

# A systematic literature review on the efficacy–effectiveness gap: comparison of randomized controlled trials and observational studies of glucose-lowering drugs

Mikkel Z Ankarfeldt<sup>1,2</sup>  
Erpur Adalsteinsson<sup>1</sup>  
Rolf HH Groenwold<sup>2,3</sup>  
M Sanni Ali<sup>2,3,4</sup>  
Olaf H Klungel<sup>2,3</sup>

On behalf of GetReal Work  
Package 2

<sup>1</sup>Novo Nordisk A/S, <sup>2</sup>Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, <sup>3</sup>Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Utrecht, the Netherlands; <sup>4</sup>Nuffield Department of Orthopaedics, Rheumatology, Musculoskeletal Sciences, University of Oxford, Oxford, UK

**Aim:** To identify a potential efficacy–effectiveness gap and possible explanations (drivers of effectiveness) for differences between results of randomized controlled trials (RCTs) and observational studies investigating glucose-lowering drugs.

**Methods:** A systematic literature review was conducted in English language articles published between 1 January, 2000 and 31 January, 2015 describing either RCTs or observational studies comparing glucagon-like peptide-1 analogs (GLP-1) with insulin or comparing dipeptidyl peptidase-4 inhibitors (DPP-4i) with sulfonylurea, all with change in glycated hemoglobin (HbA1c) as outcome. Medline, Embase, Current Content, and Biosis were searched. Information on effect estimates, baseline characteristics of the study population, publication year, study duration, and number of patients, and for observational studies, characteristics related to confounding adjustment and selection- and information bias were extracted.

**Results:** From 312 hits, 11 RCTs and 7 observational studies comparing GLP-1 with insulin, and from 474 hits, 16 RCTs and 4 observational studies comparing DPP-4i with sulfonylurea were finally included. No differences were observed in baseline characteristics of the study populations (age, sex, body mass index, time since diagnosis of type 2 diabetes mellitus, and HbA1c) or effect sizes across study designs. Mean effect sizes ranged from –0.43 to 0.91 and from –0.80 to 1.13 in RCTs and observational studies, respectively, comparing GLP-1 with insulin, and from –0.13 to 2.70 and –0.20 to 0.30 in RCTs and observational studies, respectively, comparing DPP-4i and sulfonylurea. Generally, the identified observational studies held potential flaws with regard to confounding adjustment and selection- and information bias.

**Conclusions:** Neither potential drivers of effectiveness nor an efficacy–effectiveness gap were identified. However, the limited number of studies and potential problems with confounding adjustment, selection- and information bias in the observational studies, may have hidden a true efficacy-effectiveness gap.

**Keywords:** efficacy–effectiveness gap, diabetes mellitus, type 2, glucose-lowering drugs, hemoglobin A1c, literature review

Correspondence: Mikkel Z Ankarfeldt  
Clinical Research Centre, Copenhagen  
University Hospital, Department 056,  
Kettegård Allé 30, 2650 Hvidovre,  
Denmark  
Tel +45 3 862 3350  
Email mikkelza@gmail.com

## Introduction

The beneficial effects of drugs can be divided into efficacy and effectiveness. The efficacy of a drug describes the biological effect and can be seen as the effect evaluated under optimal conditions in randomized controlled trials (RCTs). The effectiveness of

a drug describes the effect under circumstances of routine clinical practice. The efficacy–effectiveness gap refers to the difference between the (in theory) largest possible effect of a drug and its effect in clinical practice.<sup>1–4</sup> A comparison of RCTs and observational studies can be used as a model to investigate and better understand the efficacy–effectiveness gap.

The population in routine clinical practice may differ from the often highly selected study population included in RCTs,<sup>5–10</sup> which could be one possible reason for an efficacy–effectiveness gap. Observational studies usually reflect the population seen in clinical practice, and also other factors such as the delivery of care, adherence to treatment, and time between treatment and assessment of the outcome are often more similar to ordinary clinical practice than that which is seen in RCTs because observational studies are often based on real-world data.<sup>11</sup> Discrepancies in the results from RCTs and observational studies may be due to biases in the observational study design,<sup>12–15</sup> but may also be explained by an efficacy–effectiveness gap. An understanding of the efficacy–effectiveness gap is important for patients, health care professionals, payers, regulators, and the pharmaceutical industry to provide effective treatments.<sup>3,16</sup>

The aim of this literature review is to identify a potential efficacy–effectiveness gap, by comparing RCTs and observational studies investigating glucose-lowering drugs in relation to change in glycated hemoglobin (HbA1c), and if such a gap exists, to investigate whether it can be explained by differences in the baseline characteristics of the study populations or other features that characterize the RCTs and observational studies.

## Methods

A systematic literature search was performed to identify RCTs and observational studies fulfilling the following inclusion criteria: published between 1 January, 2000 and 31 January, 2015 in English language and compared either glucagon-like peptide-1 analogs (GLP-1) with insulin or dipeptidyl peptidase-4 inhibitors (DPP-4i) with sulfonylurea, all with change in HbA1c as an outcome. The chosen comparator groups were to compare second-line (DPP-4i and sulfonylurea) and third-line (GLP-1 and insulin) treatments, respectively.<sup>17</sup> Especially, observational studies are difficult to identify, and therefore, more search terms were used to identify such studies, and covered both prospective and retrospective studies, as well as cohort and case–control studies. The key terms and the combination of these can be found in the [supplementary material](#). The following databases

were used: Medline, Embase, Current Content, and Biosis. The search strategy was developed by one of the reviewers (MZA) and a librarian. References of the identified studies were searched to identify additional relevant studies.

The studies identified through the literature search were screened on title and abstract by two reviewers independently. Disagreements were settled through discussions and consensus. Full text was read by a single reviewer, who extracted information on the baseline characteristics of the study population, other features that described the included studies, and effect estimate from text and tables in the included studies. Some of the hits from the search were abstracts published in relation to scientific conferences. Information from conference abstracts was not included in this review. If a conference abstract seemed relevant, an attempt was made to identify the published studies related to the conference abstract by web search and by contacting the authors of the conference abstract.

Post hoc, it was decided to exclude studies comparing DPP-4i with sulfonylurea during Ramadan in Muslim populations (three RCTs and six observational studies) because we did not want to compare across fasting and nonfasting studies and to exclude studies with fast-acting insulin (five RCTs) because we did not want to compare across fast-acting and basal insulins. Studies investigating mixed insulin (combination of fast-acting and intermediate/long-acting) were included.

Post hoc exclusion criteria were applied as we gained knowledge when working on the review. Importantly, none of the post hoc exclusion criteria are in conflict with the initial inclusion criteria and they only narrow the inclusion criteria further.

If the identified RCTs and the observational studies included treatment arms of other drugs or placebo, only information about the relevant treatment arms was extracted. If several publications were based on the same study population, but with different follow-up time, the information on patient characteristics was extracted once, while each effect size at different time points was extracted. If studies included several analyses, for example, intention-to-treat and per protocol, the analysis that was reported as the primary analysis was extracted. Two RCTs<sup>18,19</sup> included a once-daily and a twice-daily insulin group; GLP-1 vs. twice-daily insulin is reported later. Two RCTs<sup>20,21</sup> included a high and a low dose of GLP-1 and DPP-4i, respectively; high dose vs. comparator is reported later. Generally, the data extraction protocol was based on the Cochrane Handbook:<sup>22</sup> Baseline characteristics were extracted as mean and standard deviation (SD) or

proportion. A few studies reported median and interquartile range, and in those cases, SD was derived by dividing interquartile range by 1.35.<sup>22</sup> The reported outcome is the difference in change in HbA1c between treatment groups. When extracting effect estimates, the following prioritization was used: 1) effect estimate and 95% confidence interval (CI) as written in text; 2) effect estimate and 95% CI as written in a table; 3) if, for example, one-sided interval was given, then the two-sided 95% CI was calculated; 4) if no effect size with CI was given, these were calculated from the effect estimate and SD or standard error of the mean (SEM) in each treatment group; 5) if no SD or SEM, but a *p* value was given, then *z* values were calculated, and from this SEM and 95% CI; and 6) if only an effect estimate was reported and no CI or a *p* value, only the point estimate was used.

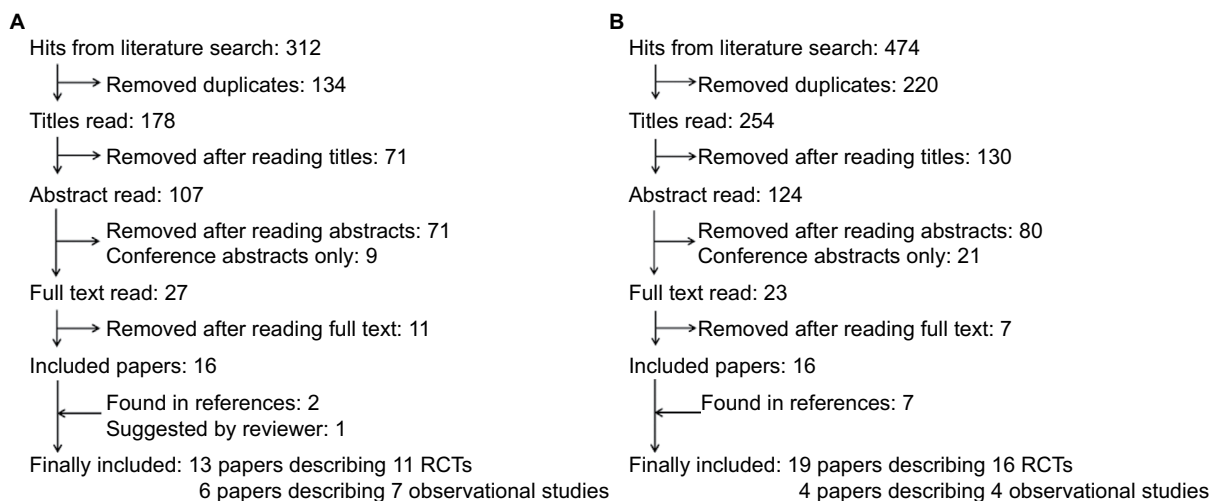
For the observational studies, additional information was extracted: confounding adjustment, analysis of initiator by having a “wash-out” period, selection bias related to clear and reasonable inclusion criteria or handling of missing data, and information bias related to the assessment of exposure and outcome. Comprehensive methods to assess quality of observational studies, such as, for example, ACROBATE-NRSI,<sup>23</sup> were not deemed necessary because the aim was not to have an estimate of the overall treatment effect across studies, but rather to look at signals of an efficacy–effectiveness gap and potential drivers of such a gap. In relation to this, pooled estimates of the study characteristics and the effect estimates were not performed. The literature search and inclusion of studies did not strive to get homogeneous studies suitable for pooled estimates. Instead, baseline characteristics and effect estimates were handled descriptively.

The overlap of patient characteristics and effect estimate was used to assess if difference was present across studies. A difference >0.4% units is acknowledged as a clinically meaningful difference in HbA1c<sup>24</sup> and was used to evaluate an efficacy–effectiveness gap.

## Results

The search for studies comparing GLP-1 with insulin showed 312 hits, of which 19 publications were included. However, the three publications by Diamant et al<sup>25–27</sup> were based on the same RCT, but with different follow-up time, and the study by Thayer et al<sup>28</sup> included two cohorts, which were reported separately later. Hence, 13 publications described 11 individual RCTs<sup>18–20,25–27,29–35</sup> and 6 publications described 7 individual observational studies<sup>28,36–40</sup> (Figure 1). The study duration ranged from 16 to 156 weeks and from 26 to 102 weeks in RCTs and observational studies, respectively, and the number of participants ranged from 69 to 1028 and from 47 to 51,977, respectively. Among the 312 hits, 9 were conference abstracts of observational studies, of which 1 was among the included observational studies as a research article. The authors of the other conference abstracts were contacted; one author replied, and no additional full-text study was identified.

The search for studies comparing DPP-4i with sulfonylurea showed 474 hits, of which 23 publications were included. However, the publications by Nauck et al,<sup>41</sup> Seck et al,<sup>42</sup> Ferrannini et al,<sup>43</sup> and Matthews et al,<sup>44</sup> and the two publications by Göke et al,<sup>45,46</sup> respectively, were based on the same RCTs with different follow-up time. Hence, 19 publications described 16 individual RCTs<sup>21,41–58</sup> and 4 publications described 4 individual observational studies<sup>59–62</sup> (Figure 1).



**Figure 1** Flow chart.

**Notes:** (A) Studies comparing glucagon-like peptide-1 with insulin. (B) Studies comparing dipeptidyl peptidase-4 inhibitors with sulfonylurea.

**Abbreviation:** RCTs, randomized controlled trials.

The study duration ranged from 4 to 104 weeks and from 24 to 52 weeks in RCTs and observational studies, respectively, and the number of participants ranged from 33 to 31 18 and from 69 to 16,832, respectively. Among the 474 hits, 4 and 17 were conference abstracts of RCTs and observational studies, respectively, of which 2 were among the included RCTs as research articles. The authors of the other conference abstracts were contacted; none of them replied, and no additional full-text study was identified.

More detailed information on the included studies is found in [Tables S1–S4](#).

Table 1 holds information on study population characteristics of the 17 individual studies (10 RCTs and 7 observational studies) and the effect estimates from the 18 publications comparing GLP-1 with insulin. Table 2 holds information on study population characteristics of the 20 individual studies (16 RCTs and 4 observational studies) and the effect estimates of the 23 publications comparing DPP-4i with sulfonylurea.

## Characteristics of observational studies

Of the 11 individual observational studies,<sup>28,36–40,59–62</sup> 4 were prospective<sup>39,60–62</sup> and 7 were based on registries.<sup>28,36–38,40,59</sup>

Information of exposure in the prospective studies was based on doctor's records of prescription, whereas exposure in registry studies was based on databases with information on prescription<sup>36–38,40,59</sup> or claims.<sup>28</sup> The outcome in all studies was based on the clinical measure of HbA1c. The inclusion criteria in the studies were primarily based on previous medication, but also age and comorbidity were used in most studies. The observational studies analyzed patients who initiated either GLP-1 or insulin, or DPP-4i or sulfonylurea, respectively. Five of the observational studies excluded patients if information was missing,<sup>28,36–38</sup> while the other six studies did not mention how missing data were handled.<sup>39,40,59–62</sup> Five of the observational studies used multivariable regression<sup>38–40,60</sup> or propensity score matching<sup>37</sup> to adjust for potential confounding, although Karagianni et al<sup>39</sup> only included body mass index (BMI) and age in the model. Unadjusted effect estimates were reported in the remaining six observational studies.<sup>28,36,59,61,62</sup> Generally, the design of the included observational studies was deemed suboptimal regarding confounding adjustment and the potential for selection- and information bias. However, two observational studies – one study<sup>37</sup> comparing GLP-1 with insulin and another study<sup>60</sup> comparing DPP-4i and sulfonylurea – were

**Table 1** Characteristics of RCTs and observational studies comparing glucagon-like peptide-1 with insulin

	Authors	Duration, weeks	N	Age, years	Men, %	Body mass index, kg/m <sup>2</sup>	Time since diagnosis of type 2 diabetes mellitus, years	Baseline HbA1c, %	Mean effect (95% confidence interval) <sup>a</sup>
Randomized controlled trials	Barnett et al <sup>30</sup>	16	138	54.9 (9.1)	47.1	31.1 (4.7)	7.4 (5.9)	9.0 (1.1)	−0.01 (−0.17, 0.15)
	Bergenstal et al <sup>18</sup>	24	248	53.0 (10.3)	48.0	33.9 (7.3)	9.3 (5.8)	10.3 (1.7)	0.91 (0.59, 1.23)
	Nauck et al <sup>20</sup>	24	667	57.5 (9.0)	51.7	32.5 (5.3)	9.5 (6.0)	8.4 (0.9)	−0.14 (−0.28, −0.01)
	Davies et al <sup>32</sup>	26	234	56.5 (9.1)	68.4	34.1 (5.3)	8.7 (4.5)	8.6 (0.7)	0.01 (−0.24, 0.26)
	Davies et al <sup>19</sup>	26	216	58.5 (10.0)	66.4	33.7 (4.7)	7.5 (5.5)	8.4 (0.9)	−0.42 (−0.63, −0.21)
	Diamant et al <sup>27</sup>	26	456	58.0 (9.5)	53.5	32.3 (5.1)	7.9 (5.0)	8.3 (1.1)	−0.16 (−0.29, −0.03)
	Heine et al <sup>29</sup>	26	535	58.9 (9.1)	55.8	31.4 (4.5)	9.6 (5.9)	8.3 (1.0)	0.02 (−0.12, 0.16)
	Inagaki et al <sup>33</sup>	26	427	56.8 (10.8)	67.9	26.2 (3.9)	9.0 (6.0)	8.5 (0.8)	−0.43 (−0.59, −0.26)
	Russell-Jones et al <sup>35</sup>	26	466	57.6 (10.0)	58.5	30.4 (5.3)	9.5 (6.1)	8.3 (0.9)	−0.24 (−0.08, −0.39)
	Bunck et al <sup>31</sup>	52	69	58.4 (8.0)	65.2	30.5 (3.8)	4.9 (4.2)	7.5 (0.6)	−0.10 (−0.54, 0.34)
	Weissman et al <sup>34</sup>	52	725	55.5 (9.5)	56.1	33.1 (5.5)	8.8 (6.3)	8.1 (0.9)	0.11 (−0.04, 0.27)
	Diamant et al <sup>26</sup>	84	456	58.0 (9.5)	53.5	32.3 (5.1)	7.9 (5.0)	8.3 (1.1)	−0.18 (−0.33, −0.02)
	Diamant et al <sup>25</sup>	156	456	58.0 (9.5)	53.5	32.3 (5.1)	7.9 (5.0)	8.3 (1.1)	−0.20 (−0.39, −0.02)
Observational studies	Karagianni et al <sup>39</sup>	26	47	62.0 (8.6)	34.0	34.4 (5.6)	11.9 (7.1)	8.4 (1.6)	−0.80 (−1.84, 0.24)
	Horton et al <sup>36</sup>	36	38,678	60.4 (13.3)	46.9	34.4 (9.1)	–	8.6 (2.2)	0.50 (0.46, 0.54)
	Thayer et al <sup>28b</sup>	36	861	53.0 (8.9)	55.7	–	–	9.0 (6.1)	0.53 (–)
	Thayer et al <sup>28b</sup>	52	1709	55.8 (11.0)	54.7	–	–	8.7 (1.7)	1.13 (–)
	Pawaskar et al <sup>37</sup>	52	5366	58.0 (–)	46.3	36.7	–	8.1 (–)	−0.20 (–)
	Hall et al <sup>38</sup>	52	2965	60.7 (11.4)	61.9	33.7 (6.5)	8.8 (5.7)	9.6 (3.8)	0.13 (−0.11, 0.38)
Bounthavong et al <sup>40</sup>	102	51,977	64.2 (10.4)	96.8	33.0 (6.7)	–	8.8 (2.0)	−0.32 (−0.47, −0.18)	

**Notes:** Data shown as mean (standard deviation) unless specified otherwise. Diamant et al<sup>25–27</sup> are based on the same RCTs, but with different follow-ups. <sup>a</sup>The difference in change in HbA1c between treatment groups. <sup>b</sup>Two cohort studies described in the same publication. – indicates data not reported.

**Abbreviations:** HbA1c, glycated hemoglobin; RCTs, randomized controlled trials.

**Table 2** Characteristics of RCTs and observational studies comparing dipeptidyl peptidase-4 inhibitors with sulfonylurea

	Authors	Duration, weeks	N	Age, years	Men, %	Body mass index, kg/m <sup>2</sup>	Time since diagnosis of type 2 diabetes mellitus, years	Baseline HbA1c, %	Mean effect (95% confidence interval) <sup>a</sup>	
Randomized controlled trials	Kim et al <sup>55</sup>	4	33	57.8 (6.7)	58.6	25.5 (2.8)	5.3 (4.7)	7.2 (0.5)	0.00 (–)	
	Shimoda et al <sup>58</sup>	12	50	63.1 (12.4)	31.0	25.1 (3.9)	–	7.4 (0.6)	2.70 (–0.10, 5.50)	
	Srivastava et al <sup>50</sup>	18	50	–	–	25.9 (3.3)	–	8.3 (0.5)	0.54 (0.02, 1.06)	
	Derosa et al <sup>57</sup>	26	167	58.1 (9.4)	49.1	27.8 (1.5)	6.7 (4.1)	7.8 (0.8)	0.00 (–)	
	Jeon and Oh <sup>49</sup>	32	101	54.5 (10.7)	64.7	22.9 (6.0)	5.9 (1.7)	8.1 (1.0)	0.06 (–0.42, 0.54)	
	Derosa et al <sup>53</sup>	52	453	–	–	49.6	27.3 (2.1)	5.0 (2.0)	8.3 (1.2)	0.20 (–1.73, 2.13)
	Rosenstock et al <sup>54</sup>	52	441	70.0 (4.3)	44.9	29.8 (4.5)	6.1 (6.3)	7.5 (0.7)	–0.05 (–0.23, 0.13)	
	Nauck et al <sup>41</sup>	52	1172	56.7 (9.6)	59.2	31.3 (5.1)	6.4 (5.8)	7.7 (0.9)	0.00 (–)	
	Göke et al <sup>45</sup>	52	858	57.6 (10.3)	51.8	31.4 (5.9)	5.5 (4.6)	7.7 (0.9)	0.06 (–0.05, 0.16)	
	Ferrannini et al <sup>43</sup>	52	3118	57.5 (9.13)	53.5	31.8 (5.3)	5.7 (5.1)	7.3 (0.7)	0.09 (0.03, 0.15)	
	Filozof and Gautier <sup>48</sup>	52	1007	59.5 (10.0)	52.0	31.0 (5.0)	6.6 (5.2)	8.5 (1.0)	0.04 (–0.11, 0.20)	
	Arjona et al <sup>52</sup>	54	426	64.5 (9.9)	57.0	26.8 (4.8)	10.4 (7.7)	7.8 (0.7)	–0.11 (–0.29, 0.06)	
	Arjona et al <sup>51</sup>	54	129	59.5 (9.5)	59.7	26.8 (5.0)	17.5 (8.9)	7.9 (0.7)	0.15 (–0.18, 0.49)	
	Ahrén et al <sup>56</sup>	104	609	54.4 (9.9)	48.8	32.5 (5.5)	5.9 (4.8)	8.1 (0.8)	0.08 (–)	
	Del Prato et al <sup>21</sup>	104	1759	55.5 (9.7)	50.8	31.2 (5.3)	5.5 (4.8)	7.6 (0.6)	–0.13 (–0.24, –0.02)	
	Foley and Sreenan <sup>47</sup>	104	1092	54.8 (10.5)	55.8	30.7 (5.3)	2.2 (3.7)	8.7 (1.1)	0.13 (–0.06, 0.33)	
	Göke et al <sup>46</sup>	104	858	57.6 (10.3)	51.8	31.4 (5.9)	5.5 (4.6)	7.7 (0.9)	–0.05 (–0.17, 0.06)	
Matthews et al <sup>44</sup>	104	3118	57.5 (9.13)	53.5	31.8 (5.3)	5.7 (5.1)	7.3 (0.7)	0.00 (0.00, 0.1)		
Seck et al <sup>42</sup>	104	1172	56.7 (9.6)	59.2	31.3 (5.1)	6.4 (5.8)	7.7 (0.9)	–0.03 (–)		
Observational studies	Lee et al <sup>60</sup>	24	69	52.3 (12.8)	58.0	26.9 (3.9)	0.5 (0.5)	8.1 (0.8)	0.07 (–0.24, 0.37)	
	Gitt et al <sup>61</sup>	52	256	65.2 (11.1)	52.0	–	5.0 (4.2)	7.4 (0.7)	–0.10 (–0.24, 0.04)	
	Göke et al <sup>62</sup>	52	7410	62.6 (11.1)	54.0	30.8 (5.5)	5.8 (4.9)	7.7 (1.2)	–0.20 (–0.22, –0.09)	
	Morgan et al <sup>59</sup>	52	16,832	61.9 (11.4)	59.8	32.1 (5.5)	4.6 (3.8)	8.7 (1.4)	0.30 (–)	

**Notes:** Data shown as mean (standard deviation) unless specified otherwise. Nauck et al<sup>41</sup> and Seck et al<sup>42</sup>; Göke et al<sup>45</sup> and Göke et al<sup>46</sup> and Ferrannini et al<sup>43</sup> and Matthews et al<sup>44</sup> are based on the same RCTs, but with different follow-ups. <sup>a</sup>The difference in change in HbA1c between treatment groups. – indicates data not reported.

**Abbreviations:** HbA1c, glycated hemoglobin; RCTs, randomized controlled trials.

explicit about the conducted analysis, including confounding adjustment, and gave information about possible selection bias and information bias. Neither the effect estimate nor the patient characteristics of these studies<sup>37,60</sup> were different from the other observational studies comparing GLP-1 with insulin or comparing DPP-4i and sulfonylurea, respectively.

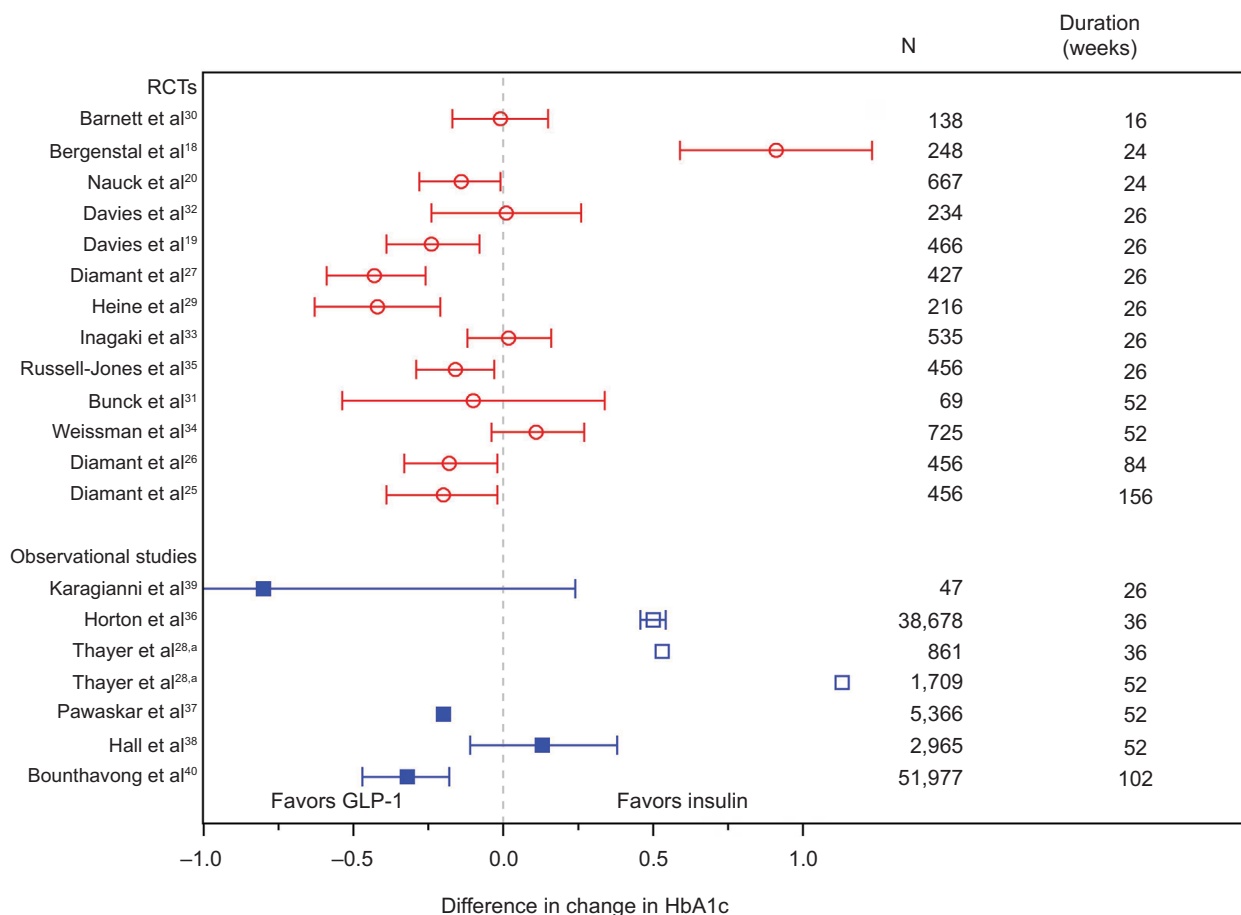
## Characteristics of the study populations across study designs

The study populations did not differ across RCTs and observational studies with regard to age, sex ratio, BMI, time since diagnosis of type 2 diabetes mellitus, and baseline HbA1c neither in the studies that compared GLP-1 with insulin nor in the studies that compared DPP-4i with sulfonylureas. Generally, this goes for both means and SDs. One exception is HbA1c among studies of GLP-1 and insulin, where the HbA1c distribution in the observational studies was more heterogeneous than in the RCTs. Also, a few outliers should be mentioned. Among studies comparing GLP-1 with insulin, the observational study by Bounthavong et al<sup>40</sup> included almost only men and BMI was low in the RCT by Inagaki et al<sup>33</sup> (explained

by a Japanese population). The range of the distribution of HbA1c is generally wider in the observational studies than in the RCTs. This indicates that the study population is more heterogeneous with regard to HbA1c in the observational studies. However, the mean of HbA1c is of similar magnitude across study designs. An outlier among the studies comparing DPP-4i with sulfonylurea is the RCT by Shimoda et al,<sup>58</sup> which included a higher proportion of women compared to the other studies. Unfortunately, information on time since diagnosis of type 2 diabetes mellitus was only available in two of the observational studies comparing GLP-1 with insulin.

## Effect estimates across study designs

Effect estimates did not differ across RCTs and observational studies, both for studies comparing GLP-1 with insulin (Figure 2) and studies comparing DPP-4i with sulfonylurea (Figure 3). Among studies comparing GLP-1 with insulin, a few studies<sup>18,28,36</sup> reported findings outside the 95% CI of the other studies; in the observational study by Horton et al<sup>36</sup> and the two cohorts in the observational study by Thayer et al,<sup>28</sup> no adjustment for confounding was done. This could



**Figure 2** Effect estimates of studies comparing glucagon-like peptide-1 with insulin.

**Notes:** Difference in mean change HbA1c  $\pm$  95% confidence interval. The difference in change in HbA1c between treatment groups. Diamant et al<sup>25-27</sup> are based on the same RCTs, but with different follow-ups. <sup>a</sup>Two different cohorts analyzed and reported in the same publication. Red circle: RCTs. Blue filled square: observational studies with confounding adjustment. Blue open square: observational studies unadjusted for confounding.

**Abbreviations:** HbA1c, glycated hemoglobin; RCTs, randomized controlled trials.

explain why the findings differ from those of the confounding adjusted observational studies and the RCTs. It should be noted that Thayer et al<sup>28</sup> did not aim for a comparison of effects across treatments. The RCT by Bergenstal et al<sup>18</sup> reported results outside the 95% CI of the other RCTs and must be seen as an outlier. Among studies comparing DPP-4i with sulfonylurea, the three observational studies reporting unadjusted effects<sup>59,61,62</sup> show effect estimates of similar magnitude to the effect estimates in the confounding adjusted observational study<sup>60</sup> and the RCTs.

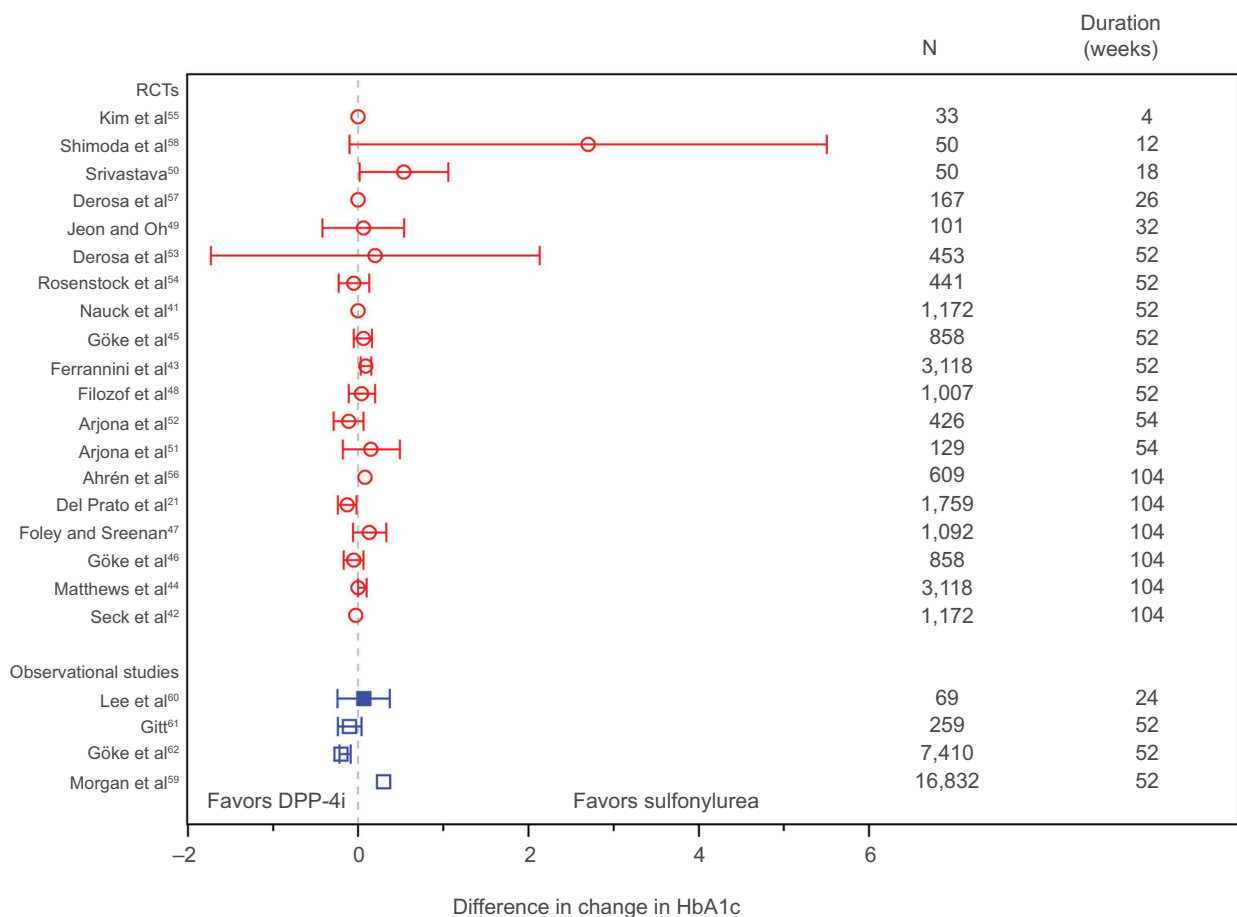
The numbers from Tables 1 and 2 are presented graphically in [Figures S1 and S2](#). More information on the RCTs and observational studies is found in [Tables S1–S4](#).

## Discussion

No clear differences in the available baseline characteristics of the study populations and in the effect estimates of the identified RCTs and observational studies were observed in this review. Hence, no efficacy–effectiveness gap was observed and no drivers of effectiveness were identified.

Despite examples where results from RCTs and observational studies seem not to agree,<sup>12–15</sup> reviews that have systematically compared the results from RCTs and observational studies have found that effect sizes from RCTs and observational studies are often similar or do not differ systematically across a range of medical subjects<sup>63,64</sup> and suggest that the theoretical efficacy–effectiveness gap may not be as widespread as often thought. This is in line with the findings in this review.

An efficacy–effectiveness gap with regard to DPP-4i (specifically vildagliptin) and sulfonylurea in relation to change in HbA1c has been investigated elsewhere,<sup>65</sup> the effect of the individual drug, that is, change from baseline of the two drugs separately, was compared across five RCTs and the one observational study. Ahrén et al<sup>65</sup> found that DPP-4i had a similar effect in the RCTs and the observational study, but that an efficacy–effectiveness gap may exist with regard to sulfonylurea because sulfonylurea proved more effective in RCTs than in the observational study. The study by Ahrén et al<sup>65</sup> is based on other data than this



**Figure 3** Effect estimates of studies comparing dipeptidyl peptidase-4 inhibitors with sulfonylurea.

**Notes:** Difference in mean change HbA1c  $\pm$ 95% confidence interval. The difference in change in HbA1c between treatment groups. Nauck et al<sup>41</sup> and Seck et al<sup>42</sup>; Göke et al<sup>45</sup> and Göke et al<sup>46</sup>; and Ferrannini et al<sup>43</sup> and Matthews et al<sup>44</sup> are based on the same RCTs, but with different follow-ups, Red circle: RCTs. Blue filled square: observational studies with confounding adjustment. Blue open square: observational studies unadjusted for confounding.

**Abbreviations:** HbA1c, glycated hemoglobin; RCTs, randomized controlled trials.

review because Ahrén et al<sup>65</sup> included RCTs that compared DPP-4i with placebo (only using data on the active arm), and because the observational data were based on the full EDGE study,<sup>66</sup> which was not included in this review because the EDGE study reported comparison of DPP-4i with other oral hypoglycemics and not specifically sulfonylurea. In this review article, the German part of the observational EDGE study<sup>62</sup> was included. Also to be mentioned, it is unclear how Ahrén et al<sup>65</sup> identified the included studies, as it was not based on a systematic literature search as in this review. This review used the comparison of two drugs as outcome (change with DPP-4i subtracted from change with sulfonylurea) and did not assess the effect of the individual drugs (change for DPP-4i and sulfonylurea, respectively) as done by Ahrén et al.<sup>65</sup>

Possible biases in this review could work in opposite directions, and thus hide an actual efficacy–effectiveness gap. No identification of an efficacy–effectiveness gap could be

a net result of such biases. Possible biases in this review are described in the following points:

1) Unmeasured confounding is always a potential problem in observational studies, and several of the observational studies reported effects not adjusted for potential confounders. Selection bias may also have been a problem in the observational studies because inclusion criteria were only partly clear in the observational studies, and all observational studies either excluded participants with missing information or did not report how missing data were handled. From this it is clear that future observational studies in the investigated area of this review can be designed to a higher degree to avoid biases and include confounding adjustment in the analyses. A descriptive approach to identify key drivers of bias was used to assess the observational studies. As stated, the aim of this review was not to assess the quality of the studies in detail with a more comprehensive and validated tool. Rather, the descriptive approach was found sufficient to identify

potential flaws in the observational studies. 2) The limited number of studies in this review may also have affected the findings. Especially, the number of observational studies was lower than the RCTs. One could speculate whether the use of a hard end point (e.g., death) would have led to a higher number of available observational studies. However, it would probably limit the amount of available RCTs. As to the number of available studies, publication bias may also have affected our results. Probably, publication bias will be most pronounced among observational studies. However, the effect estimates of the observational studies look fairly symmetric, when looking at Figures 2 and 3, which suggest no publication bias. However, a specific study on this topic is needed to draw final conclusions. It is important to note that effect estimates from the same RCTs at different follow-up time points are listed in Tables 1 and 2. However, as there was no overall effect estimated, we did not double count these studies in any pooled analysis. In the descriptive comparison of effect estimates, we wanted to make it complete, and, therefore, all effect estimates were listed. 3) Characteristics of the study populations and other features of the studies may differ in ways not quantified in the data extraction. The assessed characteristics were restricted to the information that was available in both the RCTs and the observational studies. The observational studies often included more information on patient characteristics than the RCTs, for example, distribution of comorbidities and comedication of the study population. Delivery of care and adherence to the treatments is an area where RCTs and observational studies may differ with a possible impact on treatment effect as, for example, seen in osteoporosis treatment.<sup>67</sup> However, such information was not available and, therefore, cannot be compared across study designs. Future studies based on patient-level data rather than systematic reviews may be better suited to investigate the potential drivers of effectiveness not observed in this review, for example comorbidity, comedication, delivery of care, and adherence to treatment. Studies on patient-level data are also useful to investigate effect modification of, for example, drug and patient characteristics, which will give insights in possible drivers of effectiveness. 4) It is possible that the observational studies were designed to be comparable with the RCTs with regard to, for example, the study population. If so, this would result in no efficacy–effectiveness gap because of differences in the study populations when compared in this review. However, this was neither explicitly stated in any of the observational studies nor could it be deduced from the listed inclusion criteria. 5) If the studies have had similar subgroup analyses

across RCTs and observational studies, this could be used to investigate the potential efficacy–effectiveness gap even further. However, few subgroup analyses were conducted in the included studies, and not in a way that we could compare across study designs. 6) The results of this review should be interpreted in the light of GLP-1 and DPP-4i being analyzed on drug class level. It would require many more studies to do subgroup analyses on the individual drugs, and not all observational studies give information on drug names and doses. [Tables S1–S4](#) hold the available information on drug names and doses investigated in the included studies.

In this review, HbA1c was used as outcome measure because it is the common effect measure of glucose-lowering drugs. It is important to note that this review did not aim to do a full evaluation of the included glucose-lowering drugs. Such evaluation should involve more parameters than solely change in HbA1c, for example cardiovascular events, hypoglycemic events, and weight change. We used this outcome measure as an example to study a potential efficacy–effectiveness gap. As described in the Methods section, pooled analyses were not the aim of this review. For pooled analyses to make sense, this would require more homogeneous studies, for example with regard to the duration of study, and it is likely that very few studies would be included in such analyses. Instead, the present review gives an insight into the published studies in this area, and with the inclusion of heterogenous studies, for example with varying study duration, possible explanation of an efficacy–effectiveness gap was investigated.

To conclude, no efficacy–effectiveness gap between RCTs and observational studies comparing GLP-1 with insulin or DPP-4i with sulfonylurea was observed. However, the limited number of studies and potential problems with confounding adjustment, selection- and information bias in the observational studies, may have hidden a true efficacy–effectiveness gap. Hence, the existence of an efficacy–effectiveness gap cannot be fully excluded. No potential drivers of effectiveness were identified among age, sex, BMI, time since diagnosis of type 2 diabetes mellitus, baseline HbA1c, publication year, duration of study, and number of patients in the study.

## Acknowledgments

The authors would like to thank Ida Dalgaard Pedersen from Novo Nordisk Global Information & Analysis (GLIA), Novo Nordisk A/S with help in structuring and executing the literature search. The research leading to these results was conducted as part of the GetReal consortium. For further information please refer to <https://www.imi-getreal.eu/>. This paper only reflects the personal views of the stated authors.



The work leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement no. 115546, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007–2013) and European Federation of Pharmaceutical Industries and Association (EFPIA) companies in kind contribution. In addition, as a special form of the IMI JU grant, University Medical Center Utrecht received a direct financial contribution from Novo Nordisk A/S to support work on this study. MZA and EA belong to EFPIA member companies in the IMI JU and costs related to their part in the research were carried by the respective company as in kind contribution under the IMI JU scheme.

## Author contributions

MZA, RHHG and OHK conceived the study. MZA and EA designed data collection. MZA did data extraction. All authors analyzed data and interpreted the results. MZA drafted the manuscript. All authors revised the manuscript. All authors read and approved the final manuscript and agreed to be accountable for all aspects of the work.

## Disclosure

MZA was employed by Novo Nordisk A/S as PostDoc in the IMI GetReal project. EA is employed by Novo Nordisk A/S and is a shareholder of Novo Nordisk A/S. RHHG, MSA and OHK report no conflicts of interest in this work.

## References

- Luce BR, Drummond M, Jonsson B, et al. EBM, HTA, and CER: clearing the confusion. *Milbank Q.* 2010;88(2):256–276.
- Eichler HG, Bloechl-Daum B, Abadie E, Barnett D, Konig F, Pearson S. Relative efficacy of drugs: an emerging issue between regulatory agencies and third-party payers. *Nat Rev Drug Discov.* 2010;9(4):277–291.
- Silverman E. Effectiveness/efficacy difference too often ignored. *Manag Care.* 2013;22(1):36.
- Nordon C, Karcher H, Groenwold RH, et al; GetReal consortium. The “Efficacy-Effectiveness Gap”: historical background and current conceptualization. *Value Health.* 2016;19(1):75–81.
- Davis CE. Generalizing from clinical trials. *Control Clin Trials.* 1994; 15(1):11–14.
- Bailey KR. Generalizing the results of randomized clinical trials. *Control Clin Trials.* 1994;15(1):15–23.
- Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Threats to applicability of randomised trials: exclusions and selective participation. *J Health Serv Res Policy.* 1999;4(2):112–121.
- Dowd R, Recker RR, Heaney RP. Study subjects and ordinary patients. *Osteoporos Int.* 2000;11(6):533–536.
- Khan AY, Preskorn SH, Baker B. Effect of study criteria on recruitment and generalizability of the results. *J Clin Psychopharmacol.* 2005;25(3):271–275.
- Van Spall HG, Toren A, Kiss A, Fowler RA. Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review. *JAMA.* 2007;297(11): 1233–1240.

- Sorensen HT, Lash TL, Rothman KJ. Beyond randomized controlled trials: a critical comparison of trials with nonrandomized studies. *Hepatology.* 2006;44(5):1075–1082.
- Laupacis A, Mamdani M. Observational studies of treatment effectiveness: some cautions. *Ann Intern Med.* 2004;140(11):923–924.
- Lawlor DA, Davey Smith G, Ebrahim S. Commentary: the hormone replacement–coronary heart disease conundrum: is this the death of observational epidemiology? *Int J Epidemiol.* 2004;33(3):464–467.
- Freidlin B, Korn EL. Assessing causal relationships between treatments and clinical outcomes: always read the fine print. *Bone Marrow Transplant.* 2012;47(5):626–632.
- Boyko EJ. Observational research—opportunities and limitations. *J Diabetes Complications.* 2013;27(6):642–648.
- Eichler HG, Abadie E, Breckenridge A, et al. Bridging the efficacy–effectiveness gap: a regulator's perspective on addressing variability of drug response. *Nat Rev Drug Discov.* 2011;10(7):495–506.
- The National Institute for Health and Care Excellence (NICE) [homepage on the Internet]. Managing blood glucose in adults with type 2 diabetes. Available from: <https://pathways.nice.org.uk/pathways/type-2-diabetes-in-adults#path=view%3A/pathways/type-2-diabetes-in-adults/managing-blood-glucose-in-adults-with-type-2-diabetes.xml&content=view-node%3Anodes-insulin-based-treatments>. Accessed November 25, 2016.
- Bergental R, Lewin A, Bailey T, et al. Efficacy and safety of biphasic insulin aspart 70/30 versus exenatide in subjects with type 2 diabetes failing to achieve glycemic control with metformin and a sulfonylurea. *Curr Med Res Opin.* 2009;25(1):65–75.
- Davies M, Heller S, Sreenan S, et al. Once-weekly exenatide versus once- or twice-daily insulin detemir: randomized, open-label, clinical trial of efficacy and safety in patients with type 2 diabetes treated with metformin alone or in combination with sulfonylureas. *Diabetes Care.* 2013;36(5):1368–1376.
- Nauck M, Horton E, Andjelkovic M, et al; T-emerge 5 Study Group. Taspoglutide, a once-weekly glucagon-like peptide 1 analogue, vs. insulin glargine titrated to target in patients with type 2 diabetes: an open-label randomized trial. *Diabet Med.* 2013;30(1):109–113.
- Del Prato S, Camisasca R, Wilson C, Fleck P. Durability of the efficacy and safety of alogliptin compared with glipizide in type 2 diabetes mellitus: a 2-year study. *Diabetes Obes Metab.* 2014;16(12): 1239–1246.
- Higgins JPT, Green S; Cochrane Collaboration. *Cochrane Handbook for Systematic Reviews of Interventions.* Chichester, England: Wiley – Blackwell; 2008.
- Sterne J, Higgins J, Reeves B. A Cochrane Risk of Bias Assessment Tool: for Non-Randomized Studies of Interventions (ACROBAT-NRSI), Version 1.0.0; 24 September 2014. Available from: <http://www.riskofbias.info>. Accessed October 8, 2015.
- U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER). *Guidance for Industry. Diabetes Mellitus: Developing Drugs and Therapeutic Biologics for Treatment and Prevention.* CDER: Silver Spring, MD, USA; 2008.
- Diamant M, Van Gaal L, Guerci B, et al. Exenatide once weekly versus insulin glargine for type 2 diabetes (DURATION-3): 3-year results of an open-label randomised trial. *Lancet Diabetes Endocrinol.* 2014;2(6):464–473.
- Diamant M, Van Gaal L, Stranks S, et al. Safety and efficacy of once-weekly exenatide compared with insulin glargine titrated to target in patients with type 2 diabetes over 84 weeks. *Diabetes Care.* 2012;35(4):683–689.
- Diamant M, Van Gaal L, Stranks S, et al. Once weekly exenatide compared with insulin glargine titrated to target in patients with type 2 diabetes (DURATION-3): an open-label randomised trial. *Lancet.* 2010; 375(9733):2234–2243.
- Thayer S, Wei W, Buysman E, et al. The INITIATOR study: pilot data on real-world clinical and economic outcomes in US patients with type 2 diabetes initiating injectable therapy. *Adv Ther.* 2013;30(12): 1128–1140.

29. Heine RJ, Van Gaal LF, Johns D, et al. Exenatide versus insulin glargine in patients with suboptimally controlled type 2 diabetes: a randomized trial. *Ann Intern Med.* 2005;143(8):559–569.
30. Barnett AH, Burger J, Johns D, et al. Tolerability and efficacy of exenatide and titrated insulin glargine in adult patients with type 2 diabetes previously uncontrolled with metformin or a sulfonylurea: a multinational, randomized, open-label, two-period, crossover noninferiority trial. *Clin Ther.* 2007;29(11):2333–2348.
31. Bunck MC, Diamant M, Corner A, et al. One-year treatment with exenatide improves beta-cell function, compared with insulin glargine, in metformin-treated type 2 diabetic patients: a randomized, controlled trial. *Diabetes Care.* 2009;32(5):762–768.
32. Davies MJ, Donnelly R, Barnett AH, Jones S, Nicolay C, Kilcoyne A. Exenatide compared with long-acting insulin to achieve glycaemic control with minimal weight gain in patients with type 2 diabetes: results of the helping evaluate exenatide in patients with diabetes compared with long-acting insulin (HEELA) study. *Diabetes Obes Metab.* 2009;11(12):1153–1162.
33. Inagaki N, Atsumi Y, Oura T, Saito H, Imaoka T. Efficacy and safety profile of exenatide once weekly compared with insulin once daily in Japanese patients with type 2 diabetes treated with oral antidiabetes drug(s): results from a 26-week, randomized, open-label, parallel-group, multicenter, noninferiority study. *Clin Ther.* 2012;34(9):1892–1908.
34. Weissman PN, Carr MC, Ye J, et al. HARMONY 4: randomized clinical trial comparing once-weekly albiglutide and insulin glargine in patients with type 2 diabetes inadequately controlled with metformin with or without sulfonylurea. *Diabetologia.* 2014;57(12):2475–2484.
35. Russell-Jones D, Vaag A, Schmitz O, et al; Liraglutide Effect and Action in Diabetes 5 (LEAD-5) met+SU Study Group. Liraglutide vs insulin glargine and placebo in combination with metformin and sulfonylurea therapy in type 2 diabetes mellitus (LEAD-5 met+SU): a randomised controlled trial. *Diabetologia.* 2009;52(10):2046–2055.
36. Horton ES, Silberman C, Davis KL, Berria R. Weight loss, glycemic control, and changes in cardiovascular biomarkers in patients with type 2 diabetes receiving incretin therapies or insulin in a large cohort database. *Diabetes Care.* 2010;33(8):1759–1765.
37. Pawaskar M, Li Q, Hoogwerf BJ, et al. Metabolic outcomes of matched patient populations initiating exenatide BID vs. insulin glargine in an ambulatory care setting. *Diabetes Obes Metab.* 2012;14(7):626–633.
38. Hall GC, McMahon AD, Dain MP, Wang E, Home PD. Primary-care observational database study of the efficacy of GLP-1 receptor agonists and insulin in the UK. *Diabet Med.* 2013;30(6):681–686.
39. Karagianni P, Polyzos SA, Kartali N, Zografou I, Sambanis C. Comparative efficacy of exenatide versus insulin glargine on glycemic control in type 2 diabetes mellitus patients inadequately treated with metformin monotherapy. *Adv Med Sci.* 2013;58(1):38–43.
40. Bounthavong M, Tran JN, Golshan S, et al. Retrospective cohort study evaluating exenatide twice daily and long-acting insulin analogs in a Veterans Health Administration population with type 2 diabetes. *Diabetes Metab.* 2014;40(4):284–291.
41. Nauck MA, Meininger G, Sheng D, Terranella L, Stein PP, Sitagliptin Study G. Efficacy and safety of the dipeptidyl peptidase-4 inhibitor, sitagliptin, compared with the sulfonylurea, glipizide, in patients with type 2 diabetes inadequately controlled on metformin alone: a randomized, double-blind, non-inferiority trial. *Diabetes Obes Metab.* 2007;9(2):194–205.
42. Seck T, Nauck M, Sheng D, et al; Sitagliptin Study 024 Group. Safety and efficacy of treatment with sitagliptin or glipizide in patients with type 2 diabetes inadequately controlled on metformin: a 2-year study. *Int J Clin Pract.* 2010;64(5):562–576.
43. Ferrannini E, Fonseca V, Zinman B, et al. Fifty-two-week efficacy and safety of vildagliptin vs. glimepiride in patients with type 2 diabetes mellitus inadequately controlled on metformin monotherapy. *Diabetes Obes Metab.* 2009;11(2):157–166.
44. Matthews DR, Dejager S, Ahren B, et al. Vildagliptin add-on to metformin produces similar efficacy and reduced hypoglycaemic risk compared with glimepiride, with no weight gain: results from a 2-year study. *Diabetes Obes Metab.* 2010;12(9):780–789.
45. Göke B, Gallwitz B, Eriksson J, Hellqvist A, Gause-Nilsson I; D1680C00001 Investigators. Saxagliptin is non-inferior to glipizide in patients with type 2 diabetes mellitus inadequately controlled on metformin alone: a 52-week randomised controlled trial. *Int J Clin Pract.* 2010;64(12):1619–1631.
46. Göke B, Gallwitz B, Eriksson JG, Hellqvist A, Gause-Nilsson I. Saxagliptin vs. glipizide as add-on therapy in patients with type 2 diabetes mellitus inadequately controlled on metformin alone: long-term (52-week) extension of a 52-week randomised controlled trial. *Int J Clin Pract.* 2013;67(4):307–316.
47. 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. *Horm Metab Res.* 2009;41(12):905–909.
48. Filozof C, Gautier JF. A comparison of efficacy and safety of vildagliptin and gliclazide in combination with metformin in patients with type 2 diabetes inadequately controlled with metformin alone: a 52-week, randomized study. *Diabet Med.* 2010;27(3):318–326.
49. Jeon HJ, Oh TK. Comparison of vildagliptin-metformin and glimepiride-metformin treatments in type 2 diabetic patients. *Diabetes Metab J.* 2011;35(5):529–535.
50. Srivastava S, Saxena GN, Keshwani P, Gupta R. Comparing the efficacy and safety profile of sitagliptin versus glimepiride in patients of type 2 diabetes mellitus inadequately controlled with metformin alone. *J Assoc Physicians India.* 2012;60:27–30.
51. Arjona Ferreira JC, Corry D, Mogensen CE, et al. Efficacy and safety of sitagliptin in patients with type 2 diabetes and ESRD receiving dialysis: a 54-week randomized trial. *Am J Kidney Dis.* 2013;61(4):579–587.
52. Arjona Ferreira JC, Marre M, Barzilai N, 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):1067–1073.
53. Derosa G, Cicero AF, Franzetti IG, et al. A randomized, double-blind, comparative therapy evaluating sitagliptin versus glibenclamide in type 2 diabetes patients already treated with pioglitazone and metformin: a 3-year study. *Diabetes Technol Ther.* 2013;15(3):214–222.
54. Rosenstock J, Wilson C, Fleck P. Alogliptin versus glipizide monotherapy in elderly type 2 diabetes mellitus patients with mild hyperglycaemia: a prospective, double-blind, randomized, 1-year study. *Diabetes Obes Metab.* 2013;15(10):906–914.
55. Kim HS, Shin JA, Lee SH, et al. A comparative study of the effects of a dipeptidyl peptidase-IV inhibitor and sulfonylurea on glucose variability in patients with type 2 diabetes with inadequate glycemic control on metformin. *Diabetes Technol Ther.* 2013;15(10):810–816.
56. Ahrén B, Johnson SL, Stewart M, et al; HARMONY 3 Study Group. HARMONY 3: 104-week randomized, double-blind, placebo- and active-controlled trial assessing the efficacy and safety of albiglutide compared with placebo, sitagliptin, and glimepiride in patients with type 2 diabetes taking metformin. *Diabetes Care.* 2014;37(8):2141–2148.
57. Derosa G, Bonaventura A, Bianchi L, et al. Vildagliptin compared to glimepiride on post-prandial lipemia and on insulin resistance in type 2 diabetic patients. *Metabolism.* 2014;63(7):957–967.
58. Shimoda S, Iwashita S, Sekigami T, et al. Comparison of the efficacy of sitagliptin and glimepiride dose-up in Japanese patients with type 2 diabetes poorly controlled by sitagliptin and glimepiride in combination. *J Diabetes Investig.* 2014;5(3):320–326.
59. Morgan CL, Poole CD, Evans M, Barnett AH, Jenkins-Jones S, Currie CJ. What next after metformin? a retrospective evaluation of the outcome of second-line, glucose-lowering therapies in people with type 2 diabetes. *J Clin Endocrinol Metab.* 2012;97(12):4605–4612.
60. Lee YK, Song SO, Kim KJ, et al. Glycemic effectiveness of metformin-based Ddual-combination therapies with sulphonylurea, pioglitazone, or DPP4-inhibitor in drug-naive Korean type 2 diabetic patients. *Diabetes Metab J.* 2013;37(6):465–474.
61. Gitt AK, Bramlage P, Binz C, Krekler M, Deeg E, Tschöpe D. Prognostic implications of DPP-4 inhibitor vs. sulfonylurea use on top of metformin in a real world setting – results of the 1 year follow-up of the prospective DiaRegis registry. *Int J Clin Pract.* 2013;67(10):1005–1014.

62. Göke R, Gruenberger JB, Bader G, Dworak M. Real-life efficacy and safety of vildagliptin compared with sulfonylureas as add-on to metformin in patients with type 2 diabetes mellitus in Germany. *Curr Med Res Opin.* 2014;30(5):785–789.
63. Peinemann F, Tushabe DA, Kleijnen J. Using multiple types of studies in systematic reviews of health care interventions – a systematic review. *PLoS One.* 2013;8(12):e85035.
64. Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. *Cochrane Database Syst Rev.* 2014; 4:MR000034.
65. Ahrén B, Mathieu C, Bader G, Schweizer A, Foley JE. Efficacy of vildagliptin versus sulfonylureas as add-on therapy to metformin: comparison of results from randomised controlled and observational studies. *Diabetologia.* 2014;57(7):1304–1307.
66. Mathieu C, Barnett AH, Brath H, et al. Effectiveness and tolerability of second-line therapy with vildagliptin vs. other oral agents in type 2 diabetes: a real-life worldwide observational study (EDGE). *Int J Clin Pract.* 2013;67(10):947–956.
67. Siris ES, Selby PL, Saag KG, Borgstrom F, Herings RM, Silverman SL. Impact of osteoporosis treatment adherence on fracture rates in North America and Europe. *Am J Med.* 2009;122(Suppl 2):S3–S13.

## Clinical Epidemiology

### Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification,

Submit your manuscript here: <https://www.dovepress.com/clinical-epidemiology-journal>

Dovepress

systematic reviews, risk and safety of medical interventions, epidemiology and biostatistical methods, and evaluation of guidelines, translational medicine, health policies and economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.