Validity of Meta-analysis in Diabetes: We Need to Be Aware of Its Limitations

To deliver high-quality clinical care to patients with diabetes and other chronic conditions, clinicians must understand the evidence available from studies that have been performed to address important clinical management questions. In an evidence-based approach to clinical care, the evidence from clinical research should be integrated with clinical expertise, pathophysiological knowledge, and an understanding of patient values. As such, in an effort to provide information from many studies, the publication of diabetes meta-analyses has increased markedly in the recent past, using either observational or clinical trial data. In this regard, guidelines have been developed to direct the performance of meta-analysis to provide consistency among contributions. Thus, when done appropriately, meta-analysis can provide estimates from clinically and statistically homogeneous but underpowered studies and is useful in supporting clinical decisions, guidelines, and cost-effectiveness analysis. However, often these conditions are not met, the data considered are unreliable, and the results should not be assumed to be any more valid than the data underlying the included studies. To provide an understanding of both sides of the argument, we provide a discussion of this topic as part of this two-part point-counterpoint narrative. In the point narrative as presented below, Dr. Home provides his opinion and review of the data to date showing that we need to carefully evaluate meta-analyses and to learn what results are reliable. In the counterpoint narrative following Dr. Home's contribution, Drs. Golden and Bass emphasize that an effective system exists to guide meta-analysis and that rigorously conducted, high-quality systematic reviews and meta-analyses are an indispensable tool in evidence synthesis despite their limitations.

—William T. Cefalu, md
Editor in Chief, Diabetes Care

Meta-analysis is a means of combining data from studies with a common end point with the intention of providing a more reliable estimate of the effect size of some intervention or observation (1). Effect size is useful in medicine in deciding whether an intervention has a benefit-to-risk ratio high enough to justify clinical use, and in economic analysis for cost-effectiveness. Combining data usually narrows CIs, allowing tighter limits to be set on potential outcomes and making statistical significance more likely. But early on concern was expressed about the limits of meta-analysis, concerns often now ignored (2).

The number of reports in which "meta-analysis" and "diabetes" are both contained in the title has expanded rapidly in recent years (Fig. 1). Sadly, many of these studies have significant weaknesses escaping the combined skills of peer reviewers and editors. Meanwhile the clinical community and the media seem to believe that anything that reaches statistical significance and is the result of meta-analysis has some mystical access to unchallengeable truth. This is then widely disseminated to those without the skills needed to make appropriately reserved judgments.

An approach that unintentionally promotes this misunderstanding is evidence grading. Often, this is a hierarchy in which randomized controlled trials (RCTs) are top rank, together with meta-analysis of these (3). Unfortunately RCTs are themselves of very variable quality, and often the data extracted from them for meta-analysis can be secondary and unreliable (4). The same is true of observational studies. However since the word meta-analysis is only given in the top rank of the data hierarchy, the nonexpert can be forgiven for assuming any meta-analysis must be first rank evidence.

This article, while acknowledging the usefulness of meta-analysis, attempts to highlight some of the method's weaknesses. The author is a clinical researcher not a statistician, so the technical issues are not fully covered, but instead problems accessible to the clinician.

## SOME PROBLEMS OF META-ANALYSIS

Unreliable primary data

The best data for meta-analysis is when the primary end point of a series of studies is the same. This is now beginning to be true for some cardiovascular (CV) outcomes such as myocardial infarction (MI), after years of discussion, but independent adjudication is still required (5). Otherwise, only all-cause mortality is generally regarded as reliable. CV mortality is problematic because terminal events are often CV even if another disease (e.g., cancer) is responsible, whereas sudden death is often indeterminate though usually counted as CV in the absence of other information. Furthermore investigator reporting, like death certification, is unreliable for many events—diagnoses are generally made on probabilities not certainties. Was it a stroke, a transient ischemic attack, a fit, or a syncopal event, and by what definition? This problem is compounded in meta-analyses in which the data used are trial primary or secondary end points, but are taken from adverse events, themselves subject to high levels of unreliability.

Recently, meta-analyses have been published for CV outcomes for the di-peptidyl peptidase 4 (DPP-4) inhibitors (6). Commendably the authors have often avoided use of the word "meta-analysis." The data for sitagliptin were largely taken from studies designed to address glucose control and general safety, with no special emphasis on CV safety (7). While retrospective adjudication of events can be attempted, data extracted for long past events is necessarily tenuous. Furthermore, evidence for a possible event may be missed if dedicated pages in a clinical report form were not used. More recently, to meet U.S. Food and Drug Administration requirements, such approaches to CV end points have been incorporated in glucose-lowering phase 2/3 studies, so data quality has improved. However three other problems of meta-analysis affecting the DPP-4 inhibitor CV data findings are discussed below.

An important example of "difficult data" in diabetes is hypoglycemia. Hypoglycemic symptoms are often nonspecific, particularly in those with other diseases, the elderly, and the obese, while in type 2 diabetes self-glucose testing is performed infrequently. A particular problem here is nocturnal events. The situation in type 1 diabetes is now complicated by erratic and sometimes biased use of continuous glucose monitoring. As a result of these problems, hypoglycemia event rates vary markedly between studies for what seem like similar interventions in similar populations, with no clarity as to which study

Point
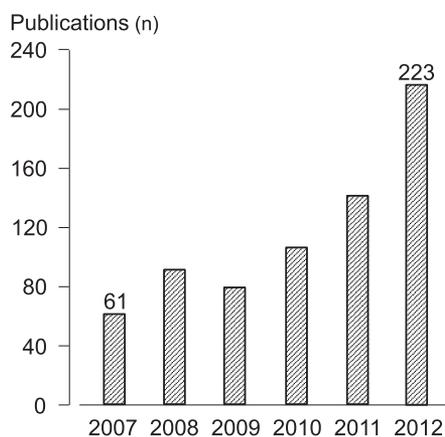
Publications (n)

Figure 1—*The number of publications in the medical literature including both the words "meta-analysis" and "diabetes" in their title for the years 2007 to 2012.*

is more valid. Reported hypoglycemia rates tend to decrease with study duration, perhaps due to reporting fatigue. The effect of this is that hypoglycemia meta-analyses may give some useful estimate of relative rate of hypoglycemia on two therapies (8) but cannot give useful estimates of the more important absolute change.

The problems of study duration
Meta-analysis can be used to disguise some of the simple study duration issues that affect single studies. For example, most phase 3 studies (median in diabetes around 26 weeks) would be too short to identify the kind of fracture/osteoporotic problems that stem from thiazolidinedione use (9). But meta-analysis of a series of short-term studies may allow sufficient numbers of fractures to be accumulated to limit the upper CI, and the agent may then be provisionally judged safe, quite inappropriately (10). Other dangers relate to oncogenic effects and work both ways: Time may be necessary to uncover

genotoxic or indeed growth-promoting effects, whereas for dapagliflozin investigations preexposure suggests that the apparent imbalance in bladder cancer malignancies was at least partly related to clinically silent disease (11).

Less obvious may be combined analyses of studies of very different duration. The type 2 diabetes MI/CV meta-analyses used four or five studies that included both the early terminated Action to Control Cardiovascular Risk in Diabetes (ACCORD; 3.5 years median) and the extension of the UK Prospective Diabetes Study (UKPDS; 15 years) (12–16). The meta-analyses agree on a 15–18% reduction in CV outcomes for a 10 mmol/mol (0.9% unit) reduction in HbA_{1c}. However beyond confirming blood glucose control matters, that is not clinically useful because of the impossibility of understanding how long such control has to be maintained to gain the benefit. Note also the problem here of the expression of the result in terms of HbA_{1c} change: 10 mmol/mol (0.9% units) is difficult to achieve early after diagnosis once lifestyle change is in place (17), but later in the disease (after 20 years) use of multiple agents including insulin may be delivering a 33 mmol/mol (3.0%) reduction—can the gain be multiplied up?

Study populations
Related to the last problem is conflation of results from different study populations. It is clear that ACCORD (obese North Americans), Action in Diabetes and Vascular Disease (ADVANCE; global, including many oriental Asians), Prospective Pioglitazone Clinical Trial in Macrovascular Events (PROactive; secondary CV), and Veterans Affairs Diabetes Trial (VADT; mainly men) are very different populations (Table 1) (15,18–20). The participants in each study were all studied relatively late in the course of their disease (e.g., at entry 8 to

12 years from diagnosis), yet combined with them is UKPDS with a population studied from 3 months after diagnosis (16). It does not follow that the results of meta-analysis will be generalizable to any particular population, or to any individual with diabetes.

This problem is not addressed by statistical tests of heterogeneity such as Cochran Q or Altman I^2 (21). These statistics consider whether the numerical results (e.g., the hazard ratio [HR] and CIs of it) of the studies are consistent or different. However they are blind to clinical issues so that, for example, no statistical heterogeneity might be found in studies done in a mixture of rats, dogs, and humans if the HRs were comparable, or when combining data from different antihypertensive drug classes. If I^2 is high, then examination of the studies will generally quickly identify why that is, but really the meta-analysis ought never to have been done in the first place.

Many of the phase 3 glucose-lowering studies used for secondary CV meta-analysis are of individuals selected as being in particularly good health, including a low prior incidence of CV disease. This may mean that meta-analyses of phase 3 studies are not suitable for assessment of CV risk, for the reports will not be reflecting risk in typical diabetes populations. A reverse issue relates to testing for CV protection. Because the evidence from UKPDS and the late megatrials is that glucose-lowering benefit only kicks in in the longer term, then no benefit is expected to be seen in early studies unless there is another pathogenetic mechanism operating. Indeed if the early indications from the incretin studies are real (6), then this would point to nonglucose-lowering mechanisms.

Tempting targets for meta-analysis exist in diabetes but have not been performed for reasons of population heterogeneity. Examples are the rosiglitazone versus

Table 1—*Summary of some characteristics and findings of the major cardiovascular outcome studies of people with type 2 diabetes as used in meta-analyses*

Outcome or study	Time from diagnosis (years)	Events (n)	Study duration (years)	HR (95% CI)	P
MI/stroke/death					
ACCORD (CV deaths) (15)	10	723	3.5	0.90 (0.78–1.04)	NS
PROactive (all deaths) (18)	8	659	2.9	0.84 (0.72–0.98)	0.027
Total CV events					
VADT (20)	12	494	6	0.88 (0.74–1.05)	NS
ADVANCE (19)	8	1,147	5	0.94 (0.84–1.06)	NS
MI only					
UKPDS extension (16)	0.3	997	17	0.84 (0.75–0.97)	0.01

metformin data for malignancy from A Diabetes Outcome Progression Trial (ADOPT) and Rosiglitazone Evaluated for Cardiovascular Outcomes and Regulation of Glycemia in Diabetes (RECORD) (22). Both studies found no difference, such that combined data would give an estimate closer to 1.00 with tighter CIs. But even before remembering one study was of monotherapy in relatively obese North Americans, and the other of combination oral agent therapy across broader European populations, is more information gained by combining the two estimates? Another thiazolidinedione example is for stroke. Using the same studies used by Nissen and Wolski (23) for MI, stroke was statistically significantly reduced by rosiglitazone compared with mixed comparators (24). Furthermore, stroke in RECORD, PROactive, and ADOPT also gave HRs below 0.80 (none statistically significant) (18,25,26), making it clear that meta-analysis would give a significant headline result (Table 2).

Such cautions did not prevent an analysis of heart failure data (27). Little was served by this—heart failure was a known problem of thiazolidinediones from the late 1990s, and adequately powered data were then available from a single RCT (28). But Singh et al. (27) included in their four selected studies one which specifically recruited people with heart failure, and thus with very different absolute outcome rates for both further failure events, CV events, and death (29). As a result the meta-analysis findings cannot be applied to any clinical population or further economic assessments.

Even measurements such as HbA_{1c} and body weight can lack validity when combined from seemingly similar studies performed as part of a phase 2/3 development program. HbA_{1c} change is known to be baseline dependent—starting therapy at higher levels results in larger falls (30,31). Accordingly, taking unadjusted data from studies with different baseline values, such as monotherapy and dual and triple combination therapy, has no useful clinical meaning but is still performed (32). Body weight change is similarly affected by baseline HbA_{1c} because of the effects of falls in calorie loss through glycosuria—simple meta-analysis is similarly incapable of giving a meaningful result.

Data snooping and multiple testing

A danger is that breach of statistical norms may be followed by publication of a meta-analysis without realization that the process of statistical testing has been undermined. One example of this is for malignancy in which multiple meta-analyses of different organ sites are followed by highlighting of a single outlier organ result. This occurred in one of the insulin glargine quartet of articles in *Diabetologia* in 2009, in which breast cancer popped up as apparently statistically significant in the Swedish database study (33).

More subtly, two other examples have occurred in diabetes care, "more subtly" because it is not possible even for the authors themselves to know the extent to which they proceeded with meta-analysis only after noting that it was likely to generate a significant and therefore publishable result. The five large glucose-lowering CV trials all delivered central HR estimates below 1.00 (Table 2) (15, 16,18–20). That in itself is close to statistical significance ($P = 0.0625$), and with upper CIs all below 1.10 or close to that, anyone familiar with statistical analysis will know that meta-analysis will give a statistically significant result. Three such publications followed the next year (12–14). Similarly the early central estimates for CV protection in all five available composite analyses of the DPP-4 inhibitors were below 1.00 (Fig. 2). Not surprisingly then, this presented an attractive target for meta-analysis and presentation and publication (6). This reviewer predicts with some confidence that this meta-analysis finding is not correct but biased toward a positive effect, as will be shown by the purpose designed RCTs (34,35).

This means that such meta-analyses are hypothesis generating, not hypothesis testing. Nissen and Wolski (23) acknowledged this in open session when discussing their rosiglitazone CV meta-analysis in 2007. Such reservations have not stopped clinical commentators in esteemed journals writing that the authors "have shown" that rosiglitazone causes increased "CV events." A hypothesis-generating article cannot "show" anything. Further, since stroke was not addressed (see above), it is not CV events that were considered, only MI and mortality.

Drug regulators have a legitimate interest in data snooping when trying to understand the importance of nonstatistically significant trends in safety issues in the limited data available in licensing application. Here the intention is not to establish the actual safety of a new medication, but rather to develop a feel for the probability of whether such a trend might reflect a real issue. Predefined meta-analysis of the insulin degludec data for CV outcomes was broadly reassuring, though based on small numbers, but data snooping in weaker data using smaller comparisons and extension studies less so (36,37). In these circumstances, the predefined analysis should be that quoted by researchers and clinicians, the secondary analyses being reserved for regulatory decision making only.

Study population sizes

Chance plays games with allocation of events to active and comparator sides of randomized study populations. Data monitoring committees often watch an adverse outcome accumulate eight events in one group and only one in the other. But multiple small studies are not immune

Table 2—*Characteristics of studies reporting stroke while taking thiazolidinediones showing clinical heterogeneity, which should prevent meta-analysis despite the face value attractiveness of similar HRs*

Study	Drug	Comparator	Event rate	Study duration (years)	Stroke (HR [95% CI])
Early rosiglitazone (23)	Rosiglitazone	Mixed	Low	<1	0.48 (0.23–0.98)
PROactive (18)	Pioglitazone	Placebo	High	2.9	0.81 (0.61–1.07)
ADOPT (25)	Rosiglitazone	Glibenclamide or metformin	Low	4.0	0.77 (0.38–1.59)
RECORD (26)	Rosiglitazone	Metformin or sulfonylurea	Low	5.5	0.72 (0.49–1.06)
Meta-analysis of above*	Mixed	Mixed	Mixed	Mixed	0.75 (0.61–0.92)

*The meta-analysis is performed purely for illustrative purposes for the current article; it is not intended to guide clinical practice.
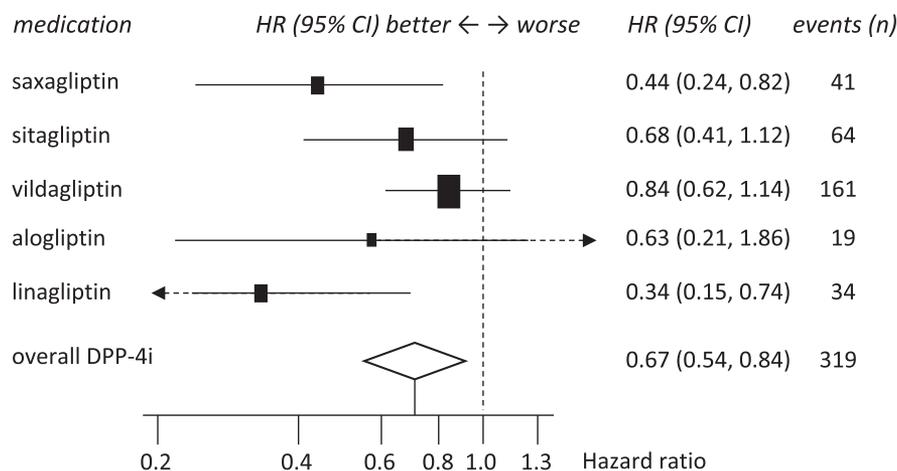
Figure 2—*Presented or published data for meta-analyses of the early studies of the five licensed and unlicensed DPP-4 inhibitors for which cardiovascular outcome data (usually composites of CV death, nonfatal MI, and nonfatal stroke) are available (for references see reference 6), and a meta-analysis of the five reports prepared by the current author (52). The findings presented should not be used to suggest efficacy when making clinical decisions (see article text).*

to this effect, as in the reported outcomes for CV effects in the early studies for saxagliptin and linagliptin (38,39). A clue here is the forest plot in which the larger composite studies, notably the study with vildagliptin, give central hazard estimates closer to 1.00 (Fig. 2) (6). Indeed, the vildagliptin finding is probably a better guide to the likely result of the formal outcome studies for the other agents than their own estimates from meta-analysis. A useful rule of thumb here is if total event number is <50, beware; if <100, be skeptical; and if <150, be cautious.

One well-known example is the meta-analysis by Nissen and Wolski (23) of the early rosiglitazone studies, after a similar analysis put in the public domain by GlaxoSmithKline some 8 months earlier (40). Nissen and Wolski were constrained to meta-analyze studies that mostly had <5 events in all, several indeed with 0 events. Because a HR cannot be calculated for the latter, they were excluded, but of course 0 versus 0 is equal hazard (and not zero information) thus inflating the final difference. Recalculation by more sophisticated methodology halved the central estimate of adverse effect and rendered it nonstatistically significant (41).

Problems inherent in included studies

Meta-analysis cannot correct for the inherent problems of the included studies, with the exception of inadequate study size (event rate and power). Some attempt is often made to deal with this by consideration of study quality, using

published criteria such as Jadad scoring or the GRADE approach (42,43). Unfortunately these criteria deal mainly with technical issues and make no approach to determining clinical quality (e.g., whether a clinically appropriate comparator is chosen, whether the enrolled population is appropriate, and duration of follow-up) (Table 3). Those without clinical expertise may then perform meta-analysis with limited generalizability to a clinical population.

A particular problem relates to meta-analysis of observational studies. Observational studies of clinical interventions, all too easy to perform in these days of large electronic clinical databases, are subject to hidden confounding, biases that which are not known (44,45). Unfortunately the same bias might be expected in a series of database studies, and when meta-analysis is then used the findings become seemingly more solid. It is difficult by definition to provide examples of something that is unknown, but a couple of semihypothetical scenarios can be given. Metformin is a long-established medication used early in the course of diabetes since the results of the UKPDS substudy published in 1998, but many practitioners were brought up understanding that contraindications included liver disease, renal disease, and cardiac disease. As a result, there is a bias in its use to younger, shorter duration patients and those with better CV outcomes (46). By contrast, insulin is used later in the course of diabetes, in individuals with more complex overall health problems and often taking multiple therapies.

Accordingly metformin does well in prescription databases, and insulin does badly in, for example, malignancy and CV disease (46,47). This will be true for

Table 3—Some factors of importance in assessing clinical trial quality of relevance to meta-analysis

Generalizability
    Appropriate comparator and statistical analyses
    Choice of end points (and definitions thereof)
    Study power and duration of study
    Representative participant population
        Age, sex, ethnic mix, geographical mix
        Duration of disease, severity of disease
        Prior disease, comorbidity
    Therapy regimens
        Dose and frequency
        Concomitant therapies
        Participant education
        Rescue therapy
    Investigator competence
Minimization of internal bias
    Appropriate randomization and blinding
    Equal provision of other aspects of care
    Systematic collection of valid outcome data
    Unbiased assessment of outcome, including outcome selection
    Withdrawal and lost to follow-up biases
    No bias in deviation from protocol
Reporting quality (e.g., by Jadad criteria [42])
    Adequate description of
        Randomization
        Blinding
        Withdrawals

meta-analysis as well as the original observational studies but is negated by properly conducted RCTs. Less obviously, two medications in one class may be marketed to different target populations, for example one to primary care and one to specialists, the latter seemingly having worse outcomes, particularly on meta-analysis. Use of newer medications by teaching hospitals, which tend also to see more critically ill and complex medical cases, is another potentially confounding example.

Available studies

As well as issues discussed above—for example, short-term studies with poor data quality—available studies may limit conclusions through problems with comparators. When metformin was compared with other oral agents for malignancy, statistical heterogeneity was good as judged by an I^2 of 13%, implying comparable results in the underlying studies, while the HR 0.98 (95% CI 0.77–1.23), implied no advantage (48). However, 67% of events come from comparisons with a thiazolidinedione, limiting the meaning of conclusions compared with other glucose-lowering therapies.

Lack of balance in study size may mean that a meta-analysis is merely a restatement of the results of the largest study. Lincoff et al. (49) published a meta-analysis of CV events with pioglitazone, in which 80% of the events came from PROactive. The remainder of the data came from several shorter studies with diverse comparators and ill-defined end point collection, but in reality the findings are a restatement of those of PROactive. If an outcome study is reasonably designed and conducted, it should displace anything based on poorer quality or observational data.

CONCLUSIONS

CONCLUSIONS—The problems discussed above suggest that the statistical and headline-grabbing power of meta-analysis must be applied and interpreted with caution. In the CV studies of glucose-lowering in which underpowering occurred, the finding of statistically significant reduction of MI and CV events is useful—but only when applied in the context of other information such as duration of the intervention and the populations that might benefit (Table 3). Taken together, the current DPP-4 inhibitor data are also positive, but the extent of the advantage is probably overestimated, such that,

beyond assurance of CV safety, there is little information that is clinically applicable (Fig. 2).

On the insulin glargine cancer issue, the careful meta-analyses by Boyle and colleagues (50,51) have been of assistance in clarifying an area of some concern even before the reporting of the Outcome Reduction with Initial Glargine Intervention (ORIGIN) study, but here again the meta-analyses of the RCTs were limited by the duration, size, and quality of the underlying studies. The care taken by Boyle and colleagues in handling the observational studies does emphasize however the importance of considering the clinical quality of the underlying data and not just using technical criteria such as Jadad scoring.

Meta-analysis will remain important in the regulatory area, in which weak data particularly on safety and long-term outcomes are usual at the time of submission for licensing and what is needed is the best possible understanding of the possibilities and probabilities of risks and benefits. For this reason alone, meta-analysis will be with us for the foreseeable future, but clinicians and the media do need to learn what results are reliable, what are merely signals of possible effect, and what are associations without likely causation. Editors and reviewers of manuscripts have a duty to ensure cautions are explicit even when the headline results are given emphasis, while those presenting results to the media sometimes need more humility in making their findings public.

PHILIP D. HOME, DM, DPHIL

From the Institute of Cellular Medicine, The Medical School, Newcastle University, Newcastle upon Tyne, U.K.
Corresponding author: Philip D. Home, philip.home@ncl.ac.uk.

References

1. Spector TD, Thompson SG. The potential and limitations of meta-analysis. J Epidemiol Community Health 1991;45:89–92
2. Thompson SG, Pocock SJ. Can meta-analyses be trusted? Lancet 1991;338:1127–1130
3. Scottish Intercollegiate Guidelines Network. Key to evidence statements and grades of recommendations. SIGN 50: A guideline developer's handbook [Internet]. Available from http://www.sign.ac.uk/guidelines/fulltext/50/annexb.html. Accessed 16 November 2012
4. Jüni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. BMJ 2001;323:42–46
5. US Food and Drug Administration. Guidance for industry: diabetes mellitus—evaluating cardiovascular risk in new antidiabetic therapies to treat type 2 diabetes [PDF online]. Available from http://www.fda.gov/downloads/Drugs/GuidanceCompliance-RegulatoryInformation/guidances/ucm071627.pdf. Accessed 1 August 2012
6. Monami M, Ahrén B, Dicembrini I, Mannucci E. Dipeptidyl peptidase-4 inhibitors and cardiovascular risk: a meta-analysis of randomized clinical trials. Diabetes Obes Metab 2013;15:112–120
7. Williams-Herman D, Engel SS, Round E, et al. Safety and tolerability of sitagliptin in clinical studies: a pooled analysis of data from 10,246 patients with type 2 diabetes. BMC Endocr Disord 2010;10:7
8. Davidson JA, Liebl A, Christiansen JS, et al. Risk for nocturnal hypoglycemia with biphasic insulin aspart 30 compared with biphasic human insulin 30 in adults with type 2 diabetes mellitus: a meta-analysis. Clin Ther 2009;31:1641–1651
9. Kahn SE, Zinman B, Lachin JM, et al.; Diabetes Outcome Progression Trial (ADOPT) Study Group. Rosiglitazone-associated fractures in type 2 diabetes: an analysis from A Diabetes Outcome Progression Trial (ADOPT). Diabetes Care 2008;31:845–851
10. Monami M, Dicembrini I, Antenore A, Mannucci E. Dipeptidyl peptidase-4 inhibitors and bone fractures: a meta-analysis of randomized clinical trials. Diabetes Care 2011;34:2474–2476
11. Burki TK. FDA rejects novel diabetes drug over safety fears. Lancet 2012;379:507
12. Ray KK, Seshasai SR, Wijesuriya S, et al. Effect of intensive control of glucose on cardiovascular outcomes and death in patients with diabetes mellitus: a meta-analysis of randomised controlled trials. Lancet 2009;373:1765–1772

13. Turnbull FM, Abraira C, Anderson RJ, et al.; Control Group. Intensive glucose control and macrovascular outcomes in type 2 diabetes. Diabetologia 2009;52:2288–2298

14. Mannucci E, Monami M, Lamanna C, Gori F, Marchionni N. Prevention of cardiovascular disease through glycemic control in type 2 diabetes: a meta-analysis of randomized clinical trials. Nutr Metab Cardiovasc Dis 2009;19:604–612

15. Gerstein HC, Miller ME, Byington RP, et al.; Action to Control Cardiovascular Risk in Diabetes Study Group. Effects of intensive glucose lowering in type 2 diabetes. N Engl J Med 2008;358:2545–2559

16. 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;359:1577–1589

17. UK Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). Lancet 1998;352: 837–853

18. Dormandy JA, Charbonnel B, Eckland DJ, et al.; PROactive investigators. 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;366: 1279–1289

19. Patel A, MacMahon S, Chalmers J, et al.; ADVANCE Collaborative Group. Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes. N Engl J Med 2008;358:2560–2572

20. Duckworth W, Abraira C, Moritz T, et al.; VADT Investigators. Glucose control and vascular complications in veterans with type 2 diabetes. N Engl J Med 2009;360: 129–139

21. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327:557–560

22. Home PD, Kahn SE, Jones NP, Noronha D, Beck-Nielsen H, Viberti G; ADOPT Study Group; RECORD Steering Committee. Experience of malignancies with oral glucose-lowering drugs in the randomised controlled ADOPT (A Diabetes Outcome Progression Trial) and RECORD (Rosiglitazone Evaluated for Cardiovascular Outcomes and Regulation of Glycaemia in Diabetes) clinical trials. Diabetologia 2010; 53:1838–1845

23. Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. N Engl J Med 2007;356:2457–2471

24. GlaxoSmithKline. Advisory Committee Briefing Document Cardiovascular Safety of Rosiglitazone, FDA Endocrinologic and Metabolic Drugs Advisory Committee and Drug Safety and Risk Management Advisory Committee. King of Prussia, PA: GlaxoSmithKline, 2007. Available from http://www.fda.gov/ohrms/dockets/ac/07/briefing/2007-4308b1-01-sponsor-backgrounder.pdf Accessed 6 January 2012

25. Kahn SE, Haffner SM, Heise MA, et al.; ADOPT Study Group. Glycemic durability of rosiglitazone, metformin, or glyburide monotherapy. N Engl J Med 2006;355: 2427–2443

26. Home PD, Pocock SJ, Beck-Nielsen H, et al.; RECORD Study Team. Rosiglitazone evaluated for cardiovascular outcomes in oral agent combination therapy for type 2 diabetes (RECORD): a multicentre, randomised, open-label trial. Lancet 2009;373: 2125–2135

27. Singh S, Loke YK, Furberg CD. Long-term risk of cardiovascular events with rosiglitazone: a meta-analysis. JAMA 2007;298: 1189–1195

28. Home PD, Pocock SJ, Beck-Nielsen H, et al.; RECORD Study Group. Rosiglitazone evaluated for cardiovascular outcomes—an interim analysis. N Engl J Med 2007;357: 28–38

29. Dargie HJ, Hildebrandt PR, Riegger GA, et al. A randomized, placebo-controlled trial assessing the effects of rosiglitazone on echocardiographic function and cardiac status in type 2 diabetic patients with New York Heart Association Functional Class I or II Heart Failure. J Am Coll Cardiol 2007;49:1696–1704

30. DeFronzo RA, Stonehouse AH, Han J, Wintle ME. Relationship of baseline HbA$_{1c}$ and efficacy of current glucose-lowering therapies: a meta-analysis of randomized clinical trials. Diabet Med 2010;27:309–317

31. Deacon CF, Mannucci E, Ahrén B. Glycaemic efficacy of glucagon-like peptide-1 receptor agonists and dipeptidyl peptidase-4 inhibitors as add-on therapy to metformin in subjects with type 2 diabetes—a review and meta analysis. Diabetes Obes Metab 2012;14:762–767

32. Karagiannis T, Paschos P, Paletas K, Matthews DR, Tsapas A. Dipeptidyl peptidase-4 inhibitors for treatment of type 2 diabetes mellitus in the clinical setting: systematic review and meta-analysis. BMJ 2012;344:e1369

33. Jonasson JM, Ljung R, Talbäck M, Haglund B, Gudbjörnsdòttir S, Steineck G. Insulin glargine use and short-term incidence of malignancies-a population-based follow-up study in Sweden. Diabetologia 2009; 52:1745–1754

34. Bethel MA, Green J, Califf RM, Holman RR. Rationale and design of the Trial Evaluating Cardiovascular Outcomes with Sitagliptin (TECOS) (Abstract). Diabetes 2009;58(Suppl. 1):A639

35. Scirica BM, Bhatt DL, Braunwald E, et al. The design and rationale of the saxagliptin assessment of vascular outcomes recorded in patients with diabetes mellitus-thrombolysis in myocardial infarction (SAVOR-TIMI) 53 study. Am Heart J 2011;162:818–825, e6

36. Food and Drugs Administration. FDA briefing document: NDA 203313 and NDA 203314: Insulin degludec and insulin degludec/aspart. Endocrinologic and Metabolic Drug Advisory Committee, 2012. Available from http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/EndocrinologicandMetabolicDrugsAdvisoryCommittee/UCM327015.pdf. Accessed 7 November 2012

37. Novo Nordisk. Insulin degludec and insulin degludec/insulin aspart treatment to improve glycemic control in patients with diabetes mellitus. NDAs 203314 and 203313. FDA Briefing Document, Endocrinologic and Metabolic Drug Advisory Committee, 2012. Available from http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/EndocrinologicandMetabolicDrugsAdvisoryCommittee/UCM327017.pdf. Accessed 7 November 2012

38. Frederich R, Alexander JH, Fiedorek FT, et al. A systematic assessment of cardiovascular outcomes in the saxagliptin drug development program for type 2 diabetes. Postgrad Med 2010;122:16–27

39. Johansen OE, Neubacher D, von Eynatten M, Patel S, Woerle H-J. Cardiovascular safety with linagliptin in patients with type 2 diabetes mellitus: a pre-specified, prospective, and adjudicated meta-analysis of a phase 3 programme. Cardiovasc Diabetol 2012;11:3

40. SmithKline Beecham: Avandia (rosiglitazone): summary of product characteristics. 2006. Welwyn Garden City, U.K.: SmithKline Beecham, 2000

41. Diamond GA, Bax L, Kaul S. Uncertain effects of rosiglitazone on the risk for myocardial infarction and cardiovascular death. Ann Intern Med 2007;147:578–581

42. Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials 1996;17:1–12

43. Balshem H, Helfand M, Schünemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 2011;64: 401–406

44. Beral V. "The practice of meta-analysis": discussion. Meta-analysis of observational studies: a case study of work in progress. J Clin Epidemiol 1995;48:165–166

45. Johnston SC. Identifying confounding by indication through blinded prospective review. Am J Epidemiol 2001;154:276–284

46. Schramm TK, Gislason GH, Vaag A, et al. Mortality and cardiovascular risk associated with different insulin secretagogues compared with metformin in type 2 diabetes, with or without a previous myocardial

infarction: a nationwide study. Eur Heart J 2011;32:1900–1908

47. Bowker SL, Majumdar SR, Veugelers P, Johnson JA. Increased cancer-related mortality for patients with type 2 diabetes who use sulfonylureas or insulin. Diabetes Care 2006;29:254–258

48. Stevens RJ, Ali R, Bankhead CR, et al. Cancer outcomes and all-cause mortality in adults allocated to metformin: systematic review and collaborative meta-analysis of randomised clinical trials. Diabetologia 2012;55:2593–2603

49. Lincoff AM, Wolski K, Nicholls SJ, Nissen SE. Pioglitazone and risk of cardiovascular events in patients with type 2 diabetes mellitus: a meta-analysis of randomized trials. JAMA 2007;298:1180–1188

50. Boyle P. Insulin glargine and cancer risk: latest data. International Diabetes Federation Congress Dubai 2011, Late breaking clinical trials. Available from http://conference2.idf.org/prog.aspx. Accessed 30 August 2012

51. Boyle P, Koechlin A, Boniol M, Robertson C, Bolli GB, Rosenstock J. Updated meta-analysis of cancer risk among users of insulin glargine (Abstract). Diabetes 2012; 61(Suppl. 1):A345

52. Home P. Improved cardiovascular outcomes with DPP-4 inhibitors in diabetes – a meta-analysis (Abstract). Atherosclerosis 2012;223:533