Exploring the Feasibility of Using Real-World Data from a Large Clinical Data Research Network to Simulate Clinical Trials of Alzheimer's Disease

Zhaoyi Chen[1*], Hansi Zhang[1*], Yi Guo[1], Thomas J George[2], Mattia Prosperi[3], William R Hogan[1], Zhe He[4], Elizabeth A Shenkman[1], Fei Wang[5], Jiang Bian[1]

[1] Health Outcomes & Biomedical Informatics, University of Florida, Gainesville, Florida, USA

[2] Hematology & Oncology, Department of Medicine, College of Medicine, University of Florida, Gainesville, Florida, USA

[3] Department of Epidemiology, College of Public Health and Health Professions & College of Medicine, University of Florida, Gainesville, Florida, USA

[4] School of Information, Florida State University, Tallahassee, Florida, USA

[5] Department of Healthcare Policy and Research, Cornell University, New York, USA

Corresponding author: Jiang Bian, PhD

College of Medicine

University of Florida

2197 Mowry Road Suite 122

PO Box 100177

Gainesville, FL 32610-0177

bianjiang@ufl.edu

352 273-8878

Keywords: trial simulation, real-world data, Alzheimer's Disease

ABSTRACT

In this study, we explored the feasibility of using real-world data (RWD) from a large clinical research network to simulate real-world clinical trials of Alzheimer's disease (AD). The target trial (i.e., NCT00478205) is a Phase III double-blind, parallel-group trial that compared the 23 mg donepezil sustained release with the 10 mg donepezil immediate release formulation in patients with moderate to severe AD. We followed the target trial's study protocol to identify the study population, treatment regimen assignments, and outcome assessments, and to set up a number of different simulation scenarios and parameters. We considered two main scenarios: (1) a one-arm simulation: simulating a standard-of-care arm that can serve as an external control arm; and (2) a two-arm simulation: simulating both intervention and control arms with proper patient matching algorithms for comparative effectiveness analysis. In the two-arm simulation scenario, we used propensity score matching controlling for baseline characteristics to simulate the randomization process. In the two-arm simulation, higher SAE rates were observed in the simulated trials than the rates reported in original trial, and a higher SAE rate was observed in the 23mg arm than the 10 mg standard-of-care arm. In the one-arm simulation scenario, similar estimates of SAE rates were observed when proportional sampling was used to control demographic variables. In conclusion, trial simulation using RWD is feasible in this example of AD trial in terms of safety evaluation. Trial simulation using RWD could be a valuable tool for post-market comparative effectiveness studies and for informing future trials' design. Nevertheless, such approach may be limited, for example, by the availability of RWD that matches the target trials of interest, and further investigations are warranted.

Introduction

Clinical trials, especially randomized controlled trials (RCTs), are critical in the drug discovery and development process to assess the efficacy and safety of the new treatment.[1] While the rigorously controlled conditions of clinical trials can reduce bias and improve the internal validity of the study results, they also come with the drawbacks of high financial costs and long execution time.[2] For example, the total cost of developing an Alzheimer's disease (AD) drug was estimated at $5.6 billion with a timeline of 13 years from the preclinical studies to approval by the Food and Drug Administration (FDA).[3] Nevertheless, no effective drugs yet have been developed for either treatment or prevention of AD thus far. Strategies that can accelerate the drug development process and reduce costs will not only be of interest to pharmaceutical companies but also ultimately benefit the patients.

Clinical trial simulation (CTS) is a valuable to assess the feasibility, investigate assumptions, and optimize study design before conducting the actual trials.[4,5] For example, Romero *et al.* conducted a CTS study to explore several design scenarios comparing the effects of donepezil with placebo.[6] Traditionally, CTS studies use virtual cohorts generated based on pharmacokinetics / pharmacodynamics models of the therapeutic agents, so these cohorts do not necessarily reflect the patients who will use the drugs in the real world. More recently, the trial emulation (i.e., "the target trial") framework—emulating hypothetical trials to establish the estimation of the casual effects, has attracted significant attention.[7] For example, Danaei *et al.* emulated a hypothetical RCT used electronic health record (EHR) data from United Kingdom (UK) to estimate the effect of statins for primary prevention of coronary heart disease.[8] Like many other emulation studies,[7,9,10] this is essentially a retrospective cohort study, where the authors followed a RCT design to identify

68 unbiased initiation of exposures and eventually to reach an unbiased estimation of the casual
69 relationship. Combining the ideas from CTS and trial emulation, a simulation study using real-
70 world data (RWD) to test different assumptions (e.g., different drop-out rates) and trial designs (e.g.,
71 different eligibility criteria) could provide insights on the effectiveness and safety of the treatments
72 to be developed in a real-world setting that reflect the patient populations who will actually use the
73 treatment.
74
75 In this study, we explored the feasibility of using RWD from the OneFlorida Clinical Research
76 Consortium—a clinical data research network funded by the Patient-Centered Outcomes Research
77 Institute (PCORI) contributing to the national Patient-Centered Clinical Research Network
78 (PCORnet)—to simulate a real-world AD RCT as a use case. We considered two main scenarios: (1)
79 a one-arm simulation: simulating a standard-of-care arm that can serve as an external control arm;
80 and (2) a two-arm simulation: simulating both intervention and control arms with proper patient
81 matching algorithms for comparative effectiveness analysis.
82
83 **Results**
84 ***Computability of eligibility criteria in the original trial (i.e., NCT00478205)***
85 In total, there are 36 eligibility criteria in trial NCT00478205, where 17 are inclusion and 19 are
86 exclusion criteria. However, not all criteria are computable against the OneFlorida patient database:
87 (1) 11 are not computable, and (2) 7 are partially computable (i.e., a part of the criterion is not
88 computable). Similar to what we have found in our prior study[11], the common reasons for not
89 computable criteria are (1) data elements needed for the criterion do not exist in the source database
90 (e.g., "*A cranial image is required, with no evidence of focal brain disease that would account for dementia.*"),

91 or (2) the criterion asked for subjective information either from the patient (e.g., *"Patients who are*
92 *unwilling or unable to fulfill the requirements of the study.") or the investigator (e.g., ""Clinical laboratory*
93 *values must be within normal limits or, if abnormal, must be judged not clinically significant by the*
94 *investigator."*). When a criterion is not computable, we consider all candidate patients met that
95 criterion (e.g., they are all willing and able to "*fulfill the requirements of the study*").
96
97 ***Characteristics of the target, study, and trial not eligible populations from OneFlorida***
98 Overall, a total of 90 and 2,048 patients were identified as the effective target populations in
99 OneFlorida for the 23 mg arm and 10 mg arm, respectively. Among them, 38 and 782 met the
100 eligibility criteria of the original target RCT for the two arms, respectively. ***Table 2*** shows the
101 demographic characteristics and SAE statistics of the original trial population as well as the effective
102 target population (TP), study population (SP), and trial not eligible population (NEP) from
103 OneFlorida.
104
105 For demographic characteristics, relative to the target RCT population, we observed a large
106 difference in race in our OneFlorida population (all p-values of race group comparisons were
107 smaller than 0.05). OneFlorida had more Hispanics (10.5% - 24.6% vs. 5.5% - 7%) and Blacks (10.5% -
108 20.1% vs. 1.9% - 2.3%), but less Whites (35.8% - 73.1% vs. 73.5% - 73.5%) or Asian/Pacific islanders
109 (0% - 1.4% vs. 16.7% - 18.5%). The age distributions were similar across all populations. For clinical
110 variables, we calculated the Charlson Comorbidity Index (CCI) of the various populations from
111 OneFlorida. Smaller CCIs were observed in the SP compared with the TP for both arms ($p<0.05$),
112 and a smaller CCI was observed in the 23mg arm compared with the 10mg arm ($p<0.05$). Our
113 primary outcomes of interest in this analysis were SAEs. Thus, we calculated the mean SAE (i.e., the

114 average number of SAEs per patient) and the number of patients who had more than 1 SAE during
115 the study period. For both 23mg and 10mg arms, the mean SAE and the number of patients with
116 SAEs were the largest in the TP, followed by the SP, and then the original trial. Consistent with the
117 original trial, populations derived from the OneFlorida data in the 23mg arm have higher numbers
118 of mean SAE and more patients with SAE compared with the 10mg arm.
119
120 ***Standard-of-care control arm (i.e., one-arm) simulation***
121 We first simulated the control arm of the original trial (i.e., the 10mg stand-of-care arm). ***Table 3***
122 displays the demographics and SAE outcomes in the simulated control arms. Here, we reported the
123 mean value and 95% confidence interval of all 1,000 bootstrap samples. Two different sampling
124 approaches were used: (1) random sampling, and (2) proportional sampling accounting for race
125 distribution. When using the random sampling approach, compared with the control arm in the
126 original trial, higher mean SAE and SAE rates were observed, in addition to discrepancies in
127 demographic variables. When using proportional sampling, the results were closer and more
128 consistent with the original trial. Notably, the SAE rates in the simulated control were similar to the
129 SAE rates from the original control (8.9% vs. 8.3%), and a z-score test for population proportion had
130 a p-value of 0.75, suggesting there were no significant differences between the two SAE rates. In
131 addition to SAE rates and mean SAE, we also explored the SAE event rates in the simulated control
132 arms stratified by the SAE category reported in the original trial. Compared with the control arm in
133 the original trial, the simulated control arms have larger SAE rates in most categories.
134
135 ***Two-arm trial simulation***

136 Because the proportional sampling had better performance in the one-arm simulation, we used this
137 sampling strategy in the two-arm simulation to match the race distribution, and tested two scenarios
138 of different matching ratios (i.e., proportional 1:1 matching and proportional 1:3 matching).
139 However, since there is no Asian/Pacific in our study population who used 23mg donepezil, all
140 Asians in the original trial and the simulated control arm were grouped into "other", and the sample
141 size of the 23mg arm was set at 30 (because of the limited number of 23mg patients in the
142 OneFlorida data). Table 4 shows our two-arm simulation results, where we show the average and
143 95% CI of all variables for the simulation arms across all 1,000 bootstrap samples. In both matching
144 scenarios, the mean SAE and SAE rates were higher in the 23mg arm than in the 10mg arm, which is
145 consistent with the original trial. However, the variance for both SAE outcomes for the 10mg arm
146 are higher in the 1:1 matching scenario than in the 1:3 matching scenario, as the sample size for the
147 10mg arm in the 1:3 matching scenario is much bigger. Because of the sample size difference,
148 estimates from the 1:3 matching scenario should be more reliable. Consistent with the original trial,
149 patients in the 23mg arm have higher event rates in most of the SAE categories compared to the
150 patients in the 10mg arm. Note that we observed no SAE events in several categories in our
151 simulation, especially in the 23mg arm, due to the limited sample size.
152
153 Finally, we conducted an additional experiment to simulate patients who withdrew from the trial.
154 In the original trial, among the 963 and 471 patients from each arm, 296 (30%) and 87 (18%) patients
155 discontinued the study for various of reason, respectively. Among the dropouts, 182 and 39 patients
156 discontinued due to AEs. We simulated the dropouts by (1) randomly removing 18% in the 10mg
157 group and 30% in the 23mg in our simulations; and (2) removing the patient after his or her first AE
158 using the same proportion as the original trial (i.e., due to small sample size, we did not simulate

159 this scenario for the two-arm simulation). The results are displayed in **Supplementary Table 3** and
160 **Supplementary Table 4**. For the random dropout scenario, similar SAE rates and smaller mean SAE
161 were observed across all scenarios. For example, in the control arm of the random dropout scenario,
162 mean SAE decreased from 0.64 to 0.19 and 0.23 in the two different dropout simulations, while in
163 the two-arm simulation, the mean SAE for the 10mg arm were 0.22 and 0.19 in the two scenarios,
164 both were much lower than simulations without dropout (0.46 and 0.47 respectively). However, the
165 effects of dropout were mostly observed in control arms, the SAE rates and mean SAE remained the
166 same for the 23mg arm before and after dropout simulation. In the AE-based dropout scenario, both
167 smaller SAE rates and smaller mean SAE were observed: the SAE rates decreased from 8.8% to 7.3%
168 and the mean SAE decreased from 0.64 to 0.23, where the mean SAE estimate was closer to the
169 original trial results.
170

171 **DISCUSSION**

172 In this work, we simulated an AD RCT utilizing RWD from the OneFlorida network—a large clinical
173 data research network, considering three different simulation scenarios. In the one-arm simulation
174 scenario, we attempted to simulate an external control arm for the original trial. We demonstrated
175 that we could achieve similar estimate of SAE rates as the original trial when proportional sampling
176 accounting for race distribution was used; and the statistics of the simulated control arm were stable
177 across all bootstrap simulation runs, which suggests that using RWD we can robustly simulate the
178 "*standard of care*" control arm. In the two-arm comparative effectiveness simulations, we used
179 propensity score matching on baseline characteristics to simulate the randomization process. It has
180 been demonstrated that propensity score matching could reduce bias in the estimate of the treatment
181 response,[12–15] and in our study, we successfully simulated two groups of patients that have similar

182 age, sex, race, and CCI distributions using propensity score matching. However, the SAE outcomes
183 in the simulated trial were still different from the original trial for various reasons (1) the original
184 trial was conducted in research settings, while RWD data reflect patients the real-world clinical
185 settings. The total time at risk for SAEs in our simulated cohort may be longer than the original trial,
186 because in clinical trials, patients may withdraw from the study once experiencing an SAE, while
187 patients in real-world setting may not. This is demonstrated by our simulation of dropouts, which
188 achieved smaller number of SAEs and closer results to the original trial; (2) sample size issue, where
189 23mg donepezil has not been the standard-of-care for AD in the real-world, leading to considerably
190 fewer patients in the 23mg arm. For the two-arm simulation, we conducted a post-hoc power
191 analysis with the SAE rates and mean SAE. Assuming a significant level at 0.05, a 65% power were
192 achieved; and (3) although propensity score matching derived two simulation arms that are
193 comparable, we were unable to compare it directly to the study population in the original trials as
194 the data for calculating propensity scores were not available from the original trial. For example, the
195 switching to the 23mg treatment after receiving at least 3 months of the 10mg does not occur at
196 random in a real-world setting, but based on clinical guidelines; and indeed, we found that patients
197 in the 23mg arm have a longer history of diagnosis (i.e., mean of days between first diagnosis and
198 first prescription in 23mg arm is 398 days vs. 128 days in the 10mg arm). Therefore, in our two-arm
199 simulation, there may be residue selection bias causing a difference between the two populations;
200 Nevertheless, this is an issue of using observational data in general; even though we can simulate
201 randomization, e.g., through propensity score matching; but trial simulations cannot replace RCTs.
202 In addition, there are still gaps, especially data gaps in RWD, that also contributed to the differences
203 between our simulation results and the original trial results. Future studies are warranted to
204 identify strategies to fill these gaps.

While simulating the original AD trial followed the study protocol in **Table 1**, we found it is difficult to replicate all the eligibility criteria of the original trial. Out of the 36 eligibility criteria, only 25 of them were computable or partially computable against the OneFlorida data. Since these criteria were used to weed out patients who are unlikely to complete the protocol (e.g., due to safety concerns), ignoring some of the criteria (not computable eligibility criteria) could potentially explain some of the increases either in the mean SAE or the SAE rates. One strategy for future simulation studies is to classify each of the eligibility criteria based on their clinical importance to the simulation study and the endpoints (i.e., effectiveness or safety) related with the criterion. By doing so, we can adjust the eligibility criteria and customize the simulation based on questions of interest. For example, efficacy-related criteria may have very small impact on a trial that is focused on examining safety and toxicity; so simulations of such trials can loosen the restrictions on efficacy-related criteria. Nevertheless, as all the patients we identified in the OneFlorida data have taken the study drugs of interest (i.e., different dosages of donepezil), they should all have been eligible to the original trial in an ideal world, where the trial participants truly reflect target population (i.e., higher trial generalizability).

Our findings are consistent with previous literature on clinical trial generalizability.[16–19] More SAEs were observed in real-world settings. In our data, the overall number of patients who had SAEs and the average number of SAEs per patient were (1) the highest in the effective target population (i.e., patients who took donepezil for AD), which is the population who actually used the medication in real-world settings, but also (2) higher in the study population—patients who used donepezil for AD and also met the original trial's eligibility criteria. Some of the differences may be due to the

228 incomputable eligibility criteria (e.g., general physical health deterioration) that we cannot account
229 for, but it is also possible that the original trial samples did not adequately reflect the TP and thus
230 there might be treatment effect heterogeneity across patient subgroups, not captured by the original
231 trial. In the two-arm simulations, large variances were observed, especially when the matched
232 sample size was small. This may also indicate the heterogeneous treatment effects of donepezil
233 when applied to different patient subgroups in real-world settings.
234
235 Our study demonstrated the feasibility of trial simulation using RWD, especially when simulating
236 external standard-of-care control arms. Our one-arm simulation provided stable and robust
237 estimates and sufficient sample sizes to compare with the original trial's control arm. The SAE rates
238 observed in the simulated control arm with proportional sampling were very close to what was
239 reported in the original trial. The mean SAE per patient and SAE event rates, however, were larger
240 in the simulated control arms, which suggested that, in a real-world setting, the patients who
241 experienced SAEs tend to have more occurrences of SAEs. On the other hand, the two-arm
242 simulation, although it provided insights, was not entirely successful. Although the randomization
243 process was effectively simulated by using propensity score matching, the outcome measures were
244 very different from the original trial. The reasons for the differences could be multi-fold (e.g.,
245 research setting vs. real-world clinical setting, difference in sample size, overly restrictive eligibility
246 criteria that limits the generalizability of the original trial), but cannot be explored due to limited
247 data reported by the original trial (i.e., no patient-level data is available). When we simulated
248 dropouts using information from the original trial, we observed similar SAE rates and mean SAE
249 compared to the original trial in both random dropout or AE-based dropout strategies, especially in
250 the control arms. This suggests that the additional simulation scenarios have led to results more

251 comparable to the cohorts in the original trial. However, due to the small sample size in the 23mg
252 arm, we were only able to simulate random dropout for the two-arm simulation, and the SAE
253 measurements did not change much comparing to the simulations without dropout. Future studies
254 with sufficient sample sizes could conduct more sophisticated analysis based on treatment delay and
255 adherence.
256
257 Compared with trial emulation, which focus on making the target trial explicitly characterized with
258 a defined protocol, our approach takes the advantage of having observational RWD, where different
259 trial protocols (i.e., simulation scenarios) with different study designs can be readily tested. For
260 example, in our current study, there are several potential simulation points that can be further
261 tuned. First, the sample size of each arm can be adjusted. In the one arm simulation (i.e., the control
262 arm of 10mg donepezil), we choose the same sample size as the original trial, but it can be adjusted
263 to increase power. Second, the eligibility criteria of selecting study population could also be
264 adjusted to test different hypotheses. For example, we can adjust the eligibility criteria in the trial
265 simulation process to assess how the original trial results may be generalized into real-world target
266 population, and provide insights on how to balance internal validity while retaining good external
267 validity.[17,20] Third, different scenarios of dropout may be simulated. The dropout rate and timing
268 can be varied, so that it can be used to simulate different patient population. Further, we can also
269 explore whether other dropout reasons such as lack of recovery and lack of access to care can be
270 simulated based on real-world data. Many other potential simulation scenarios can be tested such
271 as varying the 3-month lead time for switching from 10mg to 23mg. In this current work, as our
272 main goal was to establish the feasibility of such simulation approach, we only conducted limited
273 number of major simulation scenarios (e.g., we used two different sampling scenarios using

274 different intervention arm vs. control arm ratios). In future work, informed by literature, we shall
275 systematically simulate the different trial design scenarios, which can (1) provide critical information
276 on the comparative effectiveness of the interventions in real-world settings, but also (2) better inform
277 the study designs of future clinical trials. Last but not least, the one-arm simulation is as important
278 as the two-arm simulation, even though it does not provide comparative effectiveness results of the
279 intervention. In addition to informing future design of control arms, one-arm simulation allows us
280 to utilize readily available RWD of patients taking the standard-of-care (SOC) to determine SOC's
281 treatment effectiveness and safety profile and consider different study protocols and scenarios. The
282 demonstrated feasibility of one-arm simulation is a building block towards the potential of using
283 RWD to generate synthetic and external controls for clinical trials, leading to significant cost
284 savings.[21] Nevertheless, other issues with RWD such as its data quality (e.g. missing key measures
285 of endpoints) and the inherent biases exist in observational data, warrant further investigations.
286
287 There are some other limitations in this study. First, we only looked at one original trial for one
288 medication (i.e., donepezil). Simulations on different drugs and diseases may have different results.
289 Second, the population who took the 23mg form in our data is very small (even though the overall
290 OneFlorida population is large with more than 15 million patients), where we only identified 38
291 patients who took the 23mg donepezil and met the eligibility criteria of the original trial. The 23mg
292 donepezil form was approved by the FDA in 2012, so it is still a relatively new drug on the market,
293 and following its approval, the clinical utility of the 23mg form was called into question because of
294 its limited effectiveness and higher rates of adverse events.[22] The current practice of using the
295 donepezil 23mg form is reserved for AD patients who have been on stable donepezil 10 mg form for
296 at least 3–6 months with no significant improvement,[31,32] which limited its use in real-world clinical

297 practice. In addition, patients who switched to the 23mg treatment may have different
298 characteristics in characteristics that we did not account for in this analysis. Third, we found some
299 of the SAEs (e.g., abnormal behavior, presyncope) reported in the trial's results cannot be mapped to
300 any AE terms in CTCAE, and the definitions of AEs in the original trial were unavailable, which
301 increased the difficulty of accurately accounting for all SAEs. Further, even though trials' SAEs
302 reported in ClinicalTrials.gov largely follow the Medical Dictionary for Regulatory Activities
303 Terminology (MedDRA), not all reported SAEs were correctly defined in the trial results. For
304 example, we found "*Back pain*" and "*Fall*" were defined as SAEs in the original AD trial we modeled.
305 However, in CTCAE, there is no corresponding category 4 or 5 definition for them. More effort is
306 needed to consistently model SAEs reported in clinical trials. Finally, because of data limitations, we
307 were not able to assess the effectiveness of AD treatment (e.g., AD end points such as Mini-Mental
308 State Examination and Severe Impairment Battery are not readily available in structure EHR data,
309 but may exist in clinical narratives). Thus, we only examined safety outcomes in our current study.
310 This may also contribute to the different results we obtained from our simulation compared with
311 results from the original trial as the original trial was designed and powered with primary efficacy-
312 based outcomes. Future studies that explore the use of advanced natural language processing (NLP)
313 methods to extract these endpoint measurements from clinical notes will be important. Further,
314 variables extracted from clinical notes with NLP could also be used to render some of the
315 incomputable eligibility criteria computable.
316
317 In conclusion, in this study, we investigated the feasibility of using the existing patient records to
318 simulate clinical trials using an Alzheimer's disease trial (i.e. NCT00478205) as the use case. We
319 examined two main simulation scenarios: (1) a one-arm simulation: simulating the standard-of-care

320 arm that can serve as an external control arm; and (2) a two-arm simulation: simulating both
321 intervention and control arms with proper patient matching algorithms for comparative
322 effectiveness analysis. We have also considered a number of different simulation parameters such
323 sampling strategies, matching approaches, and dropout scenarios. In the case study, our simulation
324 can robustly simulate "standard of care" control arms (i.e. the 10mg donepezil arm) in terms of safety
325 evaluation. However, trial simulation using RWD may be limited by the availability of RWD that
326 matches the target trials of interest and may not yield reliable and consistent results if the sample
327 sizes of the interventions of interest (i.e. we found few patients were prescribed the 23mg donepezil)
328 are limited from the real-world databases. Further investigations on this topic are warranted,
329 especially how to address the data quality issues (e.g. using NLP to extract more complete patient
330 information) and reduce inherent biases (e.g. more advanced matching methods to tackle the
331 problems of high dimensionality, nonlinear/nonparallel treatment assignment, and other complex
332 confounding situations[25]) in observational RWD. Last but not least, it will also be beneficial to have
333 access more complete information (e.g. de-identified individual-level trail participant data) of the
334 target trials, so that more realistic simulation settings can be explored.
335

336 **Methods**

337 ***The target Alzheimer's disease (AD) trial and its characteristics***

338 Although there is no cure for AD yet, the U.S. FDA approved two classes of medications: (1)
339 cholinesterase inhibitors, and (2) memantine, to treat the symptoms of dementia. Donepezil
340 (Aricept®), a cholinesterase inhibitor, was the most widely tested AD drug and approved for all
341 stages of AD. The target trial NCT00478205[26] is a Phase III double-blind, double-dummy, parallel-
342 group comparison of 23 mg donepezil sustained release (SR) with the 10 mg donepezil immediate

343 release (IR) formulation (marketed as the standard-of-care) in patients with moderate to severe
344 Alzheimer's disease. Patients who have been taking 10 mg IR (or a bioequivalent generic) for at least
345 3 months prior to screening were recruited. The original trial consisted of 24 weeks of daily
346 administration of study medication, with clinic visits at screening, baseline, 3 weeks (safety only), 6
347 weeks, 12 weeks, 18 weeks, and 24 weeks or early termination. Patients received either 10 mg
348 donepezil IR in combination with the placebo corresponding to 23 mg donepezil SR, or 23 mg
349 donepezil SR in combination with the placebo corresponding to 10 mg donepezil IR. A total of 471
350 and 963 patients were enrolled from approximately 200 global sites (Asia, Oceania, Europe, India,
351 Israel, North America, South Africa, and South America). The results of the original trial yielded
352 that donepezil 23 mg/d was associated with greater benefits in cognition compared with donepezil
353 10 mg/d and led to the FDA approval of the new 23 mg dose form for treatment of AD in 2010,[27]
354 despite the debate on whether the 2.2 point of cognition improvement (on a 100 point scale) over the
355 10 mg dose form is sufficient.[22,28]
356
357 In our simulation, we followed the detailed study procedures outlined in their published article[27] to
358 formulate our simulation protocol, including the treatment regimen, population eligibility, and
359 follow-up assessments for SAEs. **Table 1** describes how the original trial design was followed in our
360 simulation.
361
362 ***Real-world patient data (RWD) from the OneFlorida network***
363 The OneFlorida data contain robust longitudinal and linked patient-level RWD of ~15 million (>60%)
364 Floridians, including data from Medicaid claims, cancer registries, vital statistics, and EHRs from its
365 clinical partners. As one of the PCORI-funded clinical research networks in the national PCORnet,

366 OneFlorida includes 12 healthcare organizations that provide care through 4,100 physicians, 914
367 clinical practices, and 22 hospitals, covering all 67 Florida counties. The OneFlorida data is a Health
368 Insurance Portability and Accountability Act (HIPAA) limited data set (i.e., dates are not shifted and
369 location data are available) that contains detailed patient characteristics and clinical variables,
370 including demographics, encounters, diagnoses, procedures, vitals, medications, and labs.[29] We
371 focused on the structured data immediately available to us formatted according to the PCORnet
372 common data model (PCORnet CDM).[30]
373
374 ***Cohort identification: the target population, the study population, and the trial not eligible***
375 ***population***
376 From the OneFlorida data, we identified three populations: the target population (TP), the study
377 population (SP), and the trial not eligible population (NEP) for the target trial following the process
378 shown in **Figure 1 Panel a**, and the relationship between these populations are displayed in **Figure 1**
379 **Panel b**. The *true* target population should be those that will benefit from the drug, thus, should be
380 broader as patients with AD in general. However, as patients who were not treated with donepezil
381 in real-world would not have any safety or effectiveness data of the drug in RWD, the effective
382 target population of interest is a constrained subset—patients who (1) had the disease of interest
383 (i.e., AD), and (2) had used the study drug (i.e., donepezil) for a specific time period according to the
384 study protocol. The 10mg donepezil is only in immediate release (IR) form while the 23mg
385 donepezil is exclusively in sustained release (SR) form, so we used the corresponding RxNorm
386 concept unique identifier (RXCUI) and the National Drug Code (NDC) to identify the two groups
387 (i.e., 10mg vs. 23mg) of patients in our data.[31,32] We then identified the study population (i.e.,
388 patients who met both the TP criteria and the trial eligibility criteria) and trial not eligible population

389 (i.e., patients who meet the TP criteria but do not meet the trial eligibility criteria) by applying the
390 eligibility criteria of the target trial to the TP. To do so, we analyzed the target trial's eligibility
391 criteria and determined the computability of each criterion. A criterion is computable when its
392 required data elements are available and clearly defined in the target patient database (i.e., the
393 OneFlorida data in our study). Then, we manually translated the computable criteria into database
394 queries against the OneFlorida database. We assumed that all patients met the non-computable
395 criteria (e.g., "*written informed consent*"), which is a limitation of our study. The full list of eligibility
396 criteria and their computability are listed in the **Supplementary Table 2**. We first decomposed each
397 criterion (e.g., "*Patients with dementia complicated by other organic disease or Alzheimer's disease with*
398 *delirium*") into smaller study traits (e.g., "dementia complicated by other organic disease" and
399 "Alzheimer's disease with delirium"). We then checked whether each of the study trait is
400 computable based on the OneFlorida data as shown in **Supplementary Table 2**. We then used the
401 computable study traits to determine patients' eligibility. Many of the incomputable study traits are
402 not clinically relevant for our studies (e.g., "*No caregiver available to meet the inclusion criteria for*
403 *caregivers.*"). Nevertheless, how computability of these study traits affect the trial simulation results
404 – a limitation of our current study – warrant further investigations in future studies.
405
406 ***Definition and identification of serious adverse events (SAE) from EHRs***
407 The target trial used Severe Impairment Battery (SIB) and the Clinician's Interview-Based
408 Impression of Change Plus Caregiver Input scale (CIBIC+; global function rating) to assess the
409 efficacy of donepezil in AD patients. Because these effectiveness data are not readily available in the
410 structured EHR data, we focused on assessing drug safety in terms of the occurrences of SAEs. To
411 define an SAE, we followed the FDA[33] definition of SAEs and the Common Terminology Criteria for

412 Adverse Events (CTCAE) version 5 – a descriptive terminology for Adverse Event (AE) reporting.
413 In CTCAE, an AE is any "*unfavorable and unintended sign, symptom, or disease temporally associated with*
414 *the use of a medical treatment or procedure that may or may not be considered related to the medical treatment*
415 *or procedure,*" and the AEs are organized based on the System Organ Class (SOC) defined in Medical
416 Dictionary for Regulatory Activities (MedDRA[34]). CTCAE also provides a grading scale for each AE
417 into Grade 1 (mild), Grade 2 (moderate), Grade 3 (severe or medically significant but not
418 immediately life-threatening), Grade 4 (life-threatening consequences), and Grade 5 (death).
419
420 We mapped each reported SAE in the trial results section of the target trial NCT00478205 on
421 ClinicalTrails.gov at https://www.clinicaltrials.gov/ct2/show/results/NCT00478205 to the CTCAE
422 term and identified the severity based on the CTCAE grading scale. We considered an AE as SAE if
423 it meets the criteria for Grade 3/4 (results in hospitalization), and Grade 5 (death). As shown in
424 **Figure 2**, to count as a SAE related to donepezil, the SAE event has to occur within 24 weeks after
425 the first donepezil prescription (which is the same follow-up period as the original trial). Note that
426 we excluded chronic conditions that happened before the study, for example, different types of
427 cancer.
428
429 ***Trial simulation***
430 ***Table 1*** shows our design of the simulated trial corresponding to the original target trial. Based on
431 the calculation from the original trial[27], a sample size of 400 and 800 were needed for the 10mg and
432 23mg arms, respectively. We first simulated the control arm of the standard therapy (i.e., the 10 mg
433 arm of the original trial), where we have a sufficiently large sample size from the OneFlorida data.
434 We designed our simulation based on the sample size of the arm in the original trial (N=400), and

435 tested two different sampling approaches: (1) random sampling, and (2) proportional sampling
436 controlling for race distribution.
437
438 Even though we did not find a sufficient number of patients who took 23mg donepezil in our data,
439 we still simulated both case-control arms using the same sampling strategy in the one-arm
440 simulation that yielded the closest effect sizes compared with the original trial. We explored two
441 different scenarios with different sample sizes: (1) the ratio of the number of subjects in the 23mg
442 arm to the 10mg arm was set as 1 to 1, and (2) the ratio was set as 1 to 3. Because of the limited
443 number of individuals who took the 23mg form, we can only increase the number of subjects in the
444 10mg arm in the second sample size scenario. We used propensity score matching (PSM) to
445 simulate randomization. The variables used for PSM included age, gender, race, and Charlson
446 comorbidity index (CCI, i.e., as a proxy for baseline overall health of the patient) prior to baseline.
447 Specifically, we fitted a logistic regression model using different treatment (i.e., case vs. control) as
448 the outcome variable and age, gender, race, and CCI as covariates to generate the logistic
449 probabilities of propensity scores individuals in the two comparison groups and then used the
450 nearest neighbor method to carry out the mapping process. The two arms were matched with the
451 propensity scores with a 1:1 or 1:3 ratio.
452
453 Specifically, we first used proportional sampling to extract a sample of patients for the 23mg study
454 population using the same race distribution as in the original trial, and then identified a matched
455 sample for the 10mg study population using PSM. We then calculated the SAEs in the 10mg vs.
456 23mg arms as the safety outcomes. The simulation process was performed 1,000 times with
457 bootstrap sampling with replacement, and the mean value and 95% confidence interval (CI) of each

458 bootstrap sample were calculated to generate the overall estimates. We focused on comparing the
459 average number of SAE per patient, the overall SAE rates (i.e., how many patients had SAEs), and
460 stratified the analysis by major SAE categories according to the CTCAE guideline. The effects of
461 PSM were evaluated by examining the distributions of propensity scores using jitter plot
462 (**Supplementary Table 1** and **Supplementary Figure 1**).
463
464
465
466 **Data availability statement**
467 OneFlorida data can be requested at https://onefloridaconsortium.org/front-door/; Since OneFlorida
468 data is a HIPAA limited data set, a data use agreement needs to be established with the OneFlorida
469 network.
470
471 **Code availability statement**
472 The data processing and analysis were conducted using R version 4.0.2. The code used in this work
473 is available upon reasonable request for academic purposes.
474
475 **Acknowledgements**
476 This work was supported in part by NIH grants R21AG068717, R21AG061431, and UL1TR001427.
477 The content is solely the responsibility of the authors and does not necessarily represent the official
478 views of the NIH.
479
480 **Author contributions**

481 ZC, HZ, and JB designed the initial concepts; ZC and HZ carried out the analysis; ZC, HZ, and JB
482 wrote the initial draft of the manuscript. TM, YG, MP, ZH, WH, FW, and ES provided critical
483 feedback and edited the manuscript.
484
485 **Competing interests**
486 The authors declared there is no competing interests.
487
488 **References**

489 1. Aronson, J. K. *What is a clinical trial? British Journal of Clinical Pharmacology* vol. 58 (John Wiley &
490 Sons, Ltd, 2004).
491 2. Sertkaya, A., Wong, H. H., Jessup, A. & Beleche, T. Key cost drivers of pharmaceutical clinical
492 trials in the United States. *Clin. Trials* (2016) doi:10.1177/1740774515625964.
493 3. Scott, T. J., O'Connor, A. C., Link, A. N. & Beaulieu, T. J. Economic analysis of opportunities to
494 accelerate Alzheimer's disease research and development. *Ann. N. Y. Acad. Sci.* **1313**, 17–34
495 (2014).
496 4. Holford, N., Ma, S. C. & Ploeger, B. A. Clinical Trial Simulation: A Review. *Clin. Pharmacol. Ther.*
497 **88**, 166–182 (2010).
498 5. Gal, J. *et al.* Optimizing drug development in oncology by clinical trial simulation: Why and
499 how? *Brief. Bioinform.* **19**, 1203–1217 (2018).
500 6. Romero, K. *et al.* The Future Is Now: Model-Based Clinical Trial Design for Alzheimer's Disease.
501 *Clin. Pharmacol. Ther.* **97**, 210–214 (2015).
502 7. Hernán, M. A. & Robins, J. M. Using Big Data to Emulate a Target Trial When a Randomized
503 Trial Is Not Available. *Am. J. Epidemiol.* (2016) doi:10.1093/aje/kwv254.

8. Danaei, G., Rodríguez, L. A. G., Cantero, O. F., Logan, R. & Hernán, M. A. Observational data for comparative effectiveness research: An emulation of randomised trials of statins and primary prevention of coronary heart disease. *Stat. Methods Med. Res.* (2013) doi:10.1177/0962280211403603.

9. Garcia-Albeniz, X., Hsu, J. & Hernan, M. A. The value of explicitly emulating a target trial when using real world evidence: an application to colorectal cancer screening. *Eur. J. Epidemiol.* **32**, 495–500 (2017).

10. Admon, A. J. *et al.* Emulating a Novel Clinical Trial Using Existing Observational Data. Predicting Results of the PreVent Study. *Ann. Am. Thorac. Soc.* **16**, 998–1007 (2019).

11. Zhang, H. *et al.* Computable Eligibility Criteria through Ontology-driven Data Access: A Case Study of Hepatitis C Virus Trials. *AMIA Annu. Symp. Proc. AMIA Symp.* **2018**, 1601–1610 (2018).

12. Brookhart, M. A., Wyss, R., Layton, J. B. & Stürmer, T. Propensity Score Methods for Confounding Control in Non-Experimental Research. *Circ. Cardiovasc. Qual. Outcomes* **6**, 604–611 (2013).

13. Dekkers, I. A. & van der Molen, A. J. Propensity Score Matching as a Substitute for Randomized Controlled Trials on Acute Kidney Injury After Contrast Media Administration: A Systematic Review. *AJR Am. J. Roentgenol.* **211**, 822–826 (2018).

14. Lin, J., Gamalo-Siebers, M. & Tiwari, R. Propensity score matched augmented controls in randomized clinical trials: A case study. *Pharm. Stat.* **17**, 629–647 (2018).

15. Hernan, M. A. & Robins, J. M. *Causal Inference: What If.* (Boca Raton: Chapman & Hall/CRC, 2020).

16. Stuart, E. A., Bradshaw, C. P. & Leaf, P. J. Assessing the generalizability of randomized trial results to target populations. *Prev. Sci. Off. J. Soc. Prev. Res.* **16**, 475–485 (2015).

527 17. He, Z. *et al.* Comparing and Contrasting A Priori and A Posteriori Generalizability Assessment
528 of Clinical Trials on Type 2 Diabetes Mellitus. *AMIA Annu. Symp. Proc. AMIA Symp.* **2017**, 849–
529 858 (2017).
530 18. Shrimanker, R., Beasley, R. & Kearns, C. Letting the right one in: evaluating the generalisability
531 of clinical trials. *Eur. Respir. J.* **52**, (2018).
532 19. Susukida, R., Crum, R. M., Ebnesajjad, C., Stuart, E. & Mojtabai, R. Generalizability of findings
533 from randomized controlled trials: Application to the National Institute of Drug Abuse Clinical
534 Trials Network. *Addiction* (2017) doi:10.1111/add.13789.
535 20. Li, Q. *et al.* Assessing the Validity of a a priori Patient-Trial Generalizability Score using Real-
536 world Data from a Large Clinical Data Research Network: A Colorectal Cancer Clinical Trial
537 Case Study. *AMIA. Annu. Symp. Proc.* **2019**, 1101–1110 (2020).
538 21. Thorlund, K., Dron, L., Park, J. J. H. & Mills, E. J. Synthetic and External Controls in Clinical
539 Trials – A Primer for Researchers. *Clin. Epidemiol.* **12**, 457–467 (2020).
540 22. Schwartz, L. M. & Woloshin, S. How the FDA forgot the evidence: the case of donepezil 23 mg.
541 *BMJ* **344**, (2012).
542 23. Deardorff, W. J., Feen, E. & Grossberg, G. T. The Use of Cholinesterase Inhibitors Across All
543 Stages of Alzheimer's Disease. *Drugs Aging* **32**, 537–547 (2015).
544 24. Cummings, J. L. *et al.* High-dose donepezil (23 mg/day) for the treatment of moderate and
545 severe Alzheimer's disease: drug profile and clinical guidelines. *CNS Neurosci. Ther.* **19**, 294–301
546 (2013).
547 25. Ghosh, S., Bian, J., Guo, Y. & Prosperi, M. Deep propensity network using a sparse autoencoder
548 for estimation of treatment effects. *J. Am. Med. Inform. Assoc.* ocaa346 (2021)
549 doi:10.1093/jamia/ocaa346.

550 26. Comparison of 23 mg Donepezil Sustained Release (SR) to 10 mg Donepezil Immediate Release
551 (IR) in Patients With Moderate to Severe Alzheimer's Disease - Full Text View -
552 ClinicalTrials.gov. https://clinicaltrials.gov/ct2/show/NCT00478205.
553 27. Farlow, M. R. *et al.* Effectiveness and tolerability of high-dose (23 mg/d) versus standard-dose
554 (10 mg/d) donepezil in moderate to severe Alzheimer's disease: A 24-week, randomized,
555 double-blind study. *Clin. Ther.* **32**, 1234–1251 (2010).
556 28. English, C. Donepezil 23 mg: Is it more advantageous compared to the original? *Ment. Health*
557 *Clin.* **1**, 272–273 (2012).
558 29. Shenkman, E. *et al.* OneFlorida Clinical Research Consortium: Linking a Clinical and
559 Translational Science Institute With a Community-Based Distributive Medical Education Model.
560 *Acad. Med. J. Assoc. Am. Med. Coll.* **93**, 451–455 (2018).
561 30. PCORnet. PCORnet Common Data Model v5.1 Specification (12 Sep 2019).
562 https://pcornet.org/data-driven-common-model/ (2019).
563 31. Gomolin, I. H., Smith, C. & Jeitner, T. M. Donepezil Dosing Strategies: Pharmacokinetic
564 Considerations. *J. Am. Med. Dir. Assoc.* **12**, 606–608 (2011).
565 32. Lee, J.-H., Jeong, S.-K., Kim, B. C., Park, K. W. & Dash, A. Donepezil across the spectrum of
566 Alzheimer's disease: dose optimization and clinical relevance. *Acta Neurol. Scand.* **131**, 259–267
567 (2015).
568 33. CFR - Code of Federal Regulations Title 21.
569 https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=314.80.
570 34. Welcome to MedDRA | MedDRA. https://www.meddra.org/.
571

572 **FIGURE LEGEND**

573 **Figure 1.** The cohort identification process for the target, study, and trial not eligible populations.

574 **Figure 2.** Follow-up window for serious adverse events (SAEs) related to treating Alzheimer's
575 disease with donepezil.

576

577

Table 1. Overall study design of the simulated trial in comparison with the original trial.

Component	Target trial (NCT00478205)	Simulated trial
Aim	Assess the safety and effectiveness of 23mg SR compared to 10 mg IR	Assess whether the simulated trial can generate similar results to the "real" trial in terms of its safety profile
Eligibility	36 eligibility criteria	25 are computable or partially computable
Treatment strategies	Randomized allocation of 23mg :10 mg ratio is 2:1	One-arm simulation of the 10 mg control arm using random sampling and proportional sampling with the same sample size as the original trial. Two-arm simulations considering different case-to-control ratios. Propensity score matching was performed on the following baseline covariates: sex, race, age, and Charlson comorbidity index (CCI).
Sampling strategies	N/A	Bootstrap with replacement was repeated 1,000 times to randomly generate the sample population, and mean value and 95% confidence interval were reported.
Follow-up	The outcomes were measured from the first dose to 24 weeks after the first dose.	
Outcome	SAE and cognition function measures	SAE [1]
Statistical analysis	Compare the average number of SAEs per patient, and the SAE rates (i.e., how many patients have SAE).	

[1] SAE: Serious Adverse Events

Table 2. Population characteristics and SAE statistics of the target trial vs. TP, SP, and NEP from OneFlorida.

	23mg Arm				10mg Arm			
	Original Trial[1]	Overall TP[2]	Overall SP[3]	Overall NEP[4]	Original Trial	Overall TP[5]	Overall SP[6]	Overall NEP[7]
# of Subject	963	90	38	52	471	2,048	782	1,266
Age Mean (SD)	73.9 (8.53)	74.3 (9.01)	73.3 (9.01)	81.6 (12.37)	73.8 (8.56)	73.4 (11.0)	74.2 (9.67)	77.1 (11.8)
Gender								
Male	356 (37.0%)	24 (26.7%)	10 (26.3%)	14 (26.9%)	177 (37.6%)	727 (35.5%)	234 (29.9%)	493 (38.9%)
Female	607 (63.0%)	66 (73.3%)	28 (73.7%)	38 (73.1%)	294 (62.4%)	1321 (64.5%)	548 (70.1%)	773 (61.1%)
Race[e]								
White	708 (73.5%)	63 (70.0%)	25 (65.8%)	38 (73.1%)	346 (73.5%)	829 (40.5%)	280 (35.8%)	549 (43.4%)
Asian/Pacific	161 (16.7%)	0 (0%)	0 (0%)	0 (0%)	87 (18.5%)	22 (1.1%)	11 (1.4%)	11 (0.9%)
Hispanic	67 (7.0%)	15 (16.7%)	4 (10.5%)	11 (21.2%)	26 (5.5%)	440 (21.5%)	192 (24.6%)	248 (19.6%)
Black	22 (2.3%)	11 (12.2%)	4 (10.5%)	7 (13.5%)	9 (1.9%)	380 (18.6%)	126 (16.1%)	254 (20.1%)
Other	5 (0.5%)	1 (1.1%)	5 (13.2%)	7 (13.5%)	3 (0.6%)	377 (18.4%)	173 (22.1%)	204 (16.1%)
CCI[f]	N/A	1.54	1.32	1.53	N/A	2.36	1.64	2.36
Mean SAE[g]	0.15	1.89	0.92	2.60	0.14	1.68	0.64	2.59
# of patients with ⩾ 1 SAE	45 (9.6%)	20 (22.2%)	4 (10.5%)	16 (30.7%)	80 (8.3%)	573 (28.0%)	121 (15.5%)	452 (35.7%)

[1]Reported in the original trial on ClinicalTrials.gov
[2]TP: Target population—patients who (1) had the disease of interest (i.e., AD), and (2) had used the study drug (i.e. donepezil) for a specific time period according to the study protocol.
[3]SP: Study population—patients in the TP who met the computable eligibility criteria of the original trial.
[4]NEP: Trial not eligible population—patients in the TP who did NOT meet the eligibility criteria of the original trial.
[5]The original trial reported Hispanic as a race, thus, we followed the same convention to make sure the results are comparable even though race and ethnicity are two different fields in OneFlorida.
[6]Charlson Comorbidity Index.
[7]Mean SAE: average number of SAEs per patient.

Table 3. One-arm simulation results for the 10mg control arm.

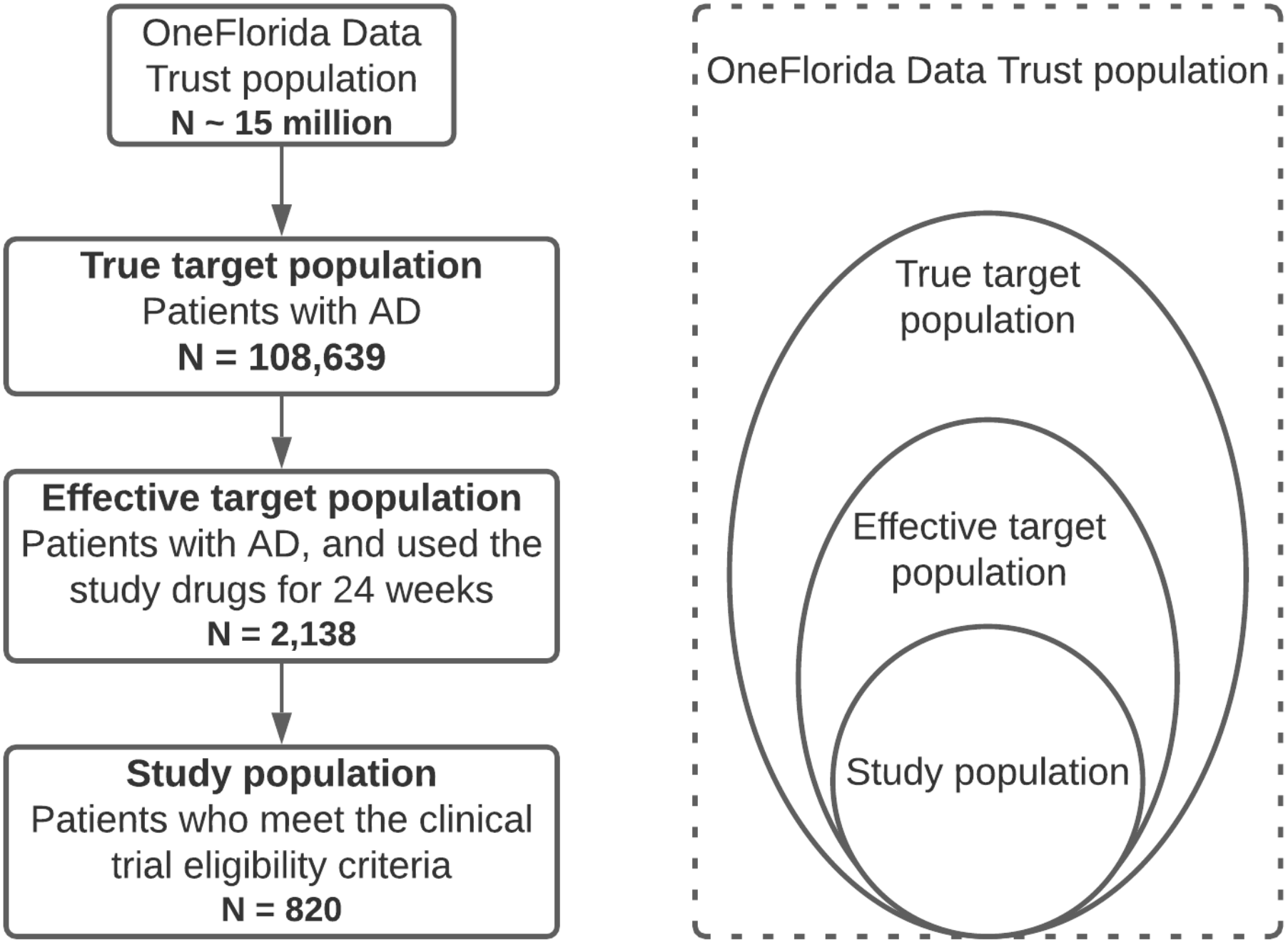
	Original control arm	Simulated control	
		Random sampling	Proportional sampling
Number of subjects	471	400	240
Mean age	73.8	78.5±0.1	78.2±0.1
Gender			
Female	62.4%	70.1%±0.1%	66.1%±0.1%
Male	37.6%	29.9%±0.1%	33.9%±0.1%
Race			
White	73.5%	28.1%±0.1%	73.3%±0.1%
Black	1.9%	16.0%±0.1%	2.1%±0.1%
Hispanic	5.5%	24.5%±0.1%	5.4%±0.1%
Asian & Other	19.1%	31.3%±0.2%	19.2%±0.1%
SAE[1] rates (patient with >= 1 SAE)	8.3%	15.5%±0.1%	8.9%±0.1%
Mean SAE (average SAE per patient)	0.14	0.64±0.01	0.48±0.01
SAE event rates by category			
Blood and lymphatic system disorders	0/471 (0.00%)	4.9%±0.1%	3.2%±0.1%
Cardiac disorders	6/471 (1.23%)	2.9%±0.1%	2.0%±0.1%
Eye disorders	1/471 (0.2%)	0	0
Gastrointestinal disorders	2/471 (0.4%)	7.0%±0.1%	5.9%±0.1%
General disorders	3/471 (0.6%)	1.7%±0.1%	1.1%±0.1%
Hepatobiliary disorders	2/471 (0.4%)	1.2%±0.1%	1.9%±0.1%
Infections and infestations	9/471 (1.9%)	5.8%±0.1%	4.7%±0.1%
Injury, poisoning and procedural complications	8/471 (1.7%)	3.2%±0.1%	3.0%±0.1%
Musculoskeletal and connective tissue disorders	2/471 (0.4%)	0.7%±0.1%	0.8%±0.1%
Neoplasms benign, malignant and unspecified	2/471 (0.4%)	0	0
Nervous system disorders	15/471 (3.3%)	6.6%±0.1%	4.8%±0.2%
Psychiatric disorders	11/471 (2.3%)	3.3%±0.1%	2.5%±0.1%
Renal and urinary disorders	2/471 (0.4%)	4.2%±0.1%	1.9%±0.1%
Reproductive system and breast disorders	0	0	0

Respiratory, thoracic and mediastinal disorders	1/471 (0.2%)	1.0%±0.1%	1.4%±0.1%
Vascular disorders	1/471 (0.2%)	0.7%±0.1%	0.3%±0.1%
[1] SAE: severe adverse events			

Table 4. Results of the two-arm simulation for proportional 1:1 and 1:3 matching.

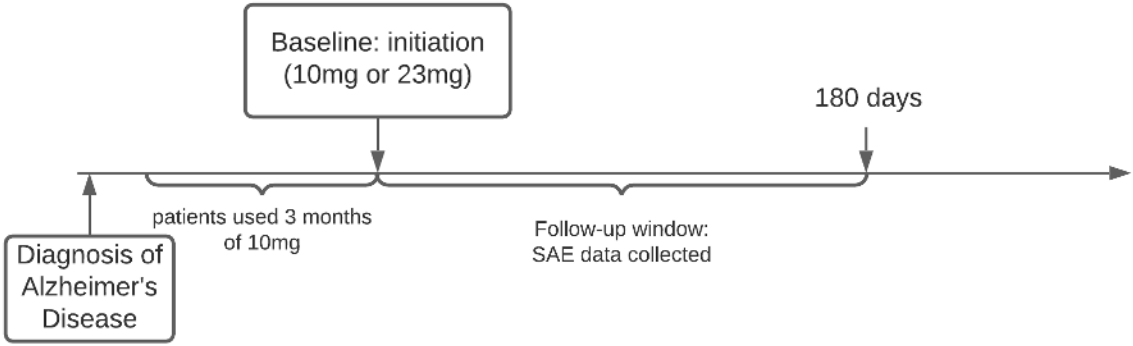
	Original trial		Simulated trial			
			Proportional 1:1 matching		Proportional 1:3 matching	
	23 mg	10mg	23mg	10mg	23mg	10mg
Number of subjects	963	471	24	24	24	72
Mean age	73.9	73.8	79.2±0.1	79.5±0.2	79.1±0.1	79.4±0.2
Gender						
Female	63.0%	62.4%	70.2%±0.6%	69.8%±0.7%	70.4%±0.6%	69.4%±0.6%
Male	37.0%	37.6%	29.8%±0.6%	30.2%±0.7%	29.6%±0.6%	30.6%±0.6%
Race						
White	73.5%	73.5%	70.8%±0.1%	70.3%±0.1%	70.8%±0.1%	70.2%±0.1%
Hispanic	7.0%	5.5%	8.3%±0.1%	7.9%±0.3%	8.3%±0.1%	8.1%±0.3%
Black	2.3%	1.9%	4.2%±0.1%	4.2%±0.3%	4.2%±0.1%	4.2%±0.2%
Asian & Other	17.2%	0.6%	16.7%±0.1%	17.6%±0.3%	16.7%±0.1%	17.6%±0.3%
Charlson Comorbidity Index	n/a		1.25±0.02	1.17±0.03	1.25±0.02	1.23±0.02
SAE[1] rates (patient with >= 1 SAE)	9.6%	8.3%	9.8%±0.4%	12.6%±0.4%	9.4%±0.2%	11.2%±0.2%
Mean SAE (average SAE per patient)	0.15	0.14	0.99±0.05	0.46±0.05	0.99±0.05	0.47±0.02
SAE Event rates by category						
Blood and lymphatic system disorders	0.2%	0	14.5%±0.9%	0.052±0.004	0.146±0.10	0.042±0.001
Cardiac disorders	1.9%	1.3%	6.7%±0.3%	0.012±0.001	0.068±0.005	0.016±0.015
Eye disorders	0	0.2%	0	0	0	0
Gastrointestinal disorders	2.4%	0.4%	0	0.082±0.004	0	0.059±0.057
General disorders	0.6%	0.6%	1.1%±0.1%	0.006±0.001	0.010±0.001	0.008±0.001
Hepatobiliary disorders	0.5%	0.4%	0	0.007±0.002	0	0.009±0.001
Infections and infestations	1.3%	1.9%	20.1%±0.1%	0.026±0.002	0.199±0.010	0.037±0.002
Injury, poisoning and procedural complications	2.5%	1.9%	0	0.014±0.003	0	0.021±0.004
Musculoskeletal and connective tissue disorders	0.3%	0.4%	0	0.019±0.001	0	0.013±0.001
Neoplasms benign, malignant and unspecified	1.5%	3.2%	0	0	0	0
Nervous system disorders	1.5%	3.2%	9.8%±0.3%	0.039±0.002	0.094±0.003	0.042±0.002
Psychiatric disorders	1.1%	2.3%	2.0%±0.5%	0.019±0.002	0.019±0.001	0.023±0.001
Renal and urinary disorders	0.4%	0.004	27.9%±1.8%	0.011±0.002	0.281±0.018	0.016±0.001

Reproductive system and breast disorders	0.1%	0	0	0	0	0
Respiratory, thoracic and mediastinal disorders	0.1%	0.2%	0	0.011±0.001	0	0.014±0.007
Vascular disorders	0.5%	0.2%	0	0.003±0.001	0	0.003±0.001
[1] SAE: severe adverse events						



Panel **a**

Panel **b**


Baseline: initiation
(10mg or 23mg)
180 days
patients used 3 months
of 10mg
Follow-up window:
SAE data collected
Diagnosis of
Alzheimer's
Disease