Manuscript Diagnostic Audit

Specimen Report

Word-Medex Pty Ltd · Directed by F. G. Issa, MD, PhD · Sydney, Australia

About this specimen

This is a genuine, unedited example of the report a client receives from a Manuscript Diagnostic Audit. It is reproduced here so that prospective clients can see exactly what they would be buying.

Client manuscripts are confidential and are never reproduced, quoted, or used as examples. This specimen was therefore prepared from a published, open-access article, assessed as though it had been submitted to Word-Medex before publication:

> Wang X, Wu N, Sun C, Jin D, Lu H. Effects of SGLT-2 inhibitors on adipose tissue distribution in patients with type 2 diabetes mellitus: a systematic review and meta-analysis of randomized controlled trials. *Diabetology & Metabolic Syndrome* 2023;15:113. doi:10.1186/s13098-023-01085-y PMID: 37254186

The authors are not clients of Word-Medex and have no connection to this firm. The article is published under a Creative Commons Attribution licence and every observation below can be checked against the public text. Because the paper is already published, this is a demonstration of method rather than a criticism of a live submission — the findings are presented as they would have been had the authors sought an opinion beforehand.

Assessed against a target of	Diabetes, Obesity and Metabolism (Q1)
Manuscript length	approx. 5,000 words excluding references
Applicable fee	Band 1 — A$250

VERDICT

Substantive revision required before submission

The review addresses a worthwhile question and the search and extraction machinery is largely sound. However, the manuscript's central conclusion is not supported by its own analysis: the subgroup findings that the Abstract and Conclusions present as "decisive factors" did not, on the paper's own numbers, explain the heterogeneity they were performed to explain. This is a reasoning failure rather than a data failure, and it is correctable — but it must be corrected before submission, because it is the first thing a methodological reviewer will test.

Domain 1 — Hypothesis clarity and positioning

Assessment: adequate, but the novelty claim is exposed.

The question is stated plainly and the clinical rationale is well constructed: adipose distribution is a cardiometabolic risk determinant, SGLT-2 inhibitors reduce weight, and whether that translates into compartment-specific fat loss is genuinely unresolved.

The difficulty is positional rather than conceptual. The Background acknowledges that Liu et al. published a meta-analysis of SGLT-2 inhibitors and adipose tissue in T2DM **four months before this submission**, covering visceral and subcutaneous fat in the same population. The stated advance over it is threefold: the addition of ectopic depots, more detailed subgroup analysis, and newer trials.

That is a defensible claim, but it is currently made in three sentences and never revisited. An editor at a competitive journal will ask, reasonably, how many trials this review contains that Liu et al. did not, and what the ectopic-fat analyses actually add. As written, the reader cannot answer either question. This is the difference between an incremental update and a genuine extension, and the manuscript does not currently make the case that it is the latter.

What would strengthen it: state the number of additional trials and name the ectopic outcomes as the primary novelty in both the Background and the Abstract. If the epicardial and liver-fat analyses are the advance, they should not be relegated to third place behind VAT and SAT throughout the paper.

Domain 2 — Scientific argument architecture

Assessment: this is where the manuscript is weakest, and it is a structural problem rather than a presentational one.

The paper's logic runs: pooled effects are large but heterogeneity is extreme → subgroup analyses are performed to find the source → subgroups identify duration, BMI, age and NAFLD status as determinants → these are reported as decisive.

The chain breaks at the third step, and the manuscript itself contains the evidence that it breaks. Within the subgroups that are described as explanatory, heterogeneity is **unchanged or higher** than in the parent analysis: I^2 = 96% for the 16–40 week VAT subgroup against 94% overall; I^2 = 97% for the NAFLD subgroup; I^2 = 95–96% across the corresponding SAT subgroups. A subgroup that leaves I^2 at 96% has not explained anything.

The Discussion appears to recognise this — it states that subgroup and sensitivity analyses failed to identify the source of heterogeneity — but the Abstract and Conclusions simultaneously present the same subgroups as having identified "the decisive factors." **Both statements cannot be true, and they appear in the same manuscript.** A reviewer will find this immediately, and it will read as either carelessness or as conclusions written to a predetermined shape.

A second architectural issue compounds it. All four subgroup variables are **study-level aggregates** — mean age, mean BMI, whether a trial enrolled NAFLD patients. Inferences drawn from them about which *patients* benefit are ecological, not individual. The Conclusions nevertheless state that the drugs work "especially in young T2DM patients with NAFLD and high BMI," which is a patient-level claim resting on study-level data. No test for subgroup interaction is reported, and with four subgrouping variables applied across five or more outcomes, the number of informal comparisons is well above twenty with no acknowledgement of multiplicity.

What would strengthen it: decide what the subgroup analyses are for. If they are exploratory hypothesis-generation — which is the defensible position — say so explicitly, move them out of the Abstract's conclusion sentence, and reframe the Conclusions in terms of associations requiring confirmation. If they are intended as explanatory, they need formal interaction tests and an honest statement that residual heterogeneity remains near-total.

Domain 3 — Strength and presentation of evidence

Assessment: the analysis is competently executed but the reporting contains errors that will damage confidence.

Terminology. Throughout the manuscript, SMD is expanded as "standard mean deviation" and MD as "mean deviation." The correct terms are *standardised mean difference* and *mean difference*. These are not synonyms for what is written, and in a methods-heavy submission the error signals to a statistical reviewer that the effect measures may not be fully understood. It appears in the Abstract, where it will be read first.

Copy-paste error in Results. The section headed *Efficacy of SGLT-2 inhibitors on SAT reduction* opens: "A total of 16 RCTs involving 937 participants estimated VAT." This is the VAT sentence duplicated. The Abstract and the

study-characteristics paragraph both state that **nine** trials reported SAT. The SAT analysis therefore has no stated sample size anywhere in the paper.

Discrepant p-value for the epicardial analysis. The Abstract reports p = 0.71; the Results report p = 0.91 for the same effect size and interval (SMD 0.03, 95% CI –0.52 to 0.58). Given those numbers the Results value is the plausible one. One of the two is wrong.

Publication-bias test is named inconsistently three times. The Abstract says Begg's test; the Methods say Egger's test; the Results say Begg's test with p = 0.251. Beyond the inconsistency, this matters substantively: the VAT funnel plot is described as asymmetric, and asymmetry is then dismissed on the basis of Begg's rank-correlation test, which is substantially less powerful than Egger's regression. With sixteen trials, Egger's is the appropriate test, and visible asymmetry dismissed by an underpowered test will not satisfy a methodological reviewer.

Search strategy as printed is not reproducible. Six of the ten drug names in the search string are misspelled — including forms such as "tofoglifloxin," "empagglifloxin," "ipraglifloxin" and "lusoglifloxin," none of which are real compounds. If these were the terms actually run, relevant trials were missed. If they are transcription errors introduced in drafting, the published search cannot be reproduced. Either way, for a systematic review — where reproducibility of the search is the foundational claim — this is a serious defect.

Data-extraction formula is garbled. The Luo mean-estimation formula is printed with a misplaced division operator, rendering it mathematically incoherent as typeset.

Risk-of-bias reporting is internally inconsistent. The text states that all trials evaluated outcomes in a blinded manner and, two sentences later, that many were downgraded because of lack of blinding among participants and assessors. It also assigns studies to a "medium risk" category immediately after stating that the Cochrane tool uses low, high, and unclear.

On the effect sizes themselves. A pooled SMD of –1.42 for visceral fat is a very large effect — larger than is typically observed for bariatric surgery in comparable syntheses. Combined with I^2 = 94%, this warrants explicit scepticism in the text rather than acceptance at face value. The manuscript does not currently interrogate whether the magnitude is biologically plausible.

Domain 4 — Critical omissions in the discussion

Assessment: three omissions that reviewers at this tier will expect to see addressed.

No certainty-of-evidence assessment. GRADE is now an expectation for systematic reviews at most Q1 journals and is required by several. Its absence is conspicuous given that the certainty here is plainly low for several outcomes — small trials, extreme heterogeneity, roughly half the included studies open-label, and the fat outcomes rarely being the trials' primary endpoints.

No prediction interval. With I^2 above 90%, the confidence interval around the pooled estimate describes the precision of the average effect, not the range of effects a clinician should expect. A prediction interval would very likely cross the null for VAT, and reporting it is the honest response to this degree of heterogeneity. Its absence lets the paper present an average that may not describe any real population.

Unsupported mechanistic speculation. The Discussion asserts that "prolonged duration may lead to resistance to SGLT2 inhibitors" to explain why the >40-week subgroup showed a smaller effect than the 16–40-week subgroup. No citation is offered and no such resistance phenomenon is established. The comparison is between different trials, not within patients over time, and the simpler explanation — that the trials differ in population and measurement method — is not considered. This single sentence then becomes a headline claim in the Conclusions and the Abstract.

Two smaller points: the epicardial analysis pools three trials with I^2 = 69%, which is arguably not poolable and should be presented descriptively; and the Data Availability statement offers the extracted dataset "on reasonable request," where most journals in this tier now expect the extraction table as a supplementary file.

Domain 5 — Journal fit

Assessment: fit is reasonable; readiness is not.

Diabetes, Obesity and Metabolism publishes meta-analyses of this type and the topic is within scope. The obstacles are the Liu et al. overlap (Domain 1) and the statistical reporting standard, which is high at this journal and which the manuscript does not currently meet.

Two structural mismatches are worth noting. The title promises adipose tissue *distribution*, but a substantial part of the Results concerns weight, BMI and waist circumference, which are not distribution measures and are well established for this drug class — this material would be better condensed. And the PROSPERO registration (CRD42023401163) postdates the search closing date of 2 January 2023; registration timing is checked at some journals and retrospective registration should be disclosed rather than left to be discovered.

Top three rejection risks

1. The subgroup analyses do not support the conclusions drawn from them

Why this is a risk. A statistical reviewer will compare the subgroup I^2 values against the parent analysis within minutes, find them unchanged at 95–97%, and then find the Discussion conceding that the source of heterogeneity was not identified while the Abstract claims decisive factors were found. That contradiction, once seen, casts doubt on everything else.

What resolving it would require. Reframing rather than reanalysis. The subgroup results are reported honestly in the Results section; it is the Abstract, Conclusions and parts of the Discussion that overstate them. Recasting these as exploratory and hypothesis-generating, adding formal interaction tests where feasible, and stating plainly that heterogeneity remains unexplained would make the paper defensible without weakening what it actually found. *Estimated effort: moderate, largely rewriting.*

2. The search strategy as published is not reproducible

Why this is a risk. For a systematic review, a reproducible search is not a detail — it is the basis on which the review claims to be systematic. Six misspelled drug names in the search string will prompt an editor to ask whether the search was actually run as printed, and there is no good answer that does not require the search to be re-run or the string corrected and verified.

What resolving it would require. Correcting the string, confirming against the search logs that the executed terms were correct, and stating the exact strings and dates per database in a supplementary table. If the misspelled terms were the ones actually used, the search must be re-run — which may change the included studies. *Estimated effort: low if a transcription error; substantial if not. This should be checked first, as the answer determines how much other work is needed.*

3. Reporting inconsistencies that undermine confidence in the analysis

Why this is a risk. Individually, the duplicated SAT sentence, the p = 0.71 versus p = 0.91 discrepancy, the Begg's/Egger's confusion, the "standard mean deviation" terminology, the garbled formula, and the contradictory risk-of-bias text are all minor. Collectively they suggest the manuscript did not receive a final consistency check before submission, and they invite a reviewer to assume there are further errors in the parts they cannot verify. This is the mechanism by which methodologically adequate papers are rejected.

What resolving it would require. A systematic consistency pass: every number in the Abstract verified against the Results, every test named identically in all three locations, terminology corrected throughout, and the risk-of-bias paragraph rewritten to the Cochrane categories actually used. *Estimated effort: low, but it must be done systematically rather than opportunistically.*

Closing note

This manuscript is not far from publishable in a good journal, and the underlying work is real. Its difficulty is that the conclusions have been written to be stronger than the analysis supports, and the surface errors invite a reviewer to look for exactly that. Both are fixable by revision rather than by new work.

Attribution and licence

The article assessed above is © The Author(s) 2023 and is published open access under a **Creative Commons Attribution 4.0 International Licence** (CC BY 4.0), which permits use, sharing, adaptation and reproduction in any medium provided appropriate credit is given, a link to the licence is provided, and changes are indicated. Licence terms: https://creativecommons.org/licenses/by/4.0/

Changes made: the original article has not been reproduced, altered, or redistributed. Short extracts totalling fewer than fifty words are quoted for the purposes of criticism and review, and the remainder of this document is original commentary by Word-Medex Pty Ltd. No figures, tables, or formatted material from the publisher have been reproduced.

Word-Medex Pty Ltd is not affiliated with the authors, with *Diabetology & Metabolic Syndrome*, or with Springer Nature, and nothing in this document implies endorsement by any of them. This commentary is the professional opinion of its author, offered in good faith on the basis of the facts stated, all of which are verifiable in the publicly available article.

A Diagnostic Audit identifies and prioritises problems. It does not edit, rewrite, format, or prepare a manuscript for submission — those are the province of the Full Pre-Submission Strategic Audit, against which the diagnostic fee is credited in full within 60 days.

Word-Medex Pty Ltd · Sydney, Australia · www.word-medex.com.au