ÖSTERREICHISCHE
AGENTUR FÜR
WISSENSCHAFTLICHE
INTEGRITÄT

Commission for
Research Integrity

Chair of the Commission:
Prof. Dr. Philipp Theisohn
Dep. Chair of the Commission
Prof. Dr. Regina Aebi-Müller

Members of the Commission:
Prof. Dr. Katrin Auspurg
Univ.-Prof. Dr. Nikolaus Forgó
Prof. Dr. Joachim Heberle
Prof. Dr. Susanne Modrow
Prof. Dr. Frits R. Rosendaal
Prof. Dr. Kerstin Schneider

Administrative Office:
Sabine Chai, Ph.D.
sabine.chai@oeawi.at

Landstrasser Hauptstrasse 9/21
1030 Vienna
T: +43/1/7106821
www.oeawi.at

PERSONAL / CONFIDENTIAL

To
The Oncologist
c/o Editor in Chief
Susan E. Bates

Columbia University

By e-mail only: seb2227@cumc.columbia.edu
cc: rebecca.simmons@oup.com

Vienna, 1st September 2022

Re: Request for Retraction | A2021-10

Dear Madam,

this is to request retraction of the article:

Homeopathic Treatment as an Add-On Therapy May Improve Quality of Life and Prolong Survival in Patients with Non-Small Cell Lung Cancer: A Prospective, Randomized, Placebo-Controlled, Double-Blind, Three-Arm, Multicenter Study
By Michael Frass, Peter Lechleitner, Christa Gründling, Claudia Pirker, Erwin Grasmuk-Siegl, Julian Domayer, Maximilian Hochmair, Katharina Gaertner, Cornelia Duscheck, Ilse Muchitsch, Christine Marosi, Michael Schumacher, Sabine Zöchbauer-Müller, Raj K. Manchanda, Andrea Schrott, Otto Burghuber
Published in The Oncologist on 7 November 2020 (2020; 25: e1930–e1955, https://doi.org/10.1002/onco.13548)

The Commission for Research Integrity of the Austrian Agency for Research Integrity (OeAWI) has, following a complaint and a request for an investigation by the Vice Rector of the Medical University of Vienna, done an extensive investigation into this publication, and found numerous manipulations that led to the recommendation to retract the article.

ÖSTERREICHISCHE AGENTUR FÜR WISSENSCHAFTLICHE INTEGRITÄT

In the attached document you will find a non-exhaustive list of these manipulations.

Kind regards,

Frits R Rosendaal
Member of the Commission of the OeAWI

Katrin Auspurg
Member of the Commission of the OeAWI

Responsible Case Manager

Prof. Dr. FR Rosendaal
Dep. of Clinical Epidemiology
Leiden University Medical Center
Leiden, the Netherlands

2nd Responsible Case Manager

Prof. Dr. Katrin Auspurg
Methods Chair, Dep. of Sociology
LMU Munich
Munich, Germany

Report on:

Homeopathic Treatment as an Add-On Therapy May Improve Quality of Life and Prolong Survival in Patients with Non-Small Cell Lung Cancer: A Prospective, Randomized, Placebo-Controlled, Double-Blind, Three-Arm, Multicenter Study

By Michael Frass, Peter Lechleitner, Christa Gründling, Claudia Pirker, Erwin Grasmuk-Siegl, Julian Domayer, Maximilian Hochmair, Katharina Gaertner, Cornelia Duscheck, Ilse Muchitsch, Christine Marosi, Michael Schumacher, Sabine Zöchbauer-Müller, Raj K. Manchanda, Andrea Schrott, Otto Burghuber

Sources for the investigation:

1. The published article
2. Audit files received from M. Frass
3. SPSS data file received from M. Frass
4. Information from ClinTrials.gov
5. Raw data on questionnaires
6. Inspection of original questionnaires (scans of answered questionnaires)
7. Inspection of archives of the Medical-Ethics Committee of the University Hospital Vienna

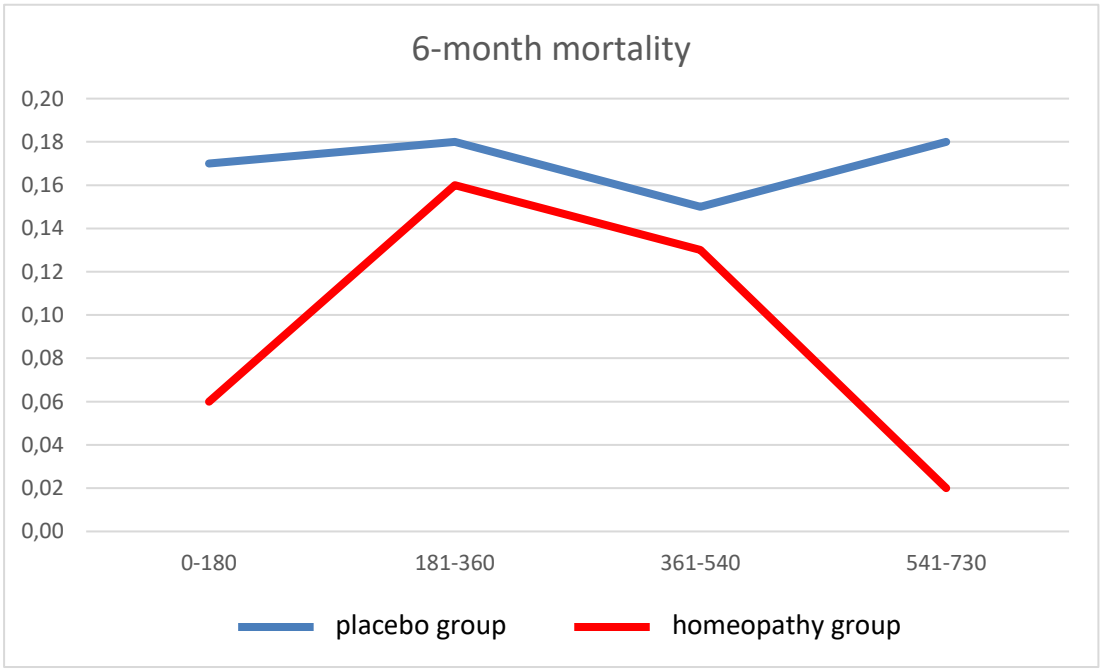
Major findings

- In the publication, the study is reported as a double-blind placebo-controlled trial. However, the study was originally registered (first registered at ClinicalTrials.gov under number NCT01509612 on 13-1-2012, and the last update is listed at 16-6-2021), and approved as an open label study by the local medial-ethics committee on 4 may 2011. The change to a placebo-controlled trial was approved by the Medical Ethics Committee (Medical University of Vienna) in July 2015, when the majority of patients had already been included. Hence, the majority of patients did not receive placebos and were not blinded to their study arm. The presentation of the study as double-blind placebo controlled is untrue.
- Following the report to the Medical-Ethics Committee by the investigators patients have been included starting 2012 through 2017. During this period of time, with the study ongoing, the study has undergone major protocol changes, including changes in the types of cancer to be included (first: advanced glioblastoma, advanced sarcoma, advanced lung cancer; later: only advanced lung cancer) and in the primary endpoint (first: quality of life, later: also survival). These changes are substantial and hardly acceptable once a study is running. Furthermore, they are not mentioned in the publication, nor is there mention of the patients with glioblastoma and sarcoma who

were initially randomized. This is suggestive of data manipulation, and the lack of transparency in the article unacceptable.

- The article states, literally, that there were “numerous exclusion criteria”, and these are listed in the article. Only two of these criteria, i.e., pregnancy and use of other alternative medication, were listed in any of the six versions of the protocol that were submitted to the Medical-Ethics Committee between July 2010 and July 2020. Given the age of the patients and the recency of diagnosis, few if any will have been pregnant or using other alternative medication. The last patient was included in 2017. The protocol with numerous exclusion criteria was submitted to ClinTrials.gov in 2020. Hence, many patients were excluded post-hoc, which is suggestive of data manipulation.
- The survival figures reported in the article can largely be reproduced from the analytical file that was provided to the Committee. This file is clearly a version that was derived from a more original data file – for instance, it includes a variable observation time, which must have been calculated from inclusion and end-of-study dates, which are not in the analysis file. However, the survival data are fully implausible, which becomes evident when we break two-year follow-up in four periods of six months. During both the first and last six-months-period, death risk is much lower in the homeopathy group than in the ‘placebo group’. However, during the intermediate full year, death risks are largely similar for both groups (see Fig. 1). This would imply that a therapy works for six months, then not for a year, and is effective again for following six months. This is highly implausible, whereas these figures are compatible with selective deletion of records.

Figure1: Mortality risk by 6-months periods



- In the audit trails it is indicated, that 21 patients were “rejected post-randomization”. The accompanying dates for these exclusions are all the same date in 2017, while these patients have randomization dates in 2014 and 2015. Thus, they were excluded after they had finished the study, which corroborates the conclusion of selective exclusion (associated with outcome), which implies data manipulation.
- The audit trails, consisting of two separate files, have several curious aspects. First, there is an end-of-study date noted in 2014, when already 75 patients had been entered. Then the study resumes in the second audit trail, where we find patients with the same patient-ID included twice, on different dates. Furthermore, on one day 71 patients were included, according to the time stamp 8.12 and 8.38 hr in the morning. This suggests that patients were entered in these audit trails from other files, at a later date than actual inclusion. Therefore, the audit trails are not reliable.
- In the publication it is stated under the Kaplan-Meier survival graphs “Crosses indicate time points of censoring, i.e., no patient died after this time in this study group”. This is obviously not the correct definition of censoring (which is simply that surviving patients reach the end-of-study-date). However, it is also not true: In fact, an additional nine and seven patients in the homeopathy and placebo group, respectively, died after two years.
- The article mentions “numerous exclusion criteria” but does not list how many patients were in fact excluded. The CONSORT diagram shown in the article states that 106 patients were randomized, and eight were excluded post-randomization because of information on mutational status arriving later. However, from the audit trail it is clear that 120 patients were randomized, and 21 patients were excluded post-randomization. These should have been listed in the CONSORT diagram, even if exclusion had been according to protocol and bona fide.
- As patients were excluded after randomization, the analysis cannot have been ‘intention-to-treat’ as is claimed in the publication.
- In the Quality of Life data there is only one missing data point in a questionnaire of 30 items filled in three times. This is highly implausible. Possibly the authors used imputation, but this is not mentioned in the publication.
- Apart from the point above, the questionnaire data are highly implausible. First, several subjects report only extreme positive values, and there is very little variation between items. Moreover, the patients in the homeopathy group report a quality of life that is much higher than that of the general population known from other surveys. For patients with stage four non-small cell lung cancer this is highly implausible. Furthermore, it is well-known that good self-reported subjective health is associated with longer survival. This is seen in the placebo group, but not in the homeopathy group. These observations are all highly suggestive for data manipulation (see the more detailed analyses in the Appendix).

APPENDIX: Analyses of questionnaire data

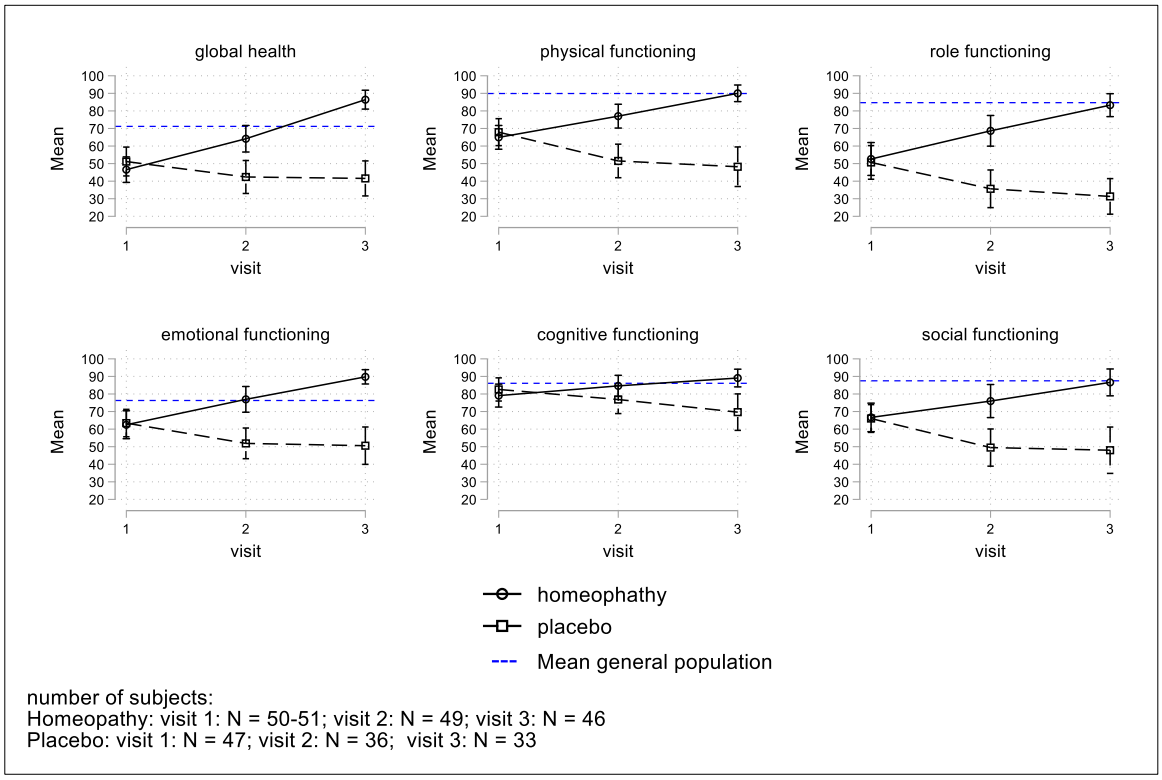
Used abbreviations:

- H = Homeopathy
- P = Placebo
- C = Control

Main Findings

1. There are statistical flaws. For example, authors partly reported one-sided and not two-sided significance tests as indicated in their method section. “HR” in the result tables probably indicate hazard rates; but for comparison of gender distributions, this statistic does not make sense. For the categorized age variable (with regular values from 1-11), there is a wild code in the data (33.4319).
2. Some aspects are not clearly documented. Only from the protocols, one can learn that authors planned to use imputations. But these are not reported in the paper. The low item-nonresponse is probably due to imputations, these should then be reported, however.
3. Many data and results are striking. For example,
    - The patients in the H-group, although suffering from lung cancer stage III/IV, report a health status that is much better than the mean and even 75% percentile of the general population after 18 weeks of treatment (see Figure 1, 2). But, some of these patients died few months later.
    - Five patients in the H-group responded the 30+36 = 66 items in the questionnaires to measure health status / quality of life throughout with the most extreme values indicating a throughout excellent health status (see Figure 3), also many other subjects in this group showed extremely good health ratings. Again, some of these patients died short time after this visit.
    - In the H-group, and only there, health status/Quality of Life (QoL) is not related to survival time. While in the P-group there is the expected pattern, that good health status prolongs survival time. (See Figure 4a, 4b; Table 1). This means that the main mechanism assumed in the paper (“A higher QoL might have contributed to the prolonged survival”, p. e1938) does not hold. It is surprising that authors did not simply test this assumption, and anyway puzzling that no correlation between these different health-related outcome variables is only evident in the H-group.
    - In the H-group, the values measured at visit 3 show much lower variance than in other groups (they are strongly inter-correlated). (See Table 2).
4. After checking the data and observing implausible patterns, we asked for the original questionnaires. These show clear indications of data manipulations, such as corrected dates and information (information was e.g. crossed out and overwritten).

Figure 1. Quality of Life (QoL) scores across visits and treatment groups, EORTC QLQ-C30 instrument (in comparison to the mean value reported for the general population[1])



There are several striking observations in Figure 1:

- The changes over time (visits) are surprisingly linear for the H-group for most outcomes, while the P-group shows more nonlinear changes. Nonlinear patterns can be expected due to diminishing marginal effects and the fixed range of the item scales: Increasing outcome values have to converge to the minimum and maximum score value [0; 100]. (Otherwise, one would expect – if the linear increase continues – values that go beyond the scale range of [0;100]).
- For some outcome measures, the values for the H-group, third visit, are (far) above the mean for the general population (and for global health, the means exceed also the 75% percentile of the general population).
- For the H-group, the confidence intervals shrink from the left to the right (1st to 3rd visit), despite decreasing numbers of observations (while for the P-group, confidence intervals increase as one would expect). This might be caused by ceiling effects, but these are not discussed in the paper. Other explanations are imputations with very low variance.

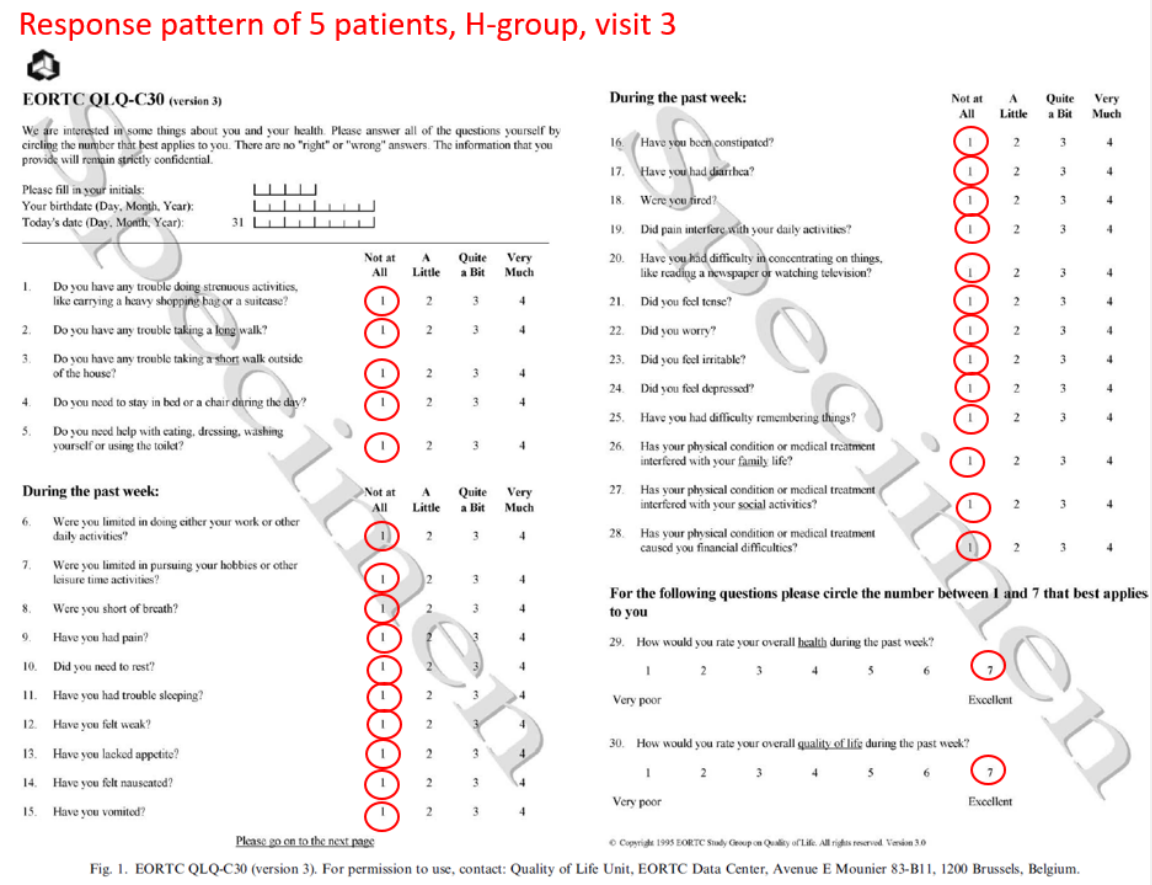
[1] Provided by the EORTC QLQ Quality of Life Group Members and other users of the QLQ-C30, based on 7,136 individuals (Scott et al. 2008).

Figure 2. Distribution of global health score values: reference values provided by the EORTC QLQ-C30Group (left panel) in contrast to the two treatment groups (middle and right panel, measured at time 0 [visit 1], after 9 [visit 2] and 18 weeks [visit 3])

- The panel on the left shows reference values (based on 802 resp. 7,136 subjects) that are provided by the EORTC QLQ-C30 Group for pre-treatment units in clinical trials on lung cancer, stage III/IV resp. the general population.[2] While the baseline distributions in the paper looks similar, there is a huge change in the H-group, with nearly half indicating an “excellent” health status (score 100) at the 3rd measurement (“visit”).

[2] Scott. N. et al. (2008) EORTC QLQ-C30 Reference Values

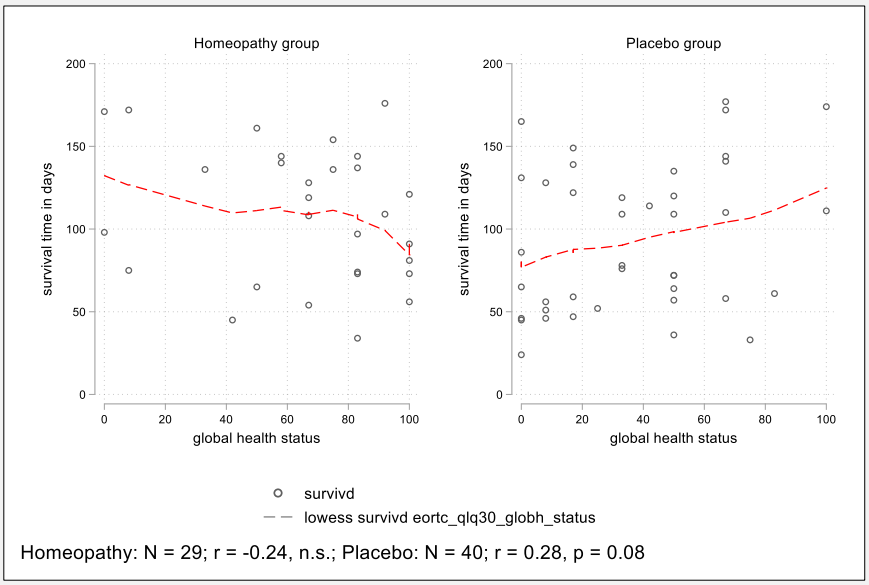
Figure 3. Response patterns of a quarter (N = 5) of the 20 patients measured at visit 3 (The case managers have reconstructed these response patterns from the data)



- Also the other individuals scored high on nearly all health items (rescaled such that high scores indicate a good health status), and indicated only the second best value for few items. (The mean value was 97.5, averaging all 30 items that are scaled [0; 100] for the 20 individuals with maximum global health status).
- Similar striking observations exist for the other instrument, SF36, the authors used to measure individuals' health status / quality of life.
- Several patients who reported maximal health status on all items died within few months afterwards (see Figure 6a, b below), which is highly implausible.

Figure 4a. Relationship between quality of life (QoL) and survival time after the visit at which QoL was measured; only subjects who died within 180 days (half a year) after the visit

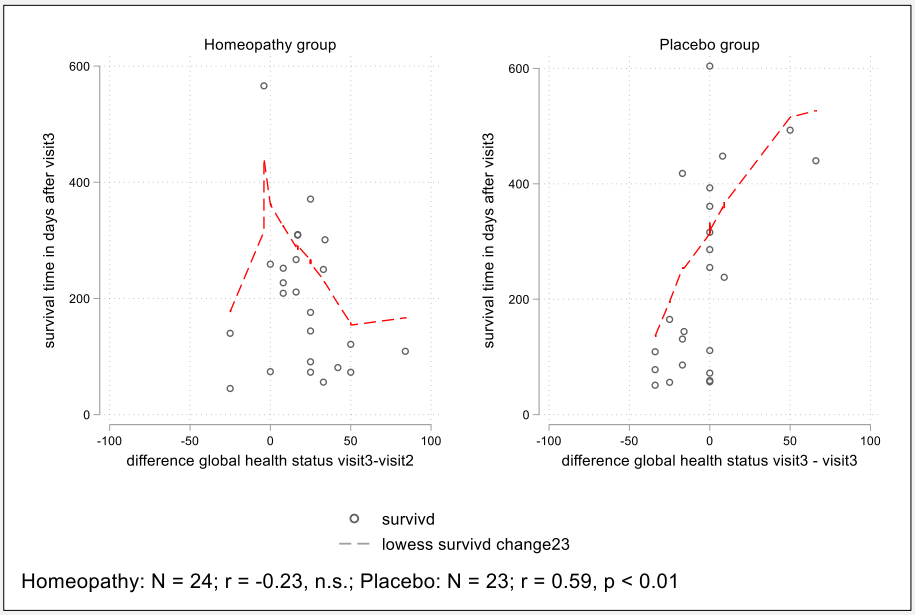
The x-axis shows the score for the global health status scale measured at the 1st, 2nd, and 3rd visit (with higher values indicating a better health status; there are up to three observations for each subject). The y-axis shows the survival time (in days) after the respective visit. The dashed line shows a locally weighted prediction of the survival time based on the health status scores (estimated with the "lowess" command in Stata).



(When using a longer survival time, there is a negative, non-significant association in the H-group, and a positive, non-significant association in the P-group.)

Figure 4b. Relationship between increase in QoL between visit 2 and 3 and survival time after visit 3

The following figure shows the increase in QoL between visit 2 and 3 on the x-axis (zero means there was no change; positive values indicate that there was an increase in QoL between the two visits, and negative values that QoL has worsened); while the y-axis shows the survival time after the second visit. Only individuals who died within an observation period of 730 days were included in these analyses (this observation window was used by the authors).



- Only for the P-group, a good health status or improvement of health status prolongs survival time. For the H-group, there is no association at all (and in tendency, there seems to be even a negative association).
- These striking results also become evident by running a Cox-regression, including censored observations (patients who did not die). Only within the P-group it is able to observe: the better the health status, the longer the survival time.

Table 1. Cox regression of risk of death on H-treatment, physical functioning measured by SF-scale, and their interaction.

```
Cox regression with Breslow method for ties

No. of subjects =     95                                  Number of obs =    252
No. of failures =     74
Time at risk    = 53,299
                                                          LR chi2(3)    =  16.49
Log likelihood = -271.42773                               Prob > chi2   = 0.0009
```

_t	Haz. ratio	Std. err.	z	P>|z|	[95% conf.	interval]
1.homeopathy	.1241236	.1090844	-2.37	0.018	.0221711	.6948991
sf36_pfi	.9828266	.0056403	-3.02	0.003	.9718337	.9939438
homeopathy#c.sf36_pfi						
1	1.025386	.0112246	2.29	0.022	1.003621	1.047624

Homeopathy leads to a lower risk of death / longer survival time (Hazard ratio < 1). But only for the P-group, a good physical function (high sf36_pfi score) lowers the risk of death. These results are also robust when running separate regressions for the two treatment groups: for the P-group, for the sf36_pfi score HR <1, $p < 0.01$; while for the H-group, HR > 1, $p = 0.901$).

- These analyses show that the main and conclusive statement of this paper, “ “homeopathy positively influences both QoL and survival”, is unsustainable.

Table 2. Reliability: Inter-Correlations (N) of both instruments by treatment and visit

	1st visit		2nd visit		3rd visit	
	P	H	P	H	P	H
Global health (sf36_ghp & eortc_qlq30_globh_status)	0.664*** (46)	0.632*** (45)	0.591*** (36)	0.762*** (45)	0.614*** (33)	0.841*** (43)
Physical functioning (sf36_pfi & eortc_qlq30_physfunction	0.889*** (46)	0.827*** (47)	0.712*** (36)	0.893*** (45)	0.895*** (33)	0.881*** (43)
Role functioning (sf36_rolph & eortc_qlq30_role_functioning)	0.581*** (46)	0.575*** (47)	0.364* (36)	0.625*** (45)	0.719*** (33)	0.815*** (43)
Social functioning (sf36_social &eortc_qlq30_social_functioning)	0.703*** (46)	0.610*** (46)	0.671*** (36)	0.625*** (45)	0.840*** (33)	0.739*** (43)
Pain[a] (sf36_pain & eortc_qlq30_pain)	-0.846*** (46)	-0.711*** (46)	-0.842*** (36)	-0.858*** (45)	-0.794*** (33)	-0.922*** (43)

[a] negative correlation – coding error in scale construction?
*** $p < 0.001$; ** $p < 0.01$* $p < 0.05$; two-sided tests.

- The reliability (inter-correlations) of the two different instruments (EORTC QLQ 30 and SF36) is in particular in the H-group higher in the 3rd than 1st or 2nd visit. The most plausible explanation for the scale values becoming much more consistent over time would be manipulations: making imputations with only the maximum value. (see Figure 2, 3).