

Online haemodiafiltration and all-cause mortality: how fragile are the results of the studies published so far?

Yuri Battaglia ^{1,2,*}, Alessandro Mantovani ^{3,*}, Rukshana Shroff ⁴, Gaetano Alfano ⁵, Björn Meijers ⁶, Casper Franssen ⁷, Christian Combe ⁸ and Carlo Basile ⁹; on behalf of the EuDial Working Group of the European Renal Association

¹University of Verona, Department of Medicine, Verona, Italy

²Pederzoli Hospital, Nephrology and Dialysis Unit, Peschiera del Garda, Italy

³Section of Endocrinology, Diabetes and Metabolism, Department of Medicine, University and Azienda Ospedaliera Universitaria Integrata of Verona, Verona, Italy

⁴Renal Unit, UCL Great Ormond Street Hospital and Institute of Child Health, London, UK

⁵Nephrology Dialysis and Transplant Unit, University Hospital of Modena, Modena, Italy

⁶Nephrology Unit, University Hospitals Leuven and Department of Microbiology, Immunology and Organ Transplantation, KU Leuven, Leuven, Belgium

⁷Department of Nephrology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

⁸Department of Nephrology, CHU de Bordeaux and INSERM U1026, University of Bordeaux, Bordeaux, France

⁹Associazione Nefrologica Gabriella Sebastio, Martina Franca, Italy

Correspondence to: Carlo Basile; E-mail: basile.miulli@libero.it

*These authors contributed equally to this work.



To the Editor,

Cardiovascular disease is a major cause of morbidity in patients on dialysis and accounts for almost 50% of deaths in these patients [1]. Although this high mortality is attributable to many reasons, the limited clearance of potentially harmful middle molecules through standard haemodialysis (HD) may be a contributor. Online haemodiafiltration (HDF) augments the clearance of middle molecular weight uraemic toxins through convective clearance while standard HD treatment is primarily based on diffusion [1].

In order to assess the potential superiority of online HDF versus HD on hard outcomes, such as all-cause mortality, we conducted a meta-analysis. Five randomized control trials (RCTs) were identified through a systematic review of three electronic databases (PubMed, Scopus, Web of Science) from inception to 1 October 2023 [2–6]. The main data of these RCTs are summarized in Table 1. The five RCTs enrolled 4143 adult patients on HD. Of these, 2078 were randomly assigned to online HDF and 2065 to high- or low-flux HD. The mean follow-up period was 2.3 years [standard deviation (SD) 0.6]. More than half of the patients were male (62.8%), 31.3% had diabetes mellitus and 83.7% had an arteriovenous fistula. An overall estimate of the effect size [expressed as the odds ratio (OR) given a similar length of follow-up across the included RCTs] was calculated using both a common effects model (Mantel–Haenszel method) [7] and a random effects model (DerSimonian–Laird method) [8] (Fig. 1). Statistical heterogeneity was assessed by the I^2 statistic [9]. Given the small number of RCTs included in this meta-analysis, we did not evaluate publi-

cation bias using a funnel plot [7, 10]. All statistical tests were two sided and used a significance level of $P < .05$. We used Stata version 17.0 (StataCorp, College Station, TX, USA) and its meta-analysis package and R version 4.2.2 (R Foundation for Statistical Computing, Vienna, Austria) for all statistical analyses.

A statistically significant reduction in all-cause mortality of the HDF group when compared with the HD group was shown [random effects OR 0.76 (95% confidence interval 0.65–0.88); $I^2 = 0\%$] (Fig. 1).

In 2014, Walsh *et al.* [11] introduced an ancillary tool that quantified the fragility (or robustness) of findings that report statistically significant P -values: they called this the Fragility Index (FI). The FI is a crucial statistical metric, easy to interpret, primarily employed to assess the robustness of research findings for dichotomous outcomes. It represents the minimum number of participants whose status needs to change from an ‘event’ to a ‘non-event’ (or vice versa), thus the results switch from being statistically significant to non-significant [11]. Although there is not an established cut-off, a higher FI score can be considered indicative of more robust results [12]. To mitigate the potential effect of the sample size on the FI, the Fragility Quotient (FQ) has been proposed and is calculated as follows: $FQ = FI/\text{sample size}$ [13].

There is no consensus on a cut-off value to indicate whether a study is fragile or robust. However, empiric observation suggests that an $FQ \leq 0.03$ raises a concern about the fragility of the findings [13]. Based on this background, we calculated the FI and FQ of the RCTs included in the meta-analysis. Table 1 reports the FI and FQ for the outcome ‘all-cause mortality’ for the two RCTs in which

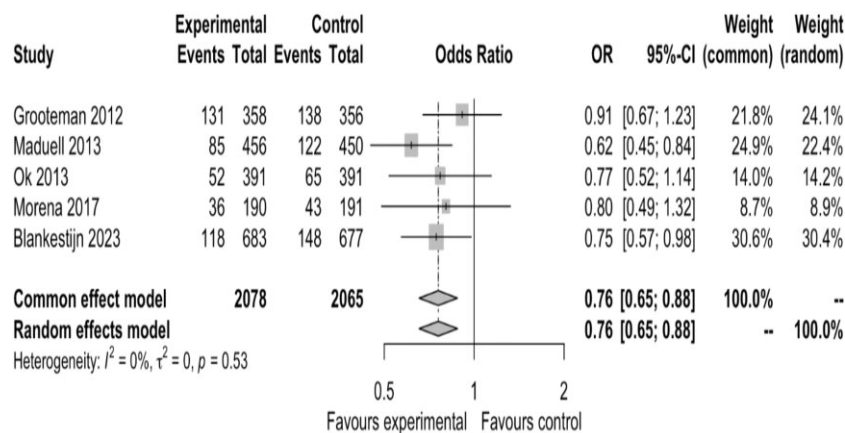
Received: October 24, 2023; Editorial decision: December 24, 2023

© The Author(s) 2024. Published by Oxford University Press on behalf of the ERA. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Table 1: Main data of the RCTs included in the meta-analysis

Author, year [reference]	Intervention versus control arms	Sample characteristics ^a	Trial duration (years), mean (SD)	FI	FQ	SIFI
Grooteman et al., 2012 [2]	Online HDF versus low-flux HD	714 patients (358 HDF, 356 HD) Age: HDF 64.1 (14.0), HD 64.0 (13.4)	3.3 (4.6)	NA	NA	NA
Maduell et al., 2013 [3]	Online HDF versus HD	906 patients (450 HDF, 456 HD) Age: HDF 64.5 (14.4), HD: 66.31 (4.3)	2.0 (1.6)	13	0.014	81
Ok et al., 2013 [4]	Online HDF versus high-flux HD	782 patients (391 HDF, 391 HD) Age: HDF 56.4 (13.0), HD: 56.5 (14.9)	1.9 (0.9)	NA	NA	NA
Morena et al., 2017 [5]	Online HDF versus high-flux HD	381 patients (190 HDF, 191 HD) Age: HDF 76.35 (6.1), HD 76.11 (6.7)	2.0 (0.5)	NA	NA	NA
Blankestijn et al., 2023 [6]	High-dose HDF versus high-flux HD	1360 patients (683 HDF, 677 HD) Age: HDF 62.5 (13.5), HD: 62.3 (13.5)	2.7 (0.7)	3	0.002	18

NA: not available.

^aAge (in years) presented as mean (SD).**Figure 1:** The forest plot and pooled estimates of the effect of online HDF (comparator arm) on all-cause mortality compared with HD (control arm).

the results were statistically significant [3, 6]. Both scores of the study by Maduell et al. were relatively low (FI = 13, FQ = 0.014) [3]. Those of the CONVINC trial (NTR7138), recently published in the *New England Journal of Medicine*, were also low (FI = 3, FQ = 0.002) [6]. This means that, for CONVINC, just three participants, i.e. 0.2% of the study population, would be required to change the results from statistically significant to non-significant. An FI was also calculated for the outcome 'all-cause mortality' for the five RCTs pooled together [14]: it was 44. This high value of FI underscores the robustness of the collective data of the RCTs. When normalized by the FQ (0.01), it appears to be influenced by the large sample size of the meta-analysis.

It must be underlined that the FI has some mathematical limitations [15, 16], including its dependence on the underlying risk, sample size, binary outcomes and use of Fisher's exact test. These factors could partly influence accurate calculation of the FI for time-to-event data [17]. In order to address this limit, a survival-inferred fragility index (SIFI) was recently developed [17]. Specifically, the SIFI is defined as the minimum number of reassignments of the best survivors (defined as the patients with the longest follow-up time, regardless of having an event or being censored; the worst survivors were defined as the patients with the earliest

event) from the experimental group to the control group, resulting in a loss of significance (in this specific case, defined as $\alpha = 0.05$ using the logrank test) [17]. Thus the SIFI can provide a measure that both accounts for events over time and retains all the advantages of the original FI. To calculate the SIFI, we first digitized the Kaplan–Meier curves in each eligible study using Digitizelt software (<http://www.digitizeit.de/>). Subsequently, with a specific code written for R software (<https://github.com/davidbomze/SIFI>), we calculated the SIFI. As reported in Table 1, the SIFI in the CONVINC trial was 18 [6], whereas the SIFI in the study by Maduell et al. [3] was 81. In other words, the statistical significance of the CONVINC trial might be lost with a change in assignment of 18 patients, representing only 1.3% of the entire population included in the RCT.

It is worth noting that Kampman et al. [18] recently published a study aimed at determining whether the statistical fragility of RCTs in high-impact journals has improved over the last decade. The FI was calculated for all eligible RCTs published from 2014 to 2021 in the *New England Journal of Medicine*, *The Lancet*, the *Journal of the American Medical Association*, the *British Medical Journal* and the *Annals of Internal Medicine*. Of 2544 screened RCTs, 643 were eligible for the FI analysis. These RCTs had a median sample size

of 625, a median FI of 12 and a median FQ of 0.015. This represents an improvement in robustness when compared with the median FI of 8 of RCTs published a decade earlier in the same five journals ($P < .001$) [11]. The study's conclusions indicate that over the last decade, the median statistical robustness of RCTs published in high-impact journals has improved. Yet the unchanged lower bound of the interquartile range reveals that statistical significance in 25% of trials is still dependent on three or fewer events [18]. Surprisingly, despite the popularity of the FI as a review subject, Kampman *et al.* [18] found only one trial in their database of 2544 RCTs in high-impact journals that mentioned this metric.

Regarding the SIFI, Bomze *et al.* [17] found that 45 phase 3 RCTs evaluating immune checkpoint inhibitors had a low SIFI, for overall survival, resulting in a loss of statistical significance with the reassignment of a small fraction of the entire sample size.

In conclusion, our study shows that the results of the RCTs and meta-analysis reporting a statistically significant reduction in the risk of all-cause mortality in patients undergoing online HDF compared with HD were not very robust [3, 6]. The FI (for binary outcomes) and the SIFI (for time-to-event data) provide additional means of assessing and communicating the strength of statistical conclusions. They should be employed as a valuable supplement to reported *P*-values in RCTs, helping readers look beyond merely reaching statistical significance.

ACKNOWLEDGEMENTS

The EuDial Working Group is a part of the European Renal Association.

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

DATA AVAILABILITY STATEMENT

The data underlying this article will be shared upon reasonable request to the corresponding author.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

REFERENCES

- Shroff R, Basile C, van der Sande F *et al.* EuDial Working Group of the European Renal Association. Haemodiafiltration for all: are we convinced? *Nephrol Dial Transplant* 2023;**38**:2663–5. <https://doi.org/10.1093/ndt/gfad136>
- Grooteman MP, van den Dorpel MA, Bots ML *et al.* Effect of online hemodiafiltration on all-cause mortality and cardiovascular outcomes. *J Am Soc Nephrol* 2012;**23**:1087–96. <https://doi.org/10.1681/ASN.2011121140>
- Maduell F, Moreso F, Pons M *et al.* High-efficiency postdilution online hemodiafiltration reduces all-cause mortality in hemodialysis patients. *J Am Soc Nephrol* 2013;**24**:487–97. <https://doi.org/10.1681/ASN.2012080875>
- Ok E, Asci G, Toz H *et al.* Mortality and cardiovascular events in online haemodiafiltration (OL-HDF) compared with high-flux dialysis: results from the Turkish OL-HDF study. *Nephrol Dial Transplant* 2013;**28**:192–202. <https://doi.org/10.1093/ndt/gfs407>
- Morena M, Jaussent A, Chalabi L *et al.* Treatment tolerance and patient-reported outcomes favor online hemodiafiltration compared to high-flux hemodialysis in the elderly. *Kidney Int* 2017;**91**:1495–509. <https://doi.org/10.1016/j.kint.2017.01.013>
- Blankestijn PJ, Vernooij RWM, Hockham C *et al.* Effect of hemodiafiltration or hemodialysis on mortality in kidney failure. *N Engl J Med* 2023;**389**:700–9. <https://doi.org/10.1056/NEJMoa2304820>
- Cumpston M, Li T, Page MJ *et al.* Updated guidance for trusted systematic reviews: a new edition of the Cochrane Handbook for Systematic Reviews of Interventions. *Cochrane Database Syst Rev* 2019;**10**:ED000142. <https://doi.org/10.1002/14651858.ED000142>
- DerSimonian R, Kacker R. Random-effects model for meta-analysis of clinical trials: an update. *Contemp Clin Trials* 2007;**28**:105–14. <https://doi.org/10.1016/j.cct.2006.04.004>
- Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;**21**:1539–58. <https://doi.org/10.1002/sim.1186>
- Egger M, Smith GD, Phillips AN. Meta-analysis: principles and procedures. *Br Med J* 1997;**315**:1533–7.
- Walsh M, Srinathan SK, McAuley DF *et al.* The statistical significance of randomized controlled trial results is frequently fragile: a case for a fragility Index. *J Clin Epidemiol* 2014;**67**:622–8. <https://doi.org/10.1016/j.jclinepi.2013.10.019>
- Andrade C. The use and limitations of the fragility index in the interpretation of clinical trial findings. *J Clin Psychiatry* 2020;**81**:20f13334. <https://doi.org/10.4088/JCP.20f13334>
- Heston TF. The Robustness Index: going beyond statistical significance by quantifying fragility. *Cureus* 2023;**15**:e44397.
- Atal I, Porcher R, Boutron I *et al.* The statistical significance of meta-analyses is frequently fragile: definition of a fragility index for meta-analyses. *J Clin Epidemiol* 2019;**111**:32–40. <https://doi.org/10.1016/j.jclinepi.2019.03.012>
- Potter GE. Dismantling the Fragility Index: a demonstration of statistical reasoning. *Stat Med* 2020;**39**:3720–31. <https://doi.org/10.1002/sim.8689>
- Baer BR, Gaudino M, Charlson M *et al.* Fragility indices for only sufficiently likely modifications. *Proc Natl Acad Sci USA* 2021;**118**:e2105254118. <https://doi.org/10.1073/pnas.2105254118>
- Bomze D, Asher N, Hasan Ali O *et al.* Survival-inferred fragility index of phase 3 clinical trials evaluating immune checkpoint inhibitors. *JAMA Netw Open* 2020;**3**:e2017675. <https://doi.org/10.1001/jamanetworkopen.2020.17675>
- Kampman JM, Turgman O, Sperma Weiland NH *et al.* Statistical robustness of randomized controlled trials in high-impact journals has improved but was low across medical specialties. *J Clin Epidemiol* 2022;**150**:165–70. <https://doi.org/10.1016/j.jclinepi.2022.07.001>