Clinical Translation of Cardiovascular Outcome Trials in Type 2 Diabetes: Is There More or Is There Less Than Meets the Eye?

Ele Ferrannini[1] and Julio Rosenstock[2]

PERSPECTIVES IN CARE

Randomized controlled trials (RCTs) have become the gold standard of clinical evidence and the staple of guided clinical practice. RCTs are based on a complex set of principles and procedures heavily strung by statistical analysis, primarily designed to answer a specific question in a clinical experiment. Readers of clinical trials need to apply critical appraisal skills before blindly accepting the results and conclusions of trials, lest they misinterpret and misapply the findings. We introduce the fundamentals of an RCT and discuss the relationship between relative risk (RR) and absolute risk (AR) in terms of the different information each conveys. The top results of some recent cardiovascular outcome trials using sodium–glucose cotransporter 2 inhibitors and glucagon-like peptide 1 receptor agonists in patients with type 2 diabetes are used to exemplify the merit of assessing both RR and AR changes for a balanced translation of findings into shrewd clinical judgment. We also suggest practical points to assist with a clinically useful interpretation of both within-trial and across-trial reports. Finally, we mention an alternative approach, namely, the restricted mean survival time, to obtaining unbiased estimates of the mean time of missed events in the treatment versus placebo arm for the duration of the trial.

Evolving treatment recommendations in type 2 diabetes are being driven by randomized controlled trials (RCTs), which have become the gold standard of clinical evidence and the staple of guided clinical practice. Reports of RCTs now crowd the pages of both general medicine and specialist journals, with a focus on cardiovascular outcome trials (CVOTs). Cardiologists have dominated the field of CVOTs, and endocrinologists/diabetologists have come relatively late into the game, when a decade ago the U.S. Food and Drug Administration mandated that all new diabetes agents demonstrate cardiovascular (CV) safety in properly powered CVOTs enriched in populations at high CV risk. Since 2015, a high number of CVOTs in type 2 diabetes have been published, and more are due to report in the next 2–4 years.

The concept of an RCT is based on a rather complex set of principles and procedures heavily strung by statistical analysis, primarily designed to answer a specific question in a clinical experiment. Over the years, it has come to have rigid rules and formalism and grown its own jargon (1). An emergent "trialist community" (1) now includes clinical investigators, regulators, developers, and payers. An industry of expensive contract research organizations for data monitoring and collection has sprung up and rapidly expanded, including large academic consortia for developing, adjudicating, conducting, and reporting CVOT results. At large, an RCT mindset has permeated the entire process of clinical investigation (project proposals, funding requests, ethics permission, adjudication committees, peer review processes, and even journal

[1]CNR Institute of Clinical Physiology, Pisa, Italy
[2]Dallas Diabetes Research Center at Medical City, Dallas, TX

Corresponding author: Ele Ferrannini, ferranni@ifc.cnr.it

instructions for authors). Yet, interpreting and translating an RCT into clinical practice are less easy than it would appear. Readers of clinical trials need to apply critical appraisal skills before blindly accepting the results and conclusions of trials, lest they misinterpret and misapply the findings. We hereby share what we have learned from having been involved in designing, conducting or interpreting, and reporting some of the CVOTs.

Relative Versus Absolute Risk

Top-line results of an RCT typically consist of Kaplan-Meier curves, i.e., estimated instantaneous incidence (hazard) event rates [IR] in the trial arms (say, placebo [IR_{Plb}] and treatment [IR_{Tx}]), thereby visually illustrating the difference of an explicitly predefined event (e.g., cardiovascular disease [CVD] death). The ratio of IR_{Tx} to IR_{Plb} cumulated over time is the relative risk (RR). The analysis (Cox proportional hazards regression) is expressed as a hazard ratio (HR) (assumed to be proportional between the two arms), its 95% CI, and P value: HR is the ratio of IR_{Tx} to IR_{Plb} at any time during follow-up. If HR is <1 and the CI does not include 1, then the treatment can be interpreted to have allowed a longer event-free survival compared with placebo (Fig. 1 in ref. 2 for CVD death). If HR exceeds 1 and the CI does not include 1, the treatment is deemed to have been harmful compared with placebo. Especially when expressed in terms of percent RR reduction, such a way of reporting may appear impressive or, even, "paradigm shifting" (e.g., reference 2). Other end points, whether primary, prespecified, or adjudicated events or secondary observations from the trial, are formally treated in the same way in hierarchical order. Canonically, if the primary end point does not reach statistical significance (the ubiquitous $P < 0.05$), the hierarchical analyses are aborted and further probability assessment (i.e., "nominal" P values) is qualified as "exploratory" and given less credit. Alternatively, one can prespecify a coprimary outcome (thereby "splitting the α," i.e., raising the statistical significance threshold) or secondary outcomes across multiple end points. Such downstream analyses can be used to at least glimpse some "hypothesis-generating" findings to earmark for direct testing in subsequent ad hoc studies or real-world data analyses. In theory, this statistical architecture should make it possible for

the clinician to assign different clinical relevance to the different outcomes of a study. In many cases, we submit that it falls short of providing a realistic clinical perspective.

Many RCTs also report estimates of the absolute risk (AR) or risk difference ($IR_{Tx} - IR_{Plb}$), in units of cumulative event rates (i.e., percentage of individuals in each arm at study end) or annualized rates (per 1,000 person-years [py]) (e.g., Table 1 in reference 2). How are RR and AR related to one another, and what differential information do they carry?

Figure 1 plots the relationship between RR and AR over a range of RRs—from favoring treatment to favoring placebo—for different base incident event rates. Note that because of randomization, IR_{Plb} can be safely taken to be the base incidence rate of the whole cohort. It can be seen that at any given base rate the relationship between RR and AR is a straight line hinged on the fulcrum (where RR is 1 and AR difference is 0). The slope of this relationship becomes steeper as the base risk rate decreases. For example (green dotted lines in Fig. 1), an RR of 0.4 (= 60% RR reduction by treatment vs. placebo) translates into a negative AR (= AR reduction) of 60/1,000 py (or 6% per

year) at a base rate of 100/1,000 py (= 10% per year) but only to an AR reduction of 12/1,000 py (= 1.2% per year) at a base rate of 20/1,000 py (= 2% per year). Thus, AR is strongly dependent on the base risk rate. The classical case in point is provided by the statin trials, where similar RR reductions in major vascular events (HR ~0.8 per each mmol/L of LDL cholesterol lowering) have been observed across a ≥10-fold range of base risk rates, thereby yielding large differences in AR reduction among trials (3). The reciprocal of a negative AR (= AR reduction) is the number needed to treat (NNT), which estimates the number of subjects to be treated to "save" (prevent or delay) one event over the duration of the trial. In the example above, in the high-risk population (base rate = 100/1,000 py) the NNT is 17 subjects, which escalates to 83 subjects in the low-risk population (base rate = 20/1,000 py). Conversely, the same AR is associated with decreasing HRs as the base risk rate increases.
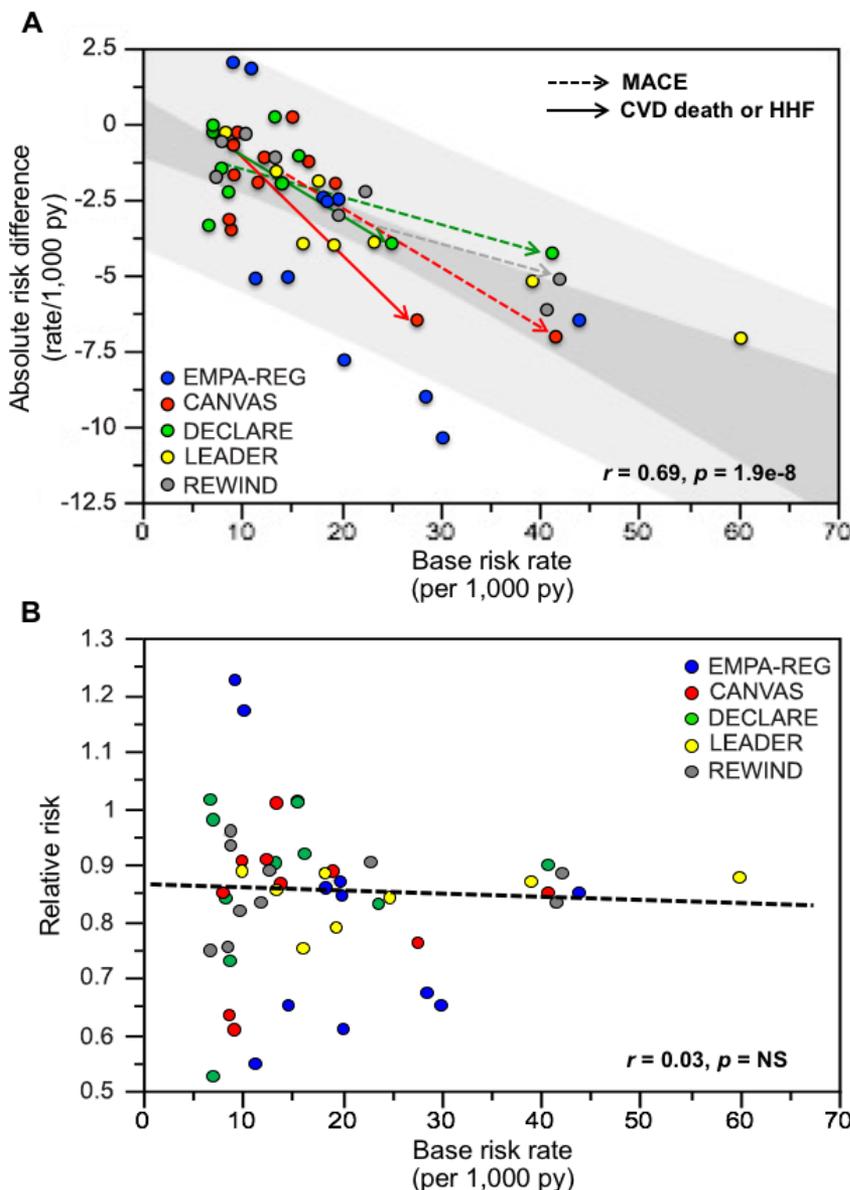
Note that on the right side of the nomogram in Fig. 1, a positive AR (= AR increase) tells the number needed to harm (NNH). For example (red dotted lines in Fig. 1), an AR increase of 30/1,000
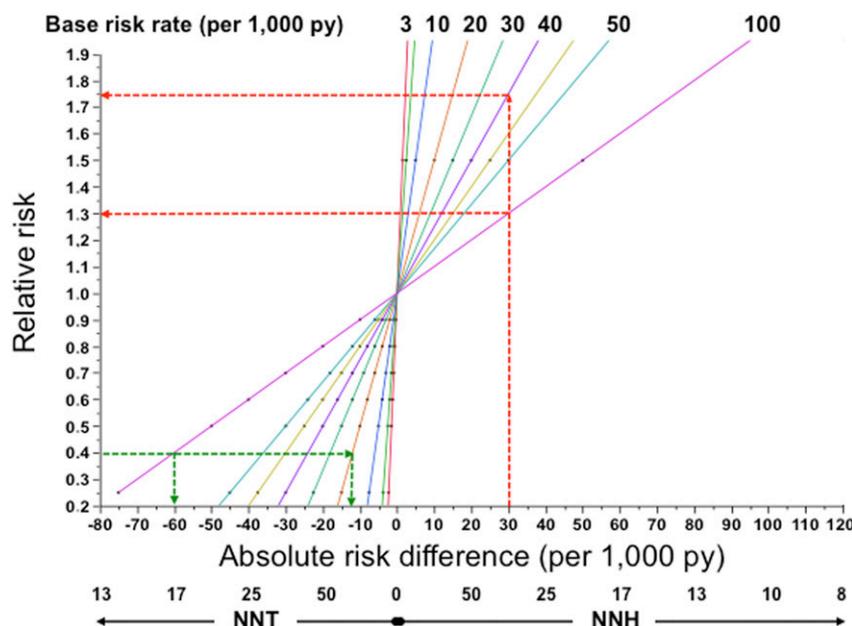
Figure 1—Relationship between RR and AR difference over a range of base risk rates. See text for further explanation.

Five CVOTs: Major Outcomes

Figure 2—*A*: Relationship between AR difference (ΔIR in Table 1) and base risk rate ($= IR_{Plb}$ in Table 1) for several clinically relevant outcomes in five recent CVOTs (listed in Table 1). The darker-shaded area includes the 95% CI of the fit and the lighter-shaded area the 95% CI of individual points. The arrows connect data from the same trial in which the analysis of major adverse CV events (MACE) was reported separately for patients with established atherosclerotic CVD (higher risk) or multiple CV risk factors (lower risk). *B*: Relationship between RR difference (HR in Table 1) and base risk rate ($= IR_{Plb}$ in Table 1) for several clinically relevant outcomes in five recent CVOTs (listed in Table 1). HHF, hospitalization for heart failure; ΔIR, change in absolute incident event rate.

py corresponds to an HR of 1.75 for a base rate of 40/1,000 py and an HR of 1.30 for a base rate of 100/1,000 py. Therefore, in trial reports both RR and AR should be highlighted and brought to the reader's attention for full understanding of the quantitative aspects of the trial and to allow comparison with existent literature (i.e., other treatments, different patient populations, etc.).

Incidentally, the balance between NNT and NNH could be used to assess the benefit-to-risk ratio. For instance, with reference to Fig. 1, suppose that in an intervention trial carried out in a high-risk population (base rate = 10% per year) the AR reduction of CV death with treatment is 6% per year, but a fatal, non-CV adverse event clearly related to the treatment is recorded in 3% per year of

the exposed group. Thus, NNT is 17 for 1 year and the NNH is 8 for 1 year. Therefore, for every 100 people treated with the theoretical intervention, 17 people who would have otherwise had a CV death will be alive, whereas 8 people who would otherwise have been alive would have died. Therefore, for the outcome of total mortality, for every 100 people, 9 people who would have otherwise died would have been alive at the end of a year. That suggests overall net benefit. The point is that one must scrutinize overall benefits and harms of an intervention and not just focus on CV benefits while ignoring non-CV harms.

Of course, this is an extreme hypothetical scenario, but it introduces a point of current special interest to the clinician, as it pertains to trials of sodium–glucose cotransporter 2 inhibitors (SGLT2i) in patients with diabetes. Use of this class of drugs has been reported to increase the incidence of so-called euglycemic diabetic ketoacidosis (DKA) in a few patients with type 2 diabetes (4) and, mainly, in a small proportion of patients with type 1 diabetes (5). DKA is rarely fatal but is a serious adverse event of major clinical impact (6). In relatively small trials of short duration, however, it is difficult to capture the real incidence of DKA precisely because it is rare. In a recent 52-week trial of sotagliflozin (a dual SGLT2/1 inhibitor) in adults with type 1 diabetes (7), the 400 mg/day dose was associated with a DKA incidence (IR_{Tx}) of 4.2% vs. the 0.4% incidence (IR_{Plb}) detected in the placebo arm, corresponding to an RR (IR_{Tx}/IR_{Plb}) of 10.5. In a recent update of the T1D Exchange clinic registry (8), the frequency (IR_{Plb}) of at least one DKA event in the 3 months prior to date of censoring in 1,525 patients (>18 years old with a complete electronic questionnaire) was 2.6%. If the RR from the placebo-controlled trial of sotagliflozin (i.e., 10.5) were to be extrapolated to the registry data above, one would predict a 2.6% × 10.5 = 27.3% incidence of DKA with use of the drug in the "real world" of the T1D Exchange registry patients, an obviously unacceptable risk. Instead, if the AR increase in the trial (4.2 − 0.4 = 3.8%) were added to the base rate of 2.6% of the registry, then the *expected* DKA rate in the registry population would be 6.4% and the corresponding RR would be 2.5 and not 10.5.

Table 1—Incidence rates (per 1,000 py) in the placebo arm (IR$_{Plb}$), rate difference (ΔIR or change in AR in percent), and HR for major clinical outcomes in five CVOTs

	EMPA-REG OUTCOME			CANVAS			DECLARE			LEADER			REWIND		
	IR$_{Plb}$	ΔIR	HR	IR$_{Plb}$	ΔIR	HR	IR$_{Plb}$	ΔIR	HR	IR$_{Plb}$	ΔIR	HR	IR$_{Plb}$	ΔIR	HR
MACE$_{CVD}$	43.9	−6.5	0.86	41.3	−7.2	0.83	41.0	−4.2	0.90	39.0	−5.0	0.87	42	−5.0	0.87
MACE$_{MRF}$	—	—	—	15.8	−0.3	1.02	13.3	0.1	1.01	—	—	—	20	−3.0	0.87
All-cause death	28.6	−9.2	0.68	19.5	−2.2	0.87	16.4	−1.3	0.92	25.0	−4.0	0.84	22.9	−2.3	0.90
CV death	20.2	−7.8	0.62	12.8	−1.2	0.87	7.1	−0.1	0.99	16.0	−4.0	0.75	13.4	−1.2	0.91
HHF	14.5	−5.1	0.65	8.7	−3.2	0.67	8.5	−2.3	0.73	14.0	−2.0	0.86	8.9	−0.6	0.93
All MI	19.3	−2.5	0.87	12.6	−1.4	0.89	13.2	−0.5	0.96	—	—	—	9.1	−0.4	0.96
Nonfatal MI	18.5	−2.5	0.87	11.6	−1.9	0.85	13.2	−1.5	0.89	18.0	−2.0	0.89	8.4	−0.4	0.96
All strokes	10.5	1.8	1.18	9.6	−1.7	0.87	—	—	—	—	—	—	8.1	−2.0	0.76
Nonfatal stroke	9.1	2.1	1.24	8.4	−1.3	0.90	6.8	0.1	1.01	10.0	−1.0	0.91	6.9	−1.7	0.76
HHF or CV death$_{CVD}$	30.1	−10.0	0.65	27.4	−6.4	0.77	23.9	−4.0	0.83	60.0*	−7.0	0.88	—	—	—
HHF or CV death$_{MRF}$	—	—	—	9.8	−0.9	0.91	8.4	−1.4	0.83	—	—	—	—	—	_
CKD progression	11.5	−5.2	0.61	9.0	−3.5	0.61	7.0	−3.3	0.53	19.0	−4.0	0.79	40.7	−6.0	0.85

MACE, hospitalized heart failure (HHF), CV death: for each of these outcomes, the subscript CVD is the value in patients with established CVD, whereas the subscript MRF is the value in patients with major CV risk factors. CKD, chronic kidney disease; ΔIR, change in absolute incident event rate; LEADER, Liraglutide Effect and Action in Diabetes: Evaluation of Cardiovascular Outcome Results; MI, myocardial infarction. *Expanded MACE.

This automatic transfer of RR across different base incidence rates should therefore be avoided.

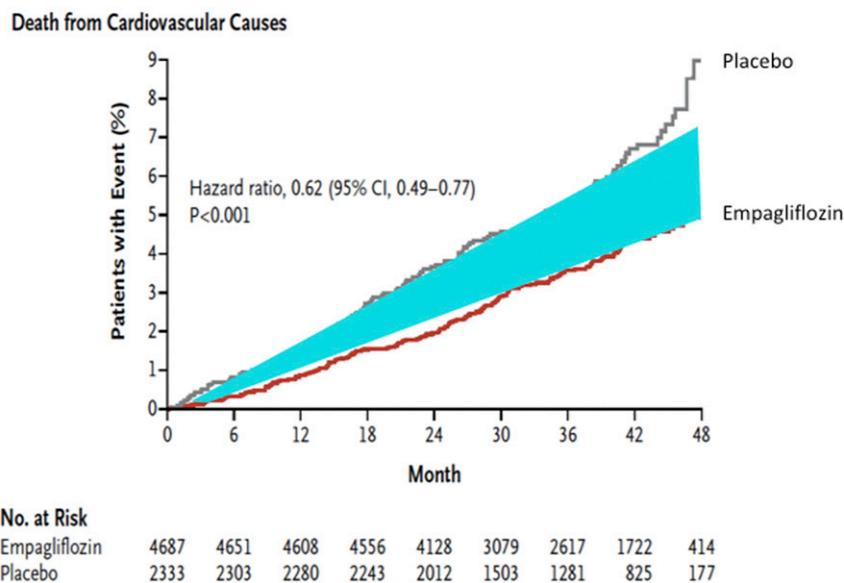### Within-Trial and Between-Trial Comparison

It is important to consider that RCTs should not be compared even when treatment consists of agents of the same class (e.g., statins, SGLT2i, glucagon-like peptide 1 receptor agonists) because of differences in population characteristics and study size and duration as well as unknown confounding. Indeed, reading the methods sections of published RCTs and, especially, the information detailed in online supplements, one finds a labyrinthian list of criteria for participant inclusion/exclusion; event adjudication; definition of primary, nonprimary, and composite end points; and adverse events. Parenthetically, a criterion implies a threshold (e.g., doubling of serum creatinine as a renal end point, QTc prolongation >450 ms on the electrocardiogram as an exclusion/inclusion criterion, etc.). Shifting any one of these thresholds predictably reclassifies patients as well as outcomes and may impact the results of a trial (and the statistics thereof). By and large, clinical medicine is about imposing thresholds onto continuous biological processes, and one has to accept that a trial—any trial—carries a load of more or less explicit assumptions, under which its outcome applies. Nevertheless, it is usually very tempting to venture into trial comparisons, in particular in trying to argue a class effect. Moreover, RCTs are large, standardized, and costly studies, which produce a huge amount of carefully collected data; in publications, the corresponding information is collapsed into just a few statistical indicators (usually, the HR of prespecified outcomes). Extracting as much information of potential clinical relevance as possible therefore is a thrifty, almost necessary, endeavor.

Commonly, summary results (less frequently, individual data) of different trials are pooled and analyzed together (by event category) to estimate *1*) consistency, *2*) average effect size, and *3*) true incidence of adverse events. Such meta-analyses have become very popular, virtually a subspecialty of sorts in all fields of medicine (>50,000 matches in a PubMed search of "meta-analysis and trial"). When all HR point estimates fall on the same side of the unity line (and heterogeneity—the I^2 value—is low), a meta-analysis yields an average effect

Restricted Mean Survival Time

Death from Cardiovascular Causes

Hazard ratio, 0.62 (95% CI, 0.49–0.77)
P<0.001

No. at Risk

Empagliflozin	4687	4651	4608	4556	4128	3079	2617	1722	414
Placebo	2333	2303	2280	2243	2012	1503	1281	825	177

Figure 3—Graphical representation of the RMST as the area between the placebo and treatment Kaplan-Meier function. Redrawn from ref. 2.

Table 2—Practical points for a correct and clinically useful interpretation of a CVOT report

- Trial design (definition of primary and secondary end points, statistical power, duration, placebo or comparator)
- Base risk rate, incident event rate (in annualized units), and cumulative events of the primary outcome, drop-out and drop-in rates, rescue treatments
- Intent-to-treat *and* per-protocol analysis
- RR and AR
- Base rate and incident rate of all nonprimary outcomes, including all systematically collected clinical observations
- Adverse events, in number and annualized rates
- Kaplan-Meier plot of major or frequent adverse events
- Context (comparison with other trials of the same or different pharmacological agent targeting the same or similar primary outcome)
- Availability of RMST differences to allow the patient and the clinician to choose a given treatment in the context of the individual patient's risk level and preferences

size and a more robust estimate of the incidence of low-frequency adverse events. With lesser consistency and greater heterogeneity of trial results, the evidence derived from a meta-analysis is less than compelling. However, clinicians pose yet different questions: they often wonder whether the same end point differs *across* trials and, when treatment induces multiple effects, whether their size is comparable *within* each trial. A simple way of providing qualitative answers to such questions is exemplified in Fig. 2*A*, which plots the change in AR against the base (= placebo arm) event rate (both in annualized rate units) for every outcome reported in the primary publications of CVOTs of SGLT2i—empagliflozin (2), canagliflozin (9), and dapagliflozin (10)—and glucagon-like peptide 1 receptor agonists liraglutide (11) and dulaglutide (12) (the corresponding numbers are in Table 1). As can be seen, across all end points there is a fairly good reciprocal association, such that for any given end point, the higher the placebo (or base) event rate, the larger the drop in AR. The data (central estimates) from all five trials fall along the fit without too much deviation from its 95% CI (dark-shaded area). One does not really need *P* values to infer that AR reduction is unlikely to be large if the base risk in the population studied is low. Actually, in Canagliflozin Cardiovascular Assessment Study (CANVAS), Dapagliflozin Effect on CardiovascuLAR Events (DECLARE), and Researching Cardiovascular Events With a Weekly INcretin in Diabetes (REWIND), the analysis of major adverse CV events was reported separately for patients with established atherosclerotic CVD (higher risk) or multiple CV risk factors (lower risk), as was "CVD death or hospitalization for heart failure" in CANVAS and DECLARE (13). As indicated by the connecting arrows in Fig. 2*A*, the within-trial results align very well onto the general regression line, confirming the expectation that the higher the base risk, the greater the AR reduction in the same trial. In contrast, plotting the corresponding RRs against the base risk rates yields a cloud (Fig. 2*B*), giving no clue as to what happened within and across trials in populations with variable base rates of the same end points. Incidentally, in trials in very-low-risk populations in which full statistical significance of the primary outcome is near missed, it is not unusual to hear that the study was not large enough: increasing the sample size in a very-low-risk population is expected to lead to narrower CIs, but the *point estimate* of the risk change may be little affected precisely because of the reciprocal relationship exemplified in Fig. 2.

Further qualitative assessment of the overall clinical benefit of the trials can be gained from Fig. 2. For example, for the three SGLT2i studies the obvious inference is that one is dealing with a class effect. In addition, the demanding clinician may set her/his own thresholds of decision making, for example, by deeming as acceptable an absolute reduction of at least 2.5/1,000 py for clinical end points with a baseline incidence of ≥5/1,000 py. In the case of SGLT2i, within such a "clinically acceptable" area one finds CVD death, hospitalization for heart failure (and their composites), and progression of nephropathy for all three trials,

confirming the relative homogeneity of the clinical impact of these three drugs.

Finally, one may misread the Kaplan-Meier plot of event-free survival, e.g., Fig. 1 in the BI 10773 (Empagliflozin) Cardiovascular Outcome Event Trial in Type 2 Diabetes Mellitus Patients (EMPA-REG OUTCOME) (2), by assuming, for example, that an HR of 0.68 for CVD death implies that empagliflozin can "save" 32% of deaths: prospective patients may exult at such news. What the result actually means is that treatment was associated with a 32% lower *population chance* of observing CVD death relative to placebo *within the trial time period* in a high-risk cohort with a base event rate of 20/1,000 py. In EMPA-REG OUTCOME, the absolute CV death risk reduction was 2.2% and the NNT to prevent one CV death over a median of 3 years was 46. A further limitation of the standard "HR approach" is that HR may change during the trial because of the early removal of high-risk patients; furthermore, the HR is influenced by treatment discontinuation or dropping out of participants (14,15). In this case, the principle of proportionality is violated and the Cox regression analysis is inappropriate; a keen reader should look and see if this condition is explicitly tested in a trial report.

Other Approaches
A more rigorous way of quantifying trial outcomes—particularly when the Kaplan-Meier functions show irregular time courses—is to consider the restricted mean survival time (RMST), which estimates the mean (and CI) time of "missed" events in the treatment versus placebo arm for the duration of the trial (16). Mathematically, RMST is the area between the placebo and treatment curve in a Kaplan-Meier plot (shaded area in Fig. 3). A formal calculation of this area in EMPA-REG OUTCOME yields a value of 0.7 months per person; i.e., empagliflozin treatment for 4 years results in the postponing of all-cause death by 21 days (95% CI 10–33) on average (17): a less impressive message indeed. The RMST metric, unlike the HR, is independent of assumptions, is more efficient in the analysis of noninferiority in low-risk populations, and can help in the benefit/risk assessment (17). A very informative analysis of RMSTs from the major diabetes CVOTs is in Table 2 of ref. 17, along with a critical

discussion of the advantages and limitations of this approach.

In conclusion, the RCT is a sophisticated clinical experiment necessary for proper hypothesis testing, currently informing guidelines, health providers, and clinical practice. In RCTs, however, there is both less and more than meets the eyes of the busy clinician dealing with lots of patients and loads of information (14–17). We suggest that there are tactics of reading into RCTs (summarized in Table 2) that go some way toward forming of an accurate, balanced, and more realistic basis for shrewd clinical judgement.

References

1. U.S. Food and Drug Administration. Multiple endpoints in clinical trials. 2017. Accessed 31 December 2020. Available from https://www.fda.gov/media/102657/download
2. Zinman B, Wanner C, Lachin JM, et al.; EMPA-REG OUTCOME Investigators. Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. N Engl J Med 2015;373:2117–2128
3. Collins R, Reith C, Emberson J, et al. Interpretation of the evidence for the efficacy and safety of statin therapy. Lancet 2016;388:2532–2561
4. Erondu N, Desai M, Ways K, Meininger G. Diabetic ketoacidosis and related events in the canagliflozin type 2 diabetes clinical program. Diabetes Care 2015;38:1680–1686
5. Peters AL, Buschur EO, Buse JB, Cohan P, Diner JC, Hirsch IB. Euglycemic diabetic ketoacidosis: a potential complication of treatment with sodium–glucose cotransporter 2 inhibition. Diabetes Care 2015;38:1687–1693
6. Rosenstock J, Ferrannini E. Euglycemic diabetic ketoacidosis: a predictable, detectable, and preventable safety concern with SGLT2 inhibitors. Diabetes Care 2015;38:1638–1642
7. Buse JB, Garg SK, Rosenstock J, et al. Sotagliflozin in combination with optimized insulin therapy in adults with type 1 diabetes: the North American inTandem1 study. Diabetes Care 2018;41:1970–1980
8. Miller KM, Foster NC, Beck RW, et al.; T1D Exchange Clinic Network. Current state of type 1 diabetes treatment in the U.S.: updated data from the T1D Exchange clinic registry. Diabetes Care 2015;38:971–978
9. Neal B, Perkovic V, Mahaffey KW, et al.; CANVAS Program Collaborative Group. Canagliflozin and cardiovascular and renal events in type 2 diabetes. N Engl J Med 2017;377:644–657
10. Wiviott SD, Raz I, Bonaca MP, et al.; DECLARE-TIMI 58 Investigators. Dapagliflozin and cardiovascular outcomes in type 2 diabetes. N Engl J Med 2019;380:347–357
11. Marso SP, Daniels GH, Brown-Frandsen K, et al.; LEADER Steering Committee; LEADER Trial Investigators. Liraglutide and cardiovascular outcomes in type 2 diabetes. N Engl J Med 2016;375:311–322
12. Gerstein HC, Colhoun HM, Dagenais GR, et al.; REWIND Investigators. Dulaglutide and cardiovascular outcomes in type 2 diabetes (REWIND): a double-blind, randomised placebo-controlled trial. Lancet 2019;394:121–130
13. Zelniker TA, Wiviott SD, Raz I, et al. SGLT2 inhibitors for primary and secondary prevention of cardiovascular and renal outcomes in type 2 diabetes: a systematic review and meta-analysis of cardiovascular outcome trials. Lancet 2019;393:31–39
14. Clagget B, Pocock S, Wei LJ, Pfeffer MA, McMurray JJV, Solomon SD. Comparison of time-to-first event and recurrent-event methods in randomized clinical trials. Circulation 2018;138:570–577
15. Stensrud MJ, Aalen JM, Aalen OO, Valberg M. Limitations of hazard ratios in clinical trials. Eur Heart J 2019;40:1378–1383
16. Kim DH, Uno H, Wei L-J. Restricted mean survival time as a measure to interpret clinical trial results. JAMA Cardiol 2017;2:1179–1180
17. Kloecker DE, Davies MJ, Khunti K, Zaccardi F. Uses and limitations of the restricted mean survival time: illustrative examples from cardiovascular outcomes and mortality trials in type 2 diabetes. Ann Intern Med 2020;172:541–552