

Psychosocial interventions indicate prolonged survival in cancer patients in a systematic review, meta-analysis, and multiverse meta-analysis of randomized controlled trials

Corresponding Author: Mr Kenji Asakawa-Haas

This file contains all editorial decision letters in order by version, followed by all author rebuttals in order by version.

Version 0:

Decision Letter:

** Please ensure you delete the link to your author homepage in this e-mail if you wish to forward it to your coauthors **

Dear Mr Asakawa-Haas,

Thank you for your patience during the peer-review process. Your manuscript titled "Psychosocial interventions and survival time in cancer patients: A coherent conclusion to a controversial topic by a systematic review, meta-analysis, and multiverse meta-analysis of randomized controlled trials" has now been seen by 4 reviewers, whose comments are appended below. You will see that they find your work of some potential interest. However, they have raised quite substantial concerns that must be addressed. In light of these comments, we cannot accept the manuscript for publication, but would be interested in considering a revised version that fully addresses these serious concerns.

We hope you will find the Reviewers' comments useful as you decide how to proceed. Should additional work allow you to address these criticisms, we would be happy to look at a substantially revised manuscript. If you choose to take up this option, please highlight all changes in the manuscript text file, and provide a detailed point-by-point reply to the reviewers.

Please bear in mind that we will be reluctant to approach the reviewers again in the absence of substantial revisions.

Editorially, we consider it important that the revised manuscript includes a better justification of the methods used (e.g., reliability of conversion to hazard ratios) and include Embase in the systematic search. Please adopt the additional analytic suggestions provided by the reviewers or provide additional clarification for why your analytic approach is more suitable, in particular R2 suggests using the ROB-2 approach, Bayesian analysis, and GRADE assessment and R4 recommends using multilevel meta-analysis and/or robust variance estimation, and multiple imputation for random forest moderator analysis.

I am attaching a checklist that details critical reporting requirements for the revised manuscript. Please attend to each item and ensure your manuscript is fully compliant. We are requesting that your manuscript aligns with these requirements as this facilitates the evaluation of your manuscript, reducing delays in re-review and potential future acceptance. If your revised manuscript is not aligned with these requests on major issues, such as those concerning statistics, it may be returned to you for further revisions without re-review. Additional information can be found in our style and formatting guide Communications Psychology formatting guide.

If the revision process takes significantly longer than five months, we will be happy to reconsider your paper at a later date, provided it still presents a significant contribution to the literature at that stage.

We would appreciate it if you could keep us informed about an estimated timescale for resubmission, to facilitate our planning.

We are committed to providing a fair and constructive peer-review process. Please do not hesitate to contact us if you wish to discuss the revision in more detail.

Please use the following link to submit your

- revised manuscript,
- point-by-point response to the referees' comments,
- cover letter (as a separate document),
- the Editorial Policy Checklist (see below),
- the Reporting Summary (see below), and
- the completed Editorial Request Table (attached):

Link Redacted

** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first **

Thank you for the opportunity to review your work.

Best regards,

Jennifer Bellingtier

Jennifer Bellingtier, PhD
Senior Editor
Communications Psychology

REVIEWER EXPERTISE:

Reviewer #1 psychosocial cancer interventions, metaanalysis
Reviewer #2 psychosocial cancer interventions, metaanalysis
Reviewer #3 psychosocial cancer interventions, metaanalysis
Reviewer #4 statistical methods, metaanalysis

REVIEWER REPORTS:

Reviewer #1 (Remarks to the Author):

This is a long awaited and admirable effort to explain past inconsistencies in results and conclusions for primary studies and meta-analyses regarding the effect of psychosocial interventions on cancer survival. The paper is an important contribution to the field. It is well-written and conducted, and I have only a few comments and questions to the authors.

Sponsorship bias (economic interest) is not the only researcher bias. Many of the primary researchers and the meta-analytic researchers probably wish to find an effect. There are probably few researchers that have put in a lot of effort trying to prove that psychosocial interventions do not have an effect on cancer. Please include a short discussion of researcher biases other than economic ones.

Why was Embase not searched? The authors cite Bramer et al (2017) as support for not including Embase. But Bramer (2017) wrote: "Optimal searches in systematic reviews should search at least Embase, MEDLINE, Web of Science, and Google Scholar as a minimum requirement to guarantee adequate and efficient coverage." (Bramer WM, Rethlefsen ML, Kleijnen J, Franco OH. Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. Syst Rev 2017;6(1):245. DOI: 10.1186/s13643-017-0644-y").

On lines 333-334 the authors state that proportional hazards were assumed. What is the evidence that the survival hazards in this field are proportional?

On line 434-435 the authors state that meta-regression was applied if at least three studies were available. This is confusing because they also acknowledge that there should be at least 10 studies per variable in a meta-regression. They also conduct other analyses (MetaForest), which are not vulnerable to this problem. Please clarify.

Table 3: "PICTOS" should be "PICOTS".

Line 999 (and several other places): The Stagl et al study had 15 dead and 105 not dead in both groups. How can this produce a hazard ratio of 0.21?

Line 1009: Why is the reference line not at 1?

On lines 111-112, the authors state that the first meta-analysis on this topic was conducted almost two decades ago 112 (Chow et al, 2004).

Please use «published» instead of «conducted» (The Smedslund meta-analysis was started early in 2001, was submitted in November 2002, accepted in September 2003, and published in August 2004, 7 months after the Chow paper.

If the authors take these comments and questions into account, I approve of the manuscript for publication in Communications Psychology.

Reviewer #2 (Remarks to the Author):

The systematic review addresses a significant public health issue: cancer survival, specifically focusing on the efficacy of psychosocial interventions in prolonging the survival of cancer patients. This is a highly controversial and unresolved topic. The results of existing studies are mixed, with nearly half of the randomized controlled trials (RCTs) indicating a life-extending effect, while the other half do not. This inconsistency is also reflected in prior meta-analyses, where some report significant effects while others do not. Tackling this issue is ambitious and commendable. The paper addresses a complex issue using highly heterogeneous data and relatively complex methodologies.

In addition to my specific comments (see below), I have a more general concern. The authors conclude that “psychosocial interventions seem to enhance cancer survival and should therefore be made accessible to all cancer patients.” I question whether this conclusion is sufficiently supported by the data. While there are many reasons to offer psychosocial interventions to cancer patients, I worry that the concluding statement may give the impression that the scientific debate has reached a coherent and final conclusion, which might seem somewhat overconfident. Key issues include the broad definitions of “psychosocial intervention” and “cancer patients,” as well as the considerable heterogeneity noted by the authors.

Overall, the manuscript has significant potential and addresses a critical and complex issue with a thoughtful approach. I hope my feedback is helpful in refining the conclusions and clarifying the strength of the evidence presented. While I may not be able to address every issue, I have tried to identify the main strengths and limitations.

1. Strengths

1.1. The introduction provides an overview of the existing meta-analyses on this topic and their strengths and limitations and presents possible reasons for conflicting results of the 10 previous meta-analyses, including the differences between the included studies, e.g., in the chosen effect size. The important issue regarding the failure to use HR as the outcome measure in several studies – and meta-analyses is highlighted. The introduction also discusses the frequent failure to examine possible publication bias and p-hacking - and when doing so – using less optimal methods, including fail-safe numbers and the trim-and-fill procedure.

1.2. The authors conducted a very comprehensive state-of-the-art meta-analysis of existing RCTs. State-of-the-art methods include prediction intervals and bias assessment with several more recently developed methods, including contour-enhanced funnel plots, the FAT-PET-PEESE regression test, and p-uniform. In addition, they apply a novel approach to exploring the causes of heterogeneity by identifying the WHICH and HOW specifications concerning the data analyses of previous meta-analyses.

2. Issues and limitations

2.1. I realize that this is not the main topic, but the references on psychological risk factors for cancer development and progression seem somewhat dated. Please include meta-analyses and more recent reviews if possible, perhaps also including more critical results. For example, it is not quite clear whether “stress”, which in itself is a very heterogeneous term and operationalized in many different ways, is a risk factor in all cancers, genders, and age groups.

2.2. The paper is highly complex. The introduction resembles a systematic overview of reviews. I would have preferred a two-tiered process, first conducting a proper systematic review of reviews, including the specification-curve plots and multiverse analysis, followed by an analysis of primary studies, guided by the results. The results could perhaps even be submitted as two separate papers.

2.3. The authors are commended for their attempts to calculate and convert various data into HRs. One wonders about the reliability of these data. Do the authors have any data on the reliability, e.g., by comparing independent raters.

2.4. The authors assessed “study quality” with the Nat Heart, Lung, and Blood Institute scale. Risk of bias and study quality are not identical. Several study quality aspects may not infer risk of bias, and one wonders why the Cochrane risk of bias (ROB-2) approach wasn’t used. Rather than a total study quality score, the ROB-2 analysis could provide a clearer picture of the risk-of-bias-related issues affecting the various studies.

2.5. I do not have the expertise to evaluate the machine-learning-based approach.

2.6. There might be some issues with publication bias assessment. The FAT-PET-PEESE approach appears to be challenged by small sample sizes and heterogeneity.

2.7. Visualizing results: Figure 1A-B is a bit difficult to follow as it includes both primary studies and meta-analyses in the same figure. Also, the reader tries to interpret why some studies are above and some below the timeline. I wonder whether the information could be presented in a more clear and easier-to-understand way. Figure 2 contains some of the same information. The note in Figure 5 states that the reference line is drawn at HR=1. It obviously isn’t – it seems to be drawn at HR=0.78.

2.8. I am a bit worried about the p-curve analysis and wonder whether the small number of studies allows a sufficiently robust analysis. Also, in this controversial area, could one imagine p-hacking in the opposite direction, i.e., in the direction of null-findings?

2.9. The overall median survival benefit was 3.8 months. While it does appear meaningful, how does it compare to, e.g., the benefits reported for new medical treatments? Some more detailed examples could be mentioned here. Also, I remain less convinced by the authors' rather strong concluding statements and recommendations given the considerable heterogeneity and very broad prediction interval.

2.10. The conventional frequentist approach has limitations which could be overcome using Bayesian approaches. Unlike frequentist methods, which provide the probabilities (p-values), given a hypothesis, Bayesian methods enable direct probability statements about the hypotheses themselves. Furthermore, unlike p-values, which only indicate the degree of "surprisingness" of the data given that the null hypothesis is true, Bayes factors quantify the support for one model compared to another, and Bayesian methods avoid other issues associated with null-hypothesis significance testing, such as the over-reliance on relatively arbitrary p-value thresholds and the dichotomization of results into "significant" and "non-significant." Such an analysis would help the reader assess the robustness of the results.

2.11. Conclusions regarding the clinical value need to be supported by an overall evaluation of the findings, e.g., with the GRADE assessment or similar. Clinicians and future attempts at clinical practice guidelines would benefit from this.

Reviewer #3 (Remarks to the Author):

I do not habitually start my reviews with "wow", but here I do.

The first 'wow' is sheer length; it is my first review of a manuscript that is over 100 pages long (among 500 + lifetime reviews)! That of course triggered consideration of "wow, nr 2": 'is there really so much important information here that this length is needed?' Overall, I will say 'yes' there is (detail to follow of course). The topic is of utmost importance to the field of Psycho-Oncology, is enormously complex, and has produced highly variable and confusing conclusions. The authors do this complexity justice with a thorough literature review and the building of a convincing rationale for this in-depth and nuanced review of outcomes and of moderators and mediators.

The issues around type of statistic used, bias control, significance fishing etc that are raised and thoroughly investigated here is impressive and does indeed lead to trustworthy conclusions. Given the many different types of cancer and the various stages when patients were entered into trials, results are bound to be variable and hard to replicate. These authors are very aware of these challenges.

Please note that there were occasions where these authors' handling of statistics exceeded my skill level even though I have published a half dozen meta-analyses myself.

Potential sources of heterogeneity are thoroughly investigated here and it comes as no surprise that: having access to more RCTs (in 2024) increases power to find small effects, that HR is more precise tool than OR and RR, and that not all cancer types and stages are associated with same sized effect.

The manuscript is very clearly written, the objective spelled out and the rationale-building is sound.

[1] As for my more critical comments, I will start with some personal opinion and context. When the first studies on this topic were planned and completed (Spiegel (a co-author of this paper), Fawzy, Cunningham, and others, all found in the tables of this Ms), there was major headscratching about how we can answer the core question with a scientifically sound approach that is still ethically acceptable. We cannot get a sound answer about 'does psychotherapy extend life' without tightly controlled RCTs where patients are randomly assigned to a wait-list, or a no treatment, or a minimal control condition. Then we have to follow all groups for many years and avoid confounding treatment; the ethics of this approach is hugely problematic. Surprisingly, the authors do not touch very much on this tricky issue.

There should be no surprise that initial studies were not perfectly tight because they also had to get approved by the ethics committees first. Critics of methodology like Ioannidis were justified in their challenges but maybe not have been overly sympathetic to the plight of researchers trying to strike the right balance of science and ethics.

Then things got really confused by Grossarth-Maticek and Eysenck who published research on psychological treatment of cardiac and cancer patients based on their personality typologies that sounded fabulous but were ultimately unbelievable. Grossarth-Maticek's effect sizes were huge and are now considered plain fraudulent (Pelosi, 2019). Unless previous meta-analyses had explicitly bracketed out the Grossarth-Maticek findings, they were inevitably over-representing the benefits of psychotherapy and should not be used for drawing conclusion. Now, two Grossarth-Maticek references are found in the reference list of this manuscript but I cannot find any mention of this work in the manuscript text or in the tables. Is that an oversight by the authors, intention but just unexplained, or an oversight on my part?

I have to admit that I am uncertain about how much of this colorful history of the field has to be described here but... it does help understanding how we got to the status quo which is rightfully and capably challenged by the researchers in this review; it has certainly colored the glasses I was wearing as a (now retired) reviewer.

[2] A second point that needs add'l attention (and will add to the length of the paper!!, so sorry ...) is that of floor effects. Psycho-oncology and Psychocardiology often do tackle psychotherapy offerings differently than psychiatry. In psychiatry, treatment of depression is offered to people who were diagnosed as depressed. But, in behavioral medicine treatment is often provided for a medical condition that may be linked with elevated rates of anxiety or depression for some, but not all patients. That means that patients without depression will be treated for depression and the possible improvements are nil or miniscule; effects get watered down. How often does this happen? In a 2008 review (Jacobson & Jim), it was revealed that 95% of all clinical trials in psycho-oncology had not first screened for distress at treatment entry, thus allowing the majority of patients to enter a trial for distress reduction although they were not overly distressed a priori. That should be seen in light of another review (Linden & Satin, 2007) that had reported on floor effects in psycho-oncological treatments and revealed that the correlation of distress scores at treatment entry with the degree of post-treatment change was $r=0.75$, thus explaining 57% of the variance in treatment outcome. We know that resulting floor effects suppress possibly good outcomes; Linden

and Girgis (2012) have shown that patients with cancer and screened for actually diagnosed depression/anxiety showed three time greater benefits on these endpoints than when all patients with cancer are treated for anxiety/depression. What we don't know but need to see addressed here is whether floor effects also suppress findings of greater longevity when patients with actually documented depression/anxiety had been recruited and treated.

[3] In my review above, I referred to the difficulty of generating a control group that is scientifically sound while ethical. The authors inform the reader that the inclusion criteria for this review include a design that has a psychologically active treatment and either a placebo or passive controls (no intervention). Being a clinical psychologist and scientist-practitioner, I will argue that there is no such thing as a true placebo control in psychotherapy; any treatment, even minimal, might lead to positive change. When that happens, effect size differences will shrink and results become more conservative. Missing here is coverage of placebo was operationalized. And how is avoidance of treatment confounding implemented during the post-treatment observation period of many months and years?

[4] minor suggestion: In the Abstract, tell the reader what an HR of .78 for longer life means in terms of months gained.

[5] The Ms length is overwhelming; 103 pages long. Mind you actual text is only 36 pages and is not wordy. I habitually push for brevity (as do editors who control precious journal pages) but these authors deserve a looong leash; I see no obvious place for text cutting. The editors may have some ideas about tables and figures that could possibly go into a supplement.

References

- Jacobsen PB, Jim HS (2008). Psychosocial Interventions for anxiety and depression in adult cancer patients: achievements and challenges. *CA Cancer J Clin* 2008; 58:214–230.
- Linden W, Girgis A (2012). Psychological treatment outcomes for cancer patients: What do meta-analyses tell us? *Psycho-Oncology*. 21, 343-350. DOI: 10.1002/po)
- Linden W, Satin JR (2007) Avoidable pitfalls in behavioural medicine outcome research. *Annals of Behavioral Medicine*, 33, 143-147
- Pelosi AJ (2019). Personality and fatal diseases: revisiting a scientific scandal. *Journal of Health Psychology*, 24 (4), 421-439.

Wolfgang Linden

Reviewer #4 (Remarks to the Author):

In this paper, the authors report on a meta-analysis evaluating the effect of psychosocial interventions on the survival time of patients with cancer.

In general, the study is done in a meticulous and systematic way. I like the insightful reflections on possible reasons why prior meta-analyses came to conflicting results; these reflections also give a clear rationale and guideline for the current study. The authors start their meta-analysis with a thorough search and selection procedure for relevant literature on the topic. Data extraction seems to be accurate, and analyses well thought-out. I appreciate the use of various techniques (for instance to explore possible publication bias) as well as the use of modern and innovative techniques (mainly the use of machine-learning techniques to evaluate moderator effects and the use of multiverse analysis approach to evaluate the dependence of the meta-analytic results on decisions taken).

Overall, I am very positive about this study. Yet, I still have several (mainly minor) suggestions and comments:

First, the authors suggest that with their study, we have a final answer to the question what the effect is of psychosocial interventions (e.g., line 5: "A coherent conclusion to a controversial topic", line 50 "... may now have arrived at a coherent and probably final conclusion"...). I agree that several of the issues in previous research are avoided. Yet, I do not believe that this study does give a final answer. It was found that there was considerable effect heterogeneity, resulting in a large prediction interval. Moreover, it appeared hard to find clear moderator effects, which means that a large part of the heterogeneity remains unexplained (by the way, it would be interesting to have this information as well). Therefore, more primary research, followed by meta-analytic research is needed to get a clearer understanding of when and for whom psychosocial interventions work. Related to that, I noted that for some kinds of cancer (e.g., brain, gastric, leukemia, ...), data from only one single study are available, which is definitely not sufficient to know what the effect is for each of the domains. I therefore do not agree with their conclusion that their findings "provided robust evidence for the existence of a generalizable true overall survival-enhancing effect" (line 834). Their recommendation that psychosocial interventions "should be made accessible to all cancer patients" (line 910) might be the best recommendation for now, but we cannot see this recommendation as the final guideline, because their prediction interval shows that effects might also be negative. I also think the authors are too optimistic when they say: "Therefore, comprehensive approaches, such as our present analysis, are often needed to get a valid overall picture to propel science forward when it seemingly got stuck" (line 904). In sum, I think some more modesty would be appropriate.

Second, like I mentioned above, I like the thorough way the data are analyzed and the fact that often multiple techniques and analyses are done and results compared. A drawback however is that the paper is very long, and the reader might get lost in all the detailed analyses and results, and lose the focus on the main findings. I therefore think that it would be a good idea to shorten the paper further, by moving some more results to the supplemental materials.

Third, the authors state that they used random and mixed effects models. 'Traditional' random and mixed effects models

assume that effect sizes are independent, which is probably not the case when there are multiple effect sizes per study. Multiple effect sizes per study is in fact rather common, for instance because effect sizes may be reported for several subpopulations or time points. If effect sizes are dependent, it is necessary to account for the dependence. An appropriate way to do so is using multilevel meta-analysis and/or robust variance estimation. The authors are not clear about whether there were multiple effect sizes per study and, if so, how they handled this issue.

For the random forest moderator analyses, missing moderator values were replaced with the median or mode. This approach has strong disadvantages: It does not take into account the uncertainty we have about the imputed values. Moreover, it decreases a possible association between the moderator and the effect sizes, therefore giving biased results. If a considerable number of scores would be missing (unfortunately, no information is given on this percentage), using multiple imputation might be a more valuable approach.

Details:

- In the text, the authors often refer to tables and figures in the supplemental materials, without saying that these can be found in the supplemental materials.
- Line 295: study screening was done independently by two researchers. It is not clear whether the whole screening process was done by two reviewers. A measure of agreement would be valuable.
- Line 425: I find the explanation of a mixed-effects model not very helpful, i.e., "using a random-effects model within, but a fixed-effect model between subgroups". Only when one already has a good understanding of a mixed-effects model, (s)he can understand what the authors mean.
- Line 453 says that 20,000 trees were used, line 635 talks about 10,000 trees.
- Line 461: what is a "50% percentile interval"?
- Line 604: the authors say that in Figure B1 no time trend can be identified. Yet, I found it striking that after 2015 all ratios were smaller than 1 (positive effect), whereas before 2015 effects were fluctuating around 1 (no effect). Also in Figure B3 (cumulative meta-analysis), it can be seen that after 2015 the mean ratio is gradually decreasing.
- Line 610 (and 735): the authors suggest that from 2015 onwards, the effect becomes statistically significant because of an increased precision due to a larger number of available studies. Yet, also here the changing mean effect size plays a role
- In some figures or tables including ORs, SMDs or HRs, it would be helpful to remind the reader how the direction of the effect size should be interpreted (e.g., does a positive or a negative SMD or a OR/HR larger or smaller than 1 indicate that the intervention is successful? (in this regard: I like the use of the 'median survival difference', because this helps the reader to understand the effect sizes).
- Line 656: not significant, or significant but in the opposite direction as one would expect.
- Line 664: the use of the Meta-Showdown Explorer is not very clear. It is suggested that "no bias" is used as input, but how comes that in that case the power can be evaluated (is this the power of the mean effect size?)?

EDITORIAL POLICIES

We ask that you ensure your manuscript complies with our editorial policies and reporting requirements.

To that end, we require revised manuscripts to be accompanied by two completed items: a reporting summary that collects information on study design and procedure, and an editorial policy checklist that verifies compliance with all required editorial policies

- <https://www.nature.com/documents/nr-reporting-summary.zip>>Nature Research Reporting Summary
- <https://www.nature.com/documents/nr-editorial-policy-checklist.pdf>>Editorial Policy Checklist

All points on the policy checklist must be addressed. Your revised manuscript can only be sent back to the referees if these checklists are completed and uploaded with the revision.

Notes: If you have submitted a Stage 1 Registered Report, Review, Primer, Comment, or Perspective you do not need to submit these forms. If you have already submitted these forms, you may disregard this request.

* **TRANSPARENT PEER REVIEW:** Communications Psychology uses a transparent peer review system. This means that we publish the editorial decision letters including Reviewers' comments to the authors and the author rebuttal letters online as a supplementary peer review file. However, on author request, confidential information and data can be removed from the published reviewer reports and rebuttal letters prior to publication. If your manuscript has been previously reviewed at another journal, those Reviewers' comments would not form part of the published peer review file.

** Visit Nature Research's author and referees' website at <http://www.nature.com/authors>>www.nature.com/authors for information about policies, services and author

benefits**

Communications Psychology is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the Manuscript Tracking System by clicking on 'Modify my Springer Nature account' and following the instructions in the link below. Please also inform all co-authors that they can add their ORCIDs to their accounts and that they must do so prior to acceptance.

<https://www.springernature.com/gp/researchers/orcid/orcid-for-nature-research>

For more information please visit <http://www.springernature.com/orcid>

If you experience problems in linking your ORCID, please contact the [Platform Support Helpdesk](http://platformsupport.nature.com/).

Version 1:

Decision Letter:

** Please ensure you delete the link to your author homepage in this e-mail if you wish to forward it to your coauthors **

Dear Mr Asakawa-Haas,

Thank you for your patience during the peer-review process. Your manuscript titled "Psychosocial interventions and survival time in cancer patients: A unifying consensus clarifying a controversial topic via systematic review, state-of-the-art meta-analysis, and multiverse meta-analysis of randomized controlled trials" has now been seen by 3 reviewers, and I include their comments at the end of this message. They find your work of interest but there are a few remaining points that need to be addressed. We are very interested in the possibility of publishing your study in Communications Psychology, but would like to consider your responses to these concerns and assess a revised manuscript before we make a final decision on publication. In particular, we ask you to attend to Reviewer 2's points 3 and 8.

We therefore invite you to revise your paper one last time to address the remaining concerns of our reviewers and a list of editorial requests. At the same time we ask that you edit your manuscript to comply with our format requirements and to maximise the accessibility and therefore the impact of your work.

EDITORIAL REQUESTS:

Please review our specific editorial comments and requests regarding your manuscript in the attached "Editorial Requests Table". Please outline your response to each request in the right hand column. Please upload the completed table with your manuscript files as a Related Manuscript file.

If you have any questions or concerns about any of our requests, please do not hesitate to contact me.

Please submit the following items:

- Revised manuscript
- Response to Reviewers
- Cover letter (as a separate document)
- <https://www.nature.com/documents/nr-reporting-summary.pdf> Nature Research Reporting Summary
- Completed Editorial Request Table (attached).

via this link: Link Redacted .

** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first **

Additional guidance is available in our style and formatting guide <https://www.nature.com/documents/commpsychol-style-formatting-guide-accept.pdf> Communications Psychology

formatting guide.

We hope to receive your revised paper within 8 weeks; please let us know if you aren't able to submit it within this time so that we can discuss how best to proceed. If we don't hear from you, and the revision process takes significantly longer, we may close your file. In this event, we will still be happy to reconsider your paper at a later date, provided it still presents a significant contribution to the literature at that stage.

We would appreciate it if you could keep us informed about an estimated timescale for resubmission, to facilitate our planning.

We look forward to seeing the revised manuscript and thank you for the opportunity to review your work.

Best regards,

Jennifer Bellingtier

Jennifer Bellingtier, PhD
Senior Editor
Communications Psychology

REVIEWER EXPERTISE:

Reviewer #2 psychosocial cancer interventions, metaanalysis

Reviewer #3 psychosocial cancer interventions, metaanalysis

Reviewer #4 statistical methods, metaanalysis

REVIEWER REPORTS:

Reviewer #2 (Remarks to the Author):

The authors have now addressed several of the issues raised by this reviewer (#2). They have included the RoB 2 and GRADE evaluations (R2.4.; R.2.11), have improved the figures (R.2.7.), and have acknowledged the challenges related to publication bias assessment (R.2.6.). They have also provided lengthy responses to some of my questions. While the manuscript thus has been improved, I still think it is a long and complex paper with many different (and possibly underpowered) analyses of robustness, including p-curve analysis (R.2.8.) and two different moderator analysis approaches (R.2.5.). Furthermore, I am not completely comfortable with the HR calculations (R.2.3.) and the machine learning moderator analysis (with a small number of studies) (R.2.5.). Finally, given the many analyses and approaches included, I still think that dividing the paper into two separate papers (a review of reviews) and a meta-analysis of primary studies will improve the readability of the study (R.2.2.). Thus, while the manuscript represents a comprehensive and thoughtful approach to an important scientific and clinical question, and several improvements have been made, there are still issues that warrant a major revision decision.

R2.1. I asked for inclusion and review of more recent reviews and meta-analyses, and the authors have now added a list of previously published meta-analyses.

R2.2. My initial concern was that the paper is highly complex – combining what resembles a review of reviews with a systematic review and meta-analysis of primary studies. Although the authors defend their choice and have made some changes to the structure, I would prefer two separate papers published simultaneously in the same journal, i.e., a paper 1 which is written as a standard umbrella review/overview of reviews and paper 2 presenting an updated meta-analysis of the primary studies included in the meta-analyses plus any additional/newer studies.

R2.3. Re. the calculation of Hazard Ratios, the authors have added a sentence describing the procedure. How HR were calculated is still unclear – and I was not able to open the calculation sheets that were linked to. I would strongly suggest that a statistician be included in the review process.

R2.4. The authors have now, as suggested, used the RoB 2 tool to evaluate the risk of bias.

R2.5. I am still not sufficiently knowledgeable about Machine-learning approaches to moderator analyses. I appreciate that the power of moderator analyses may be limited, but the very extensive moderator analysis approach with both conventional meta-regression and additional machine-learning-based approaches seems to be a bit over-reaching with increased risk of type-1 error.

R2.6. The authors have now addressed the concern by including a comment in the discussion on the challenges in assessing publication bias with the FAT-PET-PEESE approach (with small sample sizes and heterogeneity).

R2.7. The figures 1A-B have been improved and are now easier to interpret. I am still not very enthusiastic with regard to including both primary studies and meta-analyses (of these studies) in the same timeline figure – the mistake in Figure 5 has

been corrected.

R.2.8. I remain worried about drawing inferences from a p-curve assessment with only 10 significant studies.

R2.9. Regarding my question concerning interpretability of results, the authors have provided a lengthy response and have included a section on the certainty of the evidence in comparison with evidence for medical cancer treatments. One still wonders about the possible mechanisms – and the degree to which evidence for such mechanisms exist.

R2.10. I suggested the Bayesian approach as a way to evaluate the robustness of the results that go beyond “significant and non-significant”. The authors argue that given their many different types of analyses, they would prefer to stick with frequentist approaches. I will not insist, but given that the paper is highly complex with many different (and possibly underpowered) analyses of robustness (e.g., p-curve analysis, etc.), this could have simplified the paper.

R2.11. The authors have – as suggested – included a GRADE assessment of the certainty of the evidence.

Reviewer #3 (Remarks to the Author):

It may seem a contradiction, but this reviewer who first complained about the length of the manuscript and then proposed that this level of detail was nevertheless needed, is now also willing to accept the even lengthier, revised version! The authors were exceptionally thorough in their revisions. I cannot comment in great detail on the responsiveness to all other reviewers' comments, but my concerns were more than adequately dealt with. If there is an absolute need for cuts (not I urge for these) then the discussion might be a little shorter, there is some repetition of results here.

In the end, we have a blockbuster manuscript here that will become a citation classic. Every criticism one can throw at the attempt to answer the question of life extension with a psychological treatment has been tackled. I especially like the addition of highlighting not just statistical significance, but what the results mean for cancer care in the broadest sense, and how patients themselves benefit. There is no stone left unturned, no potential criticism ignored. Kudos to the authors; I bow my head to their professionalism.

Reviewer #4 (Remarks to the Author):

Thank you for the thorough way you tried to account for all comments of the reviewers.
I am happy with the result.

Just a very small detail:

Please replace

"This is a standard approach for subgroup analysis and applies a random-effects model within each subgroup to account for variation among studies, while using a fixed-effect model to compare the average effects between subgroups"

by

"This is a standard approach for subgroup analysis and applies random effects to account for variation among studies within subgroups and fixed effects to account for variation between subgroups"

* TRANSPARENT PEER REVIEW: Communications Psychology uses a transparent peer review system. This means that we publish the editorial decision letters including Reviewers' comments to the authors and the author rebuttal letters online as a supplementary peer review file. However, on author request, confidential information and data can be removed from the published reviewer reports and rebuttal letters prior to publication. If your manuscript has been previously reviewed at another journal, those Reviewers' comments would not form part of the published peer review file.

** Visit Nature Research's author and referees' website at www.nature.com/authors for information about policies, services and author benefits**

Communications Psychology is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the Manuscript Tracking System by clicking on 'Modify my Springer Nature account' and following the instructions in the link below. Please also inform all co-authors that they can add their ORCIDs to their accounts and that they must do so prior to acceptance.

<https://www.springernature.com/gp/researchers/orcid/orcid-for-nature-research>

For more information please visit <http://www.springernature.com/orcid>

If you experience problems in linking your ORCID, please contact the [Platform Support Helpdesk](http://platformsupport.nature.com/).

Version 2:

Decision Letter:

**** Please ensure you delete the link to your author homepage in this e-mail if you wish to forward it to your coauthors ****

Dear Mr Asakawa-Haas,

Your manuscript titled "Psychosocial interventions prolong survival time in cancer patients evidenced by systematic review, state-of-the-art meta-analysis, and multiverse meta-analysis of randomized controlled trials" has now been editorially reviewed, including by one of our EBM experts with expertise in meta-analyses. In light of their advice I am pleased to say that we are happy, in principle, to publish a suitably revised version in Communications Psychology.

We therefore invite you to revise your paper one last time to address the remaining editorial concerns. In particular, acceptance is conditional upon the revised manuscript addressing the presentational concerns and all requests listed in the editorial request table.

EDITORIAL REQUESTS:

Please review our specific editorial comments and requests regarding your manuscript in the attached "Editorial Requests Table". Please outline your response to each request in the right hand column. Please upload the completed table with your manuscript files as a Related Manuscript file.

If you have any questions or concerns about any of our requests, please do not hesitate to contact me.

SUBMISSION INFORMATION:

In order to accept your paper, we require the files listed here <https://www.nature.com/documents/commsj-file-checklist.pdf>.

OPEN ACCESS:

Communications Psychology is a fully open access journal. Articles are made freely accessible on publication. For further information about article processing charges, open access funding, and advice and support from Nature Research, please visit <https://www.nature.com/commspsychol/open-access>

At acceptance, you will be provided with instructions for completing the open access licence agreement on behalf of all authors. This grants us the necessary permissions to publish your paper. Additionally, you will be asked to declare that all required third party permissions have been obtained, and to provide billing information in order to pay the article-processing charge (APC).

*** TRANSPARENT PEER REVIEW:** Communications Psychology uses a transparent peer review system. On author request, confidential information and data can be removed from the published reviewer reports and rebuttal letters prior to publication. If you are concerned about the release of confidential data, please let us know specifically what information you would like to have removed. Please note that we cannot incorporate redactions for any other reasons.

*** CODE AVAILABILITY:** All Communications Psychology manuscripts must include a section titled "Code Availability" at the end of the methods section. We require that the custom analysis code supporting your conclusions is made available in a publicly accessible repository at this stage; please choose a repository that generates a digital object identifier (DOI) for the code; the link to the repository and the DOI must be included in the Code Availability statement. Publication as Supplementary Information will not suffice.

* DATA AVAILABILITY:

All Communications Psychology manuscripts must include a section titled "Data Availability" at the end of the Methods section. More information on this policy, is available in the Editorial Requests Table and at <http://www.nature.com/authors/policies/data/data-availability-statements-data-citations.pdf>.

Please use the following link to submit the above items:

Link Redacted

**** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first ****

We hope to hear from you within two weeks; please let us know if you need more time.

Best regards,

Jennifer Bellington

Jennifer Bellington, PhD
Senior Editor
Communications Psychology

** Visit Nature Research's author and referees' website at www.nature.com/authors for information about policies, services and author benefits**

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Communications Psychology

New Manuscript Title

Psychosocial interventions and survival time in cancer patients: A unifying consensus clarifying a controversial topic via systematic review, state-of-the-art meta-analysis, and multiverse meta-analysis of randomized controlled trials

Dear Reviewers,

thank you very much for taking the time and effort to review our manuscript. We greatly appreciate your insightful and constructive feedback, which has helped us substantially improve the quality and clarity of our work. We have done our best to address your suggestions in the revised version of the manuscript. We believe that the manuscript has benefited significantly from your input.

To facilitate the review process, we have labeled your comments and our corresponding responses using the following coding system: for example, Reviewer 1, Comment 1 is labeled as R1.1, and so forth. These labels are clearly indicated in the tracked version of the revised manuscript. You may use the search function (e.g., search for "R1.1") to locate each revision quickly.

We also updated our literature search as of February 1st, 2025. During this process, we removed one previously included trial (which offered psychoeducation only) upon realizing that it actually did not meet our predefined inclusion criteria for psychosocial interventions. This decision reflects the frequent challenge in distinguishing psychosocial interventions from palliative care, which are sometimes nuanced. For more detail, please refer to the Methods sections titled "Specific Inclusion Criteria" and "Specific Exclusion Criteria."

Please note that this adjustment did not lead to any major changes in the overall effect estimates. In fact, the results became slightly more conservative: the hazard ratio changed from 0.78 (95% CI [0.69, 0.88]) to 0.80 (95% CI [0.71, 0.90]), without altering the interpretation of the overall findings. Likewise, it did not lead to any substantial changes elsewhere.

Once again, we sincerely thank you for your valuable input and support in improving our manuscript.

With kind regards,
Kenji Asakawa Haas
On behalf of all co-authors

REVIEWER EXPERTISE:

Reviewer #1 psychosocial cancer interventions, metaanalysis

Reviewer #2 psychosocial cancer interventions, metaanalysis

Reviewer #3 psychosocial cancer interventions, metaanalysis

Reviewer #4 statistical methods, metaanalysis

REVIEWER REPORTS:

Reviewer #1 (Remarks to the Author):

General impression: This is a long awaited and admirable effort to explain past inconsistencies in results and conclusions for primary studies and meta-analyses regarding the effect of psychosocial interventions on cancer survival. The paper is an important contribution to the field. It is well-written and conducted, and I have only a few comments and questions to the authors.

Reply: Thank you very much for your acknowledgement.

Arguments:

R1.1: Sponsorship bias (economic interest) is not the only researcher bias. Many of the primary researchers and the meta-analytic researchers probably wish to find an effect. There are probably few researchers that have put in a lot of effort trying to prove that psychosocial interventions do not have an effect on cancer. Please include a short discussion of researcher biases other than economic ones.

Reply: A discussion of several sources of bias was already included in the Introduction section of the originally submitted manuscript but was elaborated further in the revised manuscript see Introduction “Risk of Bias Within Studies” and “Reporting Bias Across Studies” in the revised manuscript.

R1.2: Why was Embase not searched? The authors cite Bramer et al (2017) as support for not including Embase. But Bramer (2017) wrote: ““Optimal searches in systematic reviews should search at least Embase, MEDLINE, Web of Science, and Google Scholar as a minimum requirement to guarantee adequate and efficient coverage.” (Bramer WM, Rethlefsen ML, Kleijnen J, Franco OH. Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. Syst Rev 2017;6(1):245. DOI: 10.1186/s13643-017-0644-y”).

Reply: EMBASE is now included in our updated literature search – see Abstract and Methods “Information Sources” in the revised manuscript.

R1.3: On lines 333-334 the authors state that proportional hazards were assumed. What is the evidence that the survival hazards in this field are proportional?

Reply: The proportional hazards (PH) assumption in survival analysis posits that the ratio of hazard rates between two groups remains constant over time, which is essential for the validity of the Cox proportional hazards regression model. Both primary studies and prior meta-analyses in this field have adopted this assumption.

Given the unavailability of individual patient data, we assessed the validity of the PH assumption by visually inspecting published Kaplan-Meier survival curves from the included randomized controlled trials (RCTs). This approach aligns with standard practices in the absence of individual-level data. Established guidelines suggest that if these curves do not cross and maintain a relatively constant distance over time, it indicates that the PH assumption is likely valid. In our analysis, we observed no major violations of this assumption.

We acknowledge that visual inspection has limitations and that formal statistical tests are preferred when individual patient data are available. However, in the context of aggregate data, visual assessment remains a standard approach for evaluating the PH assumption – see Methods “Effect Size Measures” and Limitations “Calculation of Effect Size” in the revised manuscript.

R1.4: On line 434-435 the authors state that meta-regression was applied if at least three studies were available. This is confusing because they also acknowledge that there should be at least 10 studies per variable in a meta-regression. They also conduct other analyses (MetaForest), which are not vulnerable to this problem. Please clarify.

Reply: While the absolute minimum required to conduct a meta-regression is three studies, the Cochrane Handbook recommends including at least ten studies per covariate to reduce the risk of overfitting and enhance the reliability of the conclusions. We acknowledge that with only three studies per covariate, the results of our meta-regression analyses should be interpreted with caution. Nonetheless, we proceeded with these analyses, fully aware of the limitations imposed by the small number of studies. To promote transparency, the number of studies included in each analysis is clearly indicated in the moderator bubble plots, allowing readers to assess for themselves the strength and credibility of the presented evidence.

Additionally, we employed a MetaForest analysis, which is less susceptible to the limitations associated with meta-regression, such as the need for a large number of studies per variable. This analyses provides complementary insights and helps mitigate potential biases arising from the limited number of studies in the (simple) meta-regressions.

We have revised the manuscript to clarify these points, ensuring that readers are aware of the potential limitations and the rationale behind our analytical approach – see Methods “Effect-Moderator Analysis” and Limitations “Moderators” in the revised manuscript.

R1.5: Table 3: “PICTOS” should be “PICOTS”.

Reply: Agreed, spelling mistake on our part – see Table 3 in the revised manuscript.

R1.6: Line 999 (and several other places): The Stagl et al study had 15 dead and 105 not dead in both groups. How can this produce a hazard ratio of 0.21? Line 1009: Why is the reference line not at 1? How can this produce a hazard ratio of 0.21?

Reply: The discrepancy arises from the different methodologies used to calculate hazard ratios (HR) vs. odds (OR) or risk ratios (RR). The HR assesses the rate at which events occur over time, incorporating both the occurrence and timing of events. In contrast, the OR and RR simply compare the proportion of individuals experiencing an event in each group, focusing solely on the occurrence without considering timing.

Example Calculation:

Consider a study with two groups, each comprising 120 participants. In both groups, 15 participants experience the event (e.g., death), resulting in an odds ratio (OR) or risk ratio (RR) of 1.0, indicating no difference in overall event occurrence between the groups. However, if the timing of events differs – such that events occur earlier in the control group and later in the intervention group – the hazard ratio (HR) would capture this temporal difference. In this case, the HR could be less than 1.0, suggesting a reduced risk over time in the intervention group. This reflects the fact that, although the total number of events is equal, the delayed occurrence in the intervention group indicates a time-to-event benefit.

This illustrates that while ORs and RRs provide a straightforward comparison of event proportions, HRs offer a more nuanced understanding about survival effects by accounting for the timing of events, which can lead to very different interpretations. Thus, the HR remains the preferred method for survival analysis, as it accounts for both the occurrence AND timing of events, offering a more precise analysis and nuanced conclusions.

Why is the reference line not at 1?

Reply: That was a mistake on our part. We have corrected it – see Figure 5 in the revised manuscript.

R1.7: On lines 111-112, the authors state that the first meta-analysis on this topic was conducted almost two decades ago 112 (Chow et al, 2004). Please use «published» instead of «conducted» (The Smedslund meta-analysis was started early in 2001, was submitted in November 2002, accepted in September 2003, and published in August 2004, 7 months after the Chow paper.

Reply: We have exchanged the word “conducted” with “published” at two places in the manuscript – see Introduction “Number and Selection of Included Trials” and “Data-Analytic Decisions” in the revised manuscript.

If the authors take these comments and questions into account, I approve of the manuscript for publication in Communications Psychology.

Thank you!

Reviewer #2 (Remarks to the Author):

General impression: The systematic review addresses a significant public health issue: cancer survival, specifically focusing on the efficacy of psychosocial interventions in prolonging the survival of cancer patients. This is a highly controversial and unresolved topic. The results of existing studies are mixed, with nearly half of the randomized controlled trials (RCTs) indicating a life-extending effect, while the other half do not. This inconsistency is also reflected in prior meta-analyses, where some report significant effects while others do not. Tackling this issue is ambitious and commendable. The paper addresses a complex issue using highly heterogeneous data and relatively complex methodologies.

In addition to my specific comments (see below), I have a more general concern. The authors conclude that “psychosocial interventions seem to enhance cancer survival and should therefore be made accessible to all cancer patients.” I question whether this conclusion is sufficiently supported by the data. While there are many reasons to offer psychosocial interventions to cancer patients, I worry that the concluding statement may give the impression that the scientific debate has reached a coherent and final conclusion, which might seem somewhat overconfident. Key issues include the broad definitions of “psychosocial intervention” and “cancer patients,” as well as the considerable heterogeneity noted by the authors.

Overall, the manuscript has significant potential and addresses a critical and complex issue with a thoughtful approach. I hope my feedback is helpful in refining the conclusions and clarifying the strength of the evidence presented. While I may not be able to address every issue, I have tried to identify the main strengths and limitations.

Reply: Thank you very much, we fully acknowledge the controversial nature of the topic and agree that the considerable heterogeneity in both primary and secondary research warrants a cautious interpretation. While some journals encourage clear and decisive statements, we understand and value the importance of a nuanced framing, especially given the complexity and unresolved aspects of this field.

In response, we have revised our manuscript to better reflect the exploratory and evolving nature of evidence synthesis. Specifically, we have removed the term “final”, replaced the term “conclusion” with “consensus” (see title in the revised manuscript) and rephrased key text passages at the end of the Abstract, at the end of the Introduction section “Objectives of the Present Analysis,” and a part in the Discussion section, which has been renamed from “A Possible End to the Discussion by a Coherent Conclusion” to “A Unifying Consensus to a Controversial Topic.” These changes aim to convey a more balanced perspective while still emphasizing the relevance of our findings within the broader context of cancer care. Please also see our more elaborate reply to reviewer 4 (R4.1).

1. Strengths

1.1. The introduction provides an overview of the existing meta-analyses on this topic and their strengths and limitations and presents possible reasons for conflicting results of the 10 previous meta-analyses, including the differences between the included studies, e.g., in the chosen effect size. The important issue regarding the failure to use HR as the outcome measure in several studies – and meta-analyses is highlighted. The introduction also discusses the frequent failure to examine possible publication bias and p-hacking- and when doing so – using less optimal methods, including fail-safe numbers and the trim-and-fill procedure.

1.2. The authors conducted a very comprehensive state-of-the-art meta-analysis of existing RCTs. State-of-the-art methods include prediction intervals and bias assessment with several more recently developed methods, including contour-enhanced funnel plots, the FAT-PET-PEESE regression test, and p-uniform. In addition, they apply a novel approach to exploring the causes of heterogeneity by

identifying the WHICH and HOW specifications concerning the data analyses of previous meta-analyses.

2. Issues and limitations

R2.1: I realize that this is not the main topic, but the references on psychological risk factors for cancer development and progression seem somewhat dated. Please include meta-analyses and more recent reviews if possible, perhaps also including more critical results. For example, it is not quite clear whether “stress”, which in itself is a very heterogeneous term and operationalized in many different ways, is a risk factor in all cancers, genders, and age groups.

Reply: We have included some more modern meta-analyses to this point in the introduction – see Introduction “The Origins of a Controversial Debate” in the revised manuscript.

R2.2: The paper is highly complex. The introduction resembles a systematic overview of reviews. I would have preferred a two-tiered process, first conducting a proper systematic review of reviews, including the specification-curve plots and multiverse analysis, followed by an analysis of primary studies, guided by the results. The results could perhaps even be submitted as two separate papers.

Reply: We initially considered separating our analysis into two papers at the beginning of the project but concluded that integrating all parts into a single manuscript would provide a more cohesive and comprehensive examination of the factors contributing to the heterogeneity in existing primary (RCTs) and secondary (meta-analyses) research. This approach allowed us to trace a continuous chain of evidence, linking findings from primary studies directly to meta-analytic studies, thereby enhancing the logical flow and depth of our arguments. We believe that presenting our findings in this unified format offers readers a clearer understanding of the complexities involved and supports the development of a more unified consensus. We have now moved the entire appendix section (~100 pages) into the Supplementary Information to streamline the main manuscript. We respectfully hope that the editors of *Communications Psychology* will provide the space required to accommodate the scope and significance of this extensive project.

R2.3: The authors are commended for their attempts to calculate and convert various data into HRs. One wonders about the reliability of these data. Do the authors have any data on the reliability, e.g., by comparing independent raters.

Reply: The initial HR calculations were performed by one author (KDAH) and subsequently checked by another author (UST), who is an active researcher in the fields of quantitative methods and statistics. This process ensured that any discrepancies or errors could be identified and addressed promptly. We have added a note to the manuscript – see Methods “Effect Size Measures” in the revised manuscript.

R2.4: The authors assessed “study quality” with the Nat Heart, Lung, and Blood Institute scale. Risk of bias and study quality are not identical. Several study quality aspects may not infer risk of bias, and one wonders why the Cochrane risk of bias revised of bias (ROB-2) approach wasn’t used. Rather than a total study quality score, the ROB-2 analysis could provide a clearer picture of the risk-of-bias-related issues affecting the various studies.

Reply: Study quality (with the NHLBI Scale) is included as one (of several) of the risk of bias factors that we have investigated, alongside meta-analytic post-hoc power and risk of sponsorship bias. The study quality measure we applied enabled us to calculate a quality score for each study and subsequently examine its influence on effect sizes in the moderator analysis.

Following your suggestion, we have now supplemented our methodology by also applying the Cochrane Risk of Bias Tool 2 (RoB 2) to all primary studies. This tool offers a nuanced assessment by evaluating potential biases specific to randomized trials, including domains like the randomization process, deviations from intended interventions, missing outcome data, measurement of outcomes, and selection of reported results.

Two authors independently applied the RoB 2 tool to each study, ensuring a comprehensive evaluation of potential biases. This approach provided a clearer picture of the risk-of-bias-related issues affecting the studies included in our analysis – see Methods “Risk of Bias Assessment”, Results “Risk of Bias in Studies” and Figure B4 in the revised manuscript.

R2.5: I do not have the expertise to evaluate the machine-learning-based approach.

Reply: Alright, we have followed the closely described MetaForest approach for small-sample meta-analyses as outlined by van Lissa (2020). See Methods “Exploratory Machine-Learning-Based Search for Moderators”.

van Lissa, C. J. (2020). Small sample meta-analyses: Exploring heterogeneity using MetaForest. In R. van De Schoot & M. Miočević (Eds.), *Small sample size solutions: A guide for applied researchers and practitioners* (pp. 121–140). CRC Press. <https://www.crcpress.com/Small-Sample-Size-Solutions-Open-Access-A-Guide-for-Applied-Researchers/Schoot-Miocevic/p/book/9780367222222>

R2.6: There might some issues with publication bias assessment. The FAT-PET-PEESE approach appears to be challenged by small sample sizes and heterogeneity.

Reply: In our meta-analysis, we have found a small overall effect favoring the intervention group $\log HR = -0.23$, 95% CI $[-0.35, -0.11]$, $p < .001$, which translates into $HR = 0.80$, 95% CI $[0.71, 0.90]$. The Q test indicated significant heterogeneity ($Q(31) = 59.27$, $p < .001$), and the relative ($I^2 = 48\%$) and absolute ($\tau^2 = 0.05$, $\tau = 0.23$; log scale) proportion of true variance was moderate to substantial, respectively. The 95% prediction interval ranged from $HR 0.49$ to 1.29 .

Regarding publication bias, the funnel asymmetry test (FAT), conducted as part of the PET-PEESE procedure, yielded a non-significant result ($p = .091$) but indicated a trend toward asymmetry. We fully acknowledge that, given the moderate to substantial heterogeneity and the relatively small sample sizes in some studies, the FAT-PET-PEESE estimates should be interpreted with caution. Funnel plot asymmetry in this context could be due to genuine heterogeneity in effect sizes, outcome reporting bias, or methodological differences and not necessarily publication bias alone (Ioannidis & Trikalinos, 2007; Stanley, 2017).

We have now clarified this limitation more explicitly in the revised manuscript – see Discussion: “Reporting Biases” and Limitations: “Reporting Biases”.

Ioannidis, J. P. A., & Trikalinos, T. A. (2007). The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. *Canadian Medical Association Journal*, 176(8), 1091–1096. <https://doi.org/10.1503/cmaj.060410>

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, 5(1), 60–78. <https://doi.org/10.1002/jrsm.1095>

R2.7: Visualizing results: Figure 1A-B is a bit difficult to follow as it includes both primary studies and meta-analyses in the same figure. Also, the reader tries to interpret why some studies are above and some below the timeline. I wonder whether the information could be presented in a more clear and easier-to-understand ways. Figure 2 contains some of same information.

Reply: We have revised Figures 1A & B to enhance clarity: meta-analyses are now distinctly marked with “MA,” and we have clarified the meaning of studies positioned above and below the timeline with additional lines of text. Moreover, the size of the descriptions in the right corner have been increased to improve readability – see Figure 1A & B in the revised manuscript.

Regarding the comparison between Figures 1A & B to Figure 2, we believe they serve distinct purposes. Figures 1A & B provide a chronological overview of published randomized controlled trials (RCTs) and meta-analyses regarding the research question over the past four decades, illustrating the occurrence frequency and timing of significant versus non-significant findings, which enables the reader (e.g., researchers, practitioners and patients) to get a quick overview and “feel” for the ongoing controversy in the field. In contrast, Figure 2 presents an overlap matrix of included primary studies in prior meta-analyses, including our own, utilizing the GROOVE tool's matrix of evidence. This figure

aims to depict the interrelationships and possible redundancies with respect to the included primary studies (RCTs) across different meta-analyses.

We hope that these revisions and clarifications address your concerns and enhance the comprehensibility of our visual data presentations.

The note in Figure 5 states that the reference line is drawn at HR=1. It obviously isn't – it seems to be drawn at HR=0.78.

Reply: That was a mistake on our part. We have corrected it – see Figure 5 in the revised manuscript.

R2.8: I am a bit worried about the p-curve analysis and wonder whether the small number of studies allows a sufficiently robust analysis. Also, in this controversial area, could one imagine p-hacking in the opposite direction, i.e., in the direction of null-findings?

Reply: We acknowledge that the number of significant studies ($k = 10$) included in our analysis is at the lower threshold for conducting a robust p-curve assessment. The reliability and robustness of its conclusions naturally increase with a larger number of studies.

In our case, the p-curve did not indicate signs of p-hacking. The test for evidential value was significant for the half-curve ($p = .029$), suggesting the presence of true effects in the literature. However, we also observed indications of low statistical power, which could affect replicability. We have now explicitly stated in both the Discussion and Limitations that our findings are based on a limited number of significant studies and should be interpreted with appropriate caution – see Discussion “Reporting Biases” and Limitations “Reporting Biases” in the revised manuscript.

Regarding the possibility of p-hacking in the direction of null results, we acknowledge that this is an interesting theoretical consideration. However, we believe the actual incentive for such practices is limited in this field. Unlike areas with strong commercial or regulatory interests, monetary or industry-related pressures are not prominent in this research domain. While personal or theoretical biases may influence the interpretation of results to some extent, we consider it rather unlikely that researchers would actively manipulate data or analyses to produce non-significant findings in a field aimed at improving cancer patient survival through psychosocial interventions. To date, we are also not aware of a method like p-curve that is able to test p-hacking into the direction of no effect.

R2.9: The overall median survival benefit was 3.8 months. While it does appear meaningful, how does it compare to, e.g., the benefits reported for new medical treatments? Some more detailed examples could be mentioned here. Also, I remain less convinced by the authors' rather strong concluding statements and recommendations given the considerable heterogeneity and very broad prediction interval.

Reply: Hazard ratios (HRs), odds ratios (ORs), and risk ratios (RRs) are the most frequently reported effect sizes in cancer survival research. However, these effect sizes are not always intuitively meaningful to clinicians and patients. To address this, we have additionally calculated an overall median survival difference (see Introduction “An Effect Size for Practitioners and Patients” and Discussion “An Easy-to-Grasp Effect Size for Survival”) using the recently developed R package *metamedian* (McGrath et al., 2018; 2019).

While the estimated median survival benefit of approximately four months is easier to grasp and communicate, it is not the ideal metric for comparing the results across meta-analyses in this field, as – to our knowledge – no other meta-analysis focusing on psychosocial interventions, chemo-, radio- or hormone therapy for cancer patients has reported results in this format.

Instead, a more appropriate comparison can be drawn with our observed **small overall HR of 0.80**, 95% CI [0.71–0.90] for cancer patients receiving psychosocial interventions. It represents a small but clinically meaningful improvement in survival (more on that below) – remarkably similar to the effect

of marriage on survival in cancer patients **HR = 0.80**, 95% CI [0.79–0.81], which was observed across ten different cancer types in a large-scale, five-year observational study (Aizer et al., 2013).

When contextualizing our finding within the broader oncology literature, the survival benefit associated with psychosocial interventions is actually comparable in magnitude to that of several established cancer treatments such as chemo-, radio- and hormone therapy:

- A recent meta-analysis on chemotherapy for pancreatic cancer found a similarly **small (to moderate) overall effect HR 0.71** 95% CI [0.62–0.82] (Evans et al., 2024).
- Another recent meta-analysis on metastatic breast cancer patients identified a **small overall effect HR 0.87**, 95% CI [0.78–0.97] (W. Ren et al., 2022).
- A meta-analysis on radiotherapy in breast cancer did not find a significant **overall effect RR 0.93** 95% CI [0.79–1.09] (Garg et al., 2017).
- A meta-analysis of radiotherapy for gallbladder cancer reported a **small overall effect OR of 0.63** 95% CI [0.50–0.81] (B. Ren et al., 2020), though further analyses suggests that this benefit may have been primarily driven by patients with lymph node-positive or margin-positive disease – who typically account for approximately 30% of cancer cases.
- A meta-analysis on hormone therapy in breast cancer yielded a **small overall effect HR of 0.79** (95% CI [0.69–0.90]; Yu et al., 2017), closely mirroring the effect observed in our analysis.

According to Azuero (2016), **HR thresholds for small, medium, and large effects are 0.77, 0.53, and 0.36** (inverse of HRs 1.3, 1.9, and 2.8, respectively), whereas according to Chen et al. (2010), **ORs thresholds are 0.60, 0.29, and 0.15** (inverse of OR 1.68, 3.47, and 6.71, respectively). We could not find such a categorization for RR.

It is important to note that our meta-analysis combines studies across several types of psychosocial interventions and encompasses various cancer types and stages – primarily due to the limited availability of primary studies – which may also explain the considerable heterogeneity and wide prediction interval. Other meta-analyses cited above focused on distinct medical treatments (probably with similar dosages) and specific cancer types (sometimes even similar stages) benefiting from a higher number of primary studies available in general regarding these treatments, thereby – at least in theory – allowing for more homogeneity in these meta-analyses. However, 4 out of 5 of these meta-analyses showed 95% prediction intervals (PI) similar to the one found in our meta-analysis (Evans et al., 2024 a 95% PI [HR 0.58 to 0.87]; W. Ren et al., 2022 a 95% PI [HR 0.66 to 1.15], Garg et al., 2017 a 95% PI [RR 0.72 to 1.21], B. Ren et al., 2020 a 95% PI [OR 0.26 to 1.54]; Yu et al., 2017 a 95% PI [HR 0.49 to 1.28]).

It is worth noting that the 95% prediction interval for the meta-analyses on chemo-, radio-, and hormone therapy often had to be calculated by ourselves, as these meta-analyses frequently did not report them directly. This omission may have contributed to the apparent unawareness that the level of certainty of the evidence regarding these treatments – despite their high cost in terms of adverse side effects for patients – is not as high as one might hope.

Nevertheless, chemo-, radio-, and hormone therapy are frequently administered to cancer patients without being required to demonstrate low threshold 95% prediction intervals despite their known side-effects such as anemia, alopecia, anxiety, cognitive dysfunction, fatigue, insomnia, immunosuppression, nausea and vomiting, low mood, mucositis, pain, and weight loss (De Ruyscher et al., 2019; Fleming et al., 2022; Nurgali et al., 2018).

In light of these considerations, we find our results not only meaningful but also clinically relevant, as psychosocial interventions appear to yield survival benefits that are comparable in magnitude to those of established medical treatments—with a similar level of certainty into the evidence—yet without any of the debilitating side effects noted above.

Moreover, unlike any pharmacological or invasive interventions, psychosocial approaches simultaneously enhance a wide range of beneficial psychological outcomes. According to several meta-analyses, psychosocial interventions also improve well-being (McLouth et al., 2020), quality of life (de la Torre-Luque et al., 2015), anxiety (Sanjida et al., 2018), depression (Hart et al., 2012), stress (C. Zhao et al., 2021), fatigue (Xunlin et al., 2019), and pain (Warth et al., 2020) in cancer patients.

Taken together, these findings demonstrate that psychosocial interventions provide a broad, multidimensional spectrum of proven benefits – including a modest and robust improvement in survival – making them not only a valuable but arguably an essential component of modern, integrative cancer care. Importantly, none of the included trials showed any harm.

Of course, further research on survival is warranted to better understand which types of psychosocial interventions work best for specific cancer types, stages, and patient characteristics, in order to optimize outcomes and personalize care.

We have added these contextualizations above into the manuscript – see Discussion ““Comparable Certainty in the Evidence to Medical Cancer Treatments” and “Comparable Survival Benefits to Medical Cancer Treatments” in the revised manuscript. Please also see our reply to reviewer 4 (R4.1).

Aizer, A. A., Chen, M.-H., McCarthy, E. P., Mendu, M. L., Koo, S., Wilhite, T. J., Graham, P. L., Choueiri, T. K., Hoffman, K. E., Martin, N. E., Hu, J. C., & Nguyen, P. L. (2013). Marital Status and Survival in Patients With Cancer. *Journal of Clinical Oncology*, 31(31), 3869–3876. <https://doi.org/10.1200/jco.2013.49.6489>

Chen, H., Cohen, P., & Chen, S. (2010). How Big is a Big Odds Ratio? Interpreting the Magnitudes of Odds Ratios in Epidemiological Studies. *Communications in Statistics - Simulation and Computation*, 39(4), 860–864. <https://doi.org/10.1080/03610911003650383>

de la Torre-Luque, A., Gambara, H., López, E., & Cruzado, J. A. (2016). Psychological treatments to improve quality of life in cancer contexts: A meta-analysis. *International Journal of Clinical and Health Psychology*, 16(2), 211–219. <https://doi.org/10.1016/j.ijchp.2015.07.005>

De Ruysscher, D., Niedermann, G., Burnet, N. G., Siva, S., Lee, A. W. M., & Hegi-Johnson, F. (2019). Radiotherapy toxicity. *Nature Reviews Disease Primers*, 5(1). <https://doi.org/10.1038/s41572-019-0064-5>

Evans, D., Ghassemi, N., Hajibandeh, S., Hajibandeh, S., Romman, S., Laing, R. W., Durkin, D., & Athwal, T. S. (2024). Meta-analysis of adjuvant chemotherapy versus no adjuvant chemotherapy for resected stage I pancreatic cancer. *Surgery*, 175(6), 1470–1479. <https://doi.org/10.1016/j.surg.2023.11.027>

Fleming, L., Agnew, S., Peddie, N., Crawford, M., Dixon, D., & MacPherson, I. (2022). The impact of medication side effects on adherence and persistence to hormone therapy in breast cancer survivors: A quantitative systematic review. *The Breast*, 64, 63–84. <https://doi.org/10.1016/j.breast.2022.04.010>

Garg, P. K., Jakhetiya, A., Pandey, R., Chishi, N., & Pandey, D. (2017). Adjuvant radiotherapy versus observation following lumpectomy in ductal carcinoma in-situ: A meta-analysis of randomized controlled trials. *The Breast Journal*, 24(3), 233–239. <https://doi.org/10.1111/tbj.12889>

Hart, S. L., Hoyt, M. A., Diefenbach, M., Anderson, D. R., Kilbourn, K. M., Craft, L. L., Steel, J. L., Cuijpers, P., Mohr, D. C., Berendsen, M., Spring, B., & Stanton, A. L. (2012). Meta-analysis of efficacy of interventions for elevated depressive symptoms in adults diagnosed with cancer. *JNCI: Journal of the National Cancer Institute*, 104(13), 990–1004. <https://doi.org/10.1093/jnci/djs256>

McLouth, L. E., Ford, C. G., Pustejovsky, J. E., Park, C. L., Sherman, A. C., Trevino, K., & Salsman, J. M. (2020). A systematic review and meta-analysis of effects of psychosocial interventions on spiritual well-being in adults with cancer. *Psycho-Oncology*, 30(2), 147–158. <https://doi.org/10.1002/pon.5562>

Nurgali, K., Jagoe, R. T., & Abalo, R. (2018). Editorial: Adverse effects of cancer chemotherapy: Anything new to improve tolerance and reduce sequelae? *Frontiers in Pharmacology*, 9(245). <https://doi.org/10.3389/fphar.2018.00245>

Ren, B., Guo, Q., Yang, Y., Liu, L., Wei, S., Chen, W., & Tian, Y. (2020). A meta-analysis of the efficacy of postoperative adjuvant radiotherapy versus no radiotherapy for extrahepatic cholangiocarcinoma and gallbladder carcinoma. *Radiation Oncology*, 15(1). <https://doi.org/10.1186/s13014-020-1459-x>

Ren, W., Yu, Y., Hong, H., Wang, Y., Gao, Q., Chen, Y., Chen, P., Zhao, J., Ou, Q., Lin, D., Fu, T., Tan, Y., Li, C., Xie, X., Ye, G., Tang, J., & Yao, H. (2022). Clinical evidence of chemotherapy or endocrine therapy maintenance in patients with metastatic breast cancer: Meta-analysis of randomized clinical trials and propensity score matching of multicenter cohort study. *Cancer Research and Treatment*, 54(4), 1038–1052. <https://doi.org/10.4143/crt.2021.698>

Sanjida, S., McPhail, S. M., Shaw, J., Couper, J., Kissane, D., Price, M. A., & Janda, M. (2018). Are psychological interventions effective on anxiety in cancer patients? A systematic review and meta-analyses. *Psycho-Oncology*, 27(9), 2063–2076. <https://doi.org/10.1002/pon.4794>

Warth, M., Zöller, J., Köhler, F., Aguilar-Raab, C., Kessler, J., & Ditzen, B. (2020). Psychosocial interventions for pain management in advanced cancer patients: A systematic review and meta-analysis. *Current Oncology Reports*, 22(1). <https://doi.org/10.1007/s11912-020-0870-7>

Xunlin, N., Lau, Y., & Klainin-Yobas, P. (2019). The effectiveness of mindfulness-based interventions among cancer patients and survivors: A systematic review and meta-analysis. *Supportive Care in Cancer*, 28(4), 1563–1578. <https://doi.org/10.1007/s00520-019-05219-9>

Yu, X., Zhou, S., Wang, J., Zhang, Q., Hou, J., Zhu, L., He, Y., Zhao, J., & Zhong, S. (2017). Hormone replacement therapy and breast cancer survival: a systematic review and meta-analysis of observational studies. *Breast Cancer*, 24(5), 643–657. <https://doi.org/10.1007/s12282-017-0789-5>

Zhao, C., Lai, L., Zhang, L., Cai, Z., Ren, Z., Shi, C., Luo, W., & Yan, Y. (2021). The effects of acceptance and commitment therapy on the psychological and physical outcomes among cancer patients: A meta-analysis with trial sequential analysis. *Journal of Psychosomatic Research*, 140, 110304. <https://doi.org/10.1016/j.jpsychores.2020.110304>

R2.10: The conventional frequentist approach has limitations which could be overcome using Bayesian approaches. Unlike frequentist methods, which provide the probabilities (p-values), given a hypothesis, Bayesian methods enable direct probability statements about the hypotheses themselves. Furthermore, unlike p-values, which only indicate the degree of “surprisingness” of the data given that the null hypothesis is true, Bayes factors quantify the support for one model compared to another, and Bayesian methods avoid other issues associated with null-hypothesis significance testing, such as the over-reliance on relatively arbitrary p-value thresholds and the dichotomization of results into “significant” and “non-significant.” Such an analysis would help the reader assess the robustness of the results.

Reply: We fully agree that Bayesian methods offer several conceptual and interpretive advantages over traditional null hypothesis significance testing, particularly in avoiding arbitrary thresholds and enabling direct probability statements about hypotheses. However, our goal in this study was to provide a comprehensive and methodologically rigorous frequentist analysis – and we believe that we have already gone considerably beyond the standard procedures commonly applied in this research domain. Specifically, we conducted a wide array of robustness checks, including sensitivity analyses, cumulative meta-analysis, the assessment of several risk of bias factors (RoB 2, study quality, meta-analytic post-hoc power, sponsorship bias), evaluations of publication bias and p-hacking, as well as a full multiverse meta-analysis. The latter, in particular, offers a powerful and transparent way to explore how analytic flexibility may impact results, and can be viewed as a complementary and equally robust alternative to Bayesian model comparisons.

Given the already substantial methodological scope of this study, we did not deem it feasible to include a time-intensive Bayesian re-analysis at this stage. We agree, however, that future research may benefit from adding a Bayesian framework to build on the current findings.

R2.11: Conclusions regarding the clinical value need to be supported by an overall evaluation of the findings, e.g., with the GRADE assessment or similar. Clinicians and future attempts at clinical practice guidelines would benefit from this.

Reply: We have now incorporated the GRADE (Grading of Recommendations, Assessment, Development, and Evaluations) approach to assess the certainty of the evidence in our meta-analysis – see Methods “Certainty of Evidence (GRADE Rating)”, Results “Certainty of Evidence (GRADE rating)” and Discussion “The Overall Effect” in the revised manuscript.

Thank you for your valuable suggestions!

Reviewer #3 (Remarks to the Author):

General impression:

I do not habitually start my reviews with: “wow”, but here I do.

The first 'wow' is sheer length; it is my first review of a manuscript that is over 100 pages long (among 500 + lifetime reviews)! That of course triggered consideration of "wow, nr 2": 'is there really so much important information here that this length is needed?' Overall, I will say 'yes' there is (detail to follow of course). The topic is of utmost importance to the field of Psycho-Oncology, is enormously complex, and has produced highly variable and confusing conclusions. The authors do this complexity justice with a thorough literature review and the building of a convincing rationale for this in-depth and nuanced review of outcomes and of moderators and mediators.

The issues around type of statistic used, bias control, significance fishing etc. that are raised and thoroughly investigated here is impressive and does indeed lead to trustworthy conclusions. Given the many different types of cancer and the various stages when patients were entered into trials, results are bound to be variable and hard to replicate. These authors are very aware of these challenges.

Please note that there were occasions where these authors' handling of statistics exceeded my skill level even though I have published a half dozen meta-analyses myself.

Potential sources of heterogeneity are thoroughly investigated here and it comes as no surprise that: having access to more RCTs (in 2024) increases power to find small effects, that HR is more precise tool than OR and RR, and that not all cancer types and stages are associated with same sized effect.

The manuscript is very clearly written, the objective are spelled out and the rationale-building is sound.

Reply: Thank you very much for your generous and thoughtful feedback. We are truly honored by your recognition of the depth and scope of our work. This project reflects seven years of continuous and dedicated effort to synthesize a highly complex and heterogeneous body of literature spanning more than four decades. We are especially grateful that you acknowledged the statistical depth of the manuscript and the challenges posed by the diversity of cancer types, stages, and interventions. Your recognition of the clarity of writing, the soundness of our rationale-building, and the rigor of our methodology is greatly appreciated. We sincerely hope that the editors of *Communications Psychology* will also recognize the exceptional nature of this comprehensive review and grant us some more flexibility than usual regarding space constraints. Condensing such a multifaceted analysis into a shorter format would risk undermining the nuance and transparency we have carefully strived to maintain.

Arguments:

R3.1: As for my more critical comments, I will start with some personal opinion and context. When the first studies on this topic were planned and completed (Spiegel (a co-author of this paper), Fawzy, Cunningham, and others, all found in the tables of this Ms), there was major headscratching about how we can answer the core question with a scientifically sound approach that is still ethically acceptable. We cannot get a sound answer about 'does psychotherapy extend life' without tightly controlled RCTs where patients are randomly assigned to a wait-list, or a no treatment, or a minimal control condition. Then we have to follow all groups for many years and avoid confounding treatment; the ethics of this approach is hugely problematic. Surprisingly, the authors do not touch very much on this tricky issue.

Reply: We fully agree that the core question of whether psychosocial interventions extend life is difficult to answer with maximal scientific rigor due to serious ethical constraints. As you are rightly pointing out, conducting long-term RCTs with passive or no-treatment control groups – which might provide the most methodologically sound evidence – would be ethically questionable given the well-established benefits of psychosocial support for emotional well-being. Withholding such support from vulnerable cancer patients for the sake of methodological purity would raise significant concerns.

We have addressed this issue, albeit briefly, by including control group type (passive vs. active) as a potential moderator in our analysis. We now also discuss the broader ethical limitations of such designs explicitly in the revised Limitations section of the manuscript – see limitations “Ethical Constraints”.

While we acknowledge that this issue cannot be fully resolved within the scope of our study, we hope that highlighting this tension between methodological rigor and ethical responsibility will encourage further reflections in the field.

There should be no surprise that initial studies were not perfectly tight because they also had to get approved by the ethics committees first. Critics of methodology like Ioannidis were justified in their challenges but maybe not have been overly sympathetic to the plight of researchers trying to strike the right balance of science and ethics.

Then things got really confused by Grossarth-Maticek and Eysenck who published research on psychological treatment of cardiac and cancer patients based on their personality typologies that sounded fabulous but were ultimately unbelievable. Grossarth-Maticek’s effect sizes were huge and are now considered plain fraudulent (Pelosi, 2019). Unless previous meta-analyses had explicitly bracketed out the Grossarth-Maticek findings, they were inevitably over-representing the benefits of psychotherapy and should not be used for drawing conclusion. Now, two Grossarth-Maticek references are found in the reference list of this manuscript but I cannot find any mention of this work in the manuscript text or in the tables. Is that an oversight by the authors, intention but just unexplained, or an oversight on my part?

I have to admit that I am uncertain about how much of this colorful history of the field has to be described here but... it does help understanding how we got to the status quo which is rightfully and capably challenged by the researchers in this review; it has certainly colored the glasses I was wearing as a (now retired) reviewer.

Reply: The Grossarth-Maticek studies were identified during our systematic search and were carefully reviewed. However, they were explicitly excluded from our meta-analysis due to not being randomized and serious methodological as well as credibility concerns (see Table A8). To support transparency and allow future researchers to build on our work, we had initially listed all reviewed (including initially retrieved) but later excluded studies in the main reference list.

We now recognize that including excluded studies in the reference section – without direct mentioning in the manuscript except for in a table (Table A8) – may be confusing to readers. In response, we have removed these citations from the references (including the Grossarth-Maticek articles) and included them into the Supplementary Information (below Table A8) listing all the references of the excluded articles. We hope this resolves any ambiguity and demonstrates our critical stance regarding the inclusion of studies with questionable scientific integrity.

R3.2: A second point that needs add’l attention (and will add to the length of the paper!!, so sorry ...) is that of floor effects. Psycho-oncology and Psychocardiology often do tackle psychotherapy offerings differently than psychiatry. In psychiatry, treatment of depression is offered to people who were diagnosed as depressed. But, in behavioral medicine treatment is often provided for a medical condition that may be linked with elevated rates of anxiety or depression for some, but not all patients. That means that patients without depression will be treated for depression and the possible improvements are nil or miniscule; effects get watered down. How often does this happen? In a 2008 review (Jacobson & Jim), it was revealed that 95% of all clinical trials in psycho-oncology had not first screened for distress at treatment entry, thus allowing the majority of patients to enter a trial for distress reduction although they were not overly distressed a priori. That should be seen in light of another review (Linden & Satin, 2007) that had reported on floor effects in psycho-oncological treatments and revealed that the correlation of distress scores at treatment entry with the degree of

post-treatment change was $r=0.75$, thus explaining 57% of the variance in treatment outcome. We know that resulting floor effects suppress possibly good outcomes; Linden and Girgis (2012) have shown that patients with cancer and screened for actually diagnosed depression/anxiety showed three time greater benefits on these endpoints than when all patients with cancer are treated for anxiety/depression. What we don't know but need to see addressed here is whether floor effects also suppress findings of greater longevity when patients with actually documented depression/anxiety had been recruited and treated.

Reply: The issue of floor effects is indeed a critical consideration when interpreting the impact of psychosocial interventions on psychological outcomes such as anxiety, depression and distress. As pointed out by you, many trials in the field of psycho-oncology did not screen for baseline distress (Jacobsen & Jim, 2008; Linden & Satin, 2007), which can attenuate observable improvements in those domains. However, the primary objective of our meta-analysis – as well as of the included RCTs – was to investigate the potential survival-enhancing effects of psychosocial interventions in general cancer populations. These samples typically consisted of “psychologically healthy” patients facing the psychological burden of a cancer diagnosis and its treatment but not selected based on elevated distress or psychiatric diagnoses. In fact, most included trials even explicitly excluded participants with current mental disorders.

We acknowledge your important point, and agree that studies specifically targeting clinically anxious, depressed or distressed cancer patients may yield even larger effects on both psychological outcomes and-potentially-even survival. We therefore perceive this to be a valuable direction for future research.

R3.3: In my review above, I referred to the difficulty of generating a control group that is scientifically sound while ethical. The authors inform the reader that the inclusion criteria for this review include a design that has a psychologically active treatment and either a placebo or passive controls (no intervention). Being a clinical psychologist and scientist-practitioner, I will argue that there is no such thing as a true placebo control in psychotherapy; any treatment, even minimal, might lead to positive change. When that happens, effect size differences will shrink and results become more conservative. Missing here is coverage of placebo was operationalized. And how is avoidance of treatment confounding implemented during the post-treatment observation period of many months and years?

Reply: We agree that there is no “true” placebo control in psychological research, but for our review, we used the standard terminology. Control groups were categorized as *active* (e.g., minimal psychosocial input such as health education, relaxation training, or guidance materials) or *passive* (treatment as usual without any structured psychosocial input). We have added a brief definition of active and passive controls to the manuscript – see Methods “Data Items” in the revised manuscript.

These active controls were designed to resemble a “placebo” in the sense that they provided some structure and attention, but not the specific therapeutic ingredients of the intervention (e.g., reminiscence therapy, CBT-based techniques, etc.). We agree that such designs likely reduce observed between-group differences and may lead to a conservative estimate of the intervention effect. Some trials even offered these passive treatment to both the intervention and the control group.

Regarding the question of treatment confounding during the follow-up phase: only five studies (Cunningham et al., 1998; Goodwin et al., 2001; Edelman et al., 1999; Spiegel et al., 2007; Ye et al., 2017) reported information on the use of psychological services after the intervention. In these cases, psychological care outside the study was recorded at follow-up and groups were subsequently checked for significant group differences. This means that in the other 27 studies, data on this issue are lacking. This limitation suggests that our reported effects may be conservative, as later additional care (especially for the control group, who would have the highest need for it) would likely dilute any

initial group differences. We have added this topic to the limitations – see Limitations “Lack of Control for Treatment Confounding”.

We have also added recommendations in this regard – see Suggestions “Suggestions for future Primary Research” in the revised manuscript.

R3.4: minor suggestion: In the Abstract, tell the reader what an HR of .78 for longer life means in terms of months gained.

Reply: We agree that it could be helpful for readers to understand the practical implications of the hazard ratio (HR). However, the HR reflects the relative risk of the event occurring at any given point in time over the entire observation period for a specific sample or comparison, rather than representing the absolute time until the event occurs. Therefore, it is not possible to directly translate the HR into time units (such as months). Therefore, to provide additional context, we have calculated an overall median survival difference – a novel measure in this field – of approximately four months, which we believe offers readers a clearer understanding of the actual effect of the intervention e.g. see Introduction “An Effect Size for Practitioners and Patients”.

R3.5: The Ms length is overwhelming; 103 pages long. Mind the actual text is only 36 pages and is not wordy. I habitually push for brevity (as do editors who control precious journal pages) but these authors deserve a looong leash; I see no obvious place for text cutting. The editors may have some ideas about tables and figures that could possibly go into a supplement.

Reply: Thank you so much for your understanding and supportive comment regarding the length of our manuscript. We deeply appreciate your recognition of the scope and depth of our work.

We have made multiple efforts to condense the manuscript without compromising clarity or completeness. However, given the complexity of our analyses and in line with current best-practice guidelines for systematic reviews and meta-analyses (e.g., APA 7, PRISMA 2020), we believe it is essential to provide a comprehensive introduction, a detailed account of our methods, and a thorough discussion of the findings. In order to streamline the revised manuscript further, we have now moved the entire Appendix into the Supplementary Information.

In fact, all of the reviewers’ comments as a whole suggest that our initial explanations may not have been sufficiently detailed, prompting us to elaborate further in several areas. This process of revision helped us identify and address remaining blind spots, reinforcing the importance of maintaining a certain level of detail.

We have done our best to strike a careful balance between necessary comprehensiveness and readability. That said, this project is indeed vast in scale: our aim is to bring clarity to a highly controversial field through an intensive analysis spanning four decades of research.

We sincerely hope that the editors of *Communications Psychology* will recognize the comprehensive nature of our contribution to the field of psycho-oncology – with the potential to initiate a paradigm shift – and consider allowing sufficient space for this extensive, yet necessary, synthesis, especially in light of the longstanding controversy and importance of the topic.

Thank you again for your thoughtful and encouraging feedback!

References

Jacobsen PB, Jim HS (2008). Psychosocial Interventions for anxiety and depression in adult cancer patients: achievements and challenges. *CA Cancer J Clin* 2008; 58:214–230.

Linden W, Girgis A (2012). Psychological treatment outcomes for cancer patients: What do meta-analyses tell us? *Psycho-Oncology*. 21, 343-350. DOI: 10.1002/po)

Linden W, Satin JR (2007) Avoidable pitfalls in behavioural medicine outcome research. *Annals of Behavioral Medicine*, 33, 143-147

Pelosi AJ (2019). Personality and fatal diseases: revisiting a scientific scandal. *Journal of Health Psychology*, 24 (4), 421-439.

Wolfgang Linden

Reviewer #4 (Remarks to the Author):

General impression:

In this paper, the authors report on a meta-analysis evaluating the effect of psychosocial interventions on the survival time of patients with cancer.

In general, the study is done in a meticulous and systematic way. I like the insightful reflections on possible reasons why prior meta-analyses came to conflicting results; these reflections also give a clear rationale and guideline for the current study. The authors start their meta-analysis with a thorough search and selection procedure for relevant literature on the topic. Data extraction seems to be accurate, and analyses well thought-out. I appreciate the use of various techniques (for instance to explore possible publication bias) as well as the use of modern and innovative techniques (mainly the use of machine-learning techniques to evaluate moderator effects and the use of multiverse analysis approach to evaluate the dependence of the meta-analytic results on decisions taken).

Overall, I am very positive about this study. Yet, I still have several (mainly minor) suggestions and comments:

[Reply: Thank you very much for your kind words.](#)

Arguments:

R4.1: First, the authors suggest that with their study, we have a final answer to the question what the effect is of psychosocial interventions (e.g., line 5: “A coherent conclusion to a controversial topic”, line 50 “... may now have arrived at a coherent and probably final conclusion”..., ...). I agree that several of the issues in previous research are avoided. Yet, I do not believe that this study does give a final answer. It was found that there was considerable effect heterogeneity, resulting in a large prediction interval.

[Reply: We fully acknowledge the controversial nature of this topic and agree that the moderate to substantial heterogeneity observed in our analysis – reflected in the wide prediction interval – necessitates a cautious interpretation. While our aim was to bring clarity to a fragmented body of research, we recognize that our findings do not represent a definitive or final answer. As is often the case with meta-analyses, conclusions must be interpreted within the limits of the available evidence and context.](#)

[In response to your and Reviewer 2’s comments \(see our elaborate answer to R2.9\), we have revised the manuscript to better reflect the evolving and exploratory nature of our findings. Specifically, we have removed the term “final” and rephrased key statements at the end of the Abstract, the Introduction “Objectives of the Present Analysis” and rephrased a title in the Discussion from “A Possible End to the Discussion by a Coherent Conclusion” to “A Unifying Consensus to a Controversial Topic.” Additionally, we have adjusted the manuscript title accordingly \(see tracked version of the revised manuscript\).](#)

[We have also added clarifying commentaries in the Discussion “Comparable Certainty in the Evidence to Medical Cancer Treatments,” “Comparable Survival Benefits to Medical Cancer Treatments,”](#)

“Additional Benefits of Psychosocial Treatment for Cancer Patients,” and “Recognizing the Role of Psycho-Oncology in Clinical Decision-Making” to more accurately position our findings within the broader scientific and clinical context.

We agree that the wide prediction interval is a key indicator of effect heterogeneity and limits the generalizability of a single point estimate across all patient populations. Noteworthy, while it is possible in theory that certain psychosocial interventions could be ineffective or even slightly harmful in specific contexts, in practice, none of the included RCTs demonstrated adverse effects on survival. As detailed in our response to Reviewer 2 (comment R2.9), the observed heterogeneity likely reflects variability in intervention types, cancer types, and disease stages – an inherent limitation given the relatively sparse and diverse evidence base in psycho-oncology. This situation contrasts research with pharmacological treatments, where greater financial incentives often enable a larger number of more homogeneous trials (e.g., similar drugs, dosages, cancer types, and disease stages), thereby – in theory – reducing between-study variance and allowing for narrower prediction intervals.

Nevertheless, our overall effect size of HR **0.80**, 95% CI [0.71, 0.90] is comparable to established medical treatments such as:

Chemotherapy for metastatic breast cancer: HR **0.87**, 95% CI [0.78, 0.97] (W. Ren et al., 2022); chemotherapy for pancreatic cancer: HR **0.71**, 95% CI [0.62, 0.82] (Evans et al., 2024); radiotherapy for breast cancer: RR of **0.93**, 95% CI [0.79, 1.09] (Garg et al., 2017), radiotherapy for gallbladder cancer: OR **0.63**, 95% CI [0.50, 0.81] (B. Ren et al., 2020); hormone therapy in breast cancer: HR **0.79**, 95% CI [0.69, 0.90] (Yu et al., 2017) and the protective effect of marriage: HR **0.80**, 95% CI [0.79, 0.81] (Aizer et al., 2013) in cancer populations.

These effects are generally accepted as clinically meaningful. Further, few meta-analyses of medical cancer treatments report 95% prediction intervals – yet our own calculations of such intervals for five recent meta-analyses revealed similarly wide ranges: W. Ren et al., 2022: 95% PI [HR **0.66, 1.15**]; Evans et al., 2024: 95% PI [HR **0.58, 0.87**]; Garg et al., 2017: 95% PI [RR **0.72, 1.21**]; B. Ren et al., 2020: 95% PI [OR **0.26, 1.54**] and Yu et al., 2017: 95% PI [HR **0.49, 1.28**].

Despite these wide 95% prediction intervals and known side effects such as anemia, alopecia, anxiety, cognitive dysfunction, fatigue, insomnia, immunosuppression, nausea and vomiting, low mood, mucositis, pain, and weight loss (De Ruyscher et al., 2019; Fleming et al., 2022; Nurgali et al., 2018) these treatments are routinely integrated into cancer care – underscoring the importance of contextual interpretation of the level of certainty and 95% prediction intervals.

The observed prediction interval in our study may not solely reflect true clinical heterogeneity. Methodological issues, particularly the small sample sizes and low statistical power of many included trials, likely inflated between-study variance. For example, the meta-analytic post-hoc power to detect the observed overall effect (HR = 0.80) across included RCTs was only 17% meaning that the variability in observed effects is likely amplified by random error.

Importantly, psychosocial interventions not only show survival benefits comparable to standard medical cancer treatments, but also demonstrate consistent improvements in psychological outcomes. Meta-analyses have found positive effects on well-being (McLouth et al., 2020), quality of life (de la Torre-Luque et al., 2015), anxiety (Sanjida et al., 2018), depression (Hart et al., 2012), stress (C. Zhao et al., 2021), fatigue (Xunlin et al., 2019), and pain (Warth et al., 2020).

Taken together, these findings offer a strong rationale for the systematic integration of psychosocial interventions into comprehensive cancer care. Nevertheless, we fully concur that further research is needed to identify which interventions are most effective for which patient populations, and we have revised the manuscript accordingly to better reflect the exploratory and provisional nature of our conclusions.

- Aizer, A. A., Chen, M.-H., McCarthy, E. P., Mendu, M. L., Koo, S., Wilhite, T. J., Graham, P. L., Choueiri, T. K., Hoffman, K. E., Martin, N. E., Hu, J. C., & Nguyen, P. L. (2013). Marital Status and Survival in Patients With Cancer. *Journal of Clinical Oncology*, 31(31), 3869–3876. <https://doi.org/10.1200/jco.2013.49.6489>
- de la Torre-Luque, A., Gambara, H., López, E., & Cruzado, J. A. (2016). Psychological treatments to improve quality of life in cancer contexts: A meta-analysis. *International Journal of Clinical and Health Psychology*, 16(2), 211–219. <https://doi.org/10.1016/j.ijchp.2015.07.005>
- De Ruyscher, D., Niedermann, G., Burnet, N. G., Siva, S., Lee, A. W. M., & Hegi-Johnson, F. (2019). Radiotherapy toxicity. *Nature Reviews Disease Primers*, 5(1). <https://doi.org/10.1038/s41572-019-0064-5>
- Evans, D., Ghassemi, N., Hajibandeh, S., Hajibandeh, S., Romman, S., Laing, R. W., Durkin, D., & Athwal, T. S. (2024). Meta-analysis of adjuvant chemotherapy versus no adjuvant chemotherapy for resected stage I pancreatic cancer. *Surgery*, 175(6), 1470–1479. <https://doi.org/10.1016/j.surg.2023.11.027>
- Fleming, L., Agnew, S., Peddie, N., Crawford, M., Dixon, D., & MacPherson, I. (2022). The impact of medication side effects on adherence and persistence to hormone therapy in breast cancer survivors: A quantitative systematic review. *The Breast*, 64, 63–84. <https://doi.org/10.1016/j.breast.2022.04.010>
- Garg, P. K., Jakheti, A., Pandey, R., Chishi, N., & Pandey, D. (2017). Adjuvant radiotherapy versus observation following lumpectomy in ductal carcinoma in-situ: A meta-analysis of randomized controlled trials. *The Breast Journal*, 24(3), 233–239. <https://doi.org/10.1111/tbj.12889>
- Hart, S. L., Hoyt, M. A., Diefenbach, M., Anderson, D. R., Kilbourn, K. M., Craft, L. L., Steel, J. L., Cuijpers, P., Mohr, D. C., Berendsen, M., Spring, B., & Stanton, A. L. (2012). Meta-analysis of efficacy of interventions for elevated depressive symptoms in adults diagnosed with cancer. *JNCI: Journal of the National Cancer Institute*, 104(13), 990–1004. <https://doi.org/10.1093/jnci/djs256>
- McLouth, L. E., Ford, C. G., Pustejovsky, J. E., Park, C. L., Sherman, A. C., Trevino, K., & Salsman, J. M. (2020). A systematic review and meta-analysis of effects of psychosocial interventions on spiritual well-being in adults with cancer. *Psycho-Oncology*, 30(2), 147–158. <https://doi.org/10.1002/pon.5562>
- Nurgali, K., Jagoe, R. T., & Abalo, R. (2018). Editorial: Adverse effects of cancer chemotherapy: Anything new to improve tolerance and reduce sequelae? *Frontiers in Pharmacology*, 9(245). <https://doi.org/10.3389/fphar.2018.00245>
- Ren, B., Guo, Q., Yang, Y., Liu, L., Wei, S., Chen, W., & Tian, Y. (2020). A meta-analysis of the efficacy of postoperative adjuvant radiotherapy versus no radiotherapy for extrahepatic cholangiocarcinoma and gallbladder carcinoma. *Radiation Oncology*, 15(1). <https://doi.org/10.1186/s13014-020-1459-x>
- Ren, W., Yu, Y., Hong, H., Wang, Y., Gao, Q., Chen, Y., Chen, P., Zhao, J., Ou, Q., Lin, D., Fu, T., Tan, Y., Li, C., Xie, X., Ye, G., Tang, J., & Yao, H. (2022). Clinical evidence of chemotherapy or endocrine therapy maintenance in patients with metastatic breast cancer: Meta-analysis of randomized clinical trials and propensity score matching of multicenter cohort study. *Cancer Research and Treatment*, 54(4), 1038–1052. <https://doi.org/10.4143/crt.2021.698>
- Sanjida, S., McPhail, S. M., Shaw, J., Couper, J., Kissane, D., Price, M. A., & Janda, M. (2018). Are psychological interventions effective on anxiety in cancer patients? A systematic review and meta-analyses. *Psycho-Oncology*, 27(9), 2063–2076. <https://doi.org/10.1002/pon.4794>
- Warth, M., Zöller, J., Köhler, F., Aguilar-Raab, C., Kessler, J., & Ditzgen, B. (2020). Psychosocial interventions for pain management in advanced cancer patients: A systematic review and meta-analysis. *Current Oncology Reports*, 22(1). <https://doi.org/10.1007/s11912-020-0870-7>
- Xunlin, N., Lau, Y., & Klainin-Yobas, P. (2019). The effectiveness of mindfulness-based interventions among cancer patients and survivors: A systematic review and meta-analysis. *Supportive Care in Cancer*, 28(4), 1563–1578. <https://doi.org/10.1007/s00520-019-05219-9>
- Yu, X., Zhou, S., Wang, J., Zhang, Q., Hou, J., Zhu, L., He, Y., Zhao, J., & Zhong, S. (2017). Hormone replacement therapy and breast cancer survival: a systematic review and meta-analysis of observational studies. *Breast Cancer*, 24(5), 643–657. <https://doi.org/10.1007/s12282-017-0789-5>
- Zhao, C., Lai, L., Zhang, L., Cai, Z., Ren, Z., Shi, C., Luo, W., & Yan, Y. (2021). The effects of acceptance and commitment therapy on the psychological and physical outcomes among cancer patients: A meta-analysis with trial sequential analysis. *Journal of Psychosomatic Research*, 140, 110304. <https://doi.org/10.1016/j.jpsychores.2020.110304>

R4.2: Moreover, it appeared hard to find clear moderator effects, which means that a large part of the heterogeneity remains unexplained (by the way, it would be interesting to have this information as well).

Reply: As outlined above, the heterogeneity observed in our meta-analysis appears to result from both clinical and methodological variability across the included primary studies. Importantly, the limited number of studies available for moderator analyses reduced our ability to systematically investigate sources of heterogeneity. For example, subgroup analyses often included only a small number of studies per category, resulting in limited statistical power to detect potential moderators. We have added a note on this limitation in the Limitations “Overall Effect and Moderators” of the revised

manuscript to clarify that the absence of clear moderator effects should not be interpreted as evidence of their non-existence, but rather as a reflection of the current evidence base.

R4.3: Therefore, more primary research, followed by meta-analytic research is needed to get a clearer understanding of when and for whom psychosocial interventions work.

Reply: We agree to this statement and have added a similar line to the abstract and conclusion – see end of Abstract and Discussion “A Unifying Consensus to a Controversial Topic” in the revised manuscript.

R4.4: Related to that, I noted that for some kinds of cancer (e.g., brain, gastric, leukemia, ...), data from only one single study are available, which is definitely not sufficient to know what the effect is for each of the domains. I therefore do not agree with their conclusion that their findings “provided robust evidence for the existence of a generalizable true overall survival-enhancing effect” (line 834).

Reply: We acknowledge that some cancer types were represented by only a single study – namely, brain (k = 1), esophageal (k = 1), gastric (k = 1), leukemia (k = 1), liver (k = 1), prostate (k = 1), and thyroid (k = 1). However, the overall dataset also included breast (k = 11), colorectal (k = 2), lung (k = 2), melanoma (k = 2), and mixed cancer type samples (k = 8), with the latter accounting for one fourth of all analyses. While evidence for individual cancer types is limited, our conclusions are primarily based on the multiverse meta-analysis, which consistently demonstrated a survival benefit across a wide range of reasonable analytical decisions. Thus, the term “generalizable” refers to the robustness of the overall effect across meta-analytic specifications within the current corpus of primary studies, rather than across each cancer type individually.

Based on the current evidence and considerations above, we recommend making psychosocial interventions accessible to cancer patients because these interventions show survival-enhancing effects comparable to medical treatment with a similar level of certainty for the evidence, are free of known adverse side effects, provide substantial psychological benefits such as well-being, quality of life, anxiety, depression, stress, fatigue, as well as pain (see R4.1 for references) and – importantly – no study has reported any harm. This recommendation is further supported by Aizer et al. (2013), who found significantly improved survival among married versus unmarried cancer patients in a large-scale observational study involving 734,889 individuals across ten major cancer types. Nevertheless, we have removed the word “all” before our recommendations to cancer patients and noted that due to the large 95% prediction interval the survival benefits may not contribute uniformly to all cancer patients (see end of Abstract and Discussion “Suggestions for Policy Makers, Clinicians, and Patients”).

R4.5: Their recommendation that psychosocial interventions “should be made accessible to all cancer patients” (line 910) might be the best recommendation for now, but we cannot see this recommendation as the final guideline, because their prediction interval shows that effects might also be negative. I also think the authors are too optimistic when they say: “Therefore, comprehensive approaches, such as our present analysis, are often needed to get a valid overall picture to propel science forward when it seemingly got stuck” (line 904). In sum, I think some more modesty would be appropriate.

Reply: As explained above, we have now made considerable revisions throughout the manuscript to adopt a more modest and nuanced tone, particularly in response to the suggestions raised by several reviewers. We have also adjusted the phrasing of the sentence in question to better reflect the tentative nature of our conclusions and the limitations of the current evidence base. Specifically, we have revised the sentence to: “Therefore, comprehensive approaches, such as our present analysis, may offer a useful step toward a more valid overall picture, especially in research fields where the existing evidence is fragmented or inconclusive.” – see Suggestions “Suggestions for Future Meta-Analytic Research” in the revised manuscript.

R4.6: Second, like I mentioned above, I like the thorough way the data are analyzed and the fact that often multiple techniques and analyses are done and results compared. A drawback however is that the paper is very long, and the reader might get lost in all the detailed analyses and results, and lose the focus on the main findings. I therefore think that it would be a good idea to shorten the paper further, by moving some more results to the supplemental materials.

Reply: We appreciate your concern regarding the manuscript's length and fully understand the risk that some readers may find it challenging to maintain focus amid the many detailed analyses. We have made multiple attempts and combined our efforts to shorten the manuscript as much as possible, while still adhering to best practice guidelines such as APA 7 and PRISMA 2020 – ensuring that readers are provided with sufficient context to follow our line of reasoning.

Nonetheless, it seems that additional explanations were still necessary to adequately address all of the reviewers' insightful comments and concerns. The manuscript is long because our project is indeed a very extensive one, encompassing both primary and secondary research in psycho-oncology conducted over the past four decades, and applying a full range of state-of-the-art meta-analytic techniques to untangle this highly complex and contested field – naturally resulting in a more comprehensive manuscript.

Reviewer 3 (comment R3.5) generously acknowledged the necessity of this length, stating: *“I habitually push for brevity (as do editors who control precious journal pages), but these authors deserve a looong leash; I see no obvious place for text cutting.”* We are sincerely grateful for this supportive assessment, while also recognizing the importance of balancing depth with readability.

In response to both reviewers' feedback, we have made further efforts to streamline the presentation wherever possible and have now moved the entire Appendix (~100 pages) into the Supplementary Information. We sincerely hope that the editors will take into account the scope and significance of this work and grant the necessary space to present this timely and meaningful contribution to the field of psycho-oncology.

R4.7: Third, the authors state that they used random and mixed effects models. ‘Traditional’ random and mixed effects models assume that effect sizes are independent, which is probably not the case when there are multiple effect sizes per study. Multiple effect sizes per study is in fact rather common, for instance because effect sizes may be reported for several subpopulations or time points. If effect sizes are dependent, it is necessary to account for the dependence. An appropriate way to do so is using multilevel meta-analysis and/or robust variance estimation. The authors are not clear about whether there were multiple effect sizes per study and, if so, how they handled this issue.

Reply: We agree that dependencies between effect sizes within studies must be carefully addressed in meta-analytic research. Regarding our random-effects meta-analysis, we followed the standard approach in which each study contributed exactly one effect size, ensuring statistical independence across effect sizes. We assume this is the analysis that you are referring to. As a sidenote, mixed effect models have been used in the moderator analysis according to current meta-analytic research practices (see Borenstein et al., 2021).

Regarding our multiverse meta-analysis, an approach that involves running many meta-analyses under varying inclusion and analytic decisions (the “garden of forking paths”), we also included only one effect size per study and per meta-analytic specification (i.e. a single meta-analysis) – e.g., the overall effect at 1-year, 2-year, 5-year follow-up, etc. – depending on the decision rule used in that analytical path. No individual meta-analytic specification ever included multiple effect sizes from the same study. Therefore, dependency of effect sizes within meta-analyses was avoided by design.

Hence, while dependencies may exist in a broader context, they do not violate the assumption of independence within any single meta-analysis reported in the multiverse. Each analysis was conducted using standard random-effects models, which are appropriate under these conditions.

We have revised the manuscript to make this point clearer and to ensure transparency about our analytic strategy. We hope this explanation resolves the concern and reassures that dependency was not an issue in the way we structured our analyses – see Methods “Specification-Curve and Multiverse Meta-Analysis Methods” in the revised manuscript.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2021). *Introduction to meta-analysis* (2nd ed.). John Wiley & Sons.

R4.8: For the random forest moderator analyses, missing moderator values were replaced with the median or mode. This approach has strong disadvantages: It does not take into account the uncertainty we have about the imputed values. Moreover, it decreases a possible association between the moderator and the effect sizes, therefore giving biased results. If a considerable number of scores would be missing (unfortunately, no information is given on this percentage), using multiple imputation might be a more valuable approach.

Reply: We have taken this concern seriously and reanalyzed the dataset, now excluding all moderators with > 10% of missing values. For the remaining moderators (all with $\leq 10\%$ of missingness), we used simple imputation via median (for continuous variables) or mode (for categorical variables).

We deliberately chose this approach over multivariate imputation for two main reasons:

- (1) Most moderators did not show sufficient intercorrelation to support a reliable multivariate imputation model, and
- (2) MetaForest does not support the incorporation of multiple imputed datasets within its resampling framework, which would have introduced considerable complications and potential bias in variable importance estimation.

We have clarified this rationale and the proportion of missing data in the revised sections of the Methods, Results, Discussion and Limitations – see Methods “Exploratory Machine-Learning-Based Search for Moderators”, Results “Identifying Relevant Moderators via Machine Learning”, Discussion “Moderators” and Limitations “Moderators” in the revised manuscript.

R4.9: Details:

- In the text, the authors often refer to tables and figures in the supplemental materials, without saying that these can be found in the supplemental materials.

Reply: As advised, all references to tables and figures in the manuscript have been revised in accordance with APA 7 style guidelines. We have also clarified that these can be found in Supplementary Information – see the revised manuscript.

- Line 295: study screening was done independently by two researchers. It is not clear whether the whole screening process was done by two reviewers. A measure of agreement would be valuable.

Reply: The literature search and study screening were completely conducted independently by two reviewers (KDAH, LB) and all disagreements were resolved through discussion and consensus – see adjustments to Methods “Selection Process of Primary Studies” in the revised manuscript. However, we did not calculate a quantitative measure of disagreement.

- Line 425: I find the explanation of a mixed-effects model not very helpful, i.e., “using a random-effects model within, but a fixed-effect model between subgroups”. Only when one already has a good understanding of a mixed-effects model, (s)he can understand what the authors mean.

Reply: Sentence was revised for more clarity: “A mixed-effects (ME) meta-analytic model was used in the subgroup analysis to identify moderators that may generalize to other populations. This is a standard approach for subgroup analysis and applies a random-effects model within each subgroup to account for variation among studies, while using a fixed-effect model to compare the average effects

between subgroups (Borenstein et al., 2021).” – see Methods “Effect-Moderator Analysis” in the revised manuscript.

- Line 453 says that 20,000 trees were used, line 635 talks about 10,000 trees.

Reply: As explained in the methods section, for the MetaForest analysis, an initial "test" model was run with a higher number of trees (20,000) to evaluate model convergence. Once it was confirmed that the model converged (i.e. for our dataset at 10,000), the number of trees was reduced for the final model to optimize computational efficiency without sacrificing the accuracy of the results. We have adjusted explanations slightly for more clarity – see Methods “Exploratory Machine-Learning-Based Search for Moderators” and Results “Identifying Relevant Moderators via Machine Learning” in the revised manuscript.

- Line 461: what is a “50% percentile interval”?

Reply: The "50% percentile interval" refers to the range between the 25th and 75th percentiles of the variable importance metrics. This interval captures the middle 50% of the distribution, providing a measure of the central tendency and variability of the importance scores. Moderators for which this interval did not include zero were retained in the model, ensuring that only those with consistent and non-zero importance were included. This is the approach that MetaForest automatically uses according to van Lissa et al. (2017; 2020). We have added a phrase to explain what is meant – see Methods “Exploratory Machine-Learning-Based Search for Moderators” in the revised manuscript.

van Lissa, C. J. (2017). MetaForest: Exploring heterogeneity in meta-analysis using random forests. *PsyArXiv*. <https://doi.org/10.31234/osf.io/myg6s>

van Lissa, C. J. (2020). Small sample meta-analyses: Exploring heterogeneity using MetaForest. In R. van De Schoot & M. Miočević (Eds.), *Small sample size solutions: A guide for applied researchers and practitioners* (pp. 121–140). CRC Press. <https://www.crcpress.com/Small-Sample-Size-Solutions-Open-Access-A-Guide-for-Applied-Researchers/Schoot-Miocevic/p/book/9780367222222>

- Line 610: the authors say that in Figure B2 no time trend can be identified. Yet, I found it striking that after 2015 all ratios were smaller than 1 (positive effect), whereas before 2015 effects were fluctuating around 1 (no effect). Also in Figure B3 (cumulative meta-analysis), it can be seen that after 2015 the mean ratio is gradually decreasing.

Note that we have changed the order of Figure B1 and B2 in the revised manuscript.

Reply: Has now been addressed and described in the results – see Results “Meta-Analytic Synthesis” and Discussion “Time Trend” in the revised manuscript.

- Line 604 (and 729): the authors suggest that from 2015 onwards, the effect becomes statistically significant because of an increased precision due to a larger number of available studies. Yet, also here the changing mean effect size plays a role.

Note that we have changed the order of Figure B1 and B2 in the revised manuscript.

Reply: Has now been addressed and described in the results – see Results “Meta-Analytic Synthesis” and Discussion “Small-Study Effects” in the revised manuscript.

- In some figures or tables including ORs, SMDs or HRs, it would be helpful to remind the reader how the direction of the effect size should be interpreted (e.g., does a positive or a negative SMD or a OR/HR larger or smaller than 1 indicate that the intervention is successful? (in this regard: I like the use of the ‘median survival difference’, because this helps the reader to understand the effect sizes).

Reply: To improve clarity and accessibility, we have revised all relevant figures and tables to include explanatory notes on the direction of the effect sizes (e.g., whether HR/OR/RR < 1 or positive SMD indicates a survival benefit). These notes have been added in line with both APA 7 and PRISMA 2020 guidelines, ensuring that each figure and table can be interpreted independently of the main text.

- Line 656: not significant, or significant but in the opposite direction as one would expect.

Reply: Agreed and changed accordingly – see Results “Publication Bias” in the revised manuscript.

- Line 664: the use of the Meta-Showdown Explorer is not very clear. It is suggested that “no bias” is used as input, but how comes that in that case the power can be evaluated (is this the power of the mean effect size?)?

Reply: As described in the Methods section, no definitive indication of publication bias was found using the pre-registered procedures p-uniform and FAT-PET-PEESE. Based on this finding, we used the Meta-Showdown Explorer to assess which estimation method would be most appropriate to estimate the overall effect – see Methods “Reporting Bias Assessment” in the revised manuscript.

To match the characteristics of our dataset, we specified the simulation parameters as follows: no publication bias, $\tau = 0.2$, $k = 30$, true effect size SMD = 0.2, and no questionable research practices. It is important to note that the Meta-Showdown Explorer only allows the selection of moderate or high levels of publication bias – a “low bias” option is not available – see Results “Correcting for Publication Bias” in the revised manuscript.

The Meta-Showdown Explorer then displays the false-positive rate (i.e., type I error rate) and the true-positive rate (indicative of statistical power) for various estimation methods, based on pre-computed simulation data. Under the specified parameters, it indicated that the naive random-effects meta-analysis (RE MA) yielded both a lower false-positive rate and a higher true-positive rate than p-uniform and FAT-PET-PEESE (see Figures E1 & E2).

In addition, regarding the precision of estimating either a true null effect ($\delta = 0$) or a small true effect ($\delta = 0.2$), the RE MA outperformed p-uniform and showed comparable performance to FAT-PET-PEESE (see Figures E1 & E2).

We have added short explanations regarding this point to the Methods “Reporting Bias Assessment” and Results “Correcting for Publication Bias” in the revised manuscript.

Thank you for your valuable suggestions!

Communications Psychology

Manuscript Title

Psychosocial interventions prolong survival time in cancer patients evidenced by systematic review, state-of-the-art meta-analysis, and multiverse meta-analysis of randomized controlled trials

REVIEWER EXPERTISE:

Reviewer #2 psychosocial cancer interventions, metaanalysis

Reviewer #3 psychosocial cancer interventions, metaanalysis

Reviewer #4 statistical methods, metaanalysis

To the reviewers: Again, we would like to sincerely thank all reviewers for their thoughtful and constructive feedback, as well as for the time and expertise they dedicated to carefully evaluating our manuscript. Their insightful comments and suggestions have been invaluable in improving the clarity, rigor, and overall quality of our work. We greatly appreciate their efforts in helping us strengthen both the scientific and practical contributions of this study. Please see the tracked version of the manuscript (if possible) to see the changes highlighted.

REVIEWER REPORTS:

Reviewer #2 (Remarks to the Author):

The authors have now addressed several of the issues raised by this reviewer (#2). They have included the RoB 2 and GRADE evaluations (R2.4.; R.2.11), have improved the figures (R.2.7.), and have acknowledged the challenges related to publication bias assessment (R.2.6.). They have also provided lengthy responses to some of my questions. While the manuscript thus has been improved, I still think it is a long and complex paper with many different (and possibly underpowered) analyses of robustness, including p-curve analysis (R.2.8.) and two different moderator analysis approaches (R.2.5.). Furthermore, I am not completely comfortable with the HR calculations (R.2.3.) and the machine learning moderator analysis (with a small number of studies) (R.2.5.). Finally, given the many analyses and approaches included, I still think that dividing the paper into two separate papers (a review of reviews) and a meta-analysis of primary studies will improve the readability of the study (R.2.2.). Thus, while the manuscript represents a comprehensive and thoughtful approach to an important scientific and clinical question, and several improvements have been made, there are still issues that warrant a major revision decision.

R.2.1. I asked for inclusion and review of more recent reviews and meta-analyses, and the authors have now added a list of previously published meta-analyses.

Reply: Yes.

R2.2. My initial concern was that the paper is highly complex – combining what resembles a review of reviews with a systematic review and meta-analysis of primary studies. Although the authors defend their choice and have made some changes to the structure, I would prefer two separate papers published simultaneously in the same journal, i.e., a paper 1 which is written as a standard

umbrella review/overview of reviews and paper 2 presenting an updated meta-analysis of the primary studies included in the meta-analyses plus any additional/newer studies.

Reply: We appreciate your suggestion. The editors have encouraged us to retain the all-in-one version, and we are grateful for their decision. Nevertheless, we acknowledge your concern and agree that your proposed alternative would also have been a valuable option.

R2.3. Re. the calculation of Hazard Ratios, the authors have added a sentence describing the procedure. How HR were calculated is still unclear – and I was not able to open the calculation sheets that were linked to. I would strongly suggest that a statistician be included in the review process.

Reply: Thank you for pointing out the need for clarification regarding our HR calculations and the difficulties in accessing the supplemental files.

We have now added a more detailed description in the Methods section (see “Effect Size Measures”) explaining in more detail how HRs were calculated from published Kaplan–Meier survival curves when values were not directly reported.

In addition, all calculation files are now publicly accessible in plain Excel format (no longer in a .rar archive) via OSF, ensuring easier accessibility and transparency (see Supplemental File S1). These files can now be directly downloaded via OSF without any compression (as several .xlsx and .docx files), ensuring straightforward accessibility. OSF’s interface may have also contributed to some of the confusion due to the prominent “Download File Metadata” button (which retrieves a .json file rather than the actual dataset). The correct download link can be found behind the three dots (...). We hope that our revised upload resolves this issue.

Finally, all HR calculations were independently verified by a co-author with expertise in quantitative methods and statistics (UST), thereby ensuring methodological rigor and minimizing the risk of error.

R2.4. The authors have now, as suggested, used the RoB 2 tool to evaluate the risk of bias.

Reply: Yes.

R2.5. I am still not sufficiently knowledgeable about Machine-learning approaches to moderator analyses. I appreciate that the power of moderator analyses may be limited, but the very extensive moderator analysis approach with both conventional meta-regression and additional machine-learning-based approaches seems to be a bit over-reaching with increased risk of type-1 error.

Reply: Thank you for raising an important point. We agree that moderator analyses carry an increased risk of type I error, especially when many moderators are tested. To address this issue in our conventional meta-regressions, we applied a Benjamini–Hochberg procedure (BHP), which adjusts for the false discovery rate and reduces the likelihood of spurious findings due to multiple testing.

With respect to the machine-learning-based moderator analysis (MetaForest), we would like to emphasize that this approach does not yield inferential p values for moderators but rather provides variable importance (VI) measures and cross-validation error estimates. As such, procedures like the BHP cannot be applied here. Instead, the evaluation of moderators in MetaForest is based on their relative importance and the predictive accuracy of the model, validated via out-of-bag error and

cross-validation. This inherently differs from conventional hypothesis testing and was used in our study – explicitly – as an exploratory complement to the confirmatory subgroup analyses and meta-regressions (see Results “Identifying Relevant Moderators via Machine Learning”).

We have clarified this distinction in the manuscript. The machine-learning analyses are presented as exploratory, hypothesis-generating tools, whereas the confirmatory conclusions are based on the classical meta-regressions corrected for type I error by the BHP (see Results “Moderator Analysis”).

It is important to note that the moderator analysis did not yield the confirmatory identification of any moderator, due to our control for type 1 errors, regardless of whether examined with the classical or the machine-learning approach. We hope this clarification adequately addresses your concern.

R2.6. The authors have now addressed the concern by including a comment in the discussion on the challenges in assessing publication bias with the FAT-PET-PEESE approach (with small sample sizes and heterogeneity).

Reply: Yes.

R2.7. The figures 1A-B have been improved and are now easier to interpret. I am still not very enthusiastic with regard to including both primary studies and meta-analyses (of these studies) in the same timeline figure – the mistake in Figure 5 has been corrected.

Reply: We acknowledge your concern. However, creating additional figures (e.g., meta-analyses from 1980–1999 and 2000–2025) would further increase the manuscript length. For this reason, we respectfully propose to retain the current format.

R.2.8. I remain worried about drawing inferences from a p-curve assessment with only 10 significant studies.

Reply: We have now clarified this in the manuscript. Specifically, we emphasize that the analysis included the minimum recommended number of significant p values ($k = 10$), which inherently limits the robustness and generalizability of p-curve inferences. Consequently, the results should be interpreted with caution and considered preliminary rather than definitive. This revised text is intended to make the limitations and interpretability of the p-curve findings transparent. (see Methods section “p-Hacking and Test for Evidentiality” and Results section “Reporting Bias Across Studies”)

R2.9. Regarding my question concerning interpretability of results, the authors have provided a lengthy response and have included a section on the certainty of the evidence in comparison with evidence for medical cancer treatments. One still wonders about the possible mechanisms – and the degree to which evidence for such mechanisms exist.

Reply: Thank you for raising this stimulating question. We agree that exploring possible mechanisms represents an important and promising line of future research.

R2.10. I suggested the Bayesian approach as a way to evaluate the robustness of the results that go beyond “significant and non-significant”. The authors argue that given their many different types of analyses, they would prefer to stick with frequentist approaches. I will not insist, but given that the

paper is highly complex with many different (and possibly underpowered) analyses of robustness (e.g., p-curve analysis, etc.), this could have simplified the paper.

Reply: Thank you for this thoughtful suggestion. While we opted to remain within a frequentist framework for the present work, we agree that a Bayesian approach could provide a valuable complementary perspective, and we will carefully consider its implementation in future research.

R2.11. The authors have – as suggested – included a GRADE assessment of the certainty of the evidence.

Reply: Yes.

Reviewer #3 (Remarks to the Author):

R3.1. It may seem a contradiction, but this reviewer who first complained about the length of the manuscript and then proposed that this level of detail was nevertheless needed, is now also willing to accept the even lengthier, revised version! The authors were exceptionally thorough in their revisions. I cannot comment in great detail on the responsiveness to all other reviewers' comments, but my concerns were more than adequately dealt with. If there is an absolute need for cuts (not I urge for these) then the discussion might be a little shorter, there is some repetition of results here. In the end, we have a blockbuster manuscript here that will become a citation classic. Every criticism one can throw at the attempt to answer the question of life extension with a psychological treatment has been tackled. I especially like the addition of highlighting not just statistical significance, but what the results mean for cancer care in the broadest sense, and how patients themselves benefit. There is no stone left unturned, no potential criticism ignored. Kudos to the authors; I bow my head to their professionalism.

Reply: Thank you so much for your exceptionally kind and encouraging feedback. We greatly appreciate your recognition of the thoroughness of our revisions and your thoughtful comments on the clinical relevance of our findings. Your endorsement is highly motivating and reinforces our commitment.

Reviewer #4 (Remarks to the Author):

Thank you for the thorough way you tried to account for all comments of the reviewers.
I am happy with the result.

R4.1. Just a very small detail:

Please replace

"This is a standard approach for subgroup analysis and applies a random-effects model within each subgroup to account for variation among studies, while using a fixed-effect model to compare the average effects between subgroups"

by

"This is a standard approach for subgroup analysis and applies random effects to account for variation among studies within subgroups and fixed effects to account for variation between subgroups"

Reply: Thank you for pointing that out. Our statistician (UST) proposed the following formulation: "It is a standard approach in subgroup analysis to model the variation within subgroups as a random

effect and the variation between subgroups as a fixed effect.” (see Methods section “Moderator Analysis”).

Other Manuscript Changes:

One previously listed co-author (Johannes K. Vilsmeier) has requested to be removed as a co-author and asked to be acknowledged in the Acknowledgements section instead, as his contribution is more appropriately recognized in this way. All remaining authors approve of this change. This adjustment does not affect the integrity of the manuscript or the contributions of the remaining authors.