



HAL
open science

Sample size estimation in clinical trials using ventilator-free days as the primary outcome: a systematic review

Laurent Renard Triché, Emmanuel Futier, Manuela de Carvalho, Nathalie Piñol-Domenech, Laëtitia Bodet-Contentin, Matthieu Jabaudon, Bruno Pereira

► To cite this version:

Laurent Renard Triché, Emmanuel Futier, Manuela de Carvalho, Nathalie Piñol-Domenech, Laëtitia Bodet-Contentin, et al.. Sample size estimation in clinical trials using ventilator-free days as the primary outcome: a systematic review. *Critical Care*, 2023, 27 (1), pp.303. 10.1186/s13054-023-04562-y . hal-04246799

HAL Id: hal-04246799

<https://hal.science/hal-04246799>

Submitted on 17 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 **SAMPLE SIZE ESTIMATION IN CLINICAL TRIALS USING VENTILATOR-FREE DAYS**
2 **AS THE PRIMARY OUTCOME: A SYSTEMATIC REVIEW.**

3 Laurent Renard Triché¹, Emmanuel Futier^{1,2}, Manuela De Carvalho³, Nathalie Piñol-
4 Domenech³, Laëtitia Bodet-Contentin^{4,5}, Matthieu Jabaudon^{1,2}, Bruno Pereira⁶.

5 ¹Department of Perioperative Medicine, CHU Clermont-Ferrand, Clermont-Ferrand, France

6 ²iGReD, CNRS, INSERM, Université Clermont Auvergne, Clermont-Ferrand, France

7 ³Université Clermont Auvergne, Health Library, Clermont-Ferrand, France

8 ⁴Medical Intensive Care Unit, CHRU de Tours, France

9 ⁵INSERM, SPHERE, UMR1246, Université de Tours et Nantes, Tours et Nantes, France

10 ⁶Biostatistics Unit, Department of Clinical Research, and Innovation (DRCI), CHU Clermont-Ferrand,
11 Clermont-Ferrand, France

12 **Corresponding Author:**

13 Laurent Renard Triché

14 Department of Perioperative Medicine, CHU Clermont-Ferrand

15 58 rue Montalembert, 63 000 Clermont-Ferrand, France

16 +33 6 67 05 08 74

17 lrenard--triche@chu-clermontferrand.fr

18 **ABSTRACT**

19 **Background**

20 Ventilator-free days (VFDs) are a composite endpoint increasingly used as the primary
21 outcome in critical care trials. However, because of the skewed distribution and competitive
22 risk between components, sample size estimation remains challenging. This systematic
23 review was conducted to systematically assess whether the sample size was congruent, as
24 calculated to evaluate VFDs in trials, with VFDs' distribution and the impact of alternative
25 methods on sample size estimation.

26 **Methods**

27 A systematic literature search was conducted within the Pubmed and Embase databases for
28 randomized clinical trials in adults with VFDs as the primary outcome until December 2021.

29 We focused on peer-reviewed journals with 2021 impact factors greater than five. After

30 reviewing definitions of VFDs, we extracted the sample size and methods used for its
31 estimation. The data were collected by two independent investigators and recorded in a
32 standardized, pilot-tested forms tool. Sample sizes were calculated using alternative
33 statistical approaches, and risks of bias were assessed with the Cochrane risk-of-bias tool.

34 **Results**

35 Of the 26 clinical trials included, 19 (73%) raised “some concerns” when assessing risks of
36 bias. Twenty-four (92%) trials were two-arm superiority trials, and 23 (89%) were conducted
37 at multiple sites. Almost all the trials (96%) were unable to consider the unique distribution
38 of VFDs and death as a competitive risk. Moreover, significant heterogeneity was found in
39 the definitions of VFDs, especially regarding varying start time and type of respiratory
40 support. Methods for sample size estimation were also heterogeneous and simple models,
41 such as the Mann–Whitney–Wilcoxon rank-sum test, were used in 14 (54%) trials. Finally,
42 the sample sizes calculated varied by a factor of 1.6 to 17.4.

43 **Conclusions**

44 A standardized definition and methodology for VFDs, including the use of a core outcome
45 set, seems to be required. Indeed, this could facilitate the interpretation of findings in
46 clinical trials, as well as their construction, especially the sample size estimation which is a
47 trade-off between cost, ethics, and statistical power.

48 **Systematic review registration:** PROSPERO ID: CRD42021282304. Registered 15 December
49 2021. (https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021282304)

50 **Keywords:** respiration, artificial · respiratory insufficiency · time factors · treatment outcome
51 · ventilators, mechanical.

52 **BACKGROUND**

53 Between a quarter and half of the patients admitted to the intensive care unit will
54 present with respiratory failure, requiring invasive mechanical ventilation[1]. These patients
55 are at risk of complications, such as ventilator-associated pneumonia and death, with related
56 health-care costs[2,3].

57 Mortality is a robust endpoint that has long been used in studies[4]. However, since
58 the improvement of therapeutics, mortality has decreased[5], and the sample size needed to
59 show a clinically relevant difference in mortality has also increased. Hence, most published
60 randomized clinical trials (RCTs) that aim to reduce mortality have produced negative
61 results[6,7]. For this reason, other outcomes have been developed, such as ventilator-free
62 days (VFDs), which are increasingly used in critical care RCTs[8]. First proposed in 1994[9],
63 VFDs were developed in studies focusing on acute respiratory distress syndrome. The
64 number of VFDs was defined as the number of days from the last day of mechanical
65 ventilation to day 28. If a patient died during the first 28 days, their number of VFDs is equal
66 to zero. This composite outcome measure (i.e., combining survival and the duration of
67 ventilation) is more appropriate than only the duration of ventilation because the latter
68 disregards the mortality rate[10].

69 In clinical research, it is not feasible, for most studies, to study the whole
70 population[11]. We therefore need to determine the sample size, which can be imprecise
71 and difficult. Indeed, it represents a trade-off between cost effectiveness (i.e., in terms of
72 time and resource), ethical concerns (e.g., an oversized experiment would result in exposure
73 of an unnecessary number of subjects) and statistical power (i.e., a small sample size could
74 make the study underpowered to show a clinically meaningful difference, if any, and to

75 detect a potentially effective treatment)[12]. Calculating this sample size involves the
76 employment of formulae designed to obtain significant results in studies that compare
77 several groups based on the primary endpoint. The test chosen to analyze the primary
78 endpoint will depend on its distribution and is part of the sample size estimation[13]. VFDs
79 do not follow a Gaussian distribution[14]; therefore, we cannot use parametric tests. Indeed,
80 the distribution is skewed with inflations, especially 0s, and represents a rather time-
81 dependent event. In their last review, Yehya et al.[8] recommended using competing risk
82 regression, such as the Fine and Gray competing risk regression[15], which considers
83 extubation success as the event of interest and death as the competing risk.

84 There appears to be a number of inconsistencies in the definitions and
85 methodologies used for VFDs in the literature[8,16]. As a result, we conducted a systematic
86 review of RCTs using VFDs as the primary outcome to evaluate them. Hence, our principal
87 objective was to investigate whether the sample size estimation of VFDs was congruent with
88 their true distribution. Indeed, incorrect sample size estimation may lead to additional costs,
89 expose an unnecessary number of subjects or decrease the power of a study. Our secondary
90 objectives were to review the definitions of VFDs and to evaluate different statistical
91 approaches to their estimation.

92 **Methods**

93 **Search strategy, study selection and inclusion criteria**

94 We searched through two databases (MEDLINE and Embase) using a combination of
95 keywords. The last literature search was done on December 31, 2021. We only focused on
96 RCTs with VFDs as the primary outcome in peer-reviewed journals with 2021 impact factors
97 greater than five. The complete list of search terms is available in the online data

98 supplement (Additional file 1: Appendix 1). Two investigators (LRT and MJ) independently
99 screened the titles and abstracts of the search results. The full text of all potentially eligible
100 studies was retrieved and reviewed for eligibility. First, we removed the duplicates between
101 the databases. Then, we excluded all trials that were not RCTs in adults and those with the
102 primary endpoint that was not VFDs. A narrative synthesis supporting the Preferred
103 Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram and the
104 PRISMA 2020 Checklist[17] were included as part of this systematic review. The study
105 protocol was registered with the International Prospective Register of Systematic Reviews
106 (PROSPERO) in December 2021 (ID: CRD42021282304)[18].

107 **Data extraction**

108 Data were extracted independently by two investigators (LRT and MJ or BP) and
109 collected into standardized forms using Research Electronic Data Capture tools[19,20]. Data
110 were cross-checked; any disagreements were resolved first by consensus, and if one or
111 several disagreements persisted, a third investigator (BP or MJ) was involved. For each
112 selected article, we recorded several items, as detailed in the online data supplement
113 (Additional file 1: Appendix 2).

114 **Risk of bias**

115 To assess the risk of bias in each study, we used Version 2 of the Cochrane risk-of-
116 bias tool (RoB2) for RCTs[21]. The studies were assessed using five fixed domains, as outlined
117 in RoB2. Each study was classified by two investigators (LRT and MJ) as having “low risk,”
118 “some concerns,” or “high risk.”

119 **Outcomes**

120 Our primary outcome was the sample size, as estimated in RCTs evaluating VFDs as
121 their primary outcome. First, we extracted the sample sizes estimated and observed among
122 the trials. Secondly, because there is heterogeneity in the tests used for this outcome, we
123 simulated other sample sizes through alternative statistical approaches. Our secondary
124 outcomes were to review the definitions of VFDs, mortality rates, statistical methods, and
125 VFDs' distributions among selected trials.

126 **Statistical analysis**

127 The different statistical approaches were as follows: the Student *t*-test and the
128 Mann–Whitney–Wilcoxon rank-sum test because these are standard tests used in several
129 studies; the Mann–Whitney–Wilcoxon rank-sum test using the Noether formula to compare
130 if the result differ from the previous one; the Cox regression because VFDs are considered by
131 some to be a time-dependent event; the zero-inflated negative binomial regression because
132 VFDs involve a zero-inflation; and finally, the Fine and Gray regression because the VFDs
133 involve death as a competitive risk. The corresponding formulae are available in the online
134 data supplement (Additional file 1: Appendix 3). All statistical analyses were performed using
135 R Core Version 4.2[22]. All packages used are listed in the online data supplement
136 (Additional file 1: Appendix 4).

137 **Results**

138 **Study selection and characteristics**

139 We identified 425 studies from 2004 to 2021. After removing duplicates, we assessed
140 269 studies. One hundred and thirty-six non-randomized studies were excluded, as well as
141 two animal studies and 29 pediatric studies. We then excluded 76 studies in which the
142 primary outcome was not VFDs. Twenty-six studies were finally included in our systematic

143 review[23–48] (Additional file 1: Figure S1). These were 24 (92%) superiority studies
144 comparing two groups (three for the two remaining) among several centers (median [IQR],
145 23 [8–42]) for 22 (85%) trials. An interim analysis was performed in 16 (58%) trials. Ten (39%)
146 of the selected studies had to be stopped early. Finally, the patient populations included in
147 these studies were heterogeneous, with a third having acute respiratory distress syndrome
148 (see Table 1 and Additional file 1: Table S1).

149 **Risk of bias and disagreements**

150 Using the Excel spreadsheet provided by the RoB2 tool, we assessed the risk of bias,
151 as summarized in Table S2. Nineteen (73%) of the studies were assessed as having “some
152 concerns,” mainly related to the randomization process and deviations from the intended
153 interventions.

154 Among all the collected items, the median [IQR] number of disagreements between
155 the two reviewers was 1 [0 – 2] out of the 26 selected studies, for a total of 23
156 disagreements out of 769 items (3%). All disagreements were resolved by consensus.

157 **Sample size estimation**

158 We extracted the estimated and observed sample sizes reported in the selected
159 studies. We subsequently estimated the sample size with parameters (e.g., risk, power,
160 mean difference) reported in two ways: using the expected parameters displayed in the
161 Material and Methods section or the observed parameters displayed in the Results section.

162 First, we reported the expected parameters proposed by the authors for the sample
163 size estimation in the Methods sections of the selected studies (see Table 2 and Additional
164 file 1: Table S3). The absolute mean difference in VFDs ranged from 0.5 to 7.0. In one

165 noninferiority study[30], the authors considered 1.6 to be the noninferiority margin,
166 whereas in one superiority study[29], the authors considered 1 to be the superiority margin.
167 These expected mean differences were only justified in eight (31%) studies. The standard
168 deviation was only reported in 23% of studies, but, when it was available, it was
169 heterogenous (median [IQR], 10.0 [6.8 – 10.5]) (Figure S2). Mortality was considered in one
170 study only[26], in which Markov chains considered the probabilities of death, getting off
171 ventilation alive, and receiving ventilation. Finally, the expected dropout rate was quite
172 diverse among studies (0% – 25%).

173 Using these parameters (i.e., mainly mean difference and standard deviation), we
174 calculated, as reported in Table S4, the different sample sizes resulting from different
175 statistical tests: the Student *t*-test, the Mann–Whitney–Wilcoxon rank-sum test using the
176 Noether formula or not, the Cox regression, the zero-inflated negative binomial (ZINB)
177 regression, and the Fine and Gray regression. Several models could not be computed
178 because of some expected parameters not being reported, such as the VFDs in the control
179 group and their standard deviation. Moreover, it was not possible to estimate the sample
180 size using the Fine and Gray regression because neither the probability of extubation nor the
181 mortality incidence was reported. For estimations using Cox and ZINB regression, in most
182 cases, the sample size was greater than with other models and slightly higher with Cox
183 regression than with ZINB regression (see Figure 1). The median [IQR] of the maximum
184 variation factor between sample size estimations was 1.9 [1.7 – 3.5], with a maximum of
185 17.4.

186 Second, we reported the observed parameters needed to estimate the sample size
187 (Additional file 1: Table S5). Standard deviations were slightly different from those estimated

188 (absolute median difference [IQR], 4.5 [1.0 – 6.9]). Furthermore, the dropout rate observed
189 was very low (0% – 2%).

190 Using these parameters, we calculated the different sample sizes using the same
191 statistical tests as above (Additional file 1: Table S6). Sample sizes could not be estimated
192 using the Fine and Gray regression model because, for some studies, VFDs and the mortality
193 incidence had different timeframes. In addition, the incidence of extubation was never
194 reported in the selected studies. We did not estimate the sample sizes in about half of the
195 studies because the observed mean difference was too low (i.e., when the mean difference
196 was less than 1). Indeed, conducting a study with such an effect size would appear irrelevant
197 and clinically unrealistic.

198 Finally, because several data useful to estimate the sample size, especially for the
199 Fine and Gray regression, were not reported, simulation was carried out from a previously
200 published dataset by Bodet-Contentin et al.[49]. We estimated the sample size using this
201 simulation and the same tests as above (Additional file 1: Table S7). Only the estimation
202 using the Fine and Gray regression model provided a realistic sample size. However, this was
203 more of a thought experiment because the effect size was low (mean difference = 0.46) and
204 further simulation studies are warranted.

205 **Definitions of ventilator-free days**

206 The definitions of VFDs across selected studies are reported in Table 3. Almost all
207 studies counted whole days without support ventilation (92%) and calculated VFDs at day 28
208 after randomization. Other definitions were heterogeneous. The onset (i.e., the beginning of
209 the period without support ventilation) was not the same across studies: 35% considered the
210 onset at extubation and 38% at 48 hours after extubation. If a patient was intubated again

211 after a period of extubation, the count of VFDs started only after the last extubation event in
212 35% of the studies, and 27% of the studies summed the different periods during which the
213 patient was extubated; the remaining studies did not specify this point. Finally, half of the
214 studies did not mention the type of respiratory support (invasive or noninvasive) used to
215 define VFDs.

216 **Statistical methods to analyze ventilator-free days: distribution and statistical tests used**

217 The proper sample size estimation necessitates a correct estimation of the
218 distribution of VFDs. Distributions of VFDs, as defined by the authors of the selected studies
219 are reported in Table 2 and Table S3. Half of the studies did not explicitly state the type of
220 assumed distribution, whereas the other half did not consider VFDs to be normally
221 distributed. A more precise description was available for some studies, with 8% assuming a
222 zero-inflated binomial distribution and 4% assuming a bimodal distribution. However, two
223 recently published studies[32,35] considered the number of VFDs to be normally distributed.
224 We therefore simulated a normal distribution of VFDs with the parameters used in some
225 trials included in our systematic review, which was not consistent with the empirical
226 distribution of VFDs found in Jabaudon et al.'s meta-analysis[50] (see Figure 2).

227 We also reported the statistical analysis methods used to assess VFDs in the selected
228 studies (see Table 2 and Additional file 1: Table S3). About one-third used the Mann–
229 Whitney–Wilcoxon rank-sum test, one-fifth did not specify the test used, and more than a
230 third used parametric tests. A minority used complex models, such as generalized linear
231 mixed models (GLMM) and generalized additive models for location, scale, and shape
232 (GAMLSS). The main effect size reported was the absolute mean difference (in 69% of
233 studies) (Additional file 1: Table S4). A significant result for VFDs was obtained only in a few

234 studies (35%), but it was significant when a complex model was used (Additional file 1: Table
235 S6).

236 **Other characteristics**

237 The power ranged from 80% (for 73% of the studies) to 90%, and the α risk ranged
238 from 2.5% to 10% (for one study[37]); however, the α risk was most frequently 5% (see
239 Table 2 and Additional file 1: Table S3). The reported α risk was one-sided in four studies
240 (two were noninferiority studies[30,33] and two were superiority studies[35,37]).

241 The most used timeframe for mortality was 28 days (Additional file 1: Table S5),
242 corresponding to the same frame as for VFDs. The timeframe was reported in days in 92% of
243 the studies; the remaining two used hospital or intensive care unit mortality. The incidence
244 of death was quite different across the studies but similar within studies, with a median
245 [IQR] of 24.9 days [19.2–30.9] for the control group and 24.3 days [19.8–30.7] for the
246 experimental group.

247 Finally, only three (12%) trials planned multiple imputation for the missing values
248 management, whereas the others did not plan any.

249 **Discussion**

250 In this systematic review, sample size estimation for assessment of VFDs in critical
251 care trials was heterogeneous and not in adequacy with the actual distribution of VFDs.
252 There was also important heterogeneity in the definitions of VFDs and in the methods used
253 for sample size estimation among trials. Sample size estimation extends beyond the VFDs to
254 all medical fields. Indeed, it is essential to have the right estimate before beginning a trial
255 because of several aspects, such as ethical, logistic, and financial concerns. When there is

256 heterogeneity of both outcome definition and methods for calculating it, the sample size
257 may be underestimated and lack the power to show a clinically meaningful difference, or it
258 may be overestimated and waste resources and expose an unnecessary number of subjects
259 to a potentially harmful treatment, or deny a potentially beneficial one[11–13].

260 **Sample size estimation: consensus definition of the outcome**

261 Following Contentin et al.[16] and Yehya et al.[8], we found important differences
262 between definitions of VFDs across trials, thus making it difficult to conduct meta-analyses,
263 as there is no common core.

264 Yehya et al. made several recommendations, including on how to explicitly define
265 VFDs[8]. Hence, a core outcome set, such as the Core Outcomes in Ventilation Trials[51],
266 could be used. This includes standardized definitions and measures for extubation,
267 reintubation, duration of mechanical ventilation, and mortality. However, although
268 Blackwood et al. defined several components of VFDs, there was no consensual definition of
269 VFDs. In future studies, all components of this outcome should be reported to facilitate the
270 preparation of the future statistical analysis plan, especially for the sample size estimation
271 because it includes some of these components.

272 Finally, other alternative approaches should be considered such as a ranked
273 composite score used in The Esophageal Pressure-Guided Ventilation 2 trial (EPVent-2)[52].
274 Alive and ventilator-free (AVF), the primary outcome used in EPVent-2, is a recent
275 hierarchical composite outcome that does not treat mortality as equivalent to prolonged
276 intubation[7]. This kind of outcome is already applied to other disciplines than critical care,
277 such as in lung and cardiovascular clinical trials[53,54]. In a simulation-based study[7], AVF
278 had higher power to detect differences in mortality than VFDs. Consequently, the sample

279 size could be lower with this outcome when there is a difference in mortality. Moreover, this
280 outcome typically requires fewer patients because its distribution is closer to a Gaussian
281 distribution than the distribution of VFDs. Finally, unlike AVF, which takes clinical priorities
282 into account, VFDs treat death or remaining intubated in the same way (i.e., if they were of
283 equal relevance). In contrast, death is considered more important than the duration of
284 mechanical ventilation in AVF, which seems more clinically relevant.

285 **Sample size estimation: methods**

286 Several parameters are required to estimate the sample size of an RCT. First is the tail
287 of the risk (one- or two-sided). In most superiority studies from our review, the authors used
288 a two-tailed test. However, if a specific and unidirectional difference is hypothesized (e.g.,
289 treatment vs placebo), a one-sided risk should be preferred to test the null hypothesis for
290 the two groups[55].

291 We also need the expected difference (i.e., the effect size) between two groups in a
292 superiority study or a loss of efficacy in a noninferiority study. Some studies from our review
293 used a superiority margin smaller than the noninferiority margin, which is hardly justifiable.
294 Furthermore, VFDs do not follow a Gaussian distribution and using the median difference as
295 the effect size when estimating sample size seems more relevant than using the mean
296 difference. In addition, the expected standard deviation was heterogeneous across studies.
297 Even if the population was different, the standard deviation should not have such a large
298 difference for the same outcome. A consensus according to the context to choose the
299 correct effect size and the related standard deviation seems to be necessary to estimate the
300 right sample size and ensure sufficient powered.

301 Finally, 73% of the included studies had chosen a power of 80% for sample size
302 estimation. Therefore, if the other parameters for this estimation are under- or
303 overestimated, there is limited room for mistake.

304 **Sample size estimation: statistical methods for ventilator-free days**

305 There are two concepts to consider when using VFDs: their distribution and the
306 presence of competitive risks. However, there are rarely reported, which may contribute to
307 the fact that most included trials were underpowered.

308 First, the unique distribution of VFDs could make their statistical analysis more
309 problematic. Indeed, some articles reported a zero-inflated beta distribution[30,31]. For this
310 distribution, a GLMM[31,43] (e.g., with a zero-inflated beta model or hurdle-negative
311 binomial model[56]) or a GAMLSS[30] can be used. These models also allow for adjusting
312 covariates, thus reducing the sample size and increasing the power[57]. Here, VFDs are
313 treated as a count outcome, where death, intubation, and extubation are treated together
314 and combined as one entity. These models assess whether there is a difference between
315 groups on distribution. However, there was important heterogeneity in the reported
316 distribution of VFDs and in the tests used among trials, which prompts further clarification.

317 Second, regarding the presence of competitive risks, Yehya et al.[8] recommended
318 reporting the mortality because the number of VFDs combines mortality and the duration of
319 ventilation. Because mortality is a competitive event of extubation, competing risk
320 regression using the Fine and Gray regression or the Cox-specific regression seem more
321 appropriate than the usual tests, such as the Student *t*-test or Mann–Whitney–Wilcoxon
322 rank-sum test[15]. In addition, not taking mortality into account may underpower the study,
323 especially if mortality is low. Moreover, these tests enable adjustment for covariates and

324 interim monitoring, which are common in RCTs. Nevertheless, the cumulative incidence
325 function provided by the Fine and Gray regression does not have a natural interpretation[58]
326 and does not take the zero inflation into account. Here, VFDs are treated as a time-to-event
327 outcome. However, in our opinion, this does not express the real definition of VFDs. Indeed,
328 it is more the extubation time that is shaped, with death as a competitive risk.

329 As a result, these two types of models are based on two distinct concepts: a count
330 outcome or a time-to-event outcome, not comparing the same things. Indeed, as mentioned
331 above, the different parts of the VFDs components (i.e., death and ventilation duration) have
332 a different importance depending on the type of model used, which is not much discussed in
333 the literature. However, a recent study looked into the count outcome[59]. No model was
334 globally recommended, the best model depends above all on the expected data distribution.
335 However, to date, there is no clear answer as to which statistical test to use.

336 Nevertheless, based on the current results, we believe the usual tests should be
337 discouraged for analysis of VFDs and more complex models considering competitive risks,
338 the unique distribution of VFDs, and any covariates, such as centers, might be more
339 appropriate for better fit the data.

340 **Sample size estimation: recent methodological approaches**

341 Common formulae can be used to estimate a trial's sample size[13]. However, for
342 complex models such as the GLMM and GAMLSS, these formulae cannot be employed, and
343 simulation-based power analyses could be useful[60].

344 The Markov chain model is another interesting approach to stochastically describe a
345 sequence of possible events in which the probability of each event depends only on the state

346 attained by the previous event[61]. In the case of VFDs, three possible states could be
347 defined: intubated, extubated, or dead.

348 The use of simulations to estimate the sample size should probably be encouraged,
349 especially for VFDs, given their complex probability distribution[8].

350 **Limitations**

351 We only selected studies involving adults to focus on one type of population and
352 reduce population heterogeneity. Therefore, only RCTs were included because these studies
353 must report sample size estimation and are less prone to bias. Moreover, because we knew
354 there was heterogeneity in how VFDs are defined, we selected studies published only in
355 journals with a 2021 impact factor greater than five in the hopes of a more rigorous
356 methodology. However, in this systematic review we found few unbiased studies.

357 The selected studies included different populations, mainly because the inclusion
358 criteria of the reported studies were quite different, thus reducing the possibility of
359 generalizing our results. However, we focused on sample size estimation, which was not
360 affected by differences between studies. Moreover, some data of interest were not reported
361 in many studies, which restrained sample size estimation. Additionally, Harhay et al.[6]
362 found that 63% of the power parameters were unreported in selected RCTs.

363 **Strengths**

364 To our knowledge, this is the first systematic review of 26 RCTs assessing how sample
365 sizes are estimated in trials with VFDs as the primary outcome. First, we followed the
366 PRISMA 2020 statement guidelines with checklists[17] (Additional file 1: Tables S8 and S9).
367 Second, two investigators independently reviewed the studies. Third, we used two

368 databases to be as comprehensive as possible. In addition, we did not place any limit on the
369 publication year: All studies referenced since the creation of the database were therefore
370 included but restricted to journals with an impact factor greater than five. Finally, we used
371 the RoB2 tool to evaluate any potential biases.

372 **Conclusions**

373 In this systematic review of RCTs with VFDs as the primary outcome, we observed
374 strong variability in the methods and results of sample size estimation, in addition to
375 heterogeneity in the definitions of VFDs. Moreover, the uncommon distribution of the
376 number of VFDs in clinical trials may have important implications which warrants further
377 investigation, such as for sample size estimation and analysis. Complex models and
378 simulation might be useful for sample size estimation when using VFDs as a primary
379 outcome in future trials. The methods used are of great importance as they directly impact
380 the number of patients to enroll and could jeopardize the feasibility of a trial, due to ethical,
381 logistical, and financial reasons.

382 **ABBREVIATIONS**

383 EPVent-2: The Esophageal Pressure-Guided Ventilation 2 trial

384 GAMLSS: generalized additive model for location, scale, and shape

385 GLMM: generalized linear mixed model

386 IQR: interquartile range

387 PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

388 RCTs: randomized clinical trials

389 RoB2: Version 2 of the Cochrane risk-of-bias tool

390 VFDs: ventilator-free days

391 ZINB: zero-inflated negative binomial

392 **DECLARATIONS**

393 Ethics approval and consent to participate: not applicable.

394 Consent for publication: not applicable.

395 Availability of data and materials: the datasets used and/or analysed during the current
396 study are available from the corresponding author on reasonable request. An annotated R
397 code for sample size simulations is available in open access in Zenodo[62].

398 Competing interests: the authors declare that they have no competing interests.

399 Funding: this work was supported by grants from the French Agence Nationale de la
400 Recherche and Direction Générale de l'Offre de Soins ("Programmes de Recherche
401 Translationnelle en Santé" ANR-13- PRTS-0010 and ANR-20-CE17-0015; "Programme
402 Hospitalier de Recherche Clinique" 2018), and the "Fonds Européen de Développement
403 Régional" (European Union, FEDER Auvergne 2020) (M. Jabaudon). The funder had no role in
404 the study design, data collection and analysis, the preparation or approval of the manuscript,
405 or the decision to submit the manuscript for publication.

406 Authors' contributions (CRediT author statement): LRT: Conceptualization, Methodology,
407 Formal analysis, Investigation, Writing – Original Draft, Visualization; EF: Validation, Writing
408 – Review & Editing, Supervision; MDC: Investigation, Resources, Writing – Review & Editing;
409 NPD: Investigation, Resources, Writing – Review & Editing; LBC: Resources, Writing – Review
410 & Editing; MJ: Conceptualization, Methodology, Validation, Writing – Review & Editing,

411 Supervision; BP: Conceptualization, Methodology, Validation, Formal analysis, Writing –
412 Review & Editing, Supervision. All authors read and approved the final manuscript.

413 Acknowledgements: the authors thank Ambre Salis for her help with programming and
414 proofreading the article, Myriam Garrido, Bruno Giraudeau and Caroline Mollevi for their
415 advice about the statistical analysis, and Marie De Antonio and Céline Lambert for their help
416 with programming. Research reported in this publication was supported by DaCCoTA (the
417 National Institute of General Medical Sciences of the National Institutes of Health under
418 Award Number U54GM128729).

419 REFERENCES

- 420 1. Marshall DC, Hatch RA, Gerry S, Young JD, Watkinson P. Conditional Survival With
421 Increasing Duration of ICU Admission: An Observational Study of Three Intensive Care
422 Databases. *Crit Care Med.* 2020;48:91–7.
- 423 2. Hill AD, Fowler RA, Burns KEA, Rose L, Pinto RL, Scales DC. Long-Term Outcomes and
424 Health Care Utilization after Prolonged Mechanical Ventilation. *Ann Am Thorac Soc.*
425 2017;14:355–62.
- 426 3. Herridge MS, Chu LM, Matte A, Tomlinson G, Chan L, Thomas C, et al. The RECOVER
427 Program: Disability Risk Groups and 1-Year Outcome after 7 or More Days of Mechanical
428 Ventilation. *Am J Respir Crit Care Med.* 2016;194:831–44.
- 429 4. Veldhoen RA, Howes D, Maslove DM. Is Mortality a Useful Primary End Point for Critical
430 Care Trials? *Chest.* 2020;158:206–11.
- 431 5. Spragg RG, Bernard GR, Checkley W, Curtis JR, Gajic O, Guyatt G, et al. Beyond Mortality:
432 future clinical research in acute lung injury. *Am J Respir Crit Care Med.* 2010;181:1121–7.

- 433 6. Harhay MO, Wagner J, Ratcliffe SJ, Bronheim RS, Gopal A, Green S, et al. Outcomes and
434 Statistical Power in Adult Critical Care Randomized Trials. *Am J Respir Crit Care Med.*
435 2014;189:1469–78.
- 436 7. Novack V, Beitler JR, Yitshak-Sade M, Thompson BT, Schoenfeld DA, Rubenfeld G, et al.
437 Alive and Ventilator-Free: A Hierarchical, Composite Outcome for Clinical Trials in the
438 Acute Respiratory Distress Syndrome. *Crit Care Med.* 2020;48:158–66.
- 439 8. Yehya N, Harhay MO, Curley MAQ, Schoenfeld DA, Reeder RW. Reappraisal of Ventilator-
440 Free Days in Critical Care Research. *Am J Respir Crit Care Med.* 2019;200:828–36.
- 441 9. Bernard GR, Artigas A, Brigham KL, Carlet J, Falke K, Hudson L, et al. The American-
442 European Consensus Conference on ARDS. Definitions, mechanisms, relevant outcomes,
443 and clinical trial coordination. *Am J Respir Crit Care Med.* 1994;149:818–24.
- 444 10. Schoenfeld DA, Bernard GR. Statistical evaluation of ventilator-free days as an efficacy
445 measure in clinical trials of treatments for acute respiratory distress syndrome. *Crit Care*
446 *Med.* 2002;30:1772–7.
- 447 11. Bolarinwa O. Sample size estimation for health and social science researchers: The
448 principles and considerations for different study designs. *Niger Postgrad Med J.*
449 2020;27:67–75.
- 450 12. Lenth RV. Some Practical Guidelines for Effective Sample Size Determination. *The*
451 *American Statistician.* 2001;55:187–93.
- 452 13. Chow S-C, Shao J, Wang H, Lokhnygina Y. *Sample Size Calculations in Clinical Research.*
453 3rd ed. New York: Chapman and Hall/CRC; 2017.
- 454 14. Béduneau G, Pham T, Schortgen F, Piquilloud L, Zogheib E, Jonas M, et al. Epidemiology
455 of Weaning Outcome according to a New Definition. The WIND Study. *Am J Respir Crit*
456 *Care Med.* 2017;195:772–83.

- 457 15. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing
458 Risk. *J Am Stat Assoc.* 1999;94:496–509.
- 459 16. Contentin L, Ehrmann S, Giraudeau B. Heterogeneity in the Definition of Mechanical
460 Ventilation Duration and Ventilator-Free Days. *Am J Respir Crit Care Med.*
461 2014;189:998–1002.
- 462 17. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The
463 PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.*
464 2021;372:n71.
- 465 18. Renard Triché L, Jabaudon M, De Carvalho M, Piñol-Domenech N, Pereira B. Sample Size
466 Estimation in Randomized Controlled Trials using Ventilator-Free Days as an Outcome:
467 Protocol for a Systematic Review. [Internet]. PROSPERO CRD42021282304. 2021 [cited
468 2022 Nov 6]. Available from:
469 https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021282304
- 470 19. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O’Neal L, et al. The REDCap
471 consortium: Building an international community of software platform partners. *J*
472 *Biomed Inform.* 2019;95:103208.
- 473 20. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data
474 capture (REDCap) -- a metadata-driven methodology and workflow process for providing
475 translational research informatics support. *J Biomed Inform.* 2009;42:377–81.
- 476 21. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised
477 tool for assessing risk of bias in randomised trials. *BMJ.* 2019;366:l4898.
- 478 22. R Core Team. R: A Language and Environment for Statistical Computing [Internet].
479 Vienna, Austria: R Foundation for Statistical Computing. [cited 2022 Feb 13]. Available
480 from: <https://www.r-project.org/>

- 481 23. Morris PE, Papadakos P, Russell JA, Wunderink R, Schuster DP, Truwit JD, et al. A double-
482 blind placebo-controlled study to evaluate the safety and efficacy of L-2-oxothiazolidine-
483 4-carboxylic acid in the treatment of patients with acute respiratory distress syndrome.
484 Crit Care Med. 2008;36:782–8.
- 485 24. Paine R, Standiford TJ, Dechert RE, Moss M, Martin GS, Rosenberg AL, et al. A
486 randomized trial of recombinant human GM-CSF for patients with acute lung injury. Crit
487 Care Med. 2012;40:90–7.
- 488 25. Mackle D, Bellomo R, Bailey M, Beasley R, Deane A, Eastwood G, et al. Conservative
489 Oxygen Therapy during Mechanical Ventilation in the ICU. N Engl J Med. 2020;382:989–
490 98.
- 491 26. Villar J, Ferrando C, Martínez D, Ambrós A, Muñoz T, Soler JA, et al. Dexamethasone
492 treatment for the acute respiratory distress syndrome: a multicentre, randomised
493 controlled trial. Lancet Respir Med. 2020;8:267–76.
- 494 27. Zhou Y, Jin X, Lv Y, Wang P, Yang Y, Liang G, et al. Early application of airway pressure
495 release ventilation may reduce the duration of mechanical ventilation in acute
496 respiratory distress syndrome. Intensive Care Med. 2017;43:1648–59.
- 497 28. Trouillet J-L, Luyt C-E, Guiguet M, Ouattara A, Vaissier E, Makri R, et al. Early
498 Percutaneous Tracheotomy Versus Prolonged Intubation of Mechanically Ventilated
499 Patients After Cardiac Surgery. Ann Intern Med. 2011;154:373–83.
- 500 29. Simonis FD, Neto AS, Binnekade JM, Braber A, Bruin KCM, Determann RM, et al. Effect of
501 a Low vs Intermediate Tidal Volume Strategy on Ventilator-Free Days in Intensive Care
502 Unit Patients Without ARDS: A Randomized Clinical Trial. JAMA. 2018;320:1872–80.

- 503 30. Algera AG, Pisani L, Neto AS, Gama de Abreu M, Pelosi P, Schultz MJ, et al. Effect of a
504 Lower vs Higher Positive End-Expiratory Pressure Strategy on Ventilator-Free Days in ICU
505 Patients Without ARDS: A Randomized Clinical Trial. *JAMA*. 2020;324:2509–20.
- 506 31. Tomazini BM, Maia IS, Cavalcanti AB, Berwanger O, Rosa RG, Veiga VC, et al. Effect of
507 Dexamethasone on Days Alive and Ventilator-Free in Patients With Moderate or Severe
508 Acute Respiratory Distress Syndrome and COVID-19: The CoDEX Randomized Clinical
509 Trial. *JAMA*. 2020;324:1307–16.
- 510 32. Grieco DL, Menga LS, Cesarano M, Rosà T, Spadaro S, Bitondo MM, et al. Effect of Helmet
511 Noninvasive Ventilation vs High-Flow Nasal Oxygen on Days Free of Respiratory Support
512 in Patients With COVID-19 and Moderate to Severe Hypoxemic Respiratory Failure: The
513 HENIVOT Randomized Clinical Trial. *JAMA*. 2021;325:1731–43.
- 514 33. van Meenen DMP, van der Hoeven SM, Binnekade JM, de Borgie CAJM, Merkus MP,
515 Bosch FH, et al. Effect of On-Demand vs Routine Nebulization of Acetylcysteine With
516 Salbutamol on Ventilator-Free Days in Intensive Care Unit Patients Receiving Invasive
517 Ventilation: A Randomized Clinical Trial. *JAMA*. 2018;319:993–1001.
- 518 34. Spragg RG, Lewis JF, Walmrath H-D, Johannigman J, Bellingan G, Laterre P-F, et al. Effect
519 of Recombinant Surfactant Protein C–Based Surfactant on the Acute Respiratory Distress
520 Syndrome. *N Engl J Med*. 2004;351:884–92.
- 521 35. Welte T, Dellinger RP, Ebelt H, Ferrer M, Opal SM, Singer M, et al. Efficacy and safety of
522 trimodulin, a novel polyclonal antibody preparation, in patients with severe community-
523 acquired pneumonia: a randomized, placebo-controlled, double-blind, multicenter,
524 phase II trial (CIGMA study). *Intensive Care Med*. 2018;44:438–48.

- 525 36. Rice TW, Wheeler AP, Thompson BT, deBoisblanc BP, Steingrub J, Rock P, et al. Enteral
526 Omega-3 Fatty Acid, γ -Linolenic Acid, and Antioxidant Supplementation in Acute Lung
527 Injury. *JAMA*. 2011;306:1574–81.
- 528 37. Bernard GR, Francois B, Mira J-P, Vincent J-L, Dellinger RP, Russell JA, et al. Evaluating the
529 Efficacy and Safety of Two Doses of the Polyclonal Anti-Tumor Necrosis Factor- α
530 Fragment Antibody AZD9773 in Adult Patients With Severe Sepsis and/or Septic Shock:
531 Randomized, Double-Blind, Placebo-Controlled Phase IIb Study*. *Crit Care Med*.
532 2014;42:504–11.
- 533 38. Chung KK, Wolf SE, Renz EM, Allan PF, Aden JK, Merrill GA, et al. High-frequency
534 percussive ventilation and low tidal volume ventilation in burns: A randomized
535 controlled trial. *Crit Care Med*. 2010;38:1970–7.
- 536 39. Rice TW, Wheeler AP, Thompson BT, Steingrub J, Hite RD, Moss M, et al. Initial Trophic vs
537 Full Enteral Feeding in Patients With Acute Lung Injury: The EDEN Randomized Trial.
538 *JAMA*. 2012;307:795–803.
- 539 40. Bein T, Weber-Carstens S, Goldmann A, Müller T, Staudinger T, Brederlau J, et al. Lower
540 tidal volume strategy (≈ 3 ml/kg) combined with extracorporeal CO₂ removal versus
541 ‘conventional’ protective ventilation (6 ml/kg) in severe ARDS: the prospective
542 randomized Xtravent-study. *Intensive Care Med*. 2013;39:847–56.
- 543 41. Hodgson CL, Cooper DJ, Arabi Y, King V, Bersten A, Bihari S, et al. Maximal Recruitment
544 Open Lung Ventilation in Acute Respiratory Distress Syndrome (PHARLAP). A Phase II,
545 Multicenter Randomized Controlled Clinical Trial. *Am J Respir Crit Care Med*.
546 2019;200:1363–72.

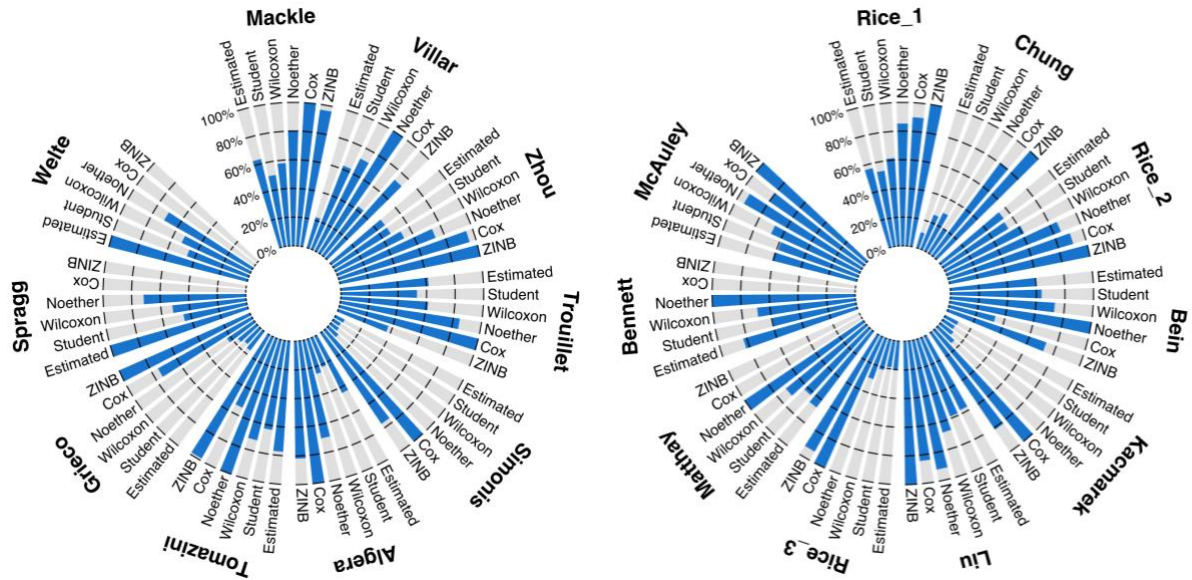
- 547 42. Kacmarek RM, Villar J, Parrilla D, Alba F, Solano R, Liu S, et al. Neurally adjusted
548 ventilatory assist in acute respiratory failure: a randomized controlled trial. *Intensive*
549 *Care Med.* 2020;46:2327–37.
- 550 43. Kacmarek RM, Wiedemann HP, Lavin PT, Wedel MK, Tütüncü AS, Slutsky AS. Partial
551 Liquid Ventilation in Adult Patients with Acute Respiratory Distress Syndrome. *Am J*
552 *Respir Crit Care Med.* 2006;173:882–9.
- 553 44. Liu KD, Levitt J, Zhuo H, Kallet RH, Brady S, Steingrub J, et al. Randomized Clinical Trial of
554 Activated Protein C for the Treatment of Acute Lung Injury. *Am J Respir Crit Care Med.*
555 2008;178:618–23.
- 556 45. Rice TW, Mogan S, Hays MA, Bernard GR, Jensen GL, Wheeler AP. Randomized Trial of
557 Initial Trophic versus Full-Energy Enteral Nutrition in Mechanically Ventilated Patients
558 with Acute Respiratory Failure. *Crit Care Med.* 2011;39:967–74.
- 559 46. Matthay MA, Brower RG, Carson S, Douglas IS, Eisner M, Hite D, et al. Randomized,
560 Placebo-controlled Clinical Trial of an Aerosolized β 2-Agonist for Treatment of Acute
561 Lung Injury. *Am J Respir Crit Care Med.* 2011;184:561–8.
- 562 47. Bennett-Guerrero E, Romeiser JL, Talbot LR, Ahmed T, Mamone LJ, Singh SM, et al.
563 Severe Acute Respiratory Syndrome Coronavirus 2 Convalescent Plasma Versus Standard
564 Plasma in Coronavirus Disease 2019 Infected Hospitalized Patients in New York: A
565 Double-Blind Randomized Trial. *Crit Care Med.* 2021;49:1015–25.
- 566 48. McAuley DF, Laffey JG, O’Kane CM, Perkins GD, Mullan B, Trinder TJ, et al. Simvastatin in
567 the Acute Respiratory Distress Syndrome. *N Engl J Med.* 2014;371:1695–703.
- 568 49. Bodet-Contentin L, Frasca D, Tavernier E, Feuillet F, Foucher Y, Giraudeau B. Ventilator-
569 Free Day Outcomes Can Be Misleading: *Crit Care Med.* 2018;46:425–9.

- 570 50. Jabaudon M, Blondonnet R, Pereira B, Cartin-Ceba R, Lichtenstern C, Mauri T, et al.
571 Plasma sRAGE is independently associated with increased mortality in ARDS: a meta-
572 analysis of individual patient data. *Intensive Care Med.* 2018;44:1388–99.
- 573 51. Blackwood B, Ringrow S, Clarke M, Marshall JC, Connolly B, Rose L, et al. A Core Outcome
574 Set for Critical Care Ventilation Trials. *Crit Care Med.* 2019;47:1324–31.
- 575 52. Beitler JR, Sarge T, Banner-Goodspeed VM, Gong MN, Cook D, Novack V, et al. Effect of
576 Titrating Positive End-Expiratory Pressure (PEEP) With an Esophageal Pressure–Guided
577 Strategy vs an Empirical High PEEP-Fio2 Strategy on Death and Days Free From
578 Mechanical Ventilation Among Patients With Acute Respiratory Distress Syndrome: A
579 Randomized Clinical Trial. *JAMA.* 2019;321:846–57.
- 580 53. Lazarus SC, Krishnan JA, King TS, Lang JE, Blake KV, Covar R, et al. Mometasone or
581 Tiotropium in Mild Asthma with a Low Sputum Eosinophil Level. *N Engl J Med.*
582 2019;380:2009–19.
- 583 54. Bartunek J, Terzic A, Davison BA, Behfar A, Sanz-Ruiz R, Wojakowski W, et al.
584 Cardiopoietic stem cell therapy in ischaemic heart failure: long-term clinical outcomes.
585 *ESC Heart Fail.* 2020;7:3345–54.
- 586 55. Ludbrook J. Should we use one-sided or two-sided P values in tests of significance? *Clin*
587 *Exp Pharmacol Physiol.* 2013;40:357–61.
- 588 56. Zeileis A, Kleiber C, Jackman S. Regression Models for Count Data in R. *Journal of*
589 *Statistical Software.* 2008;27:1–25.
- 590 57. Steingrimsson JA, Hanley DF, Rosenblum M. Improving precision by adjusting for
591 prognostic baseline variables in randomized trials with binary outcomes, without
592 regression model assumptions. *Contemp Clin Trials.* 2017;54:18–24.

- 593 58. Tai B, Chen Z, Machin D. Estimating sample size in the presence of competing risks –
594 Cause-specific hazard or cumulative incidence approach? Stat Methods Med Res.
595 2018;27:114–25.
- 596 59. Granholm A, Kaas-Hansen BS, Lange T, Munch MW, Harhay MO, Zampieri FG, et al. Use
597 of days alive without life support and similar count outcomes in randomised clinical trials
598 – an overview and comparison of methodological choices and analysis methods. BMC
599 Medical Research Methodology. 2023;23:139.
- 600 60. Kumle L, Vö ML-H, Draschkow D. Estimating power in (generalized) linear mixed models:
601 An open introduction and tutorial in R. Behav Res Methods. 2021;53:2528–43.
- 602 61. Gagniuc PA. Markov chains: from theory to implementation and experimentation.
603 Hoboken, NJ: John Wiley & Sons; 2017.
- 604 62. Renard Triché L. Sample size estimation according to different tests [Internet]. Zenodo;
605 2023 Jun. Available from: <https://zenodo.org/record/8070734>
606

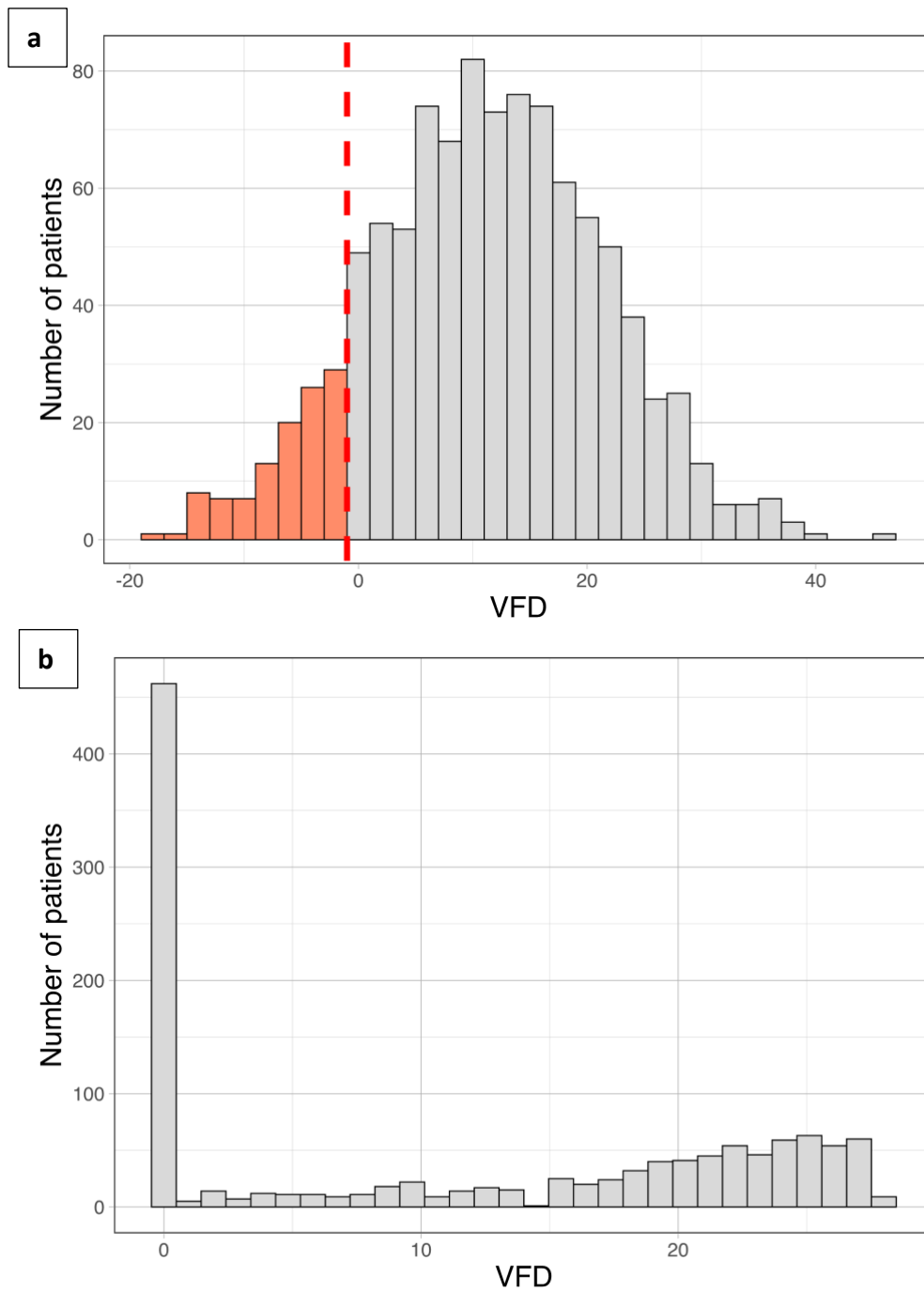
607 **FIGURES**

608 **Figure 1. Sample size estimation as reported in each trial and computed according to**
 609 **different alternative tests.**



610 For each study, sample size estimation is plotted (in blue) against the highest value among the sample size **estimated** in the study and five
 611 different tests: the **Student** *t*-test, the Mann–Whitney–**Wilcoxon** rank-sum test, the Mann–Whitney–Wilcoxon rank-sum test using the
 612 **Noether** formula, **Cox** regression and zero-inflated negative binomial (**ZINB**) regression. When an estimation is missing, the whole length of
 613 the line is gray. The estimation was only possible for the following studies: Mackle[25]; Villar[26]; Zhou[27]; Trouillet[28]; Simonis[29];
 614 Algra[30]; Tomazini[31]; Grieco[32]; Spragg[34]; Welte[35]; Rice_1[36]; Chung[38]; Rice_2[39]; Bein[40]; Kacmarek[42]; Liu[44];
 615 Rice_3[45]; Matthay[46]; Bennett[47]; and MaAuley[48].

616 **Figure 2. Distribution of ventilator-free days (VFDs) in selected trials.**



617 Histograms representing (a) a Gaussian distribution (mean of 11.7 days, standard deviation of 10.5) used in some studies for sample size
618 estimation and (b) the empirical distribution of VFDs (mean of 11.7 days, standard deviation of 10.71 and median of 12.23, interquartile
619 range 0.00–22.00) found in Jabaudon et al.'s meta-analysis.[50] The red bars correspond to the theoretical data that should be seen if the
620 distribution were normal.

621 **TABLES**622 **Table 1. Characteristics of the included studies.**

Variable	n/N (%) or median [IQR]
Year of publication	2014 [2011 – 2020]
Journal	
AIM	1/26 (4)
AJRCCM	4/26 (15)
CCM	6/26 (23)
ICM	4/26 (15)
JAMA	7/26 (27)
LRM	1/26 (4)
NEJM	3/26 (12)
Number of group – ratio (E/C)*	
2 – 1:1	23/26 (88)
2 – 4:1	1/26 (4)
3 – 1:1:1	2/26 (8)
Study type	
Superiority	24/26 (92)
noninferiority	2/26 (8)
Number of center if multicentric [†]	23 [8 – 42]
Interim analysis	16/26 (62)
Stopped earlier than expected	10/26 (39)
Population	
ARDS or ALI	14/26 (54)
COVID-19	3/26 (12)
Overall RoB2 [‡]	
low risk	3/26 (12)
some concerns	19/26 (73)
high risk	4/26 (15)

623 *AIM* Annals of internal medicine, *AJRCCM* American Journal of Respiratory and Critical Care Medicine, *ALI* acute lung injury, *ARDS* acute
624 respiratory distress syndrome, *CCM* Critical Care Medicine, *ICM* Intensive Care Medicine, *IQR* interquartile range, *JAMA* Journal of the
625 American Medical Association, *LRM* The Lancet Respiratory Medicine, *n* number of study, *N* total number of studies, *NEJM* The New
626 England Journal of Medicine.

627 *The allocation ratio (E for experimental group and C for control group).

628 [†]There were 22 (85%) studies with more than one center.

629 [‡]The overall risk-of-bias via RoB2 (Revised Cochrane risk-of bias tools for randomized trials)[21].

Table 2. Parameters reported for the sample size estimation of ventilator-free days.

Variable	n/N (%) or median [IQR]	NA (%)
Expected VFDs		
control group	14.00 [11.90 – 16.70]	10 (38)
experimental group	16.25 [14.88 – 19.77]	10 (38)
mean difference	2.60 [2.00 – 7.00]	3 (12)
standard deviation	10.00 [6.75 – 10.53]	6 (23)
VFDs distribution		13 (50)
asymmetric	3/26 (11)	
bimodal	1/26 (4)	
normal	2/26 (8)	
not normal	5/26 (19)	
zero-inflated distribution	2/26 (8)	
Correction if distribution was not considered normal		15 (57)
+ 15%	3/26 (12)	
Markov chains	1/26 (4)	
median comparison	1/26 (4)	
no correction	2/26 (8)	
nonparametric test	4/26 (15)	
Statistical model		5 (19)
ANOVA	5/26 (19)	
Kruskal-Wallis	1/26 (4)	
Student <i>t</i> -test	2/26 (8)	
Mann–Whitney–Wilcoxon rank-sum test	8/26 (30)	
quantile regression	1/26 (4)	
Cox regression	1/26 (4)	
GAMLSS	1/26 (4)	
GLM	2/26 (8)	
risk	5 [5 – 5]	0 (0)
Power (%)	80 [80 – 80]	1 (4)
Two-tailed test	21 (81)	1 (4)
Dropout rate expected	3.00 [0.00 – 9.00]	0 (0)
Mortality considered for sample size estimation	1/26 (4)	0 (0)

631

ANOVA analysis of variance, GAMLSS generalized additive model for location scale and shape, GLM generalized linear model, IQR

632

interquartile range, *n* number of study, *N* total number of studies, NA not available, VFDs ventilator-free days.

633 **Table 3. Definitions of ventilator-free days.**

Variable	n/N (%)
Onset	
no onset reported	9/26 (35)
> 4h	1/26 (4)
> 24h	3/26 (11)
> 48h	10/26 (38)
> 72h	1/26 (4)
portion of day with no onset reported	2/26 (8)
Respiratory support	
invasive ventilation	13/26 (50)
tracheostomy	12/26 (46)
ECMO	2/26 (8)
NIV	6/26 (23)
HFNO	1/26 (4)
no detail reported	13/26 (50)
Reintubation	
possible	7/26 (27)
reset	9/26 (35)
no detail reported	10/26 (38)
Day when VFDs are defined	
28	24/26 (92)
30	1/26 (4)
60	1/26 (4)

ECMO extracorporeal membrane oxygenation, HFNO high-flow nasal oxygen, *n* number of study, *N* total number of studies, NIV

noninvasive ventilation, *onset* time at which the patient is considered free of ventilation and start counting whole days except for two studies that considered portion of day, VFDs ventilator-free days.