



¹Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

²Division of Hematology, Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA

³Davidoff Cancer Center, Rabin Medical Center-Beilinson Hospital, Petah Tikva, Israel

⁴Samueli Integrative Cancer Pioneering Institute, Rabin Medical Center-Beilinson Hospital, Petah Tikva, Israel

Correspondence to: T Meirson tomermrsn@gmail.com (or @TomerMeirson on Twitter; ORCID 0000-0002-5011-5477)

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2023;383:e077329 http://dx.doi.org/10.1136/ bmi-2023-077329

Accepted: 07 November 2023

Projecting complete redaction of clinical trial protocols (RAPTURE): redacted cross sectional study

Nir Balaban, ¹ Ghulam Rehman Mohyuddin, ² Adi Kashi, ¹ Amir Massarweh, ³ Gal Markel, ^{1,3,4} David Bomze, ¹ Daniel A Goldstein, ^{1,3} Tomer Meirson^{1,3,4}

ABSTRACT

OBIECTIVES

To characterise redactions in clinical trials and estimate a time when all protocols are fully removed (RAPTURE).

DESIGN

Redacted cross sectional study.

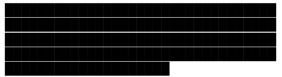
SETTING

Published phase 3 randomised controlled trials from 1 January 2010 to

PARTICIPANTS

New England Journal of Medicine, and Journal of the American Medical Association.

MAIN OUTCOME MEASURES



RESULTS

met the inclusion criteria, with 268 (56.7%) research protocols available and accessible. The rate of redactions in protocols has increased from 0 in 2010 to 60.8% in 2021 (P<0.001). The degree of data redaction has also increased, with the average cumulative redactions among industry funded trials rising from 0 in 2010 to 3.5 pages in 2021 (P<0.001). Modelling predicts that RAPTURE is expected to occur between 2073 and 2136. Redactions featured predominantly in sponsored trials and mostly occurred in the statistical design.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Sponsors claim to redact protocols to protect sensitive commercial and intellectual property

Increasing redactions in clinical trial protocols hinder accurate appraisal and replication of trials

If redaction rates continue to rise, there might come a time when protocols are fully redacted (RAPTURE—removal of amendments and protocols of trials using redactions)

WHAT THIS STUDY ADDS

Protocol redactions appear to be unique to industry funded trials, are rising consistently, and mainly occur in the statistical analysis section

At the current rate of redaction, RAPTURE is estimated to occur between 2073 and 2136

A multipronged strategy against protocol redactions is necessary to maintain the integrity of science

CONCLUSIONS

This study highlights the rise in protocol redactions and predicts that,

redacted between 2073 and 2136. A legitimate

rationale for the redactions could be found. A multipronged strategy against protocol redactions is required to maintain the integrity of science.

AVAILABILITY

This paper is partially redacted, but for the sake of a version without any redactions can be found in the supplementary material.

clinical

Introduction Redactions

and interfere with a clinical trial. ¹⁻³ There are various reasons for protocol redactions by sponsors, including fears of release of commercially sensitive information or trademarked intellectual property. ¹ ⁴ However, the authors of this study have yet to stumble upon any hidden trade secrets when reviewing many protocols during their research endeavours.

trial

protocols

Given the , we fear that a time will come when protocols will be fully redacted, which would pose a problem for all aspects of science. In religious cultures, the term rapture is often used to signify the end of time. Therefore, we aimed to predict when all protocols and amendments would become fully redacted (ie, removal of amendments and protocols of trials using redactions—RAPTURE), which would signify

, justifying the use of the word RAPTURE in this context.

sectional

study,

by assuming a constant increase in rate of redactions, predict the time when all protocols would become fully redacted (RAPTURE). Our approach was to assess the prevalence and characteristics of data redactions in phase 3 clinical trial protocols published in high impact factor journals. An unredacted version of this paper can be found in the supplementary material.

Methods

Search strategy

this

cross

This redacted cross sectional study followed the

aimed

we

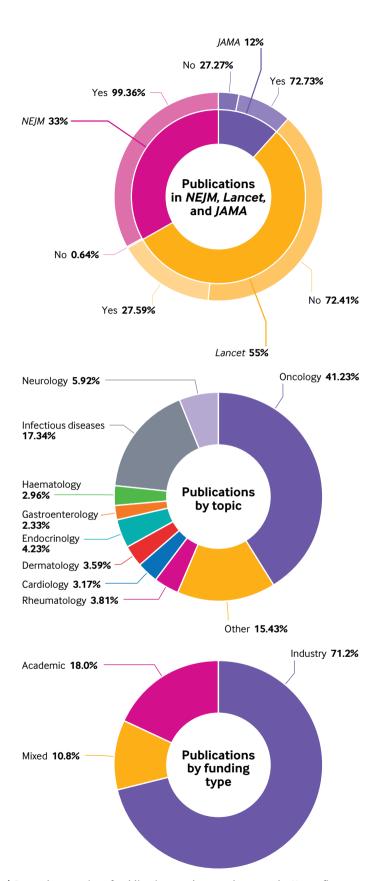
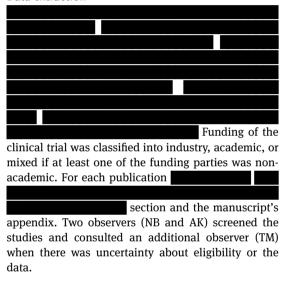


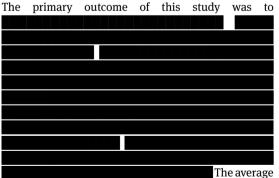
Fig 1 | General properties of publications and research protocols. Upper figure: distribution of publications in *New England Journal of Medicine (NEJM)*, *Lancet*, and *Journal of the American Medical Association (JAMA*). The additional layer depicts the presence of published and accessible protocols within each journal. Middle figure: distribution of publications by topic. Lower figure: distribution of publications by funding type

PubMed for phase 3 randomised clinical trials published between 1 January 2010 and 1 January 2022. Our analysis included studies published in the New England Journal of Medicine (NEJM), The Lancet, and the Journal of the American Medical Association (JAMA). The search terms included randomized, randomised, phase 3, phase III, Lancet (London, England), The New England journal of medicine, and JAMA (see appendix). We excluded studies reporting pooled analysis, patient reported outcomes, or post hoc analysis. When several publications for the same trial were identified (eg, long term follow-up), we included the trial only once, favouring studies with a published protocol.

Data extraction



Redaction analysis



protocol size was calculated by multiplying the mean number of lines and pages of the sampled protocols. To assess the extent of concealment, for each section, one observer (NB) with the help of a second observer (AK) went through the protocols and estimated the total amount of redactions using lines (eg, 0.5, 1, 2 lines). Large sections of redacted text were estimated using pages (eg, 0.5, 1, 2 pages). Because the variation in the number of lines within the protocols was relatively low (mean 39.5, interquartile range 36.8-42.0),

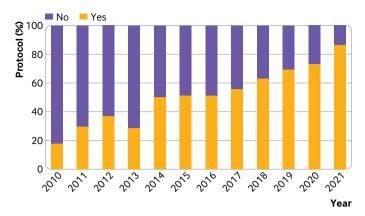
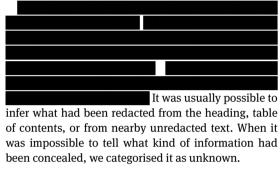


Fig 2 | Presence of protocols over time

, we used curve fitting on redaction patterns over time with linear and nonlinear regression. We used various models, including a simple linear model, a second order polynomial model, and an exponential model using nlsLM() in the R package minpack. lm^6 to fit the equation where y(t) represents mean redaction at time t; a and b are coefficients. Additionally, we applied a general

additive model with the default basis dimension for the smooth function using the R module mgcv. We calculated the coefficient of determination to assess the goodness of fit of the models. The 95% confidence intervals for the regression models were obtained using the R package insight. For the projection, we assumed that the increase in redaction extent would persist at the current rate until the entire protocol becomes completely redacted. To predict the time of RAPTURE, we calculated the time at which the mean redaction model would pass 100%, while assuming a constant average protocol size. We established the time range by identifying where the lower and upper bounds of the 95% confidence interval of the fitted model intersect with 100% mean redaction protocol rate.





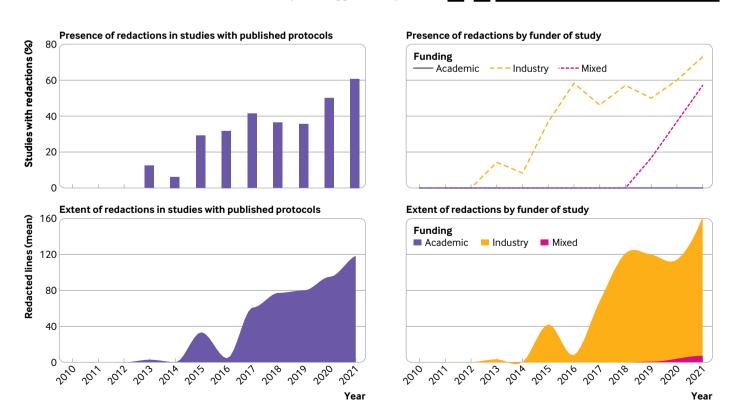
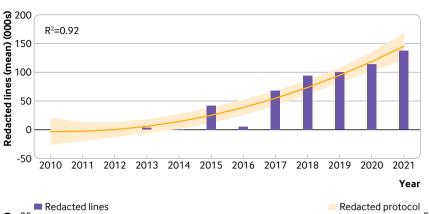


Fig 3 | Presence and extent of data redaction in research protocols over time in studies with published protocols and by funder of study



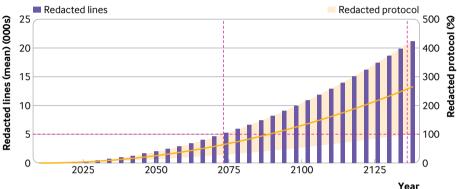


Fig 4 | Curve fitting of redaction patterns over time between 2010 and 2021 (upper panel) to extrapolate and predict date (lower panel) at which 100% of protocol (horizontal dashed line) will be fully concealed if the increase in rate of redactions and average protocol size remain constant among industry funded studies. Time range (indicated by vertical dashed lines) was estimated by identifying where the lower and upper bounds of the 95% confidence interval of the fitted model (shown in orange) intersect with estimated line representing 100% redaction of protocol. 95% confidence interval for projected RAPTURE (removal of amendments and protocols of trials using redactions) in 2088 ranges from 2073 to 2136. Average protocol size was calculated by multiplying mean number of lines and pages of sampled protocols

Patient and public involvement statement

Although patients and the public were not directly involved in the study, we did speak to patients about the study, and we asked a member of the public to read our manuscript before submission.

Results

Between 1 January 2020 and we reviewed we see the second s

that met the inclusion criteria (*NEJM* 157, *Lancet* 261, and *JAMA* 55; supplementary figure 1S). A total of 268 (56.7%) research protocols were available and accessible. In 2010, all but one study published in the *NEJM* had a protocol (99.4%), whereas protocols were included in 72.7% and 27.6% of the studies published in the *JAMA* and the *Lancet*, respectively (fig 1). Most studies were funded by industry (71.2%); the remainder had academic funding (18.0%) or mixed funding (10.8%). The most common topics were oncology (41.2%), followed by infectious diseases (17.3%), and neurology (5.9%). We estimated the average number of pages in protocols as 202.5 (interquartile range 130.5-240.2), with each page containing 39.5 lines (interquartile range 36.8-42.0).

Publication of protocols (fig 2) from 17.6% in 2010 to 86.4% in 2021. Up until 2012, none of the protocols were redacted, however from 2013 they started to contain concealed information with gradually increasing proportions (P<0.001; fig 3). By 2021, 60.8% of protocols had redacted data and the extent of redaction 3, fig 4). The mean number of cumulative redacted lines increased from 0 in 2010 to 137.4 (3.5 pages) in 2021 for industry funded trials (P<0.001). Redactions featured predominantly in industry funded trials and in only a small proportion of those with joint academic and industry funding (fig 3). In 2021, funded trials (excluding those with joint academic and industry funding) had an average of 162.0 lines (4.1 pages) of redacted data per protocol compared with 7.2 lines (0.2 page) for those with joint academic and industry funding (P<0.001). In contrast, no studies with only academic funding were found to have redactions between 2010 and 2021. These findings were consistent for the proportion of redacted protocols and the extent of redactions per protocol.

Considering that the extent of redactions has over time, larger sections of the protocol are expected to be concealed in the future. We sought to predict the time when entire protocols of industry sponsored studies will be fully redacted (RAPTURE). Linear, polynomial, exponential,

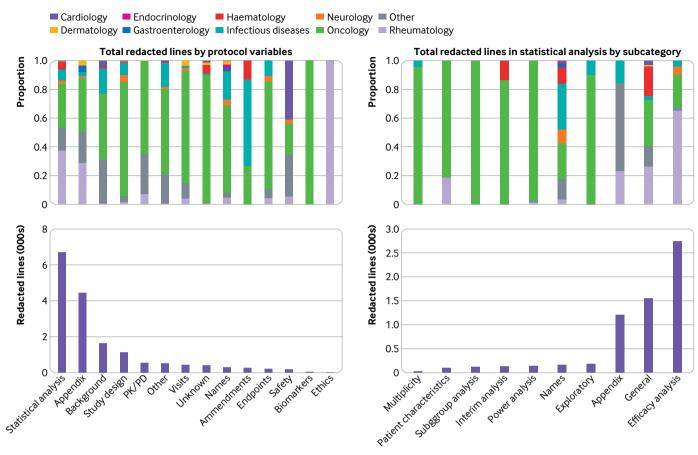


Fig 5 | Presence and extent of data redaction in research protocols by selected protocol variables and in statistical analysis section of protocols classified into subcategories. Upper panels show distribution of topics for each variable. Visits denote visit schedule and assessments; PK/PD=pharmacokinetics and pharmacodynamics

and general additive models were used to fit the data. Supplementary figure 2S shows the results of linear and nonlinear regression analysis of mean redacted lines. The second order polynomial and general additive models yielded a superior fit with an R² value of 0.92. Because both models performed equally well, we selected the simpler polynomial model to predict RAPTURE. If the increase in redaction rates remains constant, we predict RAPTURE to occur between 2073 and 2136 for industry funded studies (fig 4). To test the accuracy of our model, we sampled data from 1 January to 1 June for the years 2022 and 2023. This search identified 37 and 45 studies with protocols for each respective year. Despite the model's prediction of an expected mean number of redacted lines greater than 137.4, the values for 2022 and 2023 were 35.2 and 112.8, respectively.

Supplementary figure 3S shows the patterns of data redaction in the protocols of studies by topic. High proportions of protocol redactions were found for dermatology, paediatrics, pulmonology, and haematology. Excessive redacted text was evident for rheumatology studies, followed by pulmonology, paediatrics, nephrology, and oncology studies. Similar patterns were observed for studies published in the past three years, but the extent of the redactions has risen substantially (supplementary figure 3S).

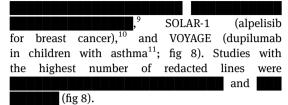
We calculated the extent of redactions according to a set of defined categories (fig 5, fig 6, fig 7; see also supplementary figure 4S). Most redactions occurred (46.4% of protocols), concealing a total of 6685 lines (169 pages). Most redactions in the statistical analysis occurred in the statistical analysis plan (95.2%). Excessive redactions were also found in the protocol's appendix (4445 lines, 113 pages), background (1638 lines, 41.4 pages), and study design (1128 lines, 28.5 pages). Redactions also occurred in sections classified as unknown (400 lines, 10.1 pages)-entire sections were so heavily redacted that their original context or category could not be identified. Redaction in the statistical analysis plan was further categorised into subcategories (fig 5). Redactions affecting the efficacy analysis were more common in rheumatology and oncology; other subcategories were more specific to oncology studies, such as exploratory analyses, power calculations, interim analyses, and subgroup analyses. To outline the variations in redactions across various sections of the protocol and between different protocols, we examined their distribution and the strength of their association (fig 6, fig 7). The analysis shows some implementing minimal redactions limited to specific sections, while others are linked to extensive



Fig 6 | Distribution of redactions among protocol variables. Each row represents a study protocol. Visits denote visit schedule and assessments; PK/PD=pharmacokinetics and pharmacodynamics

redactions encompassing several sections, notably within the statistical analysis. In most studies where parts of the appendix, background, or study design were redacted, sections of the statistical analysis were also redacted.

We ranked studies and companies according to the amount of redactions (fig 8, fig 9, fig 10). The studies that concealed the most lines include



Discussion

In this study of data redaction in phase 3 clinical trials published iournals. we found in from 0 in 2010 to 60.8% of protocols in 2021. Additionally, we found that the amount of redacted text within a protocol increased. Our mathematical model predicted that RAPTURE (the date when all protocols are fully concealed) is expected to occur between 2073 and 2136 for industry funded trials, a worrying possibility with serious implications for the scientific community.

We acknowledge that there is levity involved in our prediction of RAPTURE, and that the intention of our paper is to highlight the rise in redactions rather than believe that all protocols will be 100% redacted in the predicted timeframe. As such, our mathematical model is meant to be provocative and not . Furthermore, we only provide a timeframe for when RAPTURE might occur rather than a fixed date. Throughout history, humanity has been wrong about predicting with precision when RAPTURE will occur, mostly believing that the end is imminent, only to be proven wrong repeatedly (remember Y2K?). To avoid a similar predicament and to stay true to our scientific methods, we only provide a range for when RAPTURE might occur, rather than a precise date.

However, the rise in redactions we observed is worrying in itself even if RAPTURE does not occur. In a strange twist of events, there is a glimmer of hope-2022 and 2023 saw a slight drop in redactions. We are not sure if this is the result of a secret society of antiredaction advocates who have successfully led their campaign. Perhaps industry has suddenly paid heed to antiredaction efforts and are now willing to publicly share some of their trade secrets-that is, obscure statistical calculations. In all seriousness, we are cautious in inferring too much from this drop in redactions, and we do not know the impact the covid-19 pandemic might have had on redaction practices. We believe that advocacy against these redaction practices remains essential, even if RAPTURE appears to be less of a mathematical certainty based on recent trends.

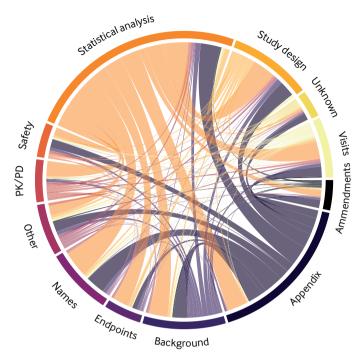


Fig 7 | Circos plot depicting associations between different sections of protocol. Broader chords indicate stronger average association among variables within protocols. Visits denote visit schedule and assessments; PK/PD=pharmacokinetics and pharmacodynamics

Our work is a recent characterisation of redactions in the scientific literature. This characterisation shows

We also found that protocol redaction appears to be unique to funded trials, and that they mostly occur in the statistical analysis section of the protocol, particularly the statistical analysis plan. We struggled to see the logic in redacting statistical analysis sections to conceal trade secrets. Surely no commercial advantage can be gained from hiding statistical power calculations.

Previous work has shown a lack of clear objective reasoning explaining why protocol redaction occurs.4 While redaction of confidential commercial information is understandable, in our study most redactions were found in the . In contrast, only a small proportion of redactions occurred in the biomarkers section, which might include commercial and innovative information. Unless the research and statistical design are fully transparent and adequately prespecified, redactions of the methods might allow enough margin to manipulate the results. 12-16 We cannot think of other reasons why these redactions occurred. It seems ironic when companies justify redactions by stating that disclosures could greatly affect their commercial interests; they are being truthful, but not in the way

What would be the scientific implications if RAPTURE were to occur? Science as we know it would

scientific experiments. Regulatory authorities would struggle with drug approvals, physicians would not know the specifics when obtaining patient consent, and patients would be at risk of harm.

How can we avert this RAPTURE? The solutions are **Example**. Editors can refuse to publish research unless a fully unredacted protocol is available for review. Governments can mandate public access to a fully unredacted protocol available for review. Although previous guidelines by regulatory authorities have addressed redaction, ¹⁷ further oversight is needed.

Strengths and limitations of study

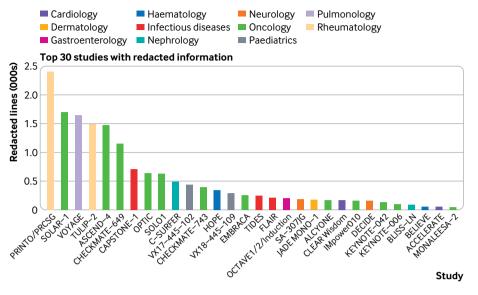
Our work has several . The degree of redaction across lines and pages was visually approximated using the dimensions of the obscured sections in relation to the text. However, small discrepancies might exist between estimated and actual concealment. The extent of redactions within a protocol does not necessarily correlate with its significance; concealing a small section or even a single phrase within the study's statistical design could hinder the ability to accurately appraise and reproduce a clinical trial. Apart from contact information, which was excluded, we regarded all sections of the protocol as potentially important because the choice of concealment was not arbitrary; however, different sections carry varying degrees of significance. Although our search strategy was designed to capture randomised controlled trials published in top journals during the given timeframe, this search strategy inherently selects industry published studies because most large randomised controlled trials are funded by industry.¹⁸ Therefore, a selection bias exists in our study because we used this sample for convenience, and our findings might not reflect redaction throughout all clinical trials.

Statistical power for findings from earlier years is limited, and it is possible that the apparent increase in redactions might not be solely caused by more editing, but rather a result of more protocols being edited out over time. While there are fewer protocols from the early years, the total number published is still substantial, with 45 protocols until 2014. The lack of redactions until that time further indicates that they are a recent phenomenon. However, there is variability in the type and extent of redaction and many protocols remain redaction free. Notably, academic sponsored trials consistently show no redactions.

The model used to predict the time of RAPTURE assumes that the increase in redactions will persist at the same rate as it did until 2021. If the rate of concealment stabilises or decreases in the future, as suggested by data from 2022 and 2023, the projected date will consequently be adjusted and pushed back. Other limitations include

The dataset does not encompass all relevant published

they intended.



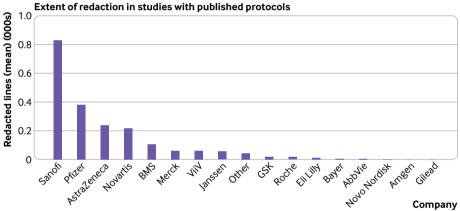


Fig 8 | Extent of data redaction in research protocols for top 30 studies with redacted information and by company

studies, particularly those that do not explicitly mention their design in the title or abstract with terms like randomised or phase 3. We could not determine why protocol redaction occurred and can only speculate on the reasoning. We did not request protocols that were missing. Previously authors have faced lawsuits⁴ when requesting access to full protocols during their analysis of redaction, and industry lawsuits might have served as a successful deterrent.

Conclusions

summary, this cross sectional study redactions. found protocol Ιf redactions increase at the same rate estimate protocols will be fully redacted between 2073 and 2136 (RAPTURE), although a limited decrease in the amount of redactions was observed in 2022 and 2023. A multipronged strategy against protocol

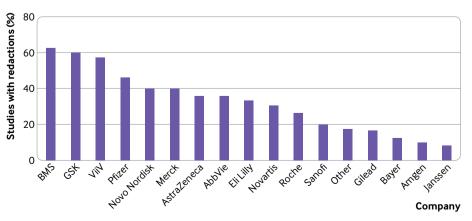


Fig 9 | Presence of redacted protocol by company with at least six trials

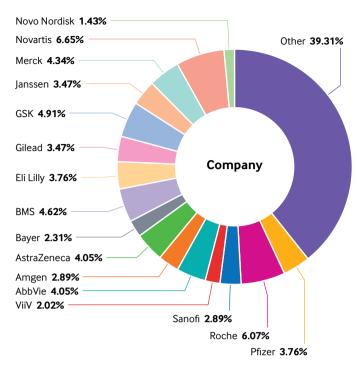


Fig 10 | Distribution of studies in dataset by company. Companies with five or fewer publications are classified as other

redactions is necessary to maintain the integrity of science.

Nir Balaban and Ghulam Rehman Mohyuddin contributed to this study equally.

Contributors: NB and GRM contributed equally and share first authorship. TM conceived the idea of the study and was primarily responsible for its design and analysis of the data. TM and DB contributed to the study conceptualisation. NB and AK contributed to data collection and curation. TM, GRM, and NB cowrote the first draft of the manuscript. TM, GM, and DAG supervised the project. All authors contributed to the interpretation of the results and reviewed and approved the final draft of the manuscript. TM and NB are the guarantors. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding: None declared.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/disclosure-of-interest/ and declare: no support from any organisation for the submitted work; TM reports receiving personal fees from Purple Biotech, outside the submitted work. GRM reports receiving royalties from MashupMD for writing, outside the submitted work. GM reports receiving personal fees from MSD and Roche; grants and personal fees from BMS and Novartis; personal fees and stock options from 4C Biomed; and stock options from Nucleai, Biond Biologics, and Ella Therapeutics, outside the submitted work.

Ethical approval:

Data sharing: Data and fully redacted version of the manuscript are available from the corresponding author (tomermrsn@gmail.com) on reasonable request.

The lead authors (the manuscript's guarantors) affirm that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Dissemination to participants and related patient and public communities: The findings of this study will be shared publicly through lay press coverage, social media, and presentations at virtual and in-person conferences.

Provenance and peer review: Not commissioned; externally peer reviewed.

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

- 1 Gøtzsche PC, Jørgensen AW. Opening up data at the European Medicines Agency. *BMJ* 2011;342:d2686. doi:10.1136/bmj.d2686
- Osipenko L. Audit of data redaction practices in NICE technology appraisals from 1999 to 2019. BMJ Open 2021;11:e051812. doi:10.1136/bmjopen-2021-051812
- 3 Wise J.BEACON trial in colorectal cancer is "single worst reported," says critic. BMJ 2019; 367.
- 4 Marquardsen M, Ogden M, Gøtzsche PC. Redactions in protocols for drug trials: what industry sponsors concealed. J R Soc Med 2018;111:136-41. doi:10.1177/0141076817750554
- 5 von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *Int J Surg* 2014;12:1495-9. doi:10.1016/j.ijsu.2014.07.013
- 6 Elzhov TV, Mullen KM, Spiess A-N, Bolker B, Mullen MKM, Suggests M. Package 'minpack. Im. Title R Interface Levenberg-Marquardt Nonlinear Least-Sq Algorithm Found MINPACK Plus Support Bounds, 2016.
- 7 Mehta CR, Patel NR, Senchaudhuri P. Exact power and samplesize computations for the Cochran-Armitage trend test. Biometrics 1998;54:1615-21. doi:10.2307/2533685.
- 8 Ali A, Rasheed A, Siddiqui AA, Naseer M, Wasim S, Akhtar W. Non-parametric test for ordered medians: The Jonckheere Terpstra test. *Int J Stat Med Res* 2015;4:203. doi:10.6000/1929-6029.2015.04.02.6.
- 9 Ruperto N, Brunner HI, Synoverska O, et al, Paediatric Rheumatology International Trials Organisation (PRINTO) and Pediatric Rheumatology Collaborative Study Group (PRCSG). Tofacitinib in juvenile idiopathic arthritis: a double-blind, placebo-controlled, withdrawal phase 3 randomised trial. *Lancet* 2021;398:1984-96. doi:10.1016/S0140-6736(21)01255-1
- 10 André F, Ciruelos E, Rubovszky G, et al, SOLAR-1 Study Group. Alpelisib for PIK3CA-mutated, hormone receptor—positive advanced breast cancer. N Engl J Med 2019;380:1929-40. doi:10.1056/ NEJMoa1813904
- Bacharier LB, Maspero JF, Katelaris CH, et al, Liberty Asthma VOYAGE Investigators. Dupilumab in children with uncontrolled moderateto-severe asthma. N Engl J Med 2021;385:2230-40. doi:10.1056/ NFIMoa2106567
- 12 Bomze D, Asher N, Ali OH, et al. Survival-inferred fragility index of phase 3 clinical trials evaluating immune checkpoint inhibitors. JAMA Netw Open 2020; 3:e2017675-e.
- 13 Chan A-W, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. IAMA 2004;291:2457-65. doi:10.1001/jama.291.20.2457
- 14 Greenberg L, Jairath V, Pearse R, Kahan BC. Pre-specification of statistical analysis approaches in published clinical trial protocols was inadequate. *J Clin Epidemiol* 2018;101:53-60. doi:10.1016/j. jclinepi.2018.05.023
- 15 Ioannidis JPA. The importance of predefined rules and prespecified statistical analyses: do not abandon significance. JAMA 2019;321:2067-8. doi:10.1001/jama.2019.4582
- 16 Patel CJ, Burford B, Ioannidis JP. Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. J Clin Epidemiol 2015;68:1046-58. doi:10.1016/j.jclinepi.2015.05.029
- 17 European Medicines Agency. External guidance on the implementation of the European Medicines Agency policy on the publication of clinical data for medicinal products for human use. 2016.
- 18 Fundytus A, Wells JC, Sharma S, et al. Industry funding of oncology randomised controlled trials: implications for design, results and interpretation. Clin Oncol (R Coll Radiol) 2022;34:28-35. doi:10.1016/j.clon.2021.08.003

Web appendix: Unredacted version of article **Web appendix:** Supplementary figures

Web appendix: Appendix