is to use a top down approach [78]. The IQWiG rejects the assumption of drug classes if measures of heterogeneity of pairwise meta-analyses exceed a threshold value (p-value for heterogeneity < 0.2 or I^2>50%). If heterogeneity in meta-analyses of single drugs is still present the IQWiG excludes this evidence trying to keep the amount of heterogeneity in the network low. This approach seems too restrictive since it downsizes the whole body of evidence. Anyway, careful attention will be given to this issue by assessing heterogeneity in pairwise meta-analyses and MTC by the corresponding statistical measures and by meta-regression.

In line with the aggregation of similar drugs into drug classes for trials including arms with different doses of the same drug these arms were aggregated into one for data entry (see 3.1.3.4). Dose dependent effects were not modeled, but overly high or low doses of the single drugs which may bias the results were subjected to sensitivity analysis.
From multi arm studies, only the eligible intervention arms were selected for inclusion. Single eligible arms were not included since they will break the principle of randomization of pairwise comparisons.

3.1.1.3 Outcomes

As mentioned before it would have been desirable to choose clinically relevant endpoints such as mortality (all-cause mortality or diabetes-related death), morbidity (impact on micro- and macrovascular complications) or HRQoL. However, due to the lack of evidence there are almost no studies that have investigated these endpoints [34,35,49]. The network structure would only consist of a few RCT and neglect the majority of diabetes drug trials that had other endpoints. Therefore studies were entitled to inclusion if they reported all of the three following outcomes: HbA1c, body weight and hypoglycemia.
Hypoglycemia was selected as third outcome since especially severe, prolonged and nightly events cause recurrent morbidity and are sometimes fatal [79]. It has to be acknowledged that not all treatments in the analysis are prone to cause hypoglycemia (e.g. gliptins, Metformin or alpha-glucosidase inhibitors). For those other treatment-specific adverse events like gastro-intestinal side effects (Metformin or alpha-glucosidase inhibitors) or upper respiratory infections (gliptins) are primarily of interest but these endpoints in turn are not relevant for other treatments. One alternative might be the frequency of adverse events (AE) and severe adverse events (SAE) and their corresponding discontinuation rates, respectively; this data was extracted but these safety endpoints are less frequent reported as hypoglycemias; moreover, they suffer from the same concerns as hypoglycemia (see below) and reporting is selective due to different thresholds of incidence of AE/SAE in the included studies (AE/SAE below these thresholds are usually not reported). Finally this safety endpoint is not very informative since it aggregates all kinds of AE/SAE and definition varies significantly between trials. Taken together, hypoglycemia appears to be the most reasonable trade-off. The rationale for including only studies that report all three outcomes simultaneously is that these studies may be less affected from selective reporting bias [80,81], acknowledging that this ruling is put up with a loss of evidence. Nevertheless this cumulative condition of reporting outcomes for inclusion allows for external cross-validation against some of the excluded studies for contrasts that are not supported by direct evidence in the network (see 3.2.3.6).

HbA1c

As outlined before HbA1c is a well analyzed surrogate for clinically relevant endpoints and accepted by regulatory authorities, even if the correlation with the relevant endpoint has not been demonstrated for every single drug [82]. Studies will be eligible for inclusion if they report the mean change from baseline along with standard error of the mean (SEM) for each eligible intervention arm (for transformation of other summary data and imputation, see 3.1.3.4).

Body weight

Weight gain is an undesired adverse effect in diabetics since dyslipidemia is associated with an increased risk for micro- and macrovascular complications. Studies will be eligible for inclusion if they report the mean change from baseline in kilogram (KG) and the SEM for each eligible intervention arm (for transformation of other summary data and imputation, see 3.1.3.4).

Hypoglycemia

Unfortunately the reporting of hypoglycemic episodes - despite existing guidelines and recommendations of regulatory bodies [79,82] – lacks consistency and completeness and differs considerably between trials. According to the working paper of the ADA working group on hypoglycemia data should be reported as followed: 1.) frequency: the workgroup recommends that both the proportion (percentage) of patients experiencing at least one episode and the event rates (e.g., episodes per patient-year) for each of the categories of hypoglycemic events should be reported; 2.) categories: to consider differences in severity of hypoglycemic episodes the ADA suggests that events should be reported separately for each of the five categories (see Table 4). At a minimum, hypoglycemic events should be reported in each of the first three categories. Thus, data was extracted for each category separately. As it was already foreseeable upfront that due to the scarcity of data within each category and the fact that the categories of clinical relevance (1 and 2) were rare in all trials (most trials reported zero event rates) and that some events as reported by the investigator could not be classified into one of the category because of different definitions, the original approach to analyze this outcome for each category had to be abandoned. Moreover, only a few studies reported episodes per patient-years in addition to the proportion of patients experiencing at least one episode, thus unfortunately, count data which would have accounted for multiple episodes could not be considered as unit of analysis.
Therefore studies will be eligible for inclusion if they reported the number of patients suffering at least one episode of any hypoglycemia for each eligible intervention arm; the trial specific definition of a hypoglycemic episode will be detailed to facilitate interpretability.

Table 4: Categories of hypoglycemia [79]

Category	Definition
1.) Severe hypoglycemia	An event requiring assistance of another person to actively administer carbohydrate, glucagons, or other resuscitative actions.
2.) Documented symptomatic hypoglycemia	An event during which typical symptoms of hypoglycemia are accompanied by a measured plasma glucose concentration $\leq$ 70 mg/dl (3.9 mmol/l)
3.) Asymptomatic hypoglycemia	An event not accompanied by typical symptoms of hypoglycemia but with a measured plasma glucose concentration $\leq$ 70 mg/dl (3.9 mmol/l).
4.) Probable symptomatic hypoglycemia	An event during which symptoms of hypoglycemia are not accompanied by a plasma glucose determination (but that was presumably caused by a plasma glucose concentration $\leq$ 70 mg/dl (3.9mmol/l).
5.) Relative hypoglycemia	An event during which the person with diabetes reports any of the typical symptoms of hypoglycemia, and interprets those as indicative of hypoglycemia, but with a measured plasma glucose concentration > 70 mg/dl (3.9 mmol/l).

Regarding the time frame studies will be considered for inclusion if the duration of the intervention (follow-up of trial drug exposure) was at least 12 weeks. Due to the progressive course of disease and the majority of studies having a follow-up between 12 and 36 weeks studies lasting longer than three years were excluded to ensure a certain degree of comparability within the study set (even though meta-regression was performed to account for potential effect modification by the covariate length of follow-up). Due to the limited glycemic durability of some of the drugs, longer studies may bias the efficacy of the results downwards.

3.1.1.4 Study design

Studies will only be considered for inclusion if they were randomized controlled trials. Cross-over studies were eligible too, if patients were randomized to intervention arms and a wash-out phase of at least 8 weeks was planned to avoid potential bias induced by time by treatment interaction and carry-over effects, respectively.

3.1.2 Searching for records

3.1.2.1 Sources to search

To design a comprehensive literature search the following sources were searched: 1.) bibliographic databases: Medline, Embase and the Cochrane Central Register for Controlled Trials (CENTRAL); 2.) clinical trials registers: ClinicalTrials.gov register and trials registers of the market authorisation holder of the relevant drugs; 3.) reviews of the assessment reports of the regulatory authorities FDA and the European Medicines Agency (EMA). Since the screening of all retrieved records is time consuming an update of the initial search will be done to capture the records that appeared in the meantime. The completeness of the retrieved records will be validated against a randomly drawn test set of trials fulfilling the search criteria.

Bibliographic databases

Medline

Medline is the U.S. National Library of Medicine premier bibliographic database from the National Institutes of Health. It can be accessed by several interfaces like Pubmed.
To date it contains more than 22 millions of records and more than 5,200 journals in 37 languages are indexed in Medline.

Embase

Embase is a biomedical bibliographic database run from the publishing company Elsevier; it can be accessed via the interface Ovid. Embase currently contains over 11 million records from 1974 onwards. Currently 4,800 journals are indexed for Embase in 30 languages.

CENTRAL

CENTRAL, which is part of the Cochrane Library serves as a comprehensive source of reports of controlled trials. It contains more than 530,000 citations a lot of them stemming from Medline and Embase.

Even if there is a huge overlap between the three sources it is unsatisfactory to search only one database. For example of the 4,800 journals in Embase more than 1,800 are not indexed in Medline. The other way around of all the journals in Medline more than 1,800 are not indexed in Embase. Moreover, approximately a third of records in CENTRAL are from other databases or hand-searching. Thus a comprehensive literature search should at least contain these three important databases.

Clinical trial registers (public and pharmaceutical industry)

Clinical trials register become an increasingly important part of the literature research. For example the ClinicalTrials.gov register contains all clinical studies conducted in the United States that require registration by law. Currently, more than 143,500 studies are listed [83]. Nowadays, expanded requirements make more trial information and results available from the database. Unfortunately the European Union lags behind and the publicly accessible parts of the clinical trial registers like EudraCT, the EU Clinical Trials Register and EudraPharm provide no relevant trial information [84-86]. Thus, in addition the trial registers from the following pharmaceutical companies were searched for records, too: Astra Zeneca, Bristol Myers Squibb, Eli Lilly, GlaxoSmithKline, Novartis and Roche [87-92].

Regulatory approval authorities

The EMA and FDA provide explicit information about the drug approval process on their homepage [93,94]. For every drug licensed in the U.S. the FDA provides access to the clinical and statistical reviews; likewise the European public assessment reports (EPAR) provide approval information about drugs holding a central market authorisation within the EU.

3.1.2.2 Defining a search strategy for bibliographic databases

For developing a search strategy it is not necessary to search for every aspect of the research question. It is common to divide the search in three domains: the condition of interest, the intervention of interest and the study design. These three domains are usually connected by the Boolean operator AND whereas a variety of different search terms within each domain are combined with the Boolean operator OR.
Searches for systematic reviews should be as extensive as possible to identify every relevant record; anyway it is necessary to balance comprehensiveness against efficiency and maintaining relevance by developing the search strategy. The quality of a literature search is defined by its sensitivity and precision. Sensitivity is defined as the number of relevant reports identified divided by the total number of relevant reports in existence and measures the degree of comprehensiveness. Precision is defined as the number of relevant reports identified divided by the total number of reports identified and measures the degree of relevance. Increasing the comprehensiveness (or sensitivity) of a search will reduce its precision and will retrieve more irrelevant articles. Anyway for systematic reviews it is accepted to apply a highly sensitive search strategy. Thus, validated and highly sensitive filters for the domain study design (RCT) were applied for Medline and Embase [95,96]. Furthermore, no limits regarding the time period or languages were set. With the help of native speakers records not published in English were also considered for inclusion. However since mistakes with respect to data extraction and bias assessment due to language barriers can´t be ruled out completely these records were excluded from sensitivity analysis. Abstracts without a full text publication were handled in the same way (see 3.1.4.4).

Medline and Embase can be searched using standardized subject terms generated by indexers. This thesaurus named MeSH (medical subject heading) in Medline and CENTRAL as well as Emtree in Embase is helpful to identify articles that may use different words to describe the same aspect going beyond the content of the title and abstract. However, one can´t rely only on this controlled vocabulary since authors not always are explicit in their scientific notation, indexers may not be familiar with every subject and not all search concepts are matched by the indexing terms. Thus, it is crucial to combine indexing and free text terms for a broad search approach. Since CENTRAL exhibits a considerable overlap with records from Medline and Embase it is possible to exclude these sources with an appropriate operator. The full search for each source including date, search strategy, syntax guide and retrieved records will be detailed in the appendix.

3.1.2.3 Managing references

The reference management software Endnote® Version X.4.0.2 was used to import and merge retrieved records with abstracts and titles from each database by customized filters [97]. Duplicate records were removed by using a filter and by manual screening. Each record was indexed by a unique record number.

3.1.2.4 Missing data

In case that outcome data is not reported in the desired format (e.g. body-mass index instead of kg or median change instead of mean change), only partly reported (e.g. outcome data only reported for some arms or the mean change reported without the measure of uncertainty) or that information is missing to judge if the study meets the eligibility criteria (e.g. ongoing antidiabetic medication apart from trial drugs) when possible the first author was addressed by email to obtain the requested information.

3.1.3. Study selection and data collection

In the review process the decision which study to include is very crucial. Even if eligibility criteria are clearly defined it involves individual judgement of the reviewer. Thus, the steps of assessment of inclusion, data collection and assessment of the risk of bias should always be undertaken by two independent reviewers for two reasons. First, careless errors that have just been overseen can be detected more easily; second if agreement can´t be achieved disagreement can be settled by a third investigator. For this thesis all steps mentioned above (except from the search of the clinical trials registers and regulatory bodies which will be conducted only by one reviewer (TB)) will be done independently by two reviewers (TB and SK); in case of disagreement this will be resolved by a third author (UM) (see 3.1.5). To account for the uncertainty included studies based upon judgement from the third author will be excluded from sensitivity analysis (see 3.1.4.4).

3.1.3.1 Screening of studies

After the removal of duplicates in the database of records, these will be screened for compliance with the eligibility criteria. This first screen serves to exclude obviously irrelevant records; at this stage reviewers should be „over-inclusive“ meaning that the eligibility criteria should not be applied to restrictive in order not to miss any relevant record upfront.
For reviews the unit of interest is the study; sometimes there may exist more reports originating from the same study either containing additional information (from other articles, abstracts or reports) or being published again with the same content.

Whereas the former may yield helpful information like missing data the latter can introduce serious bias if identical studies are appearing twice in the review. Therefore, supplementary data from different records (including the records of the search of the web sites of regulatory authorities and clinical trials registers) providing additional information will be linked together and records that are in effect duplicates will be removed.
After identifying the potentially eligible records their full text will be evaluated rigorously and a final decision regarding inclusion will be made. For the whole process of study selection, assessment of eligibility and inclusion, a study flow diagram will be provided. For both steps interrater agreement will be assessed (see 3.1.5).

3.1.3.2 Rationale for excluding studies

An overview of studies that have been assessed for full text but have been deemed ineligible will be provided by giving the primary reason for exclusion. It will encompass all studies that seem to meet the eligibility criteria on first sight but on further inspection do not, and also studies that are not eligible but are well known and likely to be expected among the selected studies.

3.1.3.3 Defining the relevant parameters and their coding for collection of data

After completion of the study selection process it is important to decide which data should be collected and which coding should be used for the different variables. The data that is required for the review is of course outcome data for the pre-specified endpoints, descriptive data to depict included studies and patient characteristics, data of observed covariates that may be potential effect modifiers and data about the methodological quality to facilitate assessment of the risk of bias (see 3.1.4). To ensure a structured and comprehensive data extraction a data collection form was designed. It covers all relevant parameters and their coding or explanation, respectively. A template will be piloted against three randomly selected studies for testing feasibility and consistent coding between reviewers.
For this work the parameters with the corresponding units, coding, measure or explanation that will be extracted are displayed in Table 5.

Table 5: Parameters for data collection

Parameter	**Explanation/ unit/ measure / coding**
Source	
Record identifiyer[a]	*Explanation:* record number assigned in the database of records from literature search (Endnote).
Study identifiyer[a]	*Explanation*: study number assigned for indexing the studies included in quantitative synthesis.
First author	-
Year of publication	-
Study duration	*Explanation:* time of trial drug exposure starting from time of randomization. *Unit:* weeks *Coding:* complete follow-up (core study/ extension).
NCT number[b]	*Explanation:* record the trial number from ClinicalTrials.gov.
Participants	
Number of randomized patients[c]	-

Diagnostic criteria of T2DM[b]	*Explanation:* criteria that were applied to confirm existence of T2DM. *Coding:* 1 ADA 1997 2 ADA 1999 3 WHO 1980 4 WHO 1985 5 WHO 1998 6 Else (state definition) 7 NR
Age[c/d]	*Unit:* years *Measure:* mean, SD
Sex[c/e]	*Explanation:* number of male patients per arm.
Diet[b]	*Explanation:* record if patients got any dietary advice in addition to the intervention. *Coding:* 1 Yes 2 No 3 NR
Duration of existing T2DM[c]	*Explanation:* record how long did the disease already exist at the time patients were enrolled for the trial. *Unit:* years *Measure:* mean, SD
Previous therapy	*Explanation:* record if patients did receive any antidiabetic therapy before the beginning of the study. *Coding:* 1 Therapy naive 2 No OAD at least 8 wks before randomization 3 Else
Baseline HbA1c[c]	*Unit:* % *Measure:*mean, SD
Baseline body weight[c]	*Unit:* kg *Measure:*mean, SD
Comorbidities[b]	*Explanation:* record if the trial was conducted in a population suffering from a special condition. *Coding:* 1 Hypertension 2 End stage renal disease 3 Chronic renal insufficiency 4 Else (state other conditions)
Interventions	
Treatment	*Explanation:* record the intervention with the cumulative dose per day and the dosing frequency: q.d; b.i.d.; t.i.d.. If reported, indicate dosing escalation and mean or median received dose. *Unit:* mg/d *Coding* (dose/d; frequency)
Number of arms[f]	*Explanation:* record the number of eligible intervention arms.

Coding of treatments[f]	*Explanation:* record the single interventions or the drug classes according to the following scheme. *Coding:* A 1 Placebo B 2 Metformin C 3 Sulfonylurea D 4 Pioglitazone E 5 Glinides F 6 Alpha-glucosidase inhibitors G 7 Gliptins
Outcomes	
HbA1c	
Number of analyzed patients[c]	*Explanation:* record the number of analyzed patients per arm.
Population[b]	*Explanation:* record in which population the analysis was done, e.g. ITT analysis, pp analysis, OC analysis or as defined by the investigator. If reported, state the imputation method for missing values.
Mean change from baseline[c]	*Explanation:* record the mean change from baseline for each arm. *Unit:* % *Measure:* mean, SD
Body weight	
Number of analyzed patients[c]	*Explanation:* record the number of analyzed patients per arm.
Population[b]	*Explanation:* record in which population the analysis was done, e.g. ITT analysis, pp analysis, OC analysis or as defined by the investigator. If reported, state the imputation method for missing values.
Mean change from baseline[c]	*Explanation:* record the mean change from baseline for each arm. *Unit:* kg *Measure:* mean, SD
Hypoglycemia	
Population[b]	*Explanation:* record in which population the safety analysis was done.
Patients with any hypoglycemia	*Explanation:* number of patients experiencing at least one event of any hypoglycemia.
Patients with severe hypoglycemia[g]	*Explanation:* number of patients experiencing at least one event of severe hypoglycemia.
Patients with documented symptomatic hypoglycemia[g]	*Explanation:* number of patients experiencing at least one event of documented symptomatic hypoglycemia.
Patients with asymptomatic hypoglycemia[g]	*Explanation:* number of patients experiencing at least one event of asymptomatic hypoglycemia.
Patients with probable symptomatic hypoglycemia[g]	*Explanation:* number of patients experiencing at least one event of probable symptomatic hypoglycemia.
Patients with relative hypoglycemia[g]	*Explanation:* number of patients experiencing at least one event of relative hypoglycemia.
Patients at risk	*Explanation:* number of patients at risk in the safety population.
Definition of hypoglycemia	*Explanation:* definition of hypoglycemia as reported in the study.

Further safety outcomes[b,h]	
Definition of AE	*Explanation:* definition of AE as reported in the study.
Patients with AE	*Explanation:* number of patients experiencing at least one AE.
Patients discontinued due to AE	*Explanation:* number of patients who were discontinued from the trial due to an AE.
Definition of SAE	*Explanation:* definition of SAE as reported in the study.
Patients with SAE	*Explanation:* number of patients experiencing at least one SAE.
Patients discontinued due to SAE	*Explanation:* number of patients who were discontinued from the trial due to an SAE.
Miscellaneous	
Remarks	*Explanation:* state any information important for evidence synthesis.

[a] The record number indexed from Endnote® wasn´t continued. Instead the study number indexing the studies included in evidence synthesis was used.
[b] For clarity purposes, only the most relevant data was displayed in the main table. Additional parameters were listed in a further table.
[c] Data was collected for each eligible arm.
[d] If the desired measure wasn´t reported the confidence interval, standard error of the mean, p-value or t-value was extracted from each intervention group, from the final value or from effect estimates between intervention groups.
[e] In the table trial and patients characteristics of included studies the proportion of men (%) was displayed.
[f] For setting up the data for MTC in a matrix containing one study per row it is necessary to denote the number of arms and to index the different treatments (see 3.2.3).
[g] Due to the scarcity of data for the different categories of hypoglycemia only the events of any hypoglycemia along with the trial specific definition were reported.
[h] Data was extracted but no analysis was performed on these endpoints (see 3.1.1.3).
HbA1c, glycated haemoglobin. T2DM, diabetes mellitus type 2. ADA, American Diabetes Association. WHO, World Health Organization. NR, not reported. SD, standard deviation. OAD, oral antidiabetic agent. Q.d., once daily. B.i.d., twice daily. T.i.d. thrice daily. Wks, weeks. Kg, kilogram. ITT, intention to treat. PP, per protocol. OC, observed cases. AE, adverse events. SAE, severe adverse events.

A table summarizing the most important trial and patient characteristics will be provided; in addition, the remaining parameters that were collected will be detailed in additional table. Amongst other things the trial information regarding the efficacy and safety populations was considered to rate the risk of bias of incomplete outcome data (see 3.1.4.2).

3.1.3.4 Transformation and imputation of data

Ideally for performing the statistical analyses outcome data should be reported as mentioned before (see 3.1.1.3). However this is not always the case, since data is missing or is reported differently. For missing data it is worthwhile to contact the authors to get the lacking information. If no data is provided for some measures like the standard deviation (SD) of the mean change from baseline imputation methods can be applied. Other target measures can be derived by transformation of reported data. Below, the procedure how to handle these issues in this thesis is described.

The standard error of the mean can be derived from the standard deviation and the sample size N.

$$SEM = \frac{SD}{\sqrt{N}} \tag{3.1}$$

If the sample size was not reported, for analysis or necessary transformations requiring the sample size the number of randomized patients or from the intention-to-treat population was taken as an approximation. For group means the standard deviations can be obtained from a confidence interval (3.2) or a p-value (3.3) and equation (3.1). For large sample sizes ($x \geq 60$) the 95% confidence interval is 3.92 standard errors wide ($3.92 = 2 \times 1.96$). For other (1-α) % CI intervals the value 3.92 has to be replaced by the corresponding z quantile of the normal distribution. If the sample size is small ($x < 60$) the quantiles of the heavier tailed t-distribution should be applied; thus the value 3.92 is replaced by $2 * t_{df\,(N-1);(1-\alpha)}$, df indicating the degrees of freedom and 1-α the quantile of the two-tailed distribution for the corresponding level of Type 1 error.

$$SD = \frac{\sqrt{N} \cdot (upper\,\mathrm{lim}\,it - lower\,\mathrm{lim}\,it)}{3.92} \tag{3.2}$$

$$SEM = \frac{MD}{t} \tag{3.3}$$

The t-value that corresponds with the p-value is read from a two-tailed t-table as follows: $t_{df\,(N-1);(1-\alpha)}$.

If no measure of uncertainty was available for the group mean but the effect estimate (MD) with a CI (3.4) or p-value was reported a within-group standard deviation as the average of the intervention arms was calculated.

$$\begin{aligned} a.)\; SEM &= \frac{(upper\,\mathrm{lim}\,it - lower\,\mathrm{lim}\,it)}{3.92} \\ b.)\; SD &= \frac{SEM}{\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} \end{aligned} \tag{3.4}$$

Again, for a small sample size the value 3.92 is replaced by $2*t_{df\,(N_1+N_2-2);(1-\alpha)}$. From a p-value the t-value is derived as follows: $t_{df(N_1+N_2-2);(1-\alpha)}$. Applying formula 3.3 and 3.4b yields the average within-group standard deviation. If a threshold of a p-value (such as $p < 0.05$) rather than the exact value is reported an conservative approach is to take the p-value at the upper limit (e.g. for $p < 0.05$, $p=0.05$); though, this assumption might increase the uncertainty artificially.

For combining continuous data from different arms (e.g. aggregating arms of different doses of one drug or aggregating all study arms to derive study level data for potential effect modifiers in meta-regression) the following formula was used.

Combined groups

$$\begin{array}{ll} \text{Sample size} & N_1 + N_2 \\ \text{Mean} & \dfrac{N_1 M_1 + N_2 M_2}{N_1 + N_2} \\ \text{SD} & \sqrt{\dfrac{(N_1 - 1)SD_1{}^2 + (N_2 - 1)SD_2{}^2 + \dfrac{N_1 N_2}{N_1 + N_2}(M_1{}^2 + M_2{}^2 - 2M_1 M_2)}{N_1 + N_2 - 1}} \end{array} \quad (3.5)$$

Furthermore where possible data reported in different units (e.g. pounds instead of kg) was transformed to the unit of analysis.

When there is no sufficient information to calculate the SD for the mean change from baseline, it can be imputed; for this thesis a correlation coefficient was used for imputation. However it must be borne in mind that this kind of imputation is only appropriate if the assumption holds that the set of studies being used for deriving a correlation coefficient is comparable to the study where the missing SD will be imputed (since correlations will differ due to different time of follow-up, interventions or patient characteristics). Thus, similar studies were selected to meet this assumption. The correlation coefficient (corr) describes how similar baseline and final measurements are across intervention groups.
It lies between -1 and 1. For each arm, it can be derived as follows:

$$Corr = \frac{SD^2{}_{bl} + SD^2{}_f - SD^2{}_{ch}}{2SD_{bl}SD_f} \quad (3.6)$$

in which SD_{bl}, SD_f and SD_{ch} denote the standard deviation of the baseline, final value and the mean change, respectively.
When either the baseline or final SD is missing, it may be substituted by the other, if it´s reasonable to assume that the intervention doesn´t impact the variability of the outcome.
If the correlation coefficients between the arms are similar they were averaged and pooled across the set of studies to obtain a global correlation coefficient.
Now, for the study with the missing SD for the mean change, it can be imputed for each intervention arm.

$$SD_{ch} = \sqrt{SD^2{}_{bl} + SD^2{}_f - (2CorrSD_{bl}SD_f)} \quad (3.7)$$

Again, missing baseline or final SD may be substituted by each other.
If calculation of a correlation coefficient seemed inappropriate (if values of less than 0.5 are obtained or correlation coefficients from two intervention groups differed obviously), a SD for the mean change from baseline from another study in the review can be taken. However, the appropriateness of this approach relies again on the similarity assumption between studies. Nevertheless, to account for the uncertainty of the latter procedure these studies were subjected to sensitivity analysis (SA) to assess the robustness of results (see 3.1.4.4).
Finally, if measures of uncertainty were not reported but displayed in a figure with its error bars, they were read from the graph.

For HbA1c and body weight, the following approach was made in subsequent order to obtain the required data according to the formulas above: preferably, the mean change from baseline together with the measure of uncertainty (SEM, SD, CI or p-value) were extracted for each eligible intervention arm; all measures of uncertainty were transformed to the SEM. Alternatively, the effect estimate and its corresponding measure of uncertainty (SEM, CI or p-value) was extracted and transformed to arm-level data. If no effect estimate was available and the measure of uncertainty for the mean change was missing this value was imputed using a correlation coefficient. If calculation of a correlation coefficient is inappropriate final values together with their measure of uncertainty (SEM, SD, CI or p-value) were extracted. If - despite randomization - baseline values differed to obviously for a valid comparison a SEM from a comparable study was taken as approximation.

For hypoglycemia, preferably, the number of patients or the proportion (%) experiencing at least one episode of any hypoglycemia and the patients at risk (sample size of the safety population) were extracted for each eligible intervention arm. Alternatively the effect estimate (odds ratio (OR)) and its corresponding measure of uncertainty (SEM, CI or p-value) was extracted and transformed to arm-level data. The odds ratio was chosen over the relative risk as unit of analysis. Under the so called “rare disease assumption” the OR provides an adequate approximation of the RR. For the incidence, a percentage of 10 % is a frequently mentioned as cut-off point to ensure a close correspondence between both measures, but even this threshold is arbitrary [98]. The non-equivalence of the risk ratio and odds ratio for common events does not indicate that either is wrong: both are entirely valid ways of describing an intervention effect, as long as they are not misinterpreted.

3.1.4 Assessing the risk of bias in included studies

The conclusions that are drawn in systematic reviews rely heavily on the validity of the data and results of the single studies. Assessing the risk of bias relates to the internal validity of study, namely if the study answers its research question accurately and is free of bias. On the other hand, external validity which will not be further addressed here means if the results of the study can be generalized and its findings can be applied to different settings (other populations, treatments, etc.).

There are two types of errors potentially affecting results, random and systematic errors. Imprecision underlying the random error means that several studies will yield different results because of sampling error but give the right answer on average (if no systematic error is present at the same time). Smaller studies are stronger affected by sampling variation and are less precise (hence have a greater SEM and confidence interval). On the other hand bias is causing systematic errors or deviations from the truth meaning that several studies biased towards the same direction would reach a wrong answer on average. The extent of bias can vary in the direction and dimension. Bias can act in both direction thus leading to an under- or overestimation of the true effect. Furthermore bias can vary in magnitude; small bias is unlikely to affect the observed effect overly, whereas a finding might be entirely attributed to a pronounced bias. Thus in meta-analysis, exploration of bias can help to prevent false positive (claiming wrongly an intervention is effective) and negative conclusions (concluding wrongly no effect) if the methodologically less rigorous studies are overestimating or underestimating the true effect, respectively; moreover its assessment can help to explain heterogeneity of the included studies. In conclusion, invalid studies may lead to misleading inferences thus it is an essential part of a systematic review to carry out a thorough assessment of the risk of bias of included studies.

3.1.4.1 Tools for assessing study quality and risk of bias

There exist a lot of quality appraisal tools for RCT in the medical literature; a review identified more than 25 scales and 9 checklists that have been proposed to assess the internal validity of trials [99]. Scales are instruments, in which various items are rated and combined to give an overall summary score of the risk of bias, whereas checklists contain different questions relating to different aspects of the risk of bias enabling operators to screen if these items were sufficiently considered or not. The disadvantage of applying scores is that they are giving different weights to the various items of the risk of bias; it´s difficult to find a justification so that assigning weights seems to be an arbitrary approach. Moreover, scales may provide unreliable assessments and are less transparent for readers of the review [100]. Although checklists avoid the issue of weighting and summarizing different components, they are incapable to rate the degree of bias within each item; thus, both tools were deemed inappropriate and hence not used in this work.

3.1.4.2 The Cochrane Collaboration's tool for assessing risk of bias

The Cochrane Collaboration's tool for assessing risk of bias is neither a scale nor a checklist. The risk of bias is assessed within seven different domains, either at study- or endpoint specific level. For each entry in the corresponding domain the assessment is split into two parts. The first part is the support for judgment which outlines what was reported in the study to enable the judgment itself. The second part, the judgment rates the risk of bias for each entry as low, high or unclear, the latter denoting either lack of information or uncertainty over the potential for bias for items that are sufficiently reported.

Domains of bias

Table 6 displays the most common types of bias occurring in RCT and shows how they translate into the corresponding domains of the Cochrane Collaboration's tool for assessing risk of bias which will be used in this work. The different domains are illustrated shortly.

Table 6: Classification scheme of bias in randomized controlled trials

Type of bias	**Description**	**Domains of the Cochrane Collaboration's tool for assessing risk of bias**
Selection bias	Systematic differences between baseline characteristics of the groups that are compared.	- Sequence generation - Allocation of concealment
Performance bias	Systematic differences between groups in the care that is provided, or in exposure to factors other than the interventions of interest.	- Blinding of participants and personnel
Detection bias	Systematic differences between groups in how outcomes are determined	- Blinding of outcome assessment
Attrition bias	Systematic differences between groups in withdrawals from a study.	- Incomplete outcome data

Reporting bias	Systematic differences between reported and unreported findings	- Selective outcome reporting
Other bias	Other potential threats to interval validity	- Other bias

Sequence generation

An adequate sequence generation is the first step in the allocation process ensuring that baseline characteristics and prognostic factors are evenly distributed between intervention groups. The use of a random component generates an unpredictable sequence and is therefore usually sufficient to provide adequate sequence generation. Alternation methods like assignment based on date of birth, case record number and date of presentation should be avoided since sequences may be anticipated or are known and selection bias can occur due to selective enrolment.

Allocation of concealment

Random sequence generation is a necessary but not a sufficient protection against selection bias. An appropriate sequence is likely to be worthless if it is not protected by adequate concealment of the allocation sequence from those involved in the enrolment and assignment of participants. On the other hand, a proper allocation of concealment alone may be insufficient, too, if the sequence is not generated at random. Due to different prognoses of patients or different beliefs of the efficacy of the compared interventions some trialists may have preferences which patient should be allocated to which arm. If this assignment becomes foreseeable patients may be rejected or delayed from study entry to meet these preferences and thereby directed to the „appropriate“ intervention.
Thus, the second step in the allocation process must guarantee that those persons involved in enrolment into the trial have no foreknowledge of future intervention assignments; adequate methods are central randomization by a third party or the use of sequentially numbered, opaque, sealed envelopes.

Blinding of participants and personnel

After enrolment into the study, blinding of study participants and personnel may prevent that the knowledge of which intervention was received, rather than the intervention itself, affects outcomes.
Lack of blinding of participants or healthcare providers, especially for patient reported outcomes like pain or depression scales could bias results whereas measured outcomes are less likely to be affected. This may be due to a lack of expectations in a control group, or due to differential drop-out, cross-over to an alternative intervention, or differential co-medication. However, for some interventions like surgery, blinding can´t be implemented. Even if the interventions are administered in a blinded fashion, unblinding can occur for example due to certain characteristic side effects enabling individuals to guess the received intervention.

Blinding of outcome assessment

Outcome assessment can be made by patients, health care providers or independent assessors. The most outcomes can be impacted by lack of blinding, although the risk of bias increases with PRO´s being more subjective. On the other hand, for example laboratory measured outcomes are less likely to be affected and some outcomes like all-cause mortality are not

affected at all. It is therefore important to consider how subjective or objective an outcome is when this item is evaluated. Within one study it can occur that the relevance of blinding of patients and personnel and outcome assessors is dependent on the endpoints. The risk of bias may be high for some outcomes and low for others, even if the same people were unblinded in the study. For example, knowledge of the assigned intervention may bias PRO´s such as reported hypoglycemia while not distorting other outcomes like HbA1c or body weight. Thus, the risk of bias resulting from lack of blinding for both categories will be made separately for the selected outcomes. However, since both HbA1c and body weight represent measured and objective outcomes and carry similar risks of bias they are grouped together.

Incomplete outcome data

Missing outcome data refers to drop-outs or exclusion of patients in a study. Attritions can arise when patients withdraw from a study, are lost to follow-up, or don´t attend follow-up visits. Exclusion can occur when patients are deemed ineligible after assignment or a certain kind of analysis like per-protocol (only patients that complied with the study protocol are included in analysis) is conducted. Missing outcome data may lead to biased results if its nature is informative implying that there are systematic differences in outcomes between patients remaining in the trial and patients leaving the trial. Uninformative missing outcome without any association between these groups will only affect precision but not magnitude in terms of direction and dimension of results. Apart from the informativeness of the missing data it is important to assess the total amount, the distribution among intervention groups and the reasons for incomplete outcome data as well as what solutions were provided by the authors to overcome this issue in the reported analyses. An intention-to-treat (ITT) analysis which analyzes patients as randomized regardless of the treatment that was finally administered is often recommended to be least biased; however, if there is substantial attrition an ITT analysis can hardly be performed without any imputation method like substituting the mean outcome (or minimum or maximum in sensitivity analyses) of an intervention group or for dichotomous outcomes assigning treatment success or failure. Even techniques like the last observation carried forward method (LOCF) (where the most recently observed outcome measure substitutes all subsequent outcome assessments) can bear a risk of bias for example in degenerative diseases with a long time period between last available measurement and end of follow-up. Generally there exists no consistent definition for the different populations analyzed. It is therefore not enough only to denote an analysis as ITT, modified ITT or per-protocol; authors should rather report their trial-specific definitions so that all relevant information for bias assessment is provided to reviewers.
Generally the risk of bias will be low, if the overall attrition rate is low, the reasons for missing outcome data are unlikely to be related to the true outcome (non-informative), informative missing data is balanced in numbers and reasons across intervention groups, per protocol analyses are done without any substantial attrition and if missing data for ITT analyses have been imputed with appropriate methods.

Again, the incompleteness of missing outcome data and its potential bias can vary among outcomes. Thus, analogous to the domains of blinding, the assessment was done separately for hypoglycemia and together for HbA1c and body weight; however, if necessary, the degree of bias for HbA1c and body weight will be reported separately.

Selective outcome reporting

Selective outcome reporting is defined as the selection of a subset of the original variables recorded and intended for reporting, on the basis of the results, for inclusion in publication of trials [101]. A main reason is that statistically non-significant results might be selectively withheld from publication. There is evidence that statistically significant outcomes had a higher odds ratio of being fully reported when compared with non-significant outcomes, both for efficacy (OR:2.4; $CI_{0.95}$:[1.4;4.0]) and for harms (OR:4.7; $CI_{0.95}$: [1.8;12]) data. Moreover, the comparison of publications with protocols revealed that 62% of trials had at least one primary outcome that was changed, introduced or omitted [80,102]. The most common reasons for non-publication of results were "lack of clinical importance" or lack of statistical significance. Thus, meta-analyses excluding unpublished outcomes are at risk to systematically over- or underestimate treatment effects.
Sometimes for reviewers it may be hard to distinguish why some outcomes are not reported, due to selective omission or due to the fact that data just has not been collected. Selective choice of data for the same outcome (such as different points in time or the use of different instruments or scales) may result in bias if authors deviate from pre-specified criteria and make decisions depending on seeing the results. Moreover, selective reporting of analyses using the same data may have an impact on results. For example, decisions about analyzing continuous outcomes as final values or mean change from baseline or dichotomizing them with different cut-off points should be defined upfront and not be driven by its findings. Another reason for selective outcome reporting is underreporting of data in inadequate detail such as p-value threshold (which requires reviewers to make assumption if they want to include the study) or just mentioning that there were no differences in results rendering a quantitative analysis impossible. A study is unlikely to be biased, if the reviewer can access the study protocol and all endpoints relevant to the review have been reported as pre-specified. Unfortunately, the study protocol isn´t always available so that reviewers have to rely on their common sense which of the relevant endpoints they reasonably expect to be reported or compare the methods section with the results section to assess the integrity of pre-specified endpoints by means of number and way of reporting (however the latter approach will only detect obvious discrepancy in the publication itself and not detect intentional data manipulation).

Other bias

The category other bias accumulates all remaining sources of bias that can impact internal validity. Some are related to specific trial designs such as a carry-over effect or naïve pre-post comparisons in cross-over trials or recruitment bias in cluster-randomized trials, whereas others can arise trial specific such as asymmetric designs favouring one of the interventions (e.g. insufficient dose escalation impairing efficacy or forced dose titration triggering side effects) or unadjusted analyses for baseline values of continuous outcomes if the baseline variables are imbalanced and may modify the treatment effect.

3.1.4.3 Presentation and summary of the assessment of risk of bias

For each study the domains sequence generation, allocation of concealment, selective outcome reporting and other bias will be assessed at the study level, whereas the remaining domains will be appraised at the outcome level (jointly for HbA1c and body weight). Each entry will comprise a judgment of the risk of bias as „low", "high" or „unclear" along with a support for the judgment - a quote from the source of information and/ or a comment from the reviewers. Assessment of bias will be done independently by two reviewers (TB and SK);

in case of disagreement this issue will be overcome by a third reviewer (UM). The risk of bias for the primary outcomes will be summarized according to the following decision rule:
The overall risk of bias for an outcome within a study will be deemed high in the presence of high bias in any domain, low if all key domains (all domains except random sequence generation and allocation concealment) will be of low bias, and unclear in all other cases. The risk of bias for each study at the study- and outcome level together with the support for judgment will be presented in a risk of bias table.
Moreover, a figure will be provided displaying the distribution of the risk of bias for all domains at study- or endpoint level across all studies.

3.1.4.4 Accounting for risk of bias in analysis

When performing and presenting meta-analyses and mixed treatment comparisons, review authors must address the risk of bias in the results of included studies. There are several ways how the risk of bias assessment can be incorporated into the statistical analysis. One possibility of visually exploring a relationship between outcomes and risks of bias is to present forest plots for the pooled treatment effects stratified according to the different categories of risk of bias. A further approach is to include a covariate for the risk of bias in meta-regression and to test if the treatment by covariate interaction is statistically significant; alternatively a test for differences across subgroups representing the different categories of risk of bias can be applied. However the last two methods suffer from low power if the number of studies is small; based on a non-significant p value, it should not be concluded that there is no association between the results and bias categories. Furthermore, studies can be weighted in evidence synthesis according to their validity [103], as it is commonly done for combining results of multiple studies by the inverse variances of their effect estimates, which gives studies with more precise results more weight. The main concern is that it requires a measure of validity for each study, and there is no established method how to weight the different domains of bias for aggregation. Bayesian approaches allow incorporating beliefs of bias from external evidence, expert opinion or reasonable guess in prior distributions driving together with the data the posterior results. However, these methods are subject of current research and are not sufficiently well developed for widespread use.

A further alternative is to evaluate the risk of bias by conducting sensitivity analyses.
If studies are of mixed quality regarding the risk of bias the different approaches to analysis involve a trade-off between bias and precision. A meta-analysis that includes all eligible studies may produce a result with high precision but bears the risks of yielding a biased estimate of the treatment effect. On the other hand, restricting the analysis to studies at low risk of bias in all domains may produce an unbiased result but ignores some of the evidence.

In this thesis the former approach will be selected for the primary analysis reporting data on all included studies; furthermore, the robustness of the results will be explored by sensitivity analysis excluding biased observations. In summary, these are

- entire studies or single endpoints according to the riks of bias assessment
- studies with disagreement regarding study inclusion, data collection and the risk of bias assessment
- studies for which only an abstract is available or which are not published in English (see 3.1.2.2)
- studies for which the SD for the mean change from baseline was borrowed from another study and
- studies reporting different categories of hypoglycemia (see Table 12b).

3.1.4.5 Publication bias

There is evidence that studies arguing a beneficial effect or a larger effect size are more often published than data showing a harmful effect or no effect at all. Hopewell et al. showed that positive results vs. negative results or results with no difference as well as significant vs. not significant results were more likely to be published (OR:3.36; $CI_{0.95:}$[1.73;6.53] and OR:3.58; $CI_{0.95:}$[1.84;6.99], respectively) [104].This issue is commonly referred to as publication bias which is somehow related to selective outcome reporting bias.

Given this situation, any evidence synthesis relying only on published results will lead to biased and spurious results. Thus it is crucial part of systematic review to explore if publication bias may be present and is likely to affect the findings.

One way to judge potential publication bias is the use of a funnel plot.
It illustrates the scatter of the intervention effect estimates on the horizontal axis against a measure of the precision of the study, usually the standard error of the mean.
The name 'funnel plot' arises from the fact that the precision of the treatment effect is positively correlated with the size of study. Effect estimates from small studies will therefore scatter more widely at the bottom, with the spread decreasing in larger studies. In the absence of publication bias the plot should approximately resemble a symmetrical inverted funnel.

If for example smaller studies without statistically significant effects are not published, this introduces bias and will lead to an asymmetrical appearance of the funnel plot with a gap in a bottom of the figure (usually at one corner of the scatter). Ratio measures should be plotted on a logarithmic scale. This ensures equidistance from opposite effects from the null effect.

However, the visual inspection of the funnel plot alone is subjective and insufficient to draw conclusions about publication bias. Thus it should only be made in combination with a statistical test of funnel plot asymmetry.
A test for funnel plot asymmetry examines whether there is a statistical significant association between estimated intervention effects and the corresponding measure of uncertainty like the standard error of the mean. Egger et al. propose a linear regression of the intervention effect estimates on their standard errors. Under the null hypothesis of no publication bias such a line would be vertical. The greater the association between intervention effect and standard error the more the slope moves towards the horizontal [105].
Since the power of such tests to detect asymmetry is rather low they should only be used if enough amount of information is available. Even if the number of studies seems adequate one should not infer that due to a non-significant result of the test publication bias is absent keeping in mind that „absence of evidence is not evidence of absence" [106].

Thus, in this work publication bias will only be assessed for pairwise comparisons informed by ten or more studies by using contour enhanced funnel plots which display different areas of statistical significance to ease assessment of potential asymmetry and the Eggers test for funnel plot asymmetry.

3.1.5 Agreement and handling of disagreement

As mentioned before, critical steps of a systematic review involving subjective judgment should be done by two independent reviewers. In addition the quantification of the amount of interrater agreement is helpful to assess the goodness of independent review. If agreement is low this for example can be due to vague formulated eligibility criteria or less sophisticated data collection forms which may be revised. If disagreement is still substantial without any obvious reason this may alert reviewers that this very step of the review procedure is susceptible to subjective judgments which must be considered by interpreting the results.

The kappa statistic is helpful tool for measuring the extent of agreement by two raters exceeding that just by chance alone. Consider the following 2x2 contingency table for two reviewers R_1 and R_2 making judgment about inclusion of studies (which can readily be extended to more than two values for the rated variable):

		R_2		
		I	E	
R_1	I	a	b	I_1
	E	c	d	E_1
		I_2	E_2	n

From the observed and expected agreement just by chance, kappa (κ) and its confidence interval can be derived (3.8).

$$
\begin{aligned}
&a.)\ p_e = \frac{I_1 \cdot I_2 + E_1 \cdot E_2}{n^2} \\
&b.)\ p_o = \frac{a+d}{n} \\
&c.)\ \kappa = \frac{p_o - p_e}{1 - p_e} \\
&d.)\ CI(\kappa)_{0.95} = \kappa \pm \sqrt{\frac{p_o(1-p_o)}{n(1-p_e)^2}}
\end{aligned}
\qquad (3.8)
$$

P_e denotes the agreement by chance, I and E stands for included and excluded studies, respectively, p_o is the observed agreement and n indicates the total number of studies. Values of kappa between 0.40 and 0.59 may be considered to correspond to fair agreement, between 0.60 and 0.74 to good agreement and 0.75 or more to reflect excellent agreement. The kappa statistic will be applied to measure agreement of inclusion of studies based upon title and abstract as well as full text, of data collection and of assessing the risk of bias.

3.2. Evidence synthesis

3.2.1 Bayesian Statistic

3.2.1.1 Inference

Bayesian statistical methods of inference combine a prior belief expressed by a prior probability distribution (which depicts possible values of a unknown parameter of interest) with a distribution based on the observed data (the likelihood) to yield a posterior probability distribution of the unknown parameter of interest [107]. The principle idea behind it goes back to 1763 when the Bayes Theorem was published. Thomas Bayes, an English mathematician and presbyterian pastor showed how existing beliefs are modified by new information to posterior results. A familiar application of its theorem today is diagnostic testing where a physician´s prior belief that a patient suffers from a disease (based on prevalence data and patients symptoms) will be modified by new information (by means of a diagnostic test) to result in the physicians posterior belief about the patient having the disease [108].
Imagine the following 2x2 contingency table where D+, D-, T+ and T- denote people with disease and without disease, positive test and negative test results, respectively. A test has a sensitivity (probability that the test is positive given that the patient is ill; p (T+|D+)) and a specificity (probability that the test is negative given that patient is healthy; p (T-|D-)) of 0.95and 0.98, respectively, and the prevalence (p (D+)) in a population of 100,000 individuals is 0.001.

	D+	D-	
T+	95	1,998	2,093
T-	5	97,902	97,907
	100	99,900	100,000

Thus the posterior probability that a patient has the disease, given that the test result is positive (also called positive predictive value (PPV)) is derived as follows:

$$p(D+|T+) = \frac{p(T+|D+)}{p(T+)} \cdot p(D+) \tag{3.9}$$

In this case the PPV is 0.045 arguing that more than 95% of test positives don´t have the disease revealing that even when the discriminatory power of the test is adequate but the prior probability is low the posterior probability will still be relatively small [109].

Generally spoken, suppose that θ is some unknown quantity, for example the treatment effect of a new drug and let p (θ) denote the prior distribution of θ. Now there is some observed evidence y, whose probability of occurrence is assumed to be dependent on θ, which is formalized by p (y|θ), the likelihood. It is the extent to which different values of θ are supported by the data. The new posterior probability for different values of θ, taking into account the evidence y, p (θ|y) is derived as follows:

$$a.)\ p(\theta|y) = \frac{p(y|\theta)}{p(y)} \cdot p(\theta)$$
$$b.)\ p(\theta|y) \propto p(y|\theta) \cdot p(\theta) \tag{3.10}$$

Generally, p (y) is just a normalizing constant to ensure that $\int p(\theta|y)\, d\theta = 1$; thus Bayes theorem is often written as in formula 3.10b) which implicates that the posterior distribution is proportional to the product of the likelihood and the prior [109].

For Bayesian inference it is important to assume that random variables are independent and identically distributed (i.i.d.). This refers to the broader term of exchangeability; it assumes that the random variables $Y_{1,....},Y_n$ are exchangeable if their joint probability function $p(y_1,.....y_n)$ is not dependent on the permutation of its indices, which is described by the theorem from De Finetti:

$$p(y_1,.....y_n) = \int \prod_{i=1}^{n} p(y_i|\theta) \cdot p(\theta) d\theta \tag{3.11}$$

From left to right it argues, that „exchangeable random quantities can be thought of as being i.i.d. variables drawn from some common distribution depending on an unknown parameter θ, which itself has a prior distribution.“ [109].

The main differences between the Bayesian and frequentist approach are outlined in Table 7.

Table 7: Comparison of the Bayesian and frequentist approach [108]

Issue	Frequentist methods	Bayesian methods
Data	Data are repeated random events with a certain frequency	Data are observed from the sample and fixed
Parameters	Unknown parameters are treated as having fixed but unknown values	Parameters are unknown quantities and are described probabilistically
Prior information other than in the study analyzed	Informally used in design	Used formally by specifying a prior probability distribution
Basic question	„How likely is the data, given a particular value of the parameter?“	„How likely is a particular value of the parameter given the data?“
Presentation of results	Likelihood functions, p values, confidence intervals	Plots of posterior distributions, calculation of probabilities of interest, credible intervals

Especially the probability statements differ between the two methods: whereas a frequentist probability in terms of a confidence interval implies that in repeated sampling x% of realized samples cover the true parameter the Bayesian probability yields credible intervals for posterior distributions that can be interpreted more intuitively like „there is a x% probability that the true parameter is in this interval“; often confidence intervals are wrongly interpreted in a Bayesian way. Another advantage of the Bayesian method is that uncertainty or scepticism about the unknown random variable can be incorporated into the model by specifying adequate priors. For example a sceptical prior about large treatment effects in a RCT might prevent to stop a trial too early due to fortuitously good interim results [108]; in random effects meta-analysis the heterogeneity variance is given a prior distribution fully acknowledging the uncertainty in this parameter whereas the frequentist approach ignores it and thus underestimates uncertainty [107].

Moreover, making predictive distributions yielding the prediction of an outcome in a future trial in a Bayesian setting is straightforward and inference from the posterior distribution leads naturally into a decision making framework. Finally Bayesian methods allow generating probability statements and ranking of treatments for different outcomes which go beyond pairwise comparisons which assists health care professionals in medical decision making.

Only in few simple examples one can derive the posterior distribution in a standard analytic form. With Markov chain Monte Carlo methods (MCMC) simulations of random numbers are estimable by iterative processes, expecting to draw from a posterior distribution p (θ|y) after a convergence time. If X_t denotes a random variable at time t and the state space s refers to the range of possible X values. The random variable is a Markov process if the transition probabilities between different values in the state space depend only on the current state p ($s_i \rightarrow j_i$) of the random variable

$$\Pr(X_{t+1} = s_j \,|\, X_0 = s_k, ..., X_t = s_i) = \Pr(X_{t+1} = s_j \,|\, X_t = s_i) \tag{3.12}$$

Thus, for a Markov random variable the only information about the past needed to predict the future is the current state of the random variable, knowledge of the value of an earlier state does not change the transition probability. A Markov chain refers to a sequence of random variables (X0,…, Xn) generated by a Markov process [110]. For computation of Bayesian models in this thesis the open source version WinBUGS of the BUGS program (Bayesian inference using Gibbs sampling) is used [111]. The Gibbs sampler is a special case of the Metropolis-Hastings algorithm and generates Markov chains to get a sample from the posterior distribution after a burn-in sample (the time before the Markov chain has reached a stationary distribution and has converged, see 3.2.1.2). After starting from an initial value random numbers are iteratively drawn from the conditional distribution given all other components [110].

For Gibbs sampling the parameter vector θ is separated in S components, $\theta = (\theta_1, \ldots, \theta_S)$ Let p (θ|y) denote the density of the posterior distribution, from which random numbers should be drawn.

1.) Define the initial values $\theta_1^{(0)}, \ldots, \theta_S^{(0)}$ and the number of iterations T. set t = 1.
2.) For s = 1,…,S: draw random numbers $\theta_s^{(t)}$ from the full conditionals

$$p(\theta_s \,|\, \theta_1^{(t)}, ..., \theta_{s-1}^{(t)}, \theta_{s+1}^{(t-1)}, ..., \theta_S^{(t-1)}, y)$$

using the current state of the other parameters
3.) if t = T, the algorithm stops, otherwise continue with step 2 and increase t by one.

After a convergence phase of t_0 iterations the random numbers can be seen as realizations $\theta_s^{(t0+1)}, \ldots, \theta_s^{(T)}$ from the marginal distribution of $\theta_S|y$, s = 1,…, S [112].

3.2.1.2 Convergence diagnostics

A crucial question is to determine whether the MCMC sampler has reached its stationary distribution. If convergence is slow and the number of iterations is not enough the drawn samples will not be representative for the target posterior distribution since the Markov chain can get stuck in a region heavily influenced by the starting distribution thus yielding misleading inferences [110]. However, it is very difficult to say conclusively that a chain (simulation) has converged, only to diagnose when it hasn´t yet. Thus, for each scalar summary of interest (all unknown quantities of interest in the model) two or more chains should be run in parallel and convergence can be assessed formally and visually [110,113]. If after the burn-in phase of the sampler, convergence has been achieved, the simulation should be run for further iterations to obtain samples that can be used for posterior inference.

The potential scale reduction factor $\sqrt{\hat{R}}$

For each parameter the scale reduction factor $\sqrt{\hat{R}}$ is essentially the square root of the variances for all simulated sequences mixed together divided by the average of the variances within the separate sequences. In the limit, when the number of iterations approaches infinity $\sqrt{\hat{R}}$ approaches 1, but if the sequences fail to reach convergence, $\sqrt{\hat{R}}$ can be much larger. It is recommended to continue simulation until $\sqrt{\hat{R}}$ is close to 1 [110].

Monitoring trace plots visually

For each simulated chain of the parameter adequate mixing can be assessed by monitoring of a trace plot of the sample values versus iteration. Convergence seems reasonable if all the chains appear to be overlapping one another [114] (see Figure 2, top). On the other hand, poor convergence (see Figure 2, bottom) requires extending the burn-in phase before sampling results from the posterior distribution.

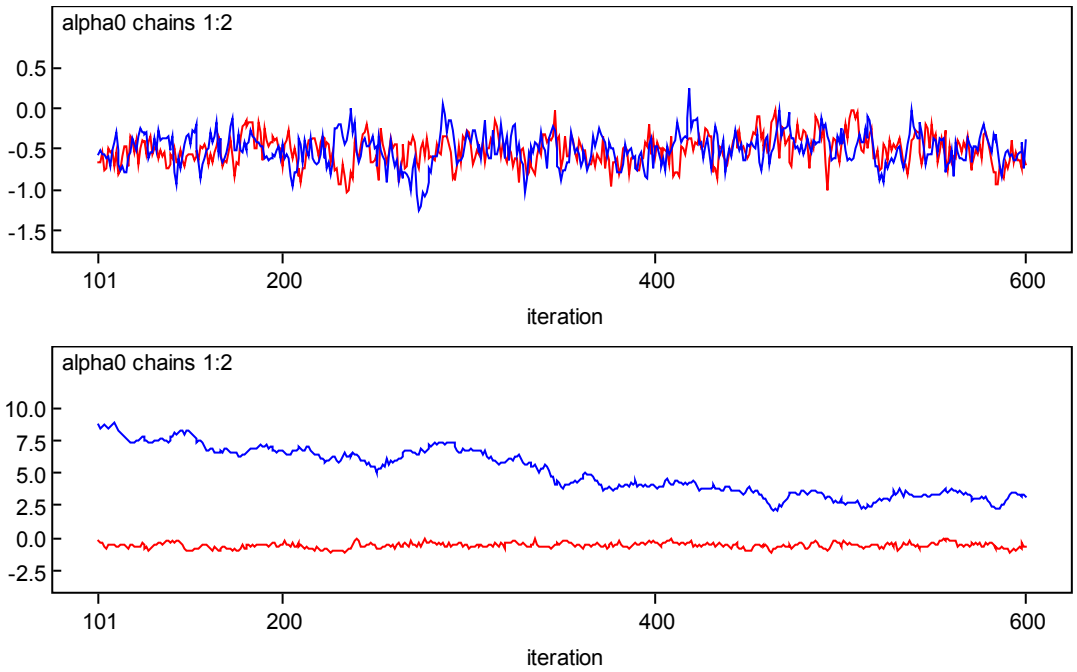


Figure 2: Trace plots for assessing convergence [114]

3.2.1.3 Model diagnostics

To check if a model fits the data well the overall residual deviance $\overline{D}_{res}$ will be considered, which is the posterior mean of the deviance under the current model minus the deviance for the saturated model. Table 8 illustrates the residual deviance for binomial and normal likelihoods which are used for the Bayesian models in this thesis.

Table 8: Formulas for the residual deviance, prediction and likelihoods [113]

Likelihood	Prediction	Residual Deviance
$r_{ik} \sim bin(p_{ik}, n_{ik})$	$\hat{r}_{ik} = n_{ik} p_{ik}$	$\overline{D}_{res} = \sum_i \sum_k 2\left(r_{ik} \log\left(\frac{r_{ik}}{\hat{r}_{ik}}\right) + (n_{ik} - r_{ik}) \log\left(\frac{n_{ik} - r_{ik}}{n_{ik} - \hat{r}_{ik}}\right)\right)$ $\overline{D}_{res} = \sum_i \sum_k \overline{dev}_{ik}$ (3.13)
$y_{ik} \sim N(\overline{y}_{ik}, se_{ik}^{\,2})$	$\overline{y}_{ik}$	$\overline{D}_{res} = \sum_i \sum_k \left(\frac{\left(yi_k - \overline{y}_{ik}\right)^2}{se_{ik}^{\,2}} \right)$ $\overline{D}_{res} = \sum_i \sum_k \overline{dev}_{ik}$ (3.14)

For a Binomial Likelihood, r_{ik} and n_{ik} are the observed number of events and patients from each trial arm, $\hat{r}_{ik} = n_{ik}p_{ik}$ is the expected number of events at each iteration; for a Normal Likelihood y_{ik} and se_{ik} are the observed mean and its standard error from each trial arm and $\overline{y}_{ik}$ is the expected mean at each iteration; $\overline{dev}_{ik}$ is the posterior mean of the deviance residual for each data point which is summarized by the posterior mean $\overline{D}_{res}$ [113].

Leverage statistics assess the influence of each data point on the model outcomes. The leverage of each data point, $leverage_{ik}$, is derived as the posterior mean of the residual deviance $\overline{dev}_{ik}$ minus the deviance at the posterior mean of the fitted values $\hat{dev}_{ik}$; this is summarized by p_D which is a term for model complexity and is equal to the effective number of parameters in the model [113].

$$p_D = \sum_i \sum_k leverage_{ik} = \sum_i \sum_k \left[\overline{dev}_{ik} - \hat{dev}_{ik}\right] \tag{3.15}$$

As an equivalent to Akaike´s Criterion (AIC) in Bayesian statistics, the Deviance Information Criterion (DIC) is the sum of the posterior mean of the residual deviance, $\overline{D}_{res}$, and the effective number of parameters, p_D. It provides a measure for the goodness of fit for the model that penalizes middle complexity and lower values of the DIC suggest a more parsimonious model. Hence, the trade-off is to use a parsimonious model that sufficiently fits the data but that provides stable parameter estimates [59].
It is helpful for model selection between different models with the same likelihood and data such as comparing fixed effects and random effects models or meta-regression models which are extended by covariates [113].

Leverage vs. residual plots can provide further information whether poorly fitting data points are substantially affecting the model parameters. Each data point´s contribution to the $leverage_{ik}$ on the vertical axis is plotted against w_{ik}, its contribution to $\overline{dev}_{ik}$ on the horizontal axis, which is expressed by $w_{ik} = \sqrt{\overline{dev}_{ik}}$. Parabolas of the form $x^2 + y = c$, c=1,2,3…, where x denotes w_{ik} and y denotes the leverage can be labeled on the plot and points lying on these curves contribute an amount c to the DIC [115]. Points which lie outside the line with c=3 can be identified as highly contributing to the models poor fit. Data points with a high leverage are influential, implicating that they have a strong influence on the model parameters [113]. In this work, the DIC will be used for model comparison and selection and studies whose arms contribute to poor model fit (that is, the data points lie outside the line with c=3) will be subjected to sensitivity analysis.

3.2.1.4 Implementation in WinBUGS and R

For computation of Bayesian models (pairwise MA, MTC, meta-regression, assessment of inconsistency, ranking of treatments and validation) the software WinBUGS was used [111]. All remaining analyses were done with the statistical software R version 2.15.2 [116]. Setting up a model and running an analysis in WinBUGS usually requires the following steps. First a model in the BUGS language has to be specified to providing all probabilistic and deterministic information such as the likelihood, the linear predictor and prior distributions. In the Bayesian setting for a normal distribution the uncertainty is parametrized by the precision tau (1/variance). Secondly, the data has to be loaded. Thirdly, the number of Markov chains which will be run in parallel to yield samples of the posterior distribution for all unknown parameters has to be determined; then, the model will be compiled. Fourthly, for each Markov chain different sets of initial values for each probabilistic parent node have to be specified. It is recommended that overdispersed starting points are used for the initial values in order to detect a potential lack of convergence and ensure that all regions of the target distribution are covered sufficiently [113].
Fifthly, the model can be run for the burn-in phase and convergence has to be monitored as mentioned above. Sixthly, after reaching convergence the model will generate values from the posterior distributions. Burn-in and posterior samples should be conservatively large enough. As a rule of thumb, iterations should be run until the Monte Carlo error which reflects both the number of iterations and autocorrelation and is an estimate of the difference between the mean of the sampled values and the true posterior mean is less than 5% of the posterior standard deviation of the parameters of interest [113].
A substantial autocorrelation of consecutively drawn parameters can impair precision of results. A thin out function using only every k-th iteration as realisation from the posterior distribution can be set-up to prevent autocorrelation. Moreover, autocorrelation plots can provide information if autocorrelation over the iterations decreases rapidly enough against null to generate uncorrelated random numbers [117]. Since a multitude of models will be run and computation in WinBUGS becomes very time consuming and cumbersome, the R2WinBUGS package will be used. It allows to call a WinBUGS model from R, summarize inferences and convergence in a table and graph, and save the simulations in arrays [118].

In this thesis the default settings for running WinBUGS models are as follows: for each model two Markov chains will be run with a burn-in phase of 30,000 iterations and a sampling phase of the posterior distribution of 70,000 iterations. A thin out function saving only every 10th sample from the posterior distribution will be used to prevent autocorrelation, thus yielding 14,000 samples from both chains for each parameter. Convergence will be assessed by monitoring trace plots for all parameters of interest.

Autocorrelation plots will be inspected at random to check if the thin out safeguards efficiently against correlation. For each parameter the Monte Carlo error is required to be less than 5% of the SD to ensure the accuracy of posterior estimates.
For DIC calculation and the construction of the leverage vs. residual plots a R function formula provided by the MTM group of the University of Ioannina was applied [119]. All relevant R and WinBUGS codes, data matrices and ancillary outputs not presented in the manuscript will be detailed as supplementary material on an electronic data carrier.

3.2.2 Pairwise meta-analysis (Bayesian and frequentistic)

Meta-analysis is used to synthesize study estimates of a particular effect of interest from related studies to obtain a summary estimate of the effect [120]; a fixed or a random effects model can be applied and meta-analysis can be performed in a frequentist or Bayesian way (see 3.2.2.1).
The advantages of conducting a MA are amongst others [121]

- To increase power. Power is the chance of detecting a real effect as statistically significant if it exists. Single studies may be too small to detect significant effects, but when several studies are combined there is a higher chance of detecting an overall significant effect.
- A gain of precision. The estimation of an intervention effect can be improved when it is based on more information rather than single studies.
- To provide answers not settled by single studies. Primary studies often involve different types of patients, received interventions, length of follow-up or other clinical characteristics. A set of studies in which these factors differ allows investigation of the consistency of effect and its generalizability, or allows reasons for differences in effect estimates to be explored.
- Conflicting results of single studies might be overcome by evidence synthesis.

However there also concerns raised when a meta-analysis might not be appropriate [121].

- A common criticism is the pooling of incomparable studies, also known as the „apples and oranges“ problem. If studies are clinically too diverse, then a meta-analysis may be meaningless, and genuine differences in effects may be masked. “This is true if apples and oranges are of intrinsic interest on their own, but may not be if they are used to contribute to a wider question about fruit” [121]. Decisions are subjective and require discussion and clinical judgment (see 3.1.1.2). However its appropriateness can be judged by statistical measures of heterogeneity (see 3.2.2.2).
- Inference of meta-analyses of studies that bear a high risk of bias may be deceptive. Present bias will compound the errors, yielding a ‘wrong’ result that may be interpreted as being more reliable.
- Serious publication and/or reporting biases are likely to produce an inappropriate summary estimate.

3.2.2.1 Fixed effects and random effects models

In a fixed effects (FE) model the underlying assumption is that all studies $\hat{\theta}_i$ are estimating the same common treatment effect θ, which implies that there is no between study heterogeneity in the true treatment effect and the observed treatment effect estimates vary only because of chance, namely their different variances sd_i^2. Thus, given that all studies had infinite sample size there would be no differences due to chance and the differences in study estimates would completely disappear [120]. The question behind any given comparison is: "What is the true treatment effect?" [121].

$$\hat{\theta}_i \sim N(\theta, sd_i^2) \tag{3.16}$$

A random effects (RE) model assumes that in addition to sampling error differences across studies are caused by heterogeneity between studies. The underlying assumption is that the study specific treatment effects θ_i are exchangeable meaning that the information provided by each study is independent of the order in which they were conducted [113]. Thus, a RE model assumes that the study specific treatment effects θ_i are drawn from a normal distribution; yielding an estimate of the average treatment effect θ and the heterogeneity variance τ^2 [107]. If all studies were infinitely large, the observed study effects would still vary because of the real differences in treatment effects, which can be caused by differences in study populations, received interventions, length of follow-up and other clinical characteristics [120]. The question for any given comparison is: "What is the average of the true treatment effects, and how much do these effects vary across trials?" [121].

$$\begin{aligned} &a.)\ \hat{\theta}_i \sim N(\theta_i, sd_i^2) \\ &b.)\ \theta_i \sim N(\theta, \tau^2) \end{aligned} \tag{3.17}$$

If heterogeneity is present the confidence interval of the pooled effect estimate from the random effects model is wider than that of the fixed effects model since it incorporates both sampling error from individual studies and between-study variance.
Often for a random effects model, only the pooled estimate and its confidence interval are reported, which is insufficient. The confidence interval for the mean treatment effect θ from a random effects meta-analysis only describes uncertainty in the estimation of the pooled effect but not the degree of heterogeneity among studies. Thus, the between-study variance in a random effects meta-analysis known as tau-squared (τ^2) should be reported, too. The square root of it is the estimated standard deviation of underlying effects across studies and prediction intervals can incorporate this between-study variance to give a range for a parameter value in a new study [120,121].

Frequentistic Approach

Meta-analysis typically consists of a two-stage process. First, based upon arm level data, a summary statistic is calculated for each study, to describe the observed intervention effect. Secondly, a summary intervention effect estimate is calculated as a weighted average of the intervention effects estimated in the individual studies [122]. For binary data consider the following 2x2 table for each study i.

Study i	Event	No event	Total
Treatment 1	a_i	b_i	n_{1i}
Treatment 2	c_i	d_i	n_{2i}

The odds ratio of treatment 1 vs. treatment 2 and its standard error are derived as follows:

$$\begin{aligned} &a.)\ OR_{21i} = \frac{a_i d_i}{b_i c_i} \\ &b.)\ se(\ln(OR_{21i})) = \sqrt{\frac{1}{a_i} + \frac{1}{b_i} + \frac{1}{c_i} + \frac{1}{d_i}} \end{aligned} \tag{3.18}$$

For continuous data consider the following table for each study i

Study i	Group size	Mean	Standard deviation
Treatment 1	n_{1i}	m_{1i}	sd_{1i}
Treatment 2	n_{2i}	m_{2i}	sd_{2i}

The difference in means and its standard error are derived as follows:

$$\begin{aligned} &a.)\ MD_{21i} = m_{1i} - m_{2i} \\ &b.)\ se(MD_{21i}) = \sqrt{\frac{sd_{1i}^{\ 2}}{n_{1i}} + \frac{sd_{2i}^{\ 2}}{n_{2i}}} \end{aligned} \tag{3.19}$$

Next, the pooled effect estimate over i studies is estimated. There are different statistical methods for summarizing the evidence such as the Mantel-Haenszel-, the Peto or the inverse-variance method. In this work, the inverse-variance method will be applied for both fixed effects and random effects models. Inverse-variance methods are used to pool effect measures like the log odds ratio (since meta-analysis requires the effect estimate to follow a normal distribution) or mean differences. The intervention effect estimate is denoted by θ_i and the effects from the single studies are given weights w_i according to the reciprocal of their variance [122]. For a fixed effects model the summary estimate $\hat{\theta}$ and its standard error are derived as follows:

$$\begin{aligned} &a.)\ \hat{\theta} = \frac{\sum w_i \hat{\theta}_i}{\sum w_i} \\ &b.)\ w_i = \frac{1}{sd_i^{\ 2}} \\ &c.)\ se(\hat{\theta}) = \frac{1}{\sqrt{\sum\ w_i}} \end{aligned} \tag{3.20}$$

For a random effects model the study specific treatment effects are drawn from superpopulation of effects ~ $N(\theta, \tau^2)$ [123]. The DerSimonian-Laird estimator $\hat{\theta}_{DL}$, its standard error and the heterogeneity variance $\hat{\tau}^2$ are derived as follows:

$$
\begin{aligned}
&a.)\ \hat{\theta}_{DL} = \frac{\sum w_i(\hat{\tau}_{DL})\hat{\theta}_i}{\sum w_i(\hat{\tau}_{DL})} \\
&b.)\ w_i(\hat{\tau}_{DL}) = \frac{1}{sd_i^{\,2} + \hat{\tau}_{DL}^{\,2}} \\
&c.)\ se(\hat{\theta}_{DL}) = \frac{1}{\sqrt{\sum w_i(\hat{\tau}_{DL})}} \\
&d.)\ \hat{\tau}^2{}_{DL} = \max\left(0, \frac{Q-(k-1)}{\sum w_i - \frac{\sum w_i^{\,2}}{\sum w_i}}\right)
\end{aligned}
\tag{3.21}
$$

where $w_i(\hat{\tau}_{DL})$ denotes the weights which are the inverse of the study specific variance sd_i^2 and the heterogeneity variance τ^2 and Q denotes the heterogeneity statistic (see 3.2.2.2). If the Q statistic is less than or equal to its degrees of freedom, τ^2 is zero and the weights correspond to those of the inverse-variance fixed effects model.

With the package meta meta-analysis with R can be performed. For binary and continuous outcomes arm-level data has to contain the number of events r_i and the patients at risk n_i and the mean change m_i, the standard deviation sd_i and the number of patients analyzed n_i for each study i, respectively. Since zero cells (e.g. no events in one group) cause problems with computation of estimates, increments of 0.5 for any affected study are added automatically to each cell of the 2×2 table as correction. Multiple arm trials are entered repeatedly into the data matrix to estimate all possible pairwise comparisons.

For frequentistic meta-analysis outcome data (the point estimate, $CI_{0.95}$ and the between-study variation τ) for fixed- and random effects models will be presented. The corresponding forest plots will be provided in the appendix.

Bayesian Approach

Bayesian meta-analysis as well as Bayesian MTC can be considered as Generalized Linear Models where the link function and the likelihood can change due to the nature of the data. For binary and continuous data, Table 9 displays the relationship.

Table 9: Link functions and their inverse for likelihoods [113]

Link	Link function $\theta = g(y)$	Inverse link function $y = g^{-1}(\theta)$	Likelihood
Identity	y	θ	Normal
Logit	$\ln(y/(1-y))$	$\frac{\exp(\theta)}{1+\exp(\theta)}$	Binomial

The generalized linear model for pairwise meta-analysis over i studies can be written as:

$$g(y) = \theta_{ik} = \mu_i + \delta_{i,1k} I_{\{k \neq 1\}} \quad where\, I_{\{u\}} \begin{cases} 1 if\ u\ is\ true \\ 0 otherwise \end{cases} \tag{3.22}$$

where g is an appropriate link function, y the likelihood of the unknown parameter y, θ_{ik} the linear predictor of a treatment effect in arm k of trial i (k=1,2; for pairwise meta-analysis), μ_i are the trial specific baseline effects in trial i, treated as nuisance parameters and $\delta_{i,12}$ are the trial specific treatment effects of treatment group (2) compared to control (1). For a random effects model the trial specific treatment effects are drawn from a normal distribution with mean d_{12} and variance σ_{12}^2.

$$\delta_{i,12} \sim N(d_{12}, {\sigma_{12}}^2) \tag{3.23}$$

For a fixed effects model, the heterogeneity variance is set to zero assuming all studies are estimating the same common treatment effect d_{12} [113]; formula 3.22 can be written as:

$$g(y) = \theta_{ik} = \mu_i + d_{1k} I_{\{k \neq 1\}} \quad where\, I_{\{u\}} \begin{cases} 1 if\ u\ is\ true \\ 0 otherwise \end{cases} \tag{3.24}$$

Binary and continuous data is modeled assuming a binomial (using logit link function to map the probabilities on a plus/minus infinity range) and a normal likelihood, respectively. The generic equation (3.22) can be written as

$$\begin{aligned} &a.)\ r_{ik} \sim bin(p_{ik}, n_{ik}) \\ &b.)\ \log it(p_{ik}) = \mu_i + \delta_{i,1k} I_{\{k \neq 1\}} \quad where\, I_{\{u\}} \begin{cases} 1 if\ u\ is\ true \\ 0 otherwise \end{cases} \\ &c.)\ y_{ik} \sim N(\theta_{ik}, {se_{ik}}^2) \\ &d.)\ \theta_{ik} = \mu_i + \delta_{i,1k} I_{\{k \neq 1\}} \quad where\, I_{\{u\}} \begin{cases} 1 if\ u\ is\ true \\ 0 otherwise \end{cases} \end{aligned} \tag{3.25}$$

for RE models with binary data (3.25a/b) and continuous data (3.25c/d), respectively; usually the uncertainty of normally distributed data in Bayesian models is expressed by the precision, which is one over the variance. If the heterogeneity variance is set to zero, the corresponding FE models are derived [113].
Special consideration must be given to the parametrization of priors for d_{12} and σ_{12}^2. Unless there is no reason for an informative prior, flat (non-informative) priors should be used to ensure that the posterior distribution is not influenced by the prior but totally driven by the data [107]. However, „there is no such thing as an uninformative prior“ and even flat priors should be reasonably selected [109]; especially the posterior of the heterogeneity variance is likely to be sensitive to its prior, and vague priors are likely to result in posteriors which allow for unrealistically high levels of heterogeneity if the amount of evidence is small [62].
In this thesis for the core analysis the treatment effects and the trial specific baseline effects are given a vague normal prior ($d \sim N(0,1.0e^{-6})$ and $\mu_i \sim N(0,1.0e^{-5})$, respectively) and the standard deviation of the heterogeneity variance σ is believed to follow a uniform distribution ($\sigma \sim unif(0,10)$; to account for the susceptibility of results, in MTC different priors for the between-study variance will be assessed in sensitivity analysis (see 3.2.3.4).

No model is assumed for the trial-specific baselines, they are treated as nuisance parameters which are estimated in the model [113]. Alternatively, one could place a second hierarchical model on the trial baselines. However, this approach is in keeping with frequentist methods in which relative effect estimates are treated as data and baselines eliminated entirely.

For analysis, WinBUGS requires binary and continuous arm-level data, namely the number of events r_i and the patients at risk n_i and the mean change m_i, and the standard error se_i for each study i, respectively. Multiple arm trials are entered repeatedly into the data matrix to estimate all possible pairwise comparisons. It must be acknowledged, that this approach ignores correlation among the effect estimates for the pairs of arms since the input data for the three direct comparisons in a three-arm trial is treated like coming from three separate studies. However, correlation will be taken into account for MTC models (see 3.2.3.1). Finally the number of studies informing each comparison must be specified. The models run for meta-analysis are adapted from the WinBUGS code provided from Dias et al. in their Technical Support Document 2 [113].

For Bayesian meta-analysis, both fixed and random effects models will be reported with the mean (if necessary the median, given a skewed posterior distribution) and the $CrI_{0.95}$; in addition, for random effects models the posterior mean of the between-trials standard deviation will be reported.

3.2.2.2 Assessment of heterogeneity

Higgins et al. proposed the following definition for heterogeneity: „…Statistical heterogeneity exists when the true effects being evaluated differ between studies, and may be detectable if the variation between the results of the studies is above that expected by chance…“ [124]. To common measures for assessment are the heterogeneity statistic Q and the quantity I^2 [123,124]. The heterogeneity test refers to the following statistic:

$$Q = \sum w_i(\hat{\theta}_i - \hat{\theta})^2 \sim \mathrm{X}_{df}{}^2 = k-1 \tag{3.26}$$

Q is the heterogeneity statistic which is assuming that under the null hypothesis all studies share a common treatment $(H_0 : \theta_1 = \theta_2 = \ldots = \theta_k = \theta)$ effect and follows a Chi-Square distribution with k-1 degrees of freedom. Under the alternative hypothesis, at least two studies differ $(H_A : \theta r \neq \theta t;\ \{r,t\} \subset \{1,\ldots,k\})$. However, if the number of studies is small, the power of the test is low to detect a significant heterogeneity. In this setting a non-significant result should not be interpreted as evidence of homogeneity [125].

The quantity I^2 describes the percentage of total variation across studies that is due to heterogeneity and goes beyond chance [124].

$$I^2 = \begin{cases} 100 \cdot \dfrac{(Q-df)}{Q}, if\ Q > df \\ 0, else \end{cases} \tag{3.27}$$

Even if there are no fixed categories I^2 values of 0.25, 0.5 and 0.75 may refer to low, moderate and high heterogeneity. The advantage of I^2 is that it can be interpreted similarly irrespective of the type of outcome data and choice of effect measure and it doesn´t depend inherently on the number of studies.

Thus, in a meta-analysis of only eight studies where the study specific odds ratios ranged from 0.07 to 0.84 the p-value of test for heterogeneity was 0.09 indicating no heterogeneity, whereas $I^2 = 0.44$ indicated a considerable amount of heterogeneity [125].

In this work the Q statistic for an alpha error level of 0.05 and I^2 will be used for assessing heterogeneity.

3.2.3 Mixed treatment comparison analysis (only Bayesian)

As outlined before (see chapter 1.6) network meta-analysis is an extension to classical pairwise meta-analysis and for connected networks it provides relative treatment effects for any possible contrast either by indirect treatment comparisons for open loop networks where either direct or indirect evidence informs evidence synthesis or by mixed treatment comparisons for networks with at least one closed loop where direct and indirect evidence (via one or several anchor treatments) yield mixed estimates. The idea of network meta-analysis in a Bayesian context was first brought up by Higgins et al. who revealed that „borrowing strength from external trials“ in a meta-analysis by jointly estimating all possible treatment effects leads to more precise posterior estimates especially with regard to the heterogeneity parameter [126]. For a simple triangular network, Bucher et al. provided the statistics to derive an indirect treatment estimate, a test for consistency to assess if direct and indirect evidence are in agreement and a mixed estimate as a weighted average of both evidence sources [127]. The rank-ordering of treatments to assign probabilities which treatment is best by relating all treatments to an overall baseline was introduced by Lu et al. in 2006 [128]. With the spread of applications and methods papers further approaches to address the issue of consistency such as the node splitting procedure or the inconsistency model were developed [128,129]; recent developments for example focus on the combination of aggregated level and individual patient data or how bias of small study effects can be accounted for in NMA reflecting that this evidence synthesis method becomes more elaborated and sophisticated [130,131]. Due to its appealing features network meta-analysis is increasingly accepted amongst researchers, health technology assessment agencies and funding bodies [60].

3.2.3.1 Fixed effects and random effects models

For simultaneously estimating all possible contrasts a statistical model can be derived that takes into account all the direct and indirect comparisons in the complete network.

When the network from equation (1.1) is generalized to any network with multiple different comparisons, it yields the following consistency equation:

$$\begin{aligned} d_{bk} &= d_{Ak} - d_{Ab} \\ d_{AA} &= 0 \end{aligned} \tag{3.28}$$

K denotes the intervention and b the baseline treatment for that trial. As long as k is alphabetically after b, in a connected network any estimate for a particular pairwise comparison can be expressed by the effect estimate of the intervention k relative to the baseline b expressing them in terms of effects in relation to an overall reference treatment A [59].

$D_b = (d_{AB}, d_{AC}, d_{AD}, \ldots, d_{Ak})$ is the vector of basic parameters of the model that are estimated based on the available studies. $D_f = (d_{BC}, d_{BD}, d_{CD}, \ldots)$ is the vector of functional parameters and can be calculated based on the estimates for the basic parameters [128].
For a network involving K treatments, there are T= K (K-1)/2 potential types of comparisons, K-1 basic parameters and T-K+1 functional parameters. In conclusion, a network meta-analysis extends a traditional meta-analysis model from one to K-1 parameters that need to be estimated to allow for multiple pairwise comparisons across the range of K interventions [59].
A fixed effects model assumes that relative treatment effects across the whole set of studies are identical.

$$g(y) = \theta_{ik} = \mu_i + d_{bk} I_{\{k \neq b\}} = \mu_i + (d_{Ak} - d_{Ab}) I_{\{k \neq b\}} \; where \; I_{\{u\}} \begin{cases} 1 \, if \; u \; is \; true \\ 0 \, otherwise \end{cases} \quad (3.29)$$
$$d_{AA} = 0$$

A random effects model assumes that the relative treatment effects relative effects across the whole set of studies are considered exchangeable and are drawn from a normal distribution.

$$a.) \; g(y) = \theta_{ik} = \mu_i + \delta_{i,bk} I_{\{k \neq b\}} \; where \; I_{\{u\}} \begin{cases} 1 \, if \; u \; is \; true \\ 0 \, otherwise \end{cases}$$
$$b.) \; \delta_{i,bk} \sim N(d_{bk}, {\sigma_{bk}}^2) = N(d_{Ak} - d_{Ab}, {\sigma_{Ab}}^2 + {\sigma_{Ak}}^2 - 2\rho_{bk}\sigma_{Ab}\sigma_{Ak}) \quad (3.30)$$
$$d_{AA} = 0$$

In this work it is assumed that between-study variances for all treatment contrast are equal, thus $\sigma_{bk}^2 = \sigma^2$. For heterogeneous variance models refer to Lu and Ades [132]. Multi arm trials induce a between arm correlation that has to be taken into account while estimating the random effects [113]. P_{bk} denotes the within trial correlation which in case of homogeneous variance is 0.5 for any two treatment comparisons from a multi-arm trial.
The vector of random effects δ_i estimated for multi arm trials follows a multivariate normal distribution:

$$\delta_i = \begin{pmatrix} \delta_{i,12} \\ \vdots \\ d_{i,1a_i} \end{pmatrix} \sim N_{a_i - 1} \left(\begin{pmatrix} d_{t_{i1},t_{i2}} \\ \vdots \\ d_{t_{i1},t_{ia_i}} \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma^2/2 & \cdots & \sigma^2/2 \\ \vdots & \vdots & \ddots & \\ \sigma^2/2 & \sigma^2/2 & \cdots & \sigma^2 \end{pmatrix} \right) \quad (3.31)$$

where a_i denotes the number of arms in trial i (a_i=2,3,…) and $d_{ti1,tik} = d_{1,tik} - d_{1,ti1}$. Then, the conditional univariate distributions for the random effects of arm k > 2 given all other arms from 2 to k-1 is [113]

$$\delta_{i,1k} \left| \begin{pmatrix} \delta_{i,12} \\ \vdots \\ \delta_{i,1(k-1)} \end{pmatrix} \right. \sim N \left(\left(d_{1,t_{ik}} - d_{1,t_{i1}} \right) + \frac{1}{k-1} \sum_{j=1}^{k-1} \left[\delta_{i,1j} - \left(d_{1,t_{ij}} - d_{1,t_{i1}} \right) \right], \frac{k}{2(k-1)} \sigma^2 \right) \quad (3.32)$$

For multi-arm trial correction in the calculated models equation (3.32) is applied to account for between-arm correlations.

For running MTC models with WinBUGS the data input consists of two components: a list containing the number of treatments and the number of studies and a data matrix that has to be set-up as follows: each row contains one study with k arms; treatment number identifiers and the number of arms are indexed as described in Table 5. Moreover, for each arm 1 to k treatment indices and data for binary and continuous outcomes are provided as described in section 3.2.2.1. Studywise for each arm treatments are always presented in ascending numerical order and treatment 1 is taken as the reference treatment. This rule is essential for MTC to make sure that the correct relative effects are estimated. Treatment A is chosen as the overall reference treatment and the studies are arranged consecutively as follows [133]:

a.) all A trials with AB trials first, then AC, AD, etc.
b.) all B trials not containing A in the same order
c.) all C trials not containing A or B, etc.

For all pairwise comparisons the outcomes for FE and RE models for MTC will be reported as described in section 3.2.2.1 together with the DIC for model comparison purposes. The between- study standard deviation will be reported with its mean and $CrI_{0.95}$. The MTC models are adapted from the WinBUGS code provided from Dias et al. [113].

3.2.3.2 Assessment of consistency and transitivity

The different assumptions underlying NMA and pairwise meta-analysis are displayed in Figure 3.

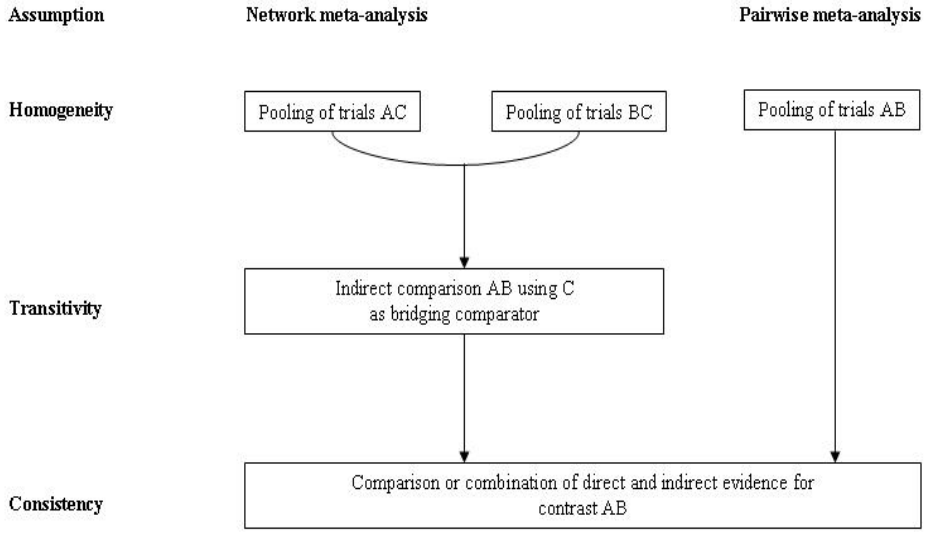


Figure 3: Assumptions underlying network meta-analysis and pairwise meta-analysis [134]

The assumption that different trials are sufficiently (not necessarily completely) homogeneous and that they estimate the same single treatment effect (fixed effect model) or different treatment effects distributed around a typical value (random effects model) holds for both, NMA and MA. For NMA, two additional assumptions must be postulated. First, the transitivity assumption implies that an indirect comparison validly estimates an unobserved head-to head comparison; thus the assumption that studies may be considered identical or exchangeable must hold across the whole set of trials in the network [60,134]. One can imagine transitivity as a network of *M* trials with *S* treatments originating from *M S*-arm trials, but that some of the arms are missing at random, meaning that the missingness of arms is not associated with the relative efficacy of the interventions [113]. Another conceivability is that transitivity is given if the distribution of effect modifiers doesn´t differ across comparisons. Transitivity can´t be tackled statistically, but reasonability can be assessed epidemiologically. Finally the consistency assumption that direct and indirect evidence are not contradictory is the prerequisite for the calculation of a mixed estimate. Consistency is the extension of transitivity to closed loops of evidence and statistical methods are at hand for verifying it [60]. Lu et al. propose the following definition: "If the discrepancy in the results of the studies goes beyond that explained by sampling error and between trial heterogeneity, then we may say the direct and indirect evidence are inconsistent as sources of effect size estimation“ [128]. Transitivity doesn´t automatically imply consistency; even if an indirect comparison via a transitive comparator C (e.g. A-C and B-C trials have the same distribution regarding the effect modifier age) yields an valid estimate but direct A-B trials are conducted in a different population regarding age, both estimates A-B_{indir} and A-B_{dir} are valid but a mixed estimate will be meaningless, since the consistency assumption is violated [60].

Lu et Ades suggest an inconsistency model where the consistency assumptions are relaxed by a random quantity ω, the inconsistency factor (ICF) and both models are compared for better model fit [128]. The number of ICF is determined by the inconsistency degrees of freedom (ICDF) which can be described by the number of independent three-way loops in the network; thus, it is derived from the number of functional parameters minus the number of loops S informed from a multi-arm trial alone, ICDF=d_f-S [135]. In their fully connected network of four interventions A represents the reference treatment, d_b= (d_{AB},d_{AC},d_{AD}), d_f= (d_{BC},d_{BD},d_{CD}) and the number of independent three way loops ICDF=3. The inconsistency equations can be parametrized as follows [128]:

$$\begin{aligned} d_{BC} &= d_{AC} - d_{AB} + \omega_{ABC} \\ d_{BD} &= d_{AD} - d_{AB} + \omega_{ABD} \\ d_{CD} &= d_{AD} - d_{AC} + \omega_{ACD} \end{aligned} \qquad (3.33)$$

Plots of the posterior mean deviance of each data point for the consistency model against the inconsistency model might allow detection of inconsistent loops in the network. Moreover, an inconsistency p-value can be derived expressing the probability given the full data y that the inconsistency variance σ_ω^2 exceeds the between trial heterogeneity σ^2 with a high value pointing to potential inconsistency ($Pr(\sigma_\omega^2 > \sigma^2 \mid y)$) [128]. However for networks with a small number of ICDF measure of variance will have wide credible intervals which handicaps detecting inconsistency [135].

The back calculation introduced by Dias et al. is an extension of the Bucher method to any network. Manipulating the basic formula for weighted averages the direct estimate $\hat{d}_{XY}^{Dir}$ and the MTC estimate $\hat{d}_{XY}^{MTC}$ for a contrast (x,y) can be transposed to derive the indirect estimate $\hat{d}_{XY}^{Rest}$ and its variance V_{XY}^{Rest}:

$$\frac{1}{V_{XY}^{Rest}} = \frac{1}{V_{XY}^{MTC}} - \frac{1}{V_{XY}^{Dir}} \quad and \quad \hat{d}_{XY}^{Rest} = \left(\frac{d_{XY}^{MTC}}{V_{XY}^{MTC}} - \frac{d_{XY}^{Dir}}{V_{XY}^{Dir}} \right) \cdot V_{XY}^{Rest} \quad (3.34)$$

Assuming a normal distribution of $\hat{d}_{XY}^{Rest}$ an inconsistency factor $\hat{\omega}_{XY}$ and its variance $Var\,\omega_{XY}$ can be derived; for the null hypothesis that there is no inconsistency (H_0: $\hat{\omega}_{XY} = 0$) a test statistic Z_{XY} which follows a standard normal distribution can be applied [129]:

$$\begin{aligned} &a.)\ \hat{\omega}_{XY} = \hat{d}_{XY}^{Dir} - \hat{d}_{XY}^{Rest} \\ &b.)\ Var(\hat{\omega}_{XY}) = Var_{XY}^{Dir} + Var_{XY}^{Rest} \\ &c.)\ Z_{XY} = \hat{\omega}_{XY} \Big/ \sqrt{Var(\hat{\omega}_{XY})} \end{aligned} \quad (3.35)$$

The node splitting method separates direct from indirect evidence for each contrast informed by both direct and indirect evidence for assessing inconsistency between the posterior distributions of $\hat{d}_{XY}^{Dir}$ and $\hat{d}_{XY}^{Ind}$ [129]. Consider the following directed acyclic graph (DAG) which reflects a random effects model for binary data for a contrast (x,y):

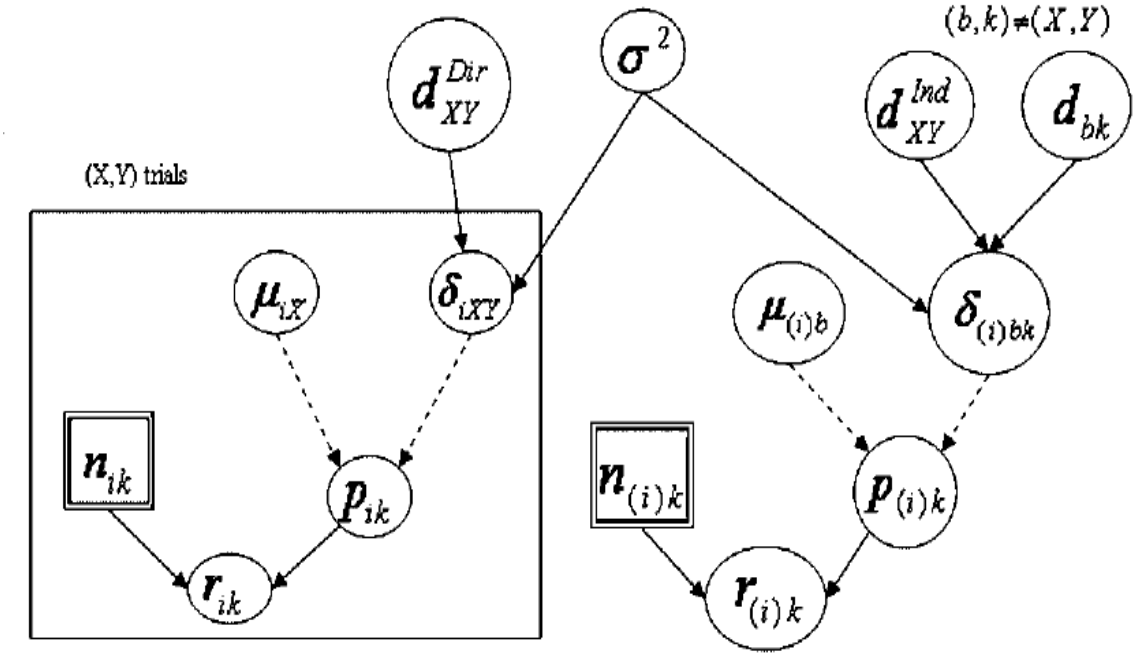


Figure 4: Directed acyclic graph for MTC for node split [129]
MTC, mixed treatment comparison.

The left side contains trials, indexed i, providing direct evidence; on the right sight the remaining MTC model is displayed with data from all other studies except XY trials. K denotes the arm of each study, b the trial specific baseline. Boxes represent constants, ellipses denote variables, either deterministic (dashed arrow indicating functional dependence) or stochastic (solid arrow) ones.

Stochastic nodes are either child nodes or parent nodes the former depending on the latter. Thus, by giving the parent nodes prior distributions the model is fully specified [117,129]. The node splitting approach draws the posterior distributions from two separate sources $\delta_{iXY} \sim N(d_{XY}^{Dir}, \sigma^2)$, based only on studies comparing the treatments X and Y directly and $\delta_{iXY} \sim N(d_{XY}^{Ind}, \sigma^2)$ based on all remaining studies in the MTC model, assuming the variance σ^2 be equal for both distributions estimated from all data.

The inconsistency parameter $\hat{\omega}_{XY} = \hat{d}_{XY}^{Dir} - \hat{d}_{XY}^{Ind}$ is examined for each node if the consistency assumption is reasonably supported by the data. A Bayesian p-value indicates the probability that $\hat{d}_{XY}^{Dir}$ exceeds $\hat{d}_{XY}^{Ind}$ (or $\hat{\omega}_{XY} > 0$); for symmetric unimodal distributions this is implemented by MCMC sampling and estimating this probability by p=2xmin (prob,1-prob) [129].

In multi-arm trials, the splitted node is used for the direct evidence $\hat{d}_{XY}^{Dir}$ whereas the remaining arms are incorporated in the indirect comparison making full use of the available evidence. For trials with more than three arms, the correlation for the unsplitted arms must be taken into account. Thus, the model for an ACD trial splitting node A-C is:

$$\begin{aligned} \theta_{jA} &= \mu_i \\ \theta_{jC} &= \mu_i + \delta_{jAC} \\ \theta_{jD} &= \mu i + \delta_{jAD} \end{aligned} \quad \text{and} \quad \begin{aligned} \delta_{jAC} &\sim N(\hat{d}_{AC}^{Dir}, \sigma^2) \\ \delta_{jAD} &\sim N(\hat{d}_{AD}, \sigma^2) \end{aligned} \tag{3.36}$$

Thus, the A-D and C-D information contributes to the indirect evidence of d_{AD} and d_{CD} if there are other A-D and C-D trials, since a three arm trial alone cannot be inconsistent [129].

The flat prior for the direct comparison parameter is assumed to follow a normal distribution ($d^{Dir} \sim N(0, 1.0e^{-4})$). The posterior mean of the deviance ($\overline{D}_{res}$), the DIC, the heterogeneity parameter $\sigma^2$and the Bayesian p-value will be examined for comparing the full MTC model with the node split models for assessing model inconsistency. The R code for data manipulation and the MTC models for node split are adapted from Dias et al. [129]. Depending on the DIC for the FE/RE MTC core models the node split will only be conducted for the model that fits the data better.

In this work transitivity for loops supported by indirect evidence only will be assessed by comparing potential effect modifiers between the sets of studies informing the indirect comparison. Consistency will be tested by the node splitting method initially and after meta-regression and sensitivity analysis to analyze if measures of reducing inconsistency have been successful.

3.2.3.3 Meta-regression of potential effect modifiers

Although a random effects model allows for incorporating heterogeneity, it doesn´t explain it. As mentioned before, MTC results will be biased if there are differences in covariates that act as treatment modifiers across the network. Thus, by incorporating covariates and modeling treatment-by-covariate interactions in a meta-regression model the impact of bias due to the violation of the similarity and/or consistency assumption might be reduced [59].

It has to be acknowledged that meta-regression suffers the limitations of observational studies, including possible bias through confounding by other study-level characteristics.

Even if individuals are randomized within trials, they are not randomized to covariate values [121]. Thus it is impossible to infer a causality from the association since an identified association may in reality reflect a true association with another (unobserved) covariate [136]. Moreover, meta-regression is vulnerable to aggregation bias, also called ecological bias or ecological fallacy. It can occur if the relationship with patient averages across trials may not be the same as the relationship for patients within trials. Thus, a within trial relationship may be masked at the study level or a study level association may not be present within the single studies each of the scenarios leading to misleading results [136]. This issue can only be overcome by performing meta-regression with individual patient data (IPD) which is generally preferable due to other features like its higher power in detecting statistic significant effects of the covariate and the possibility to re-analyze other covariates not reported at the study-level. Although meta-analysis and meta-regression on an IPD level might serve as the gold-standard it´s often unfeasible due to the unavailability of individual patient data and often the analyst has to resort to aggregated level data (AGD) [137].

Due to its low power if the number of studies in the network is limited meta-regression should only be undertaken if as a rule of thumb there are at least ten studies for each modeled covariate in the study set [121]. To keep the number of covariates at a reasonable level and to avoid „data dredging“ (namely that covariates are selected after seeing the data since the likelihood of a false positive result increases with the number of characteristics investigated), covariates that will be subjected to meta-regression should be pre-specified. Ideally, the selection of characteristics should be based on clinical grounds or supported by external evidence.

For this thesis, the following covariates were selected since it is assumed that they are likely to act as effect modifiers: for HbA1c and body weight, the covariate at baseline, the length of follow-up and the previous therapy will be evaluated; for HbA1c it is known that lower baseline levels result in a smaller treatment-induced change in HbA1c and previous therapy induces a similar effect since patients with drug history exhibit lower baseline values [138]. Moreover for some OAD´s it has been shown that they lose their efficacy and lead to secondary therapeutic failure the longer they are administered [45]; thus since length of follow-up according to the eligibility criteria ranges from 12 weeks to 3 years it may affect the relative treatment effects if it is unevenly distributed between the comparisons. For body weight, the same considerations seem plausible. For hypoglycemia, the year of study and the length of follow-up are considered for meta-regression. The rationale is that older studies may exhibit lower standards regarding the reporting of hypoglycemia compared to more recent studies (see 3.1.1.3) and that the frequency of hypoglycemia may increase with longer drug exposure especially when drugs that are prone to induce such events are compared to placebo or drugs that are unlikely to trigger hypoglycemia.
The meta-regression models for fixed (3.37) - and random effects models (3.38) are derived as follows:

$$g(y) = \theta_{ik} = \mu_i + (d_{bk} + \beta_{bk} x_i) I_{\{k \neq b\}} = \mu_i + (d_{Ak} - d_{Ab} + (\beta_{Ak} - \beta_{Ab}) x_i) I_{\{k \neq b\}} \quad where\, I_{\{u\}} \begin{cases} 1 if\ u\ is\ true \\ 0 otherwise \end{cases}$$

$$d_{AA} = 0, \beta_{AA} = 0 \qquad (3.37)$$

$$a.)\, g(y) = \theta_{ik} = \mu_i + (\delta_{i,bk} + \beta_{i,bk} x_i) I_{\{k \neq b\}} \quad where\, I_{\{u\}} \begin{cases} 1 if\ u\ is\ true \\ 0 otherwise \end{cases}$$

$$b.)\, \delta_{i,bk} \sim N(d_{bk} + \beta_{bk}, \sigma_{bk}^2) = N(d_{Ak} - d_{Ab} + (\beta_{Ak} - \beta_{AB}), \sigma_{Ab}^2 + \sigma_{Ak}^2 - 2\rho_{bk}\sigma_{Ab}\sigma_{Ak}) \quad (3.38)$$

$$d_{AA} = 0, \beta_{AA} = 0$$

X_i denotes the trial-level covariate (either categorical or continuous) and $ß_{bk}$=ß is the regression coefficient and for continuous and categorical variables, it describes, how the outcome variable (the intervention effect) changes with a unit increase in the explanatory variable and how the intervention effect differs from a reference group, respectively. Here, one common regression coefficient which assumes an identical interaction effect across all treatments with regard to the reference treatment is estimated since there is no evidence from the literature that provides a rationale for assuming independent or related regression coefficients; for those models refer to Dias et al [62]. This means all treatments related to the reference treatment exhibit the same treatment by covariate interaction and the treatment effects of functional parameters relative to each other are the same because interaction terms then cancel out (brackets for ß in equation 3.37 and 3.38 dissolve) [62].

Continuous covariates will be centered around their mean by subtracting the mean covariate value, $\bar{x}$, from each x_i to improve the mixing of the MCMC chains; thus, the treatment effects are estimated at the mean covariate value. Uncentering for any covariate value z is achieved by d-ß($\bar{x}$ - z) [62]. The categorical variable previous therapy with three values (see Table 5) is coded with k-1=2 dummy variables (d_1, d_2) with therapy naive as reference category and d_1 and d_2 taking the value 1 for no OAD at least 8 weeks before randomization and the category else, respectively. For each outcome all potential effect modifiers will be included in the full model. Variable selection will be based on the significance at an alpha level of 0.05. Thus, only covariates of which the $CrI_{0.95}$ doesn´t contain the null effect and the null hypothesis of no treatment- by- covariate interaction can be rejected will be maintained in the model. Missing values for continuous covariates will be imputed with the mean of the remaining studies. No imputation will be made for measures of uncertainty for covariate values which appear to be poorly reported since assumptions for imputation might be too strong if the proportion of missing data is substantial; therefore meta-regression for continuous covariates will be unweighted. A vague normally distributed prior will be assumed for all regression coefficients (ß~N (0,1.0e^{-4})).

Depending on the DIC for the FE/RE MTC models meta-regression will only be conducted for the model that fits the data better. To ease the assessment of plausibility of a meta-regression relationship covariate vs. outcome scatterplots will be designed for statistically significant regression coefficients. The meta-regression models provided by Dias et al. [62] were adapted and applied to the data. For all meta-regression MTC models the regression coefficients (point estimate, $CrI_{0.95}$) for all covariates will be reported to illustrate variable selection. For statistically significant treatment-by-covariate interactions the pairwise comparisons will be reported as described in section 3.2.2.1 together with the regression coefficients (point estimate, $CrI_{0.95}$) and the DIC for model comparison purposes.

3.2.3.4 Sensitivity analyses

The conduct of a systematic review involves several decisions and assumptions [121]. While many of these decisions are clearly objective, some are arbitrary or unclear and some assumptions are challengeable. It is worthwhile to prove that the results from a systematic review are not dependent on such issues. A sensitivity analysis is a repeat of the core analysis, substituting alternative decisions or ranges of values for decisions that were arbitrary or unclear or making different assumptions. A sensitivity analysis asks the question, "Are the findings robust to the decisions made in the process of obtaining them?" [121]. If core and sensitivity analyses are consistent or are conflicting, the results may be more reliable or must be interpreted more cautiously, respectively. In this review the following sensitivity analyses will be performed:

- for the endpoint HbA1c, all treatments will be assessed in SA for the appropriateness of serving as overall baseline; for each overall baseline the mean and standard deviation for the absolute efficacy and the probability of being the best treatment will be provided (see 3.2.3.5).
- the impact of the risk of bias on the robustness of results will be analyzed as described in chapter 3.1.4.4
- outliers, as described in 3.2.1.3 will be subjected to SA
- for RE models the impact of different priors on the heterogeneity parameter will be assessed. Since the posterior of the between trial standard deviation is highly sensitive to its prior and will exhibit higher heterogeneity the more vague a prior is it is advisable to subject prior assumptions to sensitivity analysis. Priors might be derived from external data like using predictive distributions from former meta-analysis on the same topic or by eliciting priors from clinical experts [109]. In this work a rather vague prior (σ~unif (0,10) reflecting a belief in a large degree of uncertainty (see 3.2.2.1) will be used in the core analysis. Tighter priors on σ (i.) (σ~unif (0,5); ii.) (σ~unif (0,2)) and a prior on the precision which follows a gamma distribution (τ~gamma (0.001,0.001)) and puts more weight on values of σ near to zero will be evaluated in sensitivity analysis [113].

3.2.3.5 Ranking of treatments

Classical MA only provides estimates of a set of pairwise comparisons and can´t yield indirect comparisons or a ranking of treatments. The output of MTC contains all possible comparisons but the outcome matrix can become confusing and difficult to interpret when the number of treatments increases.
Thus, it is helpful to provide a summary based on probabilities addressing the following questions: „What is the probability that each treatment is the best"? „What is the probability that each treatment is among the n best options"? [139]. In some cases, it may be sufficient to know which treatment is best. However, this may not be comprehensive enough. For example, the most effective treatment may be unavailable, too expensive, or associated with serious adverse effects outweighing its benefits. Moreover, a treatment may have low (high) probability to be the best but very high (low) probability of being second or third best. Also, two treatments may have similar probabilities to be the best but may have very different ranking thereafter. Thus it is important to see, what the other alternatives are and how they rank altogether.
In a Bayesian setting the ranking of treatments is implemented by choosing on overall baseline, which should ideally be connected with every other node in the network forming a

spanning tree [128]. For the baseline a separate meta-analysis is conducted including all arms with the baseline treatment to derive its absolute efficacy. Together with the estimated treatment effects for the basic parameters the arm-specific treatment effect for each node is derived and ranked; aggregation over all iterations yields the probability of each treatment for every rank. While the overall baseline isn´t interfering with the calculation of the relative treatment effects it can affect the ranking. The overall baseline is connecting the treatments across the network, so the treatments become comparable; if the baseline is poorly chosen either that it is not connected with all treatments (no spanning tree) or that the number of observations informing the absolute efficacy of the baseline treatment is low this can impact the ranking since basic parameters are missing and the baseline calculation may be instable, respectively. Moreover, choosing the treatment that has been trialed against the highest number of other treatments reduces strong correlations that may otherwise be induced between mean treatment effects for each pair of treatments which can slow convergence [113].

In this work the treatment that forms a spanning tree and includes the most observations for the calculation of the overall baseline, will be chosen as overall reference treatment. FE and RE meta-analysis will be carried out for the calculation and the DIC will be used for model selection; the predictive distributions of these posterior estimates will be used as prior for the calculation of T[k] and rank [k], the absolute efficacy and overall ranking of each treatment by adapting a code of Dias et al. [113]. The ranking of treatments will only be performed for the apparently best fitting model from preceding analyses. Depending on the outcome (if higher or lower values of T[k] are superior) the code for being the best treatment must be adapted or yields the worst treatment. The impact of different overall baselines will be evaluated exemplary for the endpoint HbA1c (see 3.2.3.4) in sensitivity analysis.

3.2.3.6 Validation of the mixed treatment comparison models

Model diagnostic by means of the DIC only allows to judge if the model fits the data well. Internal validation can be performed for the detection of outliers by the „leave one out" method or by dividing the whole study set into a test- and trainings set and calculate a prediction error. However, if the whole study set is used afterwards for estimating outcomes, it doesn´t allow to evaluate the goodness of prediction of the model regarding independent data. If the study set remains splitted, prediction is independent but a considerable amount of information is lost. Ideally, a model may be externally validated by checking its prediction against future studies within the same setting. In this work a similar approach will be accomplished; studies that don´t report all three outcomes at once will be excluded (see 3.1.1.3) from the review and can be used like having external evidence for validation. For the contrasts and outcomes that are not supported by direct evidence in the network and for which this kind of external evidence is available an external validation will be conducted by means of the „leave one out" approach.

The procedure is to compare the observed treatment effect of the excluded studies to the predictive distribution of effects that one expects based on the network analysis [62]. First the predictive treatment effect in a future trial δ_{new} is estimated from the posterior distributions of the summary estimate d and the heterogeneity parameter σ ($\delta_{new} \sim N(d,\sigma^2)$). To account for the correlation of predictive effects induced by multiple treatments and multi-arm trials a series of conditional univariate normal distributions must be used as described before (see 3.2.3.1). Predictive distributions of functional parameters are derived from the consistency equations:

$$\begin{aligned} \delta_{new_{bk}} &= \delta_{new_{Ak}} - \delta_{new_{Ab}} \\ \delta_{new_{AA}} &= 0 \end{aligned} \tag{3.39}$$

Now, the predictive treatment effect δ_{new} is applied to the baseline arm of the external trial. A flat prior is placed on the baseline risk to reflect the uncertainty about this parameter [62]. For binomial (hypoglycemia and continuous data (HbA1c and body weight), this is

$$\begin{aligned} &a.)\ p_{base} \sim Beta(a,b) \\ &b.)mc_{base} \sim N(d,\sigma^2) \end{aligned} \tag{3.40}$$

where p_{base} denotes the baseline probability of suffering a hypoglycemic event and a and b (numbers of events and non-events in the baseline arm of the external trial) represent the shape parameters and the mc_{base} is the mean change in HbA1c/ body weight of the control arm following a normal distribution with its mean d and precision $1/\sigma^2$, respectively.
The predictive outcome for the experimental arm is given by

$$\begin{aligned} &a.)\log it(p_{new}) = \log it(pbase) + \delta_{new} \\ &b.)\ mc_{new} = mc_{base} + \delta_{new} \end{aligned} \tag{3.41}$$

The predicted number of events and the mean change in HbA1c/ body weight in the experimental arm will be compared with the observed number of events and the mean change in HbA1c/ body weight of the experimental arm of the external trial. A Bayesian p-value (see 3.2.3.2) will be set-up by a step function to monitor if the predicted results are in agreement with the observed result from the external trial [62]. In addition a $CrI_{0.95}$ for the external trials will be constructed to assess if the predictive effects lie within this interval or if both $CrI_{0.95}$ overlap at least. The validation will only be performed for the apparently best fitting model from preceding analyses. The models were adapted from a code provided by Dias et al. [62].

3.2.3.7 Overview of models

Figure 5 provides on outlook over the implemented models in this thesis.

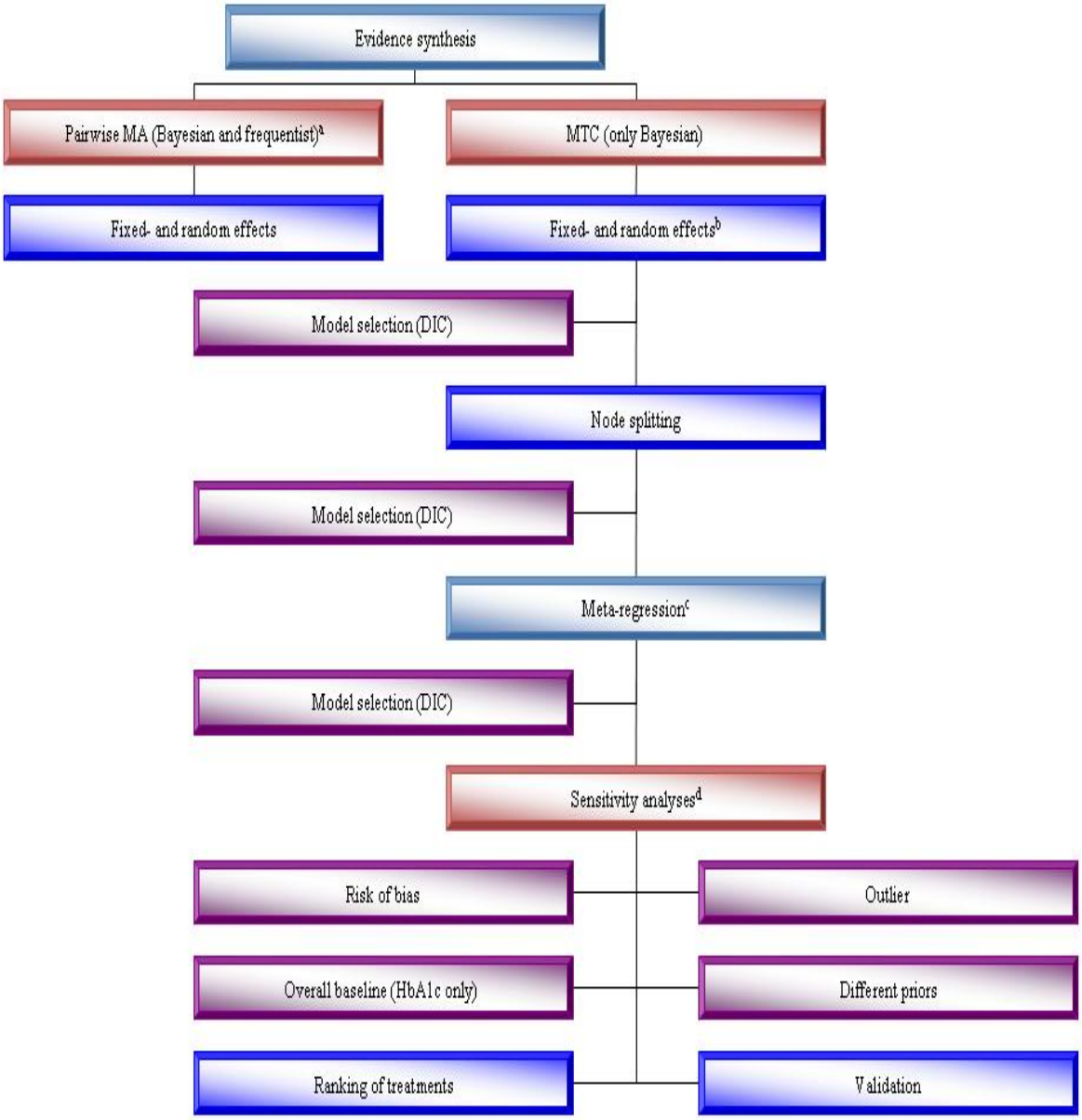


Figure 5: Model diagram
All models are run for all three outcomes HbA1c, body weight and hypoglycemia, unless otherwise specified.
[a] For frequentist models a forest plot is provided for all direct comparisons; a contour enhanced funnel plot is provided with a test for funnel plot asymmetry for direct comparisons which are supported by $\geq$10 studies (A-G and B-G).
[b] For random effects models a homogeneous variance is assumed.
[c] For meta-regression a common regression coefficient which assumes an identical interaction effect across all treatments with regard to the reference treatment was used.
[d] The apparently best fitting model from sensitivity analyses was used for ranking of treatments and validation and reanalyzing of node splitting to assess if inconsistency resolved.
MA, meta-analysis. MTC, mixed treatment comparison. DIC, deviance information criterion.

4. Results and Discussion

4.1. Systematic review

Literature search, study selection and data collection

The piloting of the eligibility criteria against a sample of clearly eligible/ not eligible studies resulted in complete agreement reasoning that eligibility criteria were precisely and adequately pre-specified. Moreover, the search strategy appeared to be comprehensive since all records from a randomly drawn test set of trials clearly fulfilling the search criteria were identified in the database of retrieved records. The literature search strategy including all information sources, the databases, the search concepts, operators and syntax is provided in Table 10 of the Appendix.

Figure 6 summarizes the flow of information from the core and updated literature research over identification of eligible trials to inclusion into evidence synthesis.

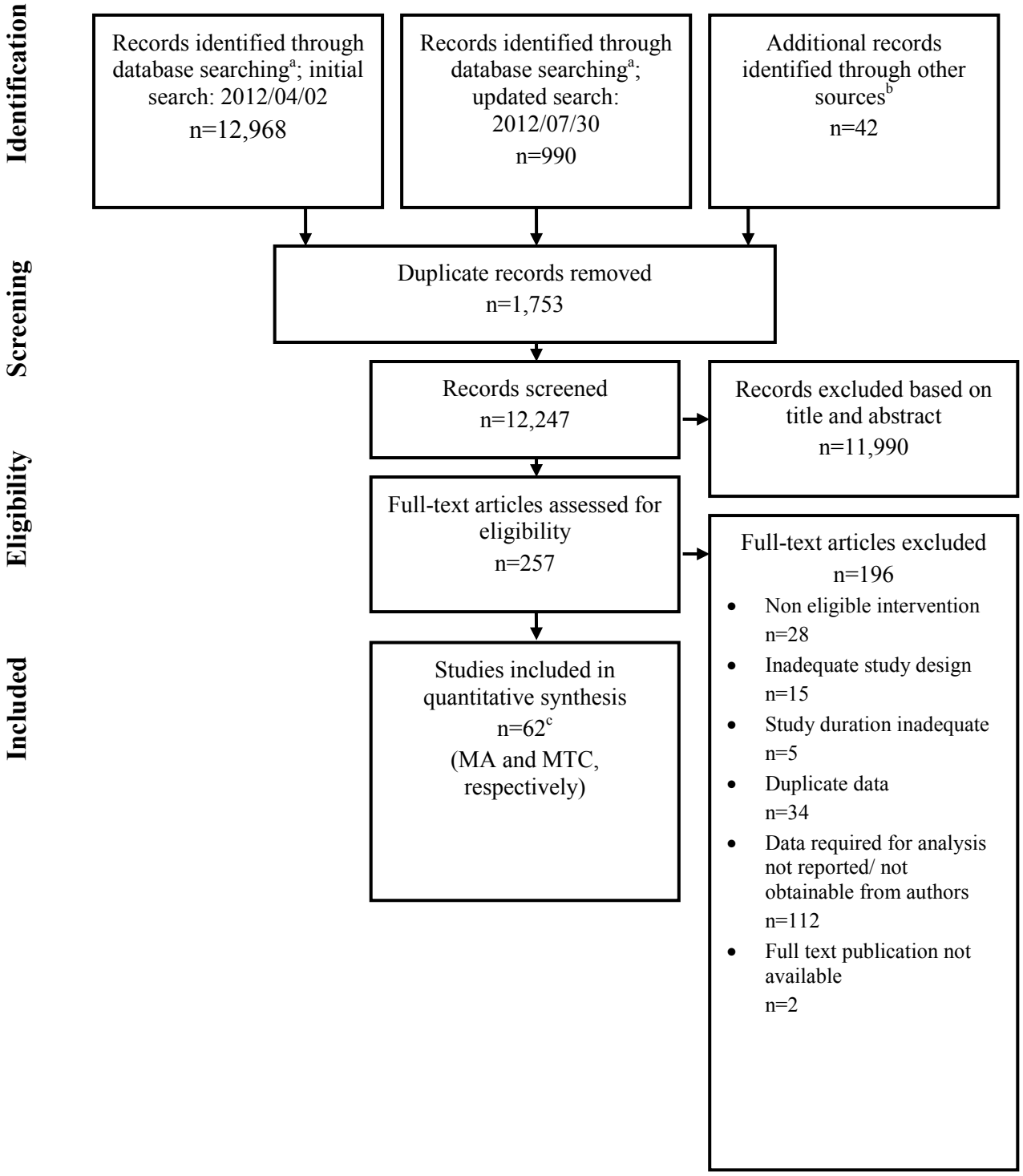


Figure 6: Flow of information through the different phases of the systematic review [76]
[a] The following bibliographic databases were searched: MEDLINE via PubMed, EMBASE and the Cochrane Central Register for controlled trials.
[b] The following additional sources were searched: ClinicalTrials.gov register, European Public Assessment Reports of the European Medicines Agency and reviews of approval of the Food and Drug Administration as well as the pharmaceutical industry trials registers of the market authorisation holder of the relevant drugs [83,87-94].
[c] One full-text article (Record No. 4945) reports two single studies, which were separately included in quantitative synthesis (Study No. 2/32).
MA, meta-analysis. MTC, mixed treatment comparison.

12,968 and 990 records were retrieved from the initial and updated literature research, respectively. Moreover, the search of other record sources resulted in 42 additional records. After removing duplicates from the merged database of all records 12,247 records were screened by study title and abstract by two reviewers (TB and SK). Finally, 257 full texts were assessed for inclusion; Japanese, Chinese, Spanish, Russian and Italian publications were evaluated with the help of native speaking scientists. 196 records had to be excluded after evaluating the full text; in Table 11 of the appendix the excluded studies are listed with the reason and a rationale for exclusion. 62 studies were eventually included in evidence synthesis.

For the majority of excluded studies (n=112) the outcome data required for analysis was not reported adequately and could not be obtained upon request from the first author (see 3.1.2.4). Nevertheless, 52 Authors were contacted to obtain missing data or information, and in 5 cases the requested piece of information was provided revealing that it is worthwhile to undertake this effort. A substantial proportion of studies (n=20) could be included by transforming the reported data to the required format or by imputation (see 3.1.3.4); therefore, it is important to consider these possibilities of data manipulation and imputation to maximize the body of evidence. The second most common reason for exclusion was duplicate data (n=34): on the one hand some records were identical assuming that the automatic removal of duplicates in the reference management software could not identify all duplicates. On the other hand some of the excluded duplicates were published twice (another journal, time of publication and study title) showing that it is enormously important to identify true duplicates in order to avoid biased outcomes. Non eligible interventions (especially due to combination therapy or studies comparing two drugs from one drug class; n=28), inadequate study design (no randomization, reviews or cross-over studies without wash out phase potentially biasing results by carry over effects; n=15), and inadequate study duration (n=5) were further reasons for exclusion. Two studies were not accessible for full-text. The well-known UKPDS 33 and 34 studies which report outcome data matching the eligibility criteria for placebo, Metformin and sulfonylureas for a follow-up of ten years were excluded since no data was obtained from the authors for the interim analysis at three years, which was set as the upper limit of eligible time frame. Since the mean follow-up of included studies is clearly shorter and below one year and most of the OAD show impaired glycemic durability due to secondary therapeutic failure, the UKPDS studies were deemed ineligible due to the distinct differences in length of follow-up.

Interestingly, the search of additional records like public and corporate clinical trial registers and regulatory authorities yielded 42 additional records some of them contributing information to other records and therefore enabling their inclusion and others providing whole summaries of clinical studies not published at all. Thus, regulatory bodies and clinical trial registers represent a useful additional source for a systematic literature search and may help to detect or avoid publication bias.

Three extension studies (study No. 31 from core study No. 30 and studies No. 44 and 45 from core study No. 43) were included, too; however regarding the latter extension studies, study arms that were switched from placebo to Metformin were not considered due to potential confounding of placebo given during core phase. Two trials with cross-over design were included (study No. 34 and 39). The former trial did not comprise a wash-out phase but outcome data was obtained from the author for the first sequence. One abstract provided sufficient information for inclusion (No. 58) and a Chinese publication could be included due to the assistance of the native speaker (No. 61).

The piloting of the data extraction form of three randomly selected studies displayed no difficulties so that the parameters and their coding could be operationalized as pre-specified and consistently between the two reviewers (TB and SK). Table 12 of the appendix depicts the core trial and patients baseline characteristics as well as the outcome data on HbA1c, body weight and hypoglycemia along with its trial-specific definition.

Overall the whole study set for quantitative analysis included 21,302 patients with a mean age of 56 (SD: 4.2) years and thereof 56% being male patients. On average, the mean duration of T2DM was 4.2 years (SD: 1.54) reflecting the fact that only a few studies included therapy naive patients (13%) and the majority of studies included patients with previous antidiabetic therapy suffering already a longer time from diabetes (87% of patients either had a pre-therapy with a wash-out of 8 weeks of their previous drug before randomization or a even shorter or no wash-out period at all). Mean baseline HbA1c and body weight were 8.2% (SD: 0.88) and 82.6 kg (SD: 8.12), respectively. Incomplete reporting for all categories of hypoglycemia (see 3.1.1.3) allowed only the analysis of data for patients suffering at least one event of any hypoglycemia; thus the other categories were not reported in this review. The trial-specific definition of hypoglycemia illustrates how inconsistently events are reported; even if probable symptomatic hypoglycemia appeared to be the prevailing category the symptoms constituting an event were often not specified. Around a quarter of all studies provided no definition at all making it difficult to assess the comparability of events between studies. Some studies reported different categories of hypoglycemia (apart from severe hypoglycemia which was reported in a lot of studies but due to rare or zero events was not adequate for analysis). Aggregation of different categories is only methodologically appropriate, if different patients are affected. Since this was never confirmed upon request from the author no aggregation was applied. If studies reported two different categories, probable symptomatic hypoglycemia (category 4) was chosen for the core analysis to ensure a certain degree of comparability since this definition occurred most often used in the included studies; to account for the potential bias, the other reported category was explored in sensitivity analysis. The mean length of follow-up was 32 weeks.
The dosing scheme of administered drugs, especially for Metformin, alpha-glucosidase inhibitors and sulfonylureas exhibited a dose titration phase and a maintenance phase which is common to increase tolerability and enhance patient compliance. Data transformation for the corresponding studies is commented in the legend of Table 12.
Table 13 of the appendix contains further parameters that were collected but are of minor importance.

Briefly, most studies did not state the criteria used to diagnose T2DM. Dietary advice was given in most of the studies which is common for study protocols in clinical diabetes trials. Diet alone was not an eligible intervention and only studies that offered dietary counseling to patients in all arms were included to avoid biased results. Clinical trials that were entirely conducted in patients with a certain condition were rare except from a few trials conducted in patients with end stage renal disease or chronic renal insufficiency since renal disorders are frequently associated with T2DM and drug elimination may be impaired. The kind of (severe) adverse events was not specified and patients discontinued for this reason were rarely reported confirming that this endpoint is less adequate for analysis.

Risk of bias and agreement

The risk of bias was assessed independently by two reviewers (TB and SK). Figure 7 illustrates the proportions of ratings of the risk of bias for all domains across all studies.

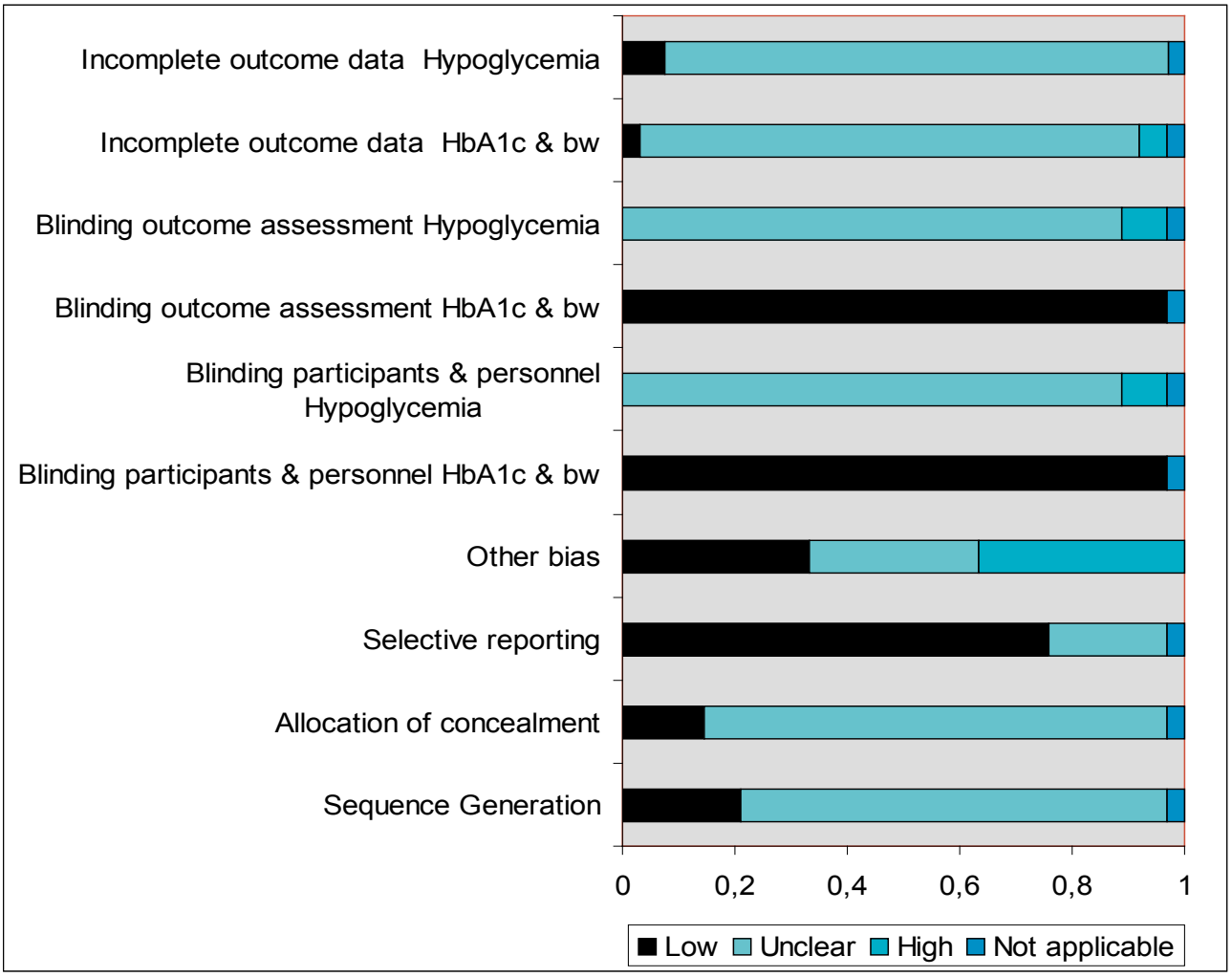


Figure 7: Methodological quality graph
Judgments about each domain of bias presented as percentages across all included studies.
If the degree of bias within one domain was reported endpoint specific the highest risk of bias (H-->U-->L) was taken for presentation.

For the domains allocation of concealment and sequence generation the risk of bias could almost never be assessed since it was not reported. Although generally accepted recommendations like the CONSORT statement clearly request authors to report the type of randomization or details of any restriction as well as mechanisms used to implement the sequence and describing the steps taken to conceal the sequence until interventions were assigned this information is still lacking in a lot of publications, even recent ones [140]. Thus, it is unclear if these points were handled properly but were just not reported or if inadequate methods possibly introduce selection bias into the analysis. The risk of bias for selective reporting was most frequently judged as low. However, the appraisal was mostly made based on whether pre-specified endpoints from the method section were fully and consistently reported in the results section. It would have been preferable to check study protocols for consistency with publication; however, there were almost no protocols available, although this is another suggestion from the CONSORT

statement. The risk of bias for HbA1c and body weight regarding the domains blinding of participants, personnel and outcome assessors was rated as low since it is unlikely to imagine that inadequate blinding or unblinding could bias these measurements. In the most studies hypoglycemia was a patient-reported outcome requiring patients to record symptoms and trialists to confirm episodes of hypoglycemia. Therefore, if drug-specific adverse events or other effects lead to unblinding, patients, personnel or outcome assessors might be influenced in recording hypoglycemic events through knowledge or belief which drug was administered. Therefore, the risk of for performance or detection bias was considered as unclear for the majority of studies.

The risk of bias for each study at the study- and outcome level together with the support for judgment is presented in Table 14 of the appendix.

In summary, according to the pre-specified criteria for the exclusion of studies at high risk for bias in sensitivity analysis (see 3.1.4.4), 17 studies were identified for HbA1c and body weight and 14 for hypoglycemia. The most common reasons for HbA1c and body weight were presumably underdosed or overdosed trial drugs impairing efficacy or enforcing side effects, respectively (HbA1c: n=6; body weight: n=1), imbalanced baseline values (HbA1c: n=3; body weight: n=6) between intervention arms (without knowledge if an adjustment was made in the statistical model for the baseline value as explanatory variable) and ongoing therapy (HbA1c/body weight: n=2) with one of the trial drugs potentially decreasing its therapeutic efficacy. For hypoglycemia, most often open label studies were assigned with a high risk of bias (n=5); other reasons were events occurring under rescue therapy (when glycemic targets weren´t met some protocols allowed adding another OAD) or an inadequate definition of the safety population potentially underestimating frequency of events (n=4).

The judgment of the risk of bias for the domain incomplete outcome data for all three outcomes was predominantly rated unclear. Although the efficacy populations and safety populations seemed to be adequately specified for most of the studies (ITT-analysis for HbA1c and body weight for patients that received the trial drug at least once and who had both baseline and at least one post-baseline measurement and safety analysis for hypoglycemia for all patients who received the trial drug at least once) the number of dropouts wasn´t reported in several studies. Even if reported numbers and reasons for each trial arm were imbalanced which might cause attrition bias due to systematic differences between groups in withdrawals from a study it was unclear if the reasons were likely to affect the outcome. For continuous outcome data the LOCF method is commonly applied for imputation to enable an ITT analysis; however its appropriateness is questionable. Early values from the trough of lowering HbA1c carried forward may overestimate the drugs efficacy in studies with long follow-up when T2DM progresses and glycemic durability of drugs is lost. On the other hand some drugs like the insulin sensitizer Pioglitazon unfold their full efficacy only after 4 weeks; thus earlier readings of HbA1c carried forward might bias its efficacy downwards. Due to this remaining uncertainty the risk of bias was deemed unclear. Some studies with substantial loss to follow-up were analyzed per protocol or only for observed cases; for HbA1c and body weight, the risk of bias was therefore considered high.

Finally two studies - one abstract and Chinese publication - were entirely excluded from sensitivity analysis, since the abstract didn´t provide enough information for appraisal and misunderstandings in the cooperation with the native speaker can´t be completely ruled out, respectively.

Interrater agreement for study selection based upon title and abstract, full-text, data collection and for assessing the risk of bias was measured using the kappa statistic. For the first three items agreement was complete with no need of involvement of the third reviewer. For the appraisal of methodological quality, interrater agreement on all domains apart other bias was complete. Still, the kappa statistic was quite high (κ=0.89; $CI_{0.95}(\kappa)$:[0.82;0.95]) reflecting reliable agreement.
For this domain the disagreement was resolved by judgment of the third reviewer (UM). The influence of disagreement was analyzed by excluding these studies from sensitivity analysis.

4.2 Evidence synthesis

4.2.1 Pairwise meta-analysis

4.2.1.1 Fixed effects and random effects models

The results of the direct pairwise frequentist and Bayesian meta-analyses are presented in Table 15 (for the corresponding forest plots, see Figure 8 of the appendix).

Table 15: Fixed effects- and random effects Bayesian and frequentist direct pairwise meta-analysis

Comparison	Frequentist MA		Bayesian MA	
1. HbA1c, %	**FE: MD [$CI_{0.95}$]**	**RE: MD [$CI_{0.95}$]; tau**	**FE: MD [$CrI_{0.95}$]**	**RE: MD [$CrI_{0.95}$]; sigma**
A-B	-1.14 [-1.27;-1.01]	-1.18 [-1.62;-0.74]; 0.47	-1.14 [-1.28;-1.01]	-1.18 [-1.95;-0.39]; 0.73
A-C	-1.07 [-1.21;-0.93]	-1.18 [-1.48;-0.88]; 0.29	-1.07 [-1.21;-0.93]	-1.2 [-1.73;-0.74]; 0.47
A-D	-0.97 [-1.2;-0.75]	-1 [-1.36;-0.65]; 0.24	-0.96 [-1.21;-0.74]	-1.03 [-3.02;0.76]; 1.02
A-E[a]	-0.6 [-0.95;-0.25]		-0.6 [-0.95;-0.25]	
A-F	-0.7 [-0.91;-0.49]	-0.79 [-1.33;-0.25]; 0.5	-0.7 [-0.91;-0.49]	-0.8 [-2.1;0.47]; 0.98
A-G	-0.73 [-0.78;-0.68]	-0.69 [-0.8;-0.58]; 0.2	-0.73 [-0.76;-0.68]	-0.69 [-0.81;-0.56]; 0.23
B-C	0.23 [0.06;0.4]	0.11 [-0.26;0.47]; 0.34	0.23 [0.06;0.4]	0,09 [-0.58;0.7]; 0.55
B-D[a]	-0.15 [-0.36;0.06]		-0.15 [-0.36;0.06]	
B-E	-0.15 [-0.41;0.11]	-0.15 [-0.41;0.11]; 0	-0.15 [-0.41;0.11]	-0.17 [-5.5;5.11]; 2.231
B-F	0.5 [0.26;0.74]	0.39 [-0.56;1.34]; 0.66	0.5 [0.26;0.74]	0.39 [-6.35;7.12]; 3.31
B-G	0.23 [0.17;0.28]	0.26 [0.18;0.35]; 0.08	0.23 [0.18;0.28]	0.26 [0.16;0.37]; 0.1
C-D	-0.05 [-0.2;0.1]	-0.05 [-0.2;0.1]; 0	-0.05 [-0.2;0.1]	-0.05 [-0.35;0.23]; 0.17
C-E	-0.03 [-0.13;0.08]	-0.08 [-0.31;0.15]; 0.19	-0.03 [-0.13;0.09]	-0.09 [-0,81;0.6]; 0.51
C-F	0.52 [0.2;0.83]	0.52 [0.2;0.83]; 0	0.51 [0.19;0.83]	0.52 [-1.41;2.34]; 0.95
C-G	0.2 [0.1;0.31]	0.17 [-0.09;0.44]; 0.21	0.2 [0.1;0.31]	0.16 [-1.46;1.72]; 0.88
D-G	0.41 [0.24;0.58]	0.41 [0.24;0.58]; 0	0.41 [0.24;0.58]	0.38 [-4.99;5.7]; 2.13
E-F[a]	0.07 [-0.16;0.3]		0.07 [-0.16;0.3]	
F-G[a]	-0.1 [-0.38;0.18]		-0.1 [-0.38;0.18]	
2. Body weight, kg	**FE: MD [$CI_{0.95}$]**	**RE: MD [$CI_{0.95}$]; tau**	**FE: MD [$CrI_{0.95}$]**	**RE: MD [$CrI_{0.95}$]; sigma**
A-B	0.04 [-0.35;0.43]	0.04 [-0.35;0.43]; 0	0.04 [-0.35;0.43]	0.01 [-0.67;0.63]; 0.43
A-C	1.95 [1.61;2.29]	2.28 [1.33;3.23]; 1.05	1.95 [1.61;2.29]	2.34 [0.78;4.07]; 1.7
A-D	2.53 [1.97;3.09]	2.53 [1.87;3.19]; 0.3	2.53 [1.96;3.09]	2.55 [-0.33;5.49]; 1.52
A-E[a]	0.1 [-0.48;0.68]		0.1 [-0.48;0.68]	
A-F	-0,03 [-0.59;0.52]	-0,03 [-0.59;0.52]; 0	-0.03 [-0.6;0.51]	-0.13 [-2.1;1.68]; 1.05
A-G	0.53 [0.4;0.66]	0.52 [0.37;0.67]; 0.15	0.53 [0.4;0.66]	0.52 [0.34;0.68]; 0.16
B-C	2.66 [2.31;3.02]	2.56 [1.69;3.43]; 0.87	2.66 [2.3;3.02]	2.57 [1.19;4.02]; 1.27
B-D[a]	3.5 [2.79;4.21]		3.5 [2.81;4.21]	
B-E	2.16 [1.46;2.87]	2.64 [0.43;4.85]; 1.49	2.16 [1.45;2.87]	2.67 [-4.93;10.27]; 4.1
B-F	0.13 [-0.56;0.82]	0.13 [-0.56;0.82]; 0	0.13 [-0.56;0.81]	0.06 [-6.28;6.28]; 2.85
B-G	1.21 [1.01;1.41]	1.34 [0.95;1.73]; 0.5	1.21 [1.01;1.41]	1.37 [0.86;1.95]; 0.7

C-D	1.3 [0.68;1.91]	1.25 [0.5;2.01]; 0.42	1.3 [0.68;1.9]	1.2 [-0.68;2.9]; 1.17
C-E	-0.32 [-0.68;0.04]	-0.32 [-0.68;.0.04]; 0	-0.32 [-0.68;0.05]	-0.3 [-1.09;0.55]; 0.52
C-F	-2.52 [-3.21;-1.82]	-2.3 [-3.87;-0.72]; 1.24	-2.52 [-3.22;-1.83]	-2.25 [-6.63;2.32]; 2.8
C-G	-1.13 [-1.47;-0.8]	-1.17 [-1.62;-0.71]; 0.25	-1.14 [-1.47;-0.8]	-1.2 [-3.46;1.07]; 1.23
D-G	-1.8 [-2.39;-1.21]	-1.8 [-2.78;-0.82]; 0.57	-1.8 [-2.4;-1.22]	-1.82 [-8.46;4.77]; 3.14
E-F[a]	-1.4 [-1.88;-0.92]		-1.4 [-1.88;-0.92]	
F-G[a]	1.3 [0.86;1.74]		1.3 [0.86;1.74]	
3. Hypoglycemia	**FE: OR [$CI_{0.95}$]**	**RE[b]: OR [$CI_{0.95}$]; tau**	**FE: OR [$CrI_{0.95}$]**	**RE[b]: OR [$CrI_{0.95}$]; sigma**
A-B	1.35 [0.64;2.84]	1.35 [0.64;2.84]; 0	1.49 [0.69;3.34]	1.61 [0.38;8.2]; 0.91
A-C	5.15 [2.51;10.58]	5.15 [2.51;10.58]; 0	7.03 [3.44;15.36]	7.93 [2.08;40.37]; 1.02
A-D	2.0 [0.3;13.36]	2.0 [0.3;13.36]; 0	2.42 [0.4;26.66]	2.74 [0.02;470.6]; 2.65
A-E[a]	0.92 [0.02;50.37]		1.02 [0.001;665.8]	
A-F	1.16 [0.53;2.52]	1.16 [0.53;2.52]; 0	1.18 [0.52;2.64]	1.18 [0.16;7.9]; 1.04
A-G	0.99 [0.6;1.63]	0.99 [0.6;1.63]; 0	1.16 [0.72;1.93]	1.13 [0.64;2.05]; 0.38
B-C	1.38 [0.8;2.37]	1.38 [0.8;2.37]; 0	1.41 [0.81;2.49]	1.35 [0.42;3.5]; 0.62
B-D[a]	0.9 [0.32;2.53]		0.87 [0.28;2.5]	
B-E	3.27 [1.78;6.03]	3.27 [1.78;6.03]; 0	3.4 [1.86;6.35]	5.69 [0.01;10,667]; 3.384
B-F	0.89 [0.33;2.42]	0.89 [0.33;2.42]; 0	0.88 [0.3;2.53]	0.87 [0.001;1,458]; 3.5
B-G	0.61 [0.39;0.94]	0.61 [0.39;0.94]; 0	0.54 [0.34;0.84]	0.55 [0.31;0.97]; 0.35
C-D	0.23 [0.15;0.35]	0.21 [0.1;0.45]; 0.56	0.21 [0.14;0.31]	0.2 [0.03;1.5]; 1.35
C-E	0.94 [0.73;1.21]	0.94 [0.73;1.21]; 0	0.94 [0.73;1.22]	0.92 [0.49;1.62]; 0.39
C-F	0.27 [0.12;0.59]	0.23 [0.07;0.75]; 0.67	0.2 [0.09;0.44]	0.15 [0.002;9.37]; 2.51
C-G	0.22 [0.14;0.34]	0.22 [0.11;0.42]; 0.41	0.21 [0.13;0.33]	0.21 [0.01;3.3]; 1.47
D-G	0.86 [0.28;2.6]	0.86 [0.28;2.6]; 0	0.85 [0.27;2.67]	0.94 [0.001;873]; 3.17
E-F[a]	0.28 [0.08;1.06]		0.25 [0.05;0.92]	
F-G[a]	0.5 [0.01;25.3]		0.51 [0.001;339]	

For frequentist and Bayesian meta-analysis the point estimate, the $CI_{0.95}$ and the heterogeneity standard deviation tau and the point estimate, the $CrI_{0.95}$ and the heterogeneity standard deviation sigma are reported, respectively.

The contrasts D-E, D-F and E-G are missing since these comparisons are not supported by direct evidence.

[a] No meta-analysis since only one study exists for these comparisons.

[b] For hypoglycemia the heterogeneity standard deviation tau and sigma are reported on the log scale, respectively.

HbA1c, glycated haemoglobin. MA, meta-analysis. FE, fixed effects model. RE, random-effects model. MD, difference in means. CI, confidence interval, CrI, credible interval. OR, odds ratio.

The comparisons D-E, D-F and E-G were not informed by direct evidence and for the contrasts A-E (Nateglinide vs. placebo), B-D (Pioglitazone vs. Metformin), E-F (Acarbose vs. Nateglinide) and F-G (Vildagliptin vs. Acarbose) only one study was available; thus meta-analysis was performed for the remaining 14 comparisons. For the drug class meglitinides and alpha-glucosidase inhibitors, the comparisons B-E, C-E and C-F were only conducted with Repaglinide and Acarbose, respectively.

Because binomial likelihoods with zero cells are tolerated, usually, no precautions need to be taken in the case of an occasional trial with a zero cell count within the Bayesian setting [113]. This is an advantage, because some frequentist methods add for log odds ratios an arbitrary constant, usually 0.5, to cells in order to obtain non-infinite estimates of treatment effects and non-infinite variance, but in so doing they might introduce bias.
However, in extreme cases of many small trials and when several trials have zero cells, models can be numerically unstable, either failing to converge, or converging to a posterior with very high standard deviation or can´t be run at all. Since for some of the comparisons WinBUGS produced an error message and didn´t run increments of 0.5 for each cell of studies containing zero events were added for all forthcoming analyses to enable computation.

For every Bayesian MA and all of the other subsequent Bayesian models convergence and the MCMC error were monitored for all relevant parameters. The corresponding trace plots displayed good mixing of the two Markov chains indicating sufficient convergence so that it can be assumed that the 14,000 iterations from the sampling phase yield reliable estimates from the posterior distributions. The MCMC error for all parameters was below 5% so that the number of iterations from the sampling phase is likely to produce stable estimates. Randomly inspected autocorrelation plots exhibited a fast decrease of correlation over iterations towards zero so that the thin-out function seems to prevent autocorrelation effectively. Density plots of the posterior distribution of outcome parameters appeared to be normally distributed. All the ancillary output together with the data matrices is provided as supplementary material (see electronic data carrier).

In general frequentist and Bayesian estimates were in agreement; for RE models Bayesian credible intervals were always wider than the corresponding frequentist confidence intervals since the former take into account the uncertainty in terms of a prior for the heterogeneity parameter whereas the latter ignore it and therefore underestimate the uncertainty of the pooled treatment effect and the between-study variance.

HbA1c

Compared to placebo all treatments showed a statistically significant difference in means; Metformin, sulfonylureas and Pioglitazone exhibited a stronger effect than Nateglinide, alpha-glucosidase inhibitors and gliptins lowering HbA1c levels by 1% and 0.6-0.7%, respectively, which is consistent with findings from other systematic reviews [6,23,37-40,43,46,48]. Contrasted against Metformin, sulfonylureas, alpha-glucosidase inhibitors and gliptins were less beneficial resulting in an moderate increase of HbA1c; for sulfonylureas and alpha-glucosidase inhibitors the difference was only statistically significant for FE models (either frequentist or Bayesian) whereas for gliptins statistical significant inferiority was proven for all models due to the large number of trials (n=17) yielding narrower $CI_{0.95}$/$CrI_{0.95}$ for RE models. Moreover, Acarbose was likely to be inferior to sulfonylureas and gliptins were less efficacious than Pioglitazone resulting in an increase in HbA1c of approximately 0.52% and 0.41%, respectively, which was significant for all models apart from the Bayesian RE model that had a very wide $CrI_{0.95}$ due to the small number of underlying studies.

The pooled treatment effects of comparisons A-B, A-C, A-F, A-G, B-C, B-F, C-E and C-G exhibited considerable heterogeneity with statistically significant tests for heterogeneity and values of I^2 between 0.69-0.93 (see Figure 8, appendix). For those comparisons, RE models provide more reliable and conservative estimates than FE models since they allow for heterogeneity. For the comparison A-D based upon the statistical test the null hypothesis (p=0.09) couldn´t be rejected, whereas the value for I^2 (0.58) pointed to a moderate extent of heterogeneity underscoring the fact that the test might sometimes be underpowered in detecting heterogeneity if the number of studies is mall.

By examining the outlying observations in the forest plots of comparisons with a high degree of heterogeneity it can be noticed that most outliers are due to unusual dosing of trial drugs or ongoing therapy with one of the trial drugs in line with the findings from the assessment of the risk of bias (see 4.1). For example, regarding the A-B comparison the study from List et al. [141] possibly yielded a very moderate effect for Metformin vs. placebo (MD: -0.55%) due to the low dose of 1,500mg/d which is usually given in doses up to 3,000mg/d; this study was also identified to be at high risk for bias and will be excluded from sensitivity analysis.

As mentioned before due to the incorporated uncertainty for the between-study variance in the Bayesian setting for RE models (for example A-D), frequentist $CI_{0.95}$ yielded sometimes a statistically significant effect, that disappeared when the corresponding $CrI_{0.95}$ were examined. Moreover, for contrasts which are only supported by a small number of studies such as B-E, B-F, C-F or D-G there is insufficient data for adequately estimating the between-trials variation resulting in overly wide $CrI_{0.95}$ compared to $CI_{0.95}$. On the one hand this might be triggered by the relatively vague uniform prior placed on σ; for example if a tighter gamma prior is instead assumed for the inverse variance (τ~gamma(0.001,0.001)) the $CrI_{0.95}$ for the relative treatment effect of B-E narrows from [-5.5;5.11] to [-1.29;1.16]. In addition, when the information increases for estimating the heterogeneity parameter as for the comparison A-G $CrI_{0.95}$ and $CI_{0.95}$ are almost of the same width. On the other hand, when there are only a few trials available one should take a sceptical view towards the frequentist CI which might produce too narrow interval estimates.

Body weight

In comparison to placebo and Metformin both sulfonylureas and Pioglitazone led to a substantial weight gain that reached statistical significance for all models except for the RE Bayesian model for A-D (which was informed only by three studies) being in accordance with earlier findings [6,45]. The weight gain was more pronounced compared to Metformin rather than placebo even if this was not evidenced by the direct comparison A-B. Meglitinides that tend to induce clinically relevant weight gain, too [6], showed an increased weight compared with placebo (Nateglinide), whereas this effect was not paralleled in the direct comparison of Repaglinide vs. Metformin. Although incretin based therapies are claimed to be beneficial regarding obese diabetics propagating a weight loss under GLP-1 analogues and a weight neutral effect of gliptins this might not hold for the latter since in this work gliptins induced moderate significant weight gain compared to placebo and Metformin [142]. The fact that alpha-glucosidase inhibitors induced weight loss in some comparisons might be attributed to their gastrointestinal side effects like nausea, diarrhea and flatulence which might in turn influence the eating behaviour of patients rather than a real beneficial effect on weight.

According to the statistical test and I^2 heterogeneity was present for the pooled treatment effects for comparisons A-C, B-C, B-E, B-G and C-F with values for I^2 ranging from 0.7-0.84.

Appraisal of potential outliers of forest plots revealed that some but not all of these studies were also identified before to bear a high risk for bias mainly due to high doses near the ceiling dose or insufficient dose escalation, pre-therapy with the trial drug or unusual high weight losses possibly caused by different adherence to diet or exercise between intervention arms.
Again, compared with frequentist models Bayes estimates for RE models with a small number of studies led to wider$CrI_{0.95}$ (in some cases non-significant e.g. A-D and B-E), whereas contrasts informed by a large study set, for example B-G yielded similar $CrI_{0.95}$. High levels of heterogeneity may partly be induced by the vague uniform prior; anew, with a tighter gamma prior for the inverse variance (τ~gamma (0.001,0.001)) the $CrI_{0.95}$ for the relative treatment effect of C-G shrinks from [-3.46;1.07] to [-1.97;-0.49].

Hypoglycemia

It is known that sulfonylureas and meglitinides bear the highest risk for suffering hypoglyemic events [6]. These results were confirmed regarding the pooled estimates of sulfonylureas and meglitinides vs. almost all other comparators resulting in mostly significant OR ranging from 3.7-7.93 and 3.57-5.69 depending on the applied model, respectively. However, for the comparisons of Nateglinide vs. placebo and sulfonylureas vs. Metformin, the direct evidence couldn´t detect a harmful effect with non-significant OR around 1.

Statistically significant heterogeneity wasn´t detectable for any comparison; merely the C-D contrast exhibited borderline heterogeneity (p=0.057; I^2=0.6). Moreover, comparisons like C-F and C-G displayed moderate heterogeneity reflected by I^2 values of 0.42 and 0.49, respectively questioning again the power of the heterogeneity test if the amount of evidence is low. Potential outliers from the forest plots were not identified by the assessment of risk of bias.

Interestingly, point and interval estimates of frequentist and Bayesian models didn´t match so well compared to the endpoints HbA1c and body weight which might due to the zero cell correction and unstable Bayesian posteriors in case of few, small studies were several trials have zero cells (e.g. Repaglinide vs. Metformin). Moreover, some Bayesian models like A-C (σ= 1.02) or A-G (σ=0.38) were heterogeneous whereas the frequentist models were not (I^2=0); this might be explained by the fact that the deviant studies were small and contributed minor weights to the summary estimate. Just as for HbA1c and body weight, tighter priors resulted in narrower credible intervals. Considering the contrast C-G if the original uniform prior (σ~unif (0,10)) is substituted by a more informative one (σ~unif(0,2)) the $CrI_{0.95}$ diminished from [0.01;3.3] to [0.06;0.71]. In light of these findings it must be conceded that the original prior might have been determined too vague and a tighter uniform prior might have been more adequate. For example for a mean treatment effect of an odds ratio of 1.5 a uniform prior of (σ~unif (0,10)) assumes a $CrI_{0.95}$ of [0.0008;27,050], whereas a uniform prior of (σ~unif(0,2)) yields an $CrI_{0.95}$ of [0.2;11], the former emanating from an unrealistic belief and the latter seeming to provide an reasonable prior [77].

In summary, regarding all endpoints some comparisons exhibit heterogeneity to a considerable extent. Some reasons like the uncommon dosing of drugs or pre-therapy with the trial drug could be identified beforehand by means of the risk of bias assessment.
Even if existent heterogeneity wasn´t further explored in meta-analysis the impact will be evaluated for the network by performing meta-regression and removing outliers from sensitivity analyses (see 4.2.2.3 and 4.2.2.4).

However there might be other reasons contributing to clinical and methodological diversity which are not covered by the bias appraisal, meta-regression or sensitivity analysis. Therefore, it must be considered that a certain degree of residual heterogeneity will be immanent in these comparisons. Thus, estimates of RE models must be interpreted with caution and the assumption of lumping treatments for analysis - such as sulfonylureas, meglitinides or gliptins - might not always be fully justified, if this aggregation into drug classes is causing the heterogeneity.

4.2.1.2 Assessment of publication bias

Contour-enhanced funnel plots for the comparisons A-G and B-G (number of observations $\geq$ 10) for all outcomes are presented in Figure 9 of the appendix. The dashed and dotted lines represent the summary effects for FE and RE models, respectively. In the graphical visualization of the test for funnel plot asymmetry, the horizontal line represents the null hypothesis of no association between estimated effects and the measure of uncertainty; thus, if the line moves towards the vertical this may be interpreted as a clue for asymmetry that might be caused by publication bias.

From the visual inspection of the comparisons A-G (for body weight and hypoglycemia) and B-G (for HbA1c and hypoglycemia) an asymmetrical scattering was deemed unlikely; the p-values of the Eggers test for funnel plot asymmetry were likewise not statistically significant ($p_{(\text{A-G body weight})}=0.56$; $p_{(\text{A-G hypoglycemia})}=0.25$; $p_{(\text{B-G HbA1c})}=0.18$; $p_{(\text{B-G hypoglycemia})}=0.53$). The pooled relative treatment effects of A-G (for HbA1c) and B-G (for body weight) appear somehow asymmetrically with borderline p-values of 0.08 and 0.09, respectively. However, the asymmetry of the funnel plot of comparison A-G seems not be caused by a gap of unpublished studies. For the comparison B-G there might be a gap at the left bottom of the funnel plot in the area of non-significant results indicating no weight gain under gliptins vs. Metformin which could be caused by unpublished results. However, since body weight was usually not the primary endpoint in these clinical trials and effect estimates located in the gap would support weight neutrality of gliptins, it is more likely that other reasons than publication bias triggered asymmetry or it occurred just by chance.

4.2.2 Mixed treatment comparison analysis

Figure 10 pictures the network diagram of the full MTC models.

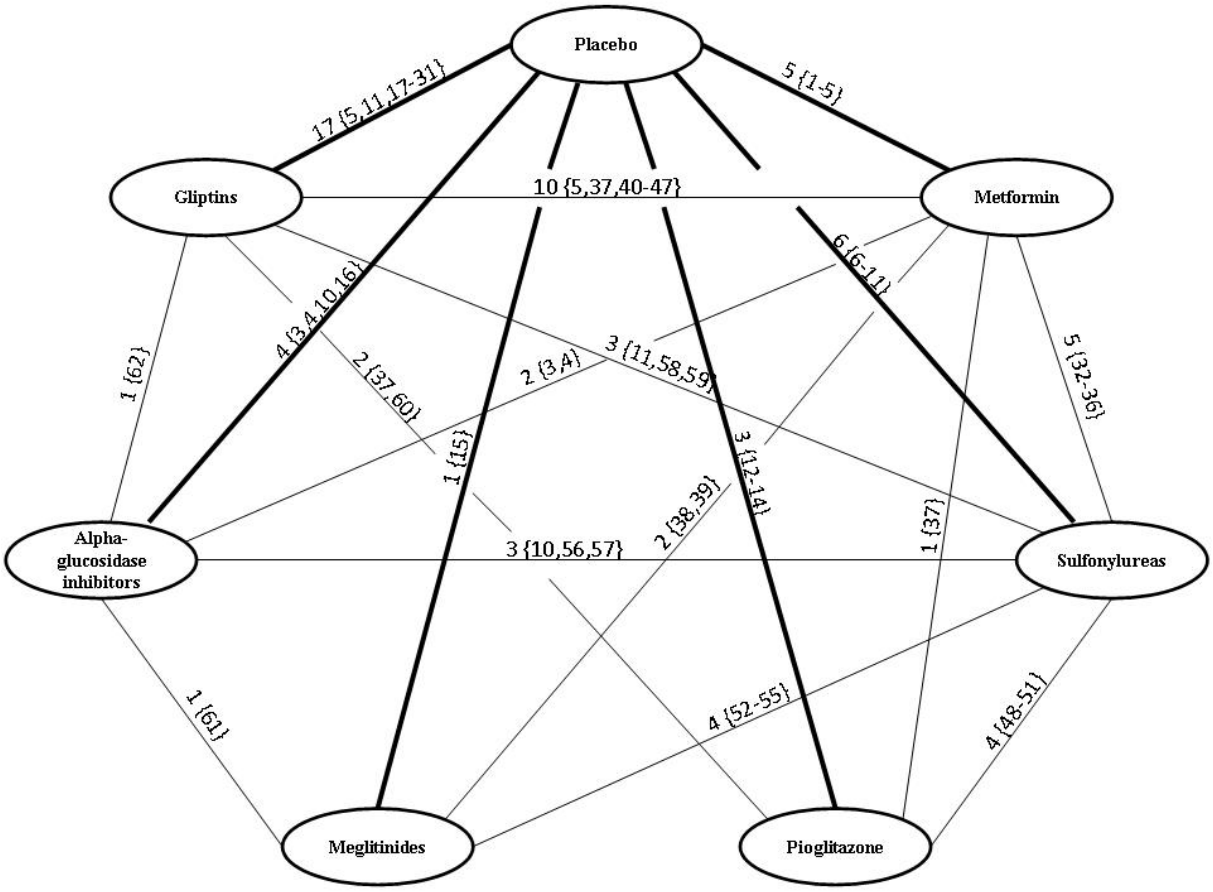


Figure 10: Network diagram of full mixed treatment comparison models.
Treatments and direct comparisons are represented by nodes and edges, respectively. Bold lines represent the basic parameters connecting all other treatments to the overall baseline (placebo).

The nodes and the edges represent the interventions and the direct comparisons, respectively and the numbers above the edges indicate the numbers of direct comparisons supporting the two linked treatments. Overall 62 studies comprising 130 arms (thereof 6 three arm trials) constitute the body of evidence. The network is well connected and only three contrasts (D-E, D-F and E-G) are not informed by head to head trials. The nodes placebo and Metformin are connected with all other treatments in the network (spanning tree) and are suited as overall baseline; in addition, with 77% of all direct comparisons these nodes account for the largest share of evidence reflecting the fact that most pivotal trials for approval were conducted against placebo and thereafter, Metformin was the second most common drug since it represents the standard of care. However, the number of placebo trials (n=37) outnumbered the Metformin trials (n=20); thus it was selected as overall baseline for analyses (bold lines). Therefore according to section 3.2.2.1 for K=7 treatments there are T=7(7-1)/2=21 potential comparisons, 7-1=6 basic parameters and 21-7+1=15 functional parameters. Taking placebo as overall baseline the vector of baseline parameters d_b and functional parameters d_f is

$$d_b = (d_{AB}, d_{AC}, d_{AD}, d_{AE}, d_{AF}, d_{AG})$$
$$d_f = (d_{BC}, d_{BD}, d_{BE}, d_{BF}, d_{BG}, d_{CD}, d_{CE}, d_{CF}, d_{CG}, d_{DE}, d_{DF}, d_{DG}, d_{EF}, d_{EG}, d_{FG})$$

The functional parameters are derived from the basic parameters through the consistency equations:

$$d_{BC} = d_{AC} - d_{AB}$$
$$d_{BD} = d_{AD} - d_{AB}$$
$$\vdots$$
$$d_{FG} = d_{AG} - d_{AF}$$

Of course, the remaining evidence is not ignored in the network. For example, the relative treatment effect of d_{BC} is not only estimated by the indirect evidence from the overall baseline A; in fact all other possible bridging comparators (D-G) which are in turn estimated via the basic parameters plus the direct evidence yield the MTC estimate of d_{BC}.

4.2.2.1 Fixed effects and random effects models

The results of the MTC analysis contrasted with the results from direct pairwise MA are displayed in Table 16.

Table 16: Fixed- and random effects Bayesian mixed treatment comparison analysis

1. HbA1c, % **FE: MD [$CrI_{0.95}$];** **RE: MD [$CrI_{0.95}$]**							
Dir / MTC	A	B	C	D	E	F	G
A		-1.14 [-1.28;-1.01]; -1.18 [-1.95;-0.39]	-1.07 [-1.21;-0.93]; -1.2 [-1.73;-0.74]	-0.96 [-1.21;-0.74]; -1.03 [-3.02;0.76]	-0.6 [-0.95;-0.25][c]	-0.7 [-0.91;-0.49]; -0.8 [-2.1;0.47]	-0,73 [-0,76;-0,68]; -0.69 [-0.81;-0.56]
B	-1 [-1.06;-0.94]; -1.05 [-1.2;-0.91]		0.23 [0.06;0.4]; 0.09 [-0.58;0.7]	-0.15 [-0.36;0.06][c]	-0.15 [-0.41;0.11]; -0.17 [-5.5;5.11]	0.5 [0.26;0.74]; 0.39 [-6.35;7.12]	0,23 [0,18;0,28]; 0.26 [0.16;0.37]
C	-0.94 [-1.02;-0.86]; -1 [-1.16;-0.84]	0.06 [-0.02;0.14]; 0.06 [-0.11;0.22]		-0.05 [-0.2;0.1]; -0.05 [-0.35;0.23]	-0.03 [-0.13;0.09]; -0.09 [-0.81;0.6]	0.51 [0.19;0.83]; 0.52 [-1.41;2.34]	0,2 [0,1;0,31]; 0.16 [-1.46;1.72]
D	-1.05 [-1.16;-0.94]; -1.06 [-1.27;-0.86]	-0.05 [-0.16;0.06]; -0.01 [-0.23;0.22]	-0.11 [-0.22;0]; -0.07 [-0.27;0.14]		-[b]	-[b]	0,41 [0,24;0,58]; 0.38 [-4.99;5.7]
E	-0.92 [-1.03;-0.81]; -0.99 [-1.22;-0.76]	0.08 [-0.03;0.19]; 0.06 [-0.17;0.29]	0.02 [-0.08;0.11]; 0 [-0.21;0.21]	0.13 [-0.01;0.26]; 0.07 [-0.21;0.35]		0.07 [-0.16;0.3][c]	-[b]
F	-0.62 [-0.75;-0.48]; -0.65 [-0.89;-0.42]	0.39 [0.25;0.52]; 0.4 [0.15;0.65]	0.33 [0.19;0.46]; 0.35 [0.1;0.59]	0.43 [0.27;0.6]; 0.41 [0.12;0.71]	0.31 [0.16;0.45]; 0.34 [0.07;0.63]		-0,1 [-0,38;0,18][c]
G	-0.75 [-0.79;-0.7]; -0.73 [-0.85;-0.62]	0.25 [0.2;0.3]; 0.32 [0.19;0.45]	0.19 [0.12;0.26]; 0.26 [0.1;0.42]	0.3 [0.2;0.41]; 0.33 [0.12;0.54]	0.17 [0.07;0.28]; 0.26 [0.03;0.49]	-0.13 [-0.27;0]; -0.09 [-0.33;0.16]	
SD[a]	0.24 [0.18;0.31]						
Model fit	$\overline{D}_{res}$	p_D	**DIC**				
FE	301	59.44	359.46				
RE	132.47	107.15	239.62				

2. Body weight, kg **FE: MD [$CrI_{0.95}$];** **RE: MD [$CrI_{0.95}$]**							
Dir / MTC	A	B	C	D	E	F	G
A		0.04 [-0.35;0.43]; 0.01 [-0.67;0.63]	1.95 [1.61;2.29]; 2.34 [0.78;4.07]	2.53 [1.96;3.09]; 2.55 [-0.33;5.49]	0.1 [-0.48;0.68][c]	-0.03 [-0.6;0.51]; -0.13 [-2.1;1.68]	0,53 [0,4;0,66]; 0,52 [0,34;0,68]
B	-0.61 [-0.79;-0.43]; -0.63 [-0.97;-0.28]		2.66 [2.3;3.02]; 2.57 [1.19;4.02]	3.5 [2.81;4.21][c]	2.16 [1.49;2.87]; 2.67 [-4.93;10.27]	0.13 [-0.56;0.81]; 0.06 [-6.28;6.28]	1,21 [1,01;1,41]; 1,37 [0,86;1,95]
C	1.78 [1.58;1.99]; 1.82 [1.47;2.19]	2.39 [2.18;2.61]; 2.45 [2.08;2.83]		1.3 [0.68;1.9]; 1.2 [-0.68;2.9]	-0.32 [-0.68;0.05]; -0.3 [-1.09;0.55]	-2.52 [-3.22;-1.83]; -2.25 [-6.63;2.32]	-1,14 [-1,47;-0,8]; -1.2 [-3.46;1.07]
D	2.67 [2.33;3]; 2.7 [2.19;3.22]	3.28 [2.93;3.62]; 3.33 [2.79;3.88]	0.88 [0.53;1.23]; 0.88 [0.35;1.4]		-[b]	-[b]	-1,8 [-2,4;-1,22]; -1.82 [-8.46;4.77]
E	1.11 [0.83;1.39]; 1.29 [0.77;1.83]	1.72 [1.43;2.01]; 1.92 [1.4;2.46]	-0.67 [-0.94;-0.41]; -0.53 [-1.01;-0.04]	-1.56 [-1.97;-1.14]; -1.41 [-2.07;-0.72]		-1.4 [-1.88;-0.92][c]	-[b]
F	-0.45 [-0.74;-0.16]; -0.35 [-0.88;0.18]	0.16 [-0.14;0.46]; 0.27 [-0.28;0.84]	-2.23 [-2.54;-1.94]; -2.18 [-2.72;-1.63]	-3.12 [-3.54;-2.69]; -3.05 [-3.75;-2.36]	-1.56 [-1.88;-1.24]; -1.65 [-2.28;-1.02]		1.3 [0.86;1.74][c]
G	0.58 [0.46;0.7]; 0.62 [0.35;0.89]	1.2 [1.03;1.36]; 1.24 [0.93;1.57]	-1.2 [-1.4;-1]; -1.21 [-1.57;-0.84]	-2.08 [-2.41;-1.74]; -2.08 [-2.61;-1.55]	-0.52 [-0.81;-0.25]; -0.68 [-1.22;-0.14]	1.04 [0.75;1.32]; 0.97 [0.43;1.51]	
SD[a]	0.53 [0.37;0.71]						

Model fit	$\overline{D}_{res}$	p_D	**DIC**
FE	237.8	67.91	305.71
RE	137.43	105.45	242.88

3. Hypoglycemia **FE: OR [$CrI_{0.95}$];** **RE: OR [$CrI_{0.95}$]**							
Dir / MTC	A	B	C	D	E	F	G
A		1.49 [0.69;3.34]; 1.61 [0.38;8.2]	7.03 [3.44;15.36]; 7.93 [2.08;40.37]	2.42 [0.4;26.66]; 2.74 [0.02;470.6]	1.02 [0.001;665.8][c]	1.18 [0.52;2.64]; 1.18 [0.16;7.9]	1.16 [0.72;1.93]; 1.13 [0.64;2.05]
B	2.3 [1.47;3.49]; 2.34 [1.47;3.66]		1.41 [0.81;2.49]; 1.35 [0.42;3.5]	0.87 [0.28;2.5][c]	3.4 [1.86;6.35]; 5.69 [0.01;10,667]	0.88 [0.3;2.53]; 0.87 [0.001;1,458]	0.54 [034;0.84]; 0.55 [0.31;0.97]

C	5.79 [3.76;8.67]; 5.8 [3.7;8.79]	2.56 [1.82;3.51]; 2.53 [1.73;3.57]		0.21 [0.14;0.31]; 0.2 [0.03;1.5]	0.94 [0.73;1.22]; 0.92 [0.49;1.62]	0.2 [0.09;0.44]; 0.15 [0.002;9.37]	0.21 [0.13;0.33]; 0.21 [0.01;3.3]
D	1.34 [0.77;2.2]; 1.35 [0.74;2.29]	0.59 [0.36;0.91]; 0.59 [0.33;0.96]	0.23 [0.16;0.33]; 0.23 [0.15;0.35]		-[b]	-[b]	0.85 [0.27;2.67]; 0.94 [0.001;873]
E	5.73 [3.55;8.98]; 5.73 [3.34;9.32]	2.52 [1.74;3.57]; 2.48 [1.6;3.7]	0.99 [0.77;1.25]; 0.99 [0.72;1.35]	4.4 [2.83;6.73]; 4.42 [2.54;7.23]		0.25 [0.05;0.92][c]	-[b]
F	1.39 [0.71;2.45]; 1.38 [0.67;2.5]	0.62 [0.33;1.04]; 0.6 [0.31;1.05]	0.24 [0.13;0.4]; 0.24 [0.13;0.41]	1.08 [0.53;1.94]; 1.07 [0.49;2.05]	0.25 [0.13;0.42]; 0.25 [0.12;0.44]		0.51 [0.001;339][c]
G	1.24 [0.83;1.82]; 1.26 [0.82;1.87]	0.55 [0.39;0.76]; 0.55 [0.37;0.78]	0.22 [0.15;0.3]; 0.22 [0.15;0.31]	0.97 [0.59;1.51]; 0.99 [0.57;1.64]	0.22 [0.15;0.33]; 0.23 [0.14;0.35]	0.97 [0.51;1.69]; 1 [0.5;1.84]	
SD[a]	0.19 [0.01;0.44]						

Model fit	$\overline{D}_{res}$	p_D	**DIC**
FE	116.86	71.33	188.19
RE	112.53	76.89	189.42

Results are cross-tabulated; below and above the diagonal the outcomes for MTC and direct pairwise MA are arranged, respectively.

[a] Between-study standard deviation estimated from RE MTC model assuming a homogeneous variance. For variances of pairwise comparisons, see Table 15. The SD for hypoglycemia is reported on the log scale.

[b] No direct evidence available.

[C] Only one study available for direct evidence.

HbA1c, glycated haemoglobin. MD, difference in means. FE, fixed effects. RE, random effects. CrI, credible interval. Dir, direct evidence. MTC, mixed treatment comparison. SD, standard deviation. $\overline{D}_{res}$, total residual deviance. p_D, effective number of parameters. DIC, deviance information criterion. KG, kilogram. OR, odds ratio.

Generally MTC estimates and direct evidence were consistent for most of the pairwise comparisons. Since the former are supported by one or several indirect comparisons they yield more precise interval estimates compared to the results from direct pairwise MA. Moreover, some comparisons gained statistical significance by incorporating the indirect comparisons; e.g., the difference in means for HbA1c of Pioglitazone vs. placebo was not significant when direct evidence was considered solely whereas the combination of direct and indirect evidence (via the bridging comparators B, C and G) reached statistical significance ($CrI_{0.95}^{Dir}$:[-3.02;0.76] vs. $CrI_{0.95}^{MTC}$:[-1.27;-0.86]). The increase of precision which may even lead to statistically significant results is one of the major advantages of MTC compared to common pairwise MA as long as the assumption of consistency is fulfilled (see 4.3.2).

HbA1c

For the RE model, the difference in means for the contrasts D-E, D-F and E-G was 0.07 ($CrI_{0.95}$:[-0.21;0.35]), 0.41 ($CrI_{0.95}$:[0.12;0.71]) and 0.26 ($CrI_{0.95}$:[0.03;0.49]), respectively. The statistical significant superiority of Pioglitazone vs. alpha-glucosidase inhibitors and meglitinides over gliptins seems reasonable; however these estimates were not supported by direct evidence.
For the contrast A-E the direct estimate yielded a lower efficacy compared to the MTC estimate (see 4.2.2.2.1). The posterior mean of the commonly estimated between-study standard deviation σ was 0.24 ($CrI_{0.95}$:[0.18;0.31]); contrasted to the heterogeneity of direct pairwise MA it could be estimated with considerable precision; this is another advantage of MTC models, namely that the heterogeneity parameter can be estimated relatively exact even for comparisons that are not informed by many studies by borrowing strength from other contrasts as long as the assumption of homogeneous variance holds. The total residual deviance $\overline{D}_{res}$ of the RE model is smaller compared to the FE model. Although the effective number of parameters penalizes the RE model it nevertheless yields a considerably lower DIC (DIC_{RE}:239.62 vs. DIC_{FE}:359.36) reasoning that the RE model fits the data better. Thus, the following analyses were only run for RE models.

Body weight

RE estimates for the difference in means of the comparisons D-E, D-F and E-G which are only accessible via indirect evidence were -1.41 ($CrI_{0.95}$:[-2.07;-0.72]), -3.05 ($CrI_{0.95}$:[-3.75;-2.36]) and -0.68 ($CrI_{0.95}$:[-1.22;-0.14]), respectively. Since Pioglitazone is the drug that induces the heaviest weight gain and alpha-glucosidase inhibitors lead to weight loss compared to meglitinides, the direction and magnitude of relative treatment effects $D\text{-}E_{Ind}$ and $D\text{-}F_{Ind}$ seems plausible as well as the weight gain of meglitinides in comparison to gliptins. Since Metformin is known to exert a weight loosing effect compared to placebo [42] it is remarkable that this was not reflected by the direct evidence. However, the MTC estimate confirmed the beneficial effect of Metformin on weight which was statistically significant -0.61 ($CrI_{0.95}$:[-0.79;-0.43]). Moreover, the result from direct evidence that Nateglinide induces no weight gain compared to placebo was refuted by the MTC estimate; driven by the indirect evidence meglitinides yielded a statistically significant weight gain contrasted to placebo 1.29 ($CrI_{0.95}$:[0.77;1.83]) (see. 4.2.2.2.1). The heterogeneity parameter σ was 0.53 ($CrI_{0.95}$:[0.37;0.71]). Again, regarding model fit, the RE model performed better than the FE model resulting in a smaller DIC (even though the effective number of parameters was larger) and was applied to all remaining analyses.

Hypoglycemia

RE indirect estimates for the odds ratios of D-E, D-F and E-G were 4.42 ($CrI_{0.95}$:[2.54;7.23]), 1.07 ($CrI_{0.95}$:[0.49;2.05]) and 0.23 ($CrI_{0.95}$:[0.14;0.35]), respectively. Even if direct evidence is lacking for these comparisons the findings seem realistic since Meglitinides are prone to cause hypoglycemic events in contrast to Pioglitazone or Gliptins. Regarding the comparison of Meglitinides vs. placebo and sulfonylureas vs. Metformin their harmful effect is underlined by the MTC estimates (OR_{MTC}(A-E):5.73; $CrI_{0.95}$:[3.34;9.32]; OR_{MTC}(B-C):2.53; $CrI_{0.95}$: [1.73;3.57]), whereas this was not reflected by the direct evidence (OR_{Dir}(A-E):1.02; $CrI_{0.95}$: [0.001;665.8]; OR_{Dir}(B-C):1.35; $CrI_{0.95}$:[0.42;3.5]). The between-study standard deviation σ amounted to 0.19 ($CrI_{0.95}$: [0.01;0.44]) on the log scale. Overall goodness of fit was slightly better for the FE model compared to the RE model yielding a DIC of 188.19 and 189.42, respectively; moreover, in case of a comparable goodness of fit the FE model seems preferable since it provides the best estimate of the true unknown treatment effect and thus was maintained for further analyses.

In summary, the comparison of relative treatment effects for some contrasts revealed a discrepancy between the direct evidence and the MTC outcomes. For HbA1c regarding the comparison A-E, the disagreement might be due to the fact that administered drugs from the class of meglitinides differed between direct and indirect evidence. For body weight and hypoglycemia clinical considerations and findings from other studies provide compelling reasons that the MTC results driven by the indirect evidence are more trustworthy than the direct evidence. This is a key strength of multiple treatment analyses that they are able to provide additional information when direct evidence is possibly biased. If inconsistency can be detected by statistical means it will be evaluated by the node splitting procedure.

The comparisons D-E, D-F and E-G are only derived through indirect comparisons. Based on clinical grounds they are likely to yield reasonable results. Nevertheless, the transitivity assumption must be assessed even if consistency can´t be measured statistically. This will be evaluated by comparing the distribution of potential effect modifiers between the sets of studies informing the indirect comparison. Another way to assess the validity of indirect estimates is to quantify how good they predict effects compared to evidence from external trials (see 4.2.2.6).

4.2.2.2 Assessment of consistency and transitivity

4.2.2.2.1 Consistency

Node splitting was performed for all nodes that were supported by both direct and indirect evidence (all contrasts apart from D-E, D-F and E-G). The node split was only performed for the MTC models that fitted the data better (RE for HbA1c/body weight and FE for hypoglycemia). The results of the node split procedure are summarized in Table 17.

Table 17: Node splitting procedure

	Model fit				Source of evidence/ Inconsistency estimate/ Bayesian p-value				
Model	$\overline{D}_{res}$	p_D	**DIC**	**SD**	**MTC**	**Direct** $\hat{d}_{XY}^{Dir}$	**Indirect** $\hat{d}_{XY}^{Ind}$	**Inconsistency estimate** $\hat{\omega}_{XY} = \hat{d}_{XY}^{Dir} - \hat{d}_{XY}^{Ind}$	**Inconsistency p-value[b]**
1. HbA1c, %									
MTC (RE)[a]	132.47	107.15	239.62	0.24					
Node split					**MD**				
1-2	132.56	107.34	239.91	0.23	-1.05	-1.15	-1	-0.15	0.32
1-3	132.25	106.68	238.93	0.23	-1.00	-1.16	-0.89	-0.28	0.09
1-4	132.44	107.82	240.26	0.24	-1.06	-1.01	-1.1	0.09	0.7
1-5	131.97	107.62	239.59	0.24	-0.99	-0.6	-1.06	0.46	0.15
1-6	132.85	107.79	240.65	0.24	-0.65	-0.67	-0.68	0.01	0.95
1-7	132.19	107.57	239.76	0.24	-0.73	-0.69	-0.85	0.16	0.19
2-3	132.69	107.7	240.39	0.24	0.06	0.13	0.02	0.12	0.52
2-4	133.03	107.84	240.87	0.24	-0.01	-0.15	0.03	-0.17	0.55
2-5	133.25	107.45	240.7	0.24	0.06	-0.13	0.13	-0.26	0.33
2-6	131.85	108.82	240.67	0.24	0.40	0.48	0.39	0.09	0.74
2-7	132.4	107.38	239.78	0.24	0.32	0.28	0.36	-0.08	0.54
3-4	132.72	107.89	240.6	0.24	-0.07	-0.06	-0.08	0.02	0.92
3-5	132.05	107.36	239.41	0.24	0.00	-0.09	0.14	-0.23	0.3
3-6	132.48	108.29	240.78	0.24	0.35	0.53	0.27	0.26	0.34
3-7	131.94	107.84	239.78	0.24	0.26	0.17	0.3	-0.12	0.48
4-7	132.42	107.68	240.1	0.24	0.33	0.39	0.32	0.07	0.77
5-6	132.61	106.53	239.14	0.24	0.34	0.07	0.46	-0.39	0.21
6-7	132.43	107.84	240.27	0.25	-0.09	-0.1	-0.09	-0.02	0.96

2. Body weight, kg									
MTC (RE)[a]	137.43	105.45	242.88	0.53					
Node split					**MD**				
1-2	137.2	103.97	241.17	0.5	-0,63	-0.05	-0.89	0.84	0.02
1-3	135.96	105.58	241.54	0.52	1.82	2.15	1.58	0.57	0.12
1-4	137.13	105.85	242.98	0.53	2.7	2.53	2.8	-0.27	0.62
1-5	138.1	103.67	241.77	0.49	1.29	0.1	1.57	-1.47	0.02
1-6	136.52	106.05	242.57	0.53	-0.35	-0.36	-0.16	-0.2	0.73
1-7	136.04	105.19	241.23	0.53	0.62	0.49	0.96	-0.48	0.11
2-3	137.94	105.98	243.91	0.53	2.45	2.55	2.39	0.16	0.69
2-4	137.75	106.11	243.87	0.53	3.33	3.5	3.28	0.23	0.74
2-5	136.95	105.61	242.56	0.53	1.92	2.41	1.76	0.65	0.29
2-6	137.48	105.75	243.23	0.53	0.27	0.19	0.25	-0.06	0.93
2-7	136.89	106.25	243.14	0.53	1.24	1.35	1.06	0.29	0.37
3-4	137.5	105.62	243.12	0.52	0.88	1.24	0.63	0.61	0.27
3-5	138.68	105.11	243.8	0.51	-0.53	-0.28	-0.86	0.59	0.23
3-6	139.1	106.01	245.11	0.53	-2.18	-2.25	-2.19	-0.06	0.91
3-7	137.25	106.06	243.31	0.54	-1.21	-1.23	-1.18	-0.05	0.89
4-7	137.56	105.95	243.51	0.53	-2.08	-1.78	-2.2	0.41	0.47
5-6	136.75	105.78	242.53	0.53	-1.65	-1.41	-1.75	0.34	0.62
6-7	136.96	105.53	242.49	0.53	0.97	1.3	0.88	0.43	0.51
3. Hypoglycemia									
MTC (FE)[a]	116.86	71.33	188.19						
Node split					**log OR**				
1-2	117.34	72.46	189.79		0.83	0.72	0.86	-0.14	0.74
1-3	116.73	72.15	188.88		1.76	1.93	1.61	0.32	0.32
1-4	117.68	72.37	190.05		0.29	0.91	0.21	0.7	0.54
1-5	117.34	72.46	189.79		1.75	-0.01	1.74	-1.75	0.52
1-6	116.08	72.37	188.45		0.33	0.56	-0.24	0.8	0.18
1-7	115.18	71.9	187.08		0.22	0.04	0.53	-0.49	0.12
2-3	112.3	72.13	184.44		0.94	0.35	1.18	-0.82	0.02
2-4	117.1	72.18	189.28		-0.53	-0.26	-0.64	0.39	0.48
2-5	116.05	72.11	188.16		0.92	1.23	0.73	0.5	0.2
2-6	117.6	72.2	189.8		-0.48	-0.43	-0.61	0.18	0.76

2-7	115.18	71.9	187.08		-0.60	0.04	-0.48	-0.49	0.12
3-4	117.12	72.35	189.47		-1.47	-1.56	-1.13	-0.42	0.39
3-5	116.94	72.34	189.29		-0.01	-0.06	0.28	-0.34	0.34
3-6	117.3	72.06	189.36		-1.43	-1.63	-1.32	-0.3	0.59
3-7	117.65	72.16	189.8		-1.51	-1.55	-1.54	-0.01	0.97
4-7	117.43	72.02	189.45		-0.03	0.06	-0.08	0.13	0.78
5-6	117.66	72.19	189.85		-1.39	-1.39	-1.47	0.09	0.88
6-7	118.01	72.56	190.58		-0.03	-0.68	-0.07	-0.61	0.81

[a] For HbA1c/body weight and hypoglycemia RE and FE models were split, respectively depending on which model fitted the data better (see 4.2.2.1).

[b] The two-sided Bayesian inconsistency p-value is derived as follows p=2xmin (prob,1-prob); p is the probability over all iterations that $\hat{d}_{XY}^{Dir}$ exceeds $\hat{d}_{XY}^{Ind}$ (or $\hat{\omega}_{XY} > 0$). $\overline{D}_{res}$, total residual deviance. p_D, effective number of parameters. DIC, deviance information criterion. SD, between-study standard deviation. MTC, mixed treatment comparison. RE, random effects. FE, fixed effects. MD, difference in means. OR, odds ratio.

The left hand side of the table contains $\overline{D}_{res}$, p_D, DIC and the between-study standard deviation σ for model comparison between full MTC models and the node split models. A decrease of the DIC or σ indicates that a certain node that is freed from the consistency equation results in a better overall fit [129]. On the right hand side of the table the relative treatment effects of the full MTC models and the node split models separating the whole evidence into direct and indirect parts are compared together with the inconsistency estimate $\hat{\omega}_{XY}$ and the two-sided Bayesian p-value. Moreover it becomes clear how the MTC estimate is influenced by both sources of evidence.

HbA1c

Model comparison via the DIC revealed that no splitted node remarkably improved model fit (changes in σ weren´t detectable, either). This is paralleled by the inconsistency p-value which reached never statistical significance implicating that the null hypothesis of consistency can´t be rejected. However, for some comparisons like A-E, A-C or A-G the p-value came very close to the threshold of 0.05 suggesting that some degree of inconsistency might exist. Since this test suffers from low power if both sources of evidence are only supported from a small number of studies and since present heterogeneity may hinder to detect inconsistency a non-significant p-value should not be interpreted like inconsistency can be ruled out definitely. Moreover, for A-E the relative treatment effect showed that the MTC estimate is dominated by the indirect evidence.

Although proof or hint of inconsistency suggests that pooling of direct and indirect evidence might be doubtful it doesn´t settle the question which one is more reliable. This must be carefully contemplated by reconsidering all the evidence from a clinical perspective. Examination of baseline characteristics between the sets of evidence might yield obvious reasons for discrepancy but there may be unknown covariates which might cause inconsistency as well.
However, post-hoc explanations for inconsistency are of limited value and suitable measures like adjusting for potential effect modifiers and accounting for bias possibly introducing inconsistency should be determined prior to synthesis [129].
For the comparison A-E, the inconsistency might be attributed to the fact that the node meglitinides was differently defined between direct and indirect comparisons. Whereas the direct RCT was undertaken with Nateglinide, all other bridging trials (study No. 38,39, 52-55) apart from one (No. 61) included Repaglinide which seems to be superior to Nateglinide; thus indirect evidence yields more favourable results than direct evidence. In this case inconsistency is more a subject of node definition rather than which source of evidence is more trustworthy.
This highlights again the importance that nodes in a network have to be reasonably specified; separating the drug class meglinitindes into two independent nodes Nateglinide and Repaglinide might resolve inconsistency but increases complexity.

Body weight

For body weight, neither comparing the DIC nor σ between the full MTC model and the node split models yielded a model that could benefit from separation direct and indirect evidence. The apparent conflict comparing MTC and direct estimates (see 4.2.2.1) for the comparison A-B and A-E was confirmed by the node splitting procedure, resulting in an inconsistency estimate of 0.84 (p=0.02) and -1.47 (p=0.02), respectively; anew, both MTC estimates are driven by indirect comparisons contributing more information to the pooled estimate. Some other contrasts, e.g. B-E or C-D yielded a certain degree of discrepancy

regarding the inconsistency estimate $\hat{\omega}_{XY}$ with values of 0.65 and 0.61, respectively but p-values were not significant underlining again the issue of low power.
For the comparison A-B no baseline characteristics could be identified that might have caused inconsistency; however indirect evidence seems to be more reliable since Metformin is likely to induce weight loss. This possibility to draw conclusions based on indirect comparisons if direct evidence might be biased is another appealing feature of MTC models.
For the A-E comparison inconsistency might again be explained by the differing node definition; therefore the more potent Repaglinide is likely to trigger weight gain in indirect comparisons more intensively than Nateglinide in the placebo-controlled trial.

Hypoglycemia

The comparison of the DIC demonstrated that the node split model B-C exhibited a better model fit after cancelling the consistency equation on that node resulting in a DIC of 184.44 vs. a DIC of 188.19 for the full MTC model. This was paralleled by a significant inconsistency estimate of -0.82 (p=0.02). Strikingly, the other discrepancy for the node A-E (see 4.2.2.1) yielded a very high inconsistency estimate of -1.75 but a p-value which is far from significance (p=0.52). Even if the ratio of odds ratios for indirect vs. direct evidence of 5.75 is overly inconsistent the statistical test failed to detect it; this may be due to the fact, that the test is underpowered since there is only one direct RCT. Both MTC estimates were again controlled by the indirect evidence.

For the contrast B-C no obvious reason could be detected by comparing trial characteristics; however, the indirect evidence is likely to be more valid than the direct evidence since sulfonylureas commonly cause hypoglycemias in contrast to Metformin. Since for the node sulfonylureas five different drugs were administered in the whole study set and some of the pairwise comparisons were heterogeneous, inconsistency might be linked again to the question of how nodes were determined.
For the comparison A-E, the higher OR for hypoglycemic events from the indirect evidence compared to the direct evidence might again be caused by the stronger efficacy of Repaglinide which is accompanied by a higher potency to cause more harmful effects, too.

In summary, post-hoc explanations for inconsistency could hardly be identified apart from the fact that the node definition for some aggregated drug classes might be questionable. If the meta-regression models and the planned sensitivity analyses are able to reduce inconsistency in the network this will be evaluated by performing the node splitting procedure for the inconsistent nodes again thereafter.

4.2.2.2.2 Transitivity

For indirect comparisons without direct evidence, the transitivity assumption that an indirect comparison validly estimates an unobserved head-to head comparison must be explored on clinical grounds by assessing bridging comparators and baseline characteristics. For the comparisons of D-E, D-F and E-G Table 18 summarizes the potential effect modifiers for each relevant bridging comparator.

Table 18: Assessment of the transitivity assumption of indirect comparisons

Indirect comparison	Anchor treatment (drug,dose/d)	Follow-up, wk	Mean age (SD), y	Men, %	Mean duration of existing T2DM (SD), y	Mean bl HbA1c (SD), %	Mean bl body weight (SD), kg	Previous therapy	Definition Hypoglycemia	Heterogeneity[b]
1. Meglitinides (E) vs. Pioglitazone (D)										
1.1 Anchor: Placebo (A)										
A-D (12-14)	Placebo	16-26	55.5 (9.6)	49	NR	9.9 (1.9)	90.4 (15.8)	3	14: probably symptomatic	HbA1c
A-E (15)	Placebo	12	58.6 (10.7)	59.6	NR	7.2 (0.7)	77.4	1	Confirmed	NA
1.2 Anchor: Metformin (B)										
B-D (37)	Metformin (up to 2500mg)	26	53.7 (11)	60.3	2.7 (3.7)	8.5 (1.2)	86.8 (18.9)	2	Symptomatic	NA
B-E (38,39)	Metformin (500-3000mg)	16-20	60.5 (9.1)	69.5	5.3 (NR)	8.3 (1.3)	74.5 (9.8)	3	Symptomatic	Body weight
1.3 Anchor: Sulfonylureas (C)										
C-D (48-51)	48:Gliclazide (160-320mg; mean dose: 184mg); 49,50: Glibenclamide (1.75-15mg); 51:Glimepiride (2-8mg)	52	55.6 (10.4)	53.8	4.6 (5.4)	8.8 (1.1)	86.2 (17.8)	49: 2; 48,50,51: 3	Symptomatic, measured or confirmed	Hypoglycemia
C-E (52-55)	52: Glipizide (515mg); 53,54: Glibenclamide (1.75-15mg); 55: Gliclazide (80-160mg)	16-60	58.2 (9.6)	60.5	5.4 (NR)	7.7 (NR)	76.8 (NR)	55:1; 52-54: 3	52: not specified; 53: symptomatic, measured or confirmed; 54, 55: probably symptomatic	HbA1c

2. Alpha Glucosidase Inhibitors (F) vs. Pioglitazone (D)										
2.1 Anchor: Placebo (A)										
A-D (12-14)	See 1.1									
A-F (3,4,10,16)	Placebo	12-108	57 (NR)	55.1	3.8 (NR)	7.7 (NR)	82.7 (NR)	4, 16: 1; 3:2; 10: 3	10: probably symptomatic	HbA1c
2.2 Anchor: Metformin (B)										
B-D (37)	See 1.2									
B-F (3,4)	Metformin (1500-1700mg)	24-36	57.8 (9)	61.9	4.8 (5.3)	8.7 (0.9)	85.7 (15.6)	4:1; 3:2	Not specified	HbA1c
2.3 Anchor: Sulfonylureas (C)										
C-D (48-51)	See 1.3									
C-F (10,56,57)	10: Tolbutamide (750mg); 56: Glimepiride (1-6mg); 57: Gliclazide (80-160mg)	24-26	56.4 (NR)	53.5	4.3 (NR)	8.2 (NR)	84 (NR)	57:1; 56:2; 10: 3	Probably symptomatic	Body weight
2.4 Anchor: Gliptins (G)										
D-G (37,60)	37: Sitagliptin (100mg); 60: Vildagliptin (100mg)	24-26	53.1 (10.9)	61.5	2.4 (3.5)	8.6 (1.1)	84.9 (NR)	2	37: symptomatic; 60: confirmed	Body weight
F-G (62)	Vildagliptin (100mg)	24	51.8 (10.2)	61.1	1.2 (2.4)	8.6 (0.9)	72 (NR)	2	Confirmed	NA
3. Gliptins (G) vs. Meglitinides (E)										
3.1 Anchor: Placebo (A)										
A-E (15)	See 1.1									
A-G (5,11,17-31)	Placebo	12-108	57.1 (10.7)	55.1	3.8 (NR)	7.9 (0.9)	80.5 (NR)	19: 1; 11,30,31,20-29: 2; 5,17,18:3	5,11,17,18,20,21,25,27,28: not specified; 19,22,26: symptomatic; 30,31,23,24: confirmed; 29: symptomatic, measured or confirmed	HbA1c
3.2 Anchor: Metformin (B)										

B-E (38,39)	See 1.2									
B-G (5,37,40-47)	Metformin (500-2500mg)	24-104	55.1 (11.1)	50.7	2.3 (NR)	8.4 (1.2)	86.6 (NR)	5: 3; all other studies: 2	5: not specified; 37, 41, 44, 45,: symptomatic; 40,42,46,47: confirmed; 43: symptomatic and confirmed	Body weight
3.3 Anchor: Sulfonylureas (C)										
C-E (52-55)	See 1.3									
C-G (11,58,59)	11, 58: Glipizide (2,5-10mg); 59: Gliclazide (80-320mg)	12-104	56.7 (10.7)	56.1	3.2 (NR)	8.2 (1)	84.3 (NR)	2	11: not specified; 58: symptomatic; 59: confirmed	HbA1c
3.4 Anchor: Alpha-glucosidase Inhibitors (F)										
E-F (61)	Acarbose (up to 300mg)	12	54.8	62.9	2.6	7.8	70.3	3	Not specified	NA
F-G (62)	Acarbose (up to 300mg)	24	51.8	61.1	1.2	8.6	72	2	Confirmed	NA

For coding of covariates, see Table 12.

[a] Data was aggregated for all studies that provided the required data.

[b] Reported for outcomes that display a considerable heterogeneity (I^2>50%) in pairwise meta-analysis.

Wk, week. SD, standard deviation. Y, year. T2DM, diabetes mellitus type 2. Bl, baseline. HbA1c, glycated haemoglobin. Kg, kilogram. NR, not reported. NA, not applicable.

Meglitinides vs. Pioglitazone

Indirect comparisons were pooled over the bridging nodes placebo, Metformin and sulfonylureas.

HbA1c

In A-D (vs. A-E) and D-C (vs. E-C) studies patients had higher HbA1c baseline levels which may bias indirect estimates in favour of Pioglitazone. On the other hand, patients in A-E trials were therapy naïve whereas patients in A-D studies had OAD discontinued only eight weeks or shorter before randomization which could benefit the meglitinides in the indirect comparison. Heterogeneity was present in comparisons of A-D and C-E studies which compromises the precision of the indirect estimate and might violate the transitivity relation.

Body weight

Baseline body weight was higher regarding all bridging comparators in trials where patients received Pioglitazone; due to co-administered diet and physical exercise in most trials the potential for weight loss might be higher in patients on Pioglitazone; thus the indirect comparison possibly underestimated the weight loss of meglitinides vs. Pioglitazone; moreover heterogeneity in B-E studies might serve as further factor affecting the validity of the indirect estimate.

Hypoglycemia

The definition of hypoglycemic events was especially inconsistent in the set of trials via the bridging node placebo and sulfonylureas. Moreover, heterogeneity was present in the pairwise comparison of C-D studies. However, it is difficult to say in which direction potential violations of the transitivity assumption might have biased the indirect estimate.

Alpha-glucosidase inhibitors vs. Pioglitazone

Indirect estimates were derived via placebo, Metformin, sulfonylureas and gliptins.

HbA1c

The indirect estimate might overestimate the increase in HbA1c levels of alpha-glucosidase inhibitors vs. Pioglitazone since baseline HbA1c levels were unevenly distributed in favour of Pioglitazone in the indirect comparisons via placebo and sulfonylureas. Present heterogeneity in the pairwise comparisons of A-D, A-F and B-F might interfere with transitive effects.

Body weight

The indirect evidence might be biased downwards; since patients on Pioglitazone had a higher baseline body weight compared to patients on alpha-glucosidase inhibitors especially in the set of trials via placebo and gliptins as linkage weight increase of Pioglitazone against alpha-glucosidase inhibitors might even be larger if baselines would have been balanced. Heterogeneity in contrasts of C-F and D-G might act as another factor influencing the indirect estimate.

Hypoglycemia

Definitions of hypoglycemia were not overly deviant within the indirect comparisons. Heterogeneity was present in the pairwise comparison of C-D.

Gliptins vs. Meglitinides

Indirect evidence was synthesized by means of placebo, Metformin, sulfonylureas and alpha-glucosidase inhibitors serving as linking treatments.

HbA1c

Unbalanced HbA1c levels in favour of gliptins in the indirect comparisons via placebo, sulfonylureas and alpha-glucosidase inhibitors might underestimate the weaker glycemic control of gliptins vs. meglitinides. Heterogeneity in pairwise comparisons of A-G, C-E and C-G might impact on the transitivity hypothesis.

Body weight

The indirect estimate might potentially exaggerate the weight loosing effect of gliptins against meglitinides since patients on gliptins had a higher baseline body weight compared to patients on meglitinides in trials with the bridging comparator Metformin and sulfonylureas, respectively. Apparent heterogeneity in the summary estimates of B-E and B-G might have influenced transitivity as well.

Hypoglycemia

Hypoglycemic events were inconsistently defined over the whole set of trials involved in indirect comparisons potentially impacting transitivity.

In conclusion, some differences in potential effect modifiers across trials included in the estimate of indirect evidence could be identified which may have violated the transitivity assumption in some circumstances. However, as mentioned before from a clinical perspective (see 4.2.2.1) most of the indirect estimates were plausible so that potential bias might occur in the magnitude of relative treatment effects but is unlikely to affect the direction of results. The validation against external evidence will deliver further insights if indirect estimates may be valid.

4.2.2.3 Meta-regression of potential effect modifiers

Meta-regression MTC models were only run for the models that fitted the data better (HbA1c/body weight: RE, hypoglycemia: FE). The number of observations (62 studies) appeared large enough to include the pre-specified covariates in the full models in order to detect significant effects of the regression coefficients (the rule of thumb of having at least 10 observations for each included covariate was achieved). For studies that didn´t report the baseline body weight the mean of the remaining studies was imputed ($\overline{m} = 82.64$ kg).

The regression coefficients of the pre-specified covariates for the full MTC models for each outcome are reported in Table 19.

Table 19: Regression coefficients for meta-regression MTC models

1. HbA1c (covariate)	**Mean [$CrI_{0.95}$]**
$ß_0$ (baseline HbA1c)[a]	-0.19 [-0.32;-0.06]
$ß_1$ (length of follow-up)[a]	-0.008 [-0.004;0.006]
$ß_2$ (previous therapy)[b]	0.22 [-0.06;0.51]
$ß_3$ (previous therapy)[b]	0.21 [-0.1;0.52]
2. Body weight	
$ß_0$ (baseline body weight)[a]	0.007 [-0.007;0.02]
$ß_1$ (length of follow-up)[a]	0.003 [-0.003;0.009]
$ß_2$ (previous therapy)[b]	0.1423 [-0.15;0.44]
$ß_3$ (previous therapy)[b]	0.1 [-0.21;0.42]
3. Hypoglycemia	
$ß_0$ (year of study)[a]	-0.02 [-0.1;0.05]
$ß_1$ (length of follow-up)[a]	-0.03 [-0.07;-0.005]

[a] For continuous covariates ß indicates how the outcome variable changes with a unit increase in the explanatory variable.
[b] For categorical covariates ß indicates how the outcome variable differs from a reference group, respectively. Reference category: therapy naïve; $ß_2$: no OAD at least 8 weeks before randomization; $ß_3$: else.
HbA1c, glycated haemoglobin. MTC, mixed treatment comparison. CrI, credible interval.

For HbA1c, the regression coefficient for the covariate baseline HbA1c yielded a significant result denoting that for each increase of the baseline HbA1c by 1% the difference in means of treatments relative to placebo decreases by -0.19%. The finding sounds reasonable since it is known that the glucose lowering potential of OAD increases the higher the baseline HbA1c is. Moreover, this relationship was confirmed, if the crude differences in means of all direct comparisons were plotted against baseline levels of HbA1c and the assumption of a common regression coefficient applying to relative effects of all the treatments relative to the overall baseline placebo seemed to fit the data well (see Figure 11, appendix).

For the categorical variable previous therapy both regression coefficients indicated that compared to the reference category „therapy naive“, both categories ($ß_2$, no OAD at least 8 weeks before randomization; $ß_3$, else) led to an increase of HbA1c of approximately 0.2%. This implicates that the treatments relative to placebo are less efficacious if patients that got some kind of pre-therapy (including wash-out or continuous therapy until randomization) compared to therapy naive patients, which seems reasonable since patients without any kind of therapy exhibit higher baseline HbA1c levels and are more responsive for drug therapy. However, the $CrI_{0.95}$ of the regression coefficients for previous therapy still included the null effect so that the observed trend for previous therapy did not reach statistical significance and for the adjusted model with both covariates baseline HbA1c and previous therapy heterogeneity did not further decrease; thus the covariate previous therapy was dropped.

For body weight, all regression coefficients of the covariates were not statistically significant so that no treatment by covariate interaction could be identified.

For hypoglycemia the regression coefficient for the covariate length of follow-up was statistically significant; the value suggests that for each week of follow-up the log odds ratios of treatments relative to placebo decrease by -0.02 which seems somehow spurious and is in contrast to the hypothesis that a longer drug exposure enforces the harmful effect. A plot of the crude log odds ratios against the length of follow-up confirms that this treatment interaction might have occurred just by chance since the prediction of odds ratios smaller than one suggesting a protective effect of sulfonylureas and meglitinides vs. placebo (if follow-up is approximately longer than 80 weeks) is not plausible. This suggests that the meta-regression model is not appropriate for explaining heterogeneity and an adjustment for the

covariate length of follow is meaningless; thus no covariate was maintained for the outcome hypoglycemia.

The adjusted relative treatment effects (reported at the mean value of the covariate of 8.21%) for baseline HbA1c of the RE MTC model are depicted in Table 20.

Table 20: RE MTC meta-regression model for HbA1c adjusted for baseline HbA1c

1. HbA1c, %			
Comparison	**MD [$CrI_{0.95}$]**		
A-B	-1.07 [-1.21;-0.93]		
A-C	-1.01 [-1.16;-0.85]		
A-D	-1.01 [-1.21;-0.81]		
A-E	-1.03 [-1.26;-0.81]		
A-F	-0.69 [-0.92;-0.46]		
A-G	-0.78 [-0.89;-0.67]		
B-C	0.07 [-0.09;0.23]		
B-D	0.06 [-0.15;0.28]		
B-E	0.04 [-0.18;0.26]		
B-F	0.39 [0.14;0.62]		
B-G	0.3 [0.17;0.43]		
C-D	-0.01 [-0.21;0.2]		
C-E	-0.02 [-0.23;0.18]		
C-F	0.32 [0.08;0.56]		
C-G	0.23 [0.08;0.39]		
D-E	-0.02 [-0.29;0.25]		
D-F	0.33 [0.03;0.61]		
D-G	0.24 [0.03;0.45]		
E-F	0.34 [0.07;0.61]		
E-G	0.25 [0.03;0.48]		
F-G	-0.09 [-0.32;0.15]		
SD	0.22 [0.17;0.29]		
$ß_0$ (baseline HbA1c)	-0.18 [-0.3;-0.07]		
Model fit	$\overline{D}_{res}$	**p_D**	**DIC**
	129.62	107.35	236.97

Relative treatment effects for HbA1c are reported at the mean covariate value (8.21%).
MTC, mixed treatment comparison. HbA1c, glycated haemoglobin. RE, random effects. MD, difference in means. CrI, credible interval. SD, standard deviation. $\overline{D}_{res}$, total residual deviance. p_D, effective number of parameters. DIC, deviance information criterion.

Contrasted to the unadjusted model the DIC and SD decreased slightly arguing that adjustment for baseline HbA1c might explain heterogeneity to some content and improves model fit. However, even if baseline HbA1c is considered as covariate there are only minor, clinically irrelevant differences between the adjusted and the crude relative treatment effects (see Table 16).

In summary, only one pre-specified covariate (baseline HbA1c) could be identified as treatment modifier and the amount of heterogeneity explained by it was moderate. Thus, it remains a certain degree of residual heterogeneity which might be caused by other (unobserved) covariates. Moreover, since meta-regression is observational the observed association might be confounded by another covariate so that a causal relationship can´t be concluded. Another limitation is that without individual patient data it can´t be excluded that the finding is susceptible to ecological bias; if the interaction of baseline HbA1c and treatment effect is not present within each trial than the treatment-by-covariate interaction across trials is confounded and adjusted results might be misleading.

4.2.2.4 Sensitivity analyses

Overall baseline

For each treatment a FE and RE Bayesian meta-analysis was carried out for the arms containing the corresponding baseline and the better fitting model was selected based upon the DIC (data not shown, see electronic data carrier: Ancillary output\Sensitivity analyses\Different overall baselines); for all baselines, the RE models fitted better. The mean and the between-study standard deviation were used to sample predictive distributions for the estimation of the absolute efficacy and the ranking of treatments for each baseline (see Table 21).

Table 21: Sensitivity analysis for HbA1c: absolute efficacy and ranking of treatments with different overall baselines

	Overall baseline													
	A (0.189;13.53)[a]		B (-0.905;3.809)[a]		C (-0.782;1.427)[a]		D (-0.939; 2.702)[a]		E (-0.27; 3.871)[a]		F (-1.021;3.265)[a]		G (-0.708;7.749)[a]	
Parameter	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
T [A]	0.189	0.273	0.132	0.517	0.214	0.844	0.035	0.624	0.767	0.534	-0.372	0.57	0.04	0.363
T [B]	-0.885	0.282	-0.911	0.512	-0.83	0.845	-1.01	0.624	-0.289	0.532	-1.426	0.571	-1.025	0.365
T [C]	-0.816	0.283	-0.854	0.519	-0.788	0.84	-0.953	0.622	-0.228	0.531	-1.366	0.571	-0.963	0.368
T [D]	-0.824	0.291	-0.93	0.524	-0.876	0.846	-0.95	0.607	-0.298	0.539	-1.437	0.578	-1.034	0.373
T [E]	-0.841	0.296	-0.854	0.525	-0.744	0.848	-0.951	0.631	-0.267	0.514	-1.36	0.575	-0.96	0.377
T [F]	-0.499	0.298	-0.515	0.525	-0.427	0.85	-0.606	0.632	0.117	0.538	-1.021	0.556	-0.617	0.377
T [G]	-0.587	0.279	-0.603	0.516	-0.518	0.844	-0.693	0.623	0.032	0.534	-1.107	0.571	-0.704	0.357
Best [A]	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Best [B]	0.457	0.498	0.312	0.463	0.257	0.437	0.427	0.495	0.263	0.44	0.335	0.472	0.341	0.474
Best [C]	0.071	0.256	0.054	0.226	0.075	0.263	0.1	0.3	0.054	0.225	0.0616	0.24	0.058	0.233
Best [D]	0.205	0.404	0.464	0.499	0.567	0.496	0.267	0.443	0.376	0.484	0.441	0.497	0.433	0.495
Best [E]	0.267	0.4424	0.171	0.377	0.101	0.302	0.205	0.404	0.307	0.461	0.163	0.37	0.168	0.374
Best [F]	2.897^{e-4}	0.0169	7.143^{e-5}	0.008	1.437^{e-4}	0.012	3.571^{e-4}	0.019	2.143^{e-4}	0.015	1.429^{e-4}	0.012	2.143^{e-4}	0.015
Best [G]	0	0	0	0	0	0	0	0	0	0	0	0	0	0

The random effects MTC meta-regression model with adjustment for baseline Hba1c was used for sensitivity analysis.

[a] Values in brackets (mean, tau) are the RE estimates from the Bayesian meta-analyses for all arms of the corresponding baseline that were used to draw predictive distributions.

T [i] is the absolute efficacy of the corresponding treatment (mean change in HbA1c, %).

Best [i] is the probability that the i-th treatment is the best averaged over all iterations.

HbA1c, glycated haemoglobin. RE, random effects. MTC, mixed treatment comparison. SD, standard deviation.

It can be noticed that even if the sequence is similar the ranking of the best treatments doesn´t automatically parallel the ranking of the absolute efficacy since T [i] is the mean absolute efficacy of each treatment whereas the probability of being the best treatment is assigned for each iteration; for example for placebo as overall baseline the mean change of HbA1c for gliptins (-0.587) is superior to alpha-glucosidase inhibitors (-0.499); nevertheless, gliptins were never the best treatment compared to alpha-glucosidase inhibitors ($p_{best}=2.8^{e-4}$), even though the probability was very low.
The absolute efficacy is depending on the overall baseline. For example if alpha-glucosidase inhibitors as baseline were compared to placebo as baseline (apparently the single arms of alpha-glucosidase inhibitors exhibited a high glucose lowering potential) the mean change for placebo was -0.372 and 0.189, respectively; however the relative treatment effects were not affected therefrom. Placebo, Metformin and sufonylureas are fully connected to all other treatments and fulfill the assumption of a spanning tree; however placebo appears most appropriate, since estimates for absolute efficacy were more precise for placebo as more studies were involved in the calculation of the baseline. The remaining treatments were less appropriate since the baseline estimates were estimated with a smaller precision, too.
Different baselines also influenced the ranking. Although Metformin, Pioglitazone and meglitinides always were between the first three treatments of the ranking for the best treatment the rank and probabilities differed considerably.

Summed up the absolute efficacy and ranking of treatments were susceptible regarding different overall baselines in terms of the sequence of treatments and the magnitude and precision of effects. Therefore, it is advisable that the spanning tree assumption is fulfilled and if this holds for several options the alternative that is supported by the largest evidence base should be selected to provide reliable estimates.

Risk of bias

The network diagrams for the set of studies for the corresponding outcomes when potentially biased studies (see 3.1.4.4) were excluded are depicted in Figure 12 of the appendix.

For all networks the contrast E-F wasn´t supported anymore by direct evidence and was only estimated via indirect comparisons.
The removal of studies for the particular endpoints bearing a high risk of bias improved model fit for HbA1c as the between-study heterogeneity decreased from 0.22 (see Table 20) to 0.16 (see Table 22); the DIC couldn´t be used for model comparison since the number of studies differed.

Table 22: Sensitivity analyses for risk of bias, removal of outliers and different priors for the between-study standard deviation

	Sensitivity analyses				
	Risk of bias[a]	**Removal of outliers[b]**	**Prior between- study standard deviation[c]**		
			unif (0,5)	**unif (0,2)**	**gamma (0.001,0.001)**
1. HbA1c , % (RE)	**MD [$CrI_{0.95}$]**				
Comparison					
A-B	-1.1 [-1.23;-0.97]	-1.02 [-1.13;-0.91]	-1.02 [-1.13;-0.91]	-1.02 [-1.13;-0.91]	-1.02 [-1.14;-0.91]
A-C	-1.04 [-1.19;-0.88]	-1.01 [-1.15;-0.87]	-1.01 [-1.15;-0.87]	-1.01 [-1.15;-0.87]	-1.01 [-1.14;-0.88]
A-D	-1.11 [-1.28;-0.93]	-1.08 [-1.24;-0.93]	-1.08 [-1.24;-0.93]	-1.08 [-1.24;-0.93]	-1.08 [-1.24;-0.93]
A-E	-0.99 [-1.21;-0.77]	-0.96 [-1.15;-0.78]	-0.96 [-1.15;-0.78]	-0.96 [-1.15;-0.78]	-0.97 [-1.15;-0.78]
A-F	-0.52 [-0.76;-0.28]	-0.49 [-0.71;-0.28]	-0.49 [-0.71;-0.28]	-0.49 [-0.71;-0.28]	-0.49 [-0.7;-0.28]
A-G	-0.78 [-0.87;-0.69]	-0.76 [-0.83;-0.68]	-0.76 [-0.83;-0.68]	-0.76 [-0.83;-0.68]	-0.76 [-0.83;-0.68]
B-C	0.05 [-0.11;0.22]	0.01 [-0.13;0.15]	0.01 [-0.13;0.15]	0.01 [-0.13;0.15]	0.01 [-0.12;0.15]
B-D	-0.01 [-0.2;0.17]	-0.06 [-0.22;0.1]	-0.06 [-0.22;0.1]	-0.06 [-0.22;0.1]	-0.06 [-0.22;0.1]
B-E	0.1 [-0.11;0.32]	0.06 [-0.13;0.25]	0.06 [-0.13;0.25]	0.06 [-0.13;0.25]	0.06 [-0.12;0.24]
B-F	0.57 [0.33;0.82]	0.53 [0.31;0.75]	0.53 [0.31;0.75]	0.53 [0.31;0.75]	0.53 [0.32;0.75]
B-G	0.31 [0.21;0.42]	0.27 [0.18;0.36]	0.27 [0.18;0.36]	0.27 [0.18;0.36]	0.27 [0.18;0.36]
C-D	-0.07 [-0.26;0.12]	-0.07 [-0.23;0.09]	-0.07 [-0.23;0.09]	-0.07 [-0.23;0.09]	-0.07 [-0.23;0.09]
C-E	0.05 [-0.14;0.25]	0.05 [-0.12;0.21]	0.05 [-0.12;0.21]	0.05 [-0.12;0.21]	0.04 [-0.11;0.2]
C-F	0.52 [0.25;0.78]	0.52 [0.28;0.75]	0.52 [0.28;0.75]	0.52 [0.28;0.75]	0.52 [0.29;0.75]
C-G	0.26 [0.11;0.41]	0.25 [0.12;0.39]	0.25 [0.12;0.39]	0.25 [0.12;0.39]	0.25 [0.13;0.38]
D-E	0.12 [-0.13;0.37]	0.12 [-0.1;0.33]	0.12 [-0.1;0.33]	0.12 [-0.1;0.33]	0.12 [-0.09;0.33]
D-F	0.58 [0.3;0.87]	0.59 [0.34;0.84]	0.59 [0.34;0.84]	0.59 [0.34;0.84]	0.59 [0.35;0.84]
D-G	0.33 [0.15;0.5]	0.33 [0.17;0.48]	0.33 [0.17;0.48]	0.33 [0.17;0.48]	0.33 [0.18;0.48]
E-F	0.47 [0.16;0.78]	0.47 [0.2;0.73]	0.47 [0.2;0.73]	0.47 [0.2;0.73]	0.47 [0.21;0.73]
E-G	0.21 [-0.01;0.43]	0.21 [0.02;0.39]	0.21 [0.02;0.39]	0.21 [0.02;0.39]	0.21 [0.03;0.39]
F-G	-0.26 [-0.5;-0.02]	-0.26 [-0.47;-0.05]	-0.26 [-0.47;-0.05]	-0.26 [-0.47;-0.05]	-0.26 [-0.47;-0.06]
SD	0.16 [0.11;0.23]	0.12 [0.07;0.18]	0.12 [0.07;0.18]	0.12 [0.07;0.18]	0.12 [0.06;0.17]

2. Body weight , kg (RE)	**MD [$CrI_{0.95}$]**				
Comparison					
A-B	-0.64 [-0.99;0.29]	-0.53 [-0.81;-0.25]	-0.53 [-0.81;-0.24]	-0.53 [-0.81;-0.24]	-0.52 [-0.8;-0.25]
A-C	1.42 [1;1.86]	1.17 [0.79;1.54]	1.17 [0.79;1.54]	1.17 [0.79;1.54]	1.18 [0.81;1.53]
A-D	2.53 [2.02;3.05]	2.47 [2.02;2.91]	2.47 [2.02;2.91]	2.47 [2.02;2.91]	2.48 [2.05;2.9]
A-E	1 [0.45;1.57]	0.66 [0.19;1.12]	0.66 [0.19;1.12]	0.66 [0.19;1.12]	0.66 [0.22;1.1]
A-F	-0.91 [-1.57;-0.26]	-0.73 [-1.37;-0.09]	-0.73 [-1.37;-0.09]	-0.73 [-1.37;-0.09]	-0.73 [-1.31;-0.12]
A-G	0.58 [0.32;0.82]	0.57 [0.37;0.77]	0.57 [0.38;0.77]	0.57 [0.38;0.77]	0.57 [0.39;0.76]
B-C	2.06 [1.59;2.54]	1.7 [1.29;2.11]	1.7 [1.29;2.1]	1.7 [1.29;2.1]	1.7 [1.32;2.09]
B-D	3.17 [2.63;3.74]	3 [2.52;3.47]	2.99 [2.53;3.46]	2.99 [2.53;3.46]	3 [2.54;3.45]
B-E	1.64 [1.08;2.23]	1.18 [0.7;1.68]	1.18 [0.69;1.68]	1.18 [0.69;1.68]	1.18 [0.71;1.65]
B-F	-0.28 [-0.96;0.41]	-0.2 [-0.85;0.45]	-0.2 [-0.84;0.45]	-0.2 [-0.84;0.45]	-0.2 [-0.8;0.42]
B-G	1.21 [0.9;1.53]	1.1 [0.85;1.35]	1.1 [0.85;1.35]	1.1 [0.85;1.35]	1.09 [0.85;1.33]
C-D	1.11 [0.56;1.66]	1.3 [0.82;1.79]	1.3 [0.81;1.78]	1.3 [0.81;1.78]	1.3 [0.83;1.77]
C-E	-0.42 [-0.93;0.11]	-0.51 [-0.93;-0.08]	-0.51 [-0.94;-0.08]	-0.51 [-0.94;-0.08]	-0.52 [-0.92;-0.11]
C-F	-2.33 [-3.08;-1.6]	-1.9 [-2.61;-1.19]	-1.9 [-2.61;-1.17]	-1.9 [-2.61;-1.17]	-1.91 [-2.57;-1.21]
C-G	-0.85 [-1.29;-0.41]	-0.6 [-0.98;-0.21]	-0.6 [-0.98;-0.21]	-0.6 [-0.98;-0.21]	-0.61 [-0.97;-0.24]
D-E	-1.53 [-2.21;-0.84]	-1.81 [-2.41;-1.21]	-1.81 [-2.4;-1.22]	-1.81 [-2.4;-1.22]	-1.82 [-2.37;-1.2]
D-F	-3.44 [-4.24;-2.65]	-3.2 [-3.96;-2.45]	-3.2 [-3.95;-2.45]	-3.2 [-3.95;-2.45]	-3.2 [-3.91;-2.48]
D-G	-1.96 [-2.49;-1.44]	-1.9 [-2.34;-1.44]	-1.9 [-2.33;-1.45]	-1.9 [-2.33;-1.45]	-1.91 [-2.33;-1.48]
E-F	-1.91 [-2.76;-1.1]	-1.39 [-2.15;-0.62]	-1.39 [-2.14;-0.6]	-1.39 [-2.14;-0.6]	-1.38 [-2.12;-0.65]
E-G	-0.42 [-1.01;0.14]	-0.09 [-0.56;0.39]	-0.09 [-0.56;0.39]	-0.09 [-0.56;0.39]	-0.09 [-0.54;0.36]
F-G	1.49 [0.85;2.14]	1.3 [0.69;1.91]	1.3 [0.68;1.91]	1.3 [0.68;1.91]	1.3 [0.71;1.87]
SD	0.43 [0.24;0.65]	0.28 [0.1;0.47]	0.28 [0.09;0.47]	0.28 [0.09;0.47]	0.24 [0.05;0.43]

3. Hypoglycemia (FE)	**OR [$CrI_{0.95}$]**	
Comparison		
A-B	2.27 [1.41;3.54]	
A-C	6.07 [3.73;9.48]	
A-D	1.34 [0.73;2.26]	
A-E	5.67 [3.18;9.5]	
A-F	1.93 [0.86;3.75]	
A-G	1.14 [0.74;1.72]	
B-C	2.71 [1.85;3.87]	
B-D	0.6[0.35;0.94]	
B-E	2.53 [1.6;3.88]	
B-F	0.89 [0.4;1.63]	
B-G	0.51 [0.34;0.74]	
C-D	0.22 [0.15;0.31]	
C-E	0.94 [0.67;1.28]	
C-F	0.36 [0.15;0.62]	
C-G	0.19 [0.13;0.28]	
D-E	4.37 [2.64;6.93]	
D-F	1.52 [0.63;3.02]	
D-G	0.9 [0.53;1.45]	
E-F	0.36 [0.15;0.7]	
E-G	0.21 [0.13;0.33]	
F-G	0.67 [0.3;1.31]	

For model fit, no DIC was calculated since the sensitivity analyses included different numbers of studies.

[a] For HbA1c, relative treatment effects are reported for the unadjusted model without baseline HbA1c as covariate, since the regression coefficient was no longer statistically significant after studies at high risk of bias were removed.

[b] According to the leverage vs. residual plots the following studies were identified as outliers and excluded from further analyses. HbA1c: study No. 2; Body weight: study No. 10, 38 and 47; for hypoglycemia, no outliers were detected.

[c] For sensitivity analyses two uniform priors were placed on the between-study deviation sigma and one gamma prior on the precision (1/between-study variance).

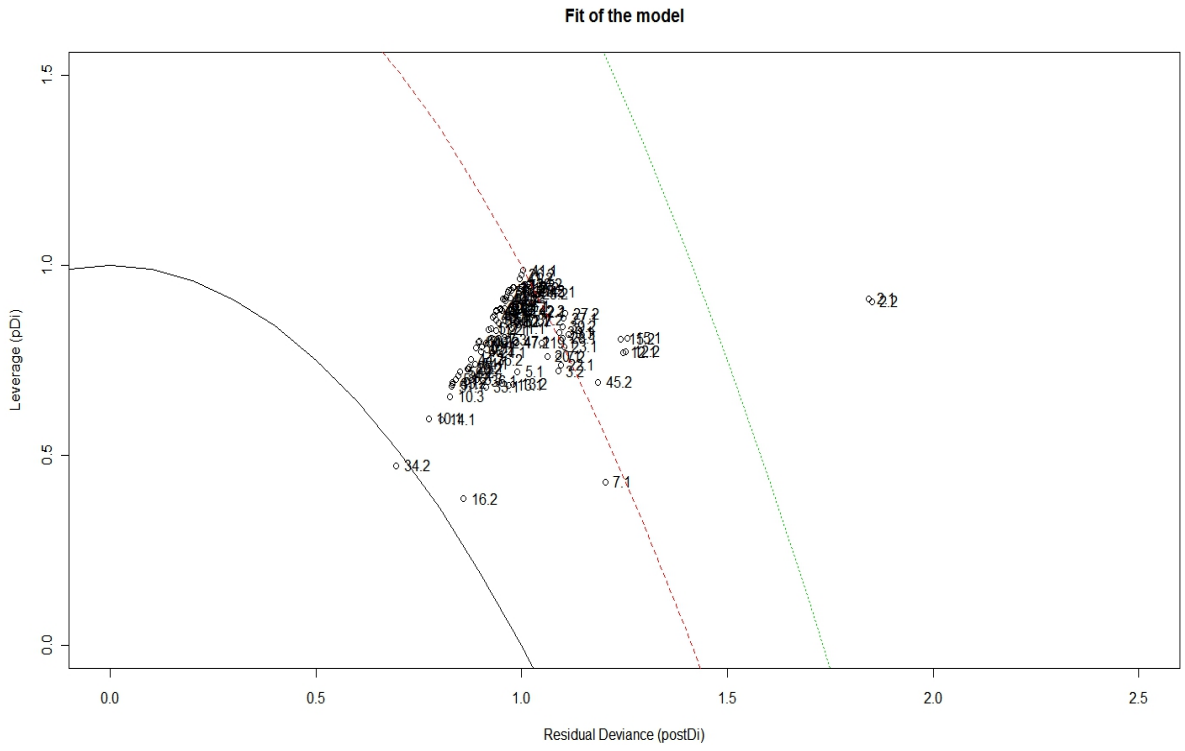
HbA1c, glycated haemoglobin. RE, random effects. MD, difference in means. CrI, credible interval. SD, standard deviation. Kg, kilogram. FE, fixed effects, OR, odds ratio.

Interestingly, the covariate for baseline HbA1c lost significance (see electronic data carrier: Ancillary output\Sensitivity analyses\Risk of bias) indicating that the potentially biased and removed studies might have caused this heterogeneity; thus the covariate was dropped. For body weight, the heterogeneity in the network decreased likewise from 0.53 (see Table 16) to 0.43 when the possibly biased studies were excluded. For the FE model for hypoglycemia, model fit could not be assessed by heterogeneity. Thus the node split on the statistically inconsistent node B-C (see Table 17) was repeated to assess if inconsistency resolved, which was not the case ($\hat{\omega}_{XY}$: -1.04; p=0.004). Moreover, relative treatment effects were comparable to the core model (see Table 16), which yielded more precise estimates and hence was used for further analyses.

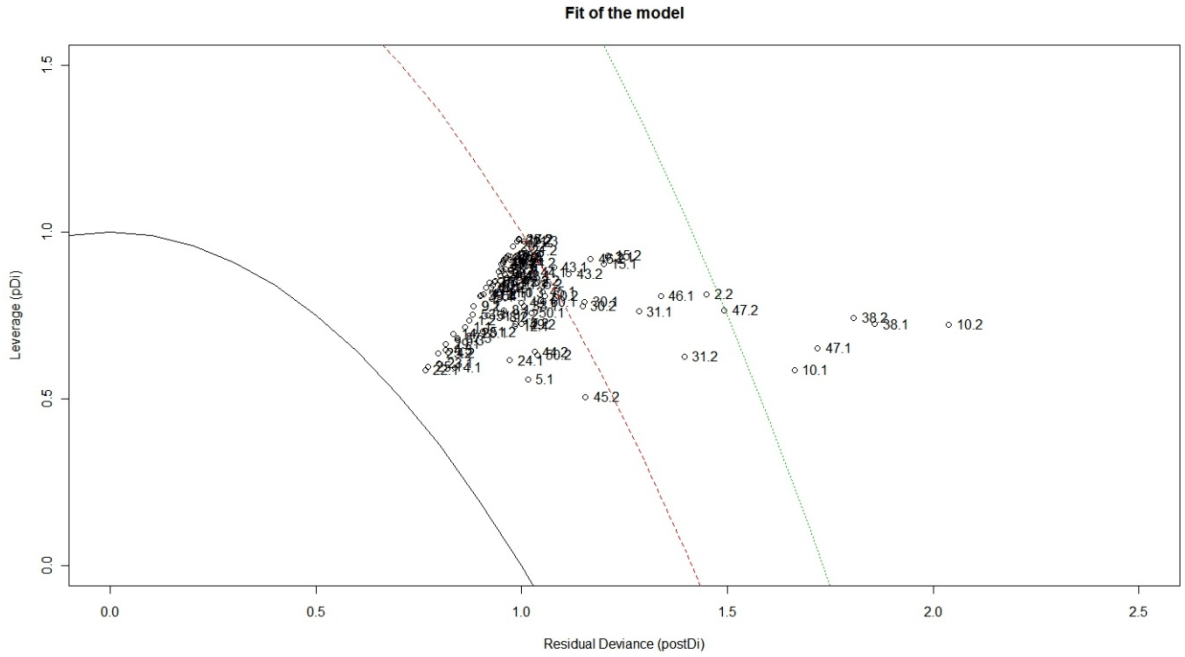
Removal of outliers

Outliers were detected from leverage vs. residual plots for HbA1c (study No. 2) and Body weight (study No. 10, 38 and 47) (see Figure 13).

a.)



b.)



c.)

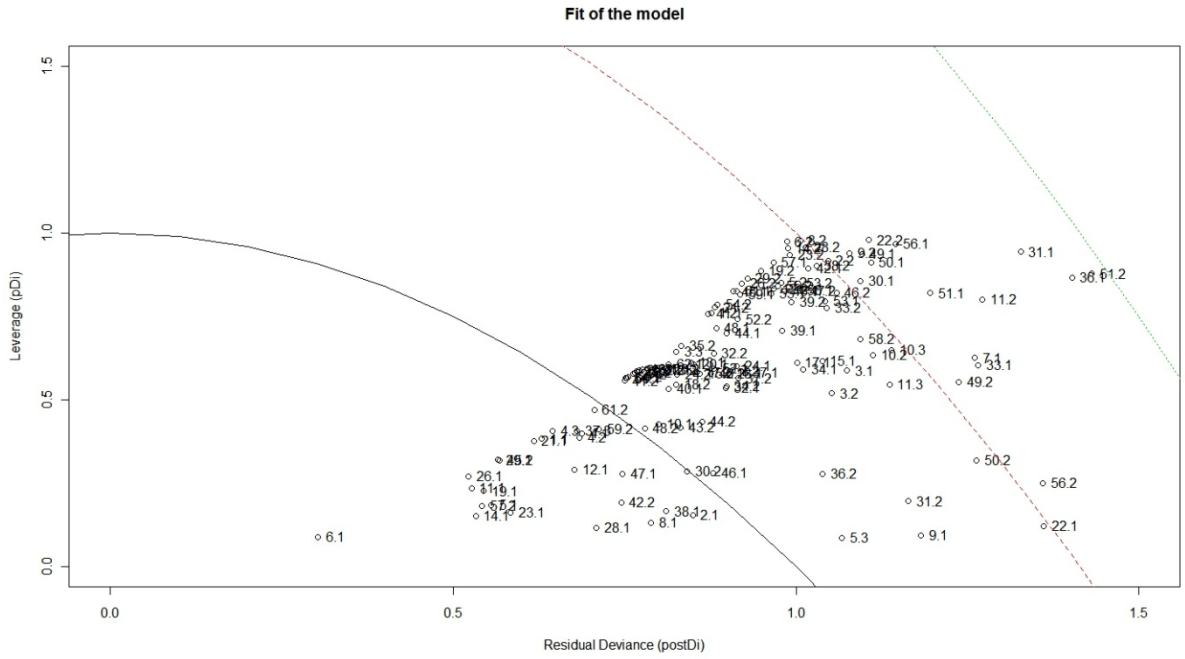


Figure 13: Leverage vs. residual plots
For a.) HbA1c, b.) body weight and c.) hypoglycemia.
Points on the black, red and green lines each contribute an amount of c=1, 2, 3, respectively to the DIC. Points with c>3 beyond the green line were classified as outliers and excluded from sensitivity analysis.

The exclusion of outliers further reduced heterogeneity in the network improving model fit (see Table 22.) Moreover, for body weight one of the statistically inconsistent nodes (1-5; see Table 17) yielded now no more disagreement between direct and indirect evidence ($\hat{\omega}_{XY}$ = -0.83; p=0.08) reasoning that the outliers introduced some inconsistency in the network.

For HbA1c, the MTC estimates from the full model (see Table 16) results were robust to sensitivity analyses. The minor differences in lowering HbA1c levels that affected some comparisons such as A-F ($MD_{core\ model}$:-0.65 vs. MD_{SA}:-0.49), C-F ($MD_{core\ model}$: 0.35 vs. MD_{SA}:0.52) or D-F ($MD_{core\ model}$:0.41 vs. MD_{SA}:0.59) were not clinically relevant.

For body weight, the results from the full MTC core model (see Table 16) were likewise unsusceptible for the different assumptions of sensitivity analyses. Changes in body weight differed slightly only for a few comparisons, e.g. A-C ($MD_{core\ model}$:1.82 vs. MD_{SA}:1.17), B-F ($MD_{core\ model}$:0.27 vs. MD_{SA}:-0.2) or C-D ($MD_{core\ model}$:0.88 vs. MD_{SA}:1.3) but were not clinically meaningful.

Thus, even though sensitivity analyses for HbA1c and body weight were able to reduce the amount of heterogeneity and/or inconsistency to some extent all core analyses yielded comparable results to different assumptions of sensitivity analyses demonstrating that the results were robust to them.

For HbA1c and body weight, the RE models excluding possibly biased studies and outliers were identified as having the best model fit. For hypoglycemia, the FE core model comprising all studies was selected for the remaining analyses.

Different priors for heterogeneity parameter

The different priors for the heterogeneity parameter which were all more informative in terms of limiting the belief in the extent of heterogeneity compared to the core model hardly influenced the estimate of the between-study heterogeneity; only for body weight, the gamma prior on the precision which gives more weight to values of σ near zero slightly decreased the between-trial standard deviation demonstrating that due to huge amount of data the posterior of the between trial standard deviation is robust to different priors (which was not the case for Bayesian pairwise meta-analysis when the number of studies was small, see 4.2.1.1).

4.2.2.5 Ranking of treatments

For the best fitting models (see 4.2.2.4) the ranking of treatments for every outcome was estimated. A new predictive value for the overall baseline placebo was calculated for the reduced study set without outliers and observations being at high risk of bias (for HbA1c and body weight).

For HbA1c Pioglitazone, Metformin and sulfonylureas were superior with a probability of being the best treatment of 0.64, 0.17 and 0.11, respectively (see Table 23).

Table 23: Ranking of treatments and number needed to harm for Hypoglycemia

1. HbA1c								
Ranking / Treatment	1	2	3	4	5	6	7	**Absolute efficacy T[i][a]**
Placebo	0	0	0	0	0	0	1	0.19
Metformin	0.17	0.38	0.29	0.17	0	0	0	-0.84
Sulfonylureas	0.1	0.3	0.43	0.18	2.14^{e-04}	0	0	-0.82
Pioglitazone	0.65	0.19	0.1	0.06	0	0	0	-0.9
Meglitinides	0.08	0.14	0.19	0.58	0.02	5.71^{e-04}	0	-0.78
Alpha-glucosidase inhibitors	0	0	7.14^{e-05}	5.0^{e-04}	0.08^{e-01}	0.99	0	-0.31
Gliptins	0	0	7.14^{e-05}	0.02	0.98	0.08^{e-01}	0	-0.57
2. Body weight								
Ranking / Treatment	1	2	3	4	5	6	7	**Absolute efficacy T[i][a]**
Placebo	0	0.13^{e-01}	0.98	0.03^{e-01}	0	0	0	-0.65
Metformin	0.26	0.74	3.57^{e-04}	0	0	0	0	-1.18
Sulfonylureas	0	0	0	2.86^{e-04}	0.01	0.99	0	0.52
Pioglitazone	0	0	0	0	0	0	1	1.82
Meglitinides	0	1.43^{e-04}	0.04^{e-01}	0.36	0.63	0.01	0	0.05^{e-01}
Alpha-glucosidase inhibitors	0.74	0.25	0.01	5.71^{e-04}	2.14^{e-04}	0	0	-1.38
Gliptins	0	0	1.43^{e-04}	0.64	0.36	0.01^{e-01}	0	-0.08
3. Hypoglycemia								
Ranking / Treatment	1	2	3	4	5	6	7	**Absolute efficacy T[i][a]**
Placebo	0.64	0.24	0.1	0.03	7.14^{e-05}	0	0	0.13^{e-01}
Metformin	0	0	0.01^{e-01}	0.04	0.96	0	0	0.03
Sulfonylureas	0	0	0	0	0	0.45	0.55	0.07
Pioglitazone	0.12	0.21	0.33	0.33	0.07^{e-01}	0	0	0.18^{e-01}
Meglitinides	0	0	0	0	0	0.55	0.45	0.07
Alpha-glucosidase inhibitors	0.15	0.19	0.22	0.41	0.04	0	0	0.18^{e-01}
Gliptins	0.1	0.37	0.35	0.19	2.86^{e-04}	0	0	0.16^{e-01}

NNH	Placebo	Metformin	Sulfonylureas	Pioglitazone	Meglitinides	Alpha-glucosidase inhibitors	Gliptins	
Placebo (A)								
Metformin (B)	61							
Sulfonylureas (C)	17	24						
Pioglitazone (D)	235	82[b]	19[b]					
Meglitinides (E)	18	25	1351[b]	19				
Alpha-glucosidase inhibitors (F)	202	87[b]	19[b]	1449	19[b]			
Gliptins (G)	324	75[b]	18[b]	855[b]	19[b]	538[b]		

For 1.) HbA1c 2.) body weight and 3.) hypoglycemia the estimated probability denotes that each treatment is ranked first (best), second, third,…,seventh (worst) in terms of reducing HbA1c, avoiding an increase in body weight and having the lowest risk for experiencing a hypoglycemic event.
The number needed to harm NNH_{bk} is derived as follows: $NNH_{bk}=1/(prob_k-prob_b)*100$, where $prob_{k/b}$ is the probability of suffering a hypoglycemic event on treatment k and b, respectively. It indicates how many patients need to be treated with treatment k compared to treatment b over a specific period to cause harm in one patient that would not otherwise have been harmed, as long as k is alphabetically after b.
[a] T [i] is the absolute efficacy of the corresponding treatment (mean change in HbA1c (%) and body weight (kg) and proportions of patients experiencing at least one event of any hypoglycemia, respectively).
[b] For these comparisons the NNH_{kb} is inversed denoting how many patients need to be treated with treatment b compared to treatment k over a specific period to cause harm in one patient that would not otherwise have been harmed.
HbA1c, glycated haemoglobin. NNH, number needed to harm.

The second and third rank was dominated by Metformin and sulfonylureas, respectively. For the fourth and fifth rank, meglitinides and gliptins were the prevailing drug class.
Apart from placebo the therapeutic efficacy was worst for alpha-glucosidase inhibitors having a 0.99 probability of being the second worst treatment.

With respect to body weight alpha-glucosidase inhibitors and Metformin were most effective in avoiding an increase of weight with a probability of 0.74 and 0.26, respectively. Whereas it is known that Metformin has a positive impact on weight loss, the apparently high effect of alpha-glucosidase inhibitors might be attributed to gastrointestinal side effects such as diarrhea and flatulence; thus this weight loss comes at the price of a worse tolerability and impaired compliance. In accordance with the MTC estimates gliptins were inferior to placebo and Metformin exhibiting a high probability of ranking fourth or fifth.
The weight gaining effect of sulfonylureas, meglitinides and especially Pioglitazone is reflected by the fact that they performed badly and were ranked among the worst options.

Regarding hypoglycemia placebo, alpha-glucosidase inhibitors and Pioglitazone were the safest drugs inducing the fewest hypoglycemic events with a probability of 0.64, 0.15 and 0.12, respectively. Obviously, therapeutic nihilism (placebo) can´t be really considered as an alternative. Moreover, one has to consider all relevant side effects and not only hypoglycemia. Although alpha-glucosidase inhibitors rank second after placebo with regard to being the safest treatment health care professionals must balance this benefit against other side effects as mentioned before. For Pioglitazone several serious side effects (like weight gain, fluid retention, edema, heart failure, fractures and a possible association with an increased risk for bladder carcinoma; see 1.4) surely outweigh its favourable effect on hypoglycemia. Thus, if these three treatments are ignored, gliptins and Metformin are likely to be the next best alternatives. For dichotomous outcomes the number needed to harm (NNH) is a helpful clinical measure to describe detrimental effects of drugs. It is defined as the inverse of the absolute risk reduction and indicates how many patients need to be exposed to a drug compared to the control over a specific period (which differs due to differences in length of follow-up of the included studies) to cause harm in one patient that would not otherwise have been harmed (the higher, the better the side effect profile of the corresponding drug). Just as the number to treat these measures are able to detect differences in absolute baseline risk whereas relative effect measures ignore them and yield identical results. Table 23 displays all pairwise NNHs; whereas some drugs like sulfonylureas and meglitinides exhibited low NNH´s compared to the other alternatives mirroring their high potential for inducing hypoglycemias, other drugs like gliptins, alpha-glucosidase inhibitors and Pioglitazone yielded relatively high NNH´s reflecting their good tolerability with respect to hypoglycemias.

The ranking of treatments apparently yields very distinct differences; e.g. for HbA1c, the probability for Metformin (p=0.17) being best is considerably lower than for Pioglitazone (p=0.64). However, the results must be scrutinized. On the one hand it is important how treatments rank over all ranks (e.g. what is the probability of ranking second best, third best, etc.) to get the whole picture. Hence, Metformin has the highest probability of ranking second (and a high chance to rank third or fourth) if Pioglitazone might not be chosen due to side effects. In addition, one must consider the relative treatment effects; the difference in means of Pioglitazone vs. Metformin is -0.06 ($CrI_{0.95}$:[-0.15;0.27]) which is clinically meaningless. Thus, in this case the ranking pretends a clear superiority of Pioglitazone which is not supported by the relative treatment effect.

Another example are hypoglycemias; Metformin seems to be inferior to gliptins since gliptins have a probability of ranking second of 0.37 whereas it is zero for Metformin. Although gliptins cause less hypoglycemic events than Metformin (OR:0.55; $CrI_{0.95}$:[0.39;0.76]), the baseline risk for Metformin is still low (the pooled incidence of experiencing such events was 0.03) which is reflected by a comparatively high NNH of 75. Alike, the benefit of gliptins compared to Metformin based only on the ranking for hypoglycemic events diminishes if the NNH is examined.

Thus, for drug choice the ranking alone might be misleading and relative and absolute treatment effects should be considered, too. Moreover, all relevant endpoints (not only the ones reported in this thesis) must of course be considered simultaneously in order to base the decision on a comprehensive benefit-risk assessment.

4.2.2.6 Validation of the mixed treatment comparison models

The contrasts D-E, D-F and E-G plus E-F (only for HbA1c and body weight) were not supported by direct evidence. From the list of excluded studies only the one by Göke et al. (record No. 9361, see Table 11 of the appendix) [143] was identified to serve as external reference and provided results for HbA1c and body weight for model prediction purposes for the contrast D-F (see Table 24).

Table 24: Validation for predictive effects of MTC models

	MC control[a]	**MD_{new}[b] [$CrI_{0.95}$]**	**MC experimental[c]**	**P-value**	**MD [$CrI_{0.95}$][d]**
1. HbA1c (%)	-1.16	0.6 [0.34;0.84]	-0.57	0.7	0.68 [0.28;1.08]
2. Body weight (kg)	1.23	-3.2 [-3.94;-2.44]	-1.96	0.84	-3.32 [-4.39;-2.25]

Predictive effects of the best fitting MTC models for contrasts that were estimated based on indirect evidence only and for which evidence from the excluded studies was available; this applied only for the contrast Pioglitazone vs. alpha-glucosidase inhibitors where one trial (record No. 9361, see Table 11 of the appendix) reported data for HbA1c and body weight for cross-validation [143].

[a] Predictive mean change for HbA1c and body weight for the control arm (Pioglitazone) of the external trial in a new study.

[b] Predictive effect of the MTC models for the difference in means for HbA1c and body weight.

[c] Predicted mean change for HbA1c and body weight for the experimental arm (alpha-glucosidase inhibitors).

[d] $CrI_{0.95}$ for the difference in means of the external trial for HbA1c and body weight.

HbA1c, glycated haemoglobin. MTC, mixed treatment comparison, MC, mean change. MD, difference in means. CrI, credible interval.

For HbA1c (MD_{new}:0.6; $CrI_{0.95new}$:[0.34;0.84] vs. MD_{ext}:0.68; $CrI_{0.95ext}$:[0.28;1.08]) and body weight (MD_{new}:-3.2; $CrI_{0.95new}$:[-3.94;-2.44] vs. MD_{ext}:-3.32; $CrI_{0.95ext}$:[-4.39;-2.25]) the predictive treatment effects were consistent compared to the treatment effect of the external trial. In addition, the Bayesian p-values (HbA1c: p=0.7; body weight: p=0.84) confirmed the good agreement and indicated that it is unlikely that predictive values from the MTC models differ from future trials.

In summary, even if only one trial was available for the validation of the indirect D-F MTC estimates for HbA1c and body weight results were satisfying in assuming that the models accurately predict these relative treatment effects in future studies.

5. Conclusion

In this systematic review the comparative effectiveness and safety of oral antihyperglycemic drugs as monotherapy in patients with type 2 diabetes mellitus in terms of HbA1c, body weight and hypoglycemia was assessed by a mixed treatment comparison analysis.

In summary, Metformin appeared to be the best OAD as first-line monotherapy in T2DM patients; it exhibited comparable glycemic control compared to sulfonylureas and Pioglitazone and induced the strongest weight loss. Even if Metformin might cause more hypoglycemias than gliptins the absolute incidence is low. Moreover, it is the only drug positively affecting micro- and macrovascular complications [35]. Contraindications like severe renal disorders must be taken into account.

If Metformin is not suitable meglitinides and sulfonylureas might be the second best option. The former might be more beneficial since both lower HbA1c and trigger hypoglycemias to a similar extent but the weight gain of meglitinides is less pronounced compared to sulfonylureas. However, their weight gaining effect and their disposition for hypoglycemias makes them a less attractive option for obese patients, elderly patients with comorbidities and patients with irregular eating behavior or increased physical activity.

If meglitinides and sulfonylureas are not an option gliptins might be considered as an alternative. Even though their HbA1c lowering effect is moderate they result in a weight loss compared to sulfonylureas and are weight neutral compared to meglinitides; moreover they cause less frequently hypoglycemias than both classes.

Finally, Pioglitazone and alpha-glucosidase inhibitors seem less appropriate. Although Pioglitazone demonstrated a promising glucose lowering efficacy, its strong weight gain and cardiovascular risk limit its use in diabetics. Alpha-glucosidase inhibitors were worst at lowering HbA1c levels and even though they don´t cause hypoglycemias very often their weight losing effect is offset particularly by the frequent gastrointestinal side effects which affect compliance negatively and can lead to discontinuation of drug therapy [6].

The analysis has some limitations which need to be addressed. One is that hypoglycemias were reported inconsistently and incompletely in the included studies. Consequently, this endpoint couldn´t be analyzed as planned for separate categories reflecting the severity of events; as a trade off any kind of hypoglycemia was used instead making it difficult to draw conclusions about the clinical relevance of this finding. Due to the heterogeneous definitions of events within the network it is unknown if differences were caused by more meaningful events like documented symptomatic hypoglycemias or less important ones such as probable symptomatic or relative hypoglycemia. Thus, it remains unclear if the elevated risk, especially for meglitinides and sulfonylureas is really a drawback or is of minor importance. Therefore it would be desirable that future clinical trials adhere stronger to the recommendations of regular approval authorities and scientific associations to improve reporting on hypoglycemic events [79,82].

A second limitation is that some residual inconsistency persisted in the network even after meta-regression and sensitivity analyses were performed as countermeasures to get rid of it. For HbA1c there were no statistically inconsistent comparisons. Even though the inconsistency estimate was close to statistical significance for some contrasts the differences were not clinically relevant.

For body weight, regarding the comparison of meglitinides vs. placebo statistically inconsistency could be overcome by sensitivity analyses, even though the inconsistency estimate remained considerably large. The conflict between direct and indirect evidence of Metformin against placebo persisted.
For hypoglycemia inconsistency on the node sulfonylureas vs. Metformin couldn´t be remedied and remained unchanged. Moreover, for the contrast meglitinides vs. placebo direct and indirect estimates were obviously not in agreement although this phenomenon couldn´t be confirmed statistically probably due to the underpowered test.

Although for body weight and hypoglycemia the mentioned discrepancies question the pooling of the direct and indirect evidence for all nodes besides the comparison of meglitinides vs. placebo indirect evidence was likely to yield more reliable results and drove clearly the MTC estimates. Thus, inference based on the MTC estimates and the ranking of treatments might be regarded as valid even though (statistically proven) inconsistency on some nodes in the network is evident.
The present inconsistency for the comparison of meglitinides vs. placebo is probably caused by the lumping of the heterogeneous drugs Nateglinide and Repaglinide into one drug class rather than by differences in the validity of direct and indirect evidence. Therefore, it would be interesting to see if inconsistency dissolves if the assumption of modeling a joint drug class of meglitinides is relaxed and both drugs are treated as independent nodes in the network.

Hence, it is very crucial to give considerable consideration prior to the evidence synthesis to the question how the treatment nodes in the network are defined and which steps will be taken to address heterogeneity and inconsistency. One the one hand, if single drugs are similar enough to aggregate them into one drug class this keeps the complexity of the network at a reasonable level and facilitates interpretation of the results. On the other hand, if this assumption might be violated the lumping of treatments may lead to imprecisely defined nodes and as a consequence heterogeneity or inconsistency might be introduced into the network.

Apart from assessing inconsistency between closed loops of evidence it is important that the transitivity assumption is also appraised for contrasts that are not informed by direct head- to-head- trials. Generally this can only be evaluated by exploring potential effect modifiers that might be unevenly distributed between the set of studies involved in the indirect comparisons. In this work some findings indicated that the transitivity relation might not always hold. However, from a clinical view, indirect estimates appeared to be reasonable and the partial validation confirming that MTC models reliably predicted results reinforced confidence in them.

Meta-regression was a helpful approach to detect and reduce heterogeneity and maybe inconsistency to a certain extent. However, the magnitude of effect was moderate and some residual heterogeneity remained although the MTC models already yielded relatively precise results of the between-study heterogeneity. Moreover, it needs to be borne in mind that meta-regression is of observational nature and can introduce ecological bias as long as no individual patient data is available [136].

The sensitivity analysis for the overall baseline revealed that it should be carefully chosen since it affects the ranking of treatments. Thus, the assumption of a spanning tree connecting the reference treatment to any other node in the network should be fulfilled to provide valid results. Although the ranking of treatments is an appealing feature of Bayesian MTC models it must be interpreted with caution. The probability statements seem to provide clear answers which treatment might be the best; however, in some cases this is not reflected by the relative or absolute treatment effects. Thus, the choice of treatment should always be based on a holistic review of all the evidence.

Furthermore, sensitivity analyses demonstrated that the exclusion of possibly biased studies and obvious outliers improved the model fit by decreasing the between-trial heterogeneity and eliminating some inconsistency in the network. However, some of the assessed domains of the risk of bias, especially the sequence generation for randomization and the allocation of concealment were very poorly reported in included studies, even in recent ones and despite well-known recommendations like the CONSORT statement [140] ; thus it couldn´t be resolved if selection bias might be present in the study set and how this might affect the findings.
Although sensitivity analyses yielded some improvements in terms of reducing the amount of heterogeneity and inconsistency in the network, the full MTC models were likely to yield comparable results exhibiting only slight but not clinically relevant differences and hence were deemed to be robust to the different assumptions made in sensitivity analyses.

In the end, the advantages of mixed treatment comparisons became evident in this work. First of all, based upon the consistency assumption indirect evidence could be obtained for comparisons that were not informed by direct evidence. Moreover, by pooling of the direct and indirect evidence MTC models improved the precision of results leading to narrower interval estimates that in some cases without indirect evidence would not have reached statistical significance. In addition, direct evidence appeared to be biased for some contrasts and indirect comparisons provided another piece of evidence which might be considered instead. Finally, interpretation of pairwise meta-analyses can become very confusing and complex if several treatment options are compared against each other; by means of an overall reference treatment the Bayesian MTC analysis allowed to generate rankings of the probability that each treatment is among the n best options which might ease choice of treatment for health care professionals.

If inference based on mixed treatment comparisons is made prudently and its limitations are adequately considered this method will become increasingly important and popular and has the potential to serve as the “next generation evidence synthesis tool” within the framework of comparative effectiveness research, benefit and health economic assessment of pharmaceuticals going beyond the requirements of drug approval [60].

Zusammenfassung

Die vorliegende systematische Übersichtsarbeit befasst sich mit der Untersuchung der Wirksamkeit und Sicherheit oraler Antidiabetika als Monotherapie bei Patienten mit Diabetes Mellitus Typ 2; die Evidenzsynthese der untersuchten Endpunkte glykosyliertes Hämoglobin, Körpergewicht und Hypoglykämien erfolgte im Rahmen einer Netzwerk Meta-Analyse.

Metformin erscheint das geeignetste orale Antidiabetikum im Hinblick auf die Erstlinien-Monotherapie bei Patienten mit T2DM zu sein. Während die antihyperglykämische Wirkung von Metformin vergleichbar mit der der Sulfonylharnstoffe und Pioglitazon ist, induziert es neben alpha- Glukosidase Inhibitoren den stärksten Gewichtsverlust unter allen Netzwerkkomparatoren. Obwohl Metformin mit mehr Hypoglykämien als die Wirkstoffgruppe der Gliptine assoziiert zu sein scheint, ist die absolute Inzidenz für diese Ereignisse gering; die klinische Relevanz dieses Befundes spielt somit eine untergeordnete Rolle. Darüber hinaus ist es das einzige Pharmakon, für das ein positiver Effekt auf die Verhinderung mikro- und makrovaskulärer Komplikationen belegt ist [35]. Bei dem Einsatz von Metformin müssen Kontraindikationen wie z.B. Nierenversagen oder Störung der Nierenfunktion (Kreatinin-Clearance < 60ml/min) berücksichtigt werden.

Wenn Metformin nicht in Frage kommt, bieten sich Glinide und Sulfonylharnstoffe als weitere Alternativen an; dabei erscheinen die Glinide im Vergleich zu den Sulfonylharnstoffen einen größeren klinischen Benefit zu haben, da sie bei vergleichbarer glykämischer Kontrolle und einem ähnlichen Schadenspotential bezüglich hypoglykämischer Episoden eine geringere Gewichtszunahme erzeugen.

Allerdings schränken deren negativen Effekte auf das Körpergewicht sowie die Neigung zu Hypoglykämien den Einsatz insbesondere bei adipösen Patienten, älteren Patienten mit Komorbiditäten sowie Patienten mit unregelmäßigem Essverhalten und starker körperlicher Aktivität ein. Bei diesen Patientengruppen kommen Gliptine als weitere Therapieoption für die orale Monotherapie in Betracht. Obwohl der HbA1c- senkende Effekt eher moderat ausgeprägt ist, bewirken sie gegenüber Sulfonylharnstoffen eine Gewichtsabnahme und sind gewichtsneutral, was den Vergleich mit den Gliniden anbelangt. Zudem sind sie durch ihr vermindertes Risiko für Hypoglykämien beiden Wirkstoffklassen überlegen.

Abschließend scheinen Pioglitazon und alpha- Glukosidase Inhibitoren für die orale Ersttherapie weniger angebracht zu sein. Wenngleich der Insulinsensitizer Pioglitazon einen sehr vielversprechenden antihyperglykämischen Effekt besitzt und in der Netzwerk Meta-Analyse den stärksten HbA1c Abfall demonstrierte, überwiegen vor allem Bedenken im Hinblick auf das kardiovaskuläre Risiko, Gewichtszunahme, eine erhöhte Frakturinzidenz sowie die im Raum stehende Assoziation mit der Entstehung von Blasenkarzinomen, so dass die Nutzen-Schaden Bilanz negativ ausfällt. Unter allen Netzwerkkomparatoren senkten alpha- Glukosidase Inhibitoren den HbA1c am geringsten. Diese Wirkstoffgruppe verursacht zwar selten hypoglykämische Episoden, allerdings scheint der Gewichtsverlust auf gastrointestinale Nebenwirkungen wie Flatulenz oder Meteorismus zurück zu führen zu sein, welche die Compliance negativ beeinflussen und Therapieabbrüche zur Folge haben können [6].

Die Arbeit weist einige Limitationen auf, die nachfolgend diskutiert werden. Ein Aspekt, der die Aussagekraft der Netzwerk Meta-Analyse einschränkt, ist dass der Sicherheitsendpunkt Hypoglykämien inkonsistent und unvollständig in den einzelnen Studien des Studienpools berichtet ist. In der Konsequenz konnte der Endpunkt nicht wie ursprünglich geplant für verschiedene Ereigniskategorien unterschiedlicher Schweregrade analysiert werden und es wurde auf die Kategorie jegliche Art von Hypoglykämien zurückgegriffen, was die Interpretation der klinischen Relevanz erschwert. Aufgrund der äußerst heterogenen Endpunktspezifizierung dieses Outcomes in den einzelnen Studien des Netzwerkes lässt sich nicht abschließend klären, ob Differenzen in den relativen Behandlungseffekten durch klinisch bedeutsame Ereignisse wie z.B. schwere oder dokumentierte symptomatische Hypoglykämien oder weniger relevante Episoden wie z.B. symptomatische oder Pseudohypoglykämien verursacht werden. Demzufolge ist es auch schwierig, das tatsächliche Schadenspotential der Sulfonylharnstoffe und Glinide, die ein erhöhtes Risiko für Hypoglykämien im Vergleich zu den anderen Interventionen besitzen, adäquat zu bewerten. Es wäre somit wünschenswert, wenn zukünftige randomisierte, kontrollierte Interventionsstudien stringenter auf die entsprechenden Empfehlungen der Zulassungsbehörden und Fachgesellschaften bei der Berichterstattung von Hypoglykämien in Publikationen achten [79,82].

Die Diskrepanz zwischen direkter und indirekter Evidenz (Inkonsistenz), die trotz Meta-Regression und Sensitivitätsanalysen für einige Ergebnisschätzer im Netzwerk persistierte, stellt eine weitere Einschränkung dar. Für den Endpunkt HbA1c war keiner der gepoolten Ergebnisschätzer statistisch signifikant inkonsistent; selbst die Inkonsistenzschätzer, deren Bayesianischer p-Wert nahe dem Signifikanzniveau lag, wiesen lediglich klinisch marginale Unterschiede auf, so dass die Konsistenzannahme gerechtfertigt erscheint.
Für den Endpunkt Körpergewicht konnte die primär vorhandene statistisch signifikante Inkonsistenz für den Behandlungsvergleich Glinide vs. Placebo mit Hilfe der durchgeführten Sensitivitätsanalysen eliminiert werden, allerdings ist der numerische Mittelwertsunterschied immer noch beträchtlich, was bei der Interpretation der Ergebnisse berücksichtigt werden sollte. Die Nichtübereinstimmung bezüglich direkter und indirekter Evidenz für den Vergleich Metformin vs. Plazebo blieb bestehen.
Auch für den Endpunkt Hypoglykämien war die Inkonsistenz des Vergleiches Sulfonylharnstoffe vs. Metformin nicht zu beheben. Darüber hinaus erweckt der Vergleich von direkter und indirekter Evidenz für Glinide vs. Plazebo trotz statistischer Unauffälligkeit den Anschein von Inkonsistenz, was vermutlich durch einen unterpowerten Test bedingt ist.

Für die inkonsistenten Netzwerkkontraste muss die Evidenzsynthese aus direkten und indirekten Vergleichen kritisch hinterfragt werden. Aus einer klinischen Perspektive (abgesehen von dem Vergleich Glinide vs. Plazebo) erscheint die indirekte Evidenz reliablere Ergebnisse als die head to head- Studien zu liefern und dominiert zudem die kombinierten Schätzer in ihrer Gewichtung. Insofern erscheint es angemessen, die Inferenz aus der Netzwerk Meta-Analyse trotz vorhandener Inkonsistenzen als valide zu betrachten.
Für den Vergleich Glinide vs. Plazebo ist die Inkonsistenz vermutlich weniger durch eine Diskrepanz von direkter und indirekter Evidenz bedingt als durch die Modellierung einer gemeinsamen Wirkstoffklasse. In diesem Fall mag die Annahme ähnlicher Effekte für die einzelnen Wirkstoffe Nateglinid und Repaglinid unzutreffend sein und die Inkonsistenz ist wohl eher “technischer” Natur, bedingt durch die zu Grunde liegende Netzwerkstruktur. Eventuell lässt sich ein Konsistenzgewinn durch eine Aufspaltung der potentiell heterogenen Wirkstoffklasse Glinide in die einzelnen Wirkstoffe erzielen, was allerdings nicht im Rahmen dieser Arbeit untersucht wurde.

Demnach ist es von großer Relevanz, im Vorfeld der Evidenzsynthese ein klinisch plausibles Netzwerk zu definieren; dabei sollten vor allem die Vorteile (überschaubares Netzwerk und leichtere Vermittelbarkeit der Ergebnisse) und Nachteile (etwaige Verletzung der Konsistenzannahme bei Gruppenaggregation heterogener Wirkstoffe) einer Analyse von Wirkstoffklassen gegeneinander abgewogen werden. Außerdem sollten a priori in einem Studienprotokoll die Maßnahmen festgelegt werden, um mit Heterogenität und Inkonsistenz umzugehen.

Für Vergleiche, die lediglich durch indirekte Evidenz gestützt sind, muss die Transitivitätsannahme ebenso evaluiert werden. In der Regel kann dies nur durch eine klinische Exploration potentieller Effektmodifikatoren, deren Verteilung zwischen den involvierten Studien differiert, untersucht werden. In dieser Arbeit ergaben sich Anhaltspunkte, dass die Annahme einer Transitivität der Effekte nicht für jeden Vergleich gleichermaßen gerechtfertigt sein mag. Allerdings erscheinen die Ergebnisse unter klinischer Betrachtung plausibel und die partielle Kreuzvalidierung mit Studien, die nicht in die Analyse mit einbezogen wurden, mag als Beleg für eine valide Prädiktion dieser Ergebnisschätzer der Netzwerk Meta-Analyse herangezogen werden.

Die durchgeführte Meta-Regression für potentielle Effektmodifikatoren erwies sich als ein hilfreicher Ansatz im Hinblick auf den Endpunkt HbA1c unter Einschluss des baseline HbA1c Wertes als erklärende Variable, um die Netzwerkheterogenität zu reduzieren; eventuell mag die Adjustierung auch dazu beigetragen haben, die Inkonsistenz positiv zu beeinflussen. Allerdings hatte die Adjustierung nur einen geringen Einfluss auf die Effektstärke, die Effektrichtung war davon nicht betroffen. Das Ausmaß der Reduktion des Heterogenitätsparameters war moderat und es verblieb eine residuale Heterogenität, die sich nicht durch die untersuchten Kovariablen erklären ließ. Allerdings war die Heterogenität bereits in den unadjustierten MTC Analysen zuvor relativ gering, so dass kein ausgeprägter Effekt durch die Adjustierung um potentielle Effektmodifikatoren zu erwarten war. Unabhängig davon ist bei der Auslegung der Ergebnisse einer Meta-Regression zu bedenken, dass sie observationeller Natur ist und ein potentieller ökologischer Bias ohne patientenindividuelle Daten nicht ausgeschlossen werden kann [136].

Die Robustheit der Wahl verschiedener Referenzbehandlungen wurde in einer Sensitivitätsanalyse untersucht. Dabei zeigte sich, dass das Ranking der Interventionen in Bezug auf Reihenfolge und Ausmaß der Wahrscheinlichkeit, welche Behandlung die n-beste ist, für verschiedene Annahmen anfällig ist und somit die Wahl der Referenzbehandlung umsichtig getroffen werden sollte; dabei sollte die Referenzbehandlung einen “spanning tree” formen, das heißt, mit allen anderen Netzwerkkomparatoren verbunden sein. Obwohl das Ranking der Behandlungsinterventionen mit Hilfe von Bayesianischen MTC Modellen aufschlussreich sein kann, müssen diese Wahrscheinlichkeitsaussagen mit Bedacht interpretiert werden; für eine umfassende Evidenzbewertung sollten sie daher nicht isoliert betrachtet werden, sondern immer in der Gesamtschau aller relativen und absoluten Behandlungseffekte.

Die Sensitivitätsanalysen, die Studien mit einem hohen Verzerrungspotential auf Endpunkt- oder Studienebene sowie potentielle Ausreißer (Devianzresiduen + Leverage > 3) ausschlossen, führten zu einer Verbesserung der Güte der Modellanpassung, was sich in einer reduzierten Heterogenität und einem Konsistenzgewinn manifestierte. Allerdings sind selbst in aktuellen Publikationen manche Domänen des Verzerrungspotentials, insbesondere die Erzeugung der Randomisierungssequenz und die Verdeckung der Gruppenzuteilung bedauerlicherweise äußerst unzureichend berichtet, obwohl dazu einschlägige Empfehlungen wie z.B. das CONSORT Statement existieren [140]. Somit kann nicht bewertet werden, ob ein Selektionsbias in dem Studienpool vorliegt und welchen Einfluss dies auf die Validität der Ergebnisse hat.

Sämtlichen Sensitivitätsanalysen ist gemein, dass deren Ergebnisschätzer nur marginale, klinisch irrelevante Unterschiede verglichen mit denen der ursprünglichen MTC Modellen aufweisen. Somit kann davon ausgegangen werden, dass die initialen MTC Modelle valide Ergebnisschätzer liefern und robust gegenüber den unterschiedlichen Annahmen in den durchgeführten Sensitivitätsanalysen sind.

Die diversen Vorteile von MTC Modellen gegenüber paarweisen Meta-Analysen einzelner Wirkstoffvergleiche lassen sich auch anhand dieser Arbeit belegen. Zunächst konnten drei der insgesamt 21 relativen Behandlungseffekte, für die keine direkten randomisierten kontrollierten Studien vorliegen basierend auf der Konsistenzannahme aus indirekten Vergleichen über die jeweiligen Brückenkomparatoren geschätzt werden. Zudem lässt sich durch die Synthese aus direkter und indirekter Evidenz ein erheblicher Präzisionsgewinn der Ergebnisschätzer feststellen, was sich in einer geringeren Standardabweichung und engeren Glaubwürdigkeitsintervallen niederschlägt. Folglich erreichten einige relative Behandlungseffekte statistische Signifikanz, die unter alleiniger Betrachtung der direkten Vergleiche nicht gegeben war. Darüber hinaus scheint die direkte Evidenz für einige Vergleiche keine validen Ergebnisse zu liefern; für diese halten indirekte Vergleiche eine weitere Informationsquelle bereit, die anstelle der direkten Evidenz betrachtet werden kann. Letztendlich erlauben die Methoden der klassischen paarweisen Meta-Analyse keine Aussage über eine Rangfolge im Hinblick auf Nutzen oder Schaden einer Therapie; ferner kann sich insbesondere bei großen Netzwerken die Interpretation der Ergebnisse reichlich kompliziert gestalten. Bayesianische MTC Modelle hingegen ermöglichen unter Bezugnahme auf eine Referenzbehandlung ein Ranking der Interventionen und erlaubenWahrscheinlichkeitsaussagen, dass die jeweilige Behandlung die n-beste Option ist und stellen somit eine weitere nützliche Entscheidungshilfe im Rahmen der Evidenzbewertung dar.

In der Zusammenschau erweisen sich Netzwerk Meta-Analysen als eine wertvolle Erweiterung des klassischen Methodenspektrums. Solange die Inferenz sorgfältig abgeleitet wird und ihre Limitationen entsprechend in Erwägung gezogen werden, stellen sie insbesondere im Bereich der Nutzenbewertung und gesundheitsökonomischen Evaluation von Arzneimitteln, deren Erfordernisse über die der Zulassung hinausgehen, einen integralen Bestandteil der Evidenzsynthese dar.

Appendix

A Tables

A.1 Table 10

Table 10: Literature search strategy

I. Medline	
Interface	PubMed
Date of search	2012/04/02 and 2012/07/30 (update)
Limits	None
Records retrieved	2012/04/02: n=7,648 201/07/30: n=315
Syntax guide	
[mh]	Medical Subject Heading (MeSH); Thesaurus, controlled vocabulary of biomedical terms issued by National Library of Medicine.
[tw]	Text word
[nm]	Supplementary concept; includes chemical, protocol or disease terms.
[pt]	Publication type
*	Truncation; use to search for one or more characters.
[mh:noexp]	Turn of feature of including more specific MeSH terms beneath the original one.
Search	
Condition	(((diabetes mellitus, type 2[MeSH Terms]) OR mody[Text Word]) OR niddm[Text Word]) OR T2DM[Text Word]
Boolean operator	AND
Drugs	((((((((((((((((((((((((((((((((((((((acarbose[MeSH Terms]) OR acarbose byproduct, component C[Supplementary Concept]) OR acarbose 7-phosphotransferase) OR acarbose 7-phosphate) OR miglitol[Supplementary Concept]) OR alpha glucosidases/ antagonists and inhibitors[MeSH Terms]) OR gliclazide[MeSH Terms]) OR gliquidone[Supplementary Concept]) OR glimepiride[Supplementary Concept]) OR hydroxyglimepiride[Supplementary Concept]) OR glyburide[MeSH Terms]) OR 4-transhydroxy glyburide[Supplementary Concept]) OR Glucovance[Supplementary Concept]) OR glibenclamide[MeSH Terms]) OR glibenclamide receptor[Supplementary Concept]) OR chlorpropamide[MeSH Terms]) OR glipizide[MeSH Terms]) OR tolbutamide[MeSH Terms]) OR tolbutamide 4-hydroxylase[Supplementary Concept]) OR carboxytolbutamide[Supplementary Concept]) OR tolazamide[MeSH Terms]) OR sulfonylurea compounds[MeSH Terms]) OR pioglitazone[Supplementary Concept]) OR nateglinide[Supplementary Concept]) OR repaglinide[Supplementary Concept]) OR 2-methoxy-4-(3-methyl-1-(2-piperidin-1-ylphenyl)butylcarbamoyl)benzoic acid[Supplementary Concept]) OR meglitinide*[Supplementary Concept]) OR sitagliptin[Supplementary Concept]) OR vildagliptin[Supplementary Concept]) OR lipoyl vildagliptin[Supplementary Concept]) OR linagliptin[Supplementary Concept]) OR saxagliptin[Supplementary Concept]) OR 5-hydroxysaxagliptin[Supplementary Concept]) OR dipeptidyl peptidase iv inhibitors[MeSH Terms]) OR LC15-0444[Supplementary Concept]) OR metformin[MeSH Terms]) OR tetrachloro(metformin)platinum(IV)[Supplementary Concept]) OR drugs, hypoglycemic[MeSH Terms]
Boolean operator	AND

Filter for randomized controlled trials[a]	(randomized controlled trial[pt] OR controlled clinical trial[pt] OR randomized controlled trials[mh] OR random allocation[mh] OR double-blind method[mh] OR single-blind method[mh] OR clinical trial[pt] OR clinical trials[mh] OR ("clinical trial"[tw]) OR ((singl*[tw] OR doubl*[tw] OR trebl*[tw] OR tripl*[tw]) AND (mask*[tw] OR blind*[tw])) OR ("latin square"[tw]) OR placebos[mh] OR placebo*[tw] OR random*[tw] OR research design[mh:noexp] OR follow-up studies[mh] OR prospective studies[mh] OR cross-over studies[mh] OR control*[tw] OR prospectiv*[tw] OR volunteer*[tw]) NOT (animal[mh] NOT human[mh])
II. Embase	
Interface	OVID
Date of search	2012/04/02 and 2012/07/30 (update)
Limits	None
Records retrieved	2012/04/02: n=5,298 201/07/30: n=653
Syntax guide	
?	Truncation symbol for one or no characters only
$	Truncation symbol, or wildcard: retrieves plural or variations of a word
adj#	Proximity operator; adjacency within # number of words (in any order)
*	Indicates that the marked subject heading is a primary topic
tw	Text word
sh	EMTREE Thesaurus terms; controlled vocabulary of biomedical terms used by Embase for indexing
rn	CAS Registry Numbers (RN); contains the Chemical Abstracts Service Registry number in an association with chemical name for a compound mentioned in an article.
Search	
Condition	(diabetes mellitus or maturity onset diabetes mellitus or non insulin dependent diabetes mellitus or lipoatrophic diabetes mellitus).sh. or ((adult or ketosis-resistant or matur* or late or non-insulin depend* or slow or stable or type 2 or type II or lipoatrophic) adj3 diabet$).tw. or (mody or niddm or T2DM).tw.
Boolean operator	AND

Drugs	metformin.sh. or metformin.tw. or (dimethylguanylguanidine or dimethylbiguanidine or glucophage).tw. or (657-24-9 or 1115-70-4).rn. or (apo-metformin or apotex or genmetformin or glucophage or glumetza or novo-metformin or nu-metformin or pms-metformin or ran-metformin or ratio-metformin or rhoxal-metformin or sandoz metformin).tw. or (Glycon or Fortamet or Riomet or Venez or Diaformina or Dimefor or Glafornil or Glucaminol or Glucofage or Diabex or Diaformin or Glucohexal or Glucomet or Novomet or Metomin or Glucamet or Metsol or Orabet).tw. or (Aron or Diabetosan or Diabex or Diformin or Diformin Retard or Dimethylbiguanide or Dmgg or Fluamine or Fortamet or Gliguanid or Glucoformin or Haurymellin or La 6023 or La6023 or Meguan or Mellittin or Metaformin or Methformin or Metiguanide or Metphormin or Dimethylguanylguanide or Nndg or Dimethylbiguanide or Dimethyl Biguanidine or Dimethylbiguanidine or Dimethyldiguanide).tw. or gliclazide.sh. or gliclazide.tw. or gliquidone.sh. or gliquidone.tw. or glimepiride.sh. or glimepiride.tw. or glyburide.sh. or glyburide.tw. or glibenclamide.sh. or glibenclamide.tw. or glybenclamide.sh. or glybenclamide.tw. or chlorpropamide.sh. or chlorpropamide.tw. or glipizide.sh. or glipizide.tw. or tolbutamide.sh. or tolbutamide.tw. or tolazamide.sh. or tolazamide.tw. or (64-77-7 or 1156-19-0 or 94-20-2 or 29094-61-9 or 10238-21-8 or 93479-97-1 or 21187-98-4).rn. or (Orinase or glyconon or Tolinase or Diabinese or glymese or Glucotrol or Diabeta or Micronase or Glynase or gen-glybe or euglucon or Amaryl or Diamicron or diaglyk or glibenese or minodiab or gen-gliclazide).tw. or sulfonylurea.sh. or sulfonylurea.tw. or pioglitazone.sh. or pioglitazone.tw. or actos.tw. or repaglinide.sh. or repaglinide.tw. or nateglinide.sh. or nateglinide.tw. or meglitinide.sh. or meglitinide.tw. or (135062-02-1 or 105816-04-4).rn. or (prandin or gluconorm or starlix or novonorm).tw. or miglitol.sh. or miglitol.tw. or acarbose.sh. or acarbose.tw. or (56180-94-0 or 72432-03-2 or 83480-29-9).rn. or (acarbose or glucobay or precose or prandase or akarbose or miglitol* or glyset or diastabol or voglibose).tw. or alpha glucosidase inhibitor.sh. or alpha glucosidase inhibitor.tw. or (((alph* adj glucos* adj inhibit*) or alf*adj glucos*) adj inhibit*).tw. or (((((sitagliptin or vildagliptin or linagliptin or saxagliptin).sh. or (sitagliptin or vildagliptin or linagliptin or saxagliptin).tw. or (januvia or galvus or trajenta or onglyza or gliptin).tw. or (486460-32-6 or 274901-16-5 or 141758-74-9 or 204656-20-2).rn. or dipeptidyl peptidase IV inhibitor.sh. or dipeptidyl peptidase IV inhibitor.tw. or dpp.mp.) adj IV.mp. adj inhibitor*.tw.) or Dipeptidyl-Peptidase.mp.) adj IV.mp. adj inhibitor*.tw.) or DPP-4 inhibitors.tw. or dipeptidyl peptidase-4 inhibitors.tw. or oral antidiabetic agent.sh. or (((Antidiabetic or anti diabetic or antihyperglycemic or anti-hyperglycemic or oral hypoglycemic or anti-diabetes or antidiabetes) adj agent) or agents or drug or drugs or compound or compounds).tw.
Boolean operator	AND
Filter for randomized controlled trials[b]	((clinical trial or randomized controlled trial or randomization or single blind procedure or double blind procedure or crossover procedure or placebo).sh. or randomi?ed controlled trial$.tw. or rct.tw. or random allocation.tw. or randomly allocated.tw. or allocated randomly.tw. or allocated ajd2 random.tw. or single blind$.tw. or double blind$.tw. or (treble or triple blind$).tw. or placebo$.tw. or prospective study.sh.) not (case study or abstract report or letter).sh.
III. Cochrane Central Register for Controlled Trials (CENTRAL)	
Interface	Cochrane Library
Date of search	2012/04/02 and 2012/07/30 (update)
Limits	Exclusion of records that have been sourced from Medline and Embase
Records retrieved	2012/04/02: n=22 201/07/30: n=22
Syntax guide	
*	Truncation; use to search for one or more characters
adj#/near#	Proximity operator; finds the terms when they are within # words of each other where # = the maximum number of words between search terms. Terms can appear in either order

ti	Title
ab	Abstract
kw	Keyword; limit includes MeSH terms but does not allow for MeSH term explosion. Also searches EMBASE keyword fields
exp	Explode search to include more specific MeSH terms beneath the original one; by default, a MeSH term will be exploded and searched in all trees
MeSH descriptor	Medical Subject Heading (MeSH); Thesaurus, controlled vocabulary of biomedical terms issued by National Library of Medicine
Search	
Condition	((adult or ketosis-resistant or matur* or late or non-insulin depend* or noninsulin depend* or slow or stable or type 2 or type II or lipoatrophic) adj3 diabet$):ti,ab,kw OR (mody or niddm or T2DM):ti,ab,kw or MeSH descriptor: [Diabetes Mellitus, Type 2] explode all trees
Boolean operator	AND
Drugs	(hypoglycemic drugs) or ((Antidiabetic or anti diabetic or antihyperglycemic or anti-hyperglycemic or oral hypoglycemic or anti-diabetes or antidiabetes) adj (agent or agents or drug or drugs or compound or compounds)) or (pioglitazone or actos) or (155141-29-0) or MeSH descriptor Metformin explode all trees or (metformin) or (dimethylguanylguanidine or dimethylbiguanidine or glucophage) or (apo-metformin or apotex or genmetformin or glucophage or glumetza or novo-metformin or nu-metformin or pms-metformin or ran-metformin or ratio-metformin or rhoxal-metformin or sandoz metformin) or (Glycon or Fortamet or Riomet or Venez or Diaformina or Dimefor or Glafornil or Glucaminol or Glucofage or Diabex or Diaformin or Glucohexal or Glucomet or Novomet or Metomin or Glucamet or Metsol or Orabet) or (Aron or Diabetosan or Diabex or Diformin or Diformin Retard or Dimethylbiguanide or Dmgg or Fluamine or Fortamet or Gliguanid or Glucoformin or Haurymellin or La 6023 or La6023 or Meguan or Mellittin or Metaformin or Methformin or Metiguanide or Metphormin or Dimethylguanylguanide or Nndg or Dimethylbiguanide or Dimethyl Biguanidine or Dimethylbiguanidine or Dimethyldiguanide) or (sitagliptin, vildagliptin, linagliptin, saxagliptin) or (januvia or janumet or galvus or eucreas or trajenta or onglyza or gliptin) or (486460-32-6 or 274901-16-5 or 141758-74-9 or 204656-20-2) or MeSH descriptor Dipeptidyl-Peptidase IV Inhibitors explode all trees or (Dipeptidyl-Peptidase adj IV adj inhibitor*) or MeSH descriptor Acarbose explode all trees or MeSH descriptor alpha-Glucosidases explode all trees or ((alph* adj glucos* adj inhibit*) or (alf* adj glucos* adj inhibit*)) or MeSH descriptor Sulfonylurea Compounds explode all trees or MeSH descriptor Gliclazide explode all trees or MeSH descriptor Glyburide explode all trees or MeSH descriptor Chlorpropamide explode all trees or MeSH descriptor Glipizide explode all trees or MeSH descriptor Tolbutamide explode all trees or MeSH descriptor Tolazamide explode all trees or (64-77-7 or 1156-19-0 or 94-20-2 or 29094-61-9 or 10238-21-8 or 93479-97-1 or 21187-98-4) or (repaglinide or nateglinide or meglitinide*) or (135062-02-1 or 105816-04-4) or (prandin or gluconorm or starlix or novonorm) or (gliclazide or gliquidone or glimepiride or glyburide or glibenclamide or glybenclamide or chlopropamide or glipizide or tolbutamide or tolazamide) or (sulfonylurea* or Orinase or glyconon or Tolinase or Diabinese or glymese or Glucotrol or Diabeta or Micronase or Glynase or gen-glybe or euglucon or Amaryl or Diamicron or diaglyk or glibenese or minodiab or gen-gliclazide) or (56180-94-0 or 72432-03-2 or 83480-29-9)
Boolean operator	NOT
Exclusion of records sourced from Medline and Embase	"accession number" near pubmed OR "accession number" near/2 embase
IV. Drug regulatory approval authorities	
Food an Drug Administration (FDA) [94]	Drugs@FDA provides information about most of the drugs approved in the US since 1939. Medical and statistical reviews for the eligible drugs were searched to find unpublished data or missing data from publications.
Date of search	2012/07/16
Records retrieved	n=7

European Medicines Agency (EMA) [93]	The publicly available European Public Assessment Reports (EPAR) for the eligible drugs were searched to find unpublished data or missing data from publications..
Date of search	2012/07/17
Records retrieved	n=11
V. Clinical trial registers (public and pharmaceutical)	
Clinical Trials.gov	ClinicalTrials.gov is a Web-based resource that provides access to clinical studies maintained by the National Library of Medicine and the National Institutes of Health. Each record presents summary information about a study protocol and includes the following: disease or condition, intervention, title, description, and design of the study, requirements for participation (eligibility criteria), locations where the study is being conducted, contact information for the study locations. Some records also include information on the results of the study, such as: description of study participants, outcomes of the study and summary of adverse events experienced by study participants. The register was searched for the eligible drugs to find unpublished data or missing data from publications
Date of search	2012/08/01
Records retrieved	n=15
Limits	Closed studies; studies with results; interventional studies; age group=adults
Pharmaceutical industry trials registers: • Astra Zeneca [87] • Bristol Myers Squibb [88] • Eli Lilly [89] • GlaxoSmithKline [90] • Novartis [91] • Roche [92]	The clinical trial registers of the mentioned manufacturers were searched for the eligible drugs to find unpublished data or missing data from publications
Date of search	2012/08/10
Records retrieved	n=9

[a] To follow a highly sensitive search strategy for randomized controlled trials a validated filter for Medline using PubMed format was used [95].
[b] To follow a highly sensitive search strategy for randomized controlled trials a validated filter for Embase was used [96].

A.2 Table 11

Table 11: List of excluded full-text articles

No.	Reference	Reason for exclusion[a]	Rationale
9	Arjona Ferreira JC, Corry D, Mogensen CE, Slone L, Xu L, Gonzalez EJ, et al. Sitagliptin Versus Glipizide in Participants With Type 2 Diabetes Mellitus and End-Stage Renal Disease. Diabetes, Stoffwechsel und Herz 2011; 6: 430.	5	Missing data for body weight.[b]
24	Pan CY, Yang W, Tou C, Gause-Nilsson I, Zhao J. Efficacy and safety of saxagliptin in drug-naïve Asian patients with type 2 diabetes mellitus: a randomized controlled trial. Diabetes Metab Res Rev 2012 Mar;28(3):268-75.	5	Missing data for HbA1c and body weight.[b]
153	Rokeya B, Parvin M, Bhowmik A, Chowdury AK. Randomized Double-Blind Clinical Trial Comparaing Efficacy and Safety of Pioglitazone and Metformin in Patients with Type 2 Diabetes. Diabetes 2011;60 (Suppl 1): A618.	5	Missing data for HbA1c and body weight; hypoglycemia not reported.
206	Rosenstock J, Gross JL, Salinas CAA, Hissa M, Berglind N, Ravichandran S, et al. Long-Term Safety and Efficacy of Saxagliptin after 4-Year Follow-Up of Patients with Type 2 Diabetes. Diabetes 2011;60 (Suppl 1): A298.	5	Missing data for hypoglycemia; HbA1c and body weight not reported.
248	Engel SS, Xu L, Andryuk PJ, Davies MJ, Amatruda J, Kaufman K et al. Efficacy and tolerability of MK-0893, a glucagon receptor antagonist (GRA), in patients with type 2 diabetes (T2DM). Diabetes 2011;60 (Suppl 1): A85.	5	Missing data for endpoint HbA1c and body weight; hypoglycemia not reported.[b]
304	Ito M, Abe M, Okada K, Sasaki H, Maruyama N, Tsuchida M et al. The dipeptidyl peptidase-4 (DPP-4) inhibitor vildagliptin improves glycemic control in type 2 diabetic patients undergoing hemodialysis. Endocr J. 2011;58(11):979-87.	1	Other OAD as baseline medication allowed.
332	Li P, Chen W, Li L, Liu C, Shan ZY, Su BL et al. Glyburide/metformin therapy for patients with type 2 diabets-a multi-center, randomized double-masked study. Chinese Pharmaceutical Journal 2011;46 (17) :1362-5.	5	Data for body weight not adequately reported (only BMI).
388	Pérez-Monteverde A, Seck T, Xu L, Lee MA, Sisk CM, Williams-Herman DE et al. Efficacy and safety of sitagliptin and the fixed-dose combination of sitagliptin and metformin vs. pioglitazone in drug-naïve patients with type 2 diabetes. Int J Clin Pract. 2011 Sep;65(9):930-8.	5	Body weight and hypoglycemia not reported.[b]
407	Kong AP, Yamasaki A, Ozaki R, Saito H, Asami T, Ohwada S et al. A randomized-controlled trial to investigate the effects of rivoglitazone, a novel PPAR gamma agonist on glucose-lipid control in type 2 diabetes. Diabetes Obes Metab. 2011 Sep;13(9):806-13.	5	Hypoglycemia not reported.[b]

560	Pfützner A, Paz-Pacheco E, Allen E, Frederich R, Chen R; CV181039 Investigators. Initial combination therapy with saxagliptin and metformin provides sustained glycaemic control and is well tolerated for up to 76 weeks. Diabetes Obes Metab. 2011 Jun;13(6):567-76.	2	Extension study: inadequate patient populations for relevant endpoints.
725	Frederich R, Öhman P, Berglind N, Allen E. Clinical characteristics and sustained glycaemic control: a 76-week, randomised, double-blind study of saxagliptin + metformin in treatment-naive patients with type 2 diabetes. Diabetologia; 54 Suppl. 1: S337.	5	HbA1c not reported.
738	Boardman MK, Hanefeld M, Kumar A,Gonzalez JG, de Teresa L, Northrup J et al. DURATION-4: Improvements in glucose control and cardiovascular risk factors in patients with type 2 diabetes treated with exenatide onceweekly, metformin, pioglitazone, or sitagliptin. Diabetologia; 54 Suppl. 1: S314.	4	Duplicate see record number 253.
753	Burant CF, Viswanathan P, Marcinak J, Cao C, Vakilynejad M, Xie B et al. TAK-875 versus placebo or glimepiride in type 2 diabetes mellitus: a phase 2, randomised, double-blind, placebo-controlled trial. Lancet. 2012 Apr 14;379(9824):1403-11.	1	Other OAD as baseline medication allowed.
866	Chou HS, Palmer JP, Jones AR, Waterhouse B, Ferreira-Cornwell C, Krebs J, et al. Initial treatment with fixed-dose combination rosiglitazone/glimepiride in patients with previously untreated type 2 diabetes. Diabetes Obes Metab. 2008 Aug;10(8):626-37.	1	Rosiglitazone not eligible.
905	Hanefeld M, Herman GA, Wu M, Mickel C, Sanchez M, Stein PP; Sitagliptin Study 014 Investigators. Once-daily sitagliptin, a dipeptidyl peptidase-4 inhibitor, for the treatment of patients with type 2 diabetes. Curr Med Res Opin. 2007 Jun;23(6):1329-39.	5	Body weight not reported.
914	Goldstein BJ, Rosenstock J, Anzalone D, Tou C, Ohman KP. Effect of tesaglitazar, a dual PPAR alpha/gamma agonist, on glucose and lipid abnormalities in patients with type 2diabetes: a 12-week dose-ranging trial. Curr Med Res Opin. 2006 Dec;22(12):2575-90.	5	Missing data for body weight.[b]
1029	Charbonnel BH, Matthews DR, Schernthaner G, Hanefeld M, Brunetti P; QUARTET Study Group. A long-term comparison of pioglitazone and gliclazide in patients with Type 2 diabetes mellitus: a randomized, double-blind, parallel-group comparison trial. Diabet Med. 2005 Apr;22(4):399-405.	5	Missing data for HbA1c and body weight.[b]
1033	Li JW, Tian HM, Yu HL, Zhang XX, Zhao GZ, Wang JN. Comparison of efficacy between nateglinide and repaglinide in treating type 2 diabetes: A randomized controlled double-blind clinical trial. Journal of Sichuan University 2005;36(2):267-270.	1	Two drugs from within the same drug class.

1039	Tan MH, Baksi A, Krahulec B, Kubalski P, Stankiewicz A, Urquhart R et al. Comparison of pioglitazone and gliclazide in sustaining glycemic control over 2 years in patients with type 2 diabetes. Diabetes Care. 2005 Mar;28(3):544-50.	5	Hypoglycemia not reported.
1052	Fujioka K, Brazg RL, Raz I, Bruce S, Joyal S, Swanink R et al. Efficacy, dose-response relationship and safety of once-daily extended-release metformin (Glucophage XR) in type 2 diabetic patients with inadequate glycaemic control despite prior treatment with diet and exercise: results from two double-blind, placebo-controlled studies. Diabetes Obes Metab. 2005 Jan;7(1):28-39.	5	Body weight and hypoglycemia not reported.
1055	Smith SR, De Jonge L, Volaufova J, Li Y, Xie H, Bray GA. Effect of pioglitazone on body composition and energy expenditure: a randomized controlled trial. Metabolism. 2005 Jan;54(1):24-32.	1	Other OAD as baseline medication allowed.
1068	Herz M, Johns D, Reviriego J, Grossman LD, Godin C, Duran S et al. A randomized, double-blind, placebo-controlled, clinical trial of the effects of pioglitazone on glycemic control and dyslipidemia in oral antihyperglycemic medication-naive patients with type 2 diabetes mellitus. Clin Ther. 2003 Apr;25(4):1074-95.	5	Missing data for HbA1c and body weight[b]
1099	Horton ES, Clinkingbeard C, Gatlin M, Foley J, Mallows S, Shen S. Nateglinide alone and in combination with metformin improves glycemic control by reducing mealtime glucose levels in type 2 diabetes. Diabetes Care.2000Nov;23(11):1660-5.	5	Missing data for body weight and hypoglycemia.[b]
1106	Blonde L, Rosenstock J, Mooradian AD, Piper BA, Henry D. Glyburide/metformin combination product is safe and efficacious in patients with type 2 diabetes failing sulphonylurea therapy. Diabetes Obes Metab. 2002 Nov;4(6):368-75.	5	HbA1c and body weight not reported.
1114	Marre M, Howlett H, Lehert P, Allavoine T. Improved glycaemic control with metformin-glibenclamide combined tablet therapy (Glucovance) in Type 2 diabetic patients inadequately controlled on metformin. Diabet Med. 2002 Aug;19(8):673-80.	5	Missing data for HbA1c and body weight.[b]
1134	Hasche H, Mertes G, Bruns C, Englert R, Genthner P, Heim D et al. Effects of acarbose treatment in Type 2 diabetic patients under dietary training: A multicentre, double-blind, placebo-controlled, 2-year study. Diabetes Nutr Metab 1999;12(4):277-285.	5	Hypoglycemia not reported.
1144	Tsumura K. Clinical evaluation of glimepiride (HOE490) in NIDDM, including a double blind comparative study versus gliclazide. Diabetes Res Clin Pract. 1995 Aug;28 Suppl:S147-9.	2	No randomization.

1165	Haidinger M, Werzowa J, Voigt HC, Pleiner J, Stemer G, Hecking M et al. A randomized, placebo-controlled, double-blind, prospective trial to evaluate the effect of vildagliptin in new-onset diabetes mellitus after kidney transplantation. Trials. 2010 Oct 6;11:91. doi: 10.1186/1745-6215-11-91.	2	Study protocol.
1265	Truitt KE, Goldberg RB, Rosenstock J, Chou HS, Merante D, Triscari J et al. A 26-week, placebo- and pioglitazone-controlled, dose-ranging study of rivoglitazone, a novel thiazolidinedione for the treatment of type 2 diabetes. Curr Med Res Opin. 2010 Jun;26(6):1321-31.	5	Missing data for body weight.[b]
1573	Pfützner A, Gurieva I, Antsiferov M, Allen E, Ravichandran S, Chen R. Saxagliptin either as add-on therapy to metformin or as initial combination therapy with metformin improves glycaemic control in patients with type 2 diabetes. Endocrine abstracts 2010; 20 Suppl.: 359.	4	Duplicate see record number 560.
1688	G. Derosa, S.A.T. Salvadeo, I. Ferrari, A. Gravina, R. Mereu, A. D'Angelo et al. Acarbose compared to placebo on insulin resistance biomarkers in a double-blind, placebo-controlled trial. Diabetologia 2010; 53 Suppl 1: S361.	4	Duplicate see record number 649.
1689	Maffioli P, Salvadeo SAT, Ferrari I, Gravina A, Mereu R, D'Angelo A. et al. Effect of acarbose on inflammatory parameters at baseline and after anoral fat load: a double-blind, placebo controlled trial. Diabetologia 2010; 53 Suppl 1: S361.	4	Duplicate see record number 649.
1698	Ferrannini E, Seman LJ, Seewaldt-Becker E, Hantel S, Pinnetti S, Woerle HJ. The potent and highly selective sodium-glucose co-cransporter (SGLT-2)inhibitor BI 10773 is safe and efficacious as monotherapy in patients with type 2 diabetes mellitus. Diabetologia 2010; 53 Suppl 1: S351.	5	Missing data forHbA1c and body weight; hypoglycemia not reported.[b]
1711	Chiasson JL,Leiter LA, Forti A, Bergenstal LM, Woloschak M, Boldrin M. Taspoglutide, a once-weekly human GLP-1 analogue, is superior to Sitagliptin in improving glycaemic control and achieving weight loss in patients with type 2 diabetes: the T-emerge 4 Trial. Diabetologia 2010; 53 Suppl 1: S334	1	Other OAD as baseline medication allowed.
1720	Barnett AH, Harper R, Toorawa R, Patel S, Woerle HJ. Linagliptin monotherapy improves glycaemic control in type 2 diabetes patients for whom metformin therapy is inappropriate. Diabetologia 2010; 53 Suppl 1: S327.	5	Body weight not reported.
1722	A. Pérez-Monteverde1, T. Seck2, L. Xu2, G. Liu2, M.A. Lee2, C. McCrary et al. Efficacy and safety of sitagliptin and the fixed-dose combination of sitagliptin and metformin versus pioglitazone in drug-naïve patients with type 2 diabetes. Diabetologia 2010; 53 Suppl 1: S325.	4	Duplicate see record number 388.

1834	Perez A, Zhao Z, Jacks R, Spanheimer R. Efficacy and safety of pioglitazone/metformin fixed-dose combination therapy compared with pioglitazone and metformin monotherapy in treating patients with T2DM. Curr Med Res Opin. 2009 Dec;25(12):2915-23.	5	Missing data for HbA1c and body weight.[b]
1942	Rosenstock J, Aguilar-Salinas C, Klein E, Nepal S, List J, Chen R; CV181-011 Study Investigators. Effect of saxagliptin monotherapy in treatment-naïve patients with type 2 diabetes. Curr Med Res Opin. 2009 Oct;25(10):2401-11.	5	Missing data for body weight.[b]
2164	Spanheimer R, Zhao Z, Perez A. Evaluating Effect of Insulin Resistance and ß-Cell Function in a Pioglitazone and Metformin Fixed-Dose Combination Study. Diabetologia 2009; 52 Suppl. 1: S212.	4	Duplicate see record number 1834.
2171	Spanheimer R, Zhao Z, Perez A. Effect of Pioglitazone and Metformin Fixed-Dose Combination on Glycemic Control. Diabetologia 2009; 52 Suppl. 1: S335.	4	Duplicate see record number 1834.
2173	DeFronzo R, Burant C, Fleck P, Wilson C, Mekki Q, R. Pratley R. Effect of alogliptin combined with pioglitazone on glycaemic control in metformin-treated patients with type 2 diabetes. Diabetologia 2009; 52 Suppl. 1: S295.	1	Other OAD as baseline medication allowed.
2240	Spanheimer R, Zhao Z, Perez A. Effect of Pioglitazone and Metformin Fixed-Dose Combination on Glycemic Control. Diabetologia 2009; 52 Suppl. 1: S335.	4	Duplicate see record number 1834.
2269	Spanheimer R, Zhao Z, Perez A. Evaluating Effect of Insulin Resistance and ß-Cell Function in a Pioglitazone and Metformin Fixed-Dose Combination Study. Diabetologia 2009; 52 Suppl. 1: S212.	4	Duplicate see record number 1834.
2306	Spanheimer R, Zhao Z, Perez A. Improvement of glycaemic control via reducing insulin resistance with pioglitazone and metformin fixed-dose combination therapy. Canadian Journal of Diabetes 2009; September: IDF 2009 20th World Diabetes Congress North American Poster and Oral Abstracts: 273	4	Duplicate see record number 1834.
2333	Gurieva I, Pfuetzner A, Antsiferov M, Allen E, Ravichandran S, Chen R. Saxagliptin improves glycaemic control either as add-on therapy to metformin or as initial combination therapy with metformin in patients with type 2 diabetes. Eur Heart J 2009; 30 (suppl 1): 226.	4	Duplicate see record number 1995.
2484	Nissen SE, Nicholls SJ, Wolski K, Nesto R, Kupfer S, Perez A et al. Comparison of pioglitazone vs glimepiride on progression of coronary atherosclerosis in patients with type 2 diabetes: the PERISCOPE randomized controlled trial. JAMA 2008 Apr 2;299(13):1561-73.	1	Other OAD as baseline medication allowed.

2569	Schwarz SL, Gerich JE, Marcellari A, Jean-Louis L, Purkayastha D, Baron MA. Nateglinide, alone or in combination with metformin, is effective and well tolerated in treatment-naïve elderly patients with type 2 diabetes. Diabetes Obes Metab 2008 Aug;10(8):652-60.	5	Body weight not reported.
2583	Chan JC, Scott R, Arjona Ferreira JC, Sheng D, Gonzalez E, Davies MJ et al. Safety and efficacy of sitagliptin in patients with type 2 diabetes and chronic renal insufficiency. Diabetes Obes Metab 2008 Jul;10(7):545-55.	4	Duplicate see record number 7670.
2876	Goldstein BJ, Feinglos MN, Lunceford JK, Johnson J, Williams-Herman DE; Sitagliptin 036 Study Group. Effect of initial combination therapy with sitagliptin, a dipeptidyl peptidase-4 inhibitor, and metformin on glycemic control in patients with type 2 diabetes. Diabetes Care 2007 Aug;30(8):1979-87.	5	Missing data for body weight.[b]
2898	Li J, Tian H, Li Q, Wang N, Wu T, Liu Y et al. Improvement of insulin sensitivity and beta-cell function by nateglinide and repaglinide in type 2 diabetic patients - a randomized controlled double-blind and double-dummy multicentre clinical trial. Diabetes Obes Metab 2007 Jul;9(4):558-65.	1	Other OAD as baseline medication allowed.
3156	Hao Y, Han Q, Yuan W, He X. Effects of pioglitazone on elderly patients with type 2 diabetes mellitus. Medical Journal of Wuhan University 2006;27(1):104-7.	5	Body weight not reported.
3309	Goto, T. Clinical efficacy of miglitol in type 2 Japanese diabetics with diet alone - Phase II dose response study. JapPharmTher 2005; 33 (11):1099-1111.	5	Body weight and hypoglycemia not reported.
3612	Miyazaki Y, Matsuda M, DeFronzo RA. Dose-response effect of pioglitazone on insulin sensitivity and insulin secretion in type 2 diabetes. Diabetes Care 2002 Mar;25(3):517-23.	5	Hypoglycemia not reported.
3654	Schernthaner G, Matthews DR, Charbonnel B, Hanefeld M, Brunetti P; Quartet [corrected] Study Group. Efficacy and safety of pioglitazone versus metformin in patients with type 2 diabetes mellitus: a double-blind, randomized trial. Clin Endocrinol Metab 2004 Dec;89(12):6068-76.	5	Missing data for body weight[b]; hypoglycemia not reported.
3783	Horton ES, Foley JE, Shen SG, Baron MA. Efficacy and tolerability of initial combination therapy with nateglinide and metformin in treatment-naïve patients with type 2 diabetes. Curr Med Res Opin 2004 Jun;20(6):883-9.	5	Missing data for body weight.[b]
3884	van de Laar FA, Lucassen PL, Kemp J, van de Lisdonk EH, van Weel C, Rutten GE. Is acarbose equivalent to tolbutamide as first treatment for newly diagnosed type 2 diabetes in general practice? A randomised controlled trial. Diabetes Res Clin Pract 2004 Jan;63(1):57-65.	5	Missing data for body weight[b] (only BMI).

3898	Gulias-Herrero A, Aguilar-Salinas CA, Gomez-Perez FJ, Rull JA. The combination metformin/glyburide exerts its hypoglycemic effect mainly by increasing insulin secretion: A controlled, randomized, double-blind, crossover study. Diabetes Nutr Metab 2003;16(5-6):268-276.	3	Cross over periods only 10 weeks.
3996	Pavo I, Jermendy G, Varkonyi TT, Kerenyi Z, Gyimesi A, Shoustov S et al. Effect of pioglitazone compared with metformin on glycemic control and indicators of insulin sensitivity in recently diagnosed patients with type 2 diabetes. J Clin Endocrinol Metab 2003 Apr;88(4):1637-45.	5	Hypoglycemia not reported.
4024	Del Prato S, Erkelens DW, Leutenegger M. Six-month efficacy of benfluorex vs. placebo or metformin in diet-failed type 2 diabetic patients. Acta Diabetol 2003 Mar;40(1):20-7.	5	Hypoglycemia not reported.
4025	Fischer S, Patzak A, Rietzsch H, Schwanebeck U, Köhler C, Wildbrett J et al. Influence of treatment with acarbose or glibenclamide on insulin sensitivity in type 2 diabetic patients. Diabetes Obes Metab 2003 Jan;5(1):38-44.	5	Hypoglycemia not reported.
4161	Josse RG, Chiasson JL, Ryan EA, Lau DC, Ross SA, Yale JF et al. Acarbose in the treatment of elderly patients with type 2 diabetes. Diabetes Res Clin Pract 2003 Jan;59(1):37-42.	5	Body weight not reported.
4167	Scherbaum WA, Göke B; German Pioglitazone Study Group. Metabolic efficacy and safety of once-daily pioglitazone monotherapy in patients with type 2 diabetes: a double-blind, placebo-controlled study. Horm Metab Res 2002 Oct;34(10):589-95.	5	Missing data for body weight[b]; hypoglycemia not reported.
4171	Drent ML, Tollefsen AT,van Heusden FH, Hoenderdos EB, Jonker JJ, van der Veen EA. Dose-dependent efficacy of miglitol, an alpha-glucosidase inhibitor, in type 2 diabetic patients on diet alone: results of a 24-week double-blind placebo-controlled study. Diabetes, nutrition & metabolism 2002; 15(3):152-9.	5	Body weight and hypoglycemia not reported.
4208	Hällsten K, Virtanen KA, Lönnqvist F, Sipilä H, Oksanen A, Viljanen T et al. Rosiglitazone but not metformin enhances insulin- and exercise-stimulated skeletal muscle glucose uptake in patients with newly diagnosed type 2 diabetes. Diabetes 2002 Dec;51(12):3479-85.	5	Hypoglycemia not reported.
4229	Saloranta C, Hershon K, Ball M, Dickinson S, Holmes D. Efficacy and safety of nateglinide in type 2 diabetic patients with modest fasting hyperglycemia. J Clin Endocrinol Metab 2002 Sep;87(9):4171-6.	5	Missing data for body weight.[b]
4263	Garber AJ, Larsen J, Schneider SH, Piper BA, Henry D; Glyburide/Metformin Initial Therapy Study Group. Simultaneous glyburide/metformin therapy is superior to component monotherapy as an initial pharmacological treatment for type 2 diabetes. Diabetes Obes Metab 2002 May;4(3):201-8.	5	Missing data for HbA1c and body weight.[b]

4427	Moses RG, Gomis R, Frandsen KB, Schlienger JL, Dedov I. Flexible meal-related dosing with repaglinide facilitates glycemic control in therapy-naive type 2 diabetes. Diabetes Care 2001 Jan;24(1):11-5.	5	Missing data for HbA1c and body weight.[b]
4447	Hermann LS, Ranstam J, Valler S, Melander A. Effect of antihyperglycaemic therapies on proinsulin and relation between proinsulin and cardiovascular risk factors in type 2 diabetes. Diabetes Obes Metab 1999;1(4):227-232.	5	HbA1c, body weight and hypoglycemia not reported.
4479	Wolever T, DM, Assiff L, Basu T, Chiasson JL, Boctor M, Gerstein C et al. Miglitol, an ct-Glucosidase inhibitor, prevents the Metformin-induced fall in serum folate and vitamin B12 in subjects with type 2 diabetes. NutritionRes 2000;20(10):1447-1456.	4	Duplicate see record number see 4338.
4499	Meneilly GS, Ryan EA, Radziuk J, Lau DC, Yale JF, Morais J et al. Effect of acarbose on insulin sensitivity in elderly patients with diabetes. Diabetes Care 2000 Aug;23(8):1162-7.	5	Hypoglycemia not reported.
4539	Lindström J, Tuomilehto J, Spengler M. Acarbose treatment does not change the habitual diet of patients with Type 2 diabetes mellitus. The Finnish Acargbos Study Group. Diabet Med 2000 Jan;17(1):20-5.	1	Other OAD as baseline medication allowed.
4541	Hanefeld M, Bouter KP, Dickinson S, Guitard C. Rapid and short-acting mealtime insulin secretion with nateglinide controls both prandial and mean glycemia. Diabetes Care 2000 Feb;23(2):202-7.	5	Missing data for body weight and hypoglycemia.[b]
4574	N.N.	2	Erratum of original article.
4588	Moses R. Repaglinide in combination therapy with metformin in Type 2 diabetes. Exp Clin Endocr Diab 1999; 107 Suppl4: S136-9.	5	Missing data for HbA1c and body weight.[b]
4610	Holman RR, Cull CA, Turner RC. A randomized double-blind trial of acarbose in type 2 diabetes shows improved glycemic control over 3 years (U.K. Prospective Diabetes Study 44). Diabetes Care 1999 Jun;22(6):960-4.	5	Missing data for HbA1c and body weight[b]; hypoglycemia not reported.
4616	Scott R, Lintott CJ, Zimmet P, Campbell L, Bowen K, Welborn T. Will acarbose improve the metabolic abnormalities of insulin-resistant type 2 diabetes mellitus? Diabetes Res Clin Pract 1999 Mar;43(3):179-85.	5	Body weight and hypoglycemia not reported.
4626	Gentile S, Turco S, Guarino G, Oliviero B, Rustici A, Torella R.Non-insulin-dependent diabetes mellitus associated with nonalcoholic liver cirrhosis: an evaluation of treatment with the intestinal alpha-glucosidase inhibitor acarbose. Ann Ital Med Int 1999 Jan-Mar;14(1):7-14.	5	Body weight and hypoglycemia not reported.
4663	Lee A, Morley JE. Metformin decreases food consumption and induces weight loss in subjects with obesity with type II non-insulin-dependent diabetes. Obes Res 1998;6(1):47-53.	5	Hypoglycemia not reported.

4682	Johnston PS, Lebovitz HE, Coniff RF, Simonson DC, Raskin P, Munera CL. Advantages of alpha-glucosidase inhibition as monotherapy in elderly type 2 diabetic patients. J Clin Endocrinol Metab 1998 May;83(5):1515-22.	5	Missing data for HbA1c and body weight.[b]
4693	Goldberg RB, Einhorn D, Lucas CP, Rendell MS, Damsbo P, Huang WC et al. A randomized placebo-controlled trial of repaglinide in the treatment of type 2 diabetes. Diabetes Care 1998 Nov;21(11):1897-903.	5	Missing data for body weight and hypoglycemia.[b]
4709	Turner R, UK Prospective Diabetes Study (UKPDS) Group. Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). UK Prospective Diabetes Study (UKPDS) Group. Lancet 1998 Sep 12;352(9131):854-65.	3	10 yr follow-up data too long for reasonable comparison with included studies (follow-up: 12wks-2yrs); data for cut-off from one year not provided from authors.
4714	Cefalu WT, Bell-Farrow A, Wang ZQ, McBride D, Dalgleish D, Terry JG. Effect of Glipizide GITS on Insulin Sensitivity, Glycemic Indices, and Abdominal Fat Composition in NIDDM. Drug Develop Res 1998;44:1-7.	5	Hypoglycemia not reported.
4717	Chan JC, Chan KW, Ho LL, Fuh MM, Horn LC, Sheaves R et al. An Asian multicenter clinical trial to assess the efficacy and tolerability of acarbose compared with placebo in type 2 diabetic patients previously treated with diet. Asian Acarbose Study Group. Diabetes Care 1998 Jul;21(7):1058-61.	5	Hypoglycemia not reported.
4726	Schade DS, Jovanovic L, Schneider J. A placebo-controlled, randomized study of glimepiride in patients with type 2 diabetes mellitus for whom diet therapy is unsuccessful. J Clin Pharmacol 1998 Jul;38(7):636-41.	5	Missing data for HbA1c and body weight.[b]
4730	Fischer S, Hanefeld M, Spengler M, Boehme K. European study on dose-response relationship of acarbose as a first-line drug in non-insulin-dependent diabetes mellitus: efficacy and safety of low and high doses. Acta Diabetol 1998;35:34-40.	5	Missing data for body weight.[b]
4750	Johnston PS, Feig PU, Coniff RF, Krol A, Kelley DE, Mooradian AD. Chronic treatment of African-American type 2 diabetic patients with alpha-glucosidase inhibition. Diabetes Care 1998 Mar;21(3):416-22.	5	Missing data for body weight and hypoglycemia.[b]
4751	Johnston PS, Krol A, Feig PU, Davidson JA, Coniff RF, Haffner SM. Long-Term Titrated-Dose a-Glucosidase Inhibition in Non-Insulin-Requiring Hispanic NIDDM Patients. Diabetes Care 1998 Mar;21(3):409-15.	5	Missing data for HbA1c b; body weight and hypoglycemia not reported.
4755	UK Prospective Diabetes Study (UKPDS) Group. United Kingdom Prospective Diabetes Study 24: A 6-Year, Randomized, Controlled Trial Comparing Sulfonylurea, Insulin, and Metformin Therapy in Patients with Newly Diagnosed Type 2 Diabetes That Could Not Be Controlled with Diet Therapy. Ann Int Med 1998 Febr1; 128(3):165-75.	3	10 yr follow up data too long for reasonable comparison with included studies (follow-up: 12wks-2yrs); data for cut-off from one year not provided from authors.

4769	Garber AJ, Duncan TG, Goodman AM, Mills DJ, Rohlf JL. Efficacy of metformin in type II diabetes: results of a double-blind, placebo-controlled, dose-response trial. Am J Med 1997 Dec;103(6):491-7.	5	Missing data for HbA1c[b]; body weight not reported.
4795	Kovacevic I,Profozic V, Skrabalo Z, Cabrijan T, Zjacic-Rotkvic V, Goldoni V, Jovic-Paskvalin Lj et al. Multicentric clinical trial to assess efficacy and tolerability of Acarbose (BAY G 5421) in comparison to Glibenclamide and Placebo. Diabetologia Croatica 1997;26(2):83-9.	5	Body weight and hypoglycemia not reported.
4823	Segal P, Feig PU, Schernthaner G, Ratzmann KP, Rybka J, Petzinna D et al. The efficacy and safety of miglitol therapy compared with glibenclamide in patients with NIDDM inadequately controlled by diet alone. Diabetes Care 1997 May;20(5):687-91.	5	Missing data for body weight.[b]
4856	Braun D, Schonherr U, Mitzkat HJ. Efficacy of acarbose monotherapy in patients with type 2 diabetes: A double-blind study conducted in general practice. Endocrinol Metab 1996;3(4):275-80.	5	Hypoglycemia not reported.
4865	Rosenstock J, Samols E, Muchmore DB, Schneider J. Glimepiride, a new once-daily sulfonylurea. A double-blind placebo-controlled study of NIDDM patients. Glimepiride Study Group. Diabetes Care 1996 Nov;19(11):1194-9.	5	Missing data for HbA1c and hypoglycemia[b]; body weight not reported
4872	Dills DG, Schneider J. Clinical evaluation of glimepiride versus glyburide in NIDDM in a double-blind comparative study. Glimepiride/Glyburide Research Group. Horm Metab Res 1996 Sep;28(9):426-9.	1	Two drugs from within the same drug class.
4873	Draeger KE, Wernicke-Panten K, Lomp HJ, Schüler E, Rosskamp R. Long-term treatment of type 2 diabetic patients with the new oral antidiabetic agent glimepiride (Amaryl): a double-blind comparison with glibenclamide. Horm Metab Res 1996 Sep;28(9):419-25.	1	Two drugs from within the same drug class.
4887	Goldberg RB, Holvey SM, Schneider J. A dose-response study of glimepiride in patients with NIDDM who have previously received sulfonylurea agents. The Glimepiride Protocol #201 Study Group. Diabetes Care 1996 Aug;19(8):849-56.	5	Missing data for HbA1c and body weight.[b]
4918	Grant PJ. The effects of high- and medium-dose metformin therapy on cardiovascular risk factors in patients with type II diabetes. Diabetes Care 1996 Jan;19(1):64-6.	5	Body weight and hypoglycemia not reported.
4923	Rodger NW,Chiasson JL, Josse RG, Hunt JA, Palmason C, Ross SA et al. Clinical experience with acarbose: Results of a Canadian multicentre study. Clin Invest Med 1995;18(4):318-24.	5	Body weight and hypoglycemia not reported.
4940	Josse RG. Acarbose for the treatment of type II diabetes: the results of a Canadian multi-centre trial. Diabetes Res Clin Pract 1995 Aug;28 Suppl:S167-72.	4	Duplicate see record number 4923.

4944	Pagano G, Marena S, Corgiat-Mansin L, Cravero F, Giorda C, Bozza M et al. Comparison of miglitol and glibenclamide in diet-treated type 2 diabetic patients. Diabetes Metab 1995;21(3):162-7.	5	Body weight and hypoglycemia not reported.
4947	Dedov I, Balabolkin MI, Mkrtumyan AM, Ametov AS, Kakhnovsky IM, Chazova TE et al. Glucobai therapy of diabetes mellitus. Problemy Endokrinologii 1995;41(3):11-3.	5	Missing data for body weight[b]; hypoglycemia not reported.
4957	Coniff RF, Shapiro JA, Robbins D, Kleinfield R, Seaton TB, Beisswenger P et al. Reduction of glycosylated hemoglobin and postprandial hyperglycemia by acarbose in patients with NIDDM. A placebo-controlled dose-comparison study. Diabetes Care 1995 Jun;18(6):817-24.	5	Missing data for HbA1c and body weight[b]; hypoglycemia not reported.
4961	Banerji MA, Chaiken RL, Lebovitz HE. Prolongation of near-normoglycemic remission in black NIDDM subjects with chronic low-dose sulfonylurea treatment. Diabetes 1994;11(10):466-70.	2	No randomization, time to event data.
4965	Escobar-Jimenez F, Barajas C, de Leiva A, Cano FJ, Masoliver R, Herrera-Pombo JL et al. Efficacy and tolerability of Miglitol in the treatment of patients with non-insulin dependent diabetes mellitus. Curr Ther Res Clin E 1995;56(3):258-68.	5	Body weight not reported.
4996	Tessier D, Dawson K, Tetrault JP, Bravo G, Meneilly GS. Glibenclamide vs gliclazide in Type 2 diabetes of the elderly. Diabetic Med 1994;11(10):974-80.	1	Two drugs from within the same drug class.
4998	Chiasson JL, Josse RG, Hunt JA, Palmason C, Rodger NW, Ross SA et al. The efficacy of acarbose in the treatment of patients with non-insulin-dependent diabetes mellitus. A multicenter controlled clinical trial. Ann Intern Med 1994 Dec 15;121(12):928-35.	5	Body weight and hypoglycemia not reported.
5001	Coniff RF, Shapiro JA, Seaton TB. Long-term efficacy and safety of acarbose in the treatment of obese subjects with non-insulin-dependent diabetes mellitus. Arch Intern Med 1994 Nov 14;154(21):2442-8.	5	Missing data for HbA1c and body weight[b]; hypoglycemia not reported.
5024	Hoffmann J, Spengler M. Efficacy of 24-week monotherapy with acarbose, glibenclamide, or placebo in NIDDM patients. The Essen Study. Diabetes Care 1994 Jun;17(6):561-6.	5	Missing data for HbA1c[b]; body weight and hypoglycemia not reported.
5031	Birkeland KI, Furuseth K, Melander A, Mowinckel P, Vaaler S. Long-term randomized placebo-controlled double-blind therapeutic comparison of glipizide and glyburide. Glycemic control and insulin secretion during 15 months. Diabetes Care 1994 Jan;17(1):45-9.	5	HbA1c, body weight and hypoglycemia not reported.

5036	Marena S, Pagani A, Montegrosso G, Boella A, Pagano A, de Michieli F et al. Valutazione dell´efficacia del trattamento a lungo termine con Miglitol e con Glibenclamide in diabetici non insulino-dipendenti. Giornale Italiano di Diabetologia 1993; 13:383-88.	5	Hypoglycemia not reported.
5041	de Leiva A, Pinon F, Tebar J, Escobar-Jimenez F, de la Calle H, Herrera-Pombo JL et al. Clinical efficacy and tolerance to acarbose in the treatment of non-insulin-dependent diabetic patients. MedClin-Barcelona 1993;100(10):368-71.	6	Not available.
5047	Santeusanio F, Ventura MM, Contadini S, Compagnucci P, Moriconi V, Zaccarini P et al. Efficacy and safety of two different dosages of acarbose in non-insulin dependent diabetic patients treated by diet alone. Diabetes Nutr Metab 1993;6(3):147-54.	5	Missing data for HbA1c,body weight and hypoglycemia.[b]
5049	Johnson AB, Webster JM, Sum CF, Heseltine L, Argyraki M, Cooper BG et al. The impact of metformin therapy on hepatic glucose production and skeletal muscle glycogen synthase activity in overweight type II diabetic patients. Metabolism 1993 Sep;42(9):1217-22.	5	Hypoglycemia not reported.
5068	Hotta N, Kakuta H, Sano T, Matsumae H, Yamada H, Kitazawa S et al. Long-term effect of acarbose on glycaemic control in non-insulin-dependent diabetes mellitus: A placebo-controlled double-blind study. Diabetic Med 1993;10(2):134-8.	5	Hypoglycemia not reported.
5072	Jenney A, Proietto J, O'Dea K, Nankervis A, Traianedes K, D'Embden H. Low-dose acarbose improves glycemic control in NIDDM patients without changes in insulin sensitivity. Diabetes Care 1993 Feb;16(2):499-502.	2	No randomization.
5151	Hanefeld M, Fischer S, Schulze J, Spengler M, Wargenau M, Schollberg K et al. Therapeutic potentials of acarbose as first-line drug in NIDDM insufficiently treated with diet alone. Diabetes Care 1991 Aug;14(8):732-7.	5	Hypoglycemia not reported.
5153	Leonhardt W, Hanefeld M, Fischer S, Schulze J, Spengler M. Beneficial effects on serum lipids in noninsulin dependent diabetics by acarbose treatment. Arzneimittel-Forsch 1991;41(7):735-8.	5	HbA1c, body weight and hypoglycemia not reported.
5161	Teupe B, Bergis K. Prospective randomized two-years clinical study comparing additional metformin treatment with reducing diet in type 2 diabetes. Diabetes Metab 1991;17(1):213-7.	5	Hypoglycemia not reported.
5162	Hermann LS, Bitzen PO, Kjellstrom T, Lindgarde F, Schersten B. Comparative efficacy of metformin and glibenclamide in patients with non-insulin-dependent diabetes mellitus. Diabetes Metab 1991;17(1):201-8.	5	Hypoglycemia not reported.
5172	Dornan TL, Heller SR, Peck GM, Tattersall RB. Double-blind evaluation of efficacy and tolerability of metformin in NIDDM. Diabetes Care 1991 Apr;14(4):342-4.	5	Hypoglycemia not reported.

5208	Josephkutty S,Potter JM. Comparison of tolbutamide and metformin in elderly diabetic patients. Diabetic Med 1990;7(6):510-4.	2	Cross-over study: no wash out after 1st sequence; carry over effect could bias the outcomes.
5213	Raptis AE, Tountas N, Yalouris AG, Hadjidakis D, Zahari A, Miras K et al. Comparative study of the therapeutic effects of glibenclamide or the fixed combination of glibenclamide-phenformin with those of gliclazide or chlorpropamide. Acta Diabetol Lat 1990;27(1):11-22.	2	No randomization.
5291	Buchanan DR, Collier A, Rodrigues E, Millar AM, Gray RS, Clarke BF. Effectiveness of acarbose, an alpha-glucosidase inhibitor, in uncontrolled non-obese non-insulin dependent diabetes. Eur J Clin Pharmacol 1988;34(1):51-3.	5	Hypoglycemia not reported.
6447	Wu CZ, Pei D, Hsieh AT, Wang K, Lin JD, Lee LH et al. Comparison of insulin sensitivity, glucose sensitivity, and first phase insulin secretion in patients treated with repaglinide or gliclazide. Arch Pharm Res 2010 Mar;33(3):411-6.	2	HbA1c, body weight and hypoglycemia not reported.
7017	Henry RR, Lincoff AM, Mudaliar S, Rabbia M, Chognot C, Herz M. Effect of the dual peroxisome proliferator-activated receptor-alpha/gamma agonist aleglitazar on risk of cardiovascular disease in patients with type 2 diabetes (SYNCHRONY): a phase II, randomised, dose-ranging study. Lancet 2009 Jul 11;374(9684):126-35.	4	Duplicate see record number 2114.
7022	Jadzinsky M, Pfützner A, Paz-Pacheco E, Xu Z, Allen E, Chen R; CV181-039 Investigators. Saxagliptin given in combination with metformin as initial therapy improves glycaemic control in patients with type 2 diabetes compared with either monotherapy: a randomized controlled trial. Diabetes Obes Metab 2009 Jun;11(6):611-22.	4	Duplicate see record number 1995.
7032	Tsuchiya K, Akaza I, Yoshimoto T, Hirata Y. Pioglitazone improves endothelial function with increased adiponectin and high-density lipoprotein cholesterol levels in type 2 diabetes. Endocr J 2009;56(5):691-8.	1	Other OAD as baseline medication allowed.
7223	Gupta AK, Smith SR, Greenway FL, Bray GA. Pioglitazone treatment in type 2 diabetes mellitus when combined with portion control diet modifies the metabolic syndrome. Diabetes Obes Metab 2009 Apr;11(4):330-7.	5	Hypoglycemia not reported.
7494	Erdem G, Dogru T, Tasci I, Bozoglu E, Muhsiroglu O, Tapan S et al. The effects of pioglitazone and metformin on plasma visfatin levels in patients with treatment naive type 2 diabetes mellitus. Diabetes Res Clin Pract 2008 Nov;82(2):214-8	5	Missing data for body weight[b]; hypoglycemia not reported.
7670	Chan JC, Scott R, Arjona Ferreira JC, Sheng D, Gonzalez E, Davies MJ. Safety and efficacy of sitagliptin in patients with type 2 diabetes and chronic renal insufficiency. Diabetes Obes Metab 2008 Jul;10(7):545-55.	1	Other OAD as baseline medication allowed.

7677	Li YX, Ding GX, Li QF, Chen L, Hu GL, Ji QH et al. Clinical evaluation of efficacy and safety of nateglinide in the treatment of type 2 diabetes. Zhongguo Yi Xue Ke Xue Yuan Xue Bao 2008;30(2):211-3.	1	Other OAD as baseline medication allowed.
7798	Kusaka I, Nagasaka S, Horie H, Ishibashi S. Metformin, but not pioglitazone, decreases postchallenge plasma ghrelin levels in type 2 diabetic patients: a possible role in weight stability? Diabetes Obes Metab 2008 Nov;10(11):1039-46.	1	Other OAD as baseline medication allowed.
7976	Chien HH, Chang CT, Chu NF, Hsieh SH, Huang YY, Lee IT et al. Effect of glyburide-metformin combination tablet in patients with type 2 diabetes. J Chin Med Assoc 2007 Nov;70(11):473-80.	5	Missing data for body weight.[b]
8093	Mari A, Scherbaum WA, Nilsson PM, Lalanne G, Schweizer A, Dunning BE et al. Characterization of the influence of vildagliptin on model-assessed -cell function in patients with type 2 diabetes and mild hyperglycemia. J Clin Endocrinol Metab 2008 Jan;93(1):103-9.	4	Duplicate see record number 2568.
8109	Perriello G, Pampanelli S, Brunetti P, di Pietro C, Mariz S; Italian Pioglitazone Study Group. Long-term effects of pioglitazone versus gliclazide on hepatic and humoral coagulation factors in patients with type 2 diabetes. Diab Vasc Dis Res 2007 Sep;4(3):226-30.	4	Duplicate see record number 1009.
8297	Heliövaara MK, Herz M, Teppo AM, Leinonen E, Ebeling P. Pioglitazone has anti-inflammatory effects in patients with Type 2 diabetes. J Endocrinol Invest 2007 Apr;30(4):292-7.	5	Missing data for body weight[b]; hypoglycemia not reported.
8345	Teramoto T, Yamada N, Shirai K, Saito Y. Effects of pioglitazone hydrochloride on Japanese patients with type 2 diabetes mellitus. J Atheroscler Thromb. 2007 Apr;14(2):86-93.	5	Body weight and hypoglycemia not reported.
8499	Rosenstock J, Kim SW, Baron MA, Camisasca RP, Cressier F, Couturier A et al. Efficacy and tolerability of initial combination therapy with vildagliptin and pioglitazone compared with component monotherapy in patients with type 2 diabetes. Diabetes Obes Metab 2007 Mar;9(2):175-85.	4	Duplicate see record number 3028.
9140	Forst T, Lübben G, Hohberg C, Kann P, Sachara C, Gottschall V et al. Influence of glucose control and improvement of insulin resistance on microvascular blood flow and endothelial function in patients with diabetes mellitus type 2. Microcirculation 2005 Oct-Nov;12(7):543-50.	1	Other OAD as baseline medication allowed.

9183	Forst T, Hohberg C, Fuellert SD, Lübben G, Konrad T, Löbig M et al. Pharmacological PPARgamma stimulation in contrast to beta cell stimulation results in an improvement in adiponectin and proinsulin intact levels and reduces intima media thickness in patients with type 2 diabetes. Horm Metab Res 2005 Aug;37(8):521-7.	4	Duplicate see record number 9140.
9237	Yamanouchi T, Sakai T, Igarashi K, Ichiyanagi K, Watanabe H, Kawasaki T. Comparison of metabolic effects of pioglitazone, metformin, and glimepiride over 1 year in Japanese patients with newly diagnosed Type 2 diabetes. Diabet Med. 2005 Aug;22(8):980-5.	5	Missing data for body weight.[b]
9265	Pfützner A, Marx N, Lübben G, Langenfeld M, Walcher D, Konrad T et al. Improvement of cardiovascular risk markers by pioglitazone is independent from glycemic control: results from the pioneer study. J Am Coll Cardiol 2005 Jun 21;45(12):1925-31.	1	Other OAD as baseline medication allowed.
9322	Langenfeld MR, Forst T, Hohberg C, Kann P, Lübben G, Konrad T. Pioglitazone decreases carotid intima-media thickness independently of glycemic control in patients with type 2 diabetes mellitus: results from a controlled randomized study. Circulation 2005 May 17;111(19):2525-31.	1	Other OAD as baseline medication allowed.
9361	Göke B; German Pioglitazone Study Group. Improved glycemic control and lipid profile in a randomized study of pioglitazone compared with acarbose in patients with type 2 diabetes mellitus. Treat Endocrinol 2002;1(5):329-36.	5	Hypoglycemia not reported.
9478	Bengel FM, Abletshauser C, Neverve J, Schnell O, Nekolla SG, Standl E et al. Effects of nateglinide on myocardial microvascular reactivity in Type 2 diabetes mellitus--a randomized study using positron emission tomography. Diabet Med 2005 Feb;22(2):158-63.	5	Body weight and hypoglycemia not reported.
9485	Ramachandran A, Snehalatha C, Salini J,Vijay V. Use of Glimepiride and Insulin Sensitizers in the Treatment of Type 2 Diabetes - A Study in Indians. J Assoc Physicians India 2004 Jun;52:459-63.	5	Hypoglycemia not reported.
9510	Gonzalez-Ortiz M,Martinez-Abundis E. Efficacy and safety of glimepiride plus metformin in a single presentation, as combined therapy, in patients with type 2 diabetes mellitus and secondary failure to glibenclamide, as monotherapy. Rev Invest Clin 2004;56(3):327-33.	5	Body weight not reported.
9625	Derosa G, Franzetti I, Gadaleta G, Ciccarelli L, Fogari R. Metabolic variations with oral antidiabetic drugs in patients with Type 2 diabetes: comparison between glimepiride and metformin. Diabetes Nutr Metab 2004;17(3):143-50.	5	Missing data for body weight.[b]

9654	Schernthaner G, Grimaldi A, Di Mario U, Drzewoski J, Kempler P, Kvapil M et al. GUIDE study: double-blind comparison of once-daily gliclazide MR and glimepiride in type 2 diabetic patients. Eur J Clin Invest 2004 Aug;34(8):535-42.	1	Two drugs from within the same drug class.
9715	Rosenstock J, Hassman DR, Madder RD, Brazinsky SA, Farrell J, Khutoryansky N et al. Repaglinide versus nateglinide monotherapy: a randomized, multicenter study. Diabetes Care 2004 Jun;27(6):1265-70.	1	Two drugs from within the same drug class.
9794	Yanagawa T, Araki A, Sasamoto K, Shirabe S, Yamanouchi T. Effect of antidiabetic medications on microalbuminuria in patients with type 2 diabetes. Metabolism 2004 Mar;53(3):353-7.	5	Body weight and hypoglycemia not reported.
9853	Jovanovic L, Hassman DR, Gooch B, Jain R, Greco S, Khutoryansky N et al. Treatment of type 2 diabetes with a combination regimen of repaglinide plus pioglitazone. Diabetes Res Clin Pract 2004 Feb;63(2):127-34.	5	Missing data for body weight.[b]
9881	Lawrence JM, Reid J, Taylor GJ, Stirling C, Reckless JP. Favorable effects of pioglitazone and metformin compared with gliclazide on lipoprotein subfractions in overweight patients with early type 2 diabetes. Diabetes Care 2004 Jan;27(1):41-6.	5	Missing data for body weight[b]; hypoglycemia not reported.
9884	van de Laar FA, Lucassen PL, Kemp J, van de Lisdonk EH, van Weel C, Rutten GE. Is acarbose equivalent to tolbutamide as first treatment for newly diagnosed type 2 diabetes in general practice? A randomised controlled trial. Diabetes Res Clin Pract 2004 Jan;63(1):57-65.	4	Duplicate see record number 3884.
10018	Garber AJ, Donovan DS Jr, Dandona P, Bruce S, Park JS. Efficacy of glyburide/metformin tablets compared with initial monotherapy in type 2 diabetes. J Clin Endocrinol Metab 2003 Aug;88(8):3598-604.	5	Missing data for HbA1c and body weight.[b]
10091	Bech P, Moses R, Gomis R. The effect of prandial glucose regulation with repaglinide on treatment satisfaction, wellbeing and health status in patients with pharmacotherapy naïve Type 2 diabetes: a placebo-controlled, multicentre study. Qual Life Res 2003 Jun;12(4):413-25.	4	Duplicate see record number 4427.
10114	Derosa G, Mugellini A, Ciccarelli L, Crescenzi G, Fogari R. Comparison of glycaemic control and cardiovascular risk profile in patients with type 2 diabetes during treatment with either repaglinide or metformin. Diabetes Res Clin Pract 2003 Jun;60(3):161-9.	5	Missing data for HbA1c[b]; body weight not reported.
10122	Derosa G, Mugellini A, Ciccarelli L, Crescenzi G, Fogari R. Comparison between repaglinide and glimepiride in patients with type 2 diabetes mellitus: a one-year, randomized, double-blind assessment of metabolic parameters and cardiovascular risk factors. Clin Ther 2003 Feb;25(2):472-84.	5	Hypoglycemia not reported.

10179	Luis Bautista J, Bugos C, Dirnberger G, Atherton T. Efficacy and safety profile of glimepiride in Mexican American Patients with type 2 diabetes mellitus: a randomized, placebo-controlled study. Clin Ther 2003 Jan;25(1):194-209.	4	Duplicate see record number 1072.
10263	Cefalu WT, Schneider DJ, Carlson HE, Migdal P, Gan Lim L, Izon MP et al. Effect of combination glipizide GITS/metformin on fibrinolytic and metabolic parameters in poorly controlled type 2 diabetic subjects. Diabetes Care 2002 Dec;25(12):2123-8.	3	Follow-up only 6 weeks.
10384	Hermann LS, Lindberg G, Lindblad U, Melander A. Efficacy, effectiveness and safety of sulphonylurea-metformin combination therapy in patients with type 2 diabetes. Diabetes Obes Metab 2002 Sep;4(5):296-304.	2	Systematic review.
10386	Rosenbaum P, Peres RB, Zanella MT, Ferreira SR. Improved glycemic control by acarbose therapy in hypertensive diabetic patients: effects on blood pressure and hormonal parameters. Braz J Med Biol Res 2002 Aug;35(8):877-84.	1	Other OAD as baseline medication allowed.
10498	Takami K, Takeda N, Nakashima K, Takami R, Hayashi M, Ozeki S et al. Effects of dietary treatment alone or diet with voglibose or glyburide on abdominal adipose tissue and metabolic abnormalities in patients with newly diagnosed type 2 diabetes. Diabetes Care. 2002 Apr;25(4):658-62.	5	Hypoglycemia not reported.
10595	Uehara MH, Kohlmann NE, Zanella MT, Ferreira SR. Metabolic and haemodynamic effects of metformin in patients with type 2 diabetes mellitus and hypertension. Diabetes Obes Metab 2001 Oct;3(5):319-25.	5	Hypoglycemia not reported.
10766	Amador-Licona N, Guízar-Mendoza J, Vargas E, Sánchez-Camargo G, Zamora-Mata L. The short-term effect of a switch from glibenclamide to metformin on blood pressure and microalbuminuria in patients with type 2 diabetes mellitus. Arch Med Res 2000 Nov-Dec;31(6):571-5.	5	Hypoglycemia not reported.
10783	Ebeling P, Teppo AM, Koistinen HA, Koivisto VA. Concentration of the complement activation product, acylation-stimulating protein, is related to C-reactive protein in patients with type 2 diabetes. Metabolism 2001 Mar;50(3):283-7.	5	Body weight and hypoglycemia not reported.
10892	Meneilly GS, Ryan EA, Radziuk J, Lau DC, Yale JF, Morais J et al. Effect of acarbose on insulin sensitivity in elderly patients with diabetes. Diabetes Care 2000 Aug;23(8):1162-7.	4	Duplicate see record number 4499.
10968	Kitabchi AE, Kaminska E, Fisher JN, Sherman A, Pitts K, Bush A et al. Comparative efficacy and potency of long-term therapy with glipizide or glyburide in patients with type 2 diabetes mellitus. Am J Med Sci. 2000 Mar;319(3):143-8.	1	Two drugs from within the same drug class.
10995	Lindström J, Tuomilehto J, Spengler M. Acarbose treatment does not change the habitual diet of patients with Type 2 diabetes mellitus. The Finnish Acargbos Study Group. Diabet Med 2000 Jan;17(1):20-5.	4	Duplicate see record number 4539.

11016	Jovanovic L, Dailey G 3rd, Huang WC, Strange P, Goldstein BJ. Repaglinide in type 2 diabetes: a 24-week, fixed-dose efficacy and safety study. J Clin Pharmacol 2000 Jan;40(1):49-57.	5	Missing data for HbA1c[b]; body weight not reported.
11100	Tessier D, Maheux P, Khalil A, Fülöp T. Effects of gliclazide versus metformin on the clinical profile and lipid peroxidation markers in type 2 diabetes. Metabolism. 1999 Jul;48(7):897-903.	5	Missing data for hypoglycemia.[b]
11115	Landgraf R, Bilo HJ, Müller PG. A comparison of repaglinide and glibenclamide in the treatment of type 2 diabetic patients previously treated with sulphonylureas. Eur J Clin Pharmacol 1999 May;55(3):165-71.	5	Missing data for body weight.[b]
11172	Wolffenbuttel BH, Landgraf R. A 1-year multicenter randomized double-blind comparison of repaglinide and glyburide for the treatment of type 2 diabetes. Dutch and German Repaglinide Study Group. Diabetes Care 1999 Mar;22(3):463-7.	4	Duplicate see record number 4632.
11261	Wolever TM, Chiasson JL, Josse RG, Hunt JA, Palmason C, Rodger NW et al. No relationship between carbohydrate intake and effect of acarbose on HbA1c or gastrointestinal symptoms in type 2 diabetic subjects consuming 30-60% of energy from carbohydrate. Diabetes Care 1998 Oct;21(10):1612-8.	4	Duplicate see record number 4998.
11313	Chan JC, Chan KW, Ho LL, Fuh MM, Horn LC, Sheaves R et al. An Asian multicenter clinical trial to assess the efficacy and tolerability of acarbose compared with placebo in type 2 diabetic patients previously treated with diet. Asian Acarbose Study Group. Diabetes Care 1998 Jul;21(7):1058-61.	4	Duplicate see record number 4717.
11491	Simonson DC, Kourides IA, Feinglos M, Shamoon H, Fischette CT. Efficacy, safety, and dose-response characteristics of glipizide gastrointestinal therapeutic system on glycemic control and insulin secretion in NIDDM. Results of two multicenter, randomized, placebo-controlled clinical trials. The Glipizide Gastrointestinal Therapeutic System Study Group. Diabetes Care 1997 Apr;20(4):597-606.	4	Duplicate see record number 4829.
11558	Dills DG, Schneider J. Clinical evaluation of glimepiride versus glyburide in NIDDM in a double-blind comparative study. Glimepiride/Glyburide Research Group. Horm Metab Res 1996 Sep;28(9):426-9.	4	Duplicate see record number 4872.
11591	Calle-Pascual A, Garcia-Honduvilla J, Martin-Alvarez PJ, Calle JR, Maranes JP. Influence of 16-week monotherapy with acarbose on cardiovascular risk factors in obese subjects with non-insulin-dependent diabetes mellitus: a controlled, double-blind comparison study with placebo. Diabetes Metab 1996;22(3):201-2.	5	Hypoglycemia not reported.
11660	Rodger NW, Chiasson JL, Josse RG, Hunt JA, Palmason C, Ross CA et al. Clinical experience with acarbose: results of a Canadian multicentre study. Clin Invest Med 1995;18(4):318-24.	4	Duplicate see record number 4923.

11665	Tsumura K. Clinical evaluation of glimepiride (HOE490) in NIDDM, including a double blind comparative study versus gliclazide. Diabetes Res Clin Pract 1995 Aug;28 Suppl:S147-9.	4	Duplicate see record number 1144.
11699	Wolever TM, Radmard R, Chiasson JL, Hunt JA, Josse RG, Palmason C et al. One-year acarbose treatment raises fasting serum acetate in diabetic patients. Diabet Med 1995;12(2):164-72.	4	Duplicate see record number 4998.
11711	Harrower AD. Comparison of efficacy, secondary failure rate, and complications of sulfonylureas. J Diabetes Complications 1994 Oct-Dec;8(4):201-3.	2	Review (single studies were assessed; see 11935, 11711_1 and 11711_2).
11754	de Leiva A, Pinon F, Tebar J, Escobar-Jimenez F, de la Calle H, Herrera-Pombo JF et al. Clinical efficacy and tolerance to acarbose in the treatment of non-insulin-dependent diabetic patients. MedClin-Barcelona 1993; 100(10):368-71.	4	Duplicate see record number 5041.
11797	Spengler M, Hansel G, Boehme K. Efficacy of 6 months monotherapy with glucosidase inhibitor Acarbose versus sulphonylurea glibenclamide on metabolic control of dietary treated type II diabetics (NIDDM). Horm Metab Res Suppl 1992;	6	Not available.
11833	Harrower AD. Efficacy of gliclazide in comparison with other sulphonylureas in the treatment of NIDDM. Diabetes Res Clin Pract 1991;14 Suppl 2:S65-7.	2	Review (single studies were assessed).
11931	U.K. prospective diabetes study. II. Reduction in HbA1c with basal insulin supplement, sulfonylurea, or biguanide therapy in maturity-onset diabetes. A multicenter study. Diabetes 1984;34(8):793-798	5	Body weight and hypoglycemia not reported.
11935	Harrower AD. Comparison of diabetic control in type 2 (non-insulin dependent) diabetic patients treated with different sulphonylureas. Curr Med Res Opin 1985;9(10):676-80.	1	Five drugs from within the same drug class.
11950	Borthwick LJ, Wilson S. Diabetic control with gliquidone--a short acting sulphonylurea. Eur J Clin Pharmacol 1984;26(4):475-9.	3	Cross over periods only 10 weeks.
11966	Toyota T. Sulfonylurea drug--a new sulfonylurea drug for type 2 diabetes. Nihon Rinsho 1999 Mar;57(3):695-701.	5	Hypoglycemia not reported.
11967	Kaneko T, Sakamoto N. Clinical efficacy of glimepiride. Nihon Rinsho 1997 Nov;55 Suppl:152-7.	2	Review.
11978	Kaneko T, Sakamoto N, Nakagawa S, Goto Y, Hirata Y, Akanuma Y et al. Clinical evaluation of Glimepiride in non-insulin dependent diabetes mellitus – a double blind comparative study versus Gliclazide. Nihon Rinsho 1993 Sept;9(5): 167-80.	1	Two drugs from within the same drug class.
11981	Dalzell GW, Hadden DR, Atkinson B, Kennedy L, Weaver JA. A randomized trial of Tolbutamide and Metformin for persistent severe hpyerglycemia in non-insulin dependent diabetes mellitus (NIDDM). Ir J Med Sci 1986;9:341-2.	5	HbA1c, body weight and hypoglycemia not reported.

11711_1	Harrower AD. Comparison of efficacy, secondary failure rate, and complications of sulfonylureas. J Diabetes Complications 1994 Oct-Dec;8(4):201-3.	1	Three drugs from within the same drug class.
11711_2	Harrower AD. Comparison of efficacy, secondary failure rate, and complications of sulfonylureas. J Diabetes Complications 1994 Oct-Dec;8(4):201-3.	2	Cross-over without randomization, pre-post comparison; treatment follow-up interaction could bias the outcomes.
U1[c]	Naka KK, Papathanassiou K, Bechlioulis A, Pappas K, Kazakos N, Kanioglou C et al. Effects of pioglitazone and metformin on vascular endothelial function in patients with type 2 diabetes treated with sulfonylureas. Diab Vasc Dis Res 2012 Jan;9(1):52-8.	1	Other OAD as baseline medication allowed.
U2[c]	Uchida T, Kawai J, FujitaniY, Kawamori R, Watada H, Hirose K. Efficacy and adverse effects of low-dose Nateglinide in early type 2 diabetes: comparison with Acarbose in a cross-over study. Diabetol Int 2010;1:35-41.	2	Cross-over study: no wash out after 1st sequence; carry over effect could bias the outcomes.

[a] 1 Non eligible intervention.
2 Inadequate study design.
3 Study duration inadequate.
4 Duplicate data.
5 Data required for analysis not reported/ not provided upon request from authors.
6 Full text publication not available.

[b] Missing data was requested from first author without success.

[c] Excluded studies from updated literature research.

A.3 Table 12

Table 12a: Trial and patients characteristics of included studies

No.	Author, y	Follow-up[a], wk	Previous therapy[b]	Included study arms[c]	Rand. patients, n	Mean age (SD), y	Men, %	Mean duration of existing T2DM (SD), y	Mean bl HbA1c (SD), %	Mean bl body weight (SD), kg
1	List, 2009 [141]	12	1	Placebo	54	53 (11)	55.6	NR	7.9 (0.9)	89 (18)
				Metformin (1,500mg;q.d.)	56	54 (9)	48.2	NR	7.6 (0.8)	88 (20)
2	Defronzo, 1995 [144]	29	1	Placebo	146	53 (12.1)	42.5	6 (7.2)	8.2 (2.4)	92.2 (14.5)
				Metformin (titration: 500mg;q.d./ maintenance: 2,550mg;t.i.d.)	143	53 (11.9)	43.4	6 (6)	8.4 (1.2)	94.4 (13.2)
3	Chiasson, 2001 [145]	36	2	Placebo	83	57.7 (9.9)	67.5	5.1 (4.9)	8.1 (0.7)	88.6 (14.1)
				Metformin (1,500mg;t.i.d.)	83	57.9 (8.6)	73.5	7.5 (7.4)	8.2 (0.9)	89 (17.8)
				Miglitol (0-4wk: 25mg;q.d./ 3-12wk: 50mg;t.id./ 12-36wk: 100mg;t.i.d.)	82	57.3 (9)	78	5.2 (4.7)	8.2 (0.9)	91 (15.5)
4	Hoffmann, 1997 [146]	24	1	Placebo	32	60.2 (8.6)	37.5	3.6 (2.8)	9.4 (0.9)	74.9 (9.7)
				Metformin (1,700mg;b.i.d.)	32	55.9 (7.8)	43.8	2.1 (1.5)	9.7 (0.9)	79 (8.8)
				Acarbose (300mg;t.i.d)	32	58.9 (9.4)	18.8	3.1 (2.3)	9.6 (0.9)	73.9 (10.3)
5	Haak, 2012 [147]	24	3	Placebo	72	55.7 (11)	50	NR	8.7 (1)	76.8 (17.5)
				Metformin (500 and 1,000mg;b.i.d.)[d]	291	54.1 (10.5)	55	NR	8.6 (0.9)	80 (18.4)
				Linagliptin (5mg;q.d.)	142	56.2 (10.8)	56.3	NR	8.7 (1)	79.1 (17.3)

6[e]	Simonson, 1997 [148]	12	3	Placebo	69	60.2 (NR)	76.8	7.5 (NR)	8.3 (1.7)	87 (NR)
				Glipizide (5, 10, 15, 20, 40, 60mg;q.d.)[d]	278	58.2 (NR)	64	7.2 (NR)	8.6 (1.5)	85.3 (NR)
7	Bautista, 2003 [149]	14	2	Placebo	22	50.7 (10)	50	5.7 (8.4)	10.5 (2.2)	76.3 (18.5)
				Glimepiride (level 1: 1mg;q.d./level 2: 2mg;q.d./ level 3: 4mg;q.d.)	48	48.4 (11.7)	56.3	4.2 (5.8)	10 (1.8)	83.3 (17)
8	Madsbad, 2004 [150]	12	3	Placebo	29	57 (9.4)	69	3.4 (2.9)	7.4 (1.2)	NR
				Glimepiride (1-4mg; q.d./ mean received dose:2.7mg)	27	57 (9.2)	59.3	3.8 (3.4)	7.8 (0.9)	NR
9	Vray, 1995 [151]	12	3	Placebo	56	56.8 (6.7)	35.7	3.9 (4.5)	10 (3)	NR
				Glibenclamide (7.5mg;t.i.d.)	56	55.8 (7.5)	35.7	2.4 (3)	9.6 (3)	NR
10[f]	Connif, 1995 [152]	24	3	Placebo	62	56.3 (NR)	51.6	5.5 (NR)	7.1 (NR)	85.8 (NR)
				Tolbutamide (750mg;t.i.d./ forced titration until 1000mg;d)	66	55.4 (NR)	56.1	5.6 (NR)	7 (NR)	84.8 (NR)
				Acarbose (600mg;t.i.d.)	67	56.2 (NR)	38.8	5.1 (NR)	6.9 (NR)	81.6 (NR)
11	Scott, 2007 [153]	12	2	Placebo	125	55.3 (9.7)	62.4	4.8 (4.7)	7.9 (1)	NR
				Glipizide (5mg;q.d.)	123	54.7 (10.7)	56.9	4.7 (4.2)	7.9 (1)	NR
				Sitagliptin (5, 12,5, 25 and 50mg;b.i.d.)[d]	495	55.5 (9.3)	51.9	4.6 (4.6)	7.9 (1)	NR
12	Henry, 2009 [154]	16	3	Placebo	55	56.5 (8.9)	41.8	NR	8.1 (0.8)	NR
				Pioglitazone (45mg;q.d.)	57	54.4 (8.8)	42.1	NR	8 (0.8)	NR

13	Rosenblatt, 2001 [155]	16	3	Placebo	96	55.2 (10)	56.3	NR	10.4 (1.7)	87.2 (18.4)
				Pioglitazone (30mg;q.d.)	101	53.8 (10)	50.5	NR	10.7 (1.8)	89.8 (18)
14	Aronoff, 2000 [156]	26	3	Placebo	79	NR	NR	NR	10.4 (2)	90.4 (13.1)
				Pioglitazone (7,5, 15, 30, 45mg;q.d.)[d]	329	NR	NR	NR	10.2 (2)	91.4 (14.7)
15	Gonzalez, 2008 [157]	12	1	Placebo	54	57.2 (10.7)	63	NR	7.1 (0.7)	77.1 (11)
				Nateglinide (480mg;t.i.d.)	55	59.9 (10.6)	56.4	NR	7.2 (0.6)	77.6 (11.5)
16	Derosa, 2011 [158]	28	1	Placebo	92	56.7 (7)	48.9	0.4 (0.1)	6.7 (0.5)	76.5 (8.2)
				Acarbose (150mg;t.i.d./ uptitration after 1 month to 300mg;t.i.d.)	96	56.7 (7)	49	0.4 (0.1)	6.8 (0.6)	75.1 (7.7)
17	Kawamori, 2012 [159]	12	3	Placebo	80	59.7 (8.9)	71.3	NR	8 (0.7)	65.3 (11.6)
				Linagliptin (5 and 10mg;q.d.)[d]	319	60.8 (9.7)	69.9	NR	8 (0.7)	65.4 (12.2)
18	Del Prato, 2011 [160]	24	3	Placebo	167	54.4 (10.3)	47.3	NR	8 (0.9)	79.2 (16)
				Linagliptin (5mg;q.d.)	336	56.4 (10.1)	48.8	NR	8 (0.9)	78.5 (16.7)
19	Kikuchi, 2009 [161]	12	1	Placebo	72	60.4 (8.1)	63.9	7.1 (5.5)	7.4 (0.8)	63.8 (10.1)
				Vildagliptin (10, 25,50mg;b.i.d.)[d]	219	58.5 (8.5)	68	4.6 (4.3)	7.4 (0.8)	63.5 (9.8)
20	Mohan, 2009 [162]	18	2	Placebo	178	50.9 (9.3)	59.6	1.9 (1.6)	8.8 (1.1)	66.6 (11.4)
				Sitagliptin (100mg;q.d.)	352	50.9 (9.3)	56.8	2.1 (1.7)	8.7 (1)	66.8 (10.2)
21	Aschner, 2006 [163]	24	2	Placebo	253	54.3 (10.1)	51.4	4.6 (4.7)	8 (0.8)	85 (18.1)
				Sitagliptin (100 and 200mg;q.d.)[d]	488	54.2 (9.8)	51.8	4.3 (4.8)	8.1 (0.9)	84.3 (18.8)

22	Rosen-stock, 2008 [164]	12	2	Placebo	67	55.2 (9.8)	62.7	NR	8 (1)	93.1 (19.2)
				Saxagliptin (2.5, 5, 10, 20 and 40mg;q.d.)[d]	271	53.7 (9.7)	57.2	NR	7.9 (1)	89 (16.6)
23	Dejager, 2007 [165]	24	2	Placebo	94	52.2 (11.2)	47.9	1.6 (2.5)	8.4 (0.8)	NR
				Vildagliptin (50 and 100mg;q.d./ 50mg;b.i.d.)[d]	286	54 (11.1)	46.9	2.2 (3.7)	8.4 (0.8)	NR
24	Pi-Sunyer, 2007 [166]	24	2	Placebo	92	52 (12)	54.3	2.5 (3.7)	8.5 (0.8)	93 (23.2)
				Vildagliptin (50 and 100mg;q.d./ 50mg;b.i.d.)[d]	262	51 (11)	55.3	2.1 (2.9)	8.4 (0.9)	90.4 (20.2)
25	Barzilai, 2011 [167]	24	2	Placebo	104	72.1 (6)	47.1	7 (7.5)	7.8 (0.7)	85.8 (16.5)
				Sitagliptin (50 or 100mg;q.d./lower dose for patients with creatinin clearance <50 mL/min)	102	71.6 (6.1)	47.1	7.2 (7.3)	7.8 (0.8)	85.6 (16.6)
26	Iwamoto, 2010 [168]	12	2	Placebo	73	60.2 (8)	68.5	6.4 (5.5)	7.7 (0.9)	NR
				Sitagliptin (25, 50, 100 and 200mg;q.d.)[d]	290	59.6 (8.6)	60	5.2 (5.3)	7.6 (0.8)	NR
27	Nonaka, 2008 [169]	12	2	Placebo	76	55 (8)	65.8	4.1 (4.6)	7.7 (0.9)	NR
				Sitagliptin (100mg;q.d.)	76	55.6 (8.6)	59.2	4 (4.1)	7.5 (0.9)	NR
28	Raz, 2006 [170]	18	2	Placebo	110	55.5 (10.1)	62.7	4.7 (5)	8.1 (0.9)	92.8 (18.8)
				Sitagliptin (100 and 200mg;q.d.)[d]	411	55 (9.6)	52.1	4.5 (4.1)	8.1 (0.9)	89.6 (19.2)
29	Ristic, 2005 [171]	12	2	Placebo	58	54.6 (10.6)	56.9	2.3 (3)	7.8 (0.8)	92 (15.6)
				Vildagliptin (25, 50 and 100mg;q.d./ 25mg;b.i.d.)[d]	221	56.5 (10.3)	53.8	3 (4.2)	7.7 (0.8)	90 (17.1)

30	Scher-baum, 2008 [172]	52	2	Placebo	150	62.8 (11)	59.3	2.7 (3.2)	6.8 (0.4)	85 (1.4)
				Vildagliptin (50mg;q.d.)	156	63.3 (10.2)	59.6	2.5 (2.9)	6.7 (0.4)	86 (1.2)
31	Scher-baum, 2008 [173]	108 (52/56)	2	Placebo	63	63.2 (10)	58.7	2.5 (2.6)	6.7 (0.4)	NR
				Vildagliptin (50mg;q.d.)	68	63.1 (9.6)	60.3	2.1 (2.1)	6.6 (0.4)	NR
32	Defronzo, 1995 [144]	29	3	Metformin (titration: 500mg;q.d./ maintenance:2,500mg; five times a day	210	55 (14.5)	45.7	8.4 (5.8)	8.9 (1.4)	92.6 (14.5)
				Glibenclamide (20mg; four times a day)	209	56 (14.5)	49.3	8.7 (5.8)	8.5 (1.4)	92.6 (14.5)
33	Charpen-tier, 2001 [174]	20	3	Metformin (2,550mg;t.i.d.)	75	56.7 (NR)	60	7 (NR)	6.8 (1.2)	82.2 (NR)
				Glimepiride (1-6mg; q.d.)	150	55.4 (NR)	58	5.3 (NR)	6.5 (1.1)	81 (NR)
34	Tosi, 2003 [175]	23	3	Metformin (500-3,000mg;nr)	19	NR	NR	NR	7.7 (0.9)	73 (10.7)
				Glibenclamide (5-15mg;nr)	20	NR	NR	NR	7.9 (1)	71.7 (8.3)
35	Hermann, 1994 [176]	24	2	Metformin (1,000-3,000mg;nr)	38	NR	NR	NR	6.9 (1.3)	78.6 (12.6)
				Glibenclamide (3.5-10.5mg;nr)	34	NR	NR	NR	6.7 (1.3)	86.2 (14.4)
36	Goldstein, 2003 [177]	18	3	Metformin (500mg;q.d./uptitra-tion to 2,000mg; four times a day)	76	56.6 (9.7)	61.8	7.3 (4.9)	8.7 (1.2)	93.8 (17)
				Glipizide (30mg;b.i.d.)	84	57.4 (9.2)	64.3	6.5 (4.4)	8.9 (1.1)	89.9 (17.3)

37	Russell-Jones, 2012 [178]	26	2	Metformin (up to 2,500mg; nr)	246	54 (11)	62.6	2.6 (3.6)	8.6 (1.2)	85.9 (19.6)
				Pioglitazone (up to 45mg;nr)	163	55 (11)	59.5	2.7 (3.7)	8.5 (1.2)	86.1 (17.8)
				Sitagliptin (100mg;q.d.)	163	52 (11)	57.7	2.7 (3.7)	8.5 (1.3)	88.7 (18.7)
38	Moses, 1999 [179]	20	3	Metformin (1,000-3,000mg;nr)	27	57.8 (9.5)	63	8 (6.2)	8.6 (1.1)	NR
				Repaglinide (1.5-12mg;t.i.d.)	28	60.3 (7.7)	53.6	7 (5.2)	8.6 (1.3)	NR
39	Lund, 2007 [180]	16	3	Metformin (500mg;q.d./ uptitra-tion to2,000mg;b.i.d.)	48	59.5 (9.3)	77.1	3 (NR)	8.1 (1.1)	74.5 (9.8)
				Repaglinide (1mg;q.d./ uptitration to 6mg;t.i.d.)	48	63.3 (8.9)	75	5 (NR)	8.1 (1.1)	74.5 (9.8)
40	Bosi, 2009 [181]	24	2	Metformin (500mg;q.d./ uptitra-tion to 2,000mg; four times a day)	294	52.4 (10.7)	58.2	2.2 (3.3)	8.6 (0.9)	88.4 (17.4)
				Vildagliptin (100mg;b.i.d.)	300	53.5 (11)	60	2.2 (3.3)	8.7 (1)	87.8 (17.4)
41	Aschner, 2010 [182]	24	2	Metformin (500mg;q.d./ uptitration to 2,000mg;b.i.d.)	522	55.7 (10.3)	37.2	2.1 (3.5)	7.2 (0.7)	84.6 (17.2)
				Sitagliptin (100mg;q.d.)	528	56.3 (10.7)	41.1	2.6 (3.9)	7.2 (0.7)	84.9 (17.7)

42	Schweizer, 2009 [183]	24	2	Metformin (500mg;q.d./ uptitration to 1,500mg;t.i.d.)	166	70.2 (5.1)	53	3 (4.7)	7.7 (0.6)	NR
				Vildagliptin (100mg;q.d.)	169	71.6 (5.2)	44.4	2.9 (4.2)	7.8 (0.6)	NR
43	Jadzinsky, 2009 [184]	24	2	Metformin (500mg;q.d./uptitra-tion to 2,000mg;four times a day)	328	51.8 (10.7)	49.7	1.7 (3.1)	9.4 (1.3)	82.8 (17.5)
				Saxagliptin (10mg;q.d.)	335	52.1 (10.2)	49.3	1.7 (2.8)	9.6 (1.3)	83.1 (16.9)
44[g]	Williams-Herman, 2009 [185]	54 (24/30)	2	Metformin (1,000 and 2,000mg;b.i.d.)[d]	259	54 (9.7)	46.3	4.1 (3.9)	8.6 (0.9)	NR
				Sitagliptin (100mg;q.d.)	106	53.5 (9.1)	51.9	3.9 (4.6)	8.7 (1)	NR
45[g]	Williams-Herman, 2010 [186]	104 (24/30/ 50)	2	Metformin (1,000 and 2,000mg;b.i.d.)[d]	153	55 (9.5)	45.1	3.9 (3.9)	8.5 (0.8)	NR
				Sitagliptin (100mg;q.d.)	52	54.1 (9.1)	57.7	3.7 (4.9)	8.5 (0.9)	NR
46	Schweizer, 2007 [187]	52	2	Metformin (2,000m;b.i.d)	254	53.6 (10.2)	57.5	1.03 (NR)	8.7 (1.1)	92.9 (1.2)
				Vildagliptin (100mg; q.d.)	526	52.8 (11.7)	52.9	1.05 (NR)	8.7 (1.1)	91.4 (0.9)
47	Goke, 2008 [188]	104 (52/52)	2	Metformin (2,000mg;b.i.d.)	158	54 (11)	NR	2.4 (3.4)	8.8 (0.1)	95.7 (1.6)
				Vildagliptin (100mg;q.d.)	305	54 (11)	NR	2.4 (3.4)	8.4 (0.1)	93.1 (1.3)

48	Perriello, 2006 [189]	52	3	Gliclazide (160-320mg;b.i.d./ mean dose: 184mg)	137	59 (7)	64.2	8.5 (4.1)	8.7 (0.9)	78.8 (10.7)
				Pioglitazone (30-45mg;q.d./ mean dose: 40mg)	146	58 (8)	66.4	9.8 (5.4)	8.8 (0.9)	81.1 (12)
49	Jain, 2006 [190]	52	2	Glibenclamide (5-15mg;nr/ range of means during maintenance phase (wk 16-52):9.9-10.5mg)	251	52.1 (12.4)	56.2	0.8 (1.3)	9.2 (1.3)	94.3 (20)
				Pioglitazone (15-45mg;nr/ range of means during maintenance phase (wk 16-52):34.9-37.6mg)	251	52.1 (11.3)	53	0.8 (1.1)	9.2 (1.2)	93.9 (19.7)
50	Tan, 2004 [191]	52	3	Glibenclamide (1.75-10.5mg;nr)	109	57.9 (9.2)	73.4	5.2 (4.7)	8.5 (0.8)	89 (16)
				Pioglitazone (30-45mg;q.d.)	91	60 (8.5)	61.5	4.8 (4.7)	8.4 (0.7)	88.4 (17.5)
51	Tan, 2004 [192]	52	3	Glimepiride (2-8mg;q.d./ mean dose: 6mg)	123	55.7 (9.3)	52.8	6.8 (6.9)	8.5 (0.9)	74.5 (10.8)
				Pioglitazone (15-45mg;q.d./ mean dose: 27mg)	121	55.1 (8)	44.6	6.5 (6.6)	8.5 (1)	74.2 (10.5)

52	Madsbad, 2001 [193]	58	3	Glipizide (5-15mg;nr/ uptitration)	81	62 (8.8)	64.2	7 (4.9)	7.2 (1.4)	83.6 (14.5)
				Repaglinide (1.5-12mg;t.i.d./ uptitration)	175	60.2 (8.1)	61.1	8.1 (6)	7.3 (1.2)	82.9 (13.4)
53	Marbury, 1999 [194]	60	3	Glibenclamide (2.5-15mg;nr)	193	58.7 (9)	62.2	8.3 (6.8)	8.9 (1.6)	NR
				Repaglinide (1.5-12mg;t.i.d./uptitration)	383	58.3 (9.4)	63.2	7.2 (6.2)	8.7 (1.7)	NR
54	Wolffen-büttel, 1999 [195]	60	3	Glibenclamide (175-10.5mg;nr)	139	61 (9)	68.3	6 (NR)	7 (1.2)	81.3 (12.2)
				Repaglinide (1.5-12mg;t.i.d./uptitration)	286	61 (9)	62.2	6 (NR)	7.1 (1.4)	81.5 (13.4)
55	NA [196]	16	1	Gliclazide (80mg;q.d./ uptitration to 160mg;b.i.d.)	222	53.5 (9.9)	55.4	0.8 (1.8)	7.2 (NR)	68.4 (11.2)
				Repaglinide (1mg;b.i.d./ uptitration to 12mg;t.i.d.)	218	54.1 (10)	50.5	1 (2)	7.2 (NR)	68.7 (11.6)
56	Feinböck, 2003 [197]	26	3	Glimepiride (1-6mg;q.d.)	111	57.7 (10.2)	59.5	3 (3.6)	9.1 (1.9)	85 (12.8)
				Acarbose (150-600mg;t.i.d.)	108	57.1 (10.7)	53.7	3.6 (4.8)	9.4 (2)	83 (12.5)
57	Salman, 2001[h] [198]	24	1	Gliclazide (80-160mg;b.i.d./ 80% of patients got 80mg b.i.d.	30	56.1 (8.7)	53.3	4.7 (5.6)	8.7 (0.6)	NR
				Acarbose (50mg;q.d./ uptitration to 300mg;t.i.d./ 96% of patients got 300mg;t.i.d.)	27	52.6 (9.1)	63	4.2 (3.4)	8.9 (0.7)	NR

58	Ferreira , 2011 [199]	54	2	Glipizide (2.5mg;q.d./ uptitration to 10mg; b.i.d.)	212	64.2 (9.4)	57.5	NR	7.8 (0.7)	NR
				Sitgaliptin (severe CRI: 25mg;q.d./ moderate CRI: 50mg;q.d.)	211	64.2 (10.7)	62.1	NR	7.8 (0.7)	NR
59	Foley, 2009 [200]	104	2	Gliclazide (80-320mg;nr/ mean dose at endpoint:209mg)	546	54.3 (10.4)	52.7	1.9 (3.1)	8.7 (1.1)	84.3 (17.6)
				Vildagliptin (100mg;b.i.d.)	546	55.2 (10.6)	58.8	2.4 (4.3)	8.6 (1)	84.2 (16.3)
60	Rosen-stock, 2007 [201]	24	2	Pioglitazone (30mg;q.d.)	161	52.4 (10.3)	64	2.2 (3.3)	8.7 (1)	81 (NR)
				Vildagliptin (100mg;q.d.)	154	51.4 (10.8)	63.6	1.9 (3.1)	8.6 (1)	82 (NR)
61	Pan, 2009 [202]	12	3	Nateglinide (480mg;t.i.d.)	119	54.4 (9.4)	58.8	2.7 (3.7)	7.8 (1)	69.5 (10.2)
				Acarbose (100mg;q.d./ uptitra-tion to 300mg;t.i.d.)	118	55.1 (9)	66.9	2.6 (3.8)	7.7 (0.8)	71.2 (9.4)
62	Pan, 2008 [203]	24	2	Acarbose (uptitration to 300mg;t.i.d.)	220	51.9 (10.3)	63.2	1.3 (2.4)	8.6 (1)	72 (NR)
				Vildagliptin (100mg;b.i.d.)	441	51.8 (10.1)	60.1	1.2 (2.4)	8.6 (0.9)	72 (NR)

References for included studies are only indexed once in Table 12a.

[a] (core study/ plus extension).

[b] Coding of previous therapy: 1: therapy naive; 2: no oral antidiabetic drug at least 8 weeks before randomization; 3: else. For study No. 37 and 43 the exclusion criteria (patients were excluded if treated with any antihyperglycemic drug for > 7 days within 3 months of screening) come pretty close to category 2; thus they were included therein.

[c] (dose/d; frequency).

[d] Different doses of one drug were aggregated into one arm (see 3.1.3.4).

[e] For the two sibling studies from study no. 6 trial and patients characteristics are only reported for the aggregated Placebo and Glipizide arms.

[f] Baseline characteristics are only reported for the patients who were valid for efficacy analysis (on study medication > 35d).
[g] Study arms that were switched from placebo to Metformin in these extension studies were not included due to potential confounding of placebo given during core phase.
[h] Baseline characteristics are only reported for the patients who completed the study.
Y, year. Wk, week. D, day. Rand., randomized. SD, standard deviation. T2DM, diabetes mellitus type 2. Bl, baseline. HbA1c, glycated haemoglobin. KG, kilogram. NR, not reported. NA, not applicable. Q.d., once daily. B.i.d., twice daily. T.i.d., thrice daily. CRI, chronic renal insufficiency.

Table 12b: Outcome data for HbA1c, body weight and hypoglycemia of included studies

No.	Author, y	Included study arms	HbA1c, %			Body weight, kg			Hypoglycemia		
			n	mc	SD[a]	n	mc	SD[a]	r	n	Definition[b]
1	List, 2009	Placebo	54[c]	-0.18	0.73	54[c]	-1.07	2.61[d]	2	54	Not further specified; ae were summarized by preferred term (MedDRA, version 10).
		Metformin	56[c]	-0.73	0.75	56[c]	-1.5	2.46[d]	5	56	
2	Defronzo, 1995	Placebo	146[c]	0.4	1.21	146[c]	-1.1	2.42	0	146	Symptoms compatible with hypoglycemia without biochemical documentation.
		Metformin	143[c]	-1.4	1.2	143[c]	-0.6	3.59	3	143	
3	Chiasson, 2001	Placebo	82	0.38	1.09	82	-0.69	2.45	7	83	Not further specified.
		Metformin	81	-0.85	1.08	81	-0.79	2.97	8	83	
		Miglitol	80	0.02	0.89	80	-0.42	2.59	7	82	
4	Hoffmann, 1997	Placebo	32	0.4	1.16[f]	32	-0.2	2.36[g]	0	32	Not further specified.
		Metformin	31	-1	1.21[f]	31	-0.5	2.14[g]	0	32	
		Acarbose	31	-1.1	1.37[f]	31	-0.8	2.51[g]	0	32	
5	Haak, 2012	Placebo	65	0.1	0.81	43	-0.7	2.62	1	72	Not further specified (Hypoglycemic event intensity was graded according to the investigator's discretion).
		Metformin[e]	279	-0.85	1.21	239	-0.6	3.27	7	291	
		Linagliptin	135	-0.5	1.16	112	0.2	3.17	0	142	
6[h]	Simonson, 1997	Placebo	68	0.87	1.5	68[c]	-3.33	2.71	0[i]	69	Presence of symptoms of hypoglycemia alone, a blood glucose level < 3.3 mmol/l by home blood glucose monitoring regardless of the presence of symptoms or blood glucose < 4.4 mmol/l when tested in the clinic.
		Glipizide[e]	272	-0.8	1.49	272[c]	-0.22	2.56	11[i]	278	
7	Bautista, 2003	Placebo	18	-0.7	1.42[j]	18	-2.1	3.9[j]	0	22	Hypoglycemic episodes (i.e., self-monitored fasting blood glucose < 2.8 mmol/l) recorded in patient diaries.
		Glimepiride	42	-2.3	1.42[j]	42	2.3	3.9[j]	0	48	

8	Madsbad, 2004	Placebo	29	-0.12	0.65[j]	29	-0.16	1.83[j]	0	29	Symptomatic hypoglycemia; not further specified (for sensitivity analysis: minor hypoglycemia, defined as blood glucose < 2.8 mmol/l).
		Glimepiride	26	-0.86	0.65[j]	26	0.78	1.83[j]	5[k]	26	
9	Vray, 1995	Placebo	56	0.04	2.24	56	-0.66	2.25	0	56[c]	Symptomatic hypoglycemia; not further specified.
		Glibenclamide	52	-1.6	2.16	52	0.81	2.16	9	52[c]	
10	Connif, 1995	Placebo	62	-0.05	1.18	62	-1.38	3.6	4	72	Probably symptomatic hypoglycemia („most reported episodes were not verified by plasma glucose determinations“).
		Tolbutamide	66	-1.04	0.97	66	2.19	3.44	11	71	
		Acarbose	67	-0.6	1.31	67	-1.54	2.52	6	74	
11	Scott, 2007	Placebo	121	0.23	0.76	121[l]	-0.4	2.14[j]	3	125	Hypoglycemia was assessed by study site investigators through reviewing daily glucose logs and patient self-report of signs and symptoms of hypoglycemia.
		Glipizide	119	-0.76	0.78	119[l]	0.9	2.23	21	123	
		Sitagliptin[e]	485	-0.38	0.79	485	-0.18	2.18[j]	12	492	
12	Henry, 2009	Placebo	54	0.35	0.96	54	-0.85	2.94	0	55	Data not reported, obtained from author upon request; not further specified.
		Pioglitazone	57	-0.35	0.98	57	1.06	3.02	1	57	
13	Rosenblatt, 2001	Placebo	93	0.76	1.64	93[l]	-1.87	4.44[m]	0	55	Not further specified.
		Pioglitazone	100	-0.6	1.7	100[l]	1.35	3.31[m]	0	101	
14	Aronoff, 2000	Placebo	79	0.7	1.51	79	-1.3	3.2	0	79	Probable symptomatic hypoglycemia, not confirmed by blood glucose levels; all of the events were considered mild or moderate in intensity and none resulted in discontinuation.
		Pioglitazone[e]	320	-0.32	1.58	326	1.2	3.36	4	329	
15	Gonzalez, 2008	Placebo	54[c]	0.1	0.93[j]	54[c]	-0.3	1.55[j]	0	54	Hypoglycemia was defined as a capillary glycemia < 3.3 mmol/l with signs or symptoms of hypoglycemia.
		Nateglinide	55[c]	-0.5	0.93[j]	55[c]	-0.2	1.55[j]	0	55	
16	Derosa, 2011	Placebo	92[c]	-0.3	1.45[m]	92[c]	-5.1	24.63[m]	0	92	Not further specified.
		Acarbose	96[c]	-1.1	3.17[m]	96[c]	-5.9	29.12[m]	0	96	
17	Kawamori, 2012	Placebo	80	0.63	0.72	80	-0.39	1.43	0	80	Not further specified.
		Linagliptin[e]	316	-0.25	0.75	316	-0.05	1.51	0	319	
18[n]	Del Prato, 2011	Placebo	163	0.25	0.89	124	-0.29	2.12	1	167	Not further specified, events did not require third-party assistance.
		Linagliptin	333	-0.44	0.91	288	0	2.21	1	336	

19	Kikuchi, 2009	Placebo	71	0.28	0.57	71	-0.5	1.1	1	72	Grade 1 hypoglycemia: signs or symptoms suggestive of hypoglycemia that could be managed by the patient either with sucrose or in any other appropriate way.
		Vildagliptin[e]	218	-0.71	0.6	218	0.17	1.42	5	219	
20	Mohan, 2009	Placebo	169	0.3	1.33	169[l]	0	2.6	0	178	Not further specified.
		Sitagliptin	339	-0.7	0.94	339[l]	0.6	1.84	0	352	
21	Aschner, 2006	Placebo	244	0.18	0.96	244[l]	-1.1	3.12	2	253	Not further specified.
		Sitagliptin[e]	467	-0.69	0.96	467[l]	-0.15	3.05	5	488	
22	Rosen-stock, 2008	Placebo	62	-0.27	0.87	54	-1.03	2.87	1	67	Symptoms of hypoglycemia (such as confusion and dizziness).
		Saxagliptin[e]	254	-0.81	0.86	216	-0.48	2.47	17	271	
23	Dejager, 2007	Placebo	94	-0.3[o]	0.97	94	-1.4	3.88	0	157	Symptoms suggestive of low blood glucose confirmed by self-monitored blood glucose measurement < 3.1 mmol /l.
		Vildagliptin[e]	286	-0.83	0.98	286	-1.01	3.95	3	468	
24	Pi-Sunyer, 2007	Placebo	88	0	0.94	88	-1.4	3.75	0	92	Confirmed hypoglycemia was defined as symptoms suggestive of low blood glucose confirmed by self-monitored blood glucose (SMBG) measurement <3.1 mmol/L plasma glucose equivalent. Instances of SMBG measurement <3.1 mmol/L plasma glucose equivalent without accompanying symptoms were recorded as asymptomatic low blood glucose.
		Vildagliptin[e]	252	-0.67	0.92	252	-0.27	3.35	0	260	
25	Barzilai, 2011	Placebo	91	0.2	1.22	91[l]	-1.7	6.33	0	104	Not further specified.
		Sitagliptin	101	-0.5	1.28	101[l]	-1.1	6.41	0	102	
26	Iwamoto, 2010	Placebo	73	0.28	0.52	73[l]	-0.5	1.52	2	73	Symptomatic hypoglycemia not necessary confirmed by fingerstick blood glucose determination.
		Sitagliptin[e]	290	-0.63	0.54	290[l]	0.28	1.37	10	290	
27	Nonaka, 2008	Placebo	75	0.41	0.66	76	-0.7	1.33	0	76	Not further specified.
		Sitagliptin	75	-0.65	0.66	75	-0.1	1.55	0	75	
28	Raz, 2006	Placebo	103	0.12	0.91	103[l]	-0.7	3.11	0	110	Not further specified.
		Sitagliptin[e]	392	-0.42	0.91	392[l]	-0.4	3.05	5	411	
29	Ristic, 2005	Placebo	55	-0.13	0.74	55	-0.73	2.45	3	56	Symptomatic, asymptomatic (plasma glucose < 3.7 mmol/l) and symptomatic confirmed hypoglycemia.
		Vildagliptin[e]	217	-0.42	0.76	218	-0.13	2.38	14	220	

30	Scher-baum, 2008	Placebo	149	0.1	1.22	149	-0.2	3.66	1	150	Symptoms suggestive of low blood glucose confirmed by self-monitored blood glucose measurement < 3.1 mmol/l.
		Vildagliptin	153	-0.2	1.24	153	-0.5	3.71	0	156	
31	Scher-baum, 2008	Placebo	61	0.5	0.78	61[l]	-0.3	3.12	2	63	Symptoms suggestive of low blood glucose confirmed by self-monitored blood glucose measurement < 3.1 mmol/l.
		Vildagliptin	67	0.1	0.82	67[l]	-1.1	4.09	0	68	
32	Defronzo, 1995	Metformin	210[c]	-0.4	1.45	210[c]	-3.8	2.9	4	210	Symptoms compatible with hypoglycemia without biochemical documentation.
		Glibenclamide	209[c]	0.2	1.45	209[c]	-0.3	2.89	6	209	
33	Charpen-tier, 2001	Metformin	75[c]	0.07	1.21	75[c]	-0.74	2.58	8	75	Symptomatic hypoglycemia, including severe hypoglycemia (symptoms necessitating assis-tance from another person and loss of conscious-ness and/or medical intervention).
		Glimepiride	150[c]	0.27	1.1	150[c]	0.78	2.98	20	150	
34[p]	Tosi, 2003	Metformin	19	-0.49	1.11	19	-1.48	1.75[g]	2	19	Symptomatic and severe hypoglycemia.
		Glibenclamide	20	-0.5	1.5	20	0.7	1.28[g]	2	20	
35	Hermann, 1994	Metformin	19	-0.9	0.87	19	-0.8	2.18	8	38	Interpreted by the investigator on the basis of clinical findings and any available blood glucose measurements.
		Glibenclamide	19	-1.3	0.87	19	2.8	3.05	12	34	
36	Goldstein, 2003	Metformin	71	-0.3	1.24[g]	75	-2.7	2.6	1	75	Symptoms of hypoglycemia were confirmed by a fingerstick blood glucose level < 2,8 mmol/l.
		Glipizide	79	-0.4	1.19[g]	83	-0.4	2.73	0	84	
37	Russell-Jones, 2012	Metformin	246[c]	-1.48	1.1	246[c]	-2	3.14	10	246	Symptoms of hypoglycemia, not confirmed by blood glucose measurement.
		Pioglitazone	163[c]	-1.63	1.02	163[c]	1.5	3.83	6	163	
		Sitagliptin	163	-1.15	1.02	163	-0.8	3.83	5	163	
38	Moses, 1999	Metformin	27	-0.33	1.25	27	-0.86	2.65	0	27	Symptomatic hypoglycemia; not further specified.
		Repaglinide	28	-0.38	1.22	28	2.98	2.59	3	29	
39	Lund, 2007	Metformin	83	-0.16	0.93	82	-0.88	2.67	23[k]	92	Mild hypoglycemic symptoms treated by the patient (for sensitivity analysis: biochemical hypoglycemia, defined as plasma glucose < 3.5 mmol/l).
		Repaglinide	82	-0.33	0.92	82	0.7	2.68	46[k]	89	
40	Bosi, 2009	Metformin	285	-1.4	1.01	285[l]	-1.62	3.71	2	292	Symptoms suggestive of hypoglycemia and confirmed by self-monitored plasma glucose < 3.1 mmol/l.
		Vildagliptin	287	-1.1	1.02	287[l]	-0.59	3.73	2	297	
41	Aschner, 2010	Metformin	498	-0.55	0.63	446	-1.9	2.65	17	522	Symptomatic hypoglycemia; not further specified.
		Sitagliptin	512	-0.38	0.64	458	-0.6	2.73	9	528	

42	Schweizer, 2009	Metformin	166[c]	-0.75	0.9	166[c]	-1.25	2.45	2	165	Symptoms suggestive of hypoglycemia and confirmed by self-monitored plasma glucose < 3.1 mmol/l.
		Vildagliptin	169[c]	-0.64	0.91	169[c]	-0.45	2.6	0	167	
43	Jadzinsky, 2009	Metformin	313	-2	1.47[q]	313[l]	-1.6	3.18	13	328	Reported hypoglycemia was defined as events consistent with signs or symptoms of hypoglycemia with or without documented blood glucose levels < 2,8 mmol/l.
		Saxagliptin	317	-1.7	2.23[q]	317[l]	-1.1	3.74	5	335	
44	Williams-Herman, 2009	Metformin[e]	251	-1.16	1.02	248	-1.27	3.98	4	364	Symptomatic hypoglycemia, documentation of a glucose determination at the time the patient had symptoms was not required. Events of hypo-glycemia were defined as follows: those not requiring assistance or those requiring the (nonmedical) assistance of others.
		Sitagliptin	106	-0.8	1.05	100	0.6	4.08	2	179	
45	Williams-Herman, 2010	Metformin	151	-1.22	0.76	140	-1.73	4.24[m]	7	364	See study No. 1919.
		Sitagliptin	50	-1.2	0.9	50	0.5	4.22[m]	2	179	
46	Schweizer, 2007	Metformin	249	-1.4	1.58	249	-1.9	4.73	1	252	Symptoms suggestive of hypoglycemia and confirmed by self-monitored plasma glucose < 3.1 mmol/l.
		Vildagliptin	511	-1	2.26	511	0.3	4.52	3	519	
47	Goke, 2008	Metformin	158	-1.5	1.26	158[l]	-2.5	6.28	0	159	See study No. 2862.
		Vildagliptin	300	-1	1.73	300[l]	0.5	6.93	1	304	
48	Perriello, 2006	Gliclazide	135	-0.79	1.21[g]	135[l]	2	8.14[m]	2	137	Not specified.
		Pioglitazone	140	-0.79	1.3[g]	140[l]	2	8.3[m]	1	146	
49	Jain, 2006	Glibenclamide	251	-2.02	1.31[j]	251[l]	1.95	5.35	61	251	Hypoglycemic events were defined as follows: two or more simultaneous symptoms of hypoglycemia, one symptom before ingesting a glucose- or lactose-containing substance, a blood glucose level < 3,3mmol/l (home monitoring) or < 3.9 mmol/l (clinical laboratory test).
		Pioglitazone	251	-2.07	1.31[j]	251[l]	3.66	6.14	11	251	
50	Tan, 2004	Glibenclamide	96	-0.4	1.23	108	1.1	4.16	32	109	Hypoglycemic episodes were defined by either signs or symptoms of hypoglycemia as reported by the patient, or blood glucose levels ≤ 2.8 mmol/ l, regardless of the presence of hypoglycemic signs or symptoms.
		Pioglitazone	83	-0.5	1.34	90	3	4.74	4	91	

51	Tan, 2004	Glimepiride	99	-0.68	1.68	99[l]	0.79	3.87[m]	38	123	Hypoglycemic episodes were defined either subjectively (ie, in terms of reported signs or symptoms) or by a blood glucose level ≤ 2.8 mmol/L, regardless of the presence of subjective symptoms.
		Pioglitazone	109	-0.78	1.69	109[l]	1.49	4.6[m]	19	121	
52	Madsbad, 2001	Glipizide	81	0.78	0.96	81	-0.9	3.93[g]	15	81	Minor hypoglycemic events, not further specified.
		Repaglinide	175	0.19	2.16	175	-0.7	3.64[g]	26	175	
53	Marbury, 1999	Glibenclamide	171	0.1	1.44	171[l]	0.05	4.09[q]	37	193	Mild-to-moderate hypoglycemic events were defined as symptoms of sweating, strong hunger, dizziness, tremors, and/or a blood glucose level < 3,9mmol/l.
		Repaglinide	338	0.08	1.29	338[l]	-0.22	3.85[q]	59	383	
54	Wolffen-büttel, 1999	Glibenclamide	139	0.45	1.41	139[l]	0.7	3.28[g]	13	139	Not further specified, probably mild and mode-rate symptoms that could be treated/corrected by the patient without the assistance of other individuals.
		Repaglinide	286	0.58	1.51	286[l]	0	3.65[g]	26	286	
55	NA	Gliclazide	202	-0.87	0.72	194	-0.51	2.99	68	219	Hypoglycemic episodes based on subjects self-reports.
		Repaglinide	206	-0.86	0.73	196	-0.75	3.05	74	216	
56	Feinböck, 2003	Glimepiride	111[c]	-2.5	2.2	111[c]	-0.4	5.2	20	111	Mild-to-moderate hypoglycemic episodes, not further defined.
		Acarbose	108[c]	-1.8	2.2	108[c]	-1.9	3.9	2	108	
57	Salman, 2001	Gliclazide	30	-2.2	3.29[m]	30	0.4	2.83[j]	3	30	Mild-to-moderate hypoglycemic episodes, not further defined.
		Acarbose	27	-1.8	2.38[m]	27	-1.1	2.83[j]	0	27	
58[n]	Ferreira, 2011	Glipizide	142	-0.62	0.91	148	1.2	3.65	36	212	Symptomatic hypoglycemia, not further specified.
		Sitgaliptin	135	-0.7	0.8	143	-0.6	3.59	13	210	
59[n]	Foley, 2009	Gliclazide	545	-0.71	1.63	409	1.6	4.65	14	545	Grade 1 hypoglycemia: symptoms suggestive of hypoglycemia and confirmed by self-monitored plasma glucose < 3.1 mmol/l not requiring the assistance of another party.
		Vildagliptin	543	-0.51	1.63	409	0.75	4.85	4	545	
60	Rosen-stock, 2007	Pioglitazone	157	-1.4	1.25	157[l]	1.5	3.76	1	161	Symptoms suggestive of hypoglycemia and confirmed by self-monitored plasma glucose < 3.1 mmol/l.
		Vildagliptin	150	-1.1	1.22	150[l]	0.2	3.67	1	153	
61	Pan, 2009	Nateglinide	119	-0.9	0.98	119	-0.66	1.79	10	119	Not further specified.
		Acarbose	118	-0.83	0.81	118	-2.06	2	3	118	

62	Pan, 2008	Acarbose	216	-1.3	1.47	216[l]	-1.7	2.94	0	220	Hypoglycemia was defined as symptoms suggestive of low blood glucose confirmed by SMBG measurement < 3.1 mmol/l plasma glucose equivalent.
		Vildagliptin	431	-1.4	2.08	431[l]	-0.4	2.08	0	440	

N denotes the numbers of patients analyzed for HbA1c and body weight, respectively; regarding hypogylcemia r indicates the number of patients experiencing at least one hypoglycemic episode and n indicates the number of patients at risk in the safety population.

[a] For quantitative synthesis the standard error of the mean was extracted or calculated (see 3.1.3.4).

[b] Unfortunately the reporting of hypoglycemic episodes - despite existing guidelines and recommendations of regulatory bodies (ADA working paper 2005, EMA) – lacks consistency and completeness and differs considerably between trials. According to the working paper of the ADA working group on hypoglycemia data was extracted for the following categories to account for differences in severity of hypoglycemic episodes: 1, severe hypoglycemia; 2, documented symptomatic hypoglycemia; 3, asymptomatic hypoglycemia; 4, probable symptomatic hypoglycemia; 5, relative hypoglycemia. Due to the scarcity of data within each category and the fact that the categories of clinical relevance (1 and 2) were rare in all trials (most trials reported zero event rates) the original approach to analyze this outcome for each category was abandoned. Therefore any hypoglycemia was taken as outcome and its trial specific definition (if available) was detailed to facilitate interpretability.

[c] Sample size not reported. For analysis or necessary transformations requiring the sample size the number of randomized patients was taken as an approximation.

[d] The standard deviation was derived from the corresponding confidence interval, unless other indicated a 0.95 % confidence interval (see 3.1.3.4). For sample sizes smaller than 60 a t-distribution was used instead of standardized normal distribution.

[e] Different doses of one drug were aggregated into one arm (see 3.1.3.4).

[f] Measure of uncertainty for mean change from baseline not reported; since imputation by means of a correlation coefficient wasn´t possible and final values couldn´t be analyzed due to imbalanced baseline values the averaged standard deviation from similar studies with the same intervention, the same degree of measurement error and a similar time period was used instead. However, the appropriateness of this method is unclear, thus these studies were excluded from sensitivity analysis to account for the potential bias introduced by this approach.

[g] Measure of uncertainty for mean change from baseline not reported; the standard deviation for each arm was imputed by using correlation coefficients from similar studies with the same intervention, the same degree of measurement error and a similar time period (see 3.1.3.4). For study No. 4 (5), 34 (2, 35), 36 (3, 33), 48 (49, 51), 52 (55) and 54 (55) a correlation coefficient was calculated from a similar set of studies consisting of the studies in brackets; most of the correlation coefficients were greater than 0.9 indicating a reasonable measure of the similarity of baseline and final measurements for imputation.

[h] For the two sibling studies from study number 6, trial outcome data was aggregated into one single study.

[i] The section of adverse events usually covers all kinds of events independent of withdrawals. Consequently the number of hypoglycemic events reported as adverse events should exceed the number of withdrawals due to hypoglycemic events. If this relationship is the other way around, categories may be reported separately or a data error occurred. For the core analysis the greater proportion is extracted (withdrawals) and this issue of ambiguous reporting will be subjected to sensitivity analysis (see Table 14).

[j] Measure of uncertainty for mean change from baseline not reported; the average standard deviation for each arm was derived from the confidence interval, the exact p value or its threshold of the differences of means (see 3.1.3.4). For sample sizes smaller than 60 a t distribution was used instead of standardized normal distribution.

[k] Hypoglycemias are separately reported for different categories of hypoglycemia; aggregation of events is inappropriate since it is unknown if single patients experienced events of both categories; to ensure a certain degree of comparability symptoms of hypoglycemia was chosen for the core analysis since this definition of hypoglycemia was most often used in the included studies; to account for the potential bias, the other reported category was explored in sensitivity analysis.

[l] Sample size not reported. For analysis or necessary transformations requiring the sample size the number of patients from the intention-to-treat population was taken as an approximation.

A.4 Table 13

Table 13: Further Trial and patients characteristics of included studies (arm level)

No.	Included study arms[a]	T2DM[b]	Diet[c]	Other condi-tions[d]	HbA1c[e]	Body weight[e]	Hypoglycemia[e]	AE	D AE[f]	SAE	D SAE[f]
1	Placebo	7	3	4	All randomly assigned and treated patients	All randomly assigned and treated patients	All randomly assigned and treated patients	29	1	0	NR
	Metformin (1500mg;q.d.)							38	1	1	NR
2	Placebo	8	1	4	ITT, LOCF	ITT, LOCF	ITT, as randomized	NR	NR	NR	NR
	Metformin (titration: 500mg;q.d./mainte-nance: 2,550mg;t.i.d.)							NR	NR	NR	NR
3	Placebo	7	1	4	ITT, LOCF	ITT, LOCF	Safety set, received at least one trial dose	71	2	NR	NR
	Metformin (1,500mg;t.i.d.)							78	5	NR	NR
	Miglitol (0-4wk: 25mg;q.d./ 3-12wk: 50mg;t.id./ 12-36wk: 100mg;t.i.d.)							79	11	NR	NR
4	Placebo	7	1	4	ITT, presumably LOCF for drop outs with drug exposure of at least 18 weeks	ITT, presumably LOCF for drop outs with drug exposure of at least 18 weeks	Safety set; all randomized patients	NR	NR	NR	NR
	Metformin (1,700mg;b.i.d.)							NR	NR	NR	NR
	Acarbose (300mg;t.i.d)							NR	NR	NR	NR
5	Placebo	7	3	4	ITT, FAS (≥1 drug &post bl meas);LOCF, rescue medication excluded	FAS; observed cases (exclusion of rescue medication and drop outs)	Treated set (≥ drug once)	39	5	1	NR
	Metformin (500 and 1,000mg;b.i.d.)[g]							149	9	9	NR
	Linagliptin (5mg;q.d.)							80	6	3	NR

6	Placebo	7	1	4	Some imputation method ,since analyzed patients > rand.-drop outs	Some imputation method ,since analyzed patients > rand.-drop outs	Safety set, as randomized	NR	NR	NR	NR
	Glipizide (5, 10, 15, 20, 40, 60mg;q.d.)[g]							NR	NR	NR	NR
7	Placebo	1	1	4	ITT, imputation done method unclear	ITT, imputation done method unclear	Treated set (≥ drug once)	13	NR	1	NR
	Glimepiride (level 1: 1mg;q.d./level 2: 2mg;q.d./ level 3: 4mg;q.d.)							27	NR	1	NR
8	Placebo	7	3	4	ITT, imputation unclear	ITT, imputation unclear	ITT, imputation unclear	16	0	NR	NR
	Glimepiride (1-4mg; q.d./ mean received dose:2.7mg)							9	0	NR	NR
9	Placebo	7	3	4	ITT, imputation method NR	ITT, imputation method NR	All patients, presumably safety set, equals at least ITT population	NR	NR	NR	NR
	Glibenclamide (7.5mg;t.i.d.)							NR	NR	NR	NR
10	Placebo	7	1	4	ITT, LOCF	ITT, LOCF	Presumably all patients treated	31	4	NR	NR
	Tolbutamide (750mg;t.i.d./ forced titration until 1,000mg;d)							42	3	NR	NR
	Acarbose (600mg;t.i.d.)							67	7	NR	NR
11	Placebo	7	1	4	ITT, LOCF	NR	Treated set (≥ drug once)	67	0	4	0
	Glipizide (5mg;q.d.)							77	7	6	3
	Sitagliptin (5, 12.5, 25 and 50mg;b.i.d.) [g]							284	7	10	0
12	Placebo	7	1	4	ITT, LOCF	ITT, LOCF	Treated set (≥ drug once)	25	3	0	NR
	Pioglitazone (45mg;q.d.)							30	2	1	NR
13	Placebo	8	1	4	ITT, LOCF	ITT, LOCF	Treated set (≥ drug once)	NR	NR	NR	NR
	Pioglitazone (30mg;q.d.)							NR	NR	NR	NR

14	Placebo	7	1	4	ITT, LOCF	ITT, LOCF	Treated set (≥ drug once)	60	2	NR	NR
	Pioglitazone (7.5, 15, 30, 45mg;q.d.) [g]							280	13	NR	NR
15	Placebo	7	1	1	ITT, imputation method NR	ITT	ITT	NR	1	NR	NR
	Nateglinide (480mg;t.i.d.)							NR	1	NR	NR
16	Placebo	8	1	4	ITT, imputation method NR	ITT, imputation method NR	ITT	NR	NR	NR	NR
	Acarbose (150mg;t.i.d./ uptitration after 1month to 300mg;t.i.d.)							NR	NR	NR	NR
17	Placebo	7	3	4	ITT (FAS); LOCF	ITT (FAS); LOCF	Treated set, observed cases	45	7	1	0
	Linagliptin (5 and 10mg;q.d.) [g]							174	7	5	2
18	Placebo	7	3	4	FAS /ITT (LOCF)	OC (exclusion of rescue medi-cation, no imputation)	Treated set (≥ drug once)	98	4	7	NR
	Linagliptin (5mg;q.d.)							176	5	10	NR
19	Placebo	7	1	4	ITT (FAS), imputation method NR	ITT(FAS), imputation method NR	Treated set (≥ drug once), as randomized	53	1	2	1
	Vildagliptin (10, 25,50mg;b.i.d.) [g]							136	3	1	0
20	Placebo	7	1	4	ITT, LOCF	NR	APAT	27	2	2	2
	Sitagliptin (100mg;q.d.)							82	5	6	3
21	Placebo	8	1	4	ITT, LOCF	NR	Treated set (≥ drug once), as randomized	167	4	9	3
	Sitagliptin (100 and 200mg;q.d.) [g]							317	9	24	6
22	Placebo	7	3	4	m-ITT: at least 6 weeks treatment + post bl measurement, LOCF	m-ITT: at least 6 weeks treatment + post bl measurement, LOCF	Treated set (≥ drug once), as randomized	53	1	1	NR
	Saxagliptin (2.5, 5, 10, 20 and 40mg;q.d.) [g]							215	5	3	NR

23	Placebo	7	1	4	Primary ITT, LOCF	Primary ITT, LOCF	Treated set (≥ drug once)	108	NR	9	NR
	Vildagliptin (50 and 100mg;q.d./ 50mg;b.i.d.) [g]							312	NR	23	NR
24	Placebo	7	3	4	ITT, LOCF	ITT, LOCF	Treated set (≥ drug once)	53	3	1	NR
	Vildagliptin (50 and 100mg;q.d./ 50mg;b.i.d.) [g]							150	2	10	NR
25	Placebo	7	3	4	ITT, LOCF	ITT, LOCF	Treated set (≥ drug once)	55	3	14	NR
	Sitagliptin (50 or 100mg;q.d./lower dose for patients with creatinin clearance <50 mL/min)							47	5	7	NR
26	Placebo	7	3	4	ITT, LOCF	APAT with at least one drug exposure	Treated set (≥ drug once), as randomized	39	0	0	0
	Sitagliptin (25, 50, 100 and 200mg;q.d.)[g]							189	1	3	0
27	Placebo	7	3	4	ITT, LOCF	ITT, LOCF	Treated set (≥ drug once)	49	2	3	NR
	Sitagliptin (100mg;q.d.)							44	0	1	0
28	Placebo	7	1	4	ITT, LOCF	APAT, imputation method NR	Treated set (≥ drug once), as randomized	57	4	3	0
	Sitagliptin (100 and 200mg;q.d.)[g]							194	5	12	3
29	Placebo	7	1	4	ITT, LOCF	ITT, LOCF	ITT, treated set	33	3	3	NR
	Vildagliptin (25, 50 and 100mg;q.d./ 25mg;b.i.d.)[g]							126	11	2	NR
30	Placebo	7	1	4	ITT, LOCF	ITT, LOCF	Treated set, as randomized	109	6	13	NR
	Vildagliptin (50mg;q.d.)							114	14	13	NR
31	Placebo	7	3	4	ITT, LOCF	NR	Treated set (≥ drug once), as randomized	56	4	12	NR
	Vildagliptin (50mg;q.d.)							57	4	14	NR

32	Metformin (titration: 500mg;q.d./ maintenance:2,500mg; five times a day)	8	1	4	ITT, LOCF	ITT, LOCF	ITT	NR	NR	NR	NR
	Glibenclamide (20mg;four times a day)							NR	NR	NR	NR
33	Metformin (2,550mg;t.i.d.)	7	1	4	ITT, imputation by taking into account FBG	ITT, imputation by taking into account FBG	Treated set (≥ drug once)	NR	NR	NR	NR
	Glimepiride (1-6mg; q.d.)							NR	NR	NR	NR
34	Metformin (500-3,000mg;nr)	7	1	4	ITT, imputation method NR	ITT, imputation method NR	ITT, imputation method NR	NR	NR	NR	NR
	Glibenclamide (5-15mg;nr)							NR	NR	NR	NR
35	Metformin (1,000-3,000mg;nr)	4	1	4	OC, visit-wise	OC, visit-wise	Safety set, as rand.	32	NR	NR	NR
	Glibenclamide (3.5-10.5mg;nr)							26	NR	NR	NR
36	Metformin (500mg;q.d./ uptitration to 2,000mg; four times a day)	7	3	4	ITT, LOCF	ITT, LOCF	Safety set	55	4	NR	NR
	Glipizide (30mg;b.i.d.)							57	3	NR	NR
37	Metformin (up to 2,500mg; nr)	7	1	4	m-ITT, LOCF	m-ITT, imputation method NR	Safety set, as rand.	NR	6	13	NR
	Pioglitazone (up to 45mg;nr)							NR	5	9	NR
	Sitagliptin (100mg;q.d.)							NR	1	3	NR
38	Metformin (1,000-3,000mg;nr)	7	3	4	ITT, imputation method NR	ITT, imputation method NR	Safety set, as rand.	NR	NR	NR	NR
	Repaglinide (1.5-12mg;t.i.d.)							NR	NR	NR	NR

39	Metformin (500mg;q.d./ uptitration to2,000mg;b.i.d.)	7	3	4	ITT, imputation method NR; all patients who completed at least one period	ITT, imputation method NR; all patients who completed at least one period	safety set; at least drug ≥ once	NR	NR	NR	NR
	Repaglinide (1mg;q.d./ uptitration to 6mg;t.i.d.)							NR	NR	NR	NR
40	Metformin (500mg;q.d./ uptitration to 2,000mg; four times a day)	7	3	4	ITT, LOCF	NR, only for primary outcome	Safety set; at least drug ≥ once	175	14	12	2
	Vildagliptin (100mg;b.i.d.)							153	9	4	4
41	Metformin (500mg;q.d./ uptitration to 2,000mg;b.i.d.)	7	1	4	ITT (FAS), LOCF	APAT with at least one post bl measurement, (LOCF)	Safety set; at least drug ≥ once (APAT)	215	19	8	3
	Sitagliptin (100mg;q.d.)							198	9	10	3
42	Metformin (500mg;q.d./ uptitration to 1,500mg;t.i.d.)	7	3	4	ITT, LOCF	ITT, LOCF	Safety set: drug exposure and at least one post bl safety measurement	83	13	6	NR
	Vildagliptin (100mg;q.d.)							74	7	5	NR
43	Metformin (500mg;q.d. /uptitration to 2,000mg; four times a day)	7	1	4	ITT, LOCF	NR	Safety set; at least drug ≥ once	192	11	8	1
	Saxagliptin (10mg;q.d.)							179	8	6	0
44	Metformin (1,000 and 2,000mg;b.i.d.)[g]	7	1	4	ITT, LOCF: continous APT (bl measurement; no rescue medication in phase A, at least one dose and postbl. measure-ment in extension study	ITT, LOCF: continous APT (bl measurement; no rescue medication in phase A, at least one dose and postbl. measure-ment in extension study	APT over 54 weeks, as rand.	243	13	9	5
	Sitagliptin (100mg;q.d.)							105	5	12	4

45	Metformin (1,000 and 2,000mg;b.i.d.)[g]	7	1	4	ITT, LOCF: continous APT (bl measurement; no rescue medication in phase A, at least one dose and postbl. measure-ment in extension study	ITT, LOCF: continous APT (bl measurement; no rescue medication in phase A, at least one dose and postbl. measure-ment in extension study	APT over 104weeks, as rand.	252	25	16	6
	Sitagliptin (100mg;q.d.)							108	14	13	4
46	Metformin (2,000m;b.i.d)	7	3	4	ITT, LOCF	ITT, LOCF	APAT (at least one dose of drug)	209	25	22	NR
	Vildagliptin (100mg; q.d.)							205	14	23	NR
47	Metformin (2,000mg;b.i.d.)	7	3	4	ITT, LOCF: continous APT (bl measurement; no rescue medication in phase A, at least one dose and postbl. measure-ment in extension study	ITT, LOCF: continous APT (bl measurement; no rescue medication in phase A, at least one dose and postbl. measure-ment in extension study	APAT over 104 weeks, only from patients included in extension study (one dose of extension study drug and one post week 52 safety assessment)	138	3	11	NR
	Vildagliptin (100mg;q.d.)							250	5	27	NR
48	Gliclazide (160-320mg;b.i.d./ mean dose: 184mg)	7	3	4	ITT, LOCF	NR	Safety set, as rand.	31	11	NR	NR
	Pioglitazone (30-45mg; q.d./ mean dose: 40mg)							40	7	NR	NR

49	Glibenclamide (5-15mg;nr/ range of means during maintenance phase (wk 16-52):9.9-10.5mg)	7	3	4	ITT, LOCF, as rand.	NR	Safety set, as rand.	209	25	22	NR
	Pioglitazone (15-45mg; nr/ range of means during maintenance phase (wk 16-52):34.9-37.6mg)							205	14	23	NR
50	Glibenclamide (1.75-10.5mg;nr)	7	1	4	ITT, LOCF	Safety set, as rand., LOCF	Safety set, as rand.	91	10	8	NR
	Pioglitazone (30-45mg;q.d.)							70	6	7	NR
51	Glimepiride (2-8mg;q.d./ mean dose: 6mg)	7	1	4	ITT, LOCF	NR	Safety set, as rand.	95	3	5	NR
	Pioglitazone (15-45mg; q.d./ mean dose: 27mg)							105	5	8	NR
52	Glipizide (5-15mg;nr/ uptitration)	7	3	4	ITT, imputation method NR	NR	NR	NR	16	NR	NR
	Repaglinide (1.5-12mg;t.i.d./ uptitration)							NR	25	NR	NR
53	Glibenclamide (2.5-15mg;nr)	5	3	4	ITT, LOCF	NR	Safety set; at least drug $\geq$ once	NR	NR	12	NR
	Repaglinide (1.5-12mg;t.i.d./ uptitration)							NR	NR	39	NR
54	Glibenclamide (1.75-10,5mg;nr)	4	3	4	ITT, LOCF	NR	Safety set equals ITT population	NR	NR	NR	NR
	Repaglinide (1.5-12mg;t.i.d./ uptitration)							NR	NR	NR	NR

55	Gliclazide (80mg;q.d./ uptitration to 160mg;b.i.d.)	7	3	4	ITT, LOCF	FAS, analysis of completers	Safety set, at least one trial dose	38	1	0	NR
	Repaglinide (1mg; b.i.d./ uptitration to 12mg;t.i.d.)							49	3	3	NR
56	Glimepiride (1-6mg; q.d.)	7	3	4	NR	NR	Safety set, as rand.	58	1	NR	NR
	Acarbose (150-600mg;t.i.d.)							88	5	NR	NR
57	Gliclazide (80-160mg;b.i.d./ 80% of patients got 80mg b.i.d.	7	1	4	OC, analysis of completers	OC, analysis of completers	Safety set, only analysis of completers	NR	NR	NR	NR
	Acarbose (50mg;q.d./ uptitration to 300mg; t.i.d./ 96% of patients got 300mg;t.i.d.)							NR	NR	NR	NR
58	Glipizide (2.5mg;q.d./ uptitration to 10mg;b.i.d.)	7	3	3	PP	As treated (without rescue medication)	APAT; all randomized participants who took at least one dose of study therapy	NR	18	38	NR
	Sitgaliptin (severe CRI: 25mg;q.d./ moderate CRI: 50mg;q.d.)							NR	16	36	NR
59	Gliclazide (80-320mg; nr/ mean dose at endpoint:209mg)	7	3	4	ITT, LOCF	PP, LOCF	Safety set; at least drug $\geq$ once	398	45	66	NR
	Vildagliptin (100mg;b.i.d.)							379	38	80	NR
60	Pioglitazone (30mg;q.d.)	7	3	4	ITT, LOCF	NR	Safety set; at least drug $\geq$ once	83	9	NR	NR
	Vildagliptin (100mg;q.d.)							78	4	NR	NR

61	Nateglinide (480mg;t.i.d.)	7	3	4	ITT, imputation method NR	ITT, imputation method NR	Safety set, as rand.	62	NR	1	1
	Acarbose (100mg;q.d./ uptitration to 300mg;t.i.d.)							63	NR	1	1
62	Acarbose (uptitration to 300mg;t.i.d.)	7	3	4	ITT, LOCF	NR	Safety set; at least drug ≥ once	113	7	2	NR
	Vildagliptin (100mg;b.i.d.)							154	11	7	NR

The NCT number was not extracted.
[a] (dose/d;frequency).
[b] Diagnostic criteria for T2DM were coded as follows: 1: ADA 1997; 2: ADA 1999; 3: WHO 1980; 4: WHO 1985; 5: WHO 1998; 6: Else (state definition); 7: NR.
[c] Diet was coded as follows: 1: Yes; 2: No; 3: NR.
[d] Comorbidities were coded as follows: 1: Hypertension; 2: End stage renal disease; 3: Chronic renal insufficiency; 4: Else (state other conditions).
[e] Efficacy- and safety populations in which the corresponding outcome was assessed; if reported the imputation method is stated.
[f] Patients who were discontinued due to AE or SAE, respectively.
[g] Different doses of one drug were aggregated into one arm (see 3.1.3.4).
T2DM, diabetes mellitus type 2. NR, not reported. (S)AE, (severe) adverse events. ITT, intention to treat-analysis. LOCF, last observation carried forward. . Q.D., once daily. B.I.D., twice daily. T.I.D., thrice daily. FAS, full-analysis set. Bl, baseline. Rand., randomized. OC, observed cases. AP(A)T, all patients as treated. FBG, fasting blood glucose. PP, per protocol-analysis.

A.5 Table 14

Table 14a: Risk of bias assessment at study and outcome level

No.	Sequence generation[a]	Allocation of concealment[a]	Selective reporting[a]	Other bias[a]	Blinding of				Incomplete outcome data[b]		Exclusion for sensitivity analysis[c]
					participants and personnel[b]		outcome assessment[b]		HbA1c & bw	Hypoglycemia	
					HbA1c& bw	Hypoglycemia	HbA1c& bw	Hypoglycemia			
1	U	U	L	**H**(HbA1c)	L	U	L	U	U	L	1
2	U	U	L	L	L	U	L	U	U	U	
3	U	U	L	**H** (bw)	L	U	L	U	U	U	4
4	U	U	L	**H**(HbA1c)	L	U	L	U	U	U	1
5	U	U	L	L	L	U	L	U	U	U	
6	U	U	L	**H**(HbA1c/ bw/hypo-glycemia)	L	U	L	U	U	L	1,2,4
7	U	U	L	**H** (bw)	L	U	L	U	U	U	4
8	U	U	L	**H**(HbA1c/ Hypo-glycemia)	L	**H**	L	**H**	U	U	1,2
9	U	U	U	**H** (HbA1c)	L	U	L	U	U	U	1
10	U	U	L	L	L	U	L	U	U	U	
11	L	U	L	L	L	U	L	U	U	U	
12	U	L	L	L	L	**H**	L	**H**	U	U	2
13	U	U	L	**H** (bw)	L	U	L	U	U	U	4
14	U	U	L	U	L	U	L	U	U	U	
15	U	L	U	U	L	U	L	U	L	L	
16	L	L	U	**H** (bw)	L	U	L	U	U	U	4
17	U	U	L	L	L	U	L	U	U	U	

18	L	U	L	**H** (hypo-glycemia)	L	U	L	U	**H** (bw)	L	2,4
19	U	U	L	L	L	U	L	U	U	L	
20	L	U	L	L	L	U	L	U	U	U	
21	U	U	L	**H** (hypo-glycemia)	L	U	L	U	U	U	2
22	U	U	L	L	L	U	L	U	U	U	
23	U	U	L	L	L	U	L	U	U	U	
24	U	U	L	L	L	U	L	U	U	U	
25	L	U	L	U	L	U	L	U	U	U	
26	U	L	L	U	L	U	L	U	L	L	
27	L	U	L	L	L	U	L	U	U	U	
28	U	U	L	U	L	U	L	U	U	U	
29	U	U	U	**H** (HbA1c)	L	U	L	U	U	U	1
30	U	U	L	**H** (HbA1c)	L	U	L	U	U	U	1
31	U	U	L	**H** (HbA1c)	L	U	L	U	U	U	1
32	U	U	L	**H**(HbA1c/bw)	L	U	L	U	U	U	1,4
33	U	L	L	U	L	U	L	U	U	U	
34	U	U	U	**H** (bw)	L	U	L	U	U	U	4
35	L	U	U	L	L	U	L	U	**H**	U	1,4
36	U	L	L	**H** (bw)	L	U	L	U	U	U	4
37	L	U	L	L	L	U	L	U	U	U	
38	U	U	L	**H**(HbA1c/bw)	L	U	L	U	U	U	1
39	L	L	U	**H** (hypo-glycemia)	L	U	L	U	U	U	2
40	U	U	L	U	L	U	L	U	U	U	
41	L	U	U	U	L	U	L	U	U	U	
42	U	U	L	L	L	U	L	U	U	U	
43	U	L	L	U	L	U	L	U	U	U	

44	L	U	U	**H** (hypo-glycemia)	L	U	L	U	U	U	2
45	L	U	L	**H** (hypo-glycemia)	L	U	L	U	U	U	2
46	U	U	L	U	L	U	L	U	U	L	
47	U	U	L	U	L	U	L	U	U	U	
48	U	U	L	**H**	L	U	L	U	U	U	3
49	U	U	U	U	L	U	L	U	U	U	
50	U	U	U	L	L	U	L	U	U	U	
51	L	L	U	L	L	U	L	U	U	U	
52	U	U	L	**H**(HbA1c/bw)	L	U	L	U	U	U	1,4
53	U	U	U	U	L	U	L	U	U	U	
54	U	U	L	U	L	U	L	U	U	U	
55	U	U	L	L	L	**H**	L	**H**	U	U	2
56	U	U	L	**H**(HbA1c/bw)	L	**H**	L	**H**	U	U	3
57	U	U	L	L	L	**H**	L	**H**	**H**	U	3
58[d]	NA	NA	NA	**H**	NA	NA	NA	NA	NA	NA	3
59	U	U	L	L	L	U	L	U	**H** (bw)	U	4
60	U	U	L	U	L	U	L	U	U	U	
61[d]	NA	NA	NA	**H**	NA	NA	NA	NA	NA	NA	3
62	U	U	L	U	L	U	L	U	U	U	

The overall risk of bias for an outcome within a study will be deemed high in the presence of high bias in any domain, low if all key domains (all domains except random sequence generation and allocation concealment) will be of low bias, and unclear in all other cases.

[a] Study specific assessment.

[b] Outcome specific assessment; HbA1c and body weight were assessed together since both represent measured outcomes and are in the same manner susceptible to the specific bias domains. If this assumption doesn´t hold for certain studies, the degree of bias for each outcome is separately evaluated.

[c] Entire Studies or outcomes at high risk for bias will be excluded from sensitivity analysis to test the robustness of results as follows: 1=HbA1c; 2=Hypoglycemia; 3=whole study; 4=body weight; in case of studies reporting different categories of hypoglycemia (study No. 8 and 39) the outcome data from the other category was used in sensitivity analysis.

[d] Reports were deemed to be at high risk for bias since they were only available as abstract (study No. 58) or in Chinese language (study No. 61).

H, high (bold letter). L, low. U, unclear. Bw, body weight. NA, not applicable.

Table 14b: Support for judgment of rating the risk of bias

No.	Sequence Generation	Allocation of concealment	Selective Reporting	Other bias	Incomplete outcome data	
					HbA1c&bw	Hypoglycemia
	Quote/ comment	Quote/ comment	Comment	Comment	Comment	Comment
1	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	- Metformin dose low (1,500mg/d) - Adjustment for bl value bw nr, balanced.	Numbers and reasons of dropouts balanced between groups, informativeness unclear; ITT analysis (all rand. patients); appropriateness of LOCF imputation method unclear.	Numbers and reasons of dropouts balanced between groups, informativeness unclear; reasons for missing data unlikely to be informative; all rand. Patients.
2	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	No hints for other bias.	LTFU substantial~25%; numbers of dropouts balanced, reasons not; relation between missing data and true outcome (informativeness) unclear; ITT analysis (all rand. patients); appropriateness of LOCF imputation method unclear.	LTFU substantial~25%; numbers of dropouts balanced, reasons not, informativeness unclear; no safety population defined; all rand. patients.
3	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	Adjustment for bl value bw nr, imbalanced.	Dropouts not reported, insufficient information for judgement; ITT analysis; appropriateness of LOCF imputation method unclear.	Dropouts not reported, insufficient information for judgement; all rand. patients.
4	„..by elecronic data“. Unsuffi-cient information for judgement	NR	Prespecified endpoints adequately reported (number & effect measure).	Adjustment for bl value bw nr, balanced; sd for mean change from bl for HbA1c nr; average sd was taken from a set of similar studies.	Dropouts not reported, insufficient information for judgement; (m)-ITT analysis (only patients with post-bl value at 18 weeks were included); appropriateness of LOCF imputation method unclear.	Dropouts not reported, insufficient information for judgement; all rand. patients.

5	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	No hints for other bias.	Numbers and reasons of dropouts balanced between groups, informativeness unclear; ITT: FAS set (received drug at least once and had bl plus at least one post-bl measurement) for HbA1c; appropriateness of LOCF imputation method unclear; OC analysis for body weight, informativeness unlikely.	Numbers and reasons of dropouts balanced between groups, informativeness unclear; treated set: all patients who received at least one dose of drug.
6	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	Adjustment for bl value HbA1c & bw nr, imbalanced; Hypoglycemic events reported as withdrawals and ae; numbers were extracted from withdrawals;ne-vertheless potential under-estimation of events could bias the results.	Numbers and reasons of dropouts imbalanced, informativeness unclear; ITT analysis; imputation method applied but nr.	Numbers and reasons of dropouts imbalanced, informativeness unlikely; patients were discontinued for hypoglycemia; bias unlikely since outcome measure wasn´t count data of episodes; all rand. patients.
7	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	Unusually high weight loss under placebo maybe due to different adherence to diet/exercise.	LTFU substantial~20%; numbers of dropouts imbalanced, reasons nr; ITT analysis (received drug at least once and had bl plus at least one post-bl measurement); imputation for missing data probably done, but method is nr.	LTFU substantial~20%; numbers of dropouts imbalanced, reasons nr; all rand. patients with drug exposure and postrand. assessment.

8	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	Glimepiride dose low (mean dose:2.7mg/d); hypoglycemias are reported in two categories; since aggregation is methodologically inappropriate, symptomatic hypoglycemia was taken for core analysis.	Reasons and numbers of dropouts imbalanced, informativeness unclear; ITT analysis (received drug at least once); no information of imputation of missing data.	Reasons and numbers of dropouts imbalanced, informativeness unclear; ITT analysis (received drug at least once).
9	NR	NR	No prespecified endpoints, but mentioned outcomes of data collection reported.	Adjustment for bl value HbA1c nr, imbalanced.	Reasons and numbers of dropouts imbalanced; informativeness unclear; ITT (endpoint analysis, comparing latest value with bl value) analysis; appropriateness of LOCF imputation method unclear.	Reasons and numbers of dropouts imbalanced; informativeness unclear; no safety population reported; all rand. patients were taken as approximation.
10	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	No hints for other bias.	Reasons for exclusion reasonable; dropouts not fully reported; unbalanced, no reasons; informativeness unclear; m-ITT (only patients who were at least 35 days on drug and had a post-bl measurement); appropriateness of LOCF imputation method unclear.	Reasons for exclusion reasonable; dropouts not fully reported; unbalanced, no reasons; informativeness unclear; safety population not defined (presumably patients with drug exposure).

11	„..computer-generated random allocation schedule“. Probably done adequately	NR	Prespecified endpoints adequately reported (number & effect measure).	No hints for other bias.	Reasons and numbers of dropouts imbalanced for AE and treatment failure: informative-ness unclear; ITT analysis (received drug at least once and had bl plus at least one post-bl measurement); appropriateness of LOCF imputation method unclear.	Reasons and numbers of dropouts imbalanced for AE and treatment failure: informativeness unclear; safety population: patients received drug at least once.
12	NR	„..via a central telephone - interactive voice-response system“. Probably done adequately.	Prespecified endpoints adequately reported (number & effect measure).	No hints for other bias	Reasons and numbers of dropouts reported, imbalanced for treatment failure: informativeness unclear; ITT analysis (received drug at least once and had bl plus at least one post-bl measurement); appropriateness of LOCF imputation method unclear.	Reasons and numbers of dropouts reported, imbalanced for treatment failure: informativeness unclear; safety population: patients received drug at least once.
13	„..in accordance with a randomization schedule“. Probably done Adequately.	NR	Prespecified endpoints adequately reported (number & effect measure).	Adjustment for bl value bw nr, imbalanced.	LTFU substantial~27%; dropouts only partly reported, imbalanced for therapy failure: informativeness unclear; ITT analysis (received drug at least once and had bl plus at least one post-bl measurement); appropriateness of LOCF imputation method unclear.	LTFU substantial~27%; dropouts only partly reported, imbalanced for therapy failure: informativeness unclear; safety population: patients received drug at least once.

14	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	Safety population not defined, rand. patients were taken as approximation.	LTFU substantial~67% placebo group and 42-56% in active control group; dropouts partly reported, imbalanced for treatment failure and ae: informativeness unclear; ; ITT analysis (received drug at least once and had bl plus at least one post-bl measurement); appropriateness of LOCF imputation method unclear.	LTFU substantial~67% placebo group and 42-56% in active control group; dropouts partly reported, imbalanced for treatment failure and ae: informativeness unclear; safety population: patients received drug at least once.
15	NR	„..central pharmacy“. Probably done adequately.	Adverse events as 2nd endpoint defined but not fully reported.	Safety population not defined, rand. patients were taken as approximation.	Numbers and reasons of dropouts balanced between groups, informativeness unclear; ITT analysis, imputation method nr.	Numbers and reasons of dropouts balanced between groups, informativeness unclear; ITT analysis.
16	„..Randomisation was done using a drawing of envelopes containing randomisation codes prepared by a statistician“. Probably done Adequately.	„..A copy of the code was provided only to the responsible person performing the statistical analysis. The code was only broken after the database lock“. Probably done adequately.	No prespecified endpoints, but mentioned outcomes of data collection reported.	Adjustment for bl value bw nr, imbalanced; weight loss in both groups extremely high maybe due to intensive diet/ exercise throughout the study.	Numbers and reasons of dropouts balanced between groups, informativeness unclear; ITT analysis (received drug at least once and had bl plus at least one post-bl measurement); imputation method nr.	Numbers and reasons of dropouts balanced between groups, informativeness unclear; safety population: patients received drug at least once and had undergone a subsequent tolerability observation.
17	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	No hints for other bias.	Numbers and reasons of dropouts reported, imbalanced for ae, informativeness unclear; ITT analysis (FAS set, received drug at least once and had bl plus at least one post-bl measurement); imputation method nr.	Numbers and reasons of dropouts reported, imbalanced for ae, informativeness unclear; treated set: patients received drug at least once.

18	„..computer-generated sequence supplied by interactive voice or web response systems“. Probably done adequately.	NR	Prespecified endpoints adequately reported (number & effect measure).	One hypoglycemic event in Placebo arm under rescue medication; potential confounder.	Numbers and reasons of dropouts reported, imbalanced for ae and lack of efficacy, informativeness unlear; ITT analysis (FAS set, received drug at least once and had bl plus at least one post-bl measurement) for HbA1c; appropriateness of LOCF imputation method unclear; OC analysis for body weight with substantial departure from the intervention received from that assigned at randomization.	Numbers and reasons of dropouts reported, imbalanced for ae and lack of efficacy, unlikely to be informative; treated set: patients received drug at least once.
19	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	No hints for other bias.	Numbers of dropouts reported, balanced, reasons not, informativeness unclear; ITT analysis (FAS set, received drug at least once and had bl plus at least one post-bl measurement); imputation method nr.	Numbers of dropouts reported, balanced, reasons not, informativeness unclear; safety population: rand. patients, not further defined
20	„..following a computer-generated schedule“. Probably done Adequately.	NR	Prespecified endpoints adequately reported (number & effect measure).	No hints for other bias.	Numbers and reasons of dropouts reported, imbalanced for therapeutic failure and withdrawal of consent, informativeness unclear; ITT analysis (FAS set, received drug at least once and had bl plus at least one post-bl measurement); appropriateness of LOCF imputation method unclear.	Numbers and reasons of dropouts reported, imbalanced for therapeutic failure and withdrawal of consent, informativeness unclear; all patients as treated set: patients received drug at least once.

21	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	One hypoglycemic event in Sitagliptin arm under rescue medication; potential confounder.	Numbers and reasons of dropouts reported, imbalanced for therapeutic failure, informativeness unclear; ITT analysis (all patients treated set, received drug at least once and had bl plus at least one post-bl measurement); appropriateness of LOCF imputation method unclear.	Numbers and reasons of dropouts reported, imbalanced for therapeutic failure, informativeness unclear; safety population: patients received drug at least once.
22	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	No hints for other bias.	LTFU substantial~20%;numbers and reasons of dropouts balanced between groups, informativeness unclear; m-ITT analysis (received drug at least 6 wks and had bl plus at least one post-bl measurement); appropriateness of LOCF imputation method unclear.	LTFU substantial~20%;numbers and reasons of dropouts balanced between groups, informativeness unclear; treated patients set: patients received drug at least once.
23	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	No hints for other bias.	LTFU substantial~20%; exclusion reasonable; dropouts partly reported, numbers balanced, reasons not, informativeness unclear; primary ITT analysis (received drug at least once and had bl plus at least one post-bl measurement; appropriateness of LOCF imputation method unclear.	LTFU substantial~20%; exclusion reasonable; dropouts partly reported, numbers balanced, reasons not, informativeness unclear; treated set: patients received drug at least once.
24	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	No hints for other bias.	Neither numbers nor reasons for dropouts reported; ITT analysis (received drug at least once and had bl plus at least one post-bl measurement); appropriateness of LOCF imputation method unclear.	Neither numbers nor reasons for dropouts reported; treated set: patients received drug at least once.

25	„..computer-generated sequence supplied by interactive voice or web response systems“. Probably done Adequately.	NR	Prespecified endpoints adequately reported (number & effect measure).	Adjustment for bl value bw nr, balanced.	Numbers and reasons of dropouts reported, imbalanced for therapeutic failure and withdrawal of consent, informativeness unclear; ITT analysis (received drug at least once and had bl plus at least one post-bl measurement); appropriateness of LOCF imputation method unclear.	Numbers and reasons of dropouts reported, imbalanced for therapeutic failure and withdrawal of consent, informativeness unclear; safety population: patients received drug at least once.
26	NR	„..Patients were allocated to treatment ..created by… thirdparty vendor.. . Numbered containers were used to implement allocation All study personnel, including investigators, study site personnel, patients, monitors, and central laboratory personnel, remained blinded to treatment allocation throughout the study“; done adequately.	Prespecified endpoints adequately reported (number & effect measure).	Adjustment for bl value bw nr, balanced.	Numbers and reasons of dropouts balanced between groups; informativeness unclear; ITT analysis (FAS set: received drug at least once and had bl plus at least one post-bl measurement); appropriateness of LOCF imputation method unclear.	Numbers and reasons of dropouts balanced between groups; informativeness unclear; all-patients-as-treated population: patients received drug at least once.

27	„..computer-generated sequence supplied by interactive voice or web response systems“. Probably done Adequately.	NR	Prespecified endpoints adequately reported (number & effect measure).	No hints for other bias.	Numbers and reasons of dropouts reported, imbalanced, informativeness unclear; ITT analysis (received drug at least once and had bl plus at least one post-bl measurement); appropriateness of LOCF imputation method unclear.	Numbers and reasons of dropouts reported, imbalanced, informativeness unclear; safety population: patients received drug at least once.
28	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	Adjustment for bl value bw nr, balanced.	Numbers and reasons of dropouts reported, imbalanced, informativeness unclear; ITT analysis (all patients treated: received drug at least once and had bl plus at least one post-bl measurement); appropriateness of LOCF imputation method unclear.	Numbers and reasons of dropouts reported, imbalanced, informativeness unclear; all-patients-as-treated population: patients received drug at least once.
29	NR	NR	No prespecified endpoints, but mentioned outcomes of data collection reported.	Dose finding study: one low dose of Vildagliptin (25mg/d).	Dropouts not reported, insufficient information for judgement; ITT analysis, not further defined; appropriateness of LOCF imputation method unclear.	Dropouts not reported, insufficient information for judgement; ITT analysis, not further defined.
30	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	Low dose of Vildagliptin (50mgm/d).	Numbers and reasons of dropouts reported, imbalanced for ae and protocol violation, informativeness unclear; ITT analysis, not further defined; appropriateness of LOCF imputation method unclear.	Numbers and reasons of dropouts reported, imbalanced for ae and protocol violation, informativeness unclear; safety population: all rand. patients.

31	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	Low dose of Vildagliptin (50mgm/d).	Numbers and reasons of dropouts reported, imbalanced for ae and protocol violation, informativeness unclear; ITT analysis, not further defined; appropriateness of LOCF imputation method unclear.	Numbers and reasons of dropouts reported, imbalanced for ae and protocol violation, informativeness unclear; safety population: all rand. patients.
32	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	Patients got pretherapy with glibenclamide; might favour outcome of Metformin on HbA1c and of Glibenclamide on bw.	LTFU substantial~21%; numbers of dropouts balanced, reasons not, informativeness unclear; ITT analysis (all rand. patients); appropriateness of LOCF imputation method unclear.	LTFU substantial~21%; numbers of dropouts balanced, reasons not, informativeness unclear; all rand. patients.
33	NR	„..central rand. via minitel to each study site“. Probably done adequately.	Prespecified endpoints adequately reported (number & effect measure).	Adjustment for bl value bw nr, balanced.	Dropouts partly reported, numbers balanced, reasons unclear, insufficient information for judgement; ITT analysis (all patients treated: received drug at least once and had bl plus at least one post-bl measurement); appropriateness of LOCF imputation method unclear.	Dropouts partly reported, numbers balanced, reasons unclear, insufficient information for judgement; safety population: patients received drug at least once.
34	NR	NR	No prespecified endpoints, but mentioned outcomes of data collection reported.	Addjustment for bl value bw nr, imbalanced.	Dropouts partly reported, numbers balanced, reasons unclear, insufficient information for judgement; cross-over design, no washout between treatment periods, potential bias due to treatment period interaction; only data from first sequence was analyzed.	Dropouts partly reported, numbers balanced, reasons unclear, insufficient information for judgement; cross-over design, no washout between treatment periods, potential bias due to treatment period interaction; only data from first sequence was analyzed; safety population: all rand. patients.

35	„..computer-generated sequence supplied by interactive voice or web response systems“. Probably done Adequately.	NR	No prespecified endpoints, but mentioned outcomes of data collection reported.	No hints for other bias.	LTFU substantial~52%; dropouts only reported for whole population; OC analysis of patients who completed the study.	LTFU substantial~52%; dropouts only reported for whole population; safety population: all rand. patients.
36	NR	„..central rand. via minitel to each study site“. Probably done adequately.	Prespecified endpoints adequately reported (number & effect measure).	Patients got pretherapy with sulfonylurea; might favour outcome of Glipizide on bw.	LTFU substantial~27% (Glibenclamide) and 34% (Metformin); dropouts partly reported, numbers and reasons imbalanced, informativeness unclear; ITT analysis (all patients treated: received drug at least once and had bl plus at least one post-bl measurement); appropriateness of LOCF imputation method unclear.	LTFU substantial~27% (Glibenclamide) and 34% (Metformin); dropouts partly reported, numbers and reasons imbalanced, informativeness unclear; safety population: patients received drug at least once.
37	„..computer-generated sequence supplied by interactive voice or web response systems“. Probably done Adequately.	NR	Prespecified endpoints adequately reported (number & effect measure).	No hints for other bias.	Numbers and reasons of dropouts reported, imbalanced for ae and patients decision, informativeness unclear; m-ITT analysis (received drug at least once and had bl plus at least one (un-) scheduled post-bl measurement); appropriateness of LOCF imputation method unclear.	Numbers and reasons of dropouts reported, imbalanced for ae and patients decision, informativeness unclear; safety population: all rand. patients.
38	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	Patients got pretherapy with Metformin; Repaglinide dose high; might favour outcome of Repaglinide on HbA1c and of Metformin on bw.	Numbers of dropouts reported, no reasons given, insufficient information for judgement; ITT analysis (received drug at least once and returned for at least one visit after the start of treatment); imputation method nr.	Numbers of dropouts reported, no reasons given, insufficient information for judgement; safety population: patients received drug at least once.

39	„..computer-generated sequence supplied by interactive voice or web response systems“. Probably done adequately.	„..doublemasked treatment using numbered drug containers (without knowledge of the assigned treatment)“. Probably done adequately.	No prespecified endpoints, but mentioned outcomes of data collection reported.	Moderate weight gain under Repaglinide unusual; maybe due to cross-over design; hypoglycemias are reported in two categories; since aggregation is methodologically inappropriate, symptomatic hypoglycemia was taken for core analysis.	Numbers and reasons of dropouts reported, imbalanced for therapeutical failure, informativeness unclear; cross-over design, wash out period one month; default model (only patients who completed at least one treatment period were evaluated for treatment efficacy; dropouts to be included in the estimates of treatment effects).	Numbers and reasons of dropouts reported, imbalanced for therapeutical failure, informativeness unclear; cross-over design, wash out period one month; safety population: patients received drug at least once.
40	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	Adjustment for bl value bw nr, balanced.	Numbers and reasons of dropouts reported, imbalanced for therapeutical failure, informativeness unclear; ITT analysis (all patients treated: received drug at least once and had bl plus at least one post-bl measurement); appropriateness of LOCF imputation method unclear.	Numbers and reasons of dropouts reported, imbalanced for therapeutical failure, informativeness unclear; safety population: patients received drug at least once and had at least one post-bl measurement.
41	„..computer-generated sequence supplied by interactive voice or web response systems“. Probably done adequately.	NR	No prespecified endpoints, but mentioned outcomes of data collection reported.	Patients bl characteristics are reported for PP population only and might differ from analysed ITT population.	Numbers and reasons of dropouts reported, imbalanced for ae and lack of efficacy, informativeness unclear; ITT analysis (FAS set: all patients treated: received drug at least once and had bl plus at least one post-bl measurement); appropriateness of LOCF imputation method unclear.	Numbers and reasons of dropouts reported, imbalanced for ae and lack of efficacy, informativeness unclear; safety population: all patients as treated, patients received drug at least once and for laboratory measurements had both bl and least one post-bl measurement.

42	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	No hints for other bias.	Numbers and reasons of dropouts reported, imbalanced for ae and lack of efficacy, informativeness unclear; ITT analysis (received drug at least once and had bl plus at least one post-bl measurement); appropriateness of LOCF imputation method unclear.	Numbers and reasons of dropouts reported, imbalanced for ae and lack of efficacy, informativeness unclear; safety population: received at least one dose of study medication and had at least one postbaseline safety assessment.
43	NR	„..central allocation via active response system“. Probably done adequately.	Prespecified endpoints adequately reported (number & effect measure).	Adjustment for bl value bw nr, balanced.	Numbers and reasons of dropouts reported, imbalanced for lack of efficacy, informativeness unclear; ITT analysis (received drug at least once and had bl plus at least one post-bl measurement); appropriateness of LOCF imputation method unclear.	Numbers and reasons of dropouts reported, imbalanced for lack of efficacy, informativeness unclear; safety population: patients received drug at least once.
44	„..computer-generated sequence supplied by interactive voice or web response systems“. Probably done Adequately.	NR	No prespecified endpoints, but mentioned outcomes of data collection reported.	Safety population: all rand. patients instead of patients entering extension phase: potential underestimation of adverse events (see incomplete outcome data hypoglycemia).	Numbers and reasons of dropouts reported, imbalanced for lack of efficacy and ae, informativeness unclear ; ITT analysis (continuation all-patients-treated population, all randomized patients who had a baseline measurement, did not receive glycemic rescue therapy in phase A, received at least one dose of study medication in the continuation phase, and had at least one efficacy measurement during the continuation phase); appropriateness of LOCF imputation method unclear.	Numbers and reasons of dropouts reported, imbalanced for lack of efficacy and ae, informativeness unclear ; safety population: all-patients-as-treated (APaT) population, which consisted of all randomized patients who received at least one dose of study medication.

45	„..computer-generated sequence supplied by interactive voice or web response systems“. Probably done adequately.	NR	Prespecified endpoints adequately reported (number & effect measure).	Safety population: all rand. patients instead of patients entering extension phase: potential underestimation of adverse events (see incomplete outcome data hypoglycemia).	LTFU substantial~23% (Metformin) - 36% (Sitagliptin); numbers and reasons of dropouts reported, imbalanced for lack of efficacy and ae, informativeness unclear; ITT analysis (continuation all-patients-treated population, all randomized patients who had a baseline measurement, did not receive glycemic rescue therapy in phase A, received at least one dose of study medication in the continuation phase, and had at least one efficacy measurement during the continuation phase); appropria-teness of LOCF imputation method unclear.	LTFU substantial~23% (Metformin) - 36% (Sitagliptin); numbers and reasons of dropouts reported, imbalanced for lack of efficacy and ae, informativeness unclear; safety population: all-patients-as-treated (APaT) population, which consisted of all randomized patients who received at least one dose of study medication.
46	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	Adjustment for bl value bw nr, balanced.	LTFU substantial~25% (Metformin) -28% (Vildagliptin); numbers and reasons of dropouts reported, imbalanced for lack of efficacy, ae and exclusion by investigator, informativeness unclear; ITT analysis (received drug at least once and had bl plus at least one post-bl measurement); appropriateness of LOCF imputation method unclear.	LTFU substantial~25% (Metformin) -28% (Vildagliptin); numbers and reasons of dropouts reported, imbalanced for lack of efficacy, ae and exclusion by investigator, informativeness unclear; safety population: received at least one dose of study medication and had at least one postbaseline safety assessment.

47	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	Adjustment for bl value bw nr, balanced.	Numbers and reasons of dropouts reported, imbalanced for therpeutical failure, informativeness unclear; extension ITT analysis (all randomized patients who had a baseline measurement, did not receive glycemic rescue therapy in core study, received at least one dose of study medication in the continuation phase, and had at least one efficacy measurement during the continuation phase);appropria-teness of LOCF imputation method unclear.	Numbers and reasons of dropouts reported, imbalanced for therpeutical failure, informativeness unclear; extension safety population: received at least one dose of extension study medication and had at least one extension post-bl safety assessment.
48	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	Unclear if other oral antidiabetic agents were allowed; authors didn´t state in which populations analyses were done.	Dropouts only partly reported for ae, insufficient information for judgement; ITT analysis (not further defined); appropriateness of LOCF imputation method unclear.	Dropouts only partly reported for ae, insufficient information for judgement; safety population: all rand. patients.
49	NR	NR	No prespecified endpoints, but mentioned outcomes of data collection reported.	Adjustment for bl value bw nr, balanced.	LTFU substantial~46% (Pioglitazone)-49% (Glyburide); numbers and reasons of dropouts reported, imbalanced for lack of efficacy and ae, informativeness unclear; ITT analysis (received at least one dose of study medication, no further definition); appropria-teness of LOCF imputation method unclear.	LTFU substantial~46% (Pioglitazone)-49% (Glyburide); numbers and reasons of dropouts reported, imbalanced for lack of efficacy and ae, informativeness unclear; safety population: ITT cohort (all rand. patients).

50	NR	NR	No prespecified endpoints, but mentioned outcomes of data collection reported.	No hints for other bias.	LTFU substantial~38% (Glibenclamide)-40% (Pioglitazone); imbalanced for lack of efficacy, further reasons not stated, informativeness unclear; ITT analysis (received drug at least once and had bl plus at least one post-bl measurement); appropria-teness of LOCF imputation method unclear	LTFU substantial~38% (Glibenclamide)-40% (Pioglitazone); imbalanced for lack of efficacy, further reasons not stated, informativeness unclear; safety population: all rand. patients.
51	„..random number table“. Probably done Adequately.	„..central allocation via active response system“. Probably done adequately.	No prespecified endpoints, but mentioned outcomes of data collection reported.	No hints for other bias.	LTFU substantial~29%; numbers and reasons of dropouts reported, imbalanced for lack of efficacy, informativeness unclear; ITT analysis (received drug at least once and had bl plus at least one post-bl measurement); appropria-teness of LOCF imputation method unclear.	LTFU substantial~29%; numbers and reasons of dropouts reported, imbalanced for lack of efficacy, informativeness unclear; safety population: all rand. patients.
52	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	Usufficient dose escalation of glipizide (5-15mg/d); might favour outcome of Repaglinide on HbA1c and of Glipizide on bw.	LTFU substantial~20% (Repaglinide) – 28% (Glipizide); numbers and reasons of dropouts balanced between groups, informativeness unclear; ITT analysis (not further defined); imputation method nr.	LTFU substantial~20% (Repaglinide) – 28% (Glipizide); numbers and reasons of dropouts balanced between groups, informativeness unclear; no safety population reported; all rand. patients were taken as approximation.
53	NR	NR	No prespecified endpoints, but mentioned outcomes of data collection reported.	Adjustment for bl value HbA1c & bw nr, balanced.		

54	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	Adjustment for bl value HbA1c & bw nr, balanced; unclear in which population assessment of bw was done.	LTFU substantial~21% (Glyburide) – 26% (Repaglinide); numbers of drop outs imbalanced, no intervention-specific reasons stated, insufficient information for judgement; ITT analysis (not further defined); appropriateness of LOCF imputation method unclear.	LTFU substantial~21% (Glyburide) – 26% (Repaglinide); numbers of drop outs imbalanced, no intervention-specific reasons stated, insufficient information for judgement; safety population equated to ITT population.
55	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	No hints for other bias.	Numbers and reasons of dropouts balanced between groups, informativeness unclear; ITT analysis (full analysis set: received drug at least once and had bl plus at least one post-bl measurement); appropriateness of LOCF imputation method unclear.	Numbers and reasons of dropouts balanced between groups, informativeness unclear; safety population: patients received drug at least once.
56	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	Adjustment for bl value HbA1c & bw nr, imbalanced.	LTFU substantial~16% (Glimepiride) – 45% (Acarbose); numbers and reasons of dropouts reported, imbalanced for withdrawal, ae and lack of efficacy; informativeness unclear; ITT analysis (not further defined); appropriateness of LOCF imputation method unclear.	LTFU substantial~16% (Glimepiride) – 45% (Acarbose); numbers and reasons of dropouts reported, imbalanced for withdrawal, ae and lack of efficacy; informativeness unclear; safety population: all rand. patients.
57	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	No hints for other bias.	Numbers and reasons of dropouts reported, imbalanced for noncompliance, informativeness unclear; only pp analysis reported.	Numbers and reasons of dropouts reported, imbalanced for noncompliance, informativeness unclear; safety population: all rand. patients.

58	NR	NR	NR	Abstract, no comprehensive assessment possible.	Abstract, no comprehensive assessment possible.	Abstract, no comprehensive assessment possible.
59	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	No hints for other bias.	LTFU substantial~25%; numbers and reasons of dropouts balanced between groups, informativeness unclear; ITT analysis for HbA1c (received drug at least once and had bl plus at least one post-bl measurement); appropriateness of LOCF imputation method unclear; pp analysis for body weight.	LTFU substantial~25%; numbers and reasons of dropouts balanced between groups, informativeness unclear; safety population: patients received drug at least once.
60	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	Adjustment for bl value bw nr, balanced.	Numbers and reasons of dropouts reported, imbalanced for withdrawal, ae and lack of efficacy; informativeness unclear; ITT analysis (received drug at least once and had bl plus at least one post-bl measurement); appropriateness of LOCF imputation method unclear.	Numbers and reasons of dropouts reported, imbalanced for withdrawal, ae and lack of efficacy; informativeness unclear; safety population not further defined, presumably all patients with exposure to study medication.
61	NR	NR	NR	Chinese publication, no comprehensive assessment possible.	Chinese publication, no comprehensive assessment possible.	Chinese publication, no comprehensive assessment possible.
62	NR	NR	Prespecified endpoints adequately reported (number & effect measure).	Adjustment for bl value bw nr, balanced.	Numbers and reasons of dropouts reported, imbalanced for ae and lack of efficacy; informativeness unclear; ITT analysis (not further defined); appropriateness of LOCF imputation method unclear.	Numbers and reasons of dropouts reported, imbalanced for ae and lack of efficacy; informativeness unclear; safety population: patients received drug at least once.

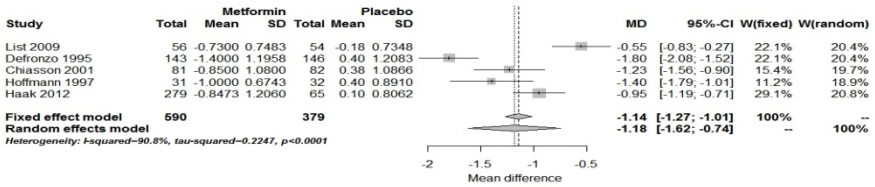
The following domains of risk of bias were left out:
- risk of bias for blinding participants and personnel and for blinding of outcome assessment for HbA1c and body weight: since these outcomes are measured they are deemed to be insusceptible to bias of unblinding or open-label studies. Thus, all studies were rated to be at low bias within these domains.
- risk of bias for blinding participants and personnel and for blinding of outcome assessment for hypoglycemia. Since this is partly a patient reported outcome knowledge of the administered drug could possibly bias the reporting of symptoms of hypoglycemia by patients and their evaluation by outcome assessors. Therefore, the risk of bias for blinded studies (n=55) was graded to be unclear since adverse reactions of the drugs could lead to unblinding; for the open-label studies (n=5) the risk of bias was judged as high.
NR, not reported. Bl, baseline. D, day. PP, per protocol analysis. (m) ITT, (modified) intention to treat analysis. FAS, full analysis set. OC, observed cases analysis. LOCF, last observation carried forward. LTFU, loss to follow-up. AE, adverse events. SD, standard deviation.

B Figures

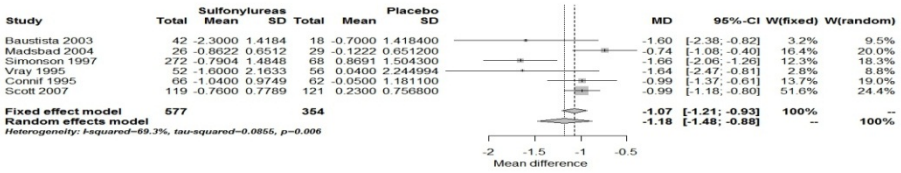
B.1 Figure 8

a.) HbA1c

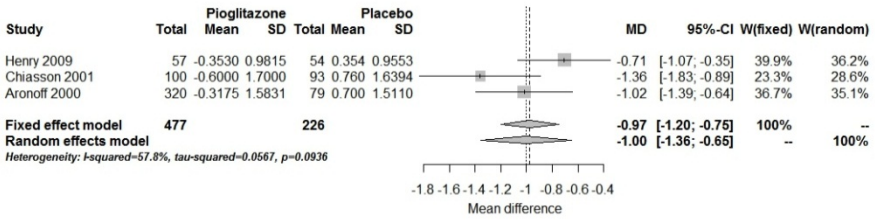
A-B



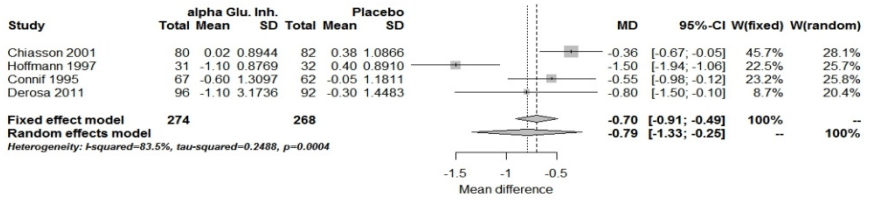
A-C



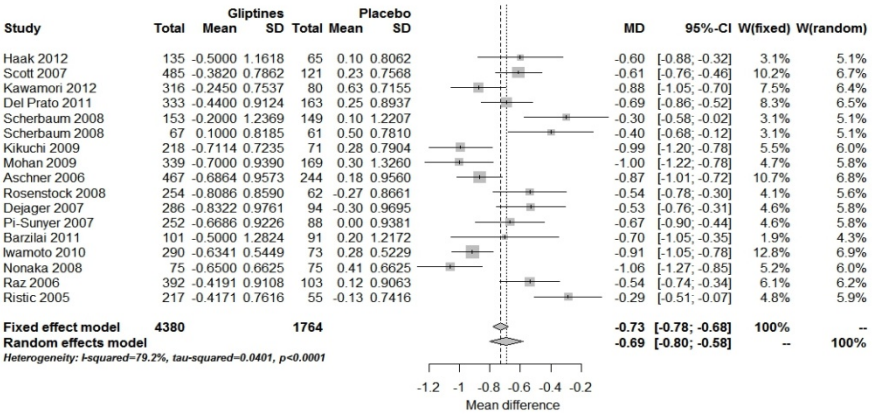
A-D



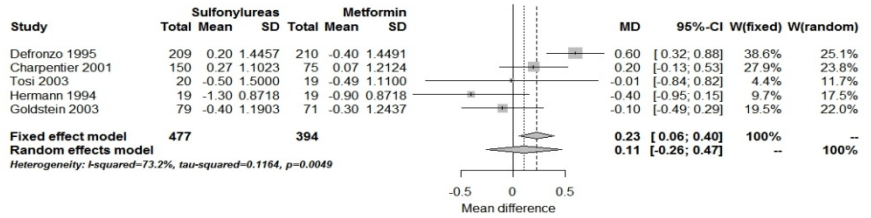
A-F


alpha Glu. Inh.
Placebo
Study Total Mean SD Total Mean SD MD 95%-CI W(fixed) W(random)
Chiasson 2001 80 0.02 0.8944 82 0.38 1.0866 -0.36 [-0.67; -0.05] 45.7% 28.1%
Hoffmann 1997 31 -1.10 0.8769 32 0.40 0.8910 -1.50 [-1.94; -1.06] 22.5% 25.7%
Connif 1995 67 -0.60 1.3097 62 -0.05 1.1811 -0.55 [-0.98; -0.12] 23.2% 25.8%
Derosa 2011 96 -1.10 3.1736 92 -0.30 1.4483 -0.80 [-1.50; -0.10] 8.7% 20.4%
Fixed effect model 274 268 -0.70 [-0.91; -0.49] 100% --
Random effects model -0.79 [-1.33; -0.25] -- 100%
Heterogeneity: I-squared=83.5%, tau-squared=0.2488, p=0.0004
-1.5 -1 -0.5
Mean difference

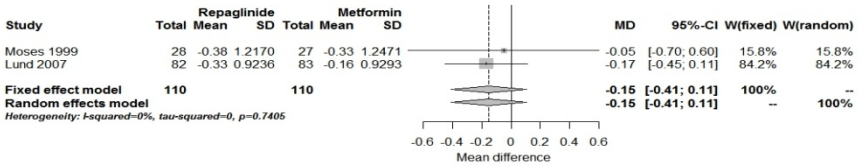
A-G


Gliptines
Placebo
Study Total Mean SD Total Mean SD MD 95%-CI W(fixed) W(random)
Haak 2012 135 -0.5000 1.1618 65 0.10 0.8062 -0.60 [-0.88; -0.32] 3.1% 5.1%
Scott 2007 485 -0.3820 0.7862 121 0.23 0.7568 -0.61 [-0.76; -0.46] 10.2% 6.7%
Kawamori 2012 316 -0.2450 0.7537 80 0.63 0.7155 -0.88 [-1.05; -0.70] 7.5% 6.4%
Del Prato 2011 333 -0.4400 0.9124 163 0.25 0.8937 -0.69 [-0.86; -0.52] 8.3% 6.5%
Scherbaum 2008 153 -0.2000 1.2369 149 0.10 1.2207 -0.30 [-0.58; -0.02] 3.1% 5.1%
Scherbaum 2008 67 0.1000 0.8185 61 0.50 0.7810 -0.40 [-0.68; -0.12] 3.1% 5.1%
Kikuchi 2009 218 -0.7114 0.7235 71 0.28 0.7904 -0.99 [-1.20; -0.78] 5.5% 6.0%
Mohan 2009 339 -0.7000 0.9390 169 0.30 1.3260 -1.00 [-1.22; -0.78] 4.7% 5.8%
Aschner 2006 467 -0.6864 0.9573 244 0.18 0.9560 -0.87 [-1.01; -0.72] 10.7% 6.8%
Rosenstock 2008 254 -0.8086 0.8590 62 -0.27 0.8661 -0.54 [-0.78; -0.30] 4.1% 5.6%
Dejager 2007 286 -0.8322 0.9761 94 -0.30 0.9695 -0.53 [-0.76; -0.31] 4.6% 5.8%
Pi-Sunyer 2007 252 -0.6686 0.9226 88 0.00 0.9381 -0.67 [-0.90; -0.44] 4.6% 5.8%
Barzilai 2011 101 -0.5000 1.2824 91 0.20 1.2172 -0.70 [-1.05; -0.35] 1.9% 4.3%
Iwamoto 2010 290 -0.6341 0.5449 73 0.28 0.5229 -0.91 [-1.05; -0.78] 12.8% 6.9%
Nonaka 2008 75 -0.6500 0.6625 75 0.41 0.6625 -1.06 [-1.27; -0.85] 5.2% 6.0%
Raz 2006 392 -0.4191 0.9108 103 0.12 0.9063 -0.54 [-0.74; -0.34] 6.1% 6.2%
Ristic 2005 217 -0.4171 0.7616 55 -0.13 0.7416 -0.29 [-0.51; -0.07] 4.8% 5.9%
Fixed effect model 4380 1764 -0.73 [-0.78; -0.68] 100% --
Random effects model -0.69 [-0.80; -0.58] -- 100%
Heterogeneity: I-squared=79.2%, tau-squared=0.0401, p<0.0001
-1.2 -1 -0.8 -0.6 -0.4 -0.2
Mean difference

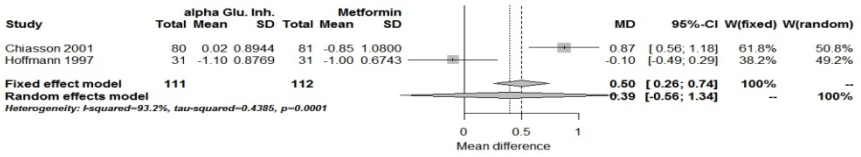
B-C


Sulfonylureas
Metformin
Study Total Mean SD Total Mean SD MD 95%-CI W(fixed) W(random)
Defronzo 1995 209 0.20 1.4457 210 -0.40 1.4491 0.60 [0.32; 0.88] 38.6% 25.1%
Charpentier 2001 150 0.27 1.1023 75 0.07 1.2124 0.20 [-0.13; 0.53] 27.9% 23.8%
Tosi 2003 20 -0.50 1.5000 19 -0.49 1.1100 -0.01 [-0.84; 0.82] 4.4% 11.7%
Hermann 1994 19 -1.30 0.8718 19 -0.90 0.8718 -0.40 [-0.95; 0.15] 9.7% 17.5%
Goldstein 2003 79 -0.40 1.1903 71 -0.30 1.2437 -0.10 [-0.49; 0.29] 19.5% 22.0%
Fixed effect model 477 394 0.23 [0.06; 0.40] 100% --
Random effects model 0.11 [-0.26; 0.47] -- 100%
Heterogeneity: I-squared=73.2%, tau-squared=0.1164, p=0.0049
-0.5 0 0.5
Mean difference

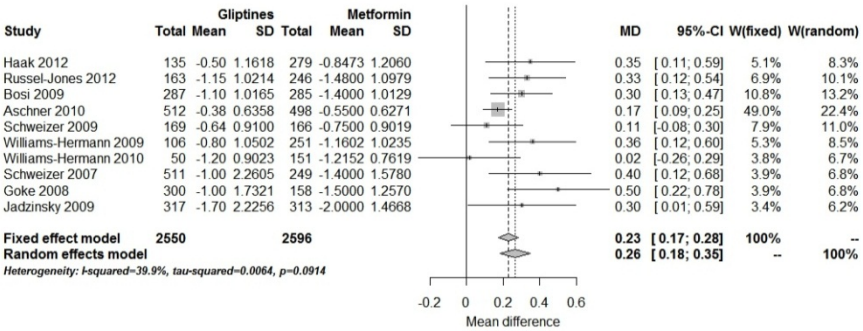
B-E


Repaglinide
Metformin
Study Total Mean SD Total Mean SD MD 95%-CI W(fixed) W(random)
Moses 1999 28 -0.38 1.2170 27 -0.33 1.2471 -0.05 [-0.70; 0.60] 15.8% 15.8%
Lund 2007 82 -0.33 0.9236 83 -0.16 0.9293 -0.17 [-0.45; 0.11] 84.2% 84.2%
Fixed effect model 110 110 -0.15 [-0.41; 0.11] 100% --
Random effects model -0.15 [-0.41; 0.11] -- 100%
Heterogeneity: I-squared=0%, tau-squared=0, p=0.7405
-0.6 -0.4 -0.2 0 0.2 0.4 0.6
Mean difference

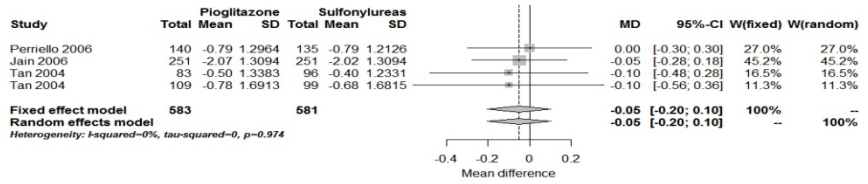
B-F


alpha Glu. Inh.
Metformin
Study Total Mean SD Total Mean SD MD 95%-CI W(fixed) W(random)
Chiasson 2001 80 0.02 0.8944 81 -0.85 1.0800 0.87 [0.56; 1.18] 61.8% 50.8%
Hoffmann 1997 31 -1.10 0.8769 31 -1.00 0.6743 -0.10 [-0.49; 0.29] 38.2% 49.2%
Fixed effect model 111 112 0.50 [0.26; 0.74] 100% --
Random effects model 0.39 [-0.56; 1.34] -- 100%
Heterogeneity: I-squared=93.2%, tau-squared=0.4385, p=0.0001
0 0.5 1
Mean difference

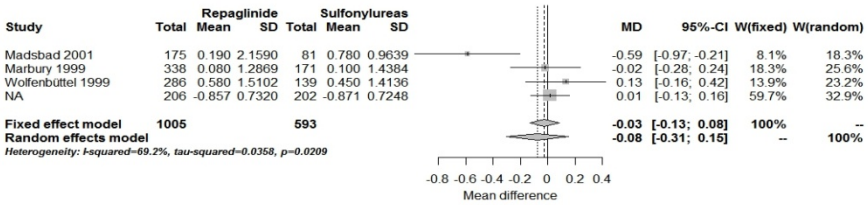
B-G


Gliptines
Metformin
Study Total Mean SD Total Mean SD MD 95%-CI W(fixed) W(random)
Haak 2012 135 -0.50 1.1618 279 -0.8473 1.2060 0.35 [0.11; 0.59] 5.1% 8.3%
Russel-Jones 2012 163 -1.15 1.0214 246 -1.4800 1.0979 0.33 [0.12; 0.54] 6.9% 10.1%
Bosi 2009 287 -1.10 1.0165 285 -1.4000 1.0129 0.30 [0.13; 0.47] 10.8% 13.2%
Aschner 2010 512 -0.38 0.6358 498 -0.5500 0.6271 0.17 [0.09; 0.25] 49.0% 22.4%
Schweizer 2009 169 -0.64 0.9100 166 -0.7500 0.9019 0.11 [-0.08; 0.30] 7.9% 11.0%
Williams-Hermann 2009 106 -0.80 1.0502 251 -1.1602 1.0235 0.36 [0.12; 0.60] 5.3% 8.5%
Williams-Hermann 2010 50 -1.20 0.9023 151 -1.2152 0.7619 0.02 [-0.26; 0.29] 3.8% 6.7%
Schweizer 2007 511 -1.00 2.2605 249 -1.4000 1.5780 0.40 [0.12; 0.68] 3.9% 6.8%
Goke 2008 300 -1.00 1.7321 158 -1.5000 1.2570 0.50 [0.22; 0.78] 3.9% 6.8%
Jadzinsky 2009 317 -1.70 2.2256 313 -2.0000 1.4668 0.30 [0.01; 0.59] 3.4% 6.2%
Fixed effect model 2550 2596 0.23 [0.17; 0.28] 100% --
Random effects model 0.26 [0.18; 0.35] -- 100%
Heterogeneity: I-squared=39.9%, tau-squared=0.0064, p=0.0914
-0.2 0 0.2 0.4 0.6
Mean difference

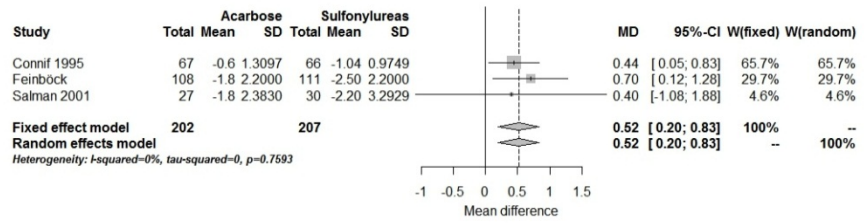
C-D


Pioglitazone
Sulfonylureas
Study Total Mean SD Total Mean SD
MD 95%-CI W(fixed) W(random)
Perriello 2006 140 -0.79 1.2964 135 -0.79 1.2126 0.00 [-0.30; 0.30] 27.0% 27.0%
Jain 2006 251 -2.07 1.3094 251 -2.02 1.3094 -0.05 [-0.28; 0.18] 45.2% 45.2%
Tan 2004 83 -0.50 1.3383 96 -0.40 1.2331 -0.10 [-0.48; 0.28] 16.5% 16.5%
Tan 2004 109 -0.78 1.6913 99 -0.68 1.6815 -0.10 [-0.56; 0.36] 11.3% 11.3%
Fixed effect model 583 581 -0.05 [-0.20; 0.10] 100% --
Random effects model -0.05 [-0.20; 0.10] -- 100%
Heterogeneity: I-squared=0%, tau-squared=0, p=0.974
-0.4 -0.2 0 0.2
Mean difference

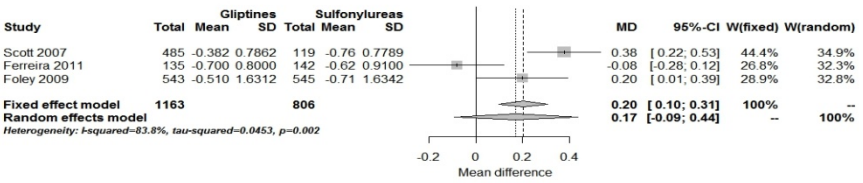
C-E


Repaglinide
Sulfonylureas
Study Total Mean SD Total Mean SD
MD 95%-CI W(fixed) W(random)
Madsbad 2001 175 0.190 2.1590 81 0.780 0.9639 -0.59 [-0.97; -0.21] 8.1% 18.3%
Marbury 1999 338 0.080 1.2869 171 0.100 1.4384 -0.02 [-0.28; 0.24] 18.3% 25.6%
Wolfenbüttel 1999 286 0.580 1.5102 139 0.450 1.4136 0.13 [-0.16; 0.42] 13.9% 23.2%
NA 206 -0.857 0.7320 202 -0.871 0.7248 0.01 [-0.13; 0.16] 59.7% 32.9%
Fixed effect model 1005 593 -0.03 [-0.13; 0.08] 100% --
Random effects model -0.08 [-0.31; 0.15] -- 100%
Heterogeneity: I-squared=69.2%, tau-squared=0.0358, p=0.0209
-0.8 -0.6 -0.4 -0.2 0 0.2 0.4
Mean difference

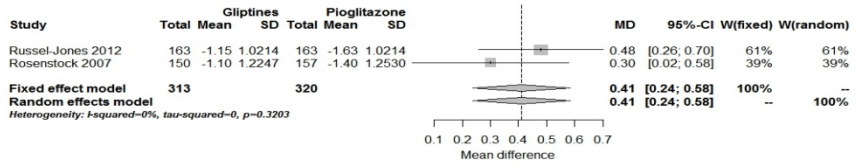
C-F


Acarbose
Sulfonylureas
Study Total Mean SD Total Mean SD
MD 95%-CI W(fixed) W(random)
Connif 1995 67 -0.6 1.3097 66 -1.04 0.9749 0.44 [0.05; 0.83] 65.7% 65.7%
Feinböck 108 -1.8 2.2000 111 -2.50 2.2000 0.70 [0.12; 1.28] 29.7% 29.7%
Salman 2001 27 -1.8 2.3830 30 -2.20 3.2929 0.40 [-1.08; 1.88] 4.6% 4.6%
Fixed effect model 202 207 0.52 [0.20; 0.83] 100% --
Random effects model 0.52 [0.20; 0.83] -- 100%
Heterogeneity: I-squared=0%, tau-squared=0, p=0.7593
-1 -0.5 0 0.5 1 1.5
Mean difference

C-G

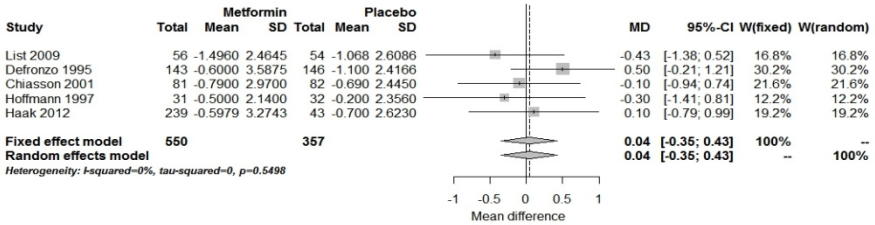


D-G

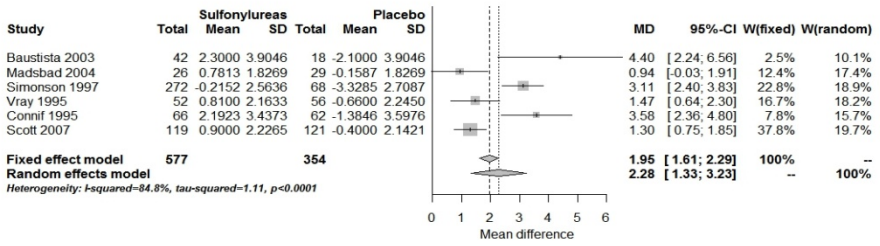


b.) Body weight

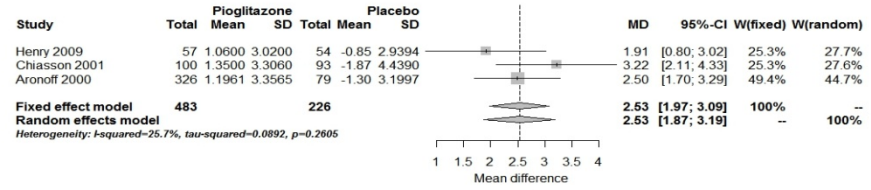
A-B



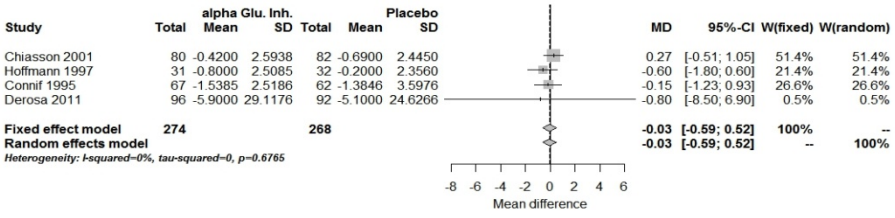
A-C


Sulfonylureas
Placebo
Study Total Mean SD Total Mean SD MD 95%-CI W(fixed) W(random)
Baustista 2003 42 2.3000 3.9046 18 -2.1000 3.9046 4.40 [2.24; 6.56] 2.5% 10.1%
Madsbad 2004 26 0.7813 1.8269 29 -0.1587 1.8269 0.94 [-0.03; 1.91] 12.4% 17.4%
Simonson 1997 272 -0.2152 2.5636 68 -3.3285 2.7087 3.11 [2.40; 3.83] 22.8% 18.9%
Vray 1995 52 0.8100 2.1633 56 -0.6600 2.2450 1.47 [0.64; 2.30] 16.7% 18.2%
Connif 1995 66 2.1923 3.4373 62 -1.3846 3.5976 3.58 [2.36; 4.80] 7.8% 15.7%
Scott 2007 119 0.9000 2.2265 121 -0.4000 2.1421 1.30 [0.75; 1.85] 37.8% 19.7%
Fixed effect model 577 354 1.95 [1.61; 2.29] 100% --
Random effects model 2.28 [1.33; 3.23] -- 100%
Heterogeneity: I-squared=84.8%, tau-squared=1.11, p<0.0001
0 1 2 3 4 5 6
Mean difference

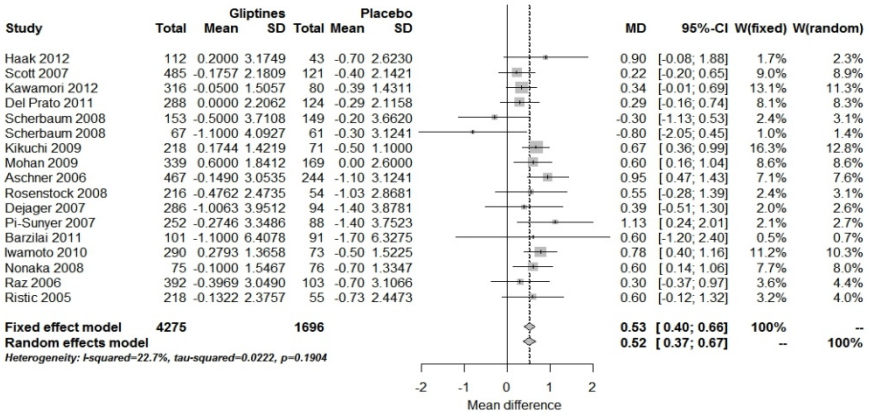
A-D


Pioglitazone
Placebo
Study Total Mean SD Total Mean SD MD 95%-CI W(fixed) W(random)
Henry 2009 57 1.0600 3.0200 54 -0.85 2.9394 1.91 [0.80; 3.02] 25.3% 27.7%
Chiasson 2001 100 1.3500 3.3060 93 -1.87 4.4390 3.22 [2.11; 4.33] 25.3% 27.6%
Aronoff 2000 326 1.1961 3.3565 79 -1.30 3.1997 2.50 [1.70; 3.29] 49.4% 44.7%
Fixed effect model 483 226 2.53 [1.97; 3.09] 100% --
Random effects model 2.53 [1.87; 3.19] -- 100%
Heterogeneity: I-squared=25.7%, tau-squared=0.0892, p=0.2605
1 1.5 2 2.5 3 3.5 4
Mean difference

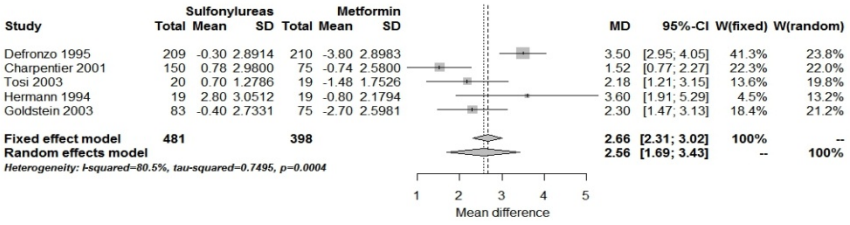
A-F


alpha Glu. Inh.
Placebo
Study Total Mean SD Total Mean SD MD 95%-CI W(fixed) W(random)
Chiasson 2001 80 -0.4200 2.5938 82 -0.6900 2.4450 0.27 [-0.51; 1.05] 51.4% 51.4%
Hoffmann 1997 31 -0.8000 2.5085 32 -0.2000 2.3560 -0.60 [-1.80; 0.60] 21.4% 21.4%
Connif 1995 67 -1.5385 2.5186 62 -1.3846 3.5976 -0.15 [-1.23; 0.93] 26.6% 26.6%
Derosa 2011 96 -5.9000 29.1176 92 -5.1000 24.6266 -0.80 [-8.50; 6.90] 0.5% 0.5%
Fixed effect model 274 268 -0.03 [-0.59; 0.52] 100% --
Random effects model -0.03 [-0.59; 0.52] -- 100%
Heterogeneity: I-squared=0%, tau-squared=0, p=0.6765
-8 -6 -4 -2 0 2 4 6
Mean difference

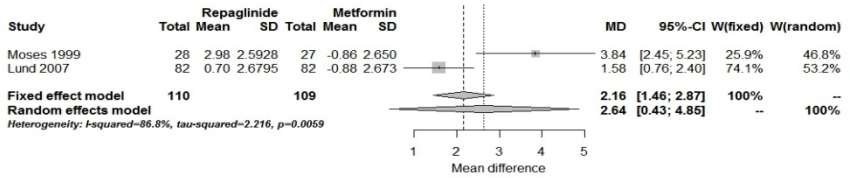
A-G


Gliptines
Placebo
Study Total Mean SD Total Mean SD MD 95%-CI W(fixed) W(random)
Haak 2012 112 0.2000 3.1749 43 -0.70 2.6230 0.90 [-0.08; 1.88] 1.7% 2.3%
Scott 2007 485 -0.1757 2.1809 121 -0.40 2.1421 0.22 [-0.20; 0.65] 9.0% 8.9%
Kawamori 2012 316 -0.0500 1.5057 80 -0.39 1.4311 0.34 [-0.01; 0.69] 13.1% 11.3%
Del Prato 2011 288 0.0000 2.2062 124 -0.29 2.1158 0.29 [-0.16; 0.74] 8.1% 8.3%
Scherbaum 2008 153 -0.5000 3.7108 149 -0.20 3.6620 -0.30 [-1.13; 0.53] 2.4% 3.1%
Scherbaum 2008 67 -1.1000 4.0927 61 -0.30 3.1241 -0.80 [-2.05; 0.45] 1.0% 1.4%
Kikuchi 2009 218 0.1744 1.4219 71 -0.50 1.1000 0.67 [0.36; 0.99] 16.3% 12.8%
Mohan 2009 339 0.6000 1.8412 169 0.00 2.6000 0.60 [0.16; 1.04] 8.6% 8.6%
Aschner 2006 467 -0.1490 3.0535 244 -1.10 3.1241 0.95 [0.47; 1.43] 7.1% 7.6%
Rosenstock 2008 216 -0.4762 2.4735 54 -1.03 2.8681 0.55 [-0.28; 1.39] 2.4% 3.1%
Dejager 2007 286 -1.0063 3.9512 94 -1.40 3.8781 0.39 [-0.51; 1.30] 2.0% 2.6%
Pi-Sunyer 2007 252 -0.2746 3.3486 88 -1.40 3.7523 1.13 [0.24; 2.01] 2.1% 2.7%
Barzilai 2011 101 -1.1000 6.4078 91 -1.70 6.3275 0.60 [-1.20; 2.40] 0.5% 0.7%
Iwamoto 2010 290 0.2793 1.3658 73 -0.50 1.5225 0.78 [0.40; 1.16] 11.2% 10.3%
Nonaka 2008 75 -0.1000 1.5467 76 -0.70 1.3347 0.60 [0.14; 1.06] 7.7% 8.0%
Raz 2006 392 -0.3969 3.0490 103 -0.70 3.1066 0.30 [-0.37; 0.97] 3.6% 4.4%
Ristic 2005 218 -0.1322 2.3757 55 -0.73 2.4473 0.60 [-0.12; 1.32] 3.2% 4.0%
Fixed effect model 4275 1696 0.53 [0.40; 0.66] 100% --
Random effects model 0.52 [0.37; 0.67] -- 100%
Heterogeneity: I-squared=22.7%, tau-squared=0.0222, p=0.1904
-2 -1 0 1 2
Mean difference

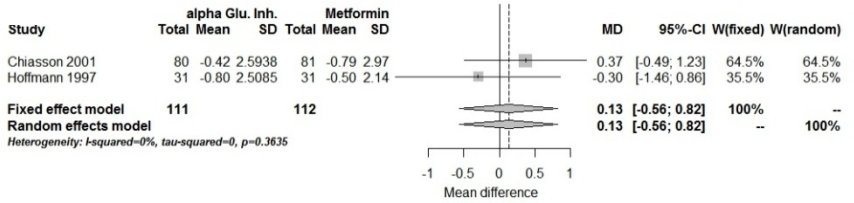
B-C


Sulfonylureas
Metformin
Study Total Mean SD Total Mean SD MD 95%-CI W(fixed) W(random)
Defronzo 1995 209 -0.30 2.8914 210 -3.80 2.8983 3.50 [2.95; 4.05] 41.3% 23.8%
Charpentier 2001 150 0.78 2.9800 75 -0.74 2.5800 1.52 [0.77; 2.27] 22.3% 22.0%
Tosi 2003 20 0.70 1.2786 19 -1.48 1.7526 2.18 [1.21; 3.15] 13.6% 19.8%
Hermann 1994 19 2.80 3.0512 19 -0.80 2.1794 3.60 [1.91; 5.29] 4.5% 13.2%
Goldstein 2003 83 -0.40 2.7331 75 -2.70 2.5981 2.30 [1.47; 3.13] 18.4% 21.2%
Fixed effect model 481 398 2.66 [2.31; 3.02] 100% --
Random effects model 2.56 [1.69; 3.43] -- 100%
Heterogeneity: I-squared=80.5%, tau-squared=0.7495, p=0.0004
1 2 3 4 5
Mean difference

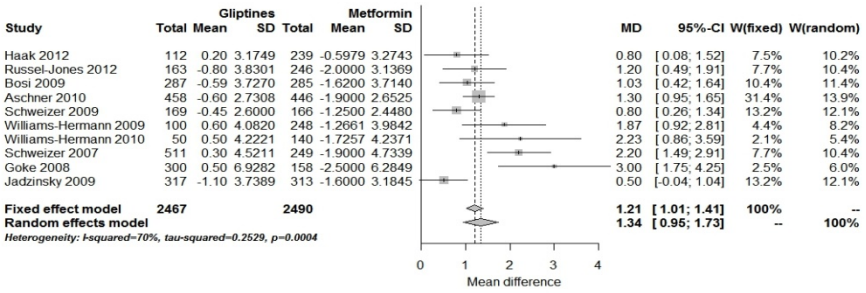
B-E


Repaglinide
Metformin
Study Total Mean SD Total Mean SD MD 95%-CI W(fixed) W(random)
Moses 1999 28 2.98 2.5928 27 -0.86 2.650 3.84 [2.45; 5.23] 25.9% 46.8%
Lund 2007 82 0.70 2.6795 82 -0.88 2.673 1.58 [0.76; 2.40] 74.1% 53.2%
Fixed effect model 110 109 2.16 [1.46; 2.87] 100% --
Random effects model 2.64 [0.43; 4.85] -- 100%
Heterogeneity: I-squared=86.8%, tau-squared=2.216, p=0.0059
1 2 3 4 5
Mean difference

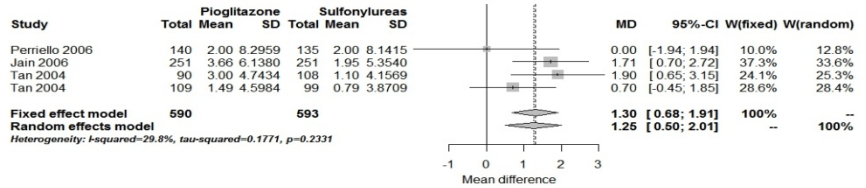
B-F


alpha Glu. Inh.
Metformin
Study Total Mean SD Total Mean SD MD 95%-CI W(fixed) W(random)
Chiasson 2001 80 -0.42 2.5938 81 -0.79 2.97 0.37 [-0.49; 1.23] 64.5% 64.5%
Hoffmann 1997 31 -0.80 2.5085 31 -0.50 2.14 -0.30 [-1.46; 0.86] 35.5% 35.5%
Fixed effect model 111 112 0.13 [-0.56; 0.82] 100% --
Random effects model 0.13 [-0.56; 0.82] -- 100%
Heterogeneity: I-squared=0%, tau-squared=0, p=0.3635
-1 -0.5 0 0.5 1
Mean difference

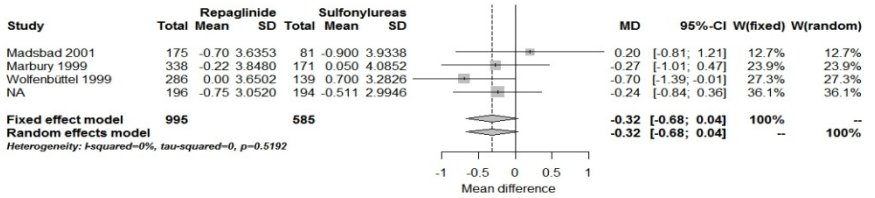
B-G


Gliptines
Metformin
Study Total Mean SD Total Mean SD MD 95%-CI W(fixed) W(random)
Haak 2012 112 0.20 3.1749 239 -0.5979 3.2743 0.80 [0.08; 1.52] 7.5% 10.2%
Russel-Jones 2012 163 -0.80 3.8301 246 -2.0000 3.1369 1.20 [0.49; 1.91] 7.7% 10.4%
Bosi 2009 287 -0.59 3.7270 285 -1.6200 3.7140 1.03 [0.42; 1.64] 10.4% 11.4%
Aschner 2010 458 -0.60 2.7308 446 -1.9000 2.6525 1.30 [0.95; 1.65] 31.4% 13.9%
Schweizer 2009 169 -0.45 2.6000 166 -1.2500 2.4480 0.80 [0.26; 1.34] 13.2% 12.1%
Williams-Hermann 2009 100 0.60 4.0820 248 -1.2661 3.9842 1.87 [0.92; 2.81] 4.4% 8.2%
Williams-Hermann 2010 50 0.50 4.2221 140 -1.7257 4.2371 2.23 [0.86; 3.59] 2.1% 5.4%
Schweizer 2007 511 0.30 4.5211 249 -1.9000 4.7339 2.20 [1.49; 2.91] 7.7% 10.4%
Goke 2008 300 0.50 6.9282 158 -2.5000 6.2849 3.00 [1.75; 4.25] 2.5% 6.0%
Jadzinsky 2009 317 -1.10 3.7389 313 -1.6000 3.1845 0.50 [-0.04; 1.04] 13.2% 12.1%
Fixed effect model 2467 2490 1.21 [1.01; 1.41] 100% --
Random effects model 1.34 [0.95; 1.73] -- 100%
Heterogeneity: I-squared=70%, tau-squared=0.2529, p=0.0004
0 1 2 3 4
Mean difference

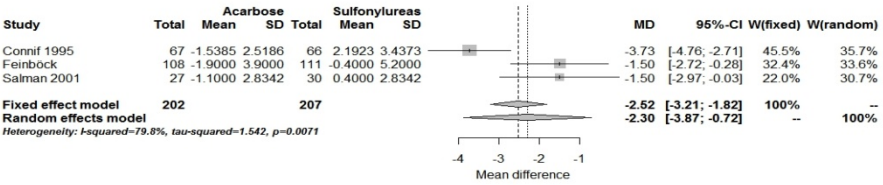
C-D


Pioglitazone
Sulfonylureas
Study Total Mean SD Total Mean SD MD 95%-CI W(fixed) W(random)
Perriello 2006 140 2.00 8.2959 135 2.00 8.1415 0.00 [-1.94; 1.94] 10.0% 12.8%
Jain 2006 251 3.66 6.1380 251 1.95 5.3540 1.71 [0.70; 2.72] 37.3% 33.6%
Tan 2004 90 3.00 4.7434 108 1.10 4.1569 1.90 [0.65; 3.15] 24.1% 25.3%
Tan 2004 109 1.49 4.5984 99 0.79 3.8709 0.70 [-0.45; 1.85] 28.6% 28.4%
Fixed effect model 590 593 1.30 [0.68; 1.91] 100% --
Random effects model 1.25 [0.50; 2.01] -- 100%
Heterogeneity: I-squared=29.8%, tau-squared=0.1771, p=0.2331
-1 0 1 2 3
Mean difference

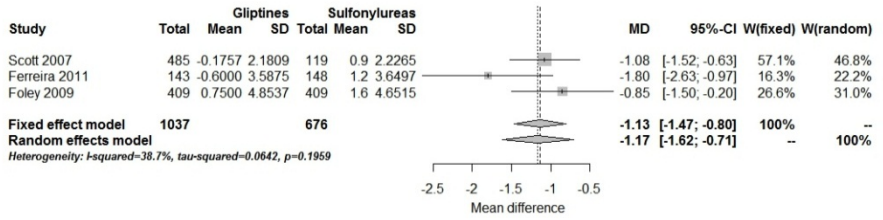
C-E


Repaglinide
Sulfonylureas
Study Total Mean SD Total Mean SD
MD 95%-CI W(fixed) W(random)
Madsbad 2001 175 -0.70 3.6353 81 -0.900 3.9338 0.20 [-0.81; 1.21] 12.7% 12.7%
Marbury 1999 338 -0.22 3.8480 171 0.050 4.0852 -0.27 [-1.01; 0.47] 23.9% 23.9%
Wolfenbüttel 1999 286 0.00 3.6502 139 0.700 3.2826 -0.70 [-1.39; -0.01] 27.3% 27.3%
NA 196 -0.75 3.0520 194 -0.511 2.9946 -0.24 [-0.84; 0.36] 36.1% 36.1%
Fixed effect model 995 585 -0.32 [-0.68; 0.04] 100% --
Random effects model -0.32 [-0.68; 0.04] -- 100%
Heterogeneity: I-squared=0%, tau-squared=0, p=0.5192
-1 -0.5 0 0.5 1
Mean difference

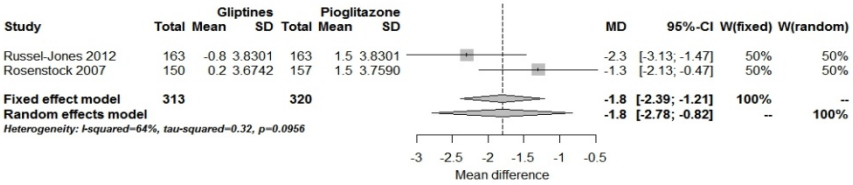
C-F


Acarbose
Sulfonylureas
Study Total Mean SD Total Mean SD
MD 95%-CI W(fixed) W(random)
Connif 1995 67 -1.5385 2.5186 66 2.1923 3.4373 -3.73 [-4.76; -2.71] 45.5% 35.7%
Feinbock 108 -1.9000 3.9000 111 -0.4000 5.2000 -1.50 [-2.72; -0.28] 32.4% 33.6%
Salman 2001 27 -1.1000 2.8342 30 0.4000 2.8342 -1.50 [-2.97; -0.03] 22.0% 30.7%
Fixed effect model 202 207 -2.52 [-3.21; -1.82] 100% --
Random effects model -2.30 [-3.87; -0.72] -- 100%
Heterogeneity: I-squared=79.8%, tau-squared=1.542, p=0.0071
-4 -3 -2 -1
Mean difference

C-G

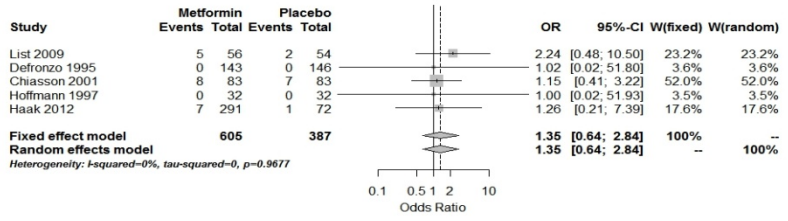

Gliptines
Sulfonylureas
Study Total Mean SD Total Mean SD
MD 95%-CI W(fixed) W(random)
Scott 2007 485 -0.1757 2.1809 119 0.9 2.2265 -1.08 [-1.52; -0.63] 57.1% 46.8%
Ferreira 2011 143 -0.6000 3.5875 148 1.2 3.6497 -1.80 [-2.63; -0.97] 16.3% 22.2%
Foley 2009 409 0.7500 4.8537 409 1.6 4.6515 -0.85 [-1.50; -0.20] 26.6% 31.0%
Fixed effect model 1037 676 -1.13 [-1.47; -0.80] 100% --
Random effects model -1.17 [-1.62; -0.71] -- 100%
Heterogeneity: I-squared=38.7%, tau-squared=0.0642, p=0.1959
-2.5 -2 -1.5 -1 -0.5
Mean difference

D-G

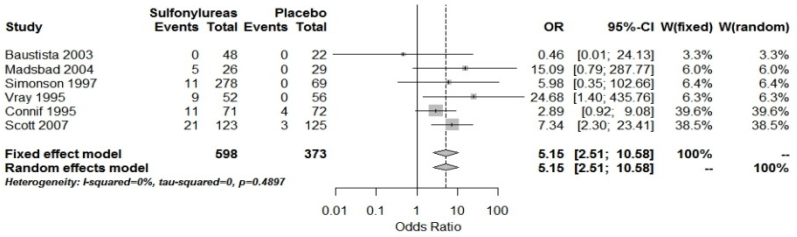


c.) Hypoglycemia

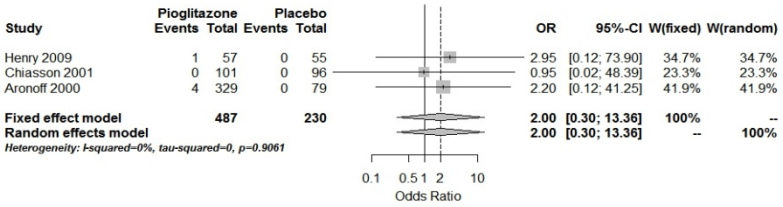
A-B



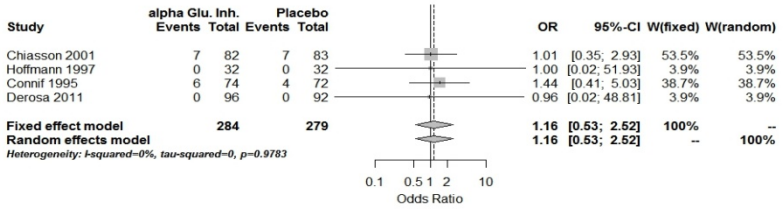
A-C



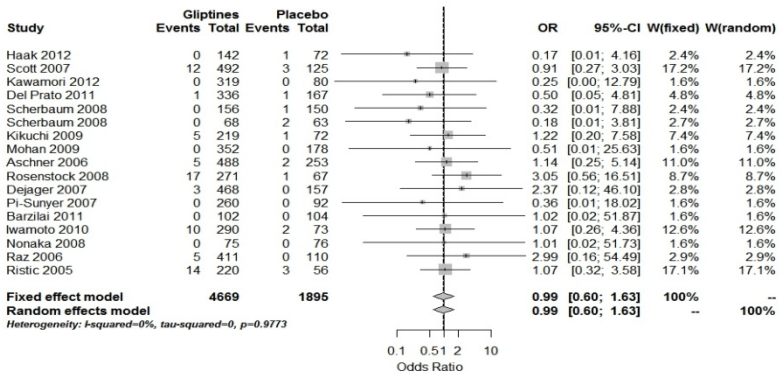
A-D


Study
Pioglitazone Events Total
Placebo Events Total
OR 95%-CI W(fixed) W(random)
Henry 2009 1 57 0 55 2.95 [0.12; 73.90] 34.7% 34.7%
Chiasson 2001 0 101 0 96 0.95 [0.02; 48.39] 23.3% 23.3%
Aronoff 2000 4 329 0 79 2.20 [0.12; 41.25] 41.9% 41.9%
Fixed effect model 487 230 2.00 [0.30; 13.36] 100% --
Random effects model 2.00 [0.30; 13.36] -- 100%
Heterogeneity: I-squared=0%, tau-squared=0, p=0.9061
0.1 0.5 1 2 10
Odds Ratio

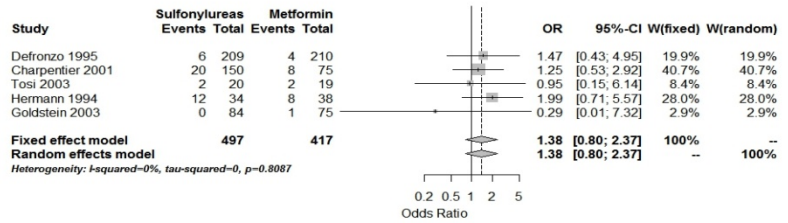
A-F


Study
alpha Glu. Inh. Events Total
Placebo Events Total
OR 95%-CI W(fixed) W(random)
Chiasson 2001 7 82 7 83 1.01 [0.35; 2.93] 53.5% 53.5%
Hoffmann 1997 0 32 0 32 1.00 [0.02; 51.93] 3.9% 3.9%
Connif 1995 6 74 4 72 1.44 [0.41; 5.03] 38.7% 38.7%
Derosa 2011 0 96 0 92 0.96 [0.02; 48.81] 3.9% 3.9%
Fixed effect model 284 279 1.16 [0.53; 2.52] 100% --
Random effects model 1.16 [0.53; 2.52] -- 100%
Heterogeneity: I-squared=0%, tau-squared=0, p=0.9783
0.1 0.5 1 2 10
Odds Ratio

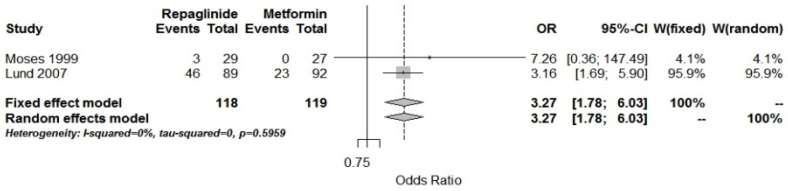
A-G


Study
Gliptines Events Total
Placebo Events Total
OR 95%-CI W(fixed) W(random)
Haak 2012 0 142 1 72 0.17 [0.01; 4.16] 2.4% 2.4%
Scott 2007 12 492 3 125 0.91 [0.27; 3.03] 17.2% 17.2%
Kawamori 2012 0 319 0 80 0.25 [0.00; 12.79] 1.6% 1.6%
Del Prato 2011 1 336 1 167 0.50 [0.05; 4.81] 4.8% 4.8%
Scherbaum 2008 0 156 1 150 0.32 [0.01; 7.88] 2.4% 2.4%
Scherbaum 2008 0 68 2 63 0.18 [0.01; 3.81] 2.7% 2.7%
Kikuchi 2009 5 219 1 72 1.22 [0.20; 7.58] 7.4% 7.4%
Mohan 2009 0 352 0 178 0.51 [0.01; 25.63] 1.6% 1.6%
Aschner 2006 5 488 2 253 1.14 [0.25; 5.14] 11.0% 11.0%
Rosenstock 2008 17 271 1 67 3.05 [0.56; 16.51] 8.7% 8.7%
Dejager 2007 3 468 0 157 2.37 [0.12; 46.10] 2.8% 2.8%
Pi-Sunyer 2007 0 260 0 92 0.36 [0.01; 18.02] 1.6% 1.6%
Barzilai 2011 0 102 0 104 1.02 [0.02; 51.87] 1.6% 1.6%
Iwamoto 2010 10 290 2 73 1.07 [0.26; 4.36] 12.6% 12.6%
Nonaka 2008 0 75 0 76 1.01 [0.02; 51.73] 1.6% 1.6%
Raz 2006 5 411 0 110 2.99 [0.16; 54.49] 2.9% 2.9%
Ristic 2005 14 220 3 56 1.07 [0.32; 3.58] 17.1% 17.1%
Fixed effect model 4669 1895 0.99 [0.60; 1.63] 100% --
Random effects model 0.99 [0.60; 1.63] -- 100%
Heterogeneity: I-squared=0%, tau-squared=0, p=0.9773
0.1 0.5 1 2 10
Odds Ratio

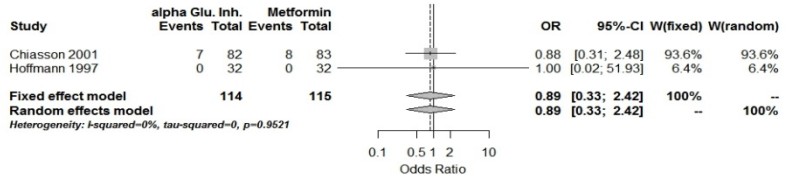
B-C


Study
Sulfonylureas Events Total
Metformin Events Total
OR 95%-CI W(fixed) W(random)
Defronzo 1995 6 209 4 210 1.47 [0.43; 4.95] 19.9% 19.9%
Charpentier 2001 20 150 8 75 1.25 [0.53; 2.92] 40.7% 40.7%
Tosi 2003 2 20 2 19 0.95 [0.15; 6.14] 8.4% 8.4%
Hermann 1994 12 34 8 38 1.99 [0.71; 5.57] 28.0% 28.0%
Goldstein 2003 0 84 1 75 0.29 [0.01; 7.32] 2.9% 2.9%
Fixed effect model 497 417 1.38 [0.80; 2.37] 100% --
Random effects model 1.38 [0.80; 2.37] -- 100%
Heterogeneity: I-squared=0%, tau-squared=0, p=0.8087
0.2 0.5 1 2 5
Odds Ratio

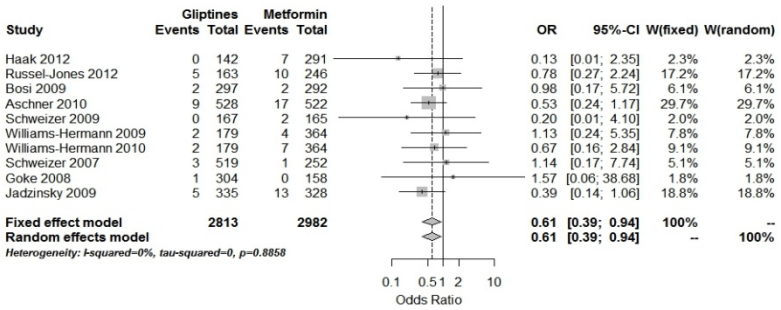
B-E


Study
Repaglinide Events Total
Metformin Events Total
OR 95%-CI W(fixed) W(random)
Moses 1999 3 29 0 27 7.26 [0.36; 147.49] 4.1% 4.1%
Lund 2007 46 89 23 92 3.16 [1.69; 5.90] 95.9% 95.9%
Fixed effect model 118 119 3.27 [1.78; 6.03] 100% --
Random effects model 3.27 [1.78; 6.03] -- 100%
Heterogeneity: I-squared=0%, tau-squared=0, p=0.5959
0.75
Odds Ratio

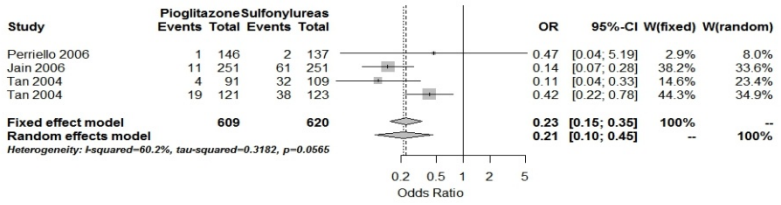
B-F


Study
alpha Glu. Inh. Events Total
Metformin Events Total
OR 95%-CI W(fixed) W(random)
Chiasson 2001 7 82 8 83 0.88 [0.31; 2.48] 93.6% 93.6%
Hoffmann 1997 0 32 0 32 1.00 [0.02; 51.93] 6.4% 6.4%
Fixed effect model 114 115 0.89 [0.33; 2.42] 100% --
Random effects model 0.89 [0.33; 2.42] -- 100%
Heterogeneity: I-squared=0%, tau-squared=0, p=0.9521
0.1 0.5 1 2 10
Odds Ratio

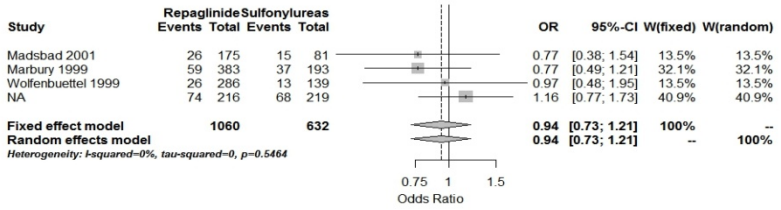
B-G


Gliptines
Metformin
Study Events Total Events Total OR 95%-CI W(fixed) W(random)
Haak 2012 0 142 7 291 0.13 [0.01; 2.35] 2.3% 2.3%
Russel-Jones 2012 5 163 10 246 0.78 [0.27; 2.24] 17.2% 17.2%
Bosi 2009 2 297 2 292 0.98 [0.17; 5.72] 6.1% 6.1%
Aschner 2010 9 528 17 522 0.53 [0.24; 1.17] 29.7% 29.7%
Schweizer 2009 0 167 2 165 0.20 [0.01; 4.10] 2.0% 2.0%
Williams-Hermann 2009 2 179 4 364 1.13 [0.24; 5.35] 7.8% 7.8%
Williams-Hermann 2010 2 179 7 364 0.67 [0.16; 2.84] 9.1% 9.1%
Schweizer 2007 3 519 1 252 1.14 [0.17; 7.74] 5.1% 5.1%
Goke 2008 1 304 0 158 1.57 [0.06; 38.68] 1.8% 1.8%
Jadzinsky 2009 5 335 13 328 0.39 [0.14; 1.06] 18.8% 18.8%
Fixed effect model 2813 2982 0.61 [0.39; 0.94] 100% --
Random effects model 0.61 [0.39; 0.94] -- 100%
Heterogeneity: I-squared=0%, tau-squared=0, p=0.8858
0.1 0.5 1 2 10
Odds Ratio

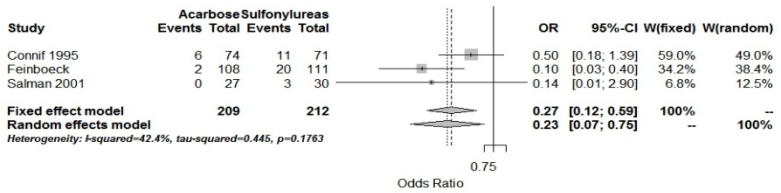
C-D


Pioglitazone
Sulfonylureas
Study Events Total Events Total OR 95%-CI W(fixed) W(random)
Perriello 2006 1 146 2 137 0.47 [0.04; 5.19] 2.9% 8.0%
Jain 2006 11 251 61 251 0.14 [0.07; 0.28] 38.2% 33.6%
Tan 2004 4 91 32 109 0.11 [0.04; 0.33] 14.6% 23.4%
Tan 2004 19 121 38 123 0.42 [0.22; 0.78] 44.3% 34.9%
Fixed effect model 609 620 0.23 [0.15; 0.35] 100% --
Random effects model 0.21 [0.10; 0.45] -- 100%
Heterogeneity: I-squared=60.2%, tau-squared=0.3182, p=0.0565
0.2 0.5 1 2 5
Odds Ratio

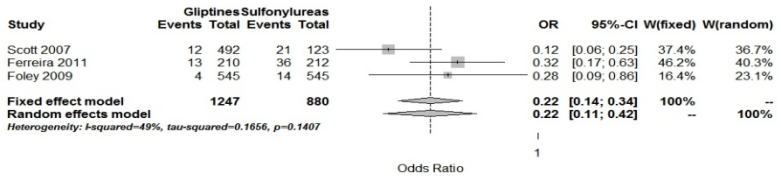
C-E


Repaglinide
Sulfonylureas
Study Events Total Events Total OR 95%-CI W(fixed) W(random)
Madsbad 2001 26 175 15 81 0.77 [0.38; 1.54] 13.5% 13.5%
Marbury 1999 59 383 37 193 0.77 [0.49; 1.21] 32.1% 32.1%
Wolfenbuettel 1999 26 286 13 139 0.97 [0.48; 1.95] 13.5% 13.5%
NA 74 216 68 219 1.16 [0.77; 1.73] 40.9% 40.9%
Fixed effect model 1060 632 0.94 [0.73; 1.21] 100% --
Random effects model 0.94 [0.73; 1.21] -- 100%
Heterogeneity: I-squared=0%, tau-squared=0, p=0.5464
0.75 1 1.5
Odds Ratio

C-F



C-G



D-G

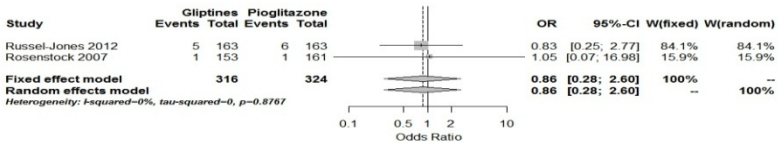
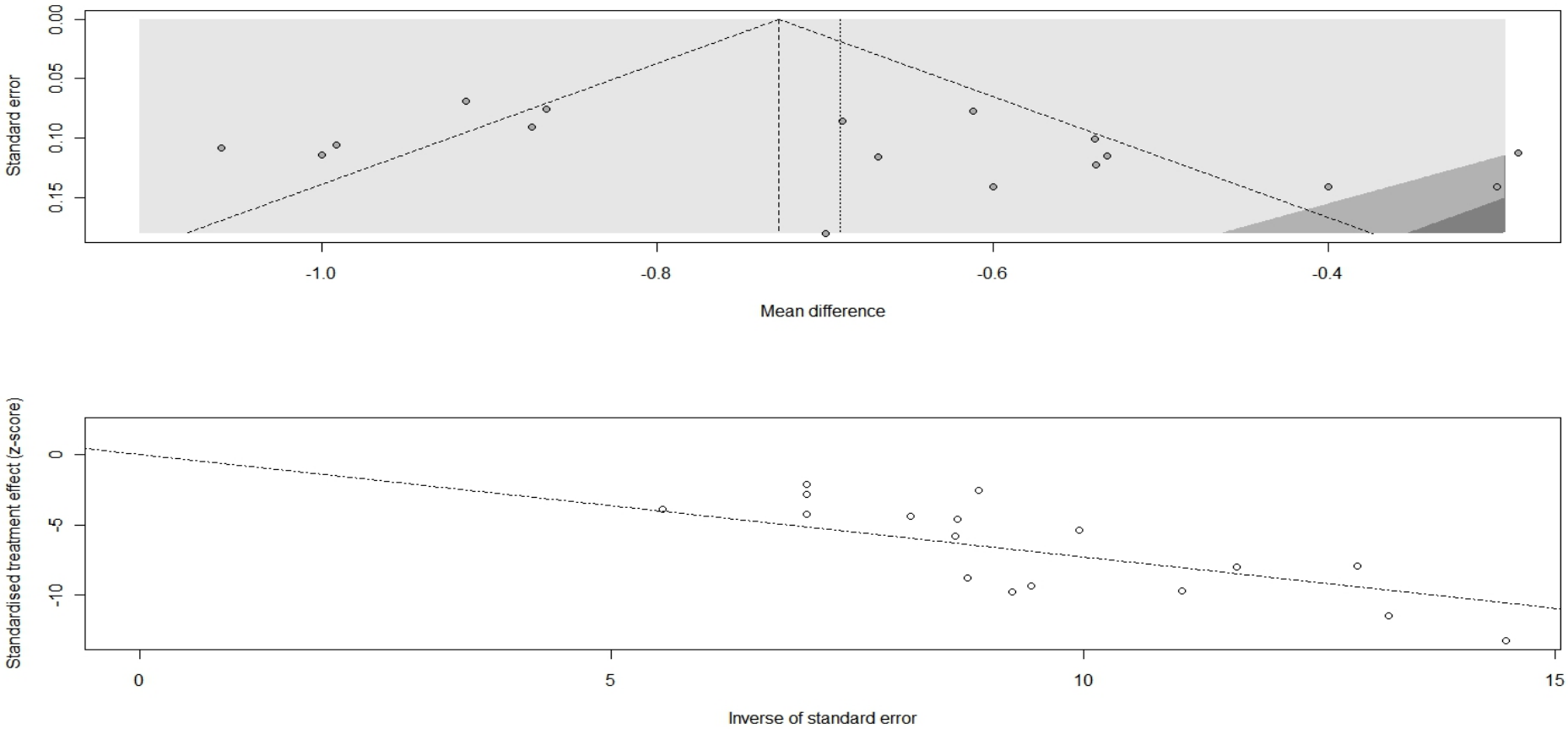


Figure 8: Forest plots of direct pairwise frequentist meta- analysis
For a.) HbA1c b.) body weight and c.) hypoglycemia. No forest plots for comparisons D-E/D-F/E-G and A-E/B-D/E-F/F-G since no direct evicende is available and direct evidence is only supported by one study, respectively.

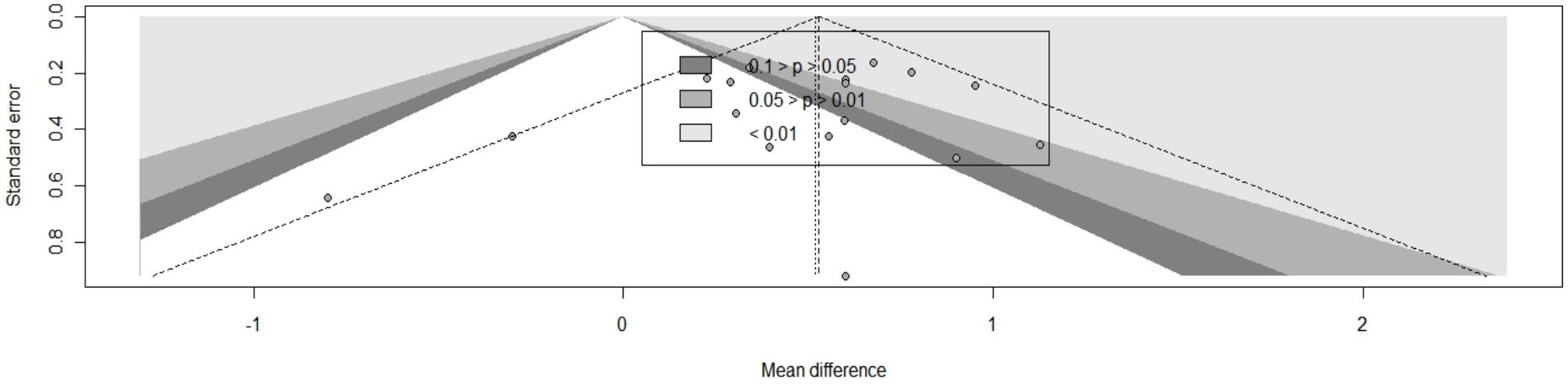
B.2 Figure 9

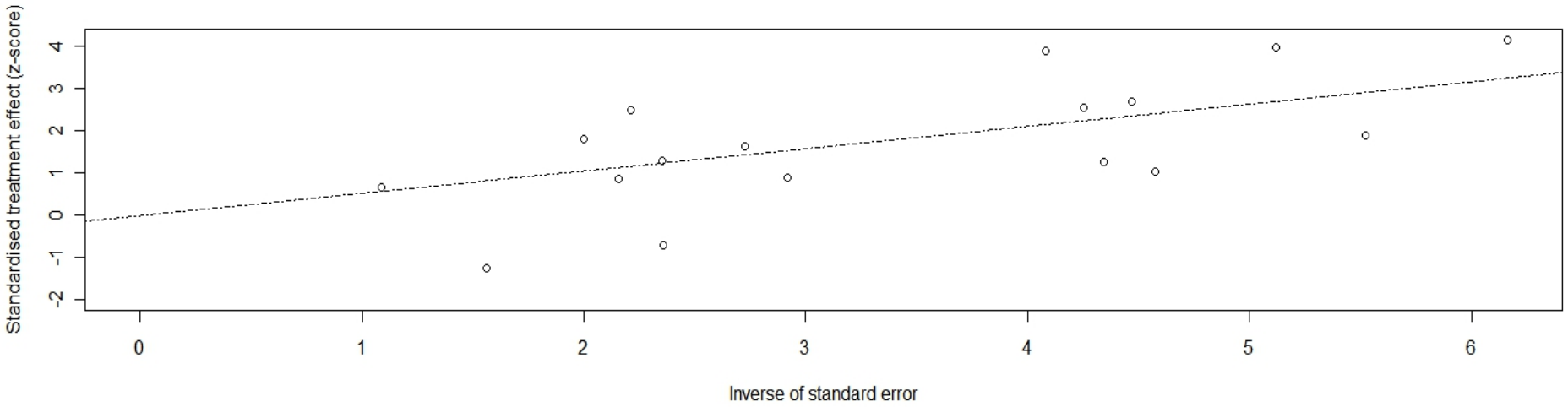
a.) Gliptins vs. placebo (A-G)

HbA1c

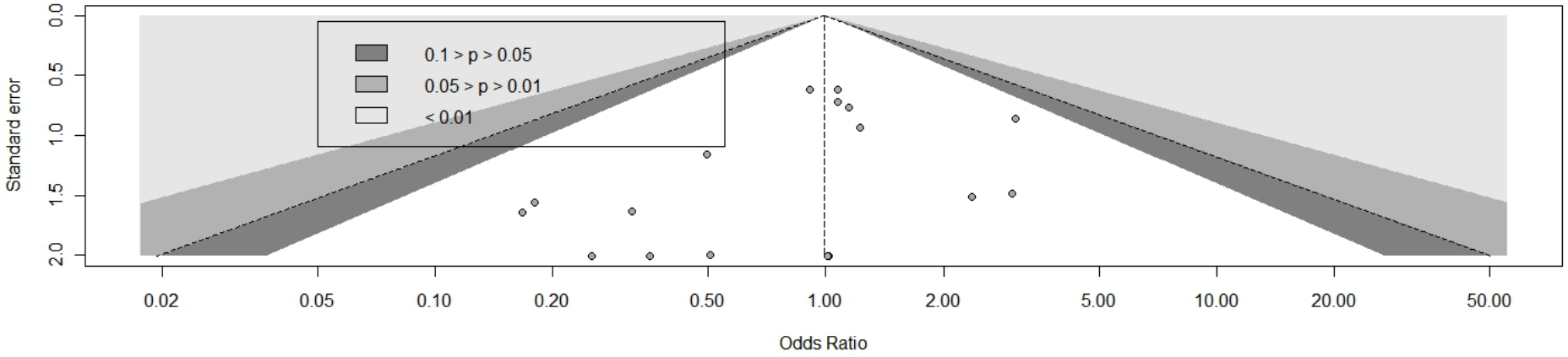


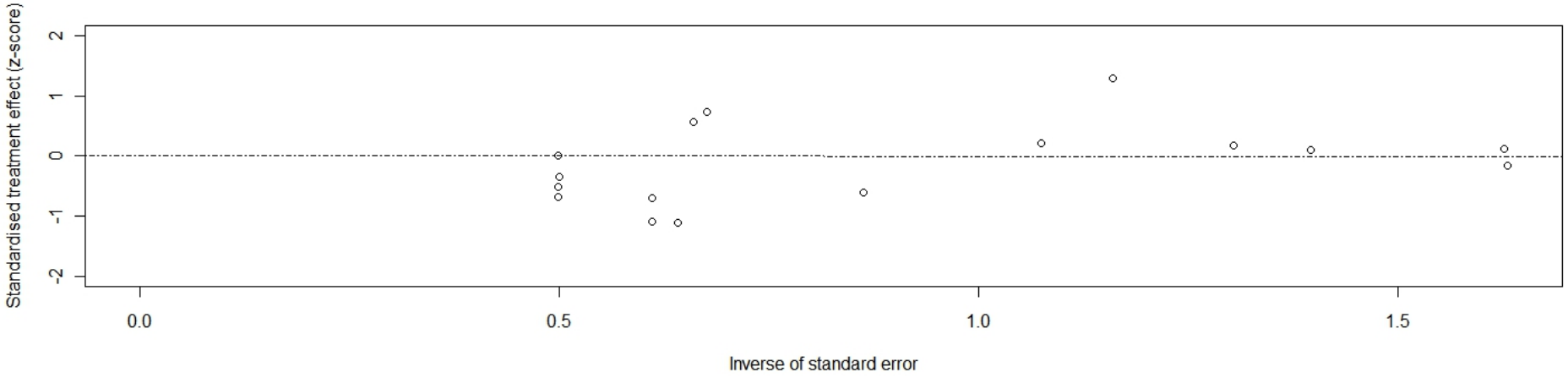
Body weight


Standard error
0.0
0.2
0.4
0.6
0.8
0.1 > p > 0.05
0.05 > p > 0.01
< 0.01
-1
0
1
2
Mean difference


Standardised treatment effect (z-score)
-2
-1
0
1
2
3
4
0
1
2
3
4
5
6
Inverse of standard error

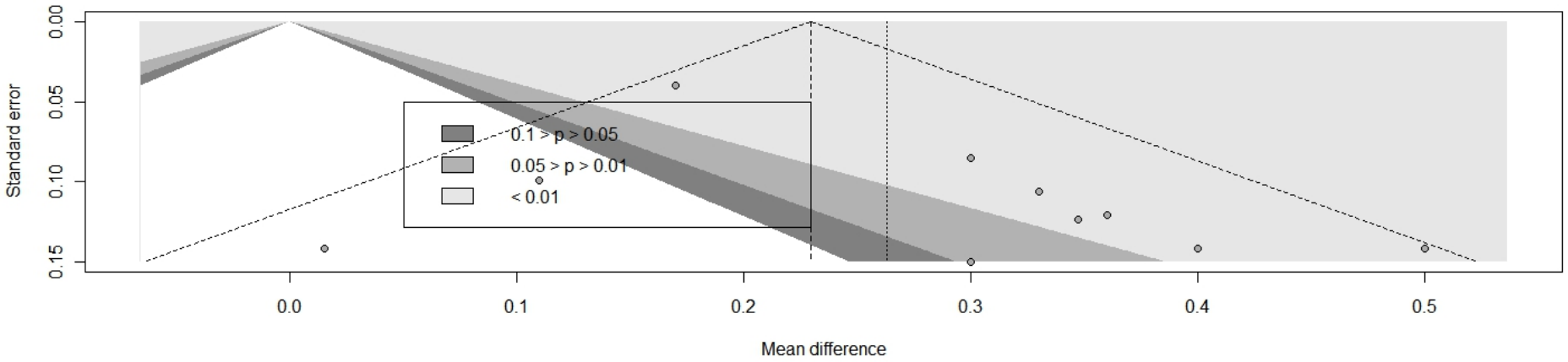
Hypoglycemia

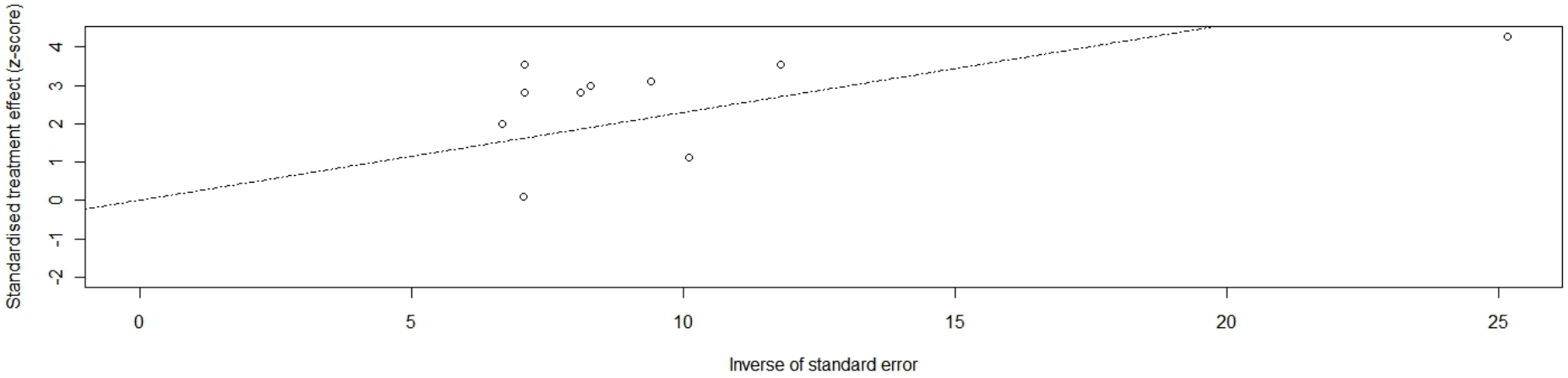




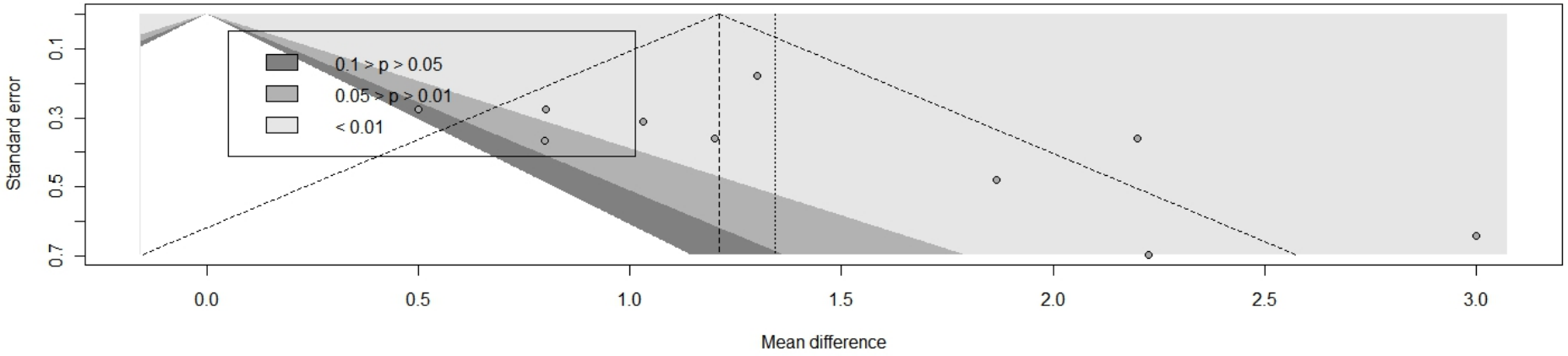
b.) Gliptins vs. Metformin (B-G)

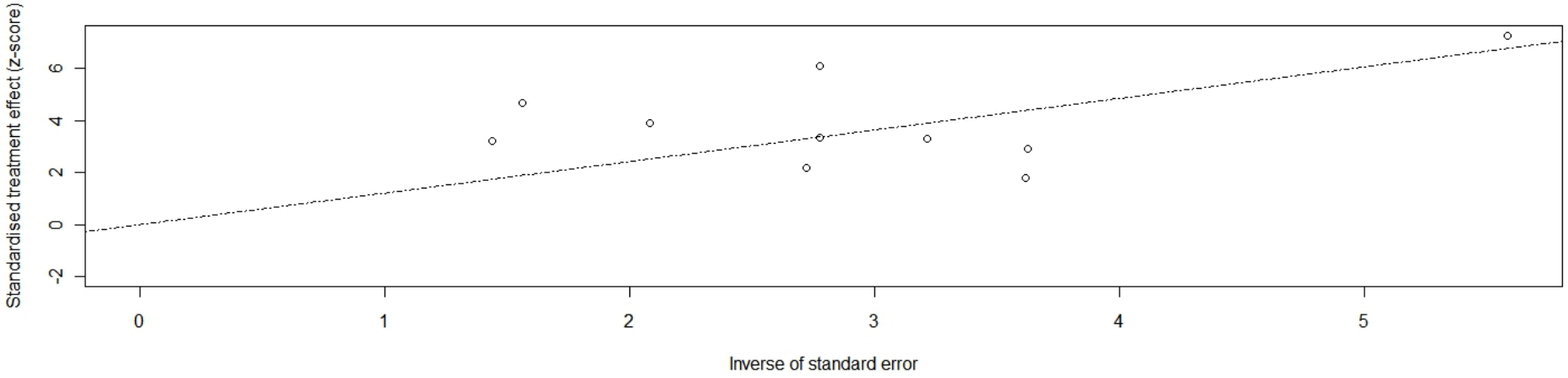
HbA1c





Body weight


0.1 > p > 0.05
0.05 > p > 0.01
< 0.01
Standard error
0.1
0.3
0.5
0.7
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Mean difference


Standardised treatment effect (z-score)
-2
0
2
4
6
0
1
2
3
4
5
Inverse of standard error

Hypoglycemia

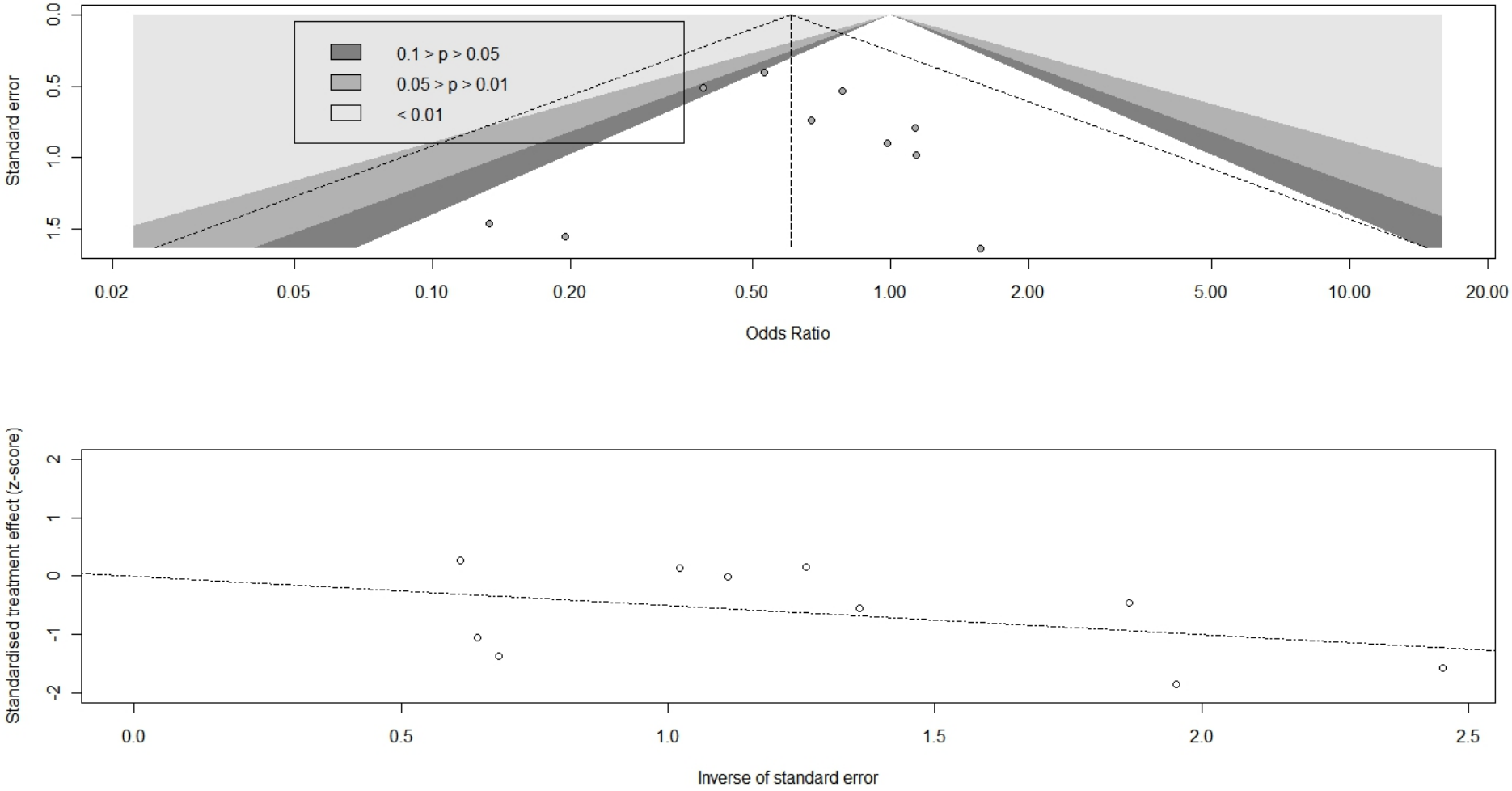
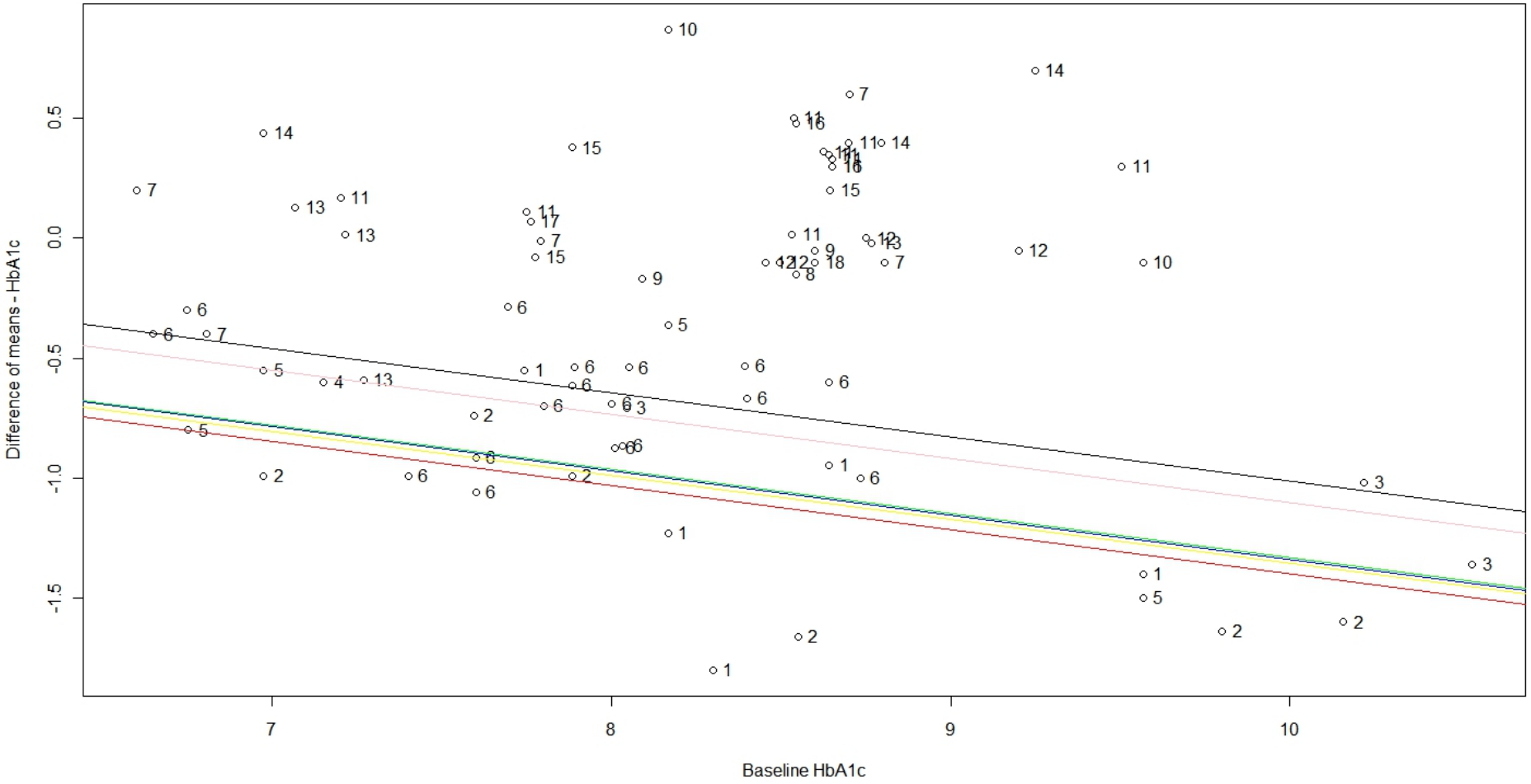


Figure 9: Contour enhanced funnel plots and test for funnel plot asymmetry
For the summary estimates of a.) gliptins vs. placebo (A-G) and b.) gliptins vs. Metformin (B-G) for HbA1c, body weight and hypoglycemia, respectively.

B.3 Figure 11

a.)



b.)

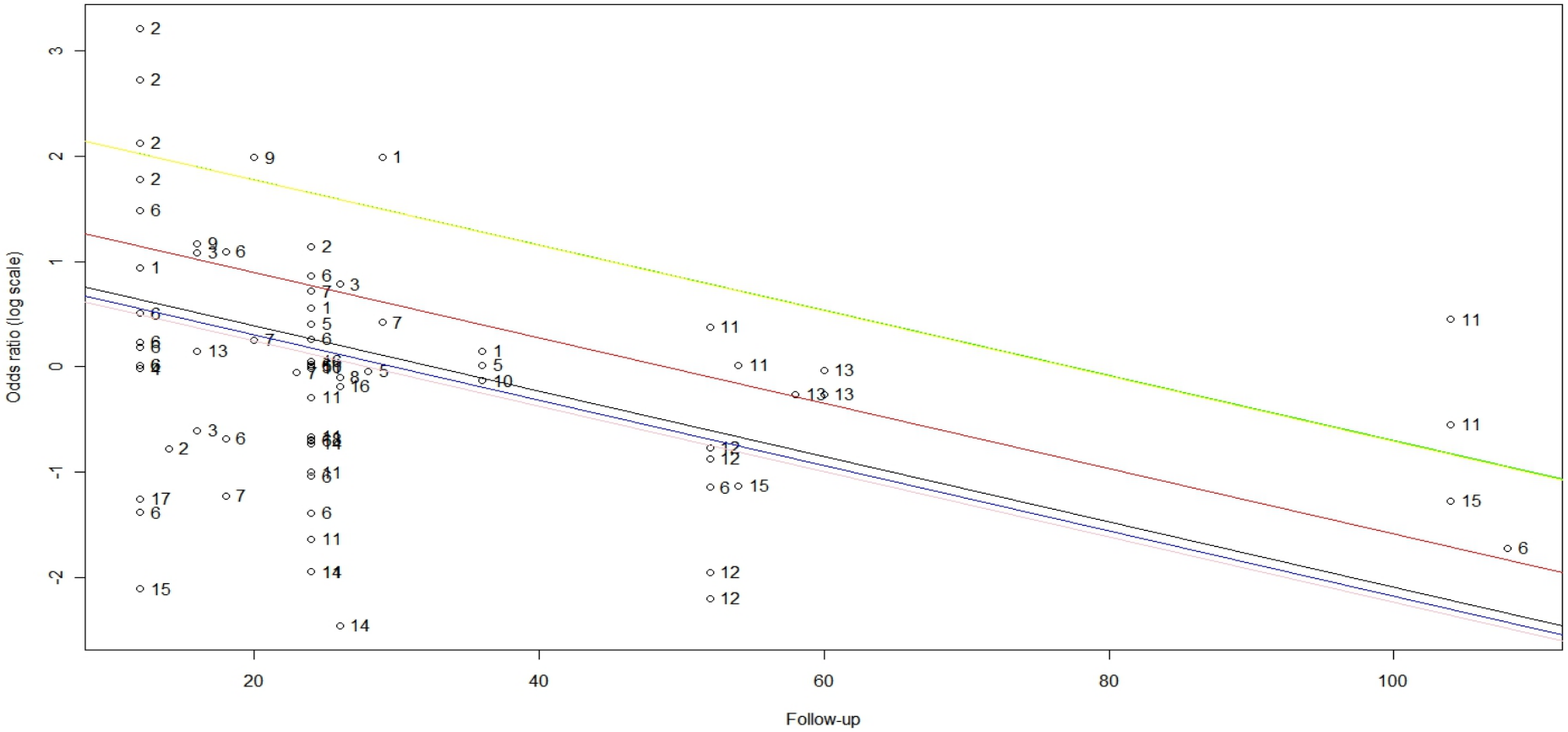
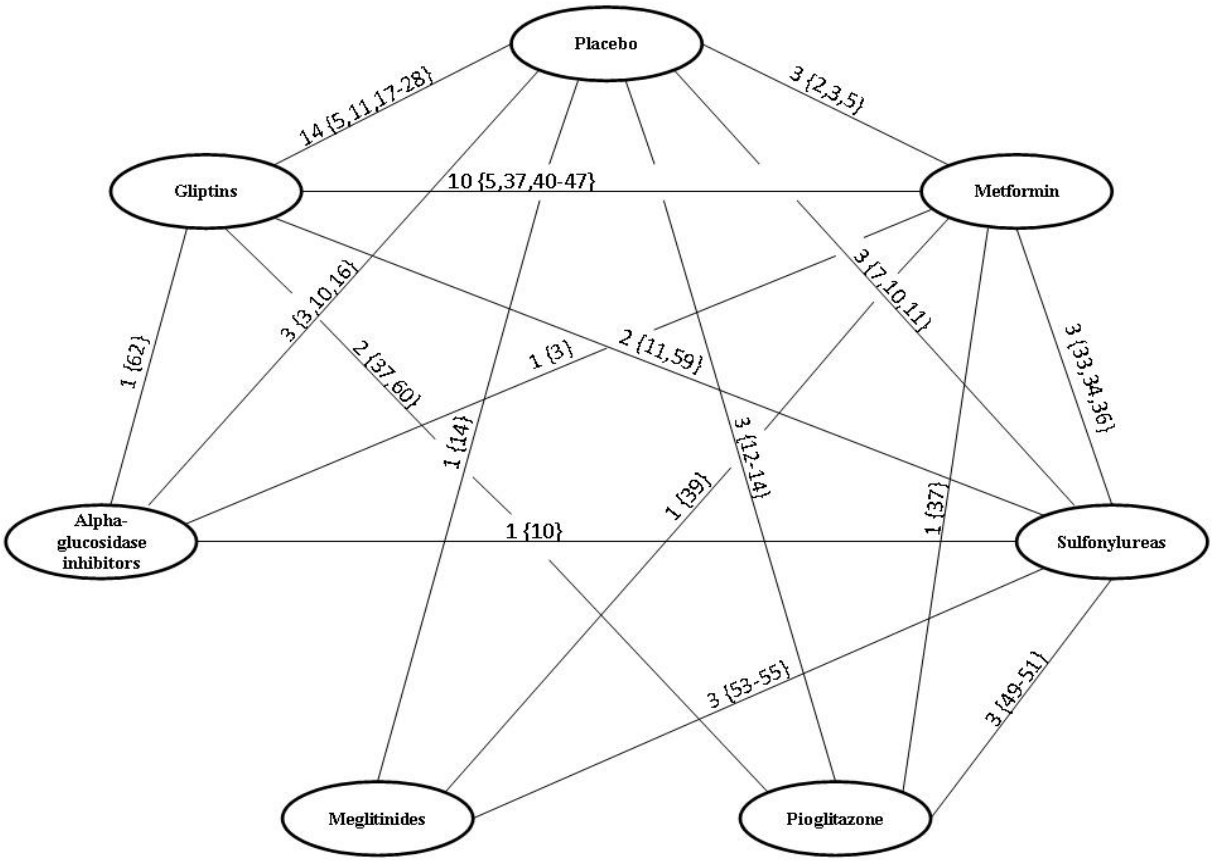


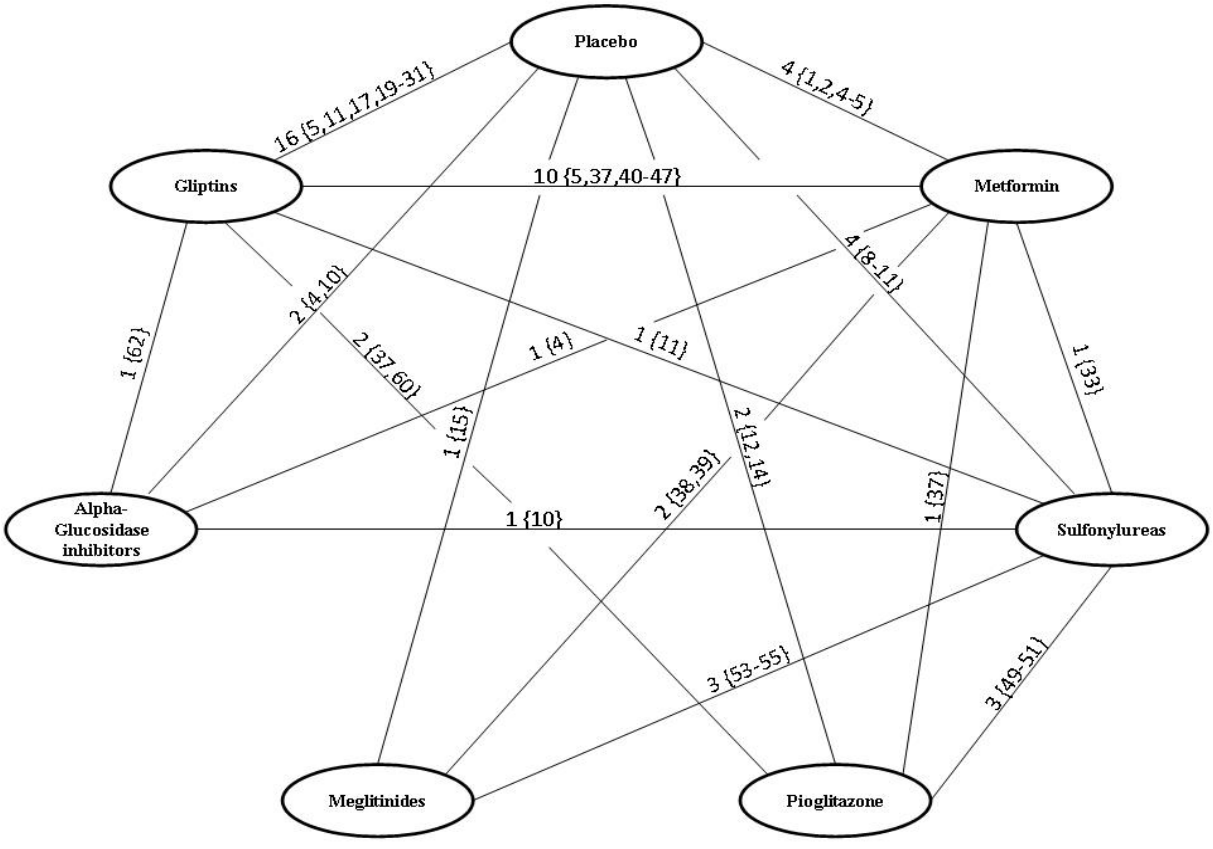
Figure 11: Covariate vs. outcome plots
Plot of a.) the difference in means against the baseline HbA1c and b.) the log odds ratios against the length of follow-up of the crude estimates of direct comparisons. The plotted numbers refer to observations comparing contrasts as follows: 1=A-B, 2=A-C, 3=A-D, 4=A-E, 5=A-F, 6=A-G, 7=B-C, 8=B-D, 9=B-E, 10=B-F, 11=B-G, 12=C-D, 13=C-E, 14=C-F, 15=C-G, 16=D-G, 17=E-F, 18=F-G. The coloured lines represent the relative effects of the treatments compared to placebo as follows: Metformin (red), sulfonylureas (green), Pioglitazone (blue), meglitinides (yellow), alpha-glucosidase inhibitors (black) and gliptins (pink).

B.4 Figure 12

a.) HbA1c



b) Body weight



c.) Hypoglycemia

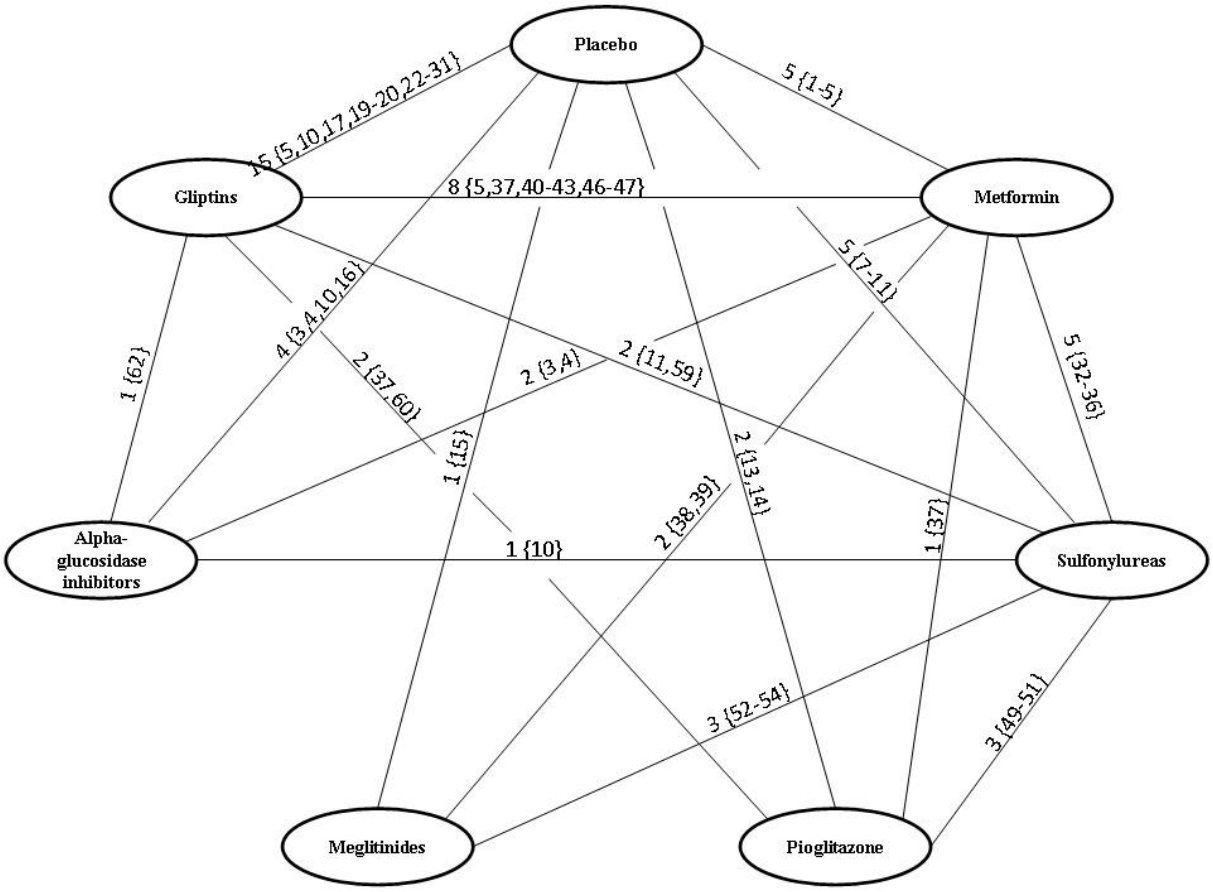


Figure 12: Network diagrams for sensitivity analysis for the risk of bias
For a.) HbA1c, b.) body weight and c.) hypoglycemia without possibly biased studies.
HbA1c, glycated haemoglobin.

References

1. Nationale Versorgungsleitlinie Therapieplanung bei Typ-2-Diabetes. Konsultationsfassung. Version 1.0. 29. August 2012: Available from: http://www.versorgungsleitlinien.de/methodik/nvl-archiv/vorversionen-dm-therapie/nvl-dm-therapie-lang-konsultation.pdf.
2. Diagnosis and classification of diabetes mellitus. Diabetes Care 2011 Jan;34 Suppl 1:S62-9.
3. Panten U RI. Pharmakologie des Glukosestoffwechsels. Antihyperglykämische Pharmakotherapie des Diabetes Mellitus. In: Aktories K FU, Hofmann F, Starke K., editor. Allgemeine und spezielle Pharmakologie und Toxikologie. 9. Auflage ed. München: Urban und Fischer Verlag; 2009. p. p. 617-36.
4. van Dieren S, Beulens JW, van der Schouw YT, Grobbee DE, Neal B. The global burden of diabetes and its complications: an emerging pandemic. Eur J Cardiovasc Prev Rehabil 2010 May;17 Suppl 1:S3-8.
5. Centers for Disease Control and Prevention. Diabetes program. Diabetes data & trends.: Available from: http://www.cdc.gov/diabetes/statistics/age/fig1.htm.
6. Arzneimittelkommission der deutschen Ärzteschaft. Empfehlungen zur Therapie zur antihyperglykämischen Therapie des Diabetes Mellitus Typ 2. 2. Auflage, März 2009. Langversion: Available from: http://www.akdae.de/Arzneimitteltherapie/TE/A-Z/PDF/Diabetes2.pdf#page=1&view=fitB.
7. Wilke T, Ahrendt P, Schwartz D, Linder R, Ahrens S, Verheyen F. [Incidence and prevalence of type 2 diabetes mellitus in Germany: an analysis based on 5.43 million patients]. Dtsch Med Wochenschr 2013 Jan;138(3):69-75.
8. Schulze MB, Hoffmann K, Boeing H, Linseisen J, Rohrmann S, Mohlig M, et al. An accurate risk score based on anthropometric, dietary, and lifestyle factors to predict the development of type 2 diabetes. Diabetes Care 2007 Mar;30(3):510-5.
9. Haffner SM, Lehto S, Ronnemaa T, Pyorala K, Laakso M. Mortality from coronary heart disease in subjects with type 2 diabetes and in nondiabetic subjects with and without prior myocardial infarction. N Engl J Med 1998 Jul 23;339(4):229-34.
10. Almdal T, Scharling H, Jensen JS, Vestergaard H. The independent effect of type 2 diabetes mellitus on ischemic heart disease, stroke, and death: a population-based study of 13,000 men and women with 20 years of follow-up. Arch Intern Med 2004 Jul 12;164(13):1422-6.
11. Colhoun HM, Lee ET, Bennett PH, Lu M, Keen H, Wang SL, et al. Risk factors for renal failure: the WHO Mulinational Study of Vascular Disease in Diabetes. Diabetologia 2001 Sep;44 Suppl 2:S46-53.
12. Zimmet P. The burden of type 2 diabetes: are we doing enough? Diabetes Metab 2003 Sep;29(4 Pt 2):6S9-18.
13. Nordrheinische Gemeinsame Einrichtung Disease-Management-Programme. Qualitätssicherungsbericht 2010. Disease-Management-Programme in Nordrhein. Brustkrebs, Diabetes mellitus Typ 1/Typ 2, Koronare Herzkrankheit, Asthma/COPD. Düsseldorf: Nordrheinische Gemeinsame Einrichtung DMP; 2011. : Available from: http://www.kvno.de/downloads/quali/qualbe_dmp11.pdf.
14. Schunk M, Reitmeir P, Schipf S, Volzke H, Meisinger C, Thorand B, et al. Health-related quality of life in subjects with and without Type 2 diabetes: pooled analysis of five population-based surveys in Germany. Diabetic medicine : a journal of the British Diabetic Association 2012 May;29(5):646-53.
15. Panzram G. Mortality and survival in type 2 (non-insulin-dependent) diabetes mellitus. Diabetologia 1987 Mar;30(3):123-31.
16. Perna L, Thien-Seitz U, Ladwig KH, Meisinger C, Mielck A. Socio-economic differences in life expectancy among persons with diabetes mellitus or myocardial infarction: results from the German MONICA/KORA study. BMC Public Health 2010;10:135.

17. Ray JA, Valentine WJ, Secnik K, Oglesby AK, Cordony A, Gordois A, et al. Review of the cost of diabetes complications in Australia, Canada, France, Germany, Italy and Spain. Current Medical Research and Opinion 2005 Oct;21(10):1617-29.
18. Brown JB, Pedula KL, Bakst AW. The progressive cost of complications in type 2 diabetes mellitus. Arch Intern Med 1999 Sep 13;159(16):1873-80.
19. Waldeyer R, Brinks R, Rathmann W, Giani G, Icks A. Projection of the burden of Type 2 diabetes mellitus in Germany: a demographic modelling approach to predict the direct medical excess costs from 2010 to 2040. Diabetic medicine : a journal of the British Diabetic Association 2013 Mar 18.
20. Liebl A. [Costs involved in the early and late phases of diabetes mellitus]. Internist (Berl) 2007 Jul;48(7):708-14.
21. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med 2002 Feb 7;346(6):393-403.
22. Lindstrom J, Ilanne-Parikka P, Peltonen M, Aunola S, Eriksson JG, Hemio K, et al. Sustained reduction in the incidence of type 2 diabetes by lifestyle intervention: follow-up of the Finnish Diabetes Prevention Study. Lancet 2006 Nov 11;368(9548):1673-9.
23. Canadian Diabetes Association 2013 Clinical Practice Guidelines for the Prevention and Management of Diabetes in Canada: pharmacologic management of type 2 diabetes.: Available from: http://guidelines.diabetes.ca/Browse/Chapter13.
24. Institute for Quality and Efficiency in Health Care. Urine and blood glucose self-measurement in diabetes mellitus type 2. Final report.: Available from: https://www.iqwig.de/download/A05-08_Abschlussbericht_Zuckerselbstmessung_bei_Diabetes_mellitus_Typ_2.pdf.
25. Federal Joint Committee. Arzneimittelrichtlinie/ Anlage III. Harn- und Blutzuckerteststreifen bei Diabetes mellitus Typ 2. : Available from: http://www.g-ba.de/downloads/39-261-1307/2011-03-17_AM-RL3_Blutzuckerteststreifen_BAnz.pdf.
26. Adler AI, Stratton IM, Neil HA, Yudkin JS, Matthews DR, Cull CA, et al. Association of systolic blood pressure with macrovascular and microvascular complications of type 2 diabetes (UKPDS 36): prospective observational study. Bmj 2000 Aug 12;321(7258):412-9.
27. Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes: UKPDS 38. UK Prospective Diabetes Study Group. Bmj 1998 Sep 12;317(7160):703-13.
28. Turner RC, Millns H, Neil HA, Stratton IM, Manley SE, Matthews DR, et al. Risk factors for coronary artery disease in non-insulin dependent diabetes mellitus: United Kingdom Prospective Diabetes Study (UKPDS: 23). Bmj 1998 Mar 14;316(7134):823-8.
29. Cushman WC, Evans GW, Byington RP, Goff DC, Jr., Grimm RH, Jr., Cutler JA, et al. Effects of intensive blood-pressure control in type 2 diabetes mellitus. N Engl J Med 2010 Apr 29;362(17):1575-85.
30. Ginsberg HN, Elam MB, Lovato LC, Crouse JR, 3rd, Leiter LA, Linz P, et al. Effects of combination lipid therapy in type 2 diabetes mellitus. N Engl J Med 2010 Apr 29;362(17):1563-74.
31. Inzucchi SE, Bergenstal RM, Buse JB, Diamant M, Ferrannini E, Nauck M, et al. Management of hyperglycemia in type 2 diabetes: a patient-centered approach: position statement of the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetes Care 2012 Jun;35(6):1364-79.
32. Stratton IM, Adler AI, Neil HA, Matthews DR, Manley SE, Cull CA, et al. Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. Bmj 2000 Aug 12;321(7258):405-12.
33. Patel A, MacMahon S, Chalmers J, Neal B, Billot L, Woodward M, et al. Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes. N Engl J Med 2008 Jun 12;358(24):2560-72.
34. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). UK Prospective Diabetes Study (UKPDS) Group. Lancet 1998 Sep 12;352(9131):837-53.

35. Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). UK Prospective Diabetes Study (UKPDS) Group. Lancet 1998 Sep 12;352(9131):854-65.
36. Gerstein HC, Miller ME, Byington RP, Goff DC, Jr., Bigger JT, Buse JB, et al. Effects of intensive glucose lowering in type 2 diabetes. N Engl J Med 2008 Jun 12;358(24):2545-59.
37. Van de Laar FA, Lucassen PL, Akkermans RP, Van de Lisdonk EH, Rutten GE, Van Weel C. Alpha-glucosidase inhibitors for type 2 diabetes mellitus. Cochrane Database Syst Rev 2005(2):CD003639.
38. Richter B, Bandeira-Echtler E, Bergerhoff K, Lerch CL. Dipeptidyl peptidase-4 (DPP-4) inhibitors for type 2 diabetes mellitus. Cochrane Database Syst Rev 2008(2):CD006739.
39. Neumiller JJ, Setter SM. Review of linagliptin for the treatment of type 2 diabetes mellitus. Clinical Therapeutics 2012 May;34(5):993-1005.
40. Kania DS, Gonzalvo JD, Weber ZA. Saxagliptin: a clinical review in the treatment of type 2 diabetes mellitus. Clinical Therapeutics 2011 Aug;33(8):1005-22.
41. Gliptine zur Behandlung von Patienten mit Diabetes mellitus Typ 2. Der Arzneimittelbrief [serial on the Internet]. 2013; 47(19): Available from: http://www.der-arzneimittelbrief.de/de/Artikel.aspx?J=2013&S=40.
42. Vague P. Is metformin more than an oral hypoglycaemic agent? Diabetes Metab 2003 Sep;29(4 Pt 2):6S5-7.
43. Saenz A, Fernandez-Esteban I, Mataix A, Ausejo M, Roque M, Moher D. Metformin monotherapy for type 2 diabetes mellitus. Cochrane Database Syst Rev 2005(3):CD002966.
44. Lipska KJ, Bailey CJ, Inzucchi SE. Use of metformin in the setting of mild-to-moderate renal insufficiency. Diabetes Care 2011 Jun;34(6):1431-7.
45. Kahn SE, Haffner SM, Heise MA, Herman WH, Holman RR, Jones NP, et al. Glycemic durability of rosiglitazone, metformin, or glyburide monotherapy. N Engl J Med 2006 Dec 7;355(23):2427-43.
46. Hirst JA, Farmer AJ, Dyar A, Lung TW, Stevens RJ. Estimating the effect of sulfonylurea on HbA1c in diabetes: a systematic review and meta-analysis. Diabetologia 2013 May;56(5):973-84.
47. Holman RR, Paul SK, Bethel MA, Matthews DR, Neil HA. 10-year follow-up of intensive glucose control in type 2 diabetes. N Engl J Med 2008 Oct 9;359(15):1577-89.
48. Black C, Donnelly P, McIntyre L, Royle PL, Shepherd JP, Thomas S. Meglitinide analogues for type 2 diabetes mellitus. Cochrane Database Syst Rev 2007(2):CD004654.
49. Dormandy JA, Charbonnel B, Eckland DJ, Erdmann E, Massi-Benedetti M, Moules IK, et al. Secondary prevention of macrovascular events in patients with type 2 diabetes in the PROactive Study (PROspective pioglitAzone Clinical Trial In macroVascular Events): a randomised controlled trial. Lancet 2005 Oct 8;366(9493):1279-89.
50. Erdmann E, Dormandy JA, Charbonnel B, Massi-Benedetti M, Moules IK, Skene AM. The effect of pioglitazone on recurrent myocardial infarction in 2,445 patients with type 2 diabetes and previous myocardial infarction: results from the PROactive (PROactive 05) Study. J Am Coll Cardiol 2007 May 1;49(17):1772-80.
51. Food and Drug administration. Drug safety communication Pioglitazone. : Available from: http://www.fda.gov/Drugs/DrugSafety/ucm226214.htm.
52. Plosker GL. Dapagliflozin: a review of its use in type 2 diabetes mellitus. Drugs 2012 Dec 3;72(17):2289-312.
53. European Medicines Agency. European Public Assessment Report. Dapagliflozin.: Available from: http://www.emea.europa.eu/ema/index.jsp?curl=pages/medicines/human/medicines/002322/human_med_001546.jsp&mid=WC0b01ac058001d124.
54. Food and Drug Administration. Dapagliflozin Advisory Committee Meeting: Available from: http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/EndocrinologicandMetabolicDrugsAdvisoryCommittee/UCM262990.pdf.
55. Neu auf dem Markt. Diabetesmittel Dapagliflozin (Forxiga).
. arznei-telegramm [serial on the Internet]. 2013; 44(1/13): Available from: http://www.arznei-telegramm.de/html/2013_01/test2.php3?&knr=200326/406430&produktid=001_01.

56. National Institute of Clinical Excellence.Clinical guideline 87. Type 2 diabetes: The management of type 2 diabetes: Available from: http://publications.nice.org.uk/type-2-diabetes-cg87.
57. International Diabetes Federation. Global Guideline for Type 2 Diabetes, 2012.: Available from: http://www.idf.org/sites/default/files/IDF-Guideline-for-Type-2-Diabetes.pdf.
58. Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. Bmj 2005 Oct 15;331(7521):897-900.
59. Jansen JP, Fleurence R, Devine B, Itzler R, Barrett A, Hawkins N, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. Value Health 2011 Jun;14(4):417-28.
60. Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. Res Synth Method 2012(3):80-97.
61. Sutton A, Ades AE, Cooper N, Abrams K. Use of indirect and mixed treatment comparisons for technology assessment. Pharmacoeconomics 2008;26(9):753-67.
62. Dias, S., Sutton, A.J., Welton, N.J., Ades, A.E. NICE DSU Technical Support Document 3:
Heterogeneity: subgroups, meta-regression, bias and bias-adjustment. 2011; last updated April 2012: Available from:
http://www.nicedsu.org.uk/TSD3%20Heterogeneity.final%20report.08.05.12.pdf.
63. The Cochrane Comparing Multiple Interventions Group: Available from: http://cmimg.cochrane.org/.
64. National Institute of Clinical Excellence. Guide to the methods of technology appraisal 2013.
65. National Institute of Clinical Excellence. Decision Support Unit. : Available from: http://www.nicedsu.org.uk/index.htm.
66. Agency for Healthcare Research and Quality. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. Prepublication Draft Copy.: Available from: http://effectivehealthcare.ahrq.gov/ehc/products/60/318/MethodsGuide_Prepublication-Draft_20120523.pdf.
67. Canadian Agency for Drugs and Technologies in Health. Indirect Evidence: Indirect Treatment Comparisons in Meta-Analysis. Technology Report. : Available from: http://www.cadth.ca/media/pdf/H0462_itc_tr_e.pdf.
68. Pharmaceutical Benefits Adivsory Committee Australia. Guidelines for preparing submissions to the PBAC - December 2008. Version 4.3.: Available from: http://www.health.gov.au/internet/main/publishing.nsf/content/pbacguidelines-index.
69. Federal Joint Committee. Benefit assessment of pharmaceuticals. : Available from: http://www.english.g-ba.de/benefitassessment/information/.
70. Institute for Quality and Efficiency in Health Care. General methods benefit assessment 4.0.: Available from: https://www.iqwig.de/download/General_Methods_4-0.pdf.
71. Institute for Quality and Efficiency in Health Care. General Methods for the Assessment of the Relation of Benefits to Costs.: Available from:
https://www.iqwig.de/download/General_Methods_for_the_Assessment_of_the_Relation_of_Benefits_to_Costs.pdf.
72. McIntosh B, Cameron C, Singh SR, Yu C, Ahuja T, Welton NJ, et al. Second-line therapy in patients with type 2 diabetes inadequately controlled with metformin monotherapy: a systematic review and mixed-treatment comparison meta-analysis. Open Med 2011;5(1):e35-48.
73. Gross JL, Kramer CK, Leitao CB, Hawkins N, Viana LV, Schaan BD, et al. Effect of antihyperglycemic agents added to metformin and a sulfonylurea on glycemic control and weight gain in type 2 diabetes: a network meta-analysis. Ann Intern Med 2011 May 17;154(10):672-9.
74. Jabbour S, Ziring B. Advantages of extended-release metformin in patients with type 2 diabetes mellitus. Postgrad Med 2011 Jan;123(1):15-23.

75. . In: Higgins JP, Green S, editors. The Cochrane Handbook for Systematic Reviews of Interventions Version 510, updated March, 2011.
76. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. Open Med 2009;3(3):e123-30.
77. Hoaglin DC, Hawkins N, Jansen JP, Scott DA, Itzler R, Cappelleri JC, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2. Value Health 2011 Jun;14(4):429-37.
78. Institute for Quality and Efficiency in Health Care. Health economic evaluation of venlafaxine, duloxetine, bupropion, and mirtazapine compared to further prescribable pharmaceutical treatments. : Available from: Health economic evaluation of venlafaxine, duloxetine, bupropion, and mirtazapine compared to further prescribable pharmaceutical treatments
79. Defining and reporting hypoglycemia in diabetes: a report from the American Diabetes Association Workgroup on Hypoglycemia. Diabetes Care 2005 May;28(5):1245-9.
80. Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. Jama 2004 May 26;291(20):2457-65.
81. Chan AW, Altman DG. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. Bmj 2005 Apr 2;330(7494):753.
82. European Medicines Agency. Guideline on clinical investigation of medicinal products in the treatment or prevention of diabetes mellitus. .
83. ClinicalTrials.gov: Available from: http://clinicaltrials.gov/ct/gui.
84. EudraCT.: Available from: https://eudract.ema.europa.eu/.
85. Eudrapharm. : Available from: http://www.eudrapharm.eu/eudrapharm/welcome.do?selectedStaticLocale.languageCode=de
.86. European Union. Clinical trials register.: Available from: https://www.clinicaltrialsregister.eu/index.html.
87. Astra Zeneca. Clinical Trials website.
88. Bristol Myers Squibb. Clinical trials.: Available from: http://www.bms.com/clinical_trials/Pages/home.aspx.
89. Eli Lilly and Company Clinical Trial Registry: Available from: Eli Lilly and Company Clinical Trial Registry.
90. GlaxoSmithKline. Clinical Study Register.: Available from: http://www.gsk-clinicalstudyregister.com/.
91. Novartis. Clinical trials.: Available from: http://www.novartis.com/innovation/research-development/clinical-trials/index.shtml.
92. Roche. Clinical Trial Protocol Registry and Results Database.: Available from: http://www.roche-trials.com/main.action.
93. European Medicines Agency. European Public Assessment Reports.: Available from: http://www.ema.europa.eu/ema/index.jsp?curl=pages/medicines/landing/epar_search.jsp&mid=WC0b01ac058001d125.
94. Food and Drug Administration. Approved drug products.: Available from: http://www.accessdata.fda.gov/scripts/cder/drugsatfda/.
95. Robinson KA, Dickersin K. Development of a highly sensitive search strategy for the retrieval of reports of controlled trials using PubMed. Int J Epidemiol 2002 Feb;31(1):150-3.
96. Scottish Intercollegiate Guidelines Network. Search filter randomized controlled trials Embase. : Available from: http://www.sign.ac.uk/methodology/filters.html#systematic.
97. Endnote X4.0.2.: Available from: http://endnote.com/.
98. Schmidt CO, Kohlmann T. When to use the odds ratio or the relative risk? Int J Public Health 2008;53(3):165-7.

99. Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. Control Clin Trials 1995 Feb;16(1):62-73.
100. Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. Jama 1999 Sep 15;282(11):1054-60.
101. Hutton JL, Williamson PR. Bias in meta-analysis due to outcome variable selection within studies. Journal of the Royal Statistical Society: Series C (Applied Statistics) 2000;49:359-70.
102. Chan AW, Krleza-Jeric K, Schmid I, Altman DG. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. Cmaj 2004 Sep 28;171(7):735-40.
103. Detsky AS, Naylor CD, O'Rourke K, McGeer AJ, L'Abbe KA. Incorporating variations in the quality of individual randomized trials into meta-analysis. J Clin Epidemiol 1992 Mar;45(3):255-65.
104. Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. Cochrane Database Syst Rev 2009(1):MR000006.
105. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. Bmj 1997 Sep 13;315(7109):629-34.
106. Altman DG, Bland JM. Absence of evidence is not evidence of absence. Bmj 1995 Aug 19;311(7003):485.
107. Jansen JP, Crawford B, Bergman G, Stam W. Bayesian meta-analysis of multiple treatment comparisons: an introduction to mixed treatment comparisons. Value Health 2008 Sep-Oct;11(5):956-64.
108. Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. Methods in health service research. An introduction to bayesian methods in health technology assessment. Bmj 1999 Aug 21;319(7208):508-12.
109. Spiegelhalter DJ AK, Myles JP. Bayesian Approaches to Clinical Trials and Health-Care Evaluation. 1st ed. ed. Cichester: John Wiley & Sons, Ltd; 2004.
110. Gelman A, Rubin DB. Markov chain Monte Carlo methods in biostatistics. Stat Methods Med Res 1996 Dec;5(4):339-55.
111. WinBUGS.: Available from: http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml.
112. L. F, T. K, S. L. Regression: Modelle, Methoden und Anwendungen.: Springer Verlag 2007.
113. Dias, S., Welton, N.J., Sutton, A.J. & Ades, A.E. NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials. 2011; last updated March 2013: Available from: http://www.nicedsu.org.uk/TSD2%20General%20meta%20analysis%20corrected%20Mar2013.pdf.
114. WinBUGS tutorial. Checking convergence.: Available from: http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml.
115. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 2002;64:583-639.
116. The R Project for Statistical Computing. Version 2.15.2.: Available from: http://www.r-project.org/.
117. WinBUGS. User manual.: Available from: http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/manual14.pdf.
118. R2WinBUGS package. Reference manual. : Available from: http://ftp5.gwdg.de/pub/misc/cran/.
119. Multiple-Treatments Meta-Analysis. A Framework for Evaluating and Ranking Multiple Healthcare Technologies. Measuring the fit of an MTM model and compare models. : Available from: http://www.mtm.uoi.gr/.
120. Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. Bmj 2011;342:d549.

121. Higgins JP, Green S, editors. The Cochrane Handbook for Systematic Reviews of Interventions Version 510, updated March, 2011.
122. Deeks JJ HJ. Statistical algorithms in Review Manager 5.2010: Available from: http://ims.cochrane.org/revman/documentation.
123. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986 Sep;7(3):177-88.
124. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002 Jun 15;21(11):1539-58.
125. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. Bmj 2003 Sep 6;327(7414):557-60.
126. Higgins JP, Whitehead A. Borrowing strength from external trials in a meta-analysis. Stat Med 1996 Dec 30;15(24):2733-49.
127. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol 1997 Jun;50(6):683-91.
128. Lu G, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. Journal of the American Statistical Association 2006;101:447-59.
129. Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. Stat Med 2010 Mar 30;29(7-8):932-44.
130. Jansen J. Network meta-analysis of individual and aggregate level data. Res Synth Method 2012;3:177-90.
131. Chaimani A SG. Using network meta-analysis to evaluate the existence of small study effects in a network of interventions. Res Synth Method 2012;3:161-76.
132. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med 2004 Oct 30;23(20):3105-24.
133. University of Bristol. School of Social and Community Medicine. Introduction to Mixed Treatment Comparisons. : Available from: http://www.bris.ac.uk/social-community-medicine/media/mpes/intro-to-mtc.pdf.
134. Song F, Loke YK, Walsh T, Glenny AM, Eastwood AJ, Altman DG. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. Bmj 2009;338:b1147.
135. Dias, S., Welton, N.J., Sutton, A.J., Caldwell, D.M., Lu, G. & Ades, A.E. NICE DSU Technical Support Document 4: Inconsistency in Networks of Evidence Based on Randomised Controlled Trials. 2011; last updated April 2012.: Available from: http://www.nicedsu.org.uk/TSD4%20Inconsistency.final.08.05.12.pdf.
136. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? Stat Med 2002 Jun 15;21(11):1559-73.
137. Simmonds MC, Higgins JP. Covariate heterogeneity in meta-analysis: criteria for deciding between meta-regression and individual patient data. Stat Med 2007 Jul 10;26(15):2982-99.
138. Chapell R, Gould AL, Alexander CM. Baseline differences in A1C explain apparent differences in efficacy of sitagliptin, rosiglitazone and pioglitazone. Diabetes, obesity & metabolism 2009 Nov;11(11):1009-16.
139. Salanti G, Ades AE, Ioannidis JP. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. J Clin Epidemiol 2011 Feb;64(2):163-71.
140. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. Ann Intern Med 2010 Jun 1;152(11):726-32.
141. List JF, Woo V, Morales E, Tang W, Fiedorek FT. Sodium-glucose cotransport inhibition with dapagliflozin in type 2 diabetes. Diabetes Care 2009 Apr;32(4):650-7.
142. Lind M. Incretin therapy and its effect on body weight in patients with diabetes. Prim Care Diabetes 2012 Oct;6(3):187-91.
143. Goke B. Improved glycemic control and lipid profile in a randomized study of pioglitazone compared with acarbose in patients with type 2 diabetes mellitus. Treat Endocrinol 2002;1(5):329-36.

144. Defronzo RA, Goodman AM. Efficacy of metformin in patients with non-insulin-dependent diabetes mellitus. New England Journal of Medicine 1995 31 Aug;333(9):541-9.
145. Chiasson JL, Naditch L. The synergistic effect of miglitol plus metformin combination therapy in the treatment of type 2 diabetes. Diabetes Care 2001 Jun;24(6):989-94.
146. Hoffmann J, Spengler M. Efficacy of 24-week monotherapy with acarbose, metformin, or placebo in dietary-treated NIDDM patients: The essen-II study. American Journal of Medicine 1997 December;103(6):483-90.
147. Haak T, Meinicke T, Jones R, Von Eynatten M, Woerle HJ. Combination of linagliptin and metformin improves glycemic control in type 2 diabetes: A randomized trial with an open-label arm in patients with poor glycemic control. Diabetes 2011;60:A77.
148. Simonson DC, Kourides IA, Feinglos M, Shamoon H, Fischette CT. Efficacy, safety, and dose-response characteristics of glipizide gastrointestinal therapeutic system on glycemic control and insulin secretion in NIDDM: Results of two multicenter, randomized, placebo-controlled clinical trials. Diabetes Care 1997 April;20(4):597-606.
149. Bautista JL, Bugos C, Dirnberger G, Atherton T. Efficacy and safety profile of glimepiride in Mexican American patients with type 2 diabetes mellitus: A randomized, placebo-controlled study. Clinical Therapeutics 2003 01 Jan;25(1):194-209.
150. Madsbad S, Schmitz O, Ranstam J, Jakobsen G, Matthews DR. Improved glycemic control with no weight increase in patients with type 2 diabetes after once-daily treatment with the long-acting glucagon-like peptide 1 analog liraglutide (NN2211): A 12-week, double-blind, randomized, controlled trial. Diabetes Care 2004 June;27(6):1335-42.
151. Vray M, Attali JR. Randomized study of glibenclamide versus traditional Chinese treatment in type 2 diabetic patients. Diabete et Metabolisme 1995;21(6):433-9.
152. Coniff RF, Shapiro JA, Seaton TB, Bray GA. Multicenter, placebo-controlled trial comparing acarbose (BAY g 5421) with placebo, tolbutamide, and tolbutamide-plus-acarbose in non-insulin-dependent diabetes mellitus. American Journal of Medicine 1995;98(5):443-51.
153. Scott R, Wu M, Sanchez M, Stein P. Efficacy and tolerability of the dipeptidyl peptidase-4 inhibitor sitagliptin as monotherapy over 12 weeks in patients with type 2 diabetes. International Journal of Clinical Practice 2007 January;61(1):171-80.
154. Henry RR, Lincoff AM, Mudaliar S, Rabbia M, Chognot C, Herz M. Aleglitazar, a dual peroxisome proliferator-activated receptor-alpha/ agonist, provides beneficial effects on glycemic control and lipids in patients with type 2 diabetes. Diabetes Conference: 69th Annual Meeting of the American Diabetes Association New Orleans, LA United States Conference Start 2009;20090605(20090609).
155. Rosenblatt S, Miskin B, Glazer NB, Prince MJ, Robertson KE. The impact of pioglitazone on glycemic control and atherogenic dyslipidemia in patients with type 2 diabetes mellitus. Coronary Artery Disease 2001;12(5):413-23.
156. Aronoff S, Rosenblatt S, Braithwaite S, Egan JW, Mathisen AL, Schneider RL. Pioglitazone hydrochloride monotherapy improves glycemic control in the treatment of patients with type 2 diabetes: A 6-month randomized placebo-controlled dose-response study. Diabetes Care 2000;23(11):1605-11.
157. Gonzalez-Clemente JM. Improvement of glycaemic control by nateglinide decreases systolic blood pressure in drug-naive patients with type 2 diabetes. European Journal of Clinical Investigation 2008 March;38(3):174-9.
158. Derosa G, Maffioli P, Ferrari I, Fogari E, D'Angelo A, Palumbo I, et al. Acarbose actions on insulin resistance and inflammatory parameters during an oral fat load. European Journal of Pharmacology 2011 25 Jan;651(1-3):240-50.
159. Kawamori R, Inagaki N, Araki E, Watada H, Hayashi N, Horie Y, et al. Linagliptin monotherapy provides superior glycaemic control versus placebo or voglibose with comparable safety in Japanese patients with type 2 diabetes: A randomized, placebo and active comparator-controlled, double-blind study. Diabetes, Obesity and Metabolism 2012 April;14(4):348-57.
160. Del Prato S, Barnett AH, Huisman H, Neubacher D, Woerle HJ, Dugi KA. Effect of linagliptin monotherapy on glycaemic control and markers of beta-cell function in patients with inadequately controlled type 2 diabetes: A randomized controlled trial. Diabetes, Obesity and Metabolism 2011 March;13(3):258-67.

161. Kikuchi M, Abe N, Kato M, Terao S, Mimori N, Tachibana H. Vildagliptin dose-dependently improves glycemic control in Japanese patients with type 2 diabetes mellitus. Diabetes Research and Clinical Practice 2009 February;83(2):233-40.
162. Mohan V, Yang W, Son HY, Xu L, Noble L, Langdon RB, et al. Efficacy and safety of sitagliptin in the treatment of patients with type 2 diabetes in China, India, and Korea. Diabetes Research and Clinical Practice 2009 January;83(1):106-16.
163. Aschner P, Kipnes MS, Lunceford JK, Sanchez M, Mickel C, Williams-Herman DE. Effect of the dipeptidyl peptidase-4 inhibitor sitagliptin as monotherapy on glycemic control in patients with type 2 diabetes. Diabetes Care 2006 December;29(12):2632-7.
164. Rosenstock J, Sankoh S, List JF. Glucose-lowering activity of the dipeptidyl peptidase-4 inhibitor saxagliptin in drug-naive patients with type 2 diabetes. Diabetes, Obesity and Metabolism 2008 May;10(5):376-86.
165. Dejager S, Razac S, Foley JE, Schweizer A. Vildagliptin in drug-naive patients with type 2 diabetes: A 24-week, double-blind, randomized, placebo-controlled, multiple-dose study. Hormone and Metabolic Research 2007 March;39(3):218-23.
166. Pi-Sunyer FX, Schweizer A, Mills D, Dejager S. Efficacy and tolerability of vildagliptin monotherapy in drug-naive patients with type 2 diabetes. Diabetes Research and Clinical Practice 2007 April;76(1):132-8.
167. Barzilai N, Guo H, Mahoney EM, Caporossi S, Golm GT, Langdon RB, et al. Efficacy and tolerability of sitagliptin monotherapy in elderly patients with type 2 diabetes: a randomized, double-blind, placebo-controlled trial. Curr Med Res Opin 2011 May;27(5):1049-58.
168. Iwamoto Y, Taniguchi T, Nonaka K, Okamoto T, Okuyama K, Arjona Ferreira JC, et al. Dose-ranging efficacy of sitagliptin, a dipeptidyl peptidase-4 inhibitor, in Japanese patients with type 2 diabetes mellitus. Endocr J 2010;57(5):383-94.
169. Nonaka K, Kakikawa T, Sato A, Okuyama K, Fujimoto G, Kato N, et al. Efficacy and safety of sitagliptin monotherapy in Japanese patients with type 2 diabetes. Diabetes Res Clin Pract 2008 Feb;79(2):291-8.
170. Raz I, Hanefeld M, Xu L, Caria C, Williams-Herman D, Khatami H. Efficacy and safety of the dipeptidyl peptidase-4 inhibitor sitagliptin as monotherapy in patients with type 2 diabetes mellitus. Diabetologia 2006 Nov;49(11):2564-71.
171. Ristic S, Byiers S, Foley J, Holmes D. Improved glycaemic control with dipeptidyl peptidase-4 inhibition in patients with type 2 diabetes: vildagliptin (LAF237) dose response. Diabetes Obes Metab 2005 Nov;7(6):692-8.
172. Scherbaum WA, Schweizer A, Mari A, Nilsson PM, Lalanne G, Jauffret S, et al. Efficacy and tolerability of vildagliptin in drug-naive patients with type 2 diabetes and mild hyperglycaemia. Diabetes, Obesity and Metabolism 2008 August;10(8):675-82.
173. Scherbaum WA, Schweizer A, Mari A, Nilsson PM, Lalanne G, Wang Y, et al. Evidence that vildagliptin attenuates deterioration of glycaemic control during 2-year treatment of patients with type 2 diabetes and mild hyperglycaemia. Diabetes, Obesity and Metabolism 2008;10(11):1114-24.
174. Charpentier G, Fleury F, Kabir M, Vaur L, Halimi S. Improved glycaemic control by addition of glimepiride to metformin monotherapy in Type 2 diabetic patients. Diabetic Medicine 2001;18(10):828-34.
175. Tosi F, Muggeo M, Brun E, Spiazzi G, Perobelli L, Zanolin E, et al. Combination treatment with metformin and glibenclamide versus single-drug therapies in type 2 diabetes mellitus: A randomized, double-blind, comparative study. Metabolism: Clinical and Experimental 2003 01 Jul;52(7):862-7.
176. Hermann LS, Schersten B, Bitzen PO, Kjellstrom T, Lindgarde F, Melander A. Therapeutic comparison of metformin and sulfonylurea, alone and in various combinations: A double-blind controlled study. Diabetes Care 1994 October;17(10):1100-9.
177. Goldstein BJ, Pans M, Rubin CJ. Multicenter, randomized, double-masked, parallel-group assessment of simultaneous glipizide/metformin as second-line pharmacologic treatment for patients with type 2 diabetes mellitus that is inadequately controlled by a sulfonylurea. Clin Ther 2003 Mar;25(3):890-903.

178. Russell-Jones D, Cuddihy RM, Hanefeld M, Kumar A, Gonzalez JG, Chan M, et al. Efficacy and safety of exenatide once weekly versus metformin, pioglitazone, and sitagliptin used as monotherapy in drug-naive patients with type 2 diabetes (DURATION-4): a 26-week double-blind study. Diabetes Care 2012 Feb;35(2):252-8.
179. Moses R, Carter J, Slobodniuk R, Donnelly T, Boyages S, Moffitt P, et al. Effect of repaglinide addition to metformin monotherapy on glycemic control in patients with type 2 diabetes. Diabetes Care 1999 January;22(1):119-24.
180. Lund SS, Tarnow L, Stehouwer CD, Schalkwijk CG, Frandsen M, Smidt UM, et al. Targeting hyperglycaemia with either metformin or repaglinide in non-obese patients with type 2 diabetes: results from a randomized crossover trial. Diabetes Obes Metab 2007 May;9(3):394-407.
181. Bosi E, Dotta F, Jia Y, Goodman M. Vildagliptin plus metformin combination therapy provides superior glycaemic control to individual monotherapy in treatment-naive patients with type 2 diabetes mellitus. Diabetes, Obesity and Metabolism 2009;11(5):506-15.
182. Aschner P, Katzeff HL, Guo H, Sunga S, Williams-Herman D, Kaufman KD, et al. Efficacy and safety of monotherapy of sitagliptin compared with metformin in patients with type 2 diabetes. Diabetes, Obesity and Metabolism 2010 March;12(3):252-61.
183. Schweizer A, Dejager S, Bosi E. Comparison of vildagliptin and metformin monotherapy in elderly patients with type 2 diabetes: A 24-week, double-blind, randomized trial. Diabetes, Obesity and Metabolism 2009;11(8):804-12.
184. Jadzinsky M, Pfutzner A, Paz-Pacheco E, Xu Z, Allen E, Chen R. Saxagliptin given in combination with metformin as initial therapy improves glycaemic control in patients with type 2 diabetes compared with either monotherapy: a randomized controlled trial. Diabetes, obesity & metabolism 2009 Jun;11(6):611-22.
185. Williams-Herman D, Johnson J, Teng R, Luo E, Davies MJ, Kaufman KD, et al. Efficacy and safety of initial combination therapy with sitagliptin and metformin in patients with type 2 diabetes: A 54-week study. Current Medical Research and Opinion 2009 March;25(3):569-83.
186. Williams-Herman D, Johnson J, Teng R, Golm G, Kaufman KD, Goldstein BJ, et al. Efficacy and safety of sitagliptin and metformin as initial combination therapy and as monotherapy over 2 years in patients with type 2 diabetes. Diabetes, Obesity and Metabolism 2010 May;12(5):442-51.
187. Schweizer A, Couturier A, Foley JE, Dejager S. Comparison between vildagliptin and metformin to sustain reductions in HbA1c over 1 year in drug-naive patients with Type 2 diabetes. Diabetic Medicine 2007 September;24(9):955-61.
188. Goke B, Hershon K, Kerr D, Calle Pascual A, Schweizer A, Foley J, et al. Efficacy and safety of vildagliptin monotherapy during 2-year treatment of drug-naive patients with type 2 diabetes: comparison with metformin. Horm Metab Res 2008 Dec;40(12):892-5.
189. Perriello G, Pampanelli S, Di Pietro C, Brunetti P. Comparison of glycaemic control over 1 year with pioglitazone or gliclazide in patients with Type 2 diabetes. Diabetic Medicine 2006 March;23(3):246-52.
190. Jain R, Osei K, Kupfer S, Perez AT, Zhang J. Long-term safety of pioglitazone versus glyburide in patients with recently diagnosed type 2 diabetes mellitus. Pharmacotherapy 2006 Oct;26(10):1388-95.
191. Tan MH, Johns D, Strand J, Halse J, Madsbad S, Eriksson JW, et al. Sustained effects of pioglitazone vs. glibenclamide on insulin sensitivity, glycaemic control, and lipid profiles in patients with Type 2 diabetes. Diabet Med 2004 Aug;21(8):859-66.
192. Tan M, Johns D, Gonzalez Galvez G, Antunez O, Fabian G, Flores-Lozano F, et al. Effects of pioglitazone and glimepiride on glycemic control and insulin sensitivity in Mexican patients with type 2 diabetes mellitus: A multicenter, randomized, double-blind, parallel-group trial. Clin Ther 2004 May;26(5):680-93.
193. Madsbad S, Kilhovd B, Lager I, Mustajoki P, Dejgaard A. Comparison between repaglinide and glipizide in Type 2 diabetes mellitus: A 1-year multicentre study. Diabetic Medicine 2001;18(5):395-401.
194. Marbury T, Huang WC, Strange P, Lebovitz H. Repaglinide versus glyburide: A one-year comparison trial. Diabetes Research and Clinical Practice 1999 March;43(3):155-66.

195. Wolffenbuttel BHR, Landgraf R. A 1-year multicenter randomized double-blind comparison of repaglinide and glyburide for the treatment of type 2 diabetes. Diabetes Care 1999 March;22(3):463-7.
196. Clinical trial report Novo Nordisk (AGEE-3783).A 16-week, multicentre, randomised, open-label, parallel group study to investigate the efficacy and safety
profiles of repaglinide monotherapy compared to gliclazide monotherapy in Chinese antidiabetic-naïve
subjects with type 2 diabetes2011: Available from: http://www.novonordisk-trials.com/website/search/trial-result-details.aspx?id=18227&p=2&search=t.
197. Feinbock C, Luger A, Klingler A, Egger T, Bielesz GK, Winkler F, et al. Prospective multicentre trial comparing the efficacy of, and compliance with, glimepiride or acarbose treatment in patients with Type 2 diabetes not controlled with diet alone. Diabetes, Nutrition and Metabolism - Clinical and Experimental 2003 August;16(4):214-21.
198. Salman S, Salman F, Satman I, Yilmaz Y, Ozer E, Sengul A, et al. Comparison of acarbose and gliclazide as first-line agents in patients with type 2 diabetes. Curr Med Res Opin 2001;16(4):296-306.
199. Ferreira JC MM, Barzilai N, Guo H, Golm GT, Sisk CM, et al. Efficacy and Safety of Sitagliptin Versus Glipizide in Patients With Type 2 Diabetes and Moderate-to-Severe Chronic Renal Insufficiency. Diabetes Care 2013;36(5):1067-73.
200. Foley JE, Sreenan S. Efficacy and safety comparison between the DPP-4 inhibitor vildagliptin and the sulfonylurea gliclazide after two years of monotherapy in drug-naive patients with type 2 diabetes. Hormone and Metabolic Research 2009;41(12):905-9.
201. Rosenstock J, Baron MA, Camisasca RP, Cressier F, Couturier A, Dejager S. Efficacy and tolerability of initial combination therapy with vildagliptin and pioglitazone compared with component monotherapy in patients with type 2 diabetes. Diabetes, Obesity and Metabolism 2007 March;9(2):175-85.
202. Pan CY, Gao Y, Li GW, Zhu XX, Gao X, Liu X. A comparison of efficacy and tolerance of nateglinide and acarbose monotherapy in type 2 diabetes mellitus. [Chinese]. Zhonghua nei ke za zhi [Chinese journal of internal medicine] 2009 Apr;48(4):304-7.
203. Pan C, Yang W, Barona JP, Niggli M, Mohideen P, Wang Y, et al. Comparison of vildagliptin and acarbose monotherapy in patients with type 2 diabetes: A 24-week, double-blind, randomized trial. Diabetic Medicine 2008 April;25(4):435-41.

Eidesstattliche Versicherung

Bartmus, Thomas, MPH

Ich erkläre hiermit an Eides statt,
dass ich die vorliegende Dissertation mit dem Thema

A systematic review and mixed treatment comparison assessing the comparative effectiveness and safety of oral antihyperglycemic drugs as monotherapy in patients with type 2 diabetes mellitus

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

Ort, Datum Unterschrift Doktorand