Extracorporeal Membrane Oxygenation for Severe Acute Respiratory Distress Syndrome and Posterior Probability of Mortality Benefit in a Post Hoc Bayesian Analysis of a Randomized Clinical Trial

Ewan C Goligher, George Tomlinson, David Hajage, Duminda N Wijeysundera, Eddy Fan, Peter Jüni, Daniel Brodie, Arthur Slutsky, Alain Combes

1 **Venovenous Extracorporeal Membrane Oxygenation for Severe Acute Respiratory Distress**

2 **Syndrome: A Post-Hoc Bayesian Analysis of a Randomized Clinical Trial**

3

4 Ewan C. Goligher MD PhD[1,2,3]

5 George Tomlinson PhD[2,3,4]

6 David Hajage PhD[5]

7 Duminda N. Wijeysundera MD PhD[4,6,7,8]

8 Eddy Fan MD PhD[1,2,4]

9 Peter Jüni MD[4,8]

10 Daniel Brodie MD[9,10]

11 Arthur S. Slutsky MD[1,11]

12 Alain Combes MD PhD[12,13]

13

14 **Institutional Affiliations**

15   1. Interdepartmental Division of Critical Care Medicine, University of Toronto, Toronto, Canada

16   2. Department of Medicine, University Health Network, Toronto, Canada

17   3. Toronto General Hospital Research Institute, Toronto, Canada

18   4. Institute for Health Policy, Management, and Evaluation, University of Toronto, Toronto, Canada

19   5. Sorbonne Université, Département Biostatistique Santé Publique et Information Médicale, Unité

20   de Recherche Clinique, Centre de Pharmacoépidémiologie (Cephepi), AP-HP, Hôpitaux

21   Universitaires Pitié Salpêtrière – Charles Foix, F-75013 Paris, France

22   6. Department of Anesthesia, University of Toronto, Toronto, Canada

23   7. Department of Anesthesia and Pain Management, Toronto General Hospital, Toronto, Canada

24   8. Applied Health Research Centre of the Li Ka Shing Knowledge Institute, St. Michael's Hospital,

25   Department of Medicine, University of Toronto, Toronto, Canada

9. Division of Pulmonary, Allergy, and Critical Care Medicine, Columbia University Medical
    Center, New York, NY

10. New York-Presbyterian Hospital, New York, NY

11. Keenan Centre for Biomedical Research, Li Ka Shing Knowledge Institute, St. Michael's
    Hospital, Toronto, Canada

12. Sorbonne Université INSERM Unité Mixte de Recherche (UMRS) 1166, Institute of
    Cardiometabolism and Nutrition, Paris, France

13. Service de Médecine Intensive-Réanimation, Institute de Cardiologie, Assistance Publique-
    Hôpitaux de Paris (APHP) Hôpital Pitié-Salpêtrière, Paris, France

Date of Revision:

September 20, 2018

Word Count:

3,560

Corresponding Author:

Dr. Ewan Goligher

Toronto General Hospital

585 University Ave.

Peter Munk Building, 11th Floor Room 192

Toronto ON, M5G 2N2

+1 (416) 340-4800 ext. 6810

ewan.goligher@utoronto.ca

2

52 **Key Points** (Word Count = 119)

53 *Question*

54 Can Bayesian analysis clarify the interpretation of clinical trial results?

55

56 *Findings*

57 In a *post hoc* Bayesian analysis of the recent EOLIA (ECMO to Rescue Acute Lung Injury) trial, the

58 posterior probability of mortality benefit (relative risk<1) ranged between 88% and 99% given a range of

59 prior assumptions reflecting varying degrees of skepticism and enthusiasm regarding previous evidence

60 for the benefit of ECMO. Probabilities varied according to the definition of minimum clinically important

61 mortality benefit; for example, the posterior probability of relative risk <0.67 ranged between 0% to 48%

62 given the same range of prior assumptions.

63

64 *Meaning*

65 Information about the posterior probability of treatment effect provided by Bayesian analysis may help to

66 clarify the interpretation of clinical trial findings.

67   **Abstract** (Word Count = 445)

68   *Importance*

69   Bayesian analysis of clinical trial data may provide useful information to aid in study interpretation,

70   especially when trial evidence suggests that the benefits of an intervention are uncertain, such as in a trial

71   that evaluated early extracorporeal membrane oxygenation (ECMO) for severe acute respiratory distress

72   syndrome (ARDS).

73

74   *Objective*

75   To demonstrate the potential utility of Bayesian analyses by estimating the posterior probability, under

76   various assumptions, that early ECMO was associated with reduced mortality in patients with very severe

77   ARDS in a recent randomized trial.

78

79   *Design and Evidence*

80   A *post hoc* Bayesian analysis of data from a randomized clinical trial (ECMO to Rescue Acute Lung

81   Injury, EOLIA) that included 249 patients with very severe ARDS who had been randomized to receive

82   early ECMO (n=124; mortality at 60 days, 35%) versus initial conventional lung-protective ventilation

83   with the option for rescue ECMO (n=125, mortality at 60 days, 46%). Statistical prior distributions were

84   specified to represent varying levels of pre-existing enthusiasm or skepticism for ECMO and by Bayesian

85   meta-analysis of previously published studies (with downweighting to account for differences between

86   studies). The relative risk (RR), credible interval (CrI), absolute risk reduction (ARR), and probability of

87   clinically important mortality benefit (varying from RR<1 to RR<0.67 and ARR from ≥2% to ≥20%)

88   were estimated with Bayesian modelling.

89

90   *Findings*

91   Combining a minimally informative prior distribution with the findings of EOLIA, the posterior

92   probability of RR < 1 for mortality at 60 days after randomization was 96% (RR 0.78, 95% CrI 0.56-

93 1.04); the posterior probability of RR<0.67 was 18%, the probability of ARR≥2% was 92%, and the

94 probability of ARR≥20% was 2%. With a moderately enthusiastic prior, equivalent to information from a

95 trial of 264 patients with an RR of 0.78, the estimated RR was 0.78 (95% CrI 0.63-0.96), the probability

96 of RR<1 was 99%, the probability of RR<0.67 was 8%, the probability of ARR≥2% was 97%, and the

97 probability of ARR≥20% was 0%. With a strongly skeptical prior, equivalent to information from a trial

98 of 264 patients with an RR of 1.0, the estimated RR was 0.88 (95% CrI 0.71-1.09), the probability of

99 RR<1 was 88%, the probability of RR<0.67 was 0%, the probability of ARR≥2% was 78%, and the

100 probability of ARR≥20% was 0%. If the prior was informed by previous studies, the estimated RR was

101 0.71 (95% CrI 0.55-0.94), the probability of RR<1 was 99%, the probability of RR<0.67 was 48%, the

102 probability of ARR≥2% was 98%, and the probability of ARR≥20% was 4%.

103

104 *Conclusion*

105 *Post hoc* Bayesian analysis of data from a randomized trial of early ECMO compared with conventional

106 lung-protective ventilation with the option for rescue ECMO among patients with very severe ARDS

107 provides information about the posterior probability of mortality benefit under a broad set of assumptions

108 that may help inform interpretation of the study findings.

109

110 *Trial Registration – this analysis was NOT registered*

111

Introduction

The conventional frequentist approach to statistical analysis of clinical trials evaluates study hypotheses *indirectly* by estimating the probability that *data* as or more extreme than the observed treatment effect size would be obtained if the null hypothesis (which generally assumes that there is no treatment effect) was true—the goal of frequentist analysis is to determine whether the evidence leads one to confidently reject the null hypothesis. In Bayesian analysis, information available prior to the trial about plausible range of values of the treatment effect (represented as a probability distribution) is updated by the data collected in the trial to produce a revised estimate of the plausible range of values of the treatment effect.[1] Bayesian analysis informs clinical decisions by *directly* estimating the probability of a hypothesized treatment effect given the observed data. [2,3] In addition, because information about treatment effect from pre-existing clinical and biological evidence is formally incorporated into statistical evaluation, Bayesian methods explicitly quantify the otherwise implicit influence of clinical judgment and prior beliefs on the interpretation of trial results.[4-6]

A recent randomized trial of extracorporeal membrane oxygenation (ECMO to Rescue Lung Injury in Severe ARDS—EOLIA)[7] offers an example of the value of Bayesian analysis. In this trial, the effect of early ECMO on mortality in very severe ARDS did not reach statistical significance ($p=0.09$ in the primary analysis). However, the clinically important point estimate of the absolute risk difference (11%), the near statistical significance of the effect despite early stopping for futility, and the wide divergence of pre-existing views regarding the benefit of ECMO[8,9] (due in part to differences between prior studies and their potential methodological limitations) have made interpretation of the trial controversial.[10-12] In this Special Communication, a *post hoc* Bayesian analysis of this trial demonstrating the potential utility of the Bayesian approach is presented.

Methods

EOLIA was a multicenter international randomized clinical trial designed to test the hypothesis that early venovenous ECMO reduces 60-day mortality in patients with very severe forms of ARDS

6

138   (PaO2/FiO2 < 50 mm Hg for >3 hours; or PaO2/FiO2 < 80 mm Hg for >6 hours; or pH<7.25 and

139   PaCO2≥60 mm Hg with a maximum plateau pressure of 32 cm H_2O and respiratory rate set at 35 breaths

140   per minute for ≥6 hours).[13] The trial was designed to detect a decrease in mortality risk from 60% to 40%

141   (absolute risk reduction [ARR] of 20%, relative risk [RR] of 0.67). The trial received ethical approval

142   from the ethics committees at all participating sites.

143   This article presents a previously unplanned re-analysis of the primary pre-specified end-point

144   conducted using Bayesian methods. The aim was to estimate the posterior probabilities that the treatment

145   effect exceeded a range of potential values for the minimum clinically important treatment effect (RR<1,

146   RR<0.9, RR<0.8, RR<0.67; and ARR≥2%, ARR≥4%, ARR≥6%, ARR≥8%, ARR≥10%, and ARR≥20%

147   assuming a baseline mortality risk of 46% based on the EOLIA control group). This range of possible

148   values for the minimum clinically important treatment effect was established from several considerations.

149   First, because the null hypothesis under frequentist conventions in the trial was 'no benefit' (RR=1), we

150   estimated the probability of any mortality benefit (RR<1). Second, we deemed ARR values of 2% to be a

151   reasonable potential minimum clinically important effect as this would be equivalent to an estimated 500

152   lives saved every year in the United States (assuming approximately 25,000 cases of very severe ARDS

153   annually in the United States based on a population of 328 million persons,[14] an annual incidence of

154   ARDS of 80/100,000 population,[15] and a prevalence of very severe ARDS of approximately 10% among

155   all cases of ARDS[16]). However, arguments can be made supporting a lower RR or larger ARR as a

156   minimal clinically important difference, and the trial was designed to detect an RR<0.67 and an

157   ARR≥20%; therefore the posterior probabilities across a range of effect sizes were computed.

158   Bayesian analysis represents one's prior beliefs about the plausible range of values for treatment

159   effect as a probability density distribution. The width (variance) of this distribution represents the

160   confidence in the treatment effect while the area under the distribution at any given value represents the

161   probability that the treatment effect is greater than or equal to that value (see **Figure 1** for examples).

162   Two approaches were used to develop prior statistical priors for this analysis. First, priors were used to

163   reflect varying degrees of enthusiasm and skepticism for the benefit of ECMO before the trial. A

164   minimally informative prior (which regards all possible log-relative risk values to be equally likely) was

165   used to produce results essentially dependent on data from the trial alone; this prior adds minimal

166   information to the trial in calculating posterior probabilities.

167   A range of reference priors were defined to represent "strongly enthusiastic", "moderately

168   enthusiastic", "skeptical", and "strongly skeptical" archetypes of prior belief about the probability of

169   benefit from early ECMO consistent with pre-existing controversy amongst experts in the field[8,17] (**Table**

170   **1**). Each prior distribution was characterized by a different assumed value for median RR (the value for

171   RR that an enthusiast or skeptic would assume to have a 50% probability of obtaining) and a different

172   width (variance, representing the magnitude of uncertainty about the plausible range of values for

173   treatment effect). To aid in understanding the strength of the enthusiasm or skepticism represented by

174   these theoretical priors, the sample size and observed RR were computed for a hypothetical clinical trial

175   achieving the same level of certainty in the treatment effect as each prior. This sample size was computed

176   by comparing the variance of each prior distribution to the variance of the log-relative risk observed in the

177   trial (**Table 1**).

178   In accordance with previously published recommendations,[9,16] the priors were defined so as to

179   represent enthusiastic or skeptical viewpoints with respect to (a) the probability that the true effect of

180   ECMO on mortality is the same or greater than that used to power the trial (i.e. RR≤0.67) or the effect

181   observed in the ARDSNet trial of low tidal volume ventilation (a classic trial in the treatment of ARDS,

182   RR≤0.78)[18] and (b) the probability that ECMO would worsen mortality (i.e. RR>1). Reference priors

183   specified on this basis are described in detail in **Table 1**. **Figure 1A** depicts the probability density

184   distribution for RR specified by each reference prior distribution.

185   Second, data-derived prior distributions were developed based on relevant studies[19-21] from a meta-

186   analysis of ECMO for ARDS.[22] The treatment effects in these previous studies were combined with the

187   observed data from this trial in a Bayesian hierarchical random effects model (that itself used non-

188   informative priors). In effect, the previous studies generated a prior for what the treatment effect in the

189   "next" study would be, a prior that is combined with data from this trial to produce an updated

190   distribution of the estimated treatment effect after this trial. To reflect concerns about possible differences

191   between the current and prior studies (e.g., non-randomized design in two studies, confounding by

192   transfer to specialist centers, suboptimal control group management), the variance of the previous studies

193   was inflated so that patients in pre-existing studies were "downweighted" to exert less influence (i.e.

194   received less weight in the analysis) on the pooled estimate of effect. Downweighting was applied to

195   varying degrees so that patients in previous studies exerted between 0% and 100% of the weight of

196   patients enrolled in the trial. It allowed us to mathematically represent the uncertainty about the estimates

197   of effect in studies given their likely differences (methodological limitations?). The effects and level of

198   uncertainty described by the data-derived priors are represented graphically in **Figure 1B**.

199   Each prior distribution for the log relative risk in the trial was included in a Bayesian model which

200   specified independent binomial sampling of the numbers of deaths in the ECMO and control groups [AU:

201   This phrase is confusing and it is unclear if the 46% mortality in the control group was held

202   constant or, instead, the data for the control group was resampled in some way.] and a uniform

203   prior on the control group risk of mortality. Markov Chain Monte Carlo modelling (with 3 chains, 20,000

204   iterations burn-in and 20,000 saved iterations per chain) was used to derive treatment effect estimates and

205   95% credible intervals (CrI) from the median, 2.5th and 97.5th percentiles of the posterior distribution, and

206   to estimate the posterior probabilities of treatment effects exceeding certain thresholds. The Gelman-

207   Rubin statistic was used to assess convergence of all models. All analyses were conducted in R (www.r-

208   project.org, Version 3.5.0) using *R2jags*[23] to run JAGS.[24]

209

210   **Results**

211   *Bayesian Analysis Using a Minimally Informative Prior*

212   Posterior probabilities of relative and absolute risk reductions in mortality for a range of priors are

213   shown in **Table 2** and **Table 3**. **Figure 2** presents both the likelihood function for the trial and the

214   posterior probability distribution for relative risk reductions for each prior. With the non-informative

215   prior, the estimated median relative risk for mortality at 60 days with early ECMO was 0.78 (95%

216   credible interval, CrI, 0.56-1.04). The posterior probability of mortality benefit with early ECMO (i.e.

217   RR<1) was 96%, the probability of RR<0.67 was 18%. Assuming a baseline mortality risk of 46%, the

218   probability of ARR≥2% was 92%, and the probability of ARR≥20% was 2% (**Table 3**).

219

220   *Bayesian Analysis Using Reference Priors*

221   The posterior probability of RR<1 exceeded 90% across the strongly enthusiastic, moderately

222   enthusiastic, and skeptical priors (**Table 2, Figure 2**). In the most extreme case of a strongly skeptical

223   prior the estimated RR was 0.88 (95% CrI 0.71-1.09), the posterior probability of RR<1 was 88%, the

224   probability of RR<0.67 was 0%, the probability of ARR≥2% was 78%, and the probability of ARR≥20%

225   was 0%.

226

227   *Bayesian Analysis Using the Data-Derived Prior*

228   When combining treatment effects from previous studies with the data from the trial in the

229   hierarchical model, estimated relative risk in the trial was 0.71 (95% CrI 0.55-0.94). With this prior, the

230   posterior probability of RR<1 was 99%, probability of RR<0.67 was 48%, the probability of ARR≥2%

231   was 98%, and the probability of ARR≥20% was 4%.

232   When the previous studies were downweighted to account for their likely differences

233   (methodological limitations?) by up to 90%, the upper limit of the 95% credible interval for treatment

234   effect fell below 1 and the probability of RR<1 exceeded 90% (**Figure 3**). The probability of RR<0.67

235   and ARR≥20% remained low across the range of downweighting (**Table 2** and **Figure 3**).

236

237   **Discussion**

238   Bayesian analysis constitutes an alternative to the conventional paradigm for the statistical

239   evaluation of medical hypotheses. Rather than estimating the probability of the *data* given the hypothesis,

240   it aims to estimate the probability of the *hypothesis* given the data. Statisticians have long identified either

241   as "Bayesians" or as "frequentists"; [2] the debate turns in part on the role of deductive vs. inductive

242   inference in scientific reasoning. [25] Many statisticians have advocated for the incorporation of Bayesian

243   analysis in trial design and interpretation to complement frequentist analysis but adoption in clinical

244   research has been limited. Recently, the United States Food & Drug Administration developed guidelines

245   for the application of Bayesian statistics in trial design and interpretation in clinical trials of medical

246   devices. [26] Bayesian analysis may suggest differing conclusions from frequentist analysis, particularly

247   when observed effect sizes are relatively large but statistical power is relatively low. [3]

248   In the original description of the trial, the investigators concluded that "early application of ECMO

249   was not associated with mortality at 60 days that was significantly lower than that in the control group."

250   This conclusion appropriately reflects the frequentist approach to hypothesis testing. The probability of

251   observing an absolute mortality difference of ≥11% under the null hypothesis of no treatment effect was

252   not sufficiently low to warrant rejection of the null hypothesis according to frequentist conventions (RR

253   0.76, 95% CI 0.55-1.04, p=0.09 in the primary analysis). This conclusion may be at variance with clinical

254   and scientific intuition as it discounts altogether the clinically relevant effect size and a 95% confidence

255   interval that lies mostly below 1. The difficulty of interpreting the results of this frequentist analysis was

256   immediately evident with one editorial concluding that "the routine use of ECMO in patients with severe

257   ARDS is not superior to the use of ECMO as a rescue maneuver"[11] while another suggested that "ECMO

258   probably has some benefit in this context."[27]

259   The statement that ECMO probably has some benefit is an intuitive expression of the Bayesian

260   approach to data analysis. The Bayesian framework aims to define the probability of a desired treatment

261   effect rather than to rule out the absence of treatment effect. Bayesian analysis of the EOLIA trial

262   demonstrates that across a range of prior assumptions about the probability of benefit from early ECMO,

263   the posterior probability of any mortality benefit (RR<1) with early ECMO is high, ranging between 88%

264   to 99%. The influence of priors on the posterior probability varied with the definition of treatment effect,

265   particularly for absolute risk reduction. For an absolute risk reduction of ≥2%, the posterior probability of

266   benefit ranged between 78% and 98%, depending on the prior. For an absolute risk reduction of ≥20%,

267   the posterior probability ranged between 0%-2%.

268   The analyses described here highlight several advantages of the Bayesian framework. First, the use

269   of statistical priors permits the wide spectrum of opinion within the clinical community regarding any

270   treatment to be formally incorporated in the analysis. This is particularly important with ECMO. In a

271   Bayesian analysis of a previous clinical trial of ECMO in children published in 1989,[20] Kass and

272   Greenhouse observed that "diverse opinions among knowledgeable and thoughtful observers arise

273   because (…) different people attach different degrees of importance to various pieces of information

274   concerning the merits of the treatment."[28] By incorporating these varying background beliefs as priors,

275   Bayesian analysis can quantify the overall strength of evidence in support of a hypothesis, complementing

276   conventional frequentist approaches to hypothesis testing in clinical trials.

277   Second, Bayesian methods directly estimate the probability that the treatment effect is larger than a

278   clinically important threshold, given prior assumptions; such information may be more directly

279   informative to clinicians and patients or families wrestling with complex treatment decisions than

280   probabilities of observing data more extreme than the observed data if there is no real treatment effect

281   quantified by frequentist p-values. The probabilistic results of Bayesian analysis naturally align with the

282   thought processes of clinicians making treatment decisions at the bedside where the probabilities of

283   various competing benefits and harms must be weighed.

284   Third, by representing what is known about the treatment effect through a probability distribution,

285   Bayesian analysis allows the probabilities for different magnitudes of treatment effect to be estimated. For

286   the purposes of analysis, we defined an absolute risk reduction of 2% as a potential threshold for

287   clinically important treatment effect. However, this threshold may be insufficient to motivate the routine

288   use of early ECMO. Indeed, with an absolute risk reduction threshold of 20%, the posterior probability

289   was 2%. Various factors must be weighed in defining the minimum clinically important effect: the

290   baseline risk of the outcomes, the relevance of the outcome under study, the resources and expertise

291   required to deliver the intervention, the risk of treatment-related adverse effects, and the effect on other

292   clinical outcomes. Given uncertainty over this value, posterior probabilities for a range of relative and

293   absolute risk reductions were reported. Further investigation using decision analysis may help to define

294   the optimal value for clinically important treatment effect.

295   There are challenges with Bayesian analysis. Given their significant influence on posterior

296   probabilities, the priors must be specified to appropriately reflect the evidence available prior to the trial.

297   Selection of priors therefore requires careful forethought. Bayesian analysis also requires decisions about

298   the minimum clinically important treatment effect, as discussed above. Because decisions about priors

299   and treatment effects inevitably incorporate an element of judgement, Bayesian analysis is sometimes

300   criticized for perceived subjectivity. To address these challenges, posterior probabilities were computed

301   for a wide range of potential values of minimum clinically important treatment effect under a range of

302   reference priors specified based on archetypal considerations and on prior data.

303   The data-derived prior was estimated based on previous studies deemed to be of acceptable

304   methodological quality (randomized trials and 'quasi-randomized' studies employing rigorous propensity-

305   score techniques for analysis). Because the methodological limitations of these studies reduced

306   confidence in their estimates of effect,[22,29] the weight of these studies was reduced in the Bayesian

307   hierarchical model to render them less informative in the construction of the prior. Reassuringly, the

308   probability of treatment benefit remained high even when these studies were heavily downweighted such

309   that a patient in the pre-existing studies contributed much less influence in comparison to a patient

310   enrolled in EOLIA.

311   Reference priors were specified based on previous recommendations for establishing representative

312   levels of enthusiasm and skepticism.[1,3] This approach permits assessment of prior probability both in

313   terms of existing clinical data and the strength of the biological plausibility. Readers should determine

314   which prior best matches their own background assessment of the prior probability of benefit from

315   ECMO in very severe ARDS and assess the posterior probability of benefit in light of EOLIA

316   accordingly. One important decision is the specification of the strongly skeptical prior; this requires a

317   judgment about the upper limit of reasonable skepticism. The strongly skeptical prior specified for this

318 analysis is equivalent to the information derived from a hypothetical trial of early ECMO enrolling 264

319 patients (6% more than EOLIA) that finds the same risk of death in treatment and control groups—as

320 there are no studies of this magnitude published in the current ECMO era, this degree of skepticism may

321 be difficult to justify. This prior distribution therefore appears to appropriately represents the upper limit

322 of reasonable prior skepticism.

323 Whether the findings of this Bayesian analysis support the routine use of early ECMO for very

324 severe ARDS remains a matter of judgment. This judgment must incorporate several considerations: the

325 distribution of prior probability, the probability of mortality benefit (level of certainty) required to

326 motivate action (i.e. should one apply a treatment that has a predicted probability of benefit of 70% vs.

327 80% vs. 90% etc.), the minimum clinically important treatment effect size, the effect on outcomes other

328 than mortality (i.e. long-term functional status, quality of life, costs, resource implications), and the risk

329 of adverse events. This is particularly important, because physicians often underestimate the risk of

330 adverse events. This complexity highlights the need for decision analyses; Bayesian posterior probability

331 distributions very naturally inform decision analysis.[1] The decision to initiate ECMO will always remain

332 complex; no clinical trial, however conclusive, can remove the role of clinical judgment in making

333 decisions about treatments. The findings of this Bayesian analysis may be helpful to inform these

334 judgments.

335 Bayesian posterior probabilities can also inform the question as to whether future trials are

336 required. For example, some might propose conducting yet another randomized trial of early ECMO to

337 confirm mortality benefit (RR<1) under frequentist conventions (i.e. p<0.05). The posterior probabilities

338 reported here can help to inform future discussions about the need for additional trials and whether the

339 ethical requirement for equipoise in a randomized trial can be satisfied. Decisions about the need for a

340 future trial depend on the definition of equipoise (probability of benefit sufficient to exclude equipoise)

341 and the definition of the minimum clinically important treatment effect.[30]

342

343 Limitations

14

344   Limitations of this analysis include those inherent in the primary trial. Premature termination and a

345   high rate of crossovers may have led to limited statistical power to detect a meaningful treatment effect.

346   Patients were enrolled from both ECMO centers and non-ECMO referral centers, resulting in delayed

347   ECMO initiation for some patients, although this reflects clinical practice given the regionalized nature of

348   ECMO services.

349   In addition, there are limitations specific to these Bayesian re-analyses. First, the present analysis

350   constitutes an unplanned *post hoc* analysis of trial data. Such analyses should generally be treated with

351   caution (i.e., regarded as hypothesis-generating only) because, among other concerns, repeated hypothesis

352   testing using different analyses increases the chance of erroneously concluding that the null hypothesis

353   can be rejected ('p-hacking').[31] Several considerations, however, suggest that the present analyses are less

354   vulnerable to these concerns. They tested the same hypothesis and analyzed the same pre-specified

355   primary end-point as in the original publication—the pre-specified hypothesis or primary outcome were

356   not revised (generally entailed in secondary analyses). In addition, under Bayesian analysis, the risk of

357   erroneously estimating the posterior probability of treatment effect arises from incorrectly specifying the

358   priors, not from repeated estimates of this probability. The capacity to allow repeated estimates of

359   posterior probability is the basis for Bayesian adaptive trial design.[32]

360   Second, because the analyses were planned after the trial was published, it was difficult to use

361   empirical methods to elicit prior beliefs about the benefit of ECMO; beliefs about benefit would

362   unavoidably be influenced by the results of EOLIA.[33] Empirically-derived priors might have helped to

363   clarify the extent to which EOLIA should modify the perceived probability of benefit. Recognizing this

364   limitation, a range of priors was specified to represent the range of potential prior beliefs about treatment

365   effect that might have been described by an empirical method.

366   Third, these analyses focused specifically on mortality and did not consider other adverse events.

367

368   **Conclusions**

369   *Post hoc* Bayesian analysis of data from a randomized trial of early ECMO compared with

370   conventional lung-protective ventilation with the option for rescue ECMO among patients with very

371   severe ARDS provides information about the posterior probability of mortality benefit under a broad set

372   of assumptions that may help inform interpretation of the study findings.

373

374

385

References

396   1.   Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. Methods in health service research. An
397   introduction to bayesian methods in health technology assessment. *BMJ*. 1999;319(7208):508-512.

398   2.   Bland JM, Altman DG. Bayesians and frequentists. *BMJ*. 1998;317(7166):1151-1160.

399   3.   Wijeysundera DN, Austin PC, Hux JE, Beattie WS, Laupacis A. Bayesian statistical inference
400   enhances the interpretation of contemporary randomized controlled trials. *J Clin Epidemiol*.
401   2009;62(1):13-21.e15. doi:10.1016/j.jclinepi.2008.07.006.

402   4.   Spiegelhalter DJ, Freedman LS, Parmar MK. Applying Bayesian ideas in drug development and
403   clinical trials. *Stat Med*. 1993;12(15-16):1501–11–discussion1513–7.

404   5.   Brophy JM, Joseph L. Placing trials in context using Bayesian analysis. GUSTO revisited by
405   Reverend Bayes. *JAMA*. 1995;273(11):871-875.

406   6.   Quintana M, Viele K, Lewis RJ. Bayesian Analysis: Using Prior Information to Interpret the
407   Results of Clinical Trials. *JAMA*. 2017;318(16):1605-1606. doi:10.1001/jama.2017.15574.

408   7.   Combes A, Hajage D, Capellier G, et al. Extracorporeal Membrane Oxygenation for Severe Acute
409   Respiratory Distress Syndrome. *N Engl J Med*. 2018;378(21):1965-1975.
410   doi:10.1056/NEJMoa1800385.

411   8.   Vincent J-L, Brochard LJ. Do we need randomized clinical trials in extracorporeal respiratory
412   support? We are not sure. *Intensive Care Medicine*. 2017;43(12):1869-1871. doi:10.1007/s00134-
413   017-4930-x.

414   9.   Dalton HJ, MacLaren G. Extracorporeal membrane oxygenation in pandemic flu: insufficient
415   evidence or worth the effort? *Critical Care Medicine*. 2010;38(6):1484-1485.
416   doi:10.1097/CCM.0b013e3181e08fff.

417   10. Gattinoni L, Vasques F, Quintel M. Use of ECMO in ARDS: does the EOLIA trial really help?
418   *Critical care (London, England)*. 2018;22(1):171. doi:10.1186/s13054-018-2098-6.

419   11. Hardin CC, Hibbert K. ECMO for Severe ARDS. *N Engl J Med*. 2018;378(21):2032-2034.
420   doi:10.1056/NEJMe1802676.

421   12. Mi MY, Matthay MA, Morris AH. Extracorporeal Membrane Oxygenation for Severe Acute
422   Respiratory Distress Syndrome. *N Engl J Med*. 2018;379(9):884-887.
423   doi:10.1056/NEJMclde1804601.

424   13. Combes A, Hajage D, Capellier G, et al. Extracorporeal Membrane Oxygenation for Severe Acute
425   Respiratory Distress Syndrome. *N Engl J Med*. 2018;378(21):1965-1975.
426   doi:10.1056/NEJMoa1800385.

427   14. United States Census Bureau. https://www.census.gov/popclock/. Accessed September 10, 2018.

428   15. Rubenfeld GD, Caldwell E, Peabody E, et al. Incidence and outcomes of acute lung injury. *N Engl
429   J Med*. 2005;353(16):1685-1693. doi:10.1056/NEJMoa050333.

430 16. Bellani G, Laffey JG, Pham T, et al. Epidemiology, Patterns of Care, and Mortality for Patients
431   With Acute Respiratory Distress Syndrome in Intensive Care Units in 50 Countries. *JAMA*.
432   2016;315(8):788-800. doi:10.1001/jama.2016.0291.

433 17. Li X, Scales DC, Kavanagh BP. Unproven and Expensive before Proven and Cheap:
434   Extracorporeal Membrane Oxygenation versus Prone Position in Acute Respiratory Distress
435   Syndrome. *Am J Respir Crit Care Med*. 2018;197(8):991-993. doi:10.1164/rccm.201711-2216CP.

436 18. Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung
437   injury and the acute respiratory distress syndrome. The Acute Respiratory Distress Syndrome
438   Network. *N Engl J Med*. 2000;342(18):1301-1308. doi:10.1056/NEJM200005043421801.

439 19. Peek GJ, Mugford M, Tiruvoipati R, et al. Efficacy and economic assessment of conventional
440   ventilatory support versus extracorporeal membrane oxygenation for severe adult respiratory
441   failure (CESAR): a multicentre randomised controlled trial. *Lancet*. 2009;374(9698):1351-1363.
442   doi:10.1016/S0140-6736(09)61069-2.

443 20. Pham T, Combes A, Rozé H, et al. Extracorporeal membrane oxygenation for pandemic influenza
444   A(H1N1)-induced acute respiratory distress syndrome: a cohort study and propensity-matched
445   analysis. *Am J Respir Crit Care Med*. 2013;187(3):276-285. doi:10.1164/rccm.201205-0815OC.

446 21. Noah MA, Peek GJ, Finney SJ, et al. Referral to an extracorporeal membrane oxygenation center
447   and mortality among patients with severe 2009 influenza A(H1N1). *JAMA*. 2011;306(15):1659-
448   1668. doi:10.1001/jama.2011.1471.

449 22. Munshi L, Telesnicki T, Walkey A, Fan E. Extracorporeal life support for acute respiratory failure.
450   A systematic review and metaanalysis. *Annals ATS*. 2014;11(5):802-810.
451   doi:10.1513/AnnalsATS.201401-012OC.

452 23. Su YS, Yajima M. R2jags: Using R to Run "JAGS." R package version 0.5-7. *httpsCRANR-*
453   *projectorgpackageRjags*. https://CRAN.R-project.org/package=R2jags.

454 24. Plummer M. JAGS: Just another Gibbs sampler. *httpmcmc-jagssourceforgenet*. 2004. http://mcmc-
455   jags.sourceforge.net.

456 25. Allmark P. Bayes and health care research. *Med Health Care Philos*. 2004;7(3):321-332.

457 26. Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials.
458   https://www.fda.gov/MedicalDevices/ucm071072.htm. Accessed September 10, 2018.

459 27. Harrington D, Drazen JM. Learning from a Trial Stopped by a Data and Safety Monitoring Board.
460   *N Engl J Med*. 2018;378(21):2031-2032. doi:10.1056/NEJMe1805123.

461 28. Kass RE, Greenhouse JB. [Investigating Therapies of Potentially Great Benefit: ECMO]:
462   Comment: A Bayesian Perspective. *Statist Sci*. 1989;4(4):310-317. doi:10.1214/ss/1177012386.

463 29. Fan E, Del Sorbo L, Goligher EC, et al. An Official American Thoracic Society/European Society
464   of Intensive Care Medicine/Society of Critical Care Medicine Clinical Practice Guideline:
465   Mechanical Ventilation in Adult Patients with Acute Respiratory Distress Syndrome. *Am J Respir*
466   *Crit Care Med*. 2017;195(9):1253-1263. doi:10.1164/rccm.201703-0548ST.

467   30.   Lilford RJ. Ethics of clinical trials from a bayesian and decision analytic perspective: whose
468   equipoise is it anyway? *BMJ*. 2003;326(7396):980-981. doi:10.1136/bmj.326.7396.980.

469   31.   Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD. The extent and consequences of p-
470   hacking in science. *PLoS Biol*. 2015;13(3):e1002106. doi:10.1371/journal.pbio.1002106.

471   32.   Berry DA. Bayesian clinical trials. *Nat Rev Drug Discov*. 2006;5(1):27-36. doi:10.1038/nrd1927.

472   33.   Johnson SR, Tomlinson GA, Hawker GA, Granton JT, Feldman BM. Methods to elicit beliefs for
473   Bayesian priors: a systematic review. *J Clin Epidemiol*. 2010;63(4):355-369.
474   doi:10.1016/j.jclinepi.2009.06.003.

475

476

Figure Legends

478

479   **Figure 1.** Graphical representation of reference priors (Left) and data-derived priors (Right). Each prior

480   distribution represents a belief about the probability of differing mortality benefits (relative risks of death)

481   with the use of early ECMO in patients with very severe acute respiratory distress syndrome. Bayesian

482   analysis combines each prior distribution with the likelihood function of the observed treatment benefit to

483   determine the posterior probability of treatment benefit. A range of reference prior distributions were

484   specified in an effort to match the spectrum of belief within the clinical community about the benefit of

485   ECMO. The minimally informative prior distribution entails that all potential values for log relative risk

486   are approximately equally likely. The data-derived priors are based on previous studies (see text for

487   details). To account for likely differences in previous studies, the weight (influence) of patients enrolled

488   in these prior studies was reduced by artificially inflating the study variance (resulting in a wider prior

489   probability density distribution).

490

491   **Figure 2.** Posterior probability distributions for relative risk (Panel A) and absolute risk reduction (Panel

492   B) obtained based on the EOLIA trial results under varying prior assumptions about the benefit of early

493   ECMO on mortality. Prior distributions (represented by the red lines) are combined with the likelihood

494   function summarizing the treatment effect observed in the trial (green shaded region) to compute the

495   posterior probability for the treatment effect. In each case the likelihood function, summarizing the trial

496   data, is the same; variation in the posterior distribution arises from variation in the prior. In the case of a

497   non-informative prior, the likelihood function and posterior distribution are identical. The median effect

498   and credible interval are shown as the black point and line below each set of distributions. This approach

499   allows assessment of the influence of prior enthusiasm or skepticism for early ECMO on the

500   interpretation of the trial.

501

502   **Figure 3.** Posterior probabilities for a reduction in mortality with VV-ECMO in very severe ARDS given

503   EOLIA and the results of previous studies. Varying degrees of weight were applied to the previous

504   studies by artificially increasing the variance (width) of their probability distribution to reflect varying

505   levels of confidence in their estimates of effect given their likely differences (and potential

506   methodological limitations). The left panel shows the resulting credible interval for the relative risk of

507   mortality for various levels of weighting of previous studies in proportion to the weight assigned to the

508   EOLIA trial. The right panel shows the resulting estimated probability of a given relative risk reduction

509   for varying weights assigned to the previous studies. EOLIA = "ECMO to Rescue Acute Lung Injury"

510   randomized trial; CrI = credible intervals; RR = relative risk.

511

512 **Table 1.** Characteristics of Reference Prior Probability Distributions Representing Prior Beliefs About
513 Mortality Benefit from ECMO
514

Prior belief	Assumed Median Relative Risk	Assumed Standard Deviation of Logarithm of Relative Risk	Prior Evidence Equivalent*	Probability of Treatment Effect Equal to or Greater Than Specified Threshold				Rationale for Specifying Distribution Characteristics
				RR < 1	RR < 0.9	RR < 0.8	RR < 0.7	
Non-informative	1.0	10	Equivalent to essentially no prior data	50%	50%	49%	49%	All possible values for treatment effect for log RR [author – correct?] are equally likely
Strongly enthusiastic	0.67	0.25	Equivalent to a previous RCT enrolling 100 patients finding a 33% relative risk reduction	95%	89%	77%	58%	Probability of observing a treatment effect equal to or greater than that assumed in EOLIA design is 50%; probability of harm (RR>1) is 5%
Moderately enthusiastic	0.78	0.15	Equivalent to a previous RCT enrolling 264 patients finding a 22% relative risk reduction	95%	83%	57%	24%	Probability of observing a treatment effect equal to or greater than that approximating effect observed in ARDSNet lower tidal volumes trial (RR=0.78) is 50%; probability of harm (RR> 1) is 5%
Skeptical	1.0	0.24	Equivalent to a previous RCT enrolling 100 patients finding a 0% relative risk reduction	50%	33%	18%	7%	Probability of observing a treatment effect equal to or greater than that assumed in EOLIA design (RR=0.67) is 5%; probability of benefit and harm are equivalent
Strongly skeptical	1.0	0.15	Equivalent to a previous RCT enrolling 264 patients finding 0% relative risk reduction	50%	24%	7%	1%	Probability of observing a treatment effect equal to or greater than that observed in the ARDSNet lower tidal volume trial (RR=0.78) is 5%

515 RR = relative risk, EOLIA = ECMO to Rescue Acute Lung Injury trial, ARDSNet = NIH/NHLBI ARDS
516 Network, RCT = randomized controlled trial
517 *"Prior evidence equivalent" communicates the level of certainty represented in each reference prior by
518 reference to the treatment effect and sample size of a hypothetical randomized trial required to generate
519 the level of informative influence on posterior probability specified by the reference prior relative to the
520 size of the EOLIA trial
521

522 **Table 2.** Probability of treatment effects estimated by Bayesian analysis using varying distributions to
523 describe prior beliefs

| Prior belief		Posterior Median Relative Risk (95% Credible Interval)	Posterior Probability that True Relative Risk Is Less Than or Equal to Specified Threshold			
			RR < 1	RR < 0.9	RR < 0.8	RR < 0.67
Reference Prior Distributions	Non-informative	0.78 (0.56-1.04)	96%	85%	60%	18%
	Strongly enthusiastic	0.74 (0.57-0.95)	99%	94%	73%	22%
	Moderately enthusiastic	0.78 (0.63-0.96)	99%	91%	61%	8%
	Skeptical	0.84 (0.64-1.07)	93%	73%	39%	5%
	Strongly skeptical	0.88 (0.71-1.09)	88%	58%	18%	0%
Data-derived Prior Distributions	No downweighting[a] of previous studies	0.71 (0.55-0.94)	99%	96%	83%	48%
	50% downweighting of previous studies	0.73 (0.56-0.96)	99%	94%	77%	40%
	75% downweighting of previous studies	0.74 (0.56-0.98)	98%	92%	72%	36%

524 RR = relative risk
525 [a]Downweighting refers to a deliberate reduction in the influence (weight) of previous studies in the
526 Bayesian hierarchical model by artificially increasing the variance of these studies. Downweighting
527 provides a method of representing uncertainty about the estimates of effect in these studies given their
528 likely differences compared to the current trial.

529 **Table 3.** Probability that early ECMO reduces mortality by a proposed minimum clinically important
530 difference according to varying possible baseline mortality rates in patients with very severe ARDS

Prior belief		Posterior Median Absolute Risk Reduction (95% Credible Interval)	Posterior Probability that Absolute Risk Reduction[a] is Greater Than or Equal to Specified Threshold					
			2%	4%	6%	8%	10%	20%
Reference Prior Distributions	Non-informative	-10.6% (-20.0% - 1.8%)	92%	86%	78%	67%	53%	2%
	Strongly enthusiastic	-12.0% (-19.9% - -2.1%)	98%	95%	89%	79%	65%	2%
	Moderately enthusiastic	-10.4% (-17.2% - -2.0%)	97%	93%	85%	71%	51%	0%
	Skeptical	-7.8% (-16.5% - 3.4%	86%	76%	62%	47%	30%	0%
	Strongly skeptical	-5.6% (-13.3% – 4.1%)	78%	63%	45%	26%	13%	0%
Data-Derived Prior Distribution	No downweighting of previous studies	-13.6% (-20.5% - -2.9%)	98%	96%	93%	88%	79%	4%
	50% downweighting of previous studies	-12.8% (-20.4% - -1.9%)	97%	95%	91%	83%	72%	3%
	75% downweighting of previous studies	-12.1% (-20.3% - -1.1%)	97%	93%	88%	79%	66%	3%

531
532 ECMO = extracorporeal membrane oxygenation; ARDS = acute respiratory distress syndrome
533 [a]Absolute risk reduction was computed assuming a baseline mortality risk of 46% (based on the mortality
534 rate in the control group of EOLIA)
535
536
537

538   Figure 1
539
540
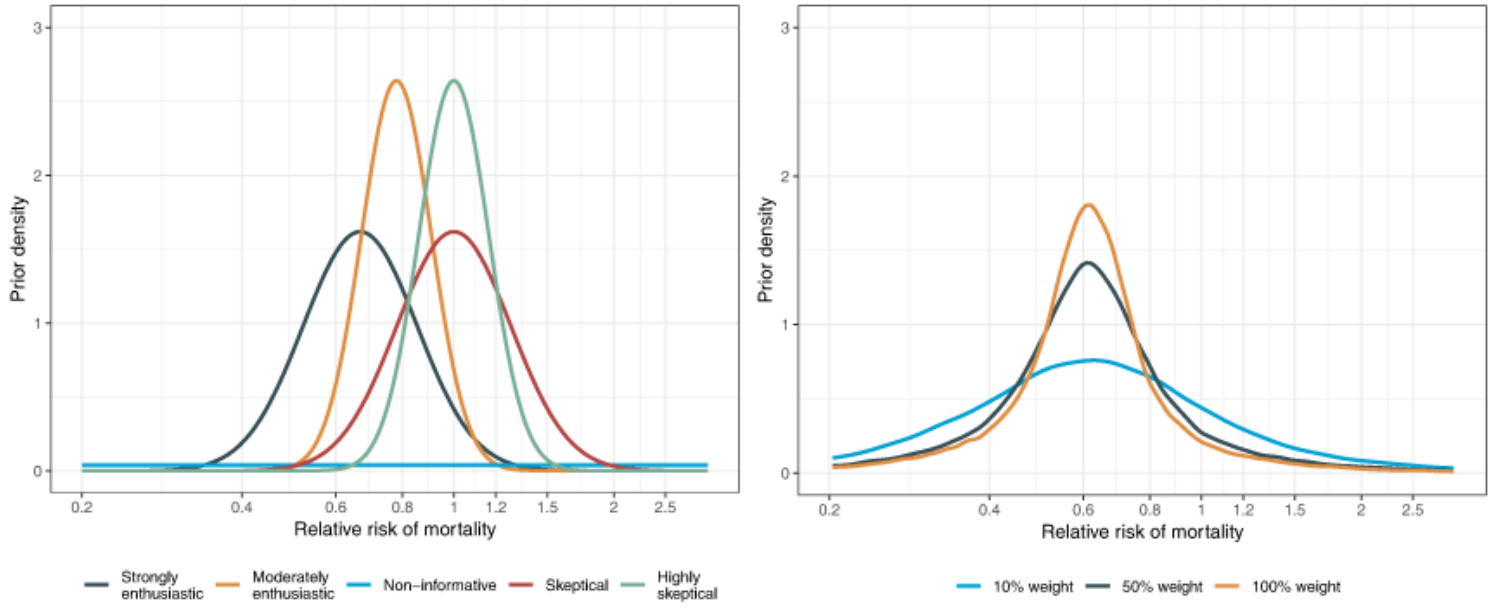
Figure 2

542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
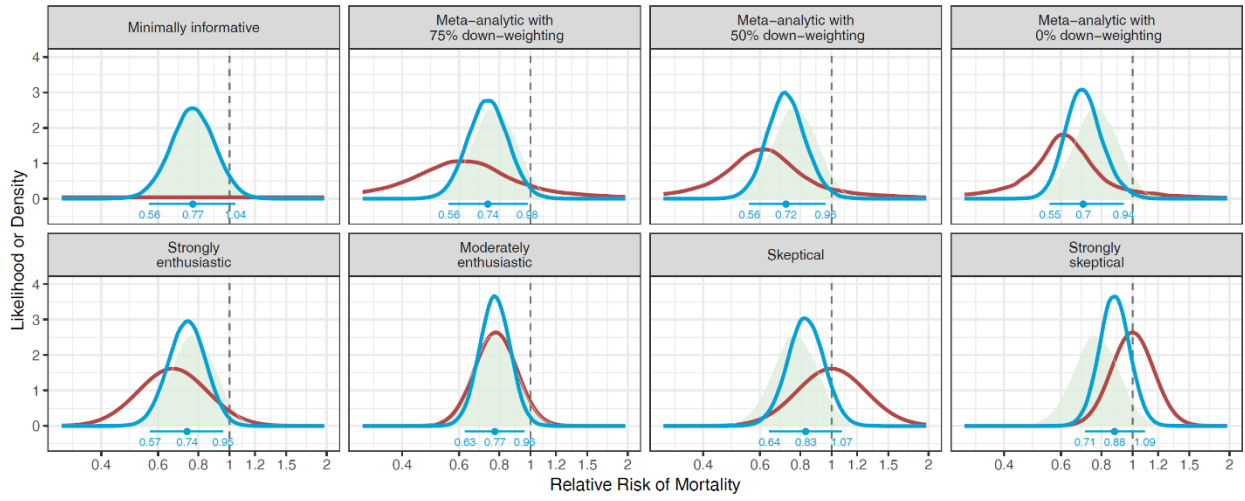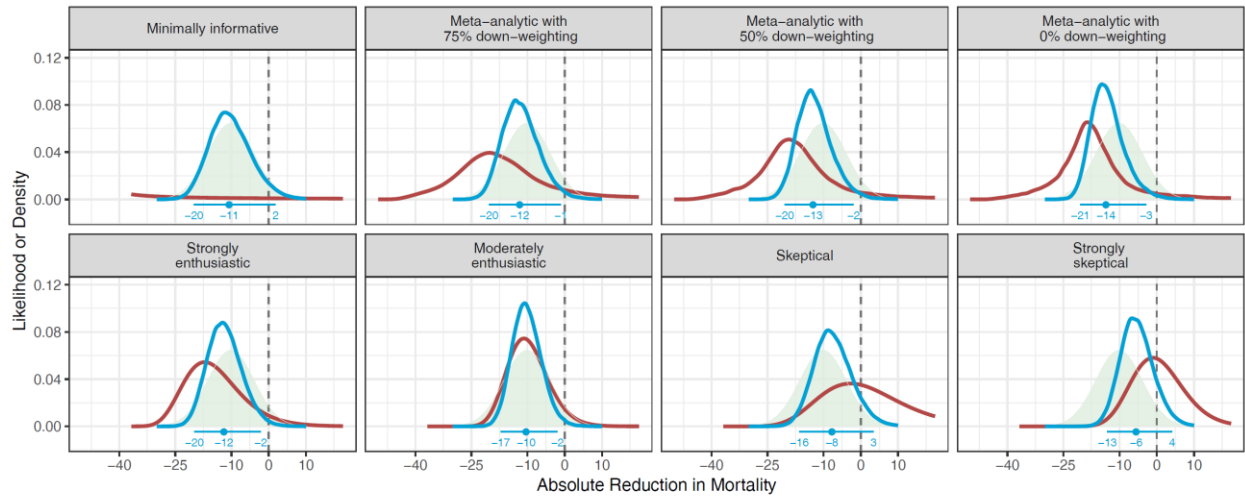566
567
568
569
570

571   Figure 3

572

573