Appendix 1: Supplementary data

More Detailed Methods

Data Analysis

Analyses were performed using Review Manager (Cochrane Collaboration, Oxford, UK). Review Manager includes continuity corrections of 0.5 in RCTs to allow inclusion of RCTs with no events in one treatment arm. Traditionally, RCTs with no events in either the intervention or control groups are excluded in binary outcome meta-analysis using RR; however, as a sensitivity analysis, continuity corrections were also used in RCTs with no events in either treatment arm to allow their inclusion as previously described.[1] These latter calculations were carried out using standard equations in Microsoft Excel and continuity corrections based on the reciprocal of the group (i.e., treatment or control) opposite the zero cell as proposed by Sweeting and colleagues[2] rather than 0.5 to minimize bias. (A recently published statistical simulation study suggests other approaches to inclusion of such studies.[3]) Random effects models[4] which incorporate between-trial heterogeneity and give wider and more conservative confidence intervals (CI) when heterogeneity is present were used for all analyses. Statistical heterogeneity among trials was assessed using the I^2 statistic, defined as the percentage of total variability across studies attributable to heterogeneity rather than chance.[5] Relative risks (RR) were used to pool outcomes. Individual trial and summary results are reported with 95% CIs. *A priori*, Z-tests of interaction were used to calculate interaction p-values comparing RRs between the larger trials with cardiovascular primary outcomes to the smaller trials, and between pairs of subgroups of RCTs using different DPP-4 inhibitors to determine whether treatment effects differed between agents; and *post hoc* between subgroups of patients with and without a previous history of HF. To assess for publication bias, a funnel plot comparing effect measure for the primary outcome of HF requiring hospitalization to

study precision was examined for evidence of asymmetry and this was tested statistically using both the Egger[6] regression and Begg and Mazumdar[7] rank correlation tests as implemented in Comprehensive Meta Analysis, Version 3.3.070 (available at www.Meta-Analysis.com).

References

1. Friedrich JO, Adhikari NK, Beyene J. Inclusion of zero total event trials in meta-analyses maintains analytic consistency and incorporates all available data. *BMC Med Res Methodol.* 2007;7:5.
2. Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Stat Med.* 2004;23:1351-75.
3. Kuss O. Statistical methods for meta-analyses including information from studies without any events-add nothing to nothing and succeed nevertheless. *Stat Med.* 2015;34:1097-116.
4. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials.* 1986;7:177-88.
5. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ.* 2003;327:557-60.
6. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ.* 1997;315:629-34.
7. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics.* 1994;50:1088-101.

Supplementary Results

Visual inspection of the funnel plot showed no evidence of asymmetry (Supplementary Figure 1) and there was no statistical evidence for publication bias (P=0.44 using the Egger regression test and P=0.59 using the Begg and Mazumdar rank correlation test).

ClinicalTrials.gov was also systematically searched using the same search strategy but limited to "Interventional" "Recruiting" (n=147) and "Active, not Recruiting" (n=35) studies. The search identified three ongoing RCTs with primary cardiovascular outcomes of which one is comparing a DPP-4 inhibitor to active control (CAROLINA, linagliptin vs. glimepiride, n=6115, estimated completion March 2019; ClinicalTrials.gov, NCT01243424), and two that are comparing DPP-4 inhibitors to placebo: 1) CARMELINA (linagliptin vs. placebo, n=8,300, targeted completion January 2018; ClinicalTrials.gov, NCT01897532) and 2) MK-3102-018 (omarigliptin vs. placebo, n=4,202, stopped in May 2016 but results currently unpublished; ClinicalTrials.gov, NCT01703208). Assuming baseline HF rates requiring hospitalization in the control group similar to the average of the large RCTs of 3.1% (2.8% for SAVOR-TIMI 53, 3.3% for EXAMINE, 3.1 % for TECOS), then an average RR of as little as 1.03 for both CARMELINA and MK-3102-018 would make the relative risk for HF requiring hospitalization statistically significant pooling only data from the large RCTs with primary cardiovascular outcomes (Supplementary Figure 2A, which assumes RR=1.03 for both CARMELINA and MK-3102-018). These hypothetical results are similar even if one uses baseline HF rates requiring hospitalization half or double the average rate of 3.1% as long as the RR of the two ongoing RCTs are similar. Alternatively, if the RRs are different, for example if the smaller RCT (MK-3102-018) has a RR of 0.9, then the larger RCT (CARMELINA) would have to have a RR at least as high as EXAMINE (RR=1.18), for the pooled random-effects RR for HF requiring hospitalization of all five cardiovascular safety RCTs to be statistically significant (Supplementary Figure 2B, which assumes RR 0.90 for MK-3102-018 and RR 1.18 for CARMELINA). (This is due to the effects of

heterogeneity: similar trial results for the ongoing RCTs lower overall heterogeneity and narrow confidence intervals when pooling the results of all five RCTs, while non-similar trial results increase heterogeneity leading to wider confidence intervals.) In any case, these calculations illustrate that the results from the two ongoing DPP-4 inhibitor vs. placebo cardiovascular-safety RCTs will be important as they could have an impact on the pooled risk estimates for HF among the cardiovascular safety RCTs.

Supplementary Table 1: Search Strategies

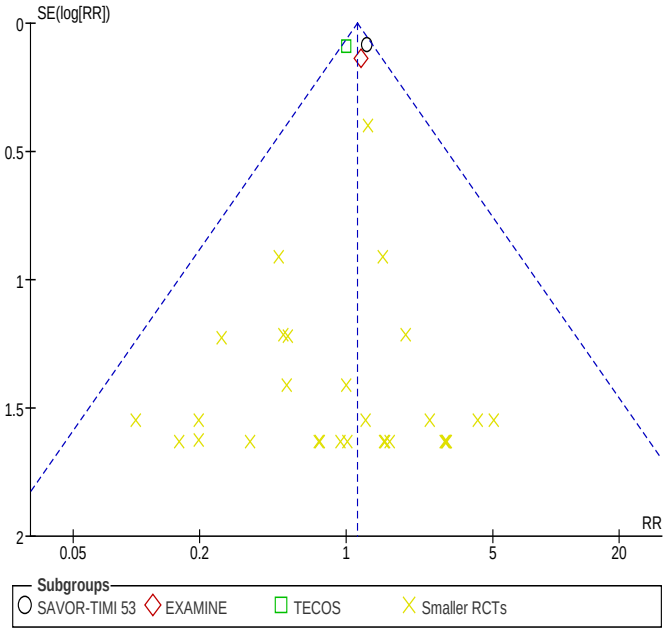
I) Medline Search Strategy:
Database: Ovid MEDLINE(R) <1946 to August Week 2 2016>, Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations <August 19, 2016>
Search Strategy:
1 Dipeptidyl-Peptidase 4 Inhibitors.mp. or exp Dipeptidyl-Peptidase IV Inhibitors/ (3124)
2 sitagliptin.mp. (1572)
3 saxagliptin.mp. (451)
4 vildagliptin.mp. (757)
5 linagliptin.mp. (399)
6 alogliptin.mp. (310)
7 gemigliptin.mp. (24)
8 teneligliptin.mp. (51)
9 anagliptin.mp. (27)
10 omarigliptin.mp. (10)
11 or/1-10 (4081)
12 limit 11 to (clinical trial, phase iii or clinical trial, phase iv or clinical trial or randomized controlled trial) (614)

II) Supplementary Embase Search Strategy:
Database: Embase <1980 to 2016 Week 34> <August 19,2016>
Search Strategy:
1 Dipeptidyl-Peptidase 4 Inhibitors.mp. or exp Dipeptidyl-Peptidase IV Inhibitors/ (11779)
2 sitagliptin.mp. (5883)
3 saxagliptin.mp. (2104)
4 vildagliptin.mp. (2916)
5 linagliptin.mp. (1452)
6 alogliptin.mp. (1159)
7 gemigliptin.mp. (79)
8 teneligliptin.mp. (120)
9 anagliptin.mp. (97)
10 omarigliptin.mp. (36)
11 or/1-10 (11916)
12 limit 11 to (clinical trial or randomized controlled trial or phase 3 clinical trial or phase 4 clinical trial) (2104)
13 limit 12 to exclude medline journals (349)

III) ClinicalTrials.gov Search Strategy: <August 22, 2016>
"dipeptidyl peptidase 4 inhibitors" OR "alogliptin" OR “SYR-322” OR "linagliptin" OR “BI-1356” OR "saxagliptin" OR “BMS-477118” OR "sitagliptin” OR “MK-0431” OR "vildagliptin" OR “LAF237” OR “gemigliptin” OR “LC15-0444” OR “teneligliptin” OR “”MP-513” OR “anagliptin” OR “CWP-0403” OR “omarigliptin” OR “MK-3102” OR “gosogliptin” OR “PF-00734200” | Interventional Studies | Completed | (528)

SUPPLEMENTARY FIGURES

Supplementary Figure 1: Funnel Plot

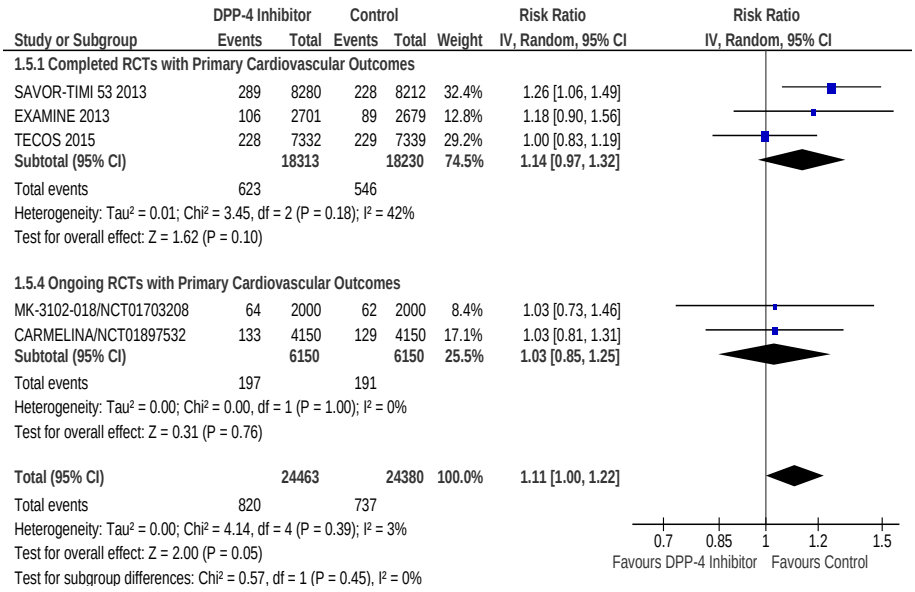


Funnel plot comparing the effect measure, relative risk (RR), for the heart failure outcome for the large RCTs with primary cardiovascular outcomes (SAVOR-TIMI 53, EXAMINE, and TECOS) and the 28 smaller RCTs with at least one patient with heart failure on the x-axis, with its precision, expressed as the standard error of the natural logarithm of RR, SE(log[RR]), on the y-axis to assess for asymmetry. There was no statistical evidence of publication bias (P=0.44 using the Egger regression test and P=0.59 using the Begg and Mazumdar rank correlation test).

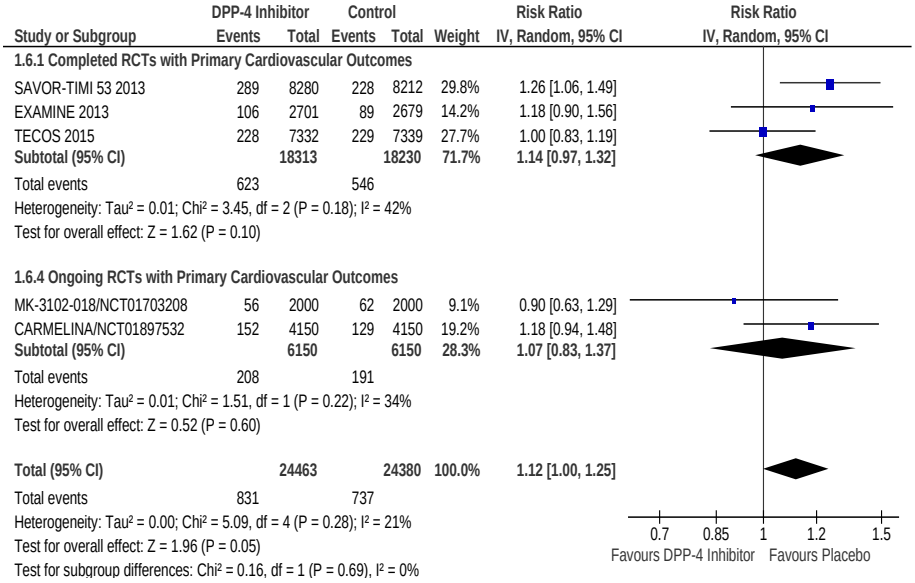
Abbreviations: RCTs, randomized controlled trials; RR, relative risk; SE, standard error.

Supplementary Figure 2: Forest Plot for Heart Failure of Large Randomized Controlled Trials with Primary Cardiovascular Outcomes: Potential Impact of Ongoing Trials

A



B



Individual and pooled risk ratios (RR) with 95% confidence intervals (CI) for completed and ongoing randomized controlled trials (RCTs) with a primary outcome that included cardiovascular outcomes and reported the number of patients in each treatment group that were hospitalized for heart failure as an adjudicated primary or secondary outcome. For the ongoing trials, CARMELINA and MK-3102-018, baseline event rates were assumed similar to SAVOR-TIMI 53, EXAMINE, and TECOS (3.1%). Panel A

shows that if the average RRs for the ongoing RCTs are identical, RRs of as little as 1.03 results in a statistically significant increased heart failure rate. Panel B shows that if the RRs for the ongoing RCTs are different, then for a RR of 0.90 for MK-3102-018, the RR for CARMELINA needs to be at least 1.18 to produce a significantly increased pooled heart failure rate for all five cardiovascular safety RCTs combined. The pooled RRs with 95% CI were calculated using random-effects models. Weight refers to the contribution of each study to the overall pooled estimate of treatment effect. Each square and horizontal line denotes the point estimate and 95% CI for each trial's RR. The diamonds signify the pooled RR; the diamond's centre denotes the point estimate and width denotes the 95% CI.