nature portfolio

Peer Review File

Effect of a behavioral counseling for adoption and maintenance of a physically active lifestyle on long-term mortality in people with type 2 diabetes: Post hoc of the Italian Diabetes and Exercise Study_2

Corresponding Author: Professor Giuseppe Pugliese

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
This study investigates whether a structured behavioural counselling intervention aimed at increasing MVPA and reducing sedentary time can reduce long-term all-cause mortality in individuals with type 2 diabetes. It is a follow-up of the IDES_2 randomized controlled trial, with a mean follow-up of 10.3 years. This is a clinically relevant analysis that would make a valuable contribution to the literature. However, in my opinion, several issues need to be clarified/addressed

* In order to aid interpretability and to investigate dose-response associations, have the authors considered reporting the HRs for change in sedentary time and PA as opposed to the average? Change could then be additionally adjusted for the baseline values in order to account for possible regression to the mean. This may also alleviate potential overstatement of the intervention efficacy— where the mortality reduction may reflect baseline fitness/PA levels rather than the intervention per se

* More widely, I appreciate the results have been reported elsewhere, but it would be extremely helpful to present the top line results for the RCT (i.e. pre and post changes for each of the behaviours).

* Line 260 - Please can the authors clarify what this means? How can you assume that accelerometer non wear time in the awake state was spent sedentary?

Also, does accelerometer wear time not need to be adjusted for in the analysis? How can you be sure that changes in these physical behaviours are not just driven by changes in how long they wore the device for? Also, on Line 257 - Consider changing from 'all day' to 'waking hours'

* There is currently no discussion of potential unmeasured confounders (e.g., diet, psychosocial, socioeconomic factors)

Reviewer #2

(Remarks to the Author)
Overall Comments
This secondary analysis of the IDES-2 trial has several important strengths including the long-term (10-year) follow-up and the choice of a "hard" and unambiguous endpoint (all-cause mortality). The use of objective accelerometer data in the original trial provides a significant advantage over self-report measures, and the authors should be commended for running a large number of sensitivity analyses.
However, the study has major limitations in its design and statistical analysis that must be addressed to support its conclusions. My concerns are detailed below.
Major Methodological and Statistical Concerns

1. Framing of the Study
Whilst the authors correctly acknowledge that "mortality was not a primary outcome of this study...", they do not state that mortality was also not a pre-specified secondary outcome. An analysis of an endpoint that was not pre-specified in the original protocol or statistical analysis plan is, by definition, post-hoc. This is fundamentally different from analyzing a pre-specified secondary endpoint, and it carries different implications for the strength of evidence.
The follow-up over a subsequent ~7-year period to the original 3-year RCT makes this a purely observational study in which the exposure is simply "prior assignment to a trial arm." During the 7 years of observational follow-up, differential behaviors and co-interventions (like medication use) can emerge.
I therefore recommend that the authors clearly and consistently acknowledge throughout the manuscript that this is a secondary, post-hoc analysis and that the long-term data collection period is observational in nature. The findings should be framed as hypothesis-generating, and not as novel evidence from an RCT.
2. The Issue of Confounding by Medication Use
This is my most important statistical concern. The manuscript states that models were sequentially adjusted for medication use "at any time throughout the study (model 2)" (page 14, line 309). This raises two interrelated problems: the analytical method and baseline imbalances.
A) Confounding by Indication and Time-Varying Covariates: If medication status was treated as a time-varying covariate, this introduces a high risk of confounding by indication. This bias arises when a patient's worsening health status causes them to initiate a new medication. Adjusting for such a post-randomization variable can lead to biased estimates (Woodman, Ther Adv Drug Saf, 2014).
B) Significant Baseline Imbalances in Cardioprotective Medications: A review of eTable 7 from the original 2019 JAMA publication's supplement reveals several potentially clinically significant, baseline imbalances that favor the intervention group:
GLP-1 Receptor Agonists (GLP-1 RA): The Intervention group had double the rate of GLP-1 RA use at baseline (13.3%) compared to the Control group (6.7%). Given the well-established mortality benefits of this drug class, this is a major potential confounder. A group that starts with twice the rate of a cardioprotective medication has a significant survival advantage that cannot be attributed to the behavioral intervention.
Other Imbalances: Other baseline differences, though smaller, consistently suggest the control group may have been at a higher baseline risk (e.g., higher use of any glucose-lowering agent, insulin, and anti-hypertensives). This may arise as a result of the relatively small trial size.
Recommendation: Given these challenges, could the authors please:
Specify precisely how medication use was incorporated into the adjusted models. The paper states “treatment with…at any time throughout the study”. Similar to use of a time-varying covariate this opens the possibility of confounding by indication. Also provide the justification for the approach used and discuss how the chosen model mitigates the risk of confounding by indication.
Present a sensitivity analysis that adjusts only for baseline characteristics (including the baseline medication imbalances) and does not adjust for any post-randomization medication changes. This would provide an estimate of the intervention effect that is less susceptible to this specific type of bias.

Minor Comments
Statistical Power: The original trial was not powered for a mortality outcome. It would strengthen the discussion to explicitly state the post-hoc power of the study to detect the observed mortality difference, or at a minimum, to discuss the implications of observing a statistically significant p-value in an underpowered analysis and the need for cautious interpretation.
Other Post Hoc Analyses: The analyses assessing the association of PA and fitness parameters with all-cause mortality (regardless of study arm) are valuable for generating hypotheses. The authors correctly label them "Post hoc analyses," and this exploratory nature should be consistently emphasized in the text to avoid any implication of confirmatory findings.
Generalizability of Findings: The study provides potentially important evidence for a specific cohort (sedentary individuals aged 40-80 with type 2 diabetes and a BMI of 27-40). The authors should consider expanding the discussion on the generalizability of these findings to broader populations, such as younger individuals, those with different BMI ranges, or individuals without diabetes.

Reviewer #3

(Remarks to the Author)
The authors present mortality follow up data from the Italian Diabetes and Exercise Study-2. This was a trial of a behavioural intervention compared to control in 300 sedentary patients with type 2 diabetes in which the primary outcome was change in activity. This has been reported previously. The protocol, which was published in 2012 lists the following secondary endpoints; physical fitness, modifiable cardiovascular risk factors, MS disturbances, well-being/depression, and health related QoL. There is no mention of mortality follow up as an endpoint.

The major observation of the paper is the apparent halving of mortality risk in the intervention group compared to control. This is a very sizeable effect.

The first consideration is whether this observation is true or could be a manifestation of chance, bias or confounding. The p value for the main comparison is 0.01 which is relatively modest, so the role of chance can't be excluded. The study was not blinded, at least for the clinicians and the patients but some of the outcome assessors were blinded. It is possible that

physicians may have altered care for patients in the intervention group. Finally although randomisation reduces the chance of measured and unmeasured confounding, it does not eliminate it. This paper does not include a comparison of characteristics of the patients at baseline. However, it is noteworthy that a previous paper showed that 32% of the control population were current smokers compared to 25% in the intervention group. The authors do present adjusted analyses, but do not appear to have adjusted for smoking. This is perhaps relevant since the big difference between groups is in cancer-related mortality (19 deaths in the control group and 5 in the intervention group).

The main result paper from this study in JAMA in 2019 contains a flow chart similar to that provided here as figure 1. However, there are some key differences. The number of individuals lost to follow up in the intervention group (n=17) and the control group (n=16) are identical between the two reports but the explanations for the reasons for lost to follow are very different. The JAMA report has no mention of any loss to follow up due to death, whereas Figure 1 in this report mentions 2 deaths in the intervention group and 6 in the control group. This difference needs some explanation. If it is correct, the authors need to explain why death is considered a "loss to follow up" in the context of a mortality outcome analysis.

The major surprise of this study is that they were able to demonstrate an apparently significant mortality impact in 300 individuals when much larger studies have been inconclusive. One explanation could be the risk profile of the individuals included. The absolute event rate in this study is 23.6 per 1000 pyfu in the control group and 11.2 in the intervention. The event rate in LOOK Ahead, by comparison is 19 per 1000 pyfu, so this is not so different.

Reviewer #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #5

(Remarks to the Author)
• Your finding of marked differences in long-term mortality between groups is substantial and fills a notable gap in the literature, where few RCTs have provided such conclusive evidence beyond observational studies.
• The tailored nature of your intervention, accounting for participants' psychosocial, socio-economic, and cultural circumstances, is remarkable and likely central to its sustained efficacy; this could be highlighted more assertively in the abstract and discussion.
• Consider clarifying, for reader benefit, the duration of both theoretical and practical content in the counselling sessions (see line 228).
• It would also be helpful to briefly address how concerns about glycaemic imbalances when increasing physical activity were managed and how this may have influenced participants' willingness and ability to engage with the intervention.
With minor clarifications as suggested above, this is a strong manuscript for publication and will be of considerable interest to the field.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
I would like to thank the reviewers for addressing the majority of my previous comments. Thanks also for the clarification regarding accelerometer wear time and its treatment in the analysis. While I appreciate the rationale for assuming that the unmeasured waking time (approximately 0.9 hours) is largely sedentary, I would suggest that not adjusting for wear time may still introduce bias, particularly if wear time varies meaningfully across participants. I would also argue that, if you wanted to avoid the influence of wear time, then it should be accounted for in the models. As a minimum, this limitation should be acknowledged in the discussion

Reviewer #2

(Remarks to the Author)

Reviewer #3

(Remarks to the Author)
The authors have added the important amendment that this is a post-hoc rather than a pre-specified analysis. They have added a more extensive discussion of the limitations of the study. They have addressed the inconsistency in the study flow diagram.

There are still a number of places in which the language is too strong and should be softened and another in which the meaning is unclear.

Abstract. Line 27. "An inverse relationship between physical activity (PA) and mortality has emerged from epidemiological surveys, but not randomized clinical trials". This doesn't make sense. A relationship doesn't emerge from a survey. This sentence should be rewritten and framed as a relationship being observed in a survey.

Abstract. Line 36. "A behavioral counselling targeting all domains of PA and sedentary behavior reduced long-term mortality risk in people with type 2 diabetes" Given the many uncertainties in this study the sentence should be reframed to state that the counselling intervention was associated with a reduction in long term risk rather than it reduced it.

Conclusions. Line 221. "This follow-up analysis of the IDES_2 showed that a three-year behavioral counselling promoting sustained changes in all domains of PA/sedentary behavior produced a ~50% reduction in all-cause mortality versus standard care." As with the sentence above, this should be rewritten to make it clear that the intervention was associated with a 50% reduction in risk rather than it produced it.

Reviewer #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #5

(Remarks to the Author)
Thank you for your thorough revisions. The manuscript is notably improved, and the presentation of your findings is considerably clearer and more balanced. By consistently highlighting throughout the text that this work represents a secondary, post hoc analysis with a long-term observational data collection period, the manuscript now has greater transparency and interpretive caution. The implications of the results are more appropriately framed, enhancing the nuance of your argument. Overall, the revised manuscript is more robust.

Senior Editor

Make sure to list in the manuscript all pre-specified primary and secondary outcomes of the trial, with a mention of/reference to what has been already published and where. If any secondary outcome is unpublished (i.e. MS disturbances), please make that clear to the reader and explain whether it will be published later.

As requested, we have listed all the pre-specified primary and secondary outcomes of the IDES_2 (line 286). We have also mentioned what was already published and where and explained why musculo-skeletal disturbances, though included among the secondary end-points, have not been analyzed and, hence, cannot be published (line 293).

Reviewer #1

This study investigates whether a structured behavioural counselling intervention aimed at increasing MVPA and reducing sedentary time can reduce long-term all-cause mortality in individuals with type 2 diabetes. It is a follow-up of the IDES_2 randomized controlled trial, with a mean follow-up of 10.3 years. This is a clinically relevant analysis that would make a valuable contribution to the literature. However, in my opinion, several issues need to be clarified/addressed.

We are indebted to the Reviewer for providing us with positive comments to our work and for raising important issues to be addressed.

In order to aid interpretability and to investigate dose-response associations, have the authors considered reporting the HRs for change in sedentary time and PA as opposed to the average? Change could then be additionally adjusted for the baseline values in order to account for possible regression to the mean. This may also alleviate potential overstatement of the intervention efficacy— where the mortality reduction may reflect baseline fitness/PA levels rather than the intervention per se

We thank the Reviewer for raising this issue, which is very important for the interpretation of the study results.

We did assess the dose response associations with mortality of baseline values, mean values throughout the study, and change from baseline (as mean values throughout the study minus baseline values).

We decided not to show the results for change from baseline for this analysis because the curve describing the relationship of the level of physical activity (PA) or physical fitness with mortality is not linear (Moore SC et al. *PLoS Med.* 2012;9:e1001335 and Ekelund U at al. *BMJ.* 2019;366:l4570). This implies that the survival benefit provided by an increase in PA or fitness is strictly dependent on the baseline values, i.e., the reduction in mortality for the same increase in PA or fitness is much higher or the increase in PA or fitness required to obtain the same reduction in mortality is much lower for individuals with lower than for those with higher baseline values. Though all participants were physically inactive and sedentary, the range of PA and fitness values at baseline was in fact quite large, with MPVA ranging from 0.6 to 21.0 $min \cdot day^{-1}$ and VO_{2max} ranging from 12.3 to 41.4 $ml \cdot min^{-1} \cdot kg^{-1}$, and the impact on mortality of increases in these values varied largely in an inverse manner and cannot be accounted for by adjusting for baseline values. However, though participants who had the highest baseline values of PA or fitness parameters did not necessarily show the highest increases in PA or fitness parameters (i.e., those falling in each quartile of baseline value distributed almost equally among the four quartiles of change from baseline), it is certainly true that mean values throughout the study reflect also the baseline values, as outlined by the Reviewer. Therefore, to assess the relative contribution to mortality reduction of mean values throughout the study and baseline values, we have run additional analyses by including both these values in the Cox proportional hazards regression models, adjusted for age and sex. The results, presented in Table 4, show that the effect of mean values throughout the study prevailed over that of baseline values for MVPA, PA volume, VO_{2max}, and lower body muscle strength, whereas the opposite occurred for LPA, with no significant association with mortality for both mean values throughout the study and baseline values observed for SED-time. This means that the associations with mortality of mean values throughout the study of MVPA, PA volume, VO_{2max}, and lower body muscle strength was independent of baseline values. As now stated in

the Discussion section (line 168), this observation argues in favor of a major role of the intervention in the reduced mortality observed in our study.

More widely, I appreciate the results have been reported elsewhere, but it would be extremely helpful to present the top line results for the RCT (i.e. pre and post changes for each of the behaviours).

Following to the Reviewer's suggestion, we have now reported the main results of the trial (line 79).

Line 260 - Please can the authors clarify what this means? How can you assume that accelerometer non wear time in the awake state was spent sedentary? Also, does accelerometer wear time not need to be adjusted for in the analysis? How can you be sure that changes in these physical behaviours are not just driven by changes in how long they wore the device for? Also, on Line 257 - Consider changing from 'all day' to 'waking hours'.

As reported in the Methods section (line 305) and the Study Protocol, we wanted to avoid the influence of the accelerometer wear time, since most of the waking time while not wearing the accelerometer is usually spent in sedentary activities, thus leading to an underestimation of total daily sedentary-time, which was a co-primary end-point of the IDES_2. To this end, participants were asked to wear the device all day (except if swimming) up to bedtime, i.e., in waking hours, as suggested by the Reviewer and adopted in the revised version. Of course, "all day" or "waking time" cannot include the sedentary activities performed immediately after waking (e.g., bathing or showering, dressing) or before going to sleep (e.g., undressing, reading in bed), when participants are awake, but cannot wear the accelerometer. Therefore, since the mean sleeping time was 8.3 hours and the mean accelerometer wear time was 14.8 hours, the remaining average 0.9 hours could be assumed to be spent in routine morning and evening sedentary activities and the analyses were not adjusted for the accelerometer wear time.

There is currently no discussion of potential unmeasured confounders (e.g., diet, psychosocial, socioeconomic factors).

Thanks for this comment. As suggested by the Reviewer, we have now expanded the discussion of potential unmeasured confounders, including diet, which was prescribed, but not considered in data analysis, though patients received dietary prescriptions and adherence to diet was verified at intermediate visits (line 212). However, regarding psychosocial, family, cultural, and socio-economic factors, these parameters were measured at baseline and no between-group difference was observed, as reported in the main IDES_2 publication (Balducci S et al. *JAMA.* 2019;321:880–890), thus reducing the risk of confounding bias.

Reviewer #2

This secondary analysis of the IDES-2 trial has several important strengths including the long-term (10-year) follow-up and the choice of a "hard" and unambiguous endpoint (all-cause mortality). The use of objective accelerometer data in the original trial provides a significant advantage over self-report measures, and the authors should be commended for running a large number of sensitivity analyses. However, the study has major limitations in its design and statistical analysis that must be addressed to support its conclusions. My concerns are detailed below.

We thank the Reviewer for providing us with useful comments and suggestions to improve our paper.

1. Framing of the Study
Whilst the authors correctly acknowledge that "mortality was not a primary outcome of this study...", they do not state that mortality was also not a pre-specified secondary outcome. An analysis of an endpoint that was not pre-specified in the original protocol or statistical analysis plan is, by definition, post-hoc. This is fundamentally different from analyzing a pre-specified secondary endpoint, and it carries different implications for the strength of evidence.
The follow-up over a subsequent ~7-year period to the original 3-year RCT makes this a purely observational study in which the exposure is simply "prior assignment to a trial arm." During the 7 years of observational follow-up, differential behaviors and co-interventions (like medication use) can emerge.

I therefore recommend that the authors clearly and consistently acknowledge throughout the manuscript that this is a secondary, post-hoc analysis and that the long-term data collection period is observational in nature. The findings should be framed as hypothesis-generating, and not as novel evidence from an RCT.

We agree with the Reviewer that this is a post hoc, not a pre-specified analysis. However, we have never stated that it was pre-specified and the assessment of mortality was in fact not included in the pre-specified analyses listed in the study protocol.
The fact that this study includes a 3-year RCT and a ~7-year follow-up makes it similar to the other lifestyle intervention studies cited in our manuscript (Qing Diabetes Prevention Outcome Study, Finnish Diabetes Prevention Study, Diabetes Prevention Program. and Look AHEAD Study) and also to the studies comparing intensive versus conventional glucose control in people with diabetes (e.g., DCCT, UKPDS, ACCORD; ADVANCE, and VADT), which have all assessed long-term mortality after a post-trial follow-up. This is because it is virtually unfeasible to conduct a RCT testing lifestyle (but also other) interventions for 10-15 years in order to establish a cause-effect relationship between the intervention and mortality. Therefore, there is no doubt that, just like in the studies mentioned above, results may have been influenced by several events (behaviors, medications, etc.) that occurred during the post-trial period.
Though these limitations had already been acknowledged in the original version, we have stressed these points in the revised versions, according to the Reviewer's suggestion (line 200).

2. The Issue of Confounding by Medication Use
This is my most important statistical concern. The manuscript states that models were sequentially adjusted for medication use "at any time throughout the study (model 2)" (page 14, line 309). This raises two interrelated problems: the analytical method and baseline imbalances.
A) Confounding by Indication and Time-Varying Covariates: If medication status was treated as a time-varying covariate, this introduces a high risk of confounding by indication. This bias arises when a patient's worsening health status causes them to initiate a new medication. Adjusting for such a post-randomization variable can lead to biased estimates (Woodman, Ther Adv Drug Saf, 2014).
B) Significant Baseline Imbalances in Cardioprotective Medications: A review of eTable 7 from the original 2019 JAMA publication's supplement reveals several potentially clinically significant, baseline imbalances that favor the intervention group:
GLP-1 Receptor Agonists (GLP-1 RA): The Intervention group had double the rate of GLP-1 RA use at baseline (13.3%) compared to the Control group (6.7%). Given the well-established mortality benefits of this drug class, this is a major potential confounder. A group that starts with twice the rate of a cardioprotective medication has a significant survival advantage that cannot be attributed to the behavioral intervention.
Other Imbalances: Other baseline differences, though smaller, consistently suggest the control group may have been at a higher baseline risk (e.g., higher use of any glucose-lowering agent, insulin, and anti-hypertensives). This may arise as a result of the relatively small trial size.
Recommendation: Given these challenges, could the authors please:
Specify precisely how medication use was incorporated into the adjusted models. The paper states "treatment with…at any time throughout the study". Similar to use of a time-varying covariate this opens the possibility of confounding by indication.
Also provide the justification for the approach used and discuss how the chosen model mitigates the risk of confounding by indication.
Present a sensitivity analysis that adjusts only for baseline characteristics (including the baseline medication imbalances) and does not adjust for any post-randomization medication changes. This would provide an estimate of the intervention effect that is less susceptible to this specific type of bias.

We thank the Reviewer for raising this important issue, which we have carefully considered when designing the statistical approach, as discussed below.
Confounding by indication certainly represents a potential bias in clinical trials. However, type 2 diabetes is a progressive disease that requires more medications over time to maintain control of glucose and other modifiable cardiovascular risk factors (Fonseca VA. *Diabetes Care.* 2009;32:S151–S156). Moreover, with a physical activity-based intervention, between-group differences in change in medication throughout the

study are mainly related to the intervention itself, as participants in the intervention group inevitably require less drugs (anti-hyperglycemic, lipid-lowering and anti-hypertensive) than those in the control group in order to achieve comparable glucose, lipid and blood pressure levels, even if, in the IDES_2, the between group differences in changes in medications were not statistically significant (Balducci S et al. *JAMA.* 2019;321:880–890). This is because, in addition to drugs, participants in the intervention group receive a non-pharmacological treatment that is also effective in controlling these risk factors through increased energy expenditure and, therefore, they do not require (or require to a much lesser extent) the increase in drug number and dosage that is necessary in participants in the control group because of the disease progression.
Imbalances in the use of cardioprotective medications at baseline may also represent a potential confounding in a study assessing mortality and, in our study, this was the case especially for GLP-1 receptor agonists (GLP-1 RAs), since the differences in other medications were not significant.
Considering both aspects, we decided to adjust for use (yes/no) of the main cardioprotective drugs at any point throughout the 3-year follow-up including baseline, instead of baseline use only (line 360), in order to account also for medications that were started during the study period to fully account for the higher exposure to these agents in the control group. This was particularly important for GLP-1 RAs and SGLT-2 inhibitors, since at the time the recruitment was started, only one GLP-1 RA with proven cardioprotective action (liraglutide) and no SGLT-2 inhibitors were available, so that only 30 patients were on liraglutide (but for $\leq$6 months) and none on an SGLT-2 inhibitor. Since other GLP-1 RAs and several SGLT-2 inhibitors became available thereafter, the use of these agents inevitably increased in the following years and we thought it was important to account for it, in addition to treatment at baseline. In this way, we could consider the participants who were subsequently treated with an SGLT-2 inhibitor (4 in the intervention and 6 in the control group) or a GLP-1 receptor agonist, the number of whom increased in the control, but not in the intervention group, for much longer (up to 3 years and possibly after trial termination) than those on a GLP-1 receptor agonist before starting the study ($\leq$6 months).
However, following to the Reviewer's recommendation, we have also run a sensitivity analysis including the baseline treatment with all medications showing even a minimal imbalance between the two groups (line 364) and found virtually identical results as with treatment with cardioprotective agents at any time throughout the study (0.418 [0.231-0.755], p=0.004) (line 96).

3. Statistical Power
The original trial was not powered for a mortality outcome. It would strengthen the discussion to explicitly state the post-hoc power of the study to detect the observed mortality difference, or at a minimum, to discuss the implications of observing a statistically significant p-value in an underpowered analysis and the need for cautious interpretation.

We thank the Reviewer for this comment and agree with her/him that the original trial was not powered for a mortality outcome and that this limitation should be strengthened more explicitly, which we have done in the revised version (line 200).

4.Other Post Hoc Analyses
The analyses assessing the association of PA and fitness parameters with all-cause mortality (regardless of study arm) are valuable for generating hypotheses. The authors correctly label them "Post hoc analyses," and this exploratory nature should be consistently emphasized in the text to avoid any implication of confirmatory findings.

Thanks also for this suggestion. We have now emphasized the exploratory nature of these post hoc analyses throughout the manuscript.

5. Generalizability of Findings
The study provides potentially important evidence for a specific cohort (sedentary individuals aged 40-80 with type 2 diabetes and a BMI of 27-40). The authors should consider expanding the discussion on the generalizability of these findings to broader populations, such as younger individuals, those with different BMI ranges, or individuals without diabetes.

We agree with the Reviewer that generalizability is an important issue, which had already been discussed in the original paper for the primary and secondary outcomes (Balducci S et al. *JAMA.* 2019;321:880–890). We have now addressed it for mortality by stating that our data cannot be readily extrapolated to people with type 2 diabetes with different characteristics or to non-diabetic individuals and, hence, that further studies are required to confirm generalizability (line 215).

Reviewer #3

The authors present mortality follow-up data from the Italian Diabetes and Exercise Study-2. This was a trial of a behavioural intervention compared to control in 300 sedentary patients with type 2 diabetes in which the primary outcome was change in activity. This has been reported previously. The protocol, which was published in 2012 lists the following secondary endpoints; physical fitness, modifiable cardiovascular risk factors, MS disturbances, well-being/depression, and health related QoL. There is no mention of mortality follow up as an endpoint. The major observation of the paper is the apparent halving of mortality risk in the intervention group compared to control. This is a very sizeable effect.

We thank the Reviewer for the comments and suggestions provided to us, which we found very useful for improving our manuscript. As stated by the Reviewer and clarified by us in the subtitle and throughout the revised version, this is a post-hoc, not a pre-specified analysis of the IDES_2, since at the time we designed the study, we did not plan to assess long-term mortality due to the relatively small sample size. Therefore, we were surprised to find such significant results (below).

The first consideration is whether this observation is true or could be a manifestation of chance, bias or confounding. The p value for the main comparison is 0.01 which is relatively modest, so the role of chance can't be excluded. The study was not blinded, at least for the clinicians and the patients but some of the outcome assessors were blinded. It is possible that physicians may have altered care for patients in the intervention group. Finally, although randomisation reduces the chance of measured and unmeasured confounding, it does not eliminate it. This paper does not include a comparison of characteristics of the patients at baseline. However, it is noteworthy that a previous paper showed that 32% of the control population were current smokers compared to 25% in the intervention group. The authors do present adjusted analyses, but do not appear to have adjusted for smoking. This is perhaps relevant since the big difference between groups is in cancer-related mortality (19 deaths in the control group and 5 in the intervention group).

We thank the Reviewer for this very well-taken comment. Of course, we have seriously considered that the sizeable between-group difference in mortality could be a manifestation of chance, bias or unmeasured confounding and, hence, we have run several sensitivity analyses to rule out or at least reduce this possibility. Regarding the specific issues raised by the Reviewer:
- The relatively modest *p* value was due to the relatively small size of the sample, since the difference in terms of number of deaths was quite wide, with almost half of deaths in the intervention than the control group, as also outlined by the Reviewer; moreover, in the Cox proportional hazards regression analysis, the *p* value further increased when adjusting for confounding.
- As stated in the Methods section (line 284) and Study Protocol, treatment was adjusted using a pre-specified, target-driven algorithm in order to avoid performance bias.
- Though randomization reduces, but does not eliminate the chance of measured and unmeasured confounding, there were no significant between-group differences in the study parameters (Balducci S et al. *JAMA.* 2019;321:880–890), including smoking, though the number (not percent) of current smokers was somewhat higher in the control (32, 21.3%) versus the intervention (25, 16.7%) group.
- Given the importance of smoking in cancer (and cardiovascular) mortality, we did adjust for it, since smoking is one of the parameters used for calculating the UKPDS risk score, which was included as covariate in the Cox proportional hazards regression models (line 355); however, following to the Reviewer's suggestion, we have repeated these analyses by including smoking and the other the individual risk factors as covariates instead of the UKPDS risk score and found almost identical results (0.434 [0.238-0.790], *p*=0.006).

- As acknowledged in the study limitations section (line 212), it remains the possibility that unmeasured confounders may have influenced the results; however, we strongly believe that a major strength of our paper was the assessment of a wide range of parameters to account for in the statistical analysis, thus reducing potential confounding bias.

The main result paper from this study in JAMA in 2019 contains a flow chart similar to that provided here as Figure 1. However, there are some key differences. The number of individuals lost to follow up in the intervention group (n=17) and the control group (n=16) are identical between the two reports but the explanations for the reasons for lost to follow are very different. The JAMA report has no mention of any loss to follow up due to death, whereas Figure 1 in this report mentions 2 deaths in the intervention group and 6 in the control group. This difference needs some explanation. If it is correct, the authors need to explain why death is considered a "loss to follow up" in the context of a mortality outcome analysis.

We thank the Reviewer for this comment and agree with her/him that reporting death as a reason for loss to follow-up is incorrect. In fact, these 8 individuals (6 in the control and 2 in the intervention group) interrupted participation in the study because (and when) they were diagnosed with a serious disease, as reported in the 2019 JAMA paper (Balducci S et al. *JAMA.* 2019;321:880–890), not because (and when) they died thereafter for it, as (erroneously) reported in the current paper. Though this was done to indicate the individuals who died during the 3-year study period and were therefore excluded from the sensitivity analysis, Figure 1 has been changed in the revised version.

The major surprise of this study is that they were able to demonstrate an apparently significant mortality impact in 300 individuals when much larger studies have been inconclusive. One explanation could be the risk profile of the individuals included. The absolute event rate in this study is 23.6 per 1000 pyfu in the control group and 11.2 in the intervention. The event rate in LOOK Ahead, by comparison is 19 per 1000 pyfu, so this is not so different.

As stated above, we agree with the Reviewer that the results of our study are unexpected, given that much larger studies have been inconclusive, and the relatively small sample size was in fact included among the study limitations (line 203). However, as outlined in the revised version (line 206), of these previous studies, the only one that reported a significant reduction in mortality with a lifestyle intervention (though it was significant in the diet and exercise and diet only, but not in the exercise only group) was the Da Qing Diabetes Prevention Outcome Study, which was also relatively small-sized (n=577). Conversely, the studies that showed no effect on mortality include the Finnish Diabetes Prevention Study (n=522), but also the much larger Diabetes Prevention Program (n= 3,234) and Look AHEAD Study (n= 5,145). This might suggest that what matters is the type of intervention and its ability to produce a sustained behavior change involving all domains of physical activity/sedentary behavior in all settings, as discussed in our manuscript (lines 162 and 189).
We also agree with the Reviewer that the results cannot be related to differences in risk profile of participants in the IDES_2 versus the other studies, given the quite similar mortality rate, a finding that further supports our interpretation.

Reviewer #4

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

We thank the Reviewer for the her/his work.

Reviewer #5

Your finding of marked differences in long-term mortality between groups is substantial and fills a notable gap in the literature, where few RCTs have provided such conclusive evidence beyond observational studies.

With minor clarifications as suggested above, this is a strong manuscript for publication and will be of considerable interest to the field.

We are indebted to the Reviewer for the positive comments and useful suggestions.

The tailored nature of your intervention, accounting for participants' psychosocial, socio-economic, and cultural circumstances, is remarkable and likely central to its sustained efficacy; this could be highlighted more assertively in the abstract and discussion.

We thank the Reviewer for this comment. As suggested, was have listed among the strengths of the study the tailored nature of the intervention (line 196) and highlighted in the Discussion section its importance in the efficacy of the intervention in producing a sustained behavior change, through the progressive improvements in psychological well-being and health related quality of life (line 127).

Consider clarifying, for reader benefit, the duration of both theoretical and practical content in the counselling sessions (line 228).

As requested, we have clarified the duration of the theoretical (15 min) and practical (75 min) parts of the counseling sessions (line 264).

It would also be helpful to briefly address how concerns about glycaemic imbalances when increasing physical activity were managed and how this may have influenced participants' willingness and ability to engage with the intervention.

As requested, we have clarified how concerns about glycemic imbalances were managed (line 271). This has certainly improved participants' willingness and ability to engage with the intervention, as attested by the >80-90% attendance to the theoretical and practical sessions and the only 8 mild hypoglycemic episodes occurring in relationship to these sessions (Balducci S et al. *JAMA.* 2019;321:880–890) as well as by the improvements in psychological well-being and health related quality of life observed in the intervention versus control group (Nicolucci A et al. *Sports Med.* 2022;52:643–654).

Reviewer #1

I would like to thank the reviewers for addressing the majority of my previous comments. Thanks also for the clarification regarding accelerometer wear time and its treatment in the analysis. While I appreciate the rationale for assuming that the unmeasured waking time (approximately 0.9 hours) is largely sedentary, I would suggest that not adjusting for wear time may still introduce bias, particularly if wear time varies meaningfully across participants. I would also argue that, if you wanted to avoid the influence of wear time, then it should be accounted for in the models. As a minimum, this limitation should be acknowledged in the discussion.

We thank the Reviewer for acknowledging our efforts to address all the issues raised.
As stated, we were concerned by the potential underestimation of total daily sedentary time (SED-time) calculated only based on the time participants wore the accelerometer. This is because, in our experience, people tend to wear the accelerometer mostly when they go out and do some physical activity (PA) and much less when they stay sedentary and possibly house dressed at home, which may be still adequate for measuring PA, but not SED-time. Moreover, the age range of participants in the IDES_2 was 40-80 years, thus including people employed (46.0%), who usually spend more time outside the house, or retired (33.7%) (the remaining were housewives [18.0%] and unemployed [2.3%], who generally spend more time at home, which may result in a large inter-individual variability in the accelerometer wear time and, hence, in the underestimation of SED-time.
However, we understand the Reviewer's concern regarding the potential bias from assuming that the unmeasured waking time was spent in sedentary activities and from not adjusting for accelerometer wear time and, therefore, the reasons why she/he considers this as a study limitation. To eliminate or at least greatly reduce this bias, we asked participants to wear the accelerometer in all waking hours and, in fact, the unmeasured waking time was always in the early morning and late evening and the accelerometer wear time was very high without significant variability (14.8±0.8 hours), implying that total SED-time had the same distribution of the accelerometer-measured SED-time. Nonetheless, when submitting the main manuscript reporting the effect of the behavioral intervention on the co-primary endpoints SED-time and PA, we showed total SED-time data calculated by summing the unmeasured waking time to the accelerometer-measured SED-time because (a) these values provide a better estimate of total SED time than accelerometer-measured values, fir the reasons outlined above; and (b) this decision was taken before initiating the study, not after terminating it, as reported in the study protocol.
This specific issue was discussed with one Reviewer, who also agreed on the rationale for choosing this method of assessing SED-time, but asked us to show that the results did not change significantly when using the accelerometer-measured SED-time instead of total SED-time (accelerometer-measured SED-time + unmeasured waking time), as we did.
In this manuscript, we used total SED-time for consistency with the main paper and the other previous papers reporting the IDES-2 results, also considering that, here, SED-time was not the primary endpoint, but only a covariate. Therefore, following to the Reviewer's suggestion, we have repeated the analyses on the relationships between baseline or mean SED-time and mortality by using the accelerometer-measured SED-time instead of total SED-time and found again similar results (see the table below).

	Total SED-time			**Accelerometer-based SED-time**		
Baseline values						
Unadjusted	1.262	1.012-1.572	0.039	1.284	1.014-1.626	0.038
Age- & sex-adj	1.246	0.987-1.572	0.065	1.243	0.975-1.585	0.080
Mean values throughout the study						
Unadjusted	1.239	1.002-1.532	0.048	1.261	1.011-1.572	0.040
Age- & sex-adj	1.231	0.984-1.541	0.069	1.256	0.984-1.601	0.064
Both						
Baseline values	1.246	0.987-1.572	0.065	1.266	0.999-1.604	0.051
Mean values	1.119	0.771-1.626	0.554	1.193	0.834-1.707	0.333

Reviewer #3

The authors have added the important amendment that this is a post-hoc rather than a pre-specified analysis. They have added a more extensive discussion of the limitations of the study. They have addressed the inconsistency in the study flow diagram. There are still a number of places in which the language is too strong and should be softened and another in which the meaning is unclear.

We thank the Reviewer for appreciating the amendments we added and for providing additional suggestions to improve our paper.

Abstract. Line 27. "An inverse relationship between physical activity (PA) and mortality has emerged from epidemiological surveys, but not randomized clinical trials". This doesn't make sense. A relationship doesn't emerge from a survey. This sentence should be rewritten and framed as a relationship being observed in a survey.

This sentence has been modified as follows "An inverse relationship between physical activity (PA) and mortality has been observed in epidemiological surveys, but in not randomized clinical trials".

Abstract. Line 36. "A behavioral counselling targeting all domains of PA and sedentary behavior reduced long-term mortality risk in people with type 2 diabetes" Given the many uncertainties in this study the sentence should be reframed to state that the counselling intervention was associated with a reduction in long term risk rather than it reduced it.

This sentence has been modified as follows "A behavioral counselling targeting all domains of PA and sedentary behavior is associated with reduced long-term mortality risk in people with type 2 diabetes".

Conclusions. Line 221. "This follow-up analysis of the IDES_2 showed that a three-year behavioral counselling promoting sustained changes in all domains of PA/sedentary behavior produced a ~50% reduction in all-cause mortality versus standard care." As with the sentence above, this should be rewritten to make it clear that the intervention was associated with a 50% reduction in risk rather than it produced it.

This sentence has been modified as follows "This follow-up analysis of the IDES_2 showed that a three-year behavioral counselling promoting sustained changes in all domains of PA/sedentary behavior was associated with a ~50% reduction in all-cause mortality versus standard care".

Reviewer #4

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

We thank the Reviewer for evaluating our manuscript.

Reviewer #5

Thank you for your thorough revisions. The manuscript is notably improved, and the presentation of your findings is considerably clearer and more balanced. By consistently highlighting throughout the text that this work represents a secondary, post hoc analysis with a long-term observational data collection period, the manuscript now has greater transparency and interpretive caution. The implications of the results are more appropriately framed, enhancing the nuance of your argument. Overall, the revised manuscript is more robust.

We thank the Reviewer for her/his positive evaluation of our revised manuscript and thank her/his again for providing useful suggestions to improve our paper.