

## ARTICLE OPEN



# 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 <sup>1</sup>, Hansi Zhang <sup>1</sup>, Yi Guo <sup>1</sup>, Thomas J. George <sup>2</sup>, Mattia Prosperi <sup>3</sup>, William R. Hogan <sup>1</sup>, Zhe He <sup>4</sup>, Elizabeth A. Shenkman <sup>1</sup>, Fei Wang <sup>5</sup> and Jiang Bian <sup>1</sup>✉

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 (SOC) 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 serious adverse event (SAE) rates were observed in the simulated trials than the rates reported in original trial, and a higher SAE rate was observed in the 23 mg arm than in the 10 mg SOC 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 an approach may be limited, for example, by the availability of RWD that matches the target trials of interest, and further investigations are warranted.

*npj Digital Medicine* (2021)4:84; <https://doi.org/10.1038/s41746-021-00452-1>

## 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<sup>1</sup>. 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<sup>2</sup>. 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)<sup>3</sup>. Nevertheless, yet no effective drugs 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 valuable to assess the feasibility, investigate the assumptions, and optimize the study design before conducting the actual trials<sup>4,5</sup>. For example, Romero et al. conducted a CTS study to explore several design scenarios comparing the effects of donepezil with placebo<sup>6</sup>. 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<sup>7</sup>. For example, Danaei et al. emulated a hypothetical RCT using electronic health record (EHR) data from United Kingdom (UK) to estimate the effect of statins for primary prevention of coronary heart disease<sup>8</sup>. Like many other emulation studies<sup>7,9,10</sup>, this is essentially a retrospective cohort study, where the authors followed a RCT design to identify unbiased initiation of exposures and eventually to reach an unbiased estimation of the casual relationship. Combining the ideas from CTS and trial emulation, a simulation study using real-world data (RWD) to test different assumptions (e.g., different drop-out rates) and trial designs (e.g., different eligibility criteria) could provide insights on the effectiveness and safety of the treatments to be developed in a real-world setting that reflect the patient populations who will actually use the treatment.

In this study, we explored the feasibility of using RWD from the OneFlorida Clinical Research Consortium—a clinical data research network funded by the Patient-Centered Outcomes Research Institute (PCORI) contributing to the national Patient-Centered Clinical Research Network (PCORnet)—to simulate a real-world AD RCT as a use case. We considered two main scenarios: (1) a one-arm simulation: simulating a standard-of-care (SOC) arm that can serve as an external control arm; and (2) a two-arm simulation:

<sup>1</sup>Health Outcomes & Biomedical Informatics, University of Florida, Gainesville, FL, USA. <sup>2</sup>Hematology & Oncology, Department of Medicine, College of Medicine, University of Florida, Gainesville, FL, USA. <sup>3</sup>Department of Epidemiology, College of Public Health and Health Professions & College of Medicine, University of Florida, Gainesville, FL, USA.

<sup>4</sup>School of Information, Florida State University, Tallahassee, FL, USA. <sup>5</sup>Department of Healthcare Policy and Research, Cornell University, New York, NY, USA.

✉email: bianjiang@ufl.edu

**Table 1.** Population characteristics and SAE statistics of the target trial vs. TP, SP, and NEP from OneFlorida.

	23 mg Arm				10 mg Arm			
	Original trial <sup>a</sup>	Overall TP <sup>b</sup>	Overall SP <sup>c</sup>	Overall NEP <sup>d</sup>	Original trial	Overall TP <sup>e</sup>	Overall SP <sup>f</sup>	Overall NEP <sup>g</sup>
No. of subjects	963	90	38	52	471	2048	782	1266
Age (years), 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								
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 <sup>h</sup>	N/A	1.54	1.32	1.53	N/A	2.36	1.64	2.36
Mean SAE <sup>i</sup>	0.15	1.89	0.92	2.60	0.14	1.68	0.64	2.59
No. of patients with $\geq 1$ 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%)

<sup>a</sup>Reported in the original trial on ClinicalTrials.gov.

<sup>b</sup>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.

<sup>c</sup>SP: Study population—patients in the TP who met the computable eligibility criteria of the original trial.

<sup>d</sup>NEP: Trial not eligible population—patients in the TP who did NOT meet the eligibility criteria of the original trial.

<sup>e</sup>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.

<sup>f</sup>Charlson Comorbidity Index.

<sup>g</sup>Mean SAE: average number of SAEs per patient.

<sup>h</sup>CCI: Charlson Comorbidity Index.

<sup>i</sup>SAE: Serious adverse event.

simulating both intervention and control arms with proper patient matching algorithms for comparative effectiveness analysis.

## RESULTS

### Computability of eligibility criteria in the original trial (i.e., NCT00478205)

In total, there are 36 eligibility criteria in trial NCT00478205, where 17 are inclusion and 19 are exclusion criteria. However, not all criteria are computable against the OneFlorida patient database: (1) 11 are not computable, and (2) 7 are partially computable (i.e., a part of the criterion is not computable). Similar to what we have found in our prior study<sup>11</sup>, the common reasons for not computable criteria are (1) data elements needed for the criterion do not exist in the source database (e.g., “A cranial image is required, with no evidence of focal brain disease that would account for dementia.”), or (2) the criterion asked for subjective information either from the patient (e.g., “Patients who are unwilling or unable to fulfill the requirements of the study.”) or the investigator (e.g., “Clinical laboratory values must be within normal limits or, if abnormal, must be judged not clinically significant by the investigator.”). When a criterion is not computable, we consider all candidate to that criterion (e.g., they are all willing and able to “fulfill the requirements of the study”).

### Characteristics of the target, study, and trial not eligible populations from OneFlorida

Overall, a total of 90 and 2048 patients were identified as the effective target populations in OneFlorida for the 23 and 10 mg arms, respectively. Among them, 38 and 782 met the eligibility criteria of the original target RCT for the two arms, respectively. Table 1 shows the demographic characteristics and serious

adverse event (SAE) statistics of the original trial population as well as the effective target population (TP), study population (SP), and trial not eligible population (NEP) from OneFlorida.

For demographic characteristics, relative to the target RCT population, we observed a large difference in race in our OneFlorida population (all  $p$ -values of race group comparisons were smaller than 0.05). OneFlorida had more Hispanics (10.5–24.6% vs. 5.5–7%) and Blacks (10.5–20.1% vs. 1.9–2.3%), but less Whites (35.8–73.1% vs. 73.5–73.5%) or Asian/Pacific islanders (0–1.4% vs. 16.7–18.5%). The age distributions were similar across all populations. For clinical variables, we calculated the Charlson Comorbidity Index (CCI) of the various populations from OneFlorida. Smaller CCIs were observed in the SP compared with the TP for both arms ( $p < 0.05$ ), and a smaller CCI was observed in the 23 mg arm compared with the 10 mg arm ( $p < 0.05$ ). Our primary outcomes of interest in this analysis were SAEs. Thus, we calculated the mean SAE (i.e., the average number of SAEs per patient) and the number of patients who had more than 1 SAE during the study period. For both 23 and 10 mg arms, the mean SAE and the number of patients with SAEs were the largest in the TP, followed by the SP, and then the original trial. Consistent with the original trial, populations derived from the OneFlorida data in the 23 mg arm have higher number of mean SAE and more patients with SAE compared with the 10 mg arm.

### Standard-of-care control arm (i.e., one-arm) simulation

We first simulated the control arm of the original trial (i.e., the 10 mg SOC arm). Table 2 displays the demographics and SAE outcomes in the simulated control arms. Here, we reported the mean value and 95% confidence interval (CI) of all 1000 bootstrap samples. Two different sampling approaches were used: (1) random sampling, and (2) proportional sampling accounting for race distribution. When using the random sampling approach,

**Table 2.** One-arm simulation results for the 10 mg control arm.

	Original control arm	Simulated control	
		Random sampling	Proportional sampling
No. of subjects	471	400	240
Mean age (years)	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 <sup>a</sup> 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%

<sup>a</sup>SAE: serious adverse event.

compared with the control arm in the original trial, higher mean SAE and SAE rates were observed, in addition to discrepancies in demographic variables. When using proportional sampling, the results were closer and more consistent with the original trial. Notably, the SAE rates in the simulated control were similar to the SAE rates from the original control (8.9% vs. 8.3%), and a z-score test for population proportion had a *p*-value of 0.75, suggesting there were no significant differences between the two SAE rates. In addition to SAE rates and mean SAE, we also explored the SAE event rates in the simulated control arms stratified by the SAE category reported in the original trial. Compared with the control arm in the original trial, the simulated control arms have larger SAE rates in most categories.

### Two-arm trial simulation

Because the proportional sampling had better performance in the one-arm simulation, we used this sampling strategy in the two-arm simulation to match the race distribution, and tested two scenarios of different matching ratios (i.e., proportional 1:1 matching and proportional 1:3 matching). However, since there is no Asian/Pacific in our SP who used 23 mg donepezil, all Asians in the original trial and the simulated control arm were grouped into "other", and the sample size of the 23 mg arm was set at 30 (because of the limited number of 23 mg patients in the

OneFlorida data). Table 3 shows our two-arm simulation results, where we show the average and 95% CI of all variables for the simulation arms across all 1000 bootstrap samples. In both matching scenarios, the mean SAE and SAE rates were higher in the 23 mg arm than in the 10 mg arm, which is consistent with the original trial. However, the variance for both SAE outcomes for the 10 mg arm are higher in the 1:1 matching scenario than in the 1:3 matching scenario, as the sample size for the 10 mg arm in the 1:3 matching scenario is much bigger. Because of the sample size difference, estimates from the 1:3 matching scenario should be more reliable. Consistent with the original trial, patients in the 23 mg arm have higher event rates in most of the SAE categories compared to the patients in the 10 mg arm. Note that we observed no SAE events in several categories in our simulation, especially in the 23 mg arm, due to the limited sample size.

Finally, we conducted an additional experiment to simulate patients who withdrew from the trial. In the original trial, among the 963 and 471 patients from each arm, 296 (30%) and 87 (18%) patients, respectively, discontinued the study for various reasons. Among the dropouts, 182 and 39 patients discontinued due to AEs. We simulated the dropouts by (1) randomly removing 18% in the 10 mg group and 30% in the 23 mg in our simulations; and (2) removing the patient after his or her first AE using the same proportion as the original trial (i.e., due to small sample size, we did not simulate this scenario for the two-arm simulation). The

**Table 3.** 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	10 mg	23 mg	10 mg	23 mg	10 mg
No. of subjects	963	471	24	24	24	72
Mean age (years)	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 <sup>a</sup> 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

<sup>a</sup>SAE: serious adverse event.

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

## DISCUSSION

In this work, we simulated an AD RCT utilizing RWD from the OneFlorida network, a large clinical data research network, considering three different simulation scenarios. In the one-arm simulation scenario, we attempted to simulate an external control

arm for the original trial. We demonstrated that we could achieve similar estimate of SAE rates as the original trial when proportional sampling accounting for race distribution was used; and the statistics of the simulated control arm were stable across all bootstrap simulation runs, which suggests that using RWD we can robustly simulate the “*standard of care*” control arm. In the two-arm comparative effectiveness simulations, we used propensity score matching (PSM) on baseline characteristics to simulate the randomization process. It has been demonstrated that PSM could reduce bias in the estimate of the treatment response<sup>12–15</sup>, and in our study, we successfully simulated two groups of patients that have similar age, sex, race, and CCI distributions using PSM. However, the SAE outcomes in the simulated trial were still different from the original trial for various reasons: (1) The original trial was conducted in research settings, while RWD data reflect patients the real-world clinical settings. The total time at risk for SAEs in our simulated cohort may be longer than the original trial, because in clinical trials, patients may withdraw from the study once experiencing an SAE, while patients in real-world setting may not. This is demonstrated by our simulation of dropouts, which achieved smaller number of SAEs and closer results to the original trial; (2) Sample size issue, where 23 mg donepezil has not been

**Table 4.** 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 23 mg 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 23 mg: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 1000 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 <sup>a</sup>
Statistical analysis	Compare the average number of SAEs per patient, and the SAE rates (i.e., how many patients have SAE).	

<sup>a</sup>SAE: serious adverse event.

the SOC for AD in the real world, leading to considerably fewer patients in the 23 mg arm. For the two-arm simulation, we conducted a post hoc power analysis with the SAE rates and mean SAE. Assuming a significant level at 0.05, a 65% power were achieved; and (3) Although PSM derived two simulation arms that are comparable, we were unable to compare it directly to the SP in the original trials as the data for calculating propensity scores were not available from the original trial. For example, the switching to the 23 mg treatment after receiving at least 3 months of the 10 mg does not occur at random in a real-world setting, but based on clinical guidelines; and indeed, we found that patients in the 23 mg arm have a longer history of diagnosis (i.e., mean of days between first diagnosis and first prescription in 23 mg arm is 398 days vs. 128 days in the 10 mg arm). Therefore, in our two-arm simulation, there may be residue selection bias causing a difference between the two populations. Nevertheless, this is an issue of using observational data in general. Even though we can simulate randomization, e.g., through PSM, trial simulations cannot replace RCTs. In addition, there are still gaps, especially data gaps in RWD, which also contributed to the differences between our simulation results and the original trial results. Future studies are warranted to identify strategies to fill these gaps.

While simulating the original AD trial that followed the study protocol in Table 4, 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 the TP (i.e., higher trial generalizability).

Our findings are consistent with previous literature on clinical trial generalizability<sup>16–19</sup>. 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 TP (i.e., patients who took donepezil for AD), which is the population who actually used the medication in real-world settings, and also (2) higher in the SP, patients who used donepezil for AD and also met the original trial’s eligibility criteria. Some of the differences may be due to the incomputable eligibility criteria (e.g., general physical health deterioration) that we cannot account for, but it is also possible that the original trial samples did not adequately reflect the TP and thus there might be treatment effect heterogeneity across patient subgroups, not captured by the original trial. In the two-arm simulations, large variances were observed, especially when the matched sample size was small. This may also indicate the heterogeneous treatment effects of donepezil when applied to different patient subgroups in real-world settings.

Our study demonstrated the feasibility of trial simulation using RWD, especially when simulating external SOC control arms. Our one-arm simulation provided stable and robust estimates and sufficient sample sizes to compare with the original trial’s control arm. The SAE rates observed in the simulated control arm with proportional sampling were very close to what was reported in the original trial. The mean SAE per patient and SAE event rates, however, were larger in the simulated control arms, which suggested that, in a real-world setting, the patients who experienced SAEs tend to have more occurrences of SAEs. On the other hand, the two-arm simulation, although it provided insights, was not entirely successful. Although the randomization process was effectively simulated by using PSM, the outcome measures were very different from the original trial. The reasons for the differences could be multi-fold (e.g., research setting vs. real-world clinical setting, difference in sample size, overly restrictive eligibility criteria that limits the generalizability of the original trial), but cannot be explored due to limited data reported by the original trial (i.e., no patient-level data are available). When we simulated dropouts using information from the original trial, we observed similar SAE rates and mean SAE compared to the original trial in both random dropout and AE-based dropout strategies, especially in the control arms. This suggests that the additional simulation scenarios have led to results more comparable to the cohorts in the original trial. However, due to the small sample size in the 23 mg arm, we were only able to simulate random dropout for the two-arm simulation, and the SAE

measurements did not change much comparing to the simulations without dropout. Future studies with sufficient sample sizes could conduct more sophisticated analysis based on treatment delay and adherence.

Compared with trial emulation, which focus on making the target trial explicitly characterized with a defined protocol, our approach takes the advantage of having observational RWD, where different trial protocols (i.e., simulation scenarios) with different study designs can be readily tested. For example, in our current study, there are several potential simulation points that can be further tuned. First, the sample size of each arm can be adjusted. In the one-arm simulation (i.e., the control arm of 10 mg donepezil), we choose the same sample size as the original trial, but it can be adjusted to increase power. Second, the eligibility criteria of selecting SP could also be adjusted to test different hypotheses. For example, we can adjust the eligibility criteria in the trial simulation process to assess how the original trial results may be generalized into real-world TP, and provide insights on how to balance internal validity while retaining good external validity<sup>17,20</sup>. Third, different scenarios of dropout may be simulated. The dropout rate and timing can be varied, so that it can be used to simulate different patient population. Further, we can also explore whether other dropout reasons such as lack of recovery and lack of access to care can be simulated based on RWD. Many other potential simulation scenarios can be tested, such as varying the 3 month lead time for switching from 10 to 23 mg. In this current work, as our main goal was to establish the feasibility of such a simulation approach, we only conducted limited number of major simulation scenarios (e.g., we used two different sampling scenarios using different intervention arm vs. control arm ratios). In future work, informed by literature, we shall systematically simulate the different trial design scenarios, which can (1) provide critical information on the comparative effectiveness of the interventions in real-world settings, and also (2) better inform the study designs of future clinical trials. Last but not the least, the one-arm simulation is as important as the two-arm simulation, even though it does not provide comparative effectiveness results of the intervention. In addition to informing future design of control arms, one-arm simulation allows us to utilize readily available RWD of patients taking the SOC to determine SOC's treatment effectiveness and safety profile, and consider different study protocols and scenarios. The demonstrated feasibility of one-arm simulation is a building block towards the potential of using RWD to generate synthetic and external controls for clinical trials, leading to significant cost savings<sup>21</sup>. Nevertheless, other issues with RWD such as its data quality (e.g., missing key measures of endpoints) and the inherent biases that exist in observational data warrant further investigations.

There are some other limitations in this study. First, we only looked at one original trial for one medication (i.e., donepezil). Simulations on different drugs and diseases may have different results. Second, the population who took the 23 mg form in our data is very small (even though the overall OneFlorida population is large with more than 15 million patients), where we only identified 38 patients who took the 23 mg donepezil and met the eligibility criteria of the original trial. The 23 mg donepezil form was approved by the FDA in 2010, so it is still a relatively new drug on the market, and following its approval, the clinical utility of the 23 mg form was called into question because of its limited effectiveness and higher rates of AEs<sup>22–24</sup>. The current practice of using the donepezil 23 mg form is reserved for AD patients who have been on stable donepezil 10 mg form for at least 3–6 months with no significant improvement<sup>25,26</sup>, which limited its use in real-world clinical practice. In addition, patients who switched to the 23 mg treatment may have different characteristics that we did not account for in this analysis. Third, we found that some of the SAEs (e.g., abnormal behavior, presyncope) reported in the trial's

results cannot be mapped to any AE terms in CTCAE, and the definitions of AEs in the original trial were unavailable, which increased the difficulty of accurately accounting for all SAEs. Further, even though trials' SAEs reported in ClinicalTrials.gov largely follow the Medical Dictionary for Regulatory Activities Terminology (MedDRA), not all reported SAEs were correctly defined in the trial results. For example, we found "Back pain" and "Fall" were defined as SAEs in the original AD trial we modeled. However, in CTCAE, there is no corresponding category 4 or 5 definition for them. More effort is needed to consistently model SAEs reported in clinical trials. Finally, because of data limitations, we were not able to assess the effectiveness of AD treatment (e.g., AD endpoints such as Mini-Mental State Examination and Severe Impairment Battery are not readily available in structure EHR data, but may exist in clinical narratives). Thus, we only examined safety outcomes in our current study. This may also contribute to the different results we obtained from our simulation compared with results from the original trial as the original trial was designed and powered with primary efficacy-based outcomes. Future studies that explore the use of advanced natural language processing (NLP) methods to extract these endpoint measurements from clinical notes will be important. Further, variables extracted from clinical notes with NLP could also be used to render some of the incomputable eligibility criteria computable.

In conclusion, in this study, we investigated the feasibility of using the existing patient records to simulate clinical trials using an AD trial (i.e., NCT00478205) as the use case. We examined two main simulation scenarios: (1) a one-arm simulation: simulating the SOC 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. We have also considered a number of different simulation parameters such as sampling strategies, matching approaches, and dropout scenarios. In the case study, our simulation can robustly simulate "standard of care" control arms (i.e., the 10 mg donepezil arm) in terms of safety evaluation. However, trial simulation using RWD may be limited by the availability of RWD that matches the target trials of interest and may not yield reliable and consistent results if the sample sizes of the interventions of interest (i.e., we found few patients were prescribed the 23 mg donepezil) are limited from the real-world databases. Further investigations on this topic are warranted, especially how to address the data quality issues (e.g., using NLP to extract more complete patient information) and reduce inherent biases (e.g., more advanced matching methods to tackle the problems of high dimensionality, nonlinear/nonparallel treatment assignment, and other complex confounding situations<sup>27</sup>) in observational RWD. Last but not the least, it will also be beneficial to have access to more complete information (e.g., de-identified individual-level trial participant data) of the target trials, so that more realistic simulation settings can be explored.

## METHODS

This study was a secondary data analysis using existing data, the study was approved by the University of Florida Institutional Review Board (IRB201902362).

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

Although there is no cure for AD yet, the U.S. FDA approved two classes of medications: (1) cholinesterase inhibitors and (2) memantine, to treat the symptoms of dementia. Donepezil (Aricept<sup>®</sup>), a cholinesterase inhibitor, was the most widely tested AD drug and approved for all stages of AD. The target trial NCT00478205<sup>28</sup> is a Phase III double-blind, double-dummy, parallel-group comparison of 23 mg donepezil sustained release (SR) with the 10 mg donepezil immediate release (IR) formulation (marketed as the SOC) in patients with moderate to severe AD. Patients who have been taking 10 mg IR (or a bioequivalent generic) for at least 3 months prior to

screening were recruited. The original trial consisted of 24 weeks of daily administration of study medication, with clinic visits at screening, baseline, 3 weeks (safety only), 6 weeks, 12 weeks, 18 weeks, and 24 weeks, or early termination. Patients received either 10 mg donepezil IR in combination with the placebo corresponding to 23 mg donepezil SR, or 23 mg donepezil SR in combination with the placebo corresponding to 10 mg donepezil IR. A total of 471 and 963 patients were enrolled from approximately 200 global sites (Asia, Oceania, Europe, India, Israel, North America, South Africa, and South America). The results of the original trial yielded that donepezil 23 mg/d was associated with greater benefits in cognition compared with donepezil 10 mg/d and led to the FDA approval of the new 23 mg dose form for the treatment of AD in 2010<sup>29</sup>, despite the debate on whether the 2.2 point of cognition improvement (on a 100 point scale) over the 10 mg dose form is sufficient<sup>22,30</sup>.

In our simulation, we followed the detailed study procedures outlined by Farlow et al.<sup>29</sup> to formulate our simulation protocol, including the treatment regimen, population eligibility, and follow-up assessments for SAEs. Table 4 describes how the original trial design was followed in our simulation.

### Real-world patient data (RWD) from the OneFlorida network

The OneFlorida data contain robust longitudinal and linked patient-level RWD of ~15 million (>60%) Floridians, including data from Medicaid claims, cancer registries, vital statistics, and EHRs from its clinical partners. As one of the PCORI-funded clinical research networks in the national PCORnet, OneFlorida includes 12 healthcare organizations that provide care through 4100 physicians, 914 clinical practices, and 22 hospitals, covering all 67 Florida counties. The OneFlorida data is a Health Insurance Portability and Accountability Act (HIPAA) limited data set (i.e., data are not shifted and location data are available) that contains detailed patient characteristics and clinical variables, including demographics, encounters, diagnoses, procedures, vitals, medications, and labs<sup>31</sup>. We focused on the structured data immediately available to us formatted according to the PCORnet common data model (PCORnet CDM)<sup>32</sup>.

### Cohort identification: the target population, the study population, and the trial not eligible population

From the OneFlorida data, we identified three populations: the target population (TP), the study population (SP), and the trial not eligible population (NEP) for the target trial following the process shown in Fig. 1a, and the relationship between these populations are displayed in Fig. 1b. The *true* TP should be those that will benefit from the drug, thus, should be broader as patients with AD in general. However, as patients who were not treated with donepezil in real world would not have any safety or

effectiveness data of the drug in RWD, the effective TP of interest is a constrained subset: 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. The 10 mg donepezil is only in IR form while the 23 mg donepezil is exclusively in SR form, so we used the corresponding RxNorm concept unique identifier (RXCU) and the National Drug Code (NDC) to identify the two groups (i.e., 10 mg vs. 23 mg) of patients in our data<sup>25,26</sup>. We then identified the SP (i.e., patients who met both the TP criteria and the trial eligibility criteria) and NEP (i.e., patients who meet the TP criteria but do not meet the trial eligibility criteria) by applying the eligibility criteria of the target trial to the TP. To do so, we analyzed the target trial's eligibility criteria and determined the computability of each criterion. A criterion is computable when its required data elements are available and clearly defined in the target patient database (i.e., the OneFlorida data in our study). Then, we manually translated the computable criteria into database queries against the OneFlorida database. We assumed that all patients met the non-computable criteria (e.g., "written informed consent"), which is a limitation of our study. The full list of eligibility criteria and their computability are listed in Supplementary Table 2. We first decomposed each criterion (e.g., "Patients with dementia complicated by other organic disease or Alzheimer's disease with delirium") into smaller study traits (e.g., "dementia complicated by other organic disease" and "Alzheimer's disease with delirium"). We then checked whether each of the study trait is computable based on the OneFlorida data as shown in Supplementary Table 2. We then used the computable study traits to determine patients' eligibility. Many of the incomputable study traits are not clinically relevant for our studies (e.g., "No caregiver available to meet the inclusion criteria for caregivers."). Nevertheless, how computability of these study traits affects the trial simulation results—a limitation of our current study—warrant further investigations in future studies.

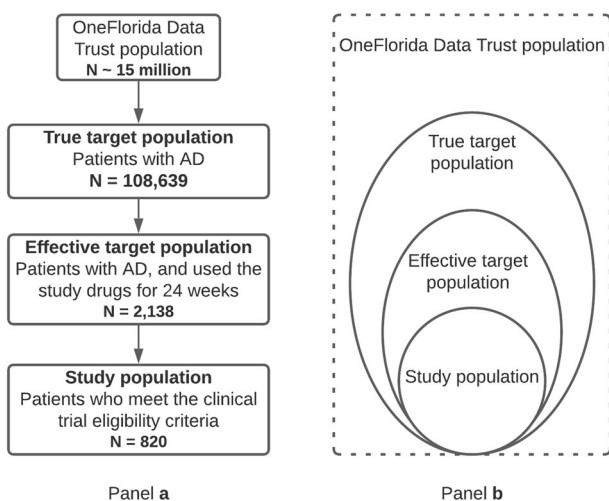
### Definition and identification of serious adverse events (SAE) from EHRs

The target trial used Severe Impairment Battery (SIB) and the Clinician's Interview-Based Impression of Change Plus Caregiver Input scale (CIBIC+; global function rating) to assess the efficacy of donepezil in AD patients. Because these effectiveness data are not readily available in the structured EHR data, we focused on assessing drug safety in terms of the occurrences of SAEs. To define an SAE, we followed the FDA<sup>33</sup> definition of SAEs and the Common Terminology Criteria for Adverse Events (CTCAE) version 5—a descriptive terminology for Adverse Event (AE) reporting. In CTCAE, an AE is any "unfavorable and unintended sign, symptom, or disease temporally associated with the use of a medical treatment or procedure that may or may not be considered related to the medical treatment or procedure," and the AEs are organized based on the System Organ Class (SOC) defined in Medical Dictionary for Regulatory Activities (MedDRA<sup>34</sup>). CTCAE also provides a grading scale for each AE into Grade 1 (mild), Grade 2 (moderate), Grade 3 (severe or medically significant but not immediately life-threatening), Grade 4 (life-threatening consequences), and Grade 5 (death).

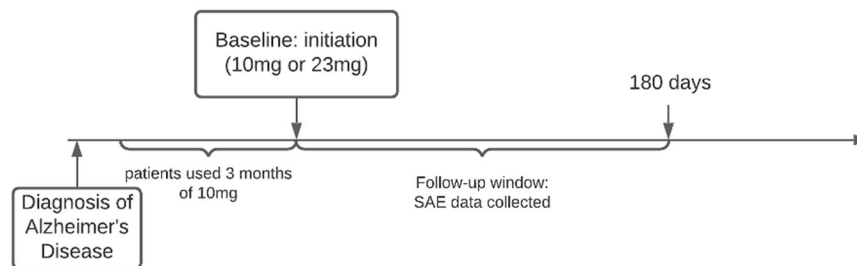
We mapped each reported SAE in the trial results section of the target trial NCT00478205 on ClinicalTrials.gov at <https://www.clinicaltrials.gov/ct2/show/results/NCT00478205> to the CTCAE term and identified the severity based on the CTCAE grading scale. We considered an AE as SAE if it meets the criteria for Grade 3/4 (results in hospitalization), and Grade 5 (death). As shown in Fig. 2, to count as an SAE related to donepezil, the SAE event has to occur within 24 weeks after the first donepezil prescription (which is the same follow-up period as the original trial). Note that we excluded chronic conditions that happened before the study, for example, different types of cancer.

### Trial simulation

Table 4 shows our design of the simulated trial corresponding to the original target trial. Based on the calculation from the original trial<sup>27</sup>, a sample size of 400 and 800 were needed for the 10 and 23 mg arms, respectively. We first simulated the control arm of the standard therapy (i.e., the 10 mg arm of the original trial), where we have a sufficiently large sample size from the OneFlorida data. We designed our simulation based on the sample size of the arm in the original trial ( $N = 400$ ), and tested two different sampling approaches: (1) random sampling, and (2) proportional sampling controlling for race distribution.



**Fig. 1** Workflow implemented to select the target, study, and trial not eligible populations are selected from the OneFlorida Data Trust. Panel a shows the definition of different population, and number of patients identified. Panel b shows the relationships between the target, study, trial not eligible, and entire OneFlorida Data Trust populations. The cohort identification process for the target, study, and trial not eligible populations.



**Fig. 2** Timeline showing how baseline and follow-up window are defined. Patients who used 3 months of 10 mg donepezil were included in the study, their first use of 10 mg or 23 mg donepezil after the 3-month were used as baseline. The follow-up window is from baseline to 180 days after drug initiation. Follow-up window for serious adverse events (SAEs) related to treating Alzheimer's disease with donepezil.

Even though we did not find a sufficient number of patients who took 23 mg donepezil in our data, we still simulated both case–control arms using the same sampling strategy in the one-arm simulation that yielded the closest effect sizes compared with the original trial. We explored two different scenarios with different sample sizes: (1) the ratio of the number of subjects in the 23 mg arm to the 10 mg arm was set as 1:1; and (2) the ratio was set as 1:3. Because of the limited number of individuals who took the 23 mg form, we can only increase the number of subjects in the 10 mg arm in the second sample size scenario. We used PSM to simulate randomization. The variables used for PSM included age, gender, race, and CCI (i.e., as a proxy for baseline overall health of the patient) prior to baseline. Specifically, we fitted a logistic regression model using different treatments (i.e., case vs. control) as the outcome variable and age, gender, race, and CCI as covariates to generate the logistic probabilities of propensity scores of individuals in the two comparison groups and then used the nearest neighbor method to carry out the mapping process. The two arms were matched with the propensity scores with a 1:1 or 1:3 ratio.

Specifically, we first used proportional sampling to extract a sample of patients for the 23 mg SP using the same race distribution as in the original trial, and then identified a matched sample for the 10 mg SP using PSM. We then calculated the SAEs in the 10 mg vs. 23 mg arms as the safety outcomes. The simulation process was performed 1000 times with bootstrap sampling with replacement, and the mean value and 95% CI of each bootstrap sample were calculated to generate the overall estimates. We focused on comparing the average number of SAE per patient, the overall SAE rates (i.e., how many patients had SAEs), and stratified the analysis by major SAE categories according to the CTCAE guideline. The effects of PSM were evaluated by examining the distributions of propensity scores using jitter plot (Supplementary Table 1 and Supplementary Fig. 1).

## Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## DATA AVAILABILITY

OneFlorida data can be requested at <https://onefloridaconsortium.org/front-door/>; since OneFlorida data is a HIPAA limited data set, a data use agreement needs to be established with the OneFlorida network.

## CODE AVAILABILITY

The data processing and analysis were conducted using R version 4.0.2. The code used in this work is available upon reasonable request for academic purposes.

Received: 28 October 2020; Accepted: 6 April 2021;

Published online: 14 May 2021

## REFERENCES

- Aronson, J. K. What is a clinical trial? *Br. J. Clin. Pharmacol.* **58**, 1–3 (2004).
- Sertkaya, A., Wong, H. H., Jessup, A. & Beleche, T. Key cost drivers of pharmaceutical clinical trials in the United States. *Clin. Trials* **13**, 117–126 (2016).

- Scott, T. J., O'Connor, A. C., Link, A. N. & Beaulieu, T. J. Economic analysis of opportunities to accelerate Alzheimer's disease research and development. *Ann. NY Acad. Sci.* **1313**, 17–34 (2014).
- Holford, N., Ma, S. C. & Ploeger, B. A. Clinical trial simulation: a review. *Clin. Pharmacol. Ther.* **88**, 166–182 (2010).
- Gal, J. et al. Optimizing drug development in oncology by clinical trial simulation: why and how? *Brief. Bioinform.* **19**, 1203–1217 (2018).
- Romero, K. et al. The future is now: model-based clinical trial design for Alzheimer's disease. *Clin. Pharmacol. Ther.* **97**, 210–214 (2015).
- Hernán, M. A. & Robins, J. M. Using big data to emulate a target trial when a randomized trial is not available. *Am. J. Epidemiol.* **183**, 758–764 (2016).
- 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.* **22**, 70–96 (2013).
- García-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).
- 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).
- Zhang, H. et al. Computable eligibility criteria through ontology-driven data access: a case study of hepatitis C virus trials. *AMIA Annu. Symp. Proc.* **2018**, 1601–1610 (2018).
- Brookhart, M. A., Wyss, R., Layton, J. B. & Stürmer, T. Propensity score methods for confounding control in non-experimental. *Res. Circ. Cardiovasc. Qual. Outcomes* **6**, 604–611 (2013).
- 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).
- 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).
- Hernan, M. A. & Robins, J. M. *Causal Inference: What If* (Chapman & Hall/CRC, Boca Raton, 2020).
- Stuart, E. A., Bradshaw, C. P. & Leaf, P. J. Assessing the generalizability of randomized trial results to target populations. *Prev. Sci.* **16**, 475–485 (2015).
- He, Z. et al. Comparing and contrasting a priori and a posteriori generalizability assessment of clinical trials on type 2 diabetes mellitus. *AMIA Annu. Symp. Proc.* **2017**, 849–858 (2017).
- Shrimanker, R., Beasley, R. & Kearns, C. Letting the right one in: evaluating the generalisability of clinical trials. *Eur. Respir. J.* **52**, 1802218 (2018).
- Susukida, R., Crum, R.M., Ebnasajjad, C., Stuart, E. & Mojtabei, R. Generalizability of findings from randomized controlled trials: application to the National Institute of Drug Abuse Clinical Trials Network. *Addiction* **112**, 1210–1219 (2017).
- Li, Q. et al. Assessing the validity of a priori patient-trial generalizability score using real-world data from a large clinical data research network: a colorectal cancer clinical trial case study. *AMIA Annu. Symp. Proc.* **2019**, 1101–1110 (2020).
- Thorlund, K., Dron, L., Park, J. J. H. & Mills, E. J. Synthetic and external controls in clinical trials – a primer for researchers. *Clin. Epidemiol.* **12**, 457–467 (2020).
- Schwartz, L. M. & Woloshin, S. How the FDA forgot the evidence: the case of donepezil 23 mg. *BMJ* **344**, e1086 (2012).
- Deardorff, W. J., Feen, E. & Grossberg, G. T. The use of cholinesterase inhibitors across all stages of Alzheimer's disease. *Drugs Aging* **32**, 537–547 (2015).
- Cummings, J. L. et al. High-dose donepezil (23 mg/day) for the treatment of moderate and severe Alzheimer's disease: drug profile and clinical guidelines. *CNS Neurosci. Ther.* **19**, 294–301 (2013).
- Gomolin, I. H., Smith, C. & Jeitner, T. M. Donepezil dosing strategies: pharmacokinetic considerations. *J. Am. Med. Dir. Assoc.* **12**, 606–608 (2011).



26. Lee, J.-H., Jeong, S.-K., Kim, B. C., Park, K. W. & Dash, A. Donepezil across the spectrum of Alzheimer's disease: dose optimization and clinical relevance. *Acta Neurol. Scand.* **131**, 259–267 (2015).
27. Ghosh, S., Bian, J., Guo, Y. & Prosperi, M. Deep propensity network using a sparse autoencoder for estimation of treatment effects. *Journal of the American Medical Informatics Association*. <https://doi.org/10.1093/jamia/ocaa346> (2021).
28. NIH. Comparison of 23 mg donepezil sustained release (SR) to 10 mg donepezil immediate release (IR) in patients with moderate to severe Alzheimer's disease. Full text view, ClinicalTrials.gov. <https://clinicaltrials.gov/ct2/show/NCT00478205> (2014). Accessed 18 Sept 2020
29. Farlow, M. R. et al. Effectiveness and tolerability of high-dose (23 mg/d) versus standard-dose (10 mg/d) donepezil in moderate to severe Alzheimer's disease: a 24-week, randomized, double-blind study. *Clin. Ther.* **32**, 1234–1251 (2010).
30. English, C. Donepezil 23 mg: is it more advantageous compared to the original?. *Ment. Health Clin.* **1**, 272–273 (2012).
31. Shenkman, E. et al. OneFlorida Clinical Research Consortium: linking a clinical and translational science institute with a community-based distributive medical education model. *Acad. Med.* **93**, 451–455 (2018).
32. PCORnet. *PCORnet Common Data Model v5.1 Specification* (12 September 2019). <https://pcornt.org/data-driven-common-model/> (2019)
33. FDA. *CFR - Code of Federal Regulations Title 21*. <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcr/CFRSearch.cfm?fr=314.80>. Accessed 18 Sept 2020.
34. MedDRA. *Welcome to MedDRA*. <https://www.meddra.org/>. Accessed 18 Sept 2020.

## ACKNOWLEDGEMENTS

This work was supported in part by NIH grants R21AG068717, R21AG061431, and UL1TR001427. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

## AUTHOR CONTRIBUTIONS

Z.C., H.Z. and J.B. designed the initial concepts; Z.C. and H.Z. carried out the analysis; Z.C., H.Z. and J.B. wrote the initial draft of the manuscript. T.M., Y.G., M.P., Z.H., W.H., F.W. and E.S. provided critical feedback and edited the manuscript.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41746-021-00452-1>.

**Correspondence** and requests for materials should be addressed to J.B.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021