Review Article

Does Sugar Control Arrest Complications in Type 2 Diabetes? Examining the Rigour in Statistical Methods and Causal Inference in Clinical Trials

Akankasha Ojha[1], Harshada Vidwans[2], Milind Watve[3]

1. Indian Institute of Science Education and Research Pune, Pune, India; 2. Independent researcher; 3. Universitat de Barcelona, Spain

Background: In contrast to type 1 diabetes mellitus (T1DM), in type 2 (T2DM) the success of intensive glucose normalisation in arresting diabetic complications is marginal and inconsistent across multiple clinical trials. However, glucose regulation still largely remains the main target of treatment for T2DM in clinical practice.

Objectives: We examine the possible causes of inconsistency across studies, the scientific rigour behind the design, conduct and inferences of 6 large institutional clinical trials targeting glucose normalisation and following up for diabetic complications and mortality.

Study design: We enumerate the possible flaws in the design, statistical treatment of the results and possible logical traps in making inferences. Further, we evaluate whether the flaws can mislead the conclusions. We also suggest a more sound statistical treatment of the data and interpret results of the trials together in a coherent way.

Results: The clinical trials for intensive glucose control suffer from a number of common problems that have not been addressed. The most important being the failure to correct for multiple outcomes. This is recognised by some reviewers but no correction is attempted. The second is the interdependence of the outcomes, owing to which statistics based on the assumption of independence can be misleading. Simulations show that the apparent inconsistency or heterogeneity between trial results can be explained by the violation of assumed independence alone. Further, the problems with placebo control, failure to recognise alternative possibilities, inability to segregate clinical significance from statistical significance and misleading reporting formats point to conformity bias and publication bias.

Conclusions: We find no support for the prevalent belief that glucose normalisation has any benefit in terms of reducing the frequency of any of the complications or mortality. It is time to reconsider the

 doi.org/10.32388/IH7KEP

glucocentric line of treatment of T2DM. Rethinking some of the fundamental beliefs about the pathophysiology of diabetic complications and facilitating novel alternative lines of research is the need of the field.

Corresponding author: Milind Watve, milind.watve@gmail.com

> "Absolute proof that good (glycaemic) control can retard or prevent the development of complications has not yet been obtained, but the assumption is accepted by most diabetologists and indeed is almost an article of faith in the current approach to achieve the best possible diabetic control" Geoffrey Gill (1991).

Diabetes was first recognised by the appearance of sugar in the urine a long time ago. Later, increased glucose in blood became known as the marker of diabetes. Insulin was discovered about 100 years ago and our thinking in diabetes has mostly revolved around the two molecules, although now over a hundred molecular, cellular and neuronal signals are known to be altered in diabetes (Watve 2013). Network models trying to integrate all these signals have raised doubts about the central and causal role of glucose and insulin in the pathophysiology of diabetes (Kulkarni et al 2017), but the mainstream clinical thinking and action continue to revolve around glucose and insulin. The differentiation of type 1 (T1DM) and type 2 diabetes mellitus (T2DM) gradually became clear over a few decades by the mid 20th century (Himsworth 1936, Dana 1954, Null et al 1973, Harris 1988, Colman et al 1999). Insulin was highly successful not only in regulating glucose but also in arresting complications in T1DM (The DCCT Research Group 1993, Nathan et al 2009, Fullerton et al 2014). Owing to the perception that T1DM and T2DM differ only in the cause of glucose dysregulation and that the downstream pathophysiology originating from chronic hyperglycaemia is almost identical in the two types, the success of glucose regulation observed in T1DM was expected to work for T2DM as well.

However, from early clinical trials of glucose regulation in type 2 diabetes, it was clear that glucose regulation in T2DM was not as effective in arresting diabetic complications as in T1DM (Meinert et al., 1970; Goldner et al., 1971; Knatterud et al., 1978; Huang et al., 2001). The debatable issue has been whether it is effective at all (Boussageon et al., 2017). The ultimate goal of drugs used in the treatment of T2DM should be the prevention of complications, but the immediate surrogate goal is believed to be the normalisation of glucose. There are two common assumptions behind treatments targeting glucose normalisation: (i) increased glucose is causal to diabetic complications and (ii) normalising glucose can

reduce the frequency of complications with statistical and clinical significance. The first major blow to this set of assumptions came from the results of the University Group Diabetes Program (UGDP), in which glucose normalisation failed to decrease complications and mortality; instead, it increased in some of the treatment arms (Miller et al., 1976; Knatterud et al., 1978). UGDP came under some criticism and controversy, and a number of other clinical trials followed in subsequent decades that used different study designs, sample sizes, drugs used, and outcome measures examined. While some of them claimed that the treatment significantly reduced certain adverse outcomes, others had a perplexing finding of increased cardiovascular and all-cause mortality in the intensively controlled group (The ACCORD study group, 2008; The NICE-SUGAR study investigators, 2009). In order to address the contradicting results, particularly in the macrovascular outcomes, and present a coherent picture, a collaboration was established between four major trials, namely UKPDS, ACCORD, ADVANCE, and VADT. A meta-analysis of the four trials concluded that only a "modest decrease" in macrovascular events was achieved by intensive glucose regulation (Turnbull et al., 2009).

However, a number of issues remain unaddressed regarding the scientific rigour in the designs of different trials, the statistical treatments used, and inferences drawn from the results. We discuss here the various possible traps and biases in the studies, possible alternative interpretations of the results, and ultimately the scientific soundness of the current set of assumptions in the treatment of type 2 diabetes. We use a comprehensive analysis of large-scale institutional, multicentre, multi-drug trials of glucose normalisation with long-term follow-up, broadening the scope of the Turnbull et al. (2009) meta-analysis, re-examining the design, statistical analysis, and inference-related issues, and clinical usefulness.

The necessity of being sceptical about published randomised clinical trials (RCT) and cross-examination of their validity has become a critical issue over the last decade or two (Ioannidis 2016, Heneghan et al 2017, Herrera-Perez et al 2019, Chow et al 2021). There have been multiple attempts to bring in reforms and rigour in RCTs. One of the measures is to make registration of the RCT mandatory before recruiting participants. Issues addressed by mandatory registration include failure to report, partial reporting, hiding inconvenient results, changing the norms of studies, issues of data transparency, etc. (Zarin and Keselman 2007, Tse et al 2009, James et al 2015, Thabane et al 2015, Ramsberg and Platt 2018). However, in spite of the increasing awareness and attempts to improve clinical trial designs, many issues still remain unaddressed. Furthermore, many of the clinical trials for glucose normalisation in T2DM began prior to the recent measures to refine clinical trial designs. Therefore, it is essential to specifically examine the

possible biases in glucose regulation trials for T2DM and their effects on the robustness of inferences and clinical usefulness.

Selection of studies and inclusion-exclusion criteria: We focus on both qualitative and quantitative issues in this article. For quantitative analysis in this study, we selected institutional clinical trials having a minimum of 100 individuals per group, a minimum follow-up of 3 years, and reporting rate ratio (RR), odds ratio (OR), or hazards ratio (HR) (or data sufficient to calculate these indices) for multiple macro and microvascular events and mortality. We excluded those that reported only surrogate outcomes. We also excluded trials conducted by pharmaceutical sector companies during drug development. The trials selected by these inclusion-exclusion criteria comprise the University Group Diabetes Program (UGDP) (Blackburn et al 2017), the UK Prospective Diabetes Study (UKPDS) (Leslie 1999), the Veterans Affairs Diabetes Trial (VADT) (Tran and Reaven 2020), the Action in Diabetes and Vascular Disease: Preterax and Diamicron Modified Release Controlled Evaluation (ADVANCE) (Heller 2009), the Action to Control Cardiovascular Risk in Diabetes (ACCORD) (The ACCORD study group 2008), and the Diabetes Prevention Programme (DPP) (Lee et al 2021).

For analysis, we take all the comparisons between a treatment or intensive treatment group with a control group without the specified treatment regime, with or without blinding/placebo. We exclude comparisons between drugs. We also exclude trial arms with specific drugs whose use was discontinued, e.g. tolbutamide in UGDP or troglitazone in DPP. Since the data needed for the intended analysis was available in the published papers, there was no need to access raw data during the study.

The endpoints/outcomes included in the analysis were those that had at least 10 events in at least one of the groups being compared. We considered all single endpoints and aggregated outcomes reported in tables, figures, as well as in the text of the published papers from the trials. It is not clear why some results were reported in tables while others were reported in the text only. However, it was deemed necessary to evaluate everything that was reported. We excluded any subgroup analysis. We also excluded any associative analysis that attempts to relate the glycaemic status with the incidence of complications by ignoring the allocation to trial regimens, such as in Hayes et al. (2013) or Stratton et al. (2000).

Since the trial designs, characteristics of patient groups, and definitions of outcomes used vary across studies, and since a meta-analysis approach has been used before, we do not use a meta-analysis approach but instead look at the results in a collective and critical way, with a focus on issues that remain unaddressed or inadequately treated in earlier literature.

We observe that the main issues unaddressed or inadequately addressed across the selected trials are as follows.

1. The total number of individually significant outcomes: Summing up over the 6 clinical trials (UGDP, UKPDS, VADT, ACCORD, ADVANCE, and DPP), out of the total 341 outcomes compared between intensive and conventional control groups, in 38 pairs, the frequency of adverse outcomes is individually significantly lower than the control, and in 26, it is individually significantly higher. However, there are repetitions and interdependence in them; for example, all-cause mortality includes cardiovascular mortality, and thereby, if the latter is significantly different, the chances of the former being significant increase. Eliminating such obvious duplications, significant reduction is seen for 17, and significant increase for 13 adverse outcomes summed up over all the 6 trials. By this simple summation, the beneficial outcomes cannot be said to be significantly greater than the harmful outcomes. If soft, subjective, and surrogate endpoints are removed, no difference in the number of beneficial and harmful results remains.
2. Appropriate use of significance level when multiple comparisons are made: When multiple statistical comparisons are made, it is possible that some of them turn out to be “significant” by chance alone. This is a well-known statistical problem, and the suggested solution is to decrease the significance level in proportion to the number of statistical tests used (Narum 2006, Rice et al. 2008). None of the clinical trials have used any of the suggested corrections for the significance level used. Turnbull et al. (2009) recognized this problem but did not attempt any solution. The total number of outcome measures reported in these trials is large. If the Bonferroni principle or any other correction for the significance level is applied, none of the effects of treatment turn out to be significant. Since the Bonferroni correction is too conservative, we should look at how many outcomes are expected to be significant at the 0.05 level by chance alone and whether the number of individually significant results observed in the trials is significantly greater than the expected.

However, when multiple statistical tests are conducted, determining how many will turn out to be individually significant by chance alone is not an easy question to answer when the different outcomes are not independent. There are two types of dependencies in the glucose normalisation trial data. One is that of repetition in aggregate endpoints as mentioned above. The other is that of having common pathways/mechanisms behind many outcomes. If the common pathway happens to differ between the two groups by chance, a number of outcomes may appear to be significant together. This problem has not been addressed in the statistical treatment of clinical trials with multiple outcome measures so far.

We address this problem using simulations (see supplementary material for details of the simulation model). The simulations generate the control and treated groups, and the outcome in each individual in the group is generated by appropriate randomisation. The ORs resulting from these simulations are studied. In the baseline simulations where the treatment effect is assumed to be zero and each outcome is assumed to be independent and have the same parameters, the ORs are distributed with a mode around unity. Being ratios, the distribution is positively skewed as expected and the ORs at both tail ends are individually significant. However, in reality, the incidence of every outcome is substantially different, for example, in the ADVANCE trial, visual impairment was seen in 54% of patients but dementia in only 0.9%. When differences in incidences are incorporated in the simulations, not only the tails but ORs with intermediate departure from unity on either side can also turn out to be significant. This is because at low incidence, a greater departure from unity is possible by chance alone, but significance is difficult owing to the smaller number of cases. On the other hand, with larger incidence, moderate departures from unity also turn out to be individually significant. This pattern matches with the clinical trials data (Figure 1).

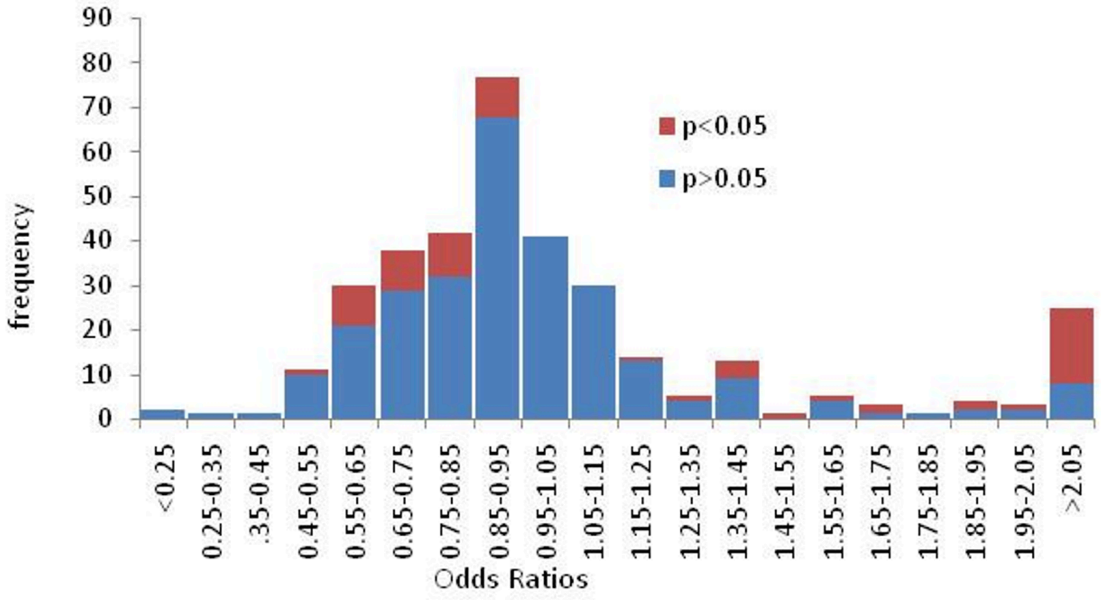

Figure 1. The frequency distribution of odds ratios for all outcomes reported in the six clinical trials. The mode and median have shifted slightly to the left from unity, but this difference cannot be said to be statistically significant. Individually significant statistical deviations are on both sides. The geometric mean of the distribution is close to one (1.01), indicating near symmetry in magnitude around “no effect”.

The simulations can incorporate interdependence in different ways. By the classical assumption, hyperglycaemia is the common trigger behind all pathophysiology of diabetic complications. If this assumption is true, all outcomes should be interdependent. Alternatively, it is also possible that some clusters of outcomes are interdependent but independent of other clusters. This assumption is relevant to T2DM since the factors involved in microvascular and macrovascular complications can be different. Further, in some tissues, capillary density appears to reduce whereas there is hyper-angiogenesis in some other complications. Simulations show that by either mode of dependence, in a large proportion of simulation runs, the proportion of outcomes individually significant can be substantially greater than the expected 5%. Therefore, having greater than 5% outcomes individually significant cannot be taken to reflect the effect of treatment. Since we do not have empirical estimates of the extent of interdependence, we cannot state how many outcomes can turn out to be individually significant by chance alone.

A possible solution to the problem is that instead of depending on individual level significance, we look at the distribution of all ORs. If there is a favourable treatment effect on the common mechanism itself, as is assumed by the classical theory of T2DM, then the entire distribution should shift to the left. The significance of this shift can be judged by multiple simulation runs. We observe that when pooled over all the six trials, the distribution of ORs has a mode and median shifted to the left by 10% and 7.3% respectively (Figure 1). By this indication, intensive glucose control may be said to have a benefit of 7 to 10%, which is compatible with the estimate of 9% obtained by Turnbull et al. (2009). This contrasts with the distribution of ORs in T1DM where almost the entire distribution of ORs lies to the left (The DCCT research group 1987, Fullerton et al. 2014, Nathan 2014). Similarly, for exercise intervention for T2DM across studies, almost all ORs are less than 1 (Sluik et al. 2012). The distribution of ORs in the T2DM trials has only a marginal leftward shift. The clear indication of this comparison is that the effects of sugar normalisation treatment are very poor and marginal compared to the effects of exercise. Whether the leftward shift in the sugar normalisation trials is significant or not is questionable. In simulations comparing 350 outcomes, comparable to the pooled data of all 6 trials, a 10% or greater shift in the mode and the median was observed in about 16% of simulation runs (see Supplementary information). Therefore, the shift observed in the pooled 6 trial data may not be considered significant.

If we assume that some clusters of outcomes have some common underlying pathway and are therefore interdependent, it is seen that a greater than expected proportion of outcomes can turn out to be individually significant by chance alone in a substantial proportion of the simulation runs. However, they

may be symmetrically or asymmetrically distributed towards the two tails. Two or more simulation runs can result in significant heterogeneity among them, although they are run at the same parameters.

Since we do not have empirical data on the extent of interdependence, simulations cannot be used to make quantitative predictions. However, they conclusively show that if the assumption of independence of chance acting on every outcome is violated, the number of false positive significances increases substantially. Unless this effect is accounted for, we cannot conclude that the reduction in some of the outcome measures claimed in the treatment groups is indeed significant and not a false positive. It is equally likely that the significant rise in mortality seen in certain trials such as the tolbutamide arm in UGDP, the metformin plus sulfonylurea group in UKPDS or in the rosiglitazone trial (RECORD) (Home et al 2009) may also be due to chance alone. If this is true, other inferences such as glucose normalising treatment being more effective for microvascular outcomes than for macrovascular ones are equally unfounded. Furthermore, the observed heterogeneity between trials (Turnbull et al 2009) is also possibly due to chance alone.

In short, since we know that certain common mechanisms are involved in the pathophysiology of diabetic complications, we should expect a large number of false positive results as well as heterogeneity between trials. Against this background, the observed individual outcomes of the trials are marginal and therefore any inferences cannot be confidently made. The shift in the mean, mode or median of the distribution is also not significant and therefore the hypothesis that hyperglycaemia is the common pathophysiological mechanism of all complications and anti-hyperglycaemic treatment can arrest complications is not supported by evidence.

3. Placebo effects: Having a double-blinded, placebo-controlled design is a standard norm for almost all clinical trials, unless a specific context makes blinding impossible. While ACCORD, ADVANCE, VADT, UGDP, and DPP had some kind of blinding or placebo control, UKPDS was an open-label trial. Plotting the distribution of ORs in the trials with and without placebo reveals that the shift in mean and median to the left is observed for UKPDS alone, which does not have a placebo control (Figure 2). The distribution of ORs in the pooled data of trials with placebo has a mode at 1 and does not show a left-shifted distribution of ORs. This raises the possibility that the apparent benefit of treatment claimed in UKPDS may only be a placebo effect. DPP allows comparison of a placebo group with a no-medication group within a single trial. There, the difference is not statistically significant but is of a comparable order (Lee et al. 2021). Some other studies have also demonstrated detectable placebo effects in diabetes treatment (Sievenpiper et al. 2007, deWit et al. 2016).

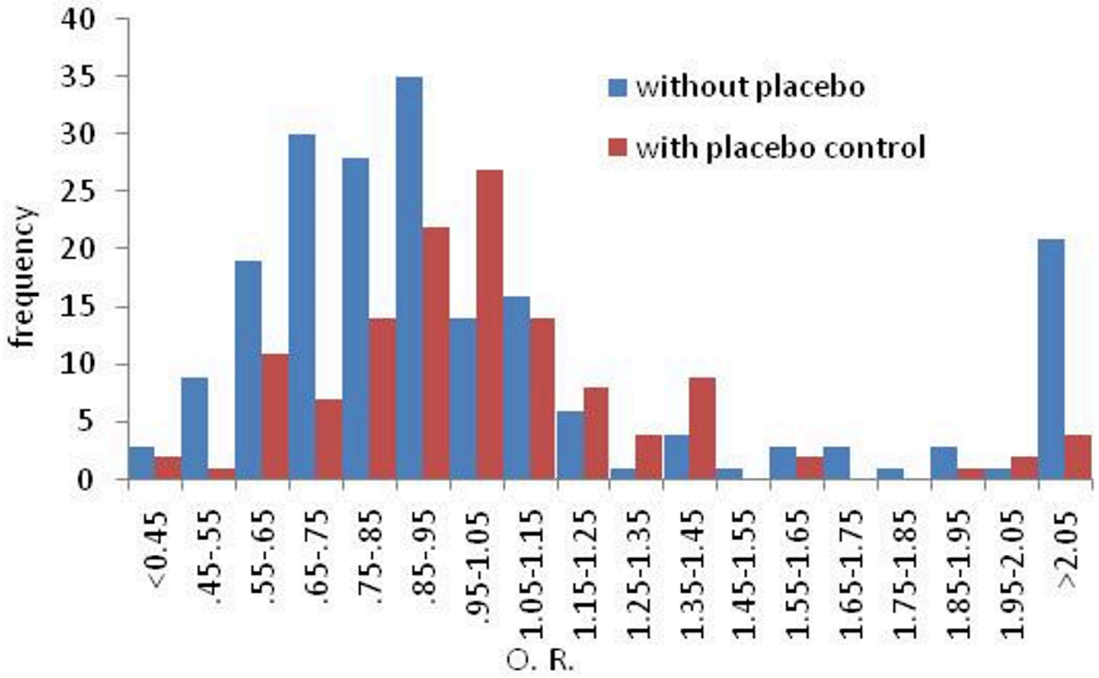

Figure 2. The distribution of ORs in trials with and without placebo control. There is a leftward shift in the distribution without placebo control. The marginal leftward shift seen in the total (Figure 1) can therefore be suspected to be a placebo effect.

In clinical trials with surrogate markers, the placebo effect can potentially operate at two distinct levels. One is in receiving the drug versus the blank. The second possible level of a placebo effect, i.e., knowing that my blood sugar is under better control, can potentially have positive psychosomatic effects. If a reduction in the frequency of complications or mortality is observed, it may be potentially explained by the second level placebo effect. So far, to the best of our knowledge, no trial has attempted to address the potential second level placebo. Unless this possibility is seriously considered, tested using appropriate controls and rejected, the marginal beneficial effects observed cannot be confidently claimed to be due to better glucose regulation.

4. Even if we assume that the marginal benefits observed in some of the outcomes are not by chance, it does not show that they are a result of glucose regulation. Across, as well as within trials, the difference in HbA1c achieved and the relative success in preventing complications do not always correlate well. Within UKPDS, the sulfonylurea and insulin arms had an HbA1c difference of 0.9, but a significant difference was seen only in the soft endpoint of the need for retinal photocoagulation (UK

Prospective Diabetes Study (UKPDS) Group 1998). In the post-trial follow-up, the difference in HbA1c had vanished, but a greater benefit of intensive control was apparent (Holman et al. 2008). Unlike UKPDS, a post-trial long-term follow-up in VADT and DPP did not show any long-term benefits (Reaven et al. 2019, Lee et al. 2021), demonstrating the lack of consistency across trials. In the overweight patients' group of UKPDS, the advantage obtained by metformin over sulphonylurea and insulin was not explicable by better glycemic control (Turner 1998). Across trials, ACCORD and ADVANCE achieved the largest HbA1c difference, but they do not exhibit the greatest success in arresting complications.

Other studies (Stratton 2000) show an association of glycaemic status with the risk of complications, but these studies cannot be used to make a causal inference. It is possible that for individuals with a more serious underlying pathology, the chances of complications are higher and glycaemic control is more difficult. This results in a correlation but does not imply that reducing glycaemia would result in risk reduction.

Many of the drugs used in the treatment of T2DM have multiple effects in the body independent of glucose regulation. Insulin is known to have multiple functions in the body, ranging from amino acid metabolism and cell division to cognitive function and ovulation regulation (Strachan 2003, Shemesh et al 2012). Metformin is shown to directly affect endothelial function (Diamanti-Kandarakis et al 2005, Heidari et al 2019). The beneficial effects of SGLT2 inhibitors are seen in both diabetics and non-diabetics alike (Ferrannini 2015, DeFronzo 2019). Therefore, even if any beneficial effects of any of the drugs in diabetes treatment are seen, it cannot be stated that the benefits are because of improved glucose regulation. Even in T1DM, whether the observed efficacy of insulin treatment is because of glucose regulation or because of the multiple other functions of insulin has not been clearly addressed and resolved.

5. The magnitude of difference: Although the number of outcomes with individually significant beneficial effects of treatment is marginally greater than the number of outcomes with increased frequency, the frequency of harmful effects is substantially greater. While the individually significant reductions in ORs range between 0.5 and 0.92, the increase ranges between 1.2 and 13.1. The harmful effects of intensive treatment not only include major hypoglycaemic events and other drug-related symptoms but also cardiovascular mortality, cancer, and all-cause mortality in different trials. By the properties of ratios, the distribution of OR is bound to be positively skewed. The geometric mean of all ORs is close to one (1.01), indicating that the net odds ratios do not deviate from unity. Therefore, it cannot be stated that the benefits, if any, of intensive glucose regulation outweigh its harmful effects.

Even if, for the time being, we ignore the harmful effects and focus solely on the beneficial ones, and assume that the claimed significance was real, the question of whether the benefits are clinically meaningful is crucial. Although in public health literature, the emphasis is on ratio-based indices such as OR, RR and HR, common people seem to prefer absolute risk reduction (ARR). In an experiment, respondents, including those who had undergone training in public health statistics, were observed to judge their own risk by probability difference rather than by probability ratio (Vidwans et al 2021). In all glucose regulation trials for T2DM, the ARR is too small to be clinically meaningful. For example, in the ADVANCE trial over a 5-year follow-up, the incidence of combined major microvascular and macrovascular events was 20% in the control group and 18.1% in the treatment group. Although by ratio-based indices this is about a 10% reduction, by difference-based indices it is only 1.9%. The number needed to treat (NNT) for preventing diabetic complications in any of the trials is 20 and above, even after ignoring the increased frequency of some adverse events. Since different trials have different follow-up periods, and the effect appears to be more or less linear consistently over time, we can express it as NTNT, i.e. number-time (person-years) of treatment needed for preventing one complication (Laupacis 1988). NTNT is fairly consistently distributed across trials and averages to about 250. That is, if 25 diabetics are treated targeting glucose regulation for 10 years, one complication in one of the 25 patients may be prevented. This is quite a low success rate. Whether at this success rate it is worth undergoing treatment for glucose control, which has a considerable frequency of adverse events, is a subjective decision that should be left to the patient after having been informed about NTNT in simple language. Failure to inform the patient about the efficacy of treatment seen in clinical trials potentially amounts to a violation of human rights.

Apart from the trials included here, there are many more that converge on the same inference. These trials did not fit into our inclusion-exclusion criteria, but their inferences converge with our analysis. The NICE-sugar study revealed that under critically ill patients, tight sugar control resulted in higher mortality (The NICE sugar study investigators 2009). Huang et al (2001) meta-analysed 5 glucose regulation trials for their effect on cardiovascular disease. Apart from UGDP and UKPDS, they included VACSDM, Kumamoto and DIGAMI trials, and the meta-analysis yielded no effect of glucose lowering on cardiovascular disease.

In short, owing to multiple flaws in the design, statistical analysis and inferential logic in the clinical trials of glucose regulation in T2DM, at present we cannot state with scientific rigour that glucose regulation has any clinical benefit in T2DM, in contrast with T1DM.

This calls into question the classical assumption that T1DM and T2DM differ only in what goes wrong with glucose regulation and that all downstream effects of hyperglycaemia are similar. It is possible that the two are fundamentally different and converge only on one of the symptoms, that is, hyperglycaemia. Hyperglycaemia may not be central to the pathophysiology of the two in a similar way. Therefore, the target of treatment for the two needs to be different.

Evidence for confirmation and conformity biases:

Although the difference in the response of T1DM and T2DM to treatment targeting glucose regulation has been evident for quite a few decades, there is a deep-rooted reluctance to accept the evidence and refute glucose lowering as the main line of treatment in T2DM. We argue that this is because of confirmation bias, which is evident in multiple ways in the analysis and presentation of data in the publications resulting from the clinical trials. Following are some of the indicators of this bias.

a. Reporting positive and negative outcomes with different statistical treatment

At times, for reporting two different results in the same paper, two different reporting formats are used without justifying the difference. Quoting verbatim from the reported UKPDS results (UK Prospective Diabetes Study (UKPDS) Group 1998),

“Compared with the conventional group, the risk in the intensive group was 12% lower (95% CI 1–21, p=0·029) for any diabetes-related endpoint; 10% lower (–11 to 27, p=0·34) for any diabetes-related death; and 6% lower (–10 to 20, p=0·44) for all-cause mortality.”

In the following paragraph, the risk of major hypoglycaemic events is reported as,

“The rates of major hypoglycaemic episodes per year were 0·7% with conventional treatment, 1·0% with chlorpropamide, 1·4% with glibenclamide, and 1·8% with insulin.”

It is strange that the risk reduction is given as a percent reduction, but the incidence of hypoglycaemic events is reported as absolute frequencies. If the risk of hypoglycaemic events is also expressed as a percentage, with chlorpropamide it is 43% higher, with glibenclamide it is 100% higher and with insulin it is 170% higher. Effectively, the increase in risk of a major hypoglycaemic event is up to an order of magnitude greater than the maximum reduction seen in any of the complications in the trial. However, since the reporting format is different, a common reader is more likely to attach greater importance to percentage figures and attach smaller importance to hypoglycaemic events since the numbers reported are fractional. The tradition of reporting favourable and unfavourable results using different statistical expressions is continued in many other clinical trials of T2DM.

Another way of differentially reporting favourable and unfavourable results is by including them in the figures or tables versus making a passing mention in the text. In UKPDS as well as in DPP, many outcomes are mentioned in the text but not included in the tables/figures. The text sometimes has incomplete information to calculate OR, confidence intervals, or statistical significance. Interestingly, not a single outcome reported in the text shows a significant favourable effect of treatment. There is no justification offered as to why these were not included in the tables. In DPP, for example, medication-related gastrointestinal symptoms are mentioned in the text but not included in the tables. Participants who were unable to continue metformin due to adverse reactions are mentioned in the text, but these adverse events do not seem to be included in the analysis (White et al., 2022).

b. Cherry-picking

The UKPDS results (UK Prospective Diabetes Study (UKPDS) Group, 1998) state that the reduction in all diabetes-related endpoints was 12% and statistically significant. However, this does not include major hypoglycaemic events. If major hypoglycaemic events are pooled with other adverse events, no difference between the control and treated group remains with respect to all outcomes. The primary reports of UKPDS enlist a large number of single point and aggregate outcome measures, but the post-trial follow-up reports only seven aggregate outcomes. There is no way to know whether only the ones showing favourable results were selectively reported.

c. Changing targets

The UKPDS study first decided to consider at least a 40% reduction in mortality or morbidity in the intensive treatment group in order to be clinically significant. This benchmark was reduced to 15% subsequently. However, UKPDS has been criticised for not maintaining even the lowered standard and later declaring a 12% reduction as significant (Ewart, 2001).

d. Recognising but failing to correct flaws

The meta-analysis of four trials (Turnbull, 2009) recognises the flaw that the necessary adjustment of significance level for multiple statistical comparisons has not been made, but does not attempt to apply any correction for it. The published reports of individual trials do not even acknowledge this serious flaw, which is quite well known in the field of statistics.

The failure to correct for multiple statistical comparisons and drawing conclusions beyond what the data actually show is a cause for paper retraction by today's standards (Polizzi di Sorrentino et al., 2021). By this

norm, most papers concluding any benefit of glucose normalisation treatment in T2DM need to be retracted or at least corrected. Unfortunately, there are no scientific norms applied consistently across the board, and they vary according to predominant beliefs in the field. Despite several flaws in the trial itself and the statistical inferences drawn from it, the UKPDS webpage of the Radcliffe Department of Medicine, University of Oxford, continues to maintain that "The UK Prospective Diabetes Study (UKPDS) was a landmark randomised, multicentre trial of glycaemic therapies in 5,102 patients with newly diagnosed type 2 diabetes. It ran for twenty years (1977 to 1997) in 23 UK clinical sites and showed conclusively that the complications of type 2 diabetes, previously often regarded as inevitable, could be reduced by improving blood glucose....." (https://www.rdm.ox.ac.uk/about/our-clinical-facilities-and-mrc-units/DTU/completed-trials/ukpds visited on 6th Nov 2023). Such unsupported statements on responsible webpages clearly indicate that prevalent beliefs have overpowered evidence.

e. Manipulating readers' perspective

Most readers stop after reading the title and abstract. Therefore, selecting what to write in the title and abstract is an effective way to manipulate readers' perspective. For example, the meta-analysis by Turnbull et al. (2009) acknowledges that since they have not corrected for multiple comparisons, their analysis should be treated as tentative and exploratory. However, this admission does not appear in the abstract, and therefore most readers, after reading only the abstract, are likely to take their results as robust and conclusive.

The most parsimonious explanation for all the reporting anomalies together is that they arise from inadvertent or intentional confirmation bias and the burden of history. Since the substantial benefit of insulin treatment in T1DM was clearly shown, it was expected that it should work for T2DM as well. As the assumed pathophysiology was similar, the expectation was strengthened. Against this background, it would be natural to be reluctant to accept counterintuitive results. Confirmation bias arises from human nature and may not be treated as a crime, but there needs to be a continued attempt to overcome it and face the reality above prior beliefs and prejudices. Going by the results of clinical trials, considering all alternative possible explanations of the results, it needs to be clearly recognised that there is no conclusive evidence so far that glucose normalisation reduces diabetic complications in the context of T2DM. The remark by Gill (1991) quoted at the top of the article is applicable even after over 30 years of research. The failure of multiple attempts to support the hypothesis that glucose regulation can arrest complications in type 2 diabetes should be taken as decisive and the target of treating T2DM changed accordingly. Although this is being increasingly recognised (Boussageon et al 2017), the recognition is still fragmentary. The

Lancet commission on diabetes (Chan et al 2020) clearly recommends “access to insulin, patient education, and tools for monitoring blood glucose” only in the context of T1DM now. For T2DM, instead of emphasising glucose regulation, it recommends addressing “diverse environmental, behavioural, and socioeconomic causes” and “sustained reduction of common cardiometabolic risk factors”. However, the changing stance in literature has not yet reflected sufficiently in clinical practice. This requires active efforts to simultaneously educate practitioners as well as the common man. Treatment of T2DM is certainly going to be a case of “medical reversal”, i.e., a long-standing treatment recommendation based on a belief is completely withdrawn based on evidence (Herrera-Perez 2019). Facilitating the process of medical reversal along with research on effective alternatives will be critical for arresting one of the greatest health concerns of the present and future.

Statements and Declarations

Acknowledgements: This study benefited from an ongoing systematic review of all clinical trials in T2DM conducted by Shubhankar Kulkarni, Pramod Patil, Vibha Bapat, Hrishikesh Chunekar, and Varada Mengale. We thank Vikrant Patil for writing the code for the simulation programme. This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Contributorship statement: HV and MW raised the questions and developed the approach, AO and MW conducted the statistical analysis, MW wrote the manuscript.

None of the authors have any competing interests.

References

- Blackburn, H., & Jacobs, D. R. (2017). The University Group Diabetes Program 1961-1978: Pioneering randomized controlled trial. International Journal of Epidemiology, *46*(5), 1354–1364. https://doi.org/10.1093/ije/dyw168
- Boussageon, R., Pouchain, D., & Renard, V. (2017). Prevention of complications in type 2 diabetes: Is drug glucose control evidence-based? *British Journal of General Practice, 67*(655), 85–87.
- Chan, J. C. N., Lim, L. L., Wareham, N. J., et al. (2020). The Lancet Commission on diabetes: Using data to transform diabetes care and patient lives. *The Lancet, 396*(10267), 2019–2022.
- Chow, S. C., Chow, S. S., & Pong, A. (2021). Review of current controversial issues in clinical trials. *General Psychiatry, 34*(5), e100540.

- Colman, P. G., Thomas, D. W., Zimmet, P. Z., et al. (1999). New classification and criteria for the diagnosis of diabetes mellitus. The Australasian Working Party on Diagnostic Criteria for Diabetes Mellitus. *New Zealand Medical Journal, 112*(1086), 139–141.
- Dana, G. W. (1954). The type of diabetes mellitus associated with diabetic retinitis. *Diabetes, 3*(3), 210–212.
- DeFronzo, R. A., Inzucchi, S., Abdul-Ghani, M., et al. (2019). Pioglitazone: The forgotten, cost-effective cardioprotective drug for type 2 diabetes. *Diabetes and Vascular Disease Research, 16*(2).
- de Wit, H. M., te Groen, M., Rovers, M. M., et al. (2016). The placebo response of injectable GLP-1 receptor agonists vs. oral DPP-4 inhibitors and SGLT-2 inhibitors: A systematic review and meta-analysis. *British Journal of Clinical Pharmacology, 82*(1), 301–314.
- Diamanti-Kandarakis, E., Alexandraki, K., Protogerou, A., et al. (2005). Metformin administration improves endothelial function in women with polycystic ovary syndrome. *European Journal of Endocrinology, 152*(5), 749–756.
- Ewart, R. M. (2001). The case against aggressive treatment of type 2 diabetes: Critique of the UK Prospective Diabetes Study. *British Medical Journal, 323*(7317), 854–858.
- Ferrannini, E., & DeFronzo, R. A. (2015). Impact of glucose-lowering drugs on cardiovascular disease in type 2 diabetes. *European Heart Journal, 36*(34).
- Fullerton, B., Jeitler, K., Seitz, M., et al. (2014). Intensive glucose control versus conventional glucose control for type 1 diabetes mellitus. *Cochrane Database of Systematic Reviews, 2016*(6).
- Gill, G. (1991). In "Textbook of diabetes," Vol I and II, Eds Pickup JC and Williams G. Blackwell Scientific. ISBN 10: 0632025948 ISBN 13: 9780632025947.
- Goldner, M. G., Knatterud, G. L., & Prout, T. E. (1971). Effects of hypoglycemic agents on vascular complications in patients with adult-onset diabetes. 3. Clinical implications of UGDP results. *JAMA, 218*(9), 1400–1410.
- Harris, M. I. (1988). Classification and diagnostic criteria for diabetes mellitus and other categories of glucose intolerance. *Primary Care, 15*(2), 205–225.
- Hayes, A. J., Leal, J., Gray, A. M., et al. (2013). UKPDS Outcomes Model 2: A new version of a model to simulate lifetime health outcomes of patients with type 2 diabetes mellitus using data from the 30-year United Kingdom Prospective Diabetes Study: UKPDS 82. *Diabetologia, 56*(9), 1925–1933. https://doi.org/10.1007/s00125-013-2940-y
- Heidari, B., Lerman, A., Lalia, A. Z., et al. (2019). Effect of metformin on microvascular endothelial function in polycystic ovary syndrome. *Mayo Clinic Proceedings, 94*(12), 2455–2466.

- Heller, S. R., & ADVANCE Collaborative Group. (2009). A summary of the ADVANCE Trial. *Diabetes Care, 32*(Suppl 2), S357-S361. https://doi.org/10.2337/dc09-S339
- Heneghan, C., Goldacre, B., & Mahtani, K. R. (2017). Why clinical trial outcomes fail to translate into benefits for patients. *Trials, 18*(1), 122.
- Herrera-Perez, D., Haslam, A., Crain, T., et al. (2019). A comprehensive review of randomized clinical trials in three medical journals reveals 396 medical reversals. *eLife, 8*. https://doi.org/10.7554/eLife.45183
- Himsworth, H. P. (2013). Diabetes mellitus: Its differentiation into insulin-sensitive and insulin-insensitive types. 1936. *International Journal of Epidemiology, 42*(6), 1594–1598.
- Holman, R. R., Paul, S. K., Bethel, M. A., et al. (2008). 10-Year Follow-up of Intensive Glucose Control in Type 2 Diabetes. *New England Journal of Medicine, 359*(15), 1577–1589.
- Home, P. D., Pocock, S. J., Beck-Nielsen, H., et al. (2009). Rosiglitazone evaluated for cardiovascular outcomes in oral agent combination therapy for type 2 diabetes (RECORD): A multicentre, randomised, open-label trial. *The Lancet, 373*(9681), 2125–2135.
- Huang, E. S., Meigs, J. B., & Singer, D. E. (2001). The effect of interventions to prevent cardiovascular disease in patients with type 2 diabetes mellitus. *American Journal of Medicine, 111*(8), 633–642.
- Ioannidis, J. P. A. (2016). Why most clinical research is not useful. *PLOS Medicine, 13*(6), e1002049.
- Ismail-Beigi, F., Craven, T., Banerji, M. A., et al. (2010). Effect of intensive treatment of hyperglycaemia on microvascular outcomes in type 2 diabetes: An analysis of the ACCORD randomised trial. *The Lancet*, 376(9739), 419–430. http://dx.doi.org/10.1016/S0140-6736(10)60576-4
- James, S., Rao, S. V., & Granger, C. B. (2015). Registry-based randomized clinical trials—a new clinical trial paradigm. *Nature Reviews Cardiology, 12*(5), 312–316.
- Knatterud, G. L., Levin, M. D. E., & Klimt, C. R., et al. (1978). Hypoglycemic agents on vascular complications in patients with adult-onset diabetes. VII. Mortality and selected nonfatal events with insulin treatment. *JAMA, 240*(1), 37-42.
- Kulkarni, S., Sharda, S., & Watve, M. (2017). Bi-stability in type 2 diabetes mellitus multi-organ signaling network. *PLoS One, 12*(8), e0181536. https://doi.org/10.1371/journal.pone.0181536
- Laupacis, A., Sackett, D. L., & Roberts, R. S. (1988). An assessment of clinically useful measures of the consequences of treatment. *New England Journal of Medicine, 318*(26), 1728–1733.
- Lee, C. G., Heckman-Stoddard, B., Dabelea, D., et al. (2021). Effect of Metformin and Lifestyle Interventions on Mortality in the Diabetes Prevention Program and Diabetes Prevention Program Outcomes Study. *Diabetes Care, 44*(12), 2775–2782.

- Leslie, R. D. (1999). United Kingdom prospective diabetes study (UKPDS): What now or so what?. *Diabetes/metabolism research and reviews, 15*(1), 65–71.
- Meinert, C. L., Knatterud, G. L., Prout, T. E., & Klimt, C. R. (1970). A study of the effects of hypoglycemic agents on vascular complications in patients with adult-onset diabetes. II. Mortality results. *Diabetes, 19*, 789–830.
- Miller, M., Knatterud, G. L., & Hawkins, B. S. (1976). A study of the effects of hypoglycemic agents on vascular complications in patients with adult-onset diabetes. VI. Supplementary report on nonfatal events in patients treated with tolbutamide. *Diabetes, 25*(12).
- Nathan, D. M., Buse, J. B., Davidson, M. B., et al. (2009). Medical management of hyperglycemia in type 2 diabetes: A consensus algorithm for the initiation and adjustment of therapy. *Diabetes Care, 32*(1), 193–203.
- Narum, S. R. (2006). Beyond Bonferroni: Less conservative analyses for conservation genetics. *Conservation Genetics, 7*(5), 783–787.
- Nathan, D. M. (2014). The Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Study at 30 Years: Overview. *Diabetes Care, 37*(1), 9–16.
- Null, J., Tillman, R. L., Shelton, A. L., et al. (1973). Pancreatic changes in maturity onset diabetes mellitus. *Journal of the National Medical Association, 65*(1), 65–67.
- Polizzi di Sorrentino, E., Herrmann, B., & Villeval, M. C. (2021). Retraction Note: Dishonesty is more affected by BMI status than by short-term changes in glucose. *Scientific Reports, 11*, 1657. https://doi.org/10.1038/s41598-020-80528-2
- Ramsberg, J., & Platt, R. (2018). Opportunities and barriers for pragmatic embedded trials: Triumphs and tribulations. *Learning Health Systems, 2*(1), e10044.
- Reaven, P. D., Emanuele, N. V., Wiitala, W. L., et al. (2019). Intensive Glucose Control in Patients with Type 2 Diabetes — 15-Year Follow-up. *New England Journal of Medicine, 380*(23), 2215–2224.
- Rice, T. K., Schork, N. J., & Rao, D. C. (2008). Methods for handling multiple testing. *Advances in Genetics, 60*, 293-308. doi: 10.1016/S0065-2660(07)00412-9.
- Shemesh, E., Rudich, A., Harman-Boehm, I., et al. (2012). Effect of Intranasal Insulin on Cognitive Function: A Systematic Review. *Journal of Clinical Endocrinology and Metabolism, 97*(2), 366–376. https://doi.org/10.1210/jc.2011-1802
- Sievenpiper, J. L., Ezatagha, A., Dascalu, A., et al. (2007). When a placebo is not a “placebo”: A placebo effect on postprandial glycaemia. *British Journal of Clinical Pharmacology, 64*(4), 546–549.

- Sluik, D., Buijsse, B., Muckelbauer, R., et al. (2012). Physical Activity and Mortality in Individuals With Diabetes Mellitus. *Archives of Internal Medicine, 172*(17), 1285.
- Strachan, M. W. J. (2003). Insulin and cognitive function. *The Lancet, 362*(9392), 1253. https://doi.org/10.1016/S0140-6736(03)14615-6
- Stratton, I. M. (2000). Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): Prospective observational study. *BMJ, 321*(7258), 405–412. https://doi.org/10.1136/bmj.321.7258.405
- Thabane, L., Kaczorowski, J., Dolovich, L., et al. (2015). Reducing the confusion and controversies around pragmatic trials: Using the Cardiovascular Health Awareness Program (CHAP) trial as an illustrative example. *Trials, 16*(1), 387.
- The ACCORD study group. (2008). Effects of intensive glucose lowering in type 2 diabetes. *New England Journal of Medicine, 358*, 2545-2559. DOI: 10.1056/NEJMoa0802743.
- The DCCT Research Group. (1987). Diabetes Control and Complications Trial (DCCT): Results of Feasibility Study. *Diabetes Care, 10*(1), 1–19.
- The Diabetes Control and Complications Trial Research Group. (1993). The Effect of Intensive Treatment of Diabetes on the Development and Progression of Long-Term Complications in Insulin-Dependent Diabetes Mellitus. *New England Journal of Medicine, 329*(14), 977–986. http://www.nejm.org/doi/abs/10.1056/NEJM199309303291401
- The NICE_SUGAR study investigators. (2009). Intensive versus Conventional Glucose Control in Critically Ill Patients. *New England Journal of Medicine, 360*(13), 1283–1297.
- Tran, K., & Reaven, P. (2020). Review of the veteran affairs diabetes trial: Lessons learned. *Reviews in Endocrine & Metabolic Disorders, 21*(4), 537–546. https://doi.org/10.1007/s11154-020-09558-5
- Tse, T., Williams, R. J., & Zarin, D. A. (2009). Reporting "Basic Results" in ClinicalTrials.gov. *Chest, 136*(1), 295–303.
- Turnbull, F. M., Abraira, C., Anderson, R. J., et al. (2009). Intensive glucose control and macrovascular outcomes in type 2 diabetes. *Diabetologia, 52*(11), 2288–2298.
- Turner, R. (1998). Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). *The Lancet, 352*(9131), 854–865.
- UK Prospective Diabetes Study (UKPDS) Group. (1998). Intensive blood-glucose control with SU or insulin compared with conventional treatment (UKPDS 33) - Lancet 1998. *The Lancet, 352*(Ukpds 33), 837–853.

- Vidwans, H., Kharate, R., & Watve, M. (2021). Probability ratio or difference: How do people perceive risk? *Resonance, 26*, 1559–1565.
- Watve, M. (2013). Doves, Diplomats, and Diabetes. New York, NY: Springer New York.
- White, N. H., Pan, Q., Knowler, W. C., et al. (2022). The Effect of Interventions to Prevent Type 2 Diabetes on the Development of Diabetic Retinopathy: The DPP/DPPOS Experience. *Diabetes Care, 45*(7), 1640–1646.
- Zarin, D. A., & Keselman, A. (2007). Registering a Clinical Trial in ClinicalTrials.gov. *Chest, 131*(3), 909–912.

Supplementary data: available at https://doi.org/10.32388/IH7KEP

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.