Contextualizing single-arm trials with real-world data: An emulated target trial comparing therapies for neovascular age-related macular degeneration

Darren S. Thomas[1] | Aaron Y. Lee[2] | Philipp L. Müller[3,4,5] | Roy Schwartz[4,5,6] | Abraham Olvera-Barrios[4,5] | Alasdair N. Warwick[4,7] | Praven J. Patel[6] | Tjebo F.C. Heeren[4,5] | Catherine Egan[4,5,6] | Paul Taylor[1] | Adnan Tufail[4,5,6] | on behalf of the UK AMD EMR Users Group[7]

[1]Institute of Health Informatics, University College London (UCL), London, UK

[2]Department of Ophthalmology, University of Washington, Seattle, Washington, USA

[3]Department of Ophthalmology, University of Bonn, Bonn, Germany

[4]Moorfields Eye Hospital NHS Foundation Trust, London, UK

[5]University College London Institute of Ophthalmology, London, UK

[6]National Institute of Health Research Biomedical Research Centre at Moorfields Eye Hospital and UCL Institute Ophthalmology, London, UK

[7]Institute of Cardiovascular Science, University College London (UCL), London, UK

Correspondence
Adnan Tufail, Moorfields Eye Hospital NHS Trust, 162 City Road, London, EC1V 2PD, UK.
Email: adnan.tufail@nhs.net

Abstract

One-in-four ophthalmology trials are single-armed, which poses challenges to their interpretation. We demonstrate how real-world cohorts used as external/synthetic control arms can contextualize such trials. We herein emulated a target trial on the intention-to-treat efficacy of off-label bevacizumab (q6w) pro re nata relative to fixed-interval aflibercept (q8w) for improving week 54 visual acuity of eyes affected by neovascular age-related macular degeneration. The bevacizumab arm ($n = 65$) was taken from the ABC randomized controlled trial. A total of 4,471 aflibercept-treated eyes aligning with the ABC trial eligibility were identified from electronic health records and synthetic control arms were created by emulating randomization conditional on age, sex, and baseline visual read via exact matching and propensity score methods. We undertook an inferiority analysis on mean difference at 54 weeks; outcomes regression on achieving a change in visual acuity of greater than or equal to 15, greater than or equal to 10, and less than or equal to −15 Early Treatment Diabetic Retinopathy (ETDRS) letters at week 54; and a time-to-event analysis on achieving a change in visual acuity of greater than or equal to 15, greater than or equal to 10, and less than or equal to −15 ETDRS letters by week 54. The findings suggest off-label bevacizumab to be neither inferior nor superior to licensed aflibercept. Our study highlights how real-world cohorts representing the counterfactual intervention could aid the interpretation of single-armed trials when analyzed in accord to the target trial framework.

INTRODUCTION

One-in-four ophthalmology trials are single-armed,[1] which poses challenges to their interpretation in the absence of a standard-of-care control arm. Nevertheless, recent conceptual advances in the methods of causal inference and in the emulation of target trials[2-5] suggests that the standard-of-care arms representing the counterfactual intervention can be approximated (the external/synthetic control arm henceforth), and that real-world analogues of trial findings are concordant with the trials they emulate.[6] Central to these successes is the emulated target trial design and the target trial framework.[7]

A target trial is the hypothetical trial that would in ideal circumstances be undertaken. The target trial framework outlines a structured approach to the design and analysis of observational research as if they were randomized trials.[7] By conceptualizing observational research as an attempt to emulate the target trial, and making explicit the experimental design and statistical analysis as one would in a trial protocol, the target trial framework can help researchers to identify and avoid biases resulting from an inappropriate selection of patients or misalignment of time zero.[8] An additional consideration for leveraging real-world cohorts as external controls is that one must also attempt conditional randomization on measured confounding, because the nonrandom treatment assignment in the real-world introduces confounding biases. This is often achieved either within subsets defined by levels of confounding through matching or within levels of confounding through inverse probability weighting.[9-12]

We herein emulated a target trial on the relative efficacy of intravitreal administration of two inhibitors of VEGF (off-label bevacizumab vs. licensed aflibercept) for improving the visual acuity of eyes affected by neovascular age-related macular degeneration (nAMD) after 54 weeks of treatment. Our emulated target trial compared the bevacizumab arm ($n = 65$) of the ABC trial[13]—a prospective, double-masked, randomized controlled trial—with a synthetic control arm from 31,151 eyes receiving aflibercept during routine care as recorded by electronic health records (EHRs). The choice of intervention and comparator is based on the lack of randomized controlled trials comparing the two therapies, despite off-label bevacizumab being the most commonly used drug to treat nAMD in the United States.[14] The size of the ABC trial bevacizumab arm is similar to that used in early phase single-armed clinical trials. Our report follows the emulated target trial framework.[7]

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?
One-in-four randomized controlled trials in ophthalmology are single-armed, which poses challenges for interpreting their efficacy relative to standard of care. Recent conceptual advances in the methods of causal inference and in the emulation of target trials suggests that the standard-of-care arms representing the counterfactual intervention can be approximated with observational data.

WHAT QUESTION DID THIS STUDY ADDRESS?
How real-world cohorts representing the counterfactual intervention can aid the interpretation of single-armed ophthalmological trials.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?
Our study highlights how real-world cohorts representing the counterfactual intervention could aid the interpretation of single-armed ophthalmological trials when undertaken in accord with the target trial framework.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?
External counterfactual arms could reduce the time and cost to reach potential regulatory approval.

after 3 loading injections (i.e., the bevacizumab trial arm) with an external synthetic control arm of eyes that received aflibercept (2 mg) during routine care on the intention to treat at fixed 8-week intervals (q8w) reflective of the posology at launch of the therapy (i.e., the aflibercept synthetic arm). The target trial protocol is outlined in Table S1.

Informed consent and ethics

The ABC trial was approved by the ethics committee at each clinical site with all patients giving written informed consent to participate. The Caldicott data protection guardian at each EHR site gave signed permission for the de-identifiable EHRs to be extracted for analysis.

Subjects

Bevacizumab trial arm

The ABC trial was a prospective, double-masked, multi-center randomized controlled trial undertaken in the United

METHODS

Study design

An emulated target trial was undertaken comparing week 54 outcomes of eyes that, during the ABC trial, received bevacizumab (1.25 mg) pro re nata at 6-week intervals (q6w)

CONTEXTUALIZING SINGLE-ARM TRIALS WITH REAL-WORLD DATA: AN EMULATED TARGET
TRIAL COMPARING THERAPIES FOR NEOVASCULAR AGE-RELATED MACULAR DEGENERATION

3

Kingdom during 2006–2008. The method has previously been reported.[15] Briefly, 131 patients were randomized 1:1 to receive in the study eye: (I) 1.25 mg bevacizumab intravitreally at q6w pro ne rata until week 48 after 3 mandatory loading injections at weeks 0, 6, and 12 ($n = 65$) or (II) the standard of care at the time ($n = 66$); the latter arm being irrelevant to this study. Patients could receive 3–9 bevacizumab treatments (0–6 maintenance injections). Administration of injection at each visit was determined pro ne rata based on the presence of subretinal fluid, development of hemorrhage, or classic choroidal neovascularization, loss of greater than five Early Treatment Diabetic Retinopathy Study (ETDRS) letters in association with new intraretinal fluid or persistent intraretinal fluid.[16] Patients underwent assessments for visual acuity at each visit before treatment was administered, with a final assessment at week 54. Best-corrected visual acuity measurements were assessed using ETDRS reading charts.

Aflibercept synthetic arm

Synthetic control arms were sampled from EHRs recording ophthalmologic care across 27 sites in England during 2012–2018 (Medisoft Ophthalmology platform (http://www.medisoft.co.uk), Leeds, UK). Medisoft EHRs hold anonymized longitudinal data on all episodes of ophthalmologic care from treatments and surgeries through to visual acuity measurements, diagnostics, and complications. All visual acuity measurements were the best-recorded unaided and were measured via, or converted to, ETDRS letters. Qualitative assessments of counting fingers, hand motion, light perception, and no light perception were converted to ETDRS letter scores of 2, 0, 0, and 0, respectively.

Study eyes were the first diagnosed (or randomly sampled for bilateral diagnosis on the same day) that initiated treatment with aflibercept (2 mg q8w fixed-dose after 3 loading injections at ~ 4 weekly intervals within 70 days) for nAMD (for EHR phenotype, see Table S2). The pool of potential synthetic controls were eyes that, as far as we could judge, would have been eligible for the ABC trial using the eligibility criteria in Table S1 (for criteria not recorded by EHRs, see Table S3). In addition to the ABC eligibility criteria, we also excluded all eyes having an incomplete loading phase; all those that had unrecorded sex, age, or baseline visual acuity required for matching or propensity scoring; and all eyes that initiated treatment after 1 December 2017 (allowing for 54 weeks follow-up).

Emulating randomization

Three methods (exact matching [EM], inverse probability treatment weighting [IPTW], and propensity score matching [PSM]) for achieving conditional exchangeability were compared with an unconditional (UC) design wherein all eyes eligible were analyzed post protocol alignment but without addressing exchangeability. Hypothesized confounders were age, sex, and visual acuity at baseline read.[17,18] The study workflow is presented in Figure 1.

Unconditional analysis

The UC analysis included all potential synthetic controls that aligned with the ABC and emulated target trial-specific criteria analyzed without emulating randomization.

Exact matching

The aflibercept arm was exactly matched 1:1 with the bevacizumab trial arm on baseline confounders (age, sex, and visual acuity). Those with greater than one exact matches were paired with one randomly selected. Those with no exact matches were excluded from the analyses.

Inverse propensity score weighting and propensity score matching

The propensity score (Pr[A|L]) describes the probability of being in the ABC bevacizumab trial arm conditional on confounding variables (age, sex, and visual acuity). The score was derived from a binomial logistic model trained on all 131 ABC trial subjects (65 and 66 randomized to receive bevacizumab and standard of care, respectively (Equation 1)). The synthetic distribution tails not overlapping with the propensity scores of the bevacizumab arm were trimmed prior to weighting and matching.

The IPTW analysis included all 65 bevacizumab-treated eyes and all aflibercept-treated eyes that fulfilled the ABC trial and target trial eligibility criteria. Outcomes for each eye were thereafter weighted on the conditional inverse probability of treatment (1/Pr[A|L] for the bevacizumab arm and 1/(1 − Pr[A|L]) for the aflibercept arm). For PSM, synthetic controls were matched 1:1 to each bevacizumab eye ($n = 65$) within a caliper of 0.1 standard deviation.

$$\text{Propensity score} = \frac{e\left(5.42 - 0.05\,(\text{age}) - 0.09\,(\text{sexM}) - 0.02\,(\text{etdrs})\right)}{1 + e\left(5.42 - 0.05\,(\text{age}) - 0.09\,(\text{sexM}) - 0.02\,(\text{etdrs})\right)} \quad (1)$$

Equation 1 = Propensity score trained on ABC trial data.

Statistical analyses

We estimated the intention-to-treat effects under each method for emulating randomization, including all eyes
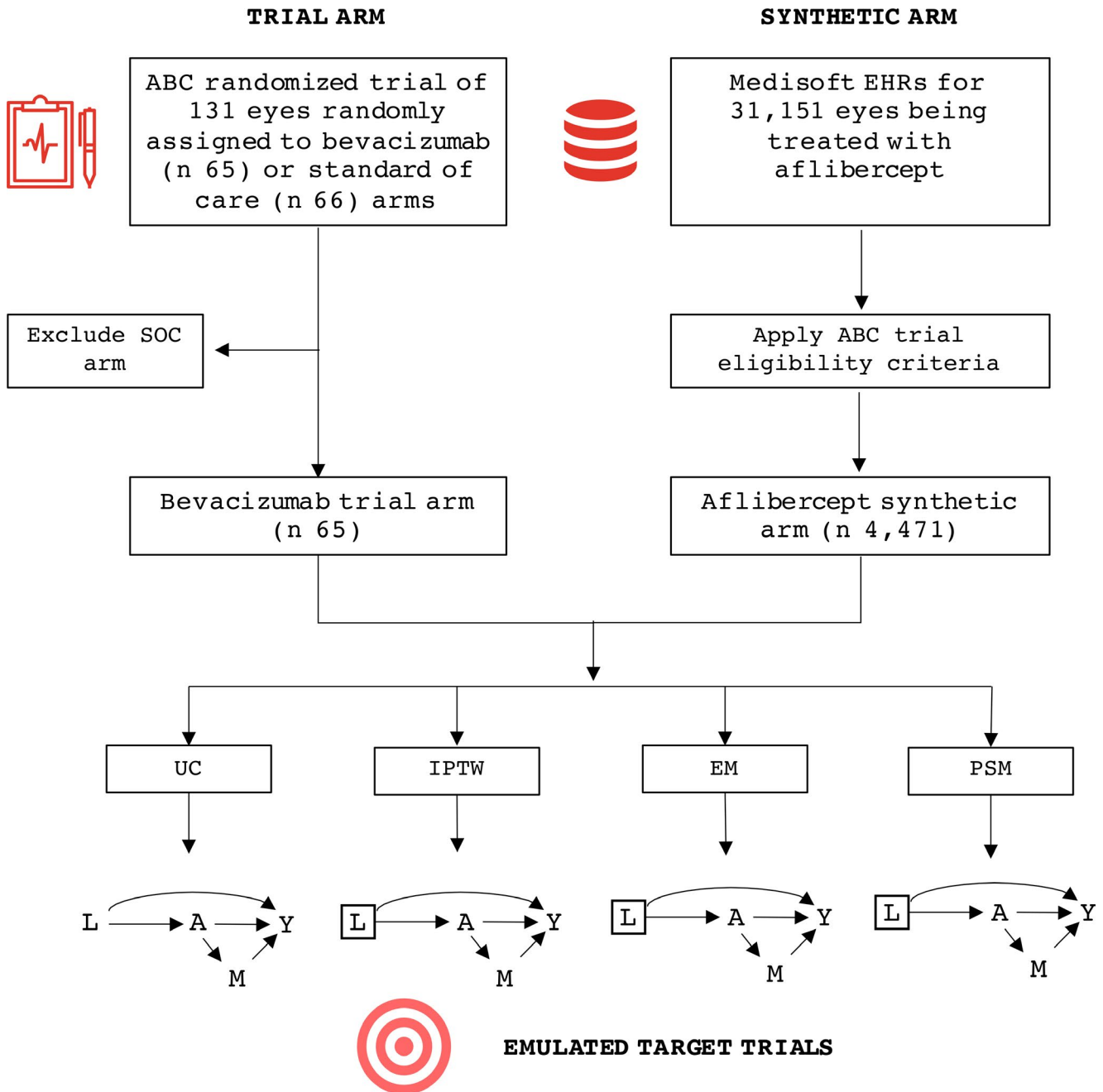
TRIAL ARM **SYNTHETIC ARM**

FIGURE 1 An emulated target trial estimating the causal effect of bevacizumab relative to aflibercept on the week 54 visual acuity of eyes affected by neovascular age-related macular degeneration. Age, sex, and ETDRS at baseline read (L) are common causes of both treatment assignment (A) and outcome (Y), whereas the number of maintenance injections received during the study period (M) mediates between A and Y. Thus, blocking the backdoor criterion L–Y via emulated randomization (denoted by □), which is assumed to naturally arise in the bevacizumab trial arm through randomization, permits unbiased estimates of the causal pathway A–Y. The backdoor criterion remains open under UC. EHRs, electronic health records; EM, exact matching; ETDRS, Early Treatment Diabetic Retinopathy; IPTW, inverse probability of treatment weighting; PSM, propensity score matching; SOC, standard of care; UC, unconditional

that started treatment with bevacizumab or aflibercept regardless of the drug load received, treatment crossover, or number of post-baseline measurements during the study period. Time zero (baseline henceforth) for all analyses was synchronized to the date of first treatment (also the day of randomization for the bevacizumab arm). The date of time zero was also the date that we established patient eligibility.

Baseline reads were taken on the date of first treatment, or the nearest up to 30 days prior for external controls (but rarely is visual acuity not measured immediately prior to treatment). Study exits were dependent on the analysis, as detailed later. Exchangeability in confounding variables was assessed pre-emulating and post-emulating randomization via standardized differences calculated as reported.[19]

CONTEXTUALIZING SINGLE-ARM TRIALS WITH REAL-WORLD DATA: AN EMULATED TARGET
TRIAL COMPARING THERAPIES FOR NEOVASCULAR AGE-RELATED MACULAR DEGENERATION

5

A standardized difference less than 0.1 was considered evidence for conditional exchangeability. Statistical inferences were made at an α threshold of less than 0.05.

Noninferiority

An ordinary least squares regression was modeled on the change in visual acuity from baseline to week 54 measurement, stratified by treatment (aflibercept arm as reference). Noninferiority was declared if the lower confidence interval (CI) bound was greater than or equal to a noninferiority margin of minus four ETDRS letters. This margin of noninferiority mirrors that of the HAWK and HARRIER randomized trials recently undertaken,[20] which we took as the current consensus. If no measurement was taken during week 54 owing to the inexactness of routine healthcare attendances, we used the measurement taken nearest to week 54 during weeks 50–58, taking the earliest date if 2 dates were equally tied. Missing data were imputed using the last observation carried forward.

Binomial regression

A binomial logistic regression, stratified by treatment, was modeled on the outcomes of achieving a change in visual acuity from baseline to week 54 measurement of greater than or equal to 15 letters, greater than or equal to 10 letters, and less than or equal to −15 letters. We present the odds ratios (ORs) and 95% CIs relative to the aflibercept synthetic arm as reference. If no measurement was taken during week 54, we used the measurement taken nearest to week 54 during weeks 50–58, taking the earliest date if 2 dates were equally tied. Missing data were imputed using the last observation carried forward.

Time-to-event

Kaplan-Meier estimators were modeled on the time from baseline to achieve a change in visual acuity of greater than or equal to 15 letters, greater than or equal to 10 letters, and less than or equal to −15 letters, stratified by treatment (aflibercept arm as reference). Right censorship was on the last visual acuity measurement taken during the study period, up to 54 weeks maximum. Inferences were made on the outputs of the log rank test comparing event distributions.

Software

EHRs were extracted by Medisoft (Leeds, UK) and stored as relational tables on a MySQL server (version 5.5.62). All analyses were undertaken with R version 4.0.2[21] with

particular use of tidyverse version 1.3.0,[22] survival version 3.2–3,[23] and fuzzyjoin version 0.1.6.[24] All code and an exhaustive listing of all R packages used are available at https://github.com/dsthom/synthetic_trial_amd.

RESULTS

Subjects

A pool of 31,151 eyes that started aflibercept treatment for nAMD was identified. Of these, 26,680 (86% of 31,151) were ineligible, for either not fulfilling the ABC ($n = 20,823$, 67%) or the emulated trial criteria ($n = 5,857$, 19%). Thus, a pool of 4,471 (14%) eyes were available for study. All baseline reads for these 4,471 eligible were taken on the same day as treatment. Of these, 186 eyes were trimmed from the IPTW and PSM analyses for having nonoverlapping propensity scores for a pool of 4,141 (13%) eyes. A stepwise breakdown of ineligibility is presented in Table S4. Eyes included in each bevacizumab trial and aflibercept synthetic arms were respectively, 65 and 4,471 for UC; 131 and 8,506 for IPTW (pseudo-population); 43 and 43 for EM; and 65 and 65 for PSM.

The aim of emulating randomization was to achieve conditional exchangeability in measured confounding (summarized in Table 1). Prior to aligning protocols, arms were imbalanced on age and baseline ETDRS, and after aligning protocols (UC) on age only. IPTW and EM were balanced for all measured confounding, whereas imbalances remained post-PSM for age and the process in itself seemed to imbalance baseline ETDRS. The causal assumption of positivity was satisfied for all levels of confounding.

Noninferiority

Approval of novel nAMD therapies is often based on proof of their noninferiority to currently approved drugs. Our results generally suggest, with exception of PSM, that bevacizumab is not inferior to aflibercept assessed on a noninferiority margin of four ETDRS letters (Figure 2). There was, however, insufficient evidence to conclude noninferiority under PSM (−0.1 mean ETDRS difference relative to aflibercept). Confidence bounds for matching methods were notably wider than those for UC and IPTW.

Superiority

We mirrored the primary analysis of the ABC trial in which the proportion of eyes in each treatment arm that gained greater than or equal to 15 ETDRS letters at week 54 were compared (Figure 3). Secondary analysis on the proportion of

Variable	Aflibercept synthetic arm	Bevacizumab trial arm	Standardized difference
Pre-aligning of trial protocols ($n = 25{,}199$)			
Arm size	25,134[a]	65	—
Sex (female)	15,860 (63%)	39 (60%)	0.06
Age in years	x̄ 80 (σx 8)	x̄ 79 (σx 8)	0.14
Baseline ETDRS	x̄ 56 (σx 17)	x̄ 51 (σx12)	0.34
Propensity score	x̄ 0.48 (σx 0.12)	x̄ 0.52 (σx 0.10)	0.34
Unconditional ($n = 4536$)			
Sample n	4,471	65	—
Sex (female)	2,817 (63%)	39 (60%)	0.06
Age in years	x̄ 80 (σx 8)	x̄ 79 (σx 8)	0.14
Baseline ETDRS	x̄ 52 (σx13)	x̄ 51 (σx 12)	0.09
Propensity score	x̄ 0.50 (σx 0.11)	x̄ 0.52 (σx 0.10)	0.17
Inverse probability of treatment weighting ($n = 8637$[b])			
Sample n	8,506[b]	131[b]	—
Sex (female)	5,372 (63%)	81 (62%)	0.03
Age in years	x̄ 80 (σx 7)	x̄ 80 (σx 7)	0.06
Baseline ETDRS	x̄ 52 (σx 13)	x̄ 52 (σx 12)	0.03
Propensity score	x̄ 0.50 (σx 0.10)	x̄ 0.49 (σx 0.10)	0.09
Exact matching ($n = 86$)			
Sample n	43	43	—
Sex (female)	29 (67%)	29 (67%)	0
Age in years	x̄ 80 (σx 6)	x̄ 80 (σx 6)	0
Baseline ETDRS	x̄ 51 (σx 11)	x̄ 51 (σx 11)	0
Propensity score	x̄ 0.50 (σx 0.09)	x̄ 0.50 (σx 0.09)	0
Propensity Score Matching ($n = 130$)			
Sample n	65	65	—
Sex (female)	40 (62%)	39 (60%)	0.03
Age in years	x̄ 78 (σx 8)	x̄ 79 (σx 8)	0.12
Baseline ETDRS	x̄ 53 (σx 12)	x̄ 51 (σx 12)	0.20
Propensity score	x̄ 0.52 (σx 0.10)	x̄ 0.52 (σx 0.10)	0

Abbreviation: ETDRS, Early Treatment Diabetic Retinopathy.

[a]There were 31,151 minus eyes with unrecorded sex, age, or baseline ETDRS.

[b]Pseudo-population created by weighting.

eyes having a change in visual acuity from baseline to week 54 measurement of greater than or equal to 10 letters and less than or equal to –15 letters are also presented. Under no trials tested was bevacizumab shown to be superior to aflibercept.

Time-to-event

It may be that the drugs themselves, or the protocols under which they were administered, are beneficial at varying rates during the study period. We explored this possibility with the Kaplan-Meier estimator modeled on the time to a change in visual acuity from baseline of greater than or equal to 15, greater than or equal to 10 and less than or equal to −15

ETDRS letters (Figure 4). The log rank statistic suggests there being no difference in the event distribution of eyes receiving bevacizumab relative to aflibercept. The Kaplan-Meier curves, too, suggest disproportional hazards over time, although this may be an artifact of the fixed-interval measurements of the trial arm.

DISCUSSION

Principal findings

One-in-four ophthalmology trials are single-armed, which poses challenges for their interpretation. Our study highlights

CONTEXTUALIZING SINGLE-ARM TRIALS WITH REAL-WORLD DATA: AN EMULATED TARGET
TRIAL COMPARING THERAPIES FOR NEOVASCULAR AGE-RELATED MACULAR DEGENERATION

7

how real-world cohorts representing the counterfactual intervention could aid the interpretation of these single-armed trials when analyzed in accord to the target trial framework. It also suggests that week 54 visual outcomes of eyes treated with off-label bevacizumab pro re nata (q6w) are neither inferior nor superior to licensed aflibercept administered at fixed intervals (q8w).

FIGURE 2 Mean ETDRS difference in change in visual acuity from baseline to week 54 measurement for bevacizumab-treated eyes relative to those treated with aflibercept. Dashed vertical line denotes a noninferiority margin of minus four ETDRS letters. EM, exact matching; ETDRS, Early Treatment Diabetic Retinopathy; IPTW, inverse probability of treatment weighting; PSM, propensity score matching; UC, unconditional

In context with the evidence base

The results suggest off-label bevacizumab administered pro re nata (q6w) to be neither inferior nor superior to licensed aflibercept administered at fixed intervals (q8w) for improving the week 54 vision of eyes affected by nAMD. Although we have no trial reference to validate the present findings with, we can attempt to contextualize ours in relation to the literature. At writing, two randomized trials have reported on the efficacy of inhibitors of VEGF for treating nAMD; each having a common comparator arm, ranibizumab, from which we can triangulate. The CATT trial demonstrated bevacizumab to be noninferior to ranibizumab.[25] The VIEW trial reported aflibercept to be noninferior to ranibizumab.[26] These findings taken together suggests there being no biological reason to suspect bevacizumab to be inferior to aflibercept, which is consistent with the results of this emulated trial. Nevertheless, it should be acknowledged that this finding remains a hypothesis that should be consolidated with randomized interventions.

Second, the methods of causal inference used herein have been validated by a number of studies that assessed the extent to which a target trial can be emulated using real-world data. In a study on statins and cancer, the authors revisited the past failures of observational analogues to arrive at the findings of clinical trials.[5] It was argued that this failure is not primarily due to confounding biases introduced by nonrandom treatment assignment, as is often thought, but also to biases introduced by subjects and analyses deviating from an explicit target trial protocol; advice that later evolved into the target trial framework[14] that we followed herein. Under the target trial approach, observational estimates were concordant with those of the trial that was emulated.[5] In another study most applicable to ours, the estimates obtained by emulated target trials composed of one trial arm and an external control arm derived from real-world data were compared with that of a reference
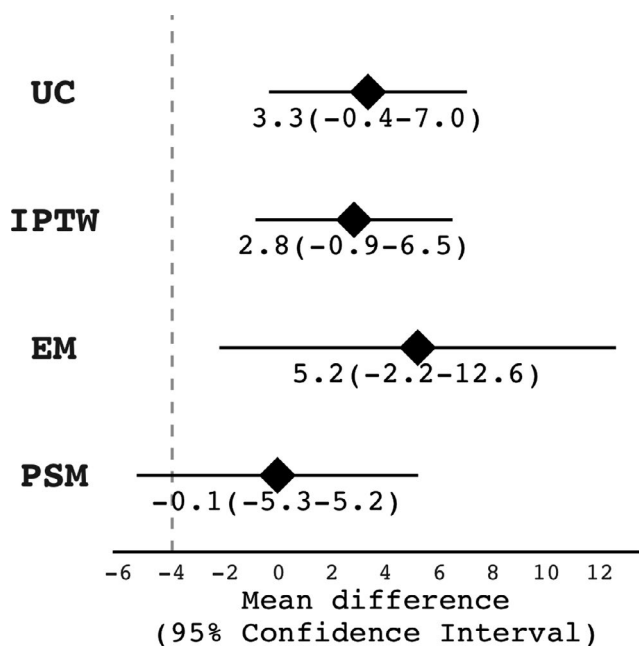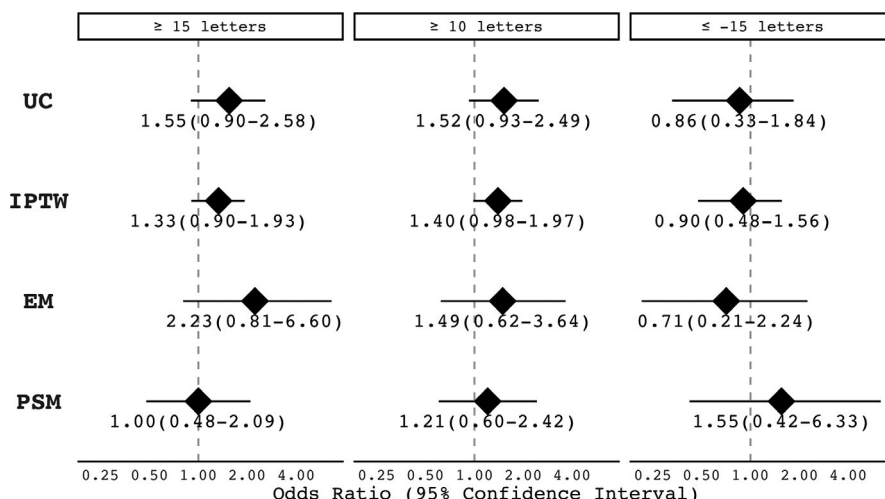

FIGURE 3 Odds of bevacizumab-treated eyes achieving a priori outcomes at week 54 relative to those treated with aflibercept. EM, exact matching; IPTW, inverse probability of treatment weighting; PSM, propensity score matching; UC, unconditional
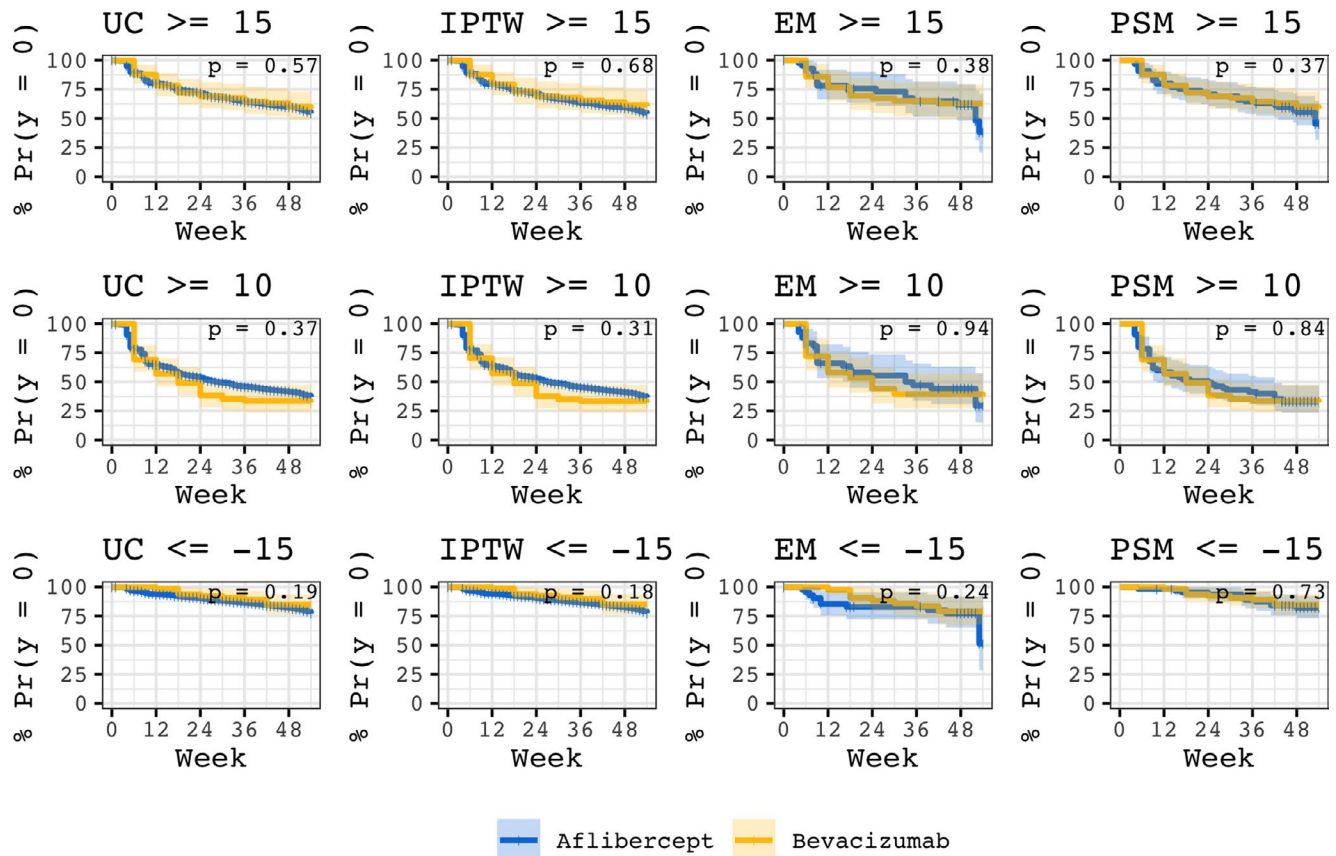
FIGURE 4 Hazards of bevacizumab-treated eyes achieving a priori outcomes during the first 54 weeks of treatment relative to those treated with aflibercept. Hued ribbons represent 95% confidence intervals. EM, exact matching; IPTW, inverse probability of treatment weighting; PSM, propensity score matching; UC, unconditional

two-armed trial.[10] The paper reports, through alignment of protocols and propensity-based adjustment, concordance in 10 out of the 11 trials emulated. It is on these validated methods that our findings are dependent on.

A criticism of PSM is that its use may inadvertently exacerbate exchangeability,[26] and we question whether there is a link between this possibility and the discordance in the PSM noninferiority estimate relative to other methods. Given the evidence base is that visual acuity on starting treatment is the most important determinant of visual improvement,[17,27,28] it may therefore be plausible that the imbalance of the PSM aflibercept arm toward greater baseline acuity—despite exchangeability in the dimension of propensity score—nullified the relative efficacy of bevacizumab. Notably, of all methods for emulating randomization, PSM was the only to lead to inexchangeability in baseline read, and more curiously was that this was being exchangeable prior to emulating randomization. Another notable limitation of matching in general is the reduced sample size leading to widened confidence bounds and potential biases from discarding unmatched observations. IPTW, in contrast, amplifies the weighting of these rare events without discarding data.

Strengths and limitations

The target trial framework provides a structured approach to improve the rigor of observational research, by applying the best practices for the design and analysis of randomized trials.[7] Adhering to the framework can help researchers identify and avoid biases in an inappropriate selection or the accrual of immortal time.[8] A major strength of our research is, because we adhered to the framework, the risk of these biases is low to nonexistent. For the vast majority of eligibility criteria, we were also able to align the synthetic arms with the ABC trial protocol so that the trial and real-world arms were comparable. For those criteria not explicitly recorded by Medisoft EHRs (Table S3), some of these would have been excluded through other criteria (exclusion of eyes ever treated for cataracts or myopia, for example); are implied through treatment not to be present (inadequate contraception in premenopausal women, no contraindications or active infection, or recent stroke or cardiac event); or do not lie on a backdoor path between treatment and outcome (concurrent warfarin use and fluorescein allergy). Any potential risk for bias would be through an inability to exclude aphakic eyes

CONTEXTUALIZING SINGLE-ARM TRIALS WITH REAL-WORLD DATA: AN EMULATED TARGET
TRIAL COMPARING THERAPIES FOR NEOVASCULAR AGE-RELATED MACULAR DEGENERATION

9

that have poorer prognoses, but given their rarity we believe that the risk of bias to be low.

Naturally, we also acknowledge several limitations. The first being the potential for residual confounding. Where trials can assume exchangeability across all confounders—measured and unmeasured—due to random treatment assignment, emulated target trials in contrast are limited to exchangeability only in measured variables. Second, the nonoverlap in treatment periods between the bevacizumab (2006–2008) and aflibercept (2012–2018) arms and differences between trial and real-world cohorts could bias our estimates. As shown in Table 1, the alignment of the real-world cohorts to the ABC trial population (mean baseline read 51) lowers the mean from 56 ETDRS in those initiating treatment in the real-world to 51–53 for the external cohorts. Because eyes with starting vision approaching the ceiling stand to gain less improvement, one would expect the pragmatic treatment effects for a contemporaneous sample to shrink toward the null relative to those reported here.

Implications for stakeholders

It is evident through the founding of the FDA Real-World Evidence Program[27] that regulatory decisions may increasingly rely on real-world evidence. Indeed, the use of external counterfactual arms has already expedited the regulatory approval of blinatumomab (Amgen)[10] and alectinib (Roche)[12] for oncologic indications. Nevertheless, barring a small evidence base, the extent to which randomized trials can be emulated by real-world data is largely unknown. In consequence, the US Food and Drug Administration (FDA) has funded the RCT DUPLICATE initiative, which aims to determine the validity of using real-world evidence for regulatory decisions.[5]

More immediately, our study has implications for the economics of treating nAMD, for bevacizumab is somewhat more cost-effective than licensed therapies.[28] For context, the treatment of retinal disorders with inhibitors of VEGF accounts for 12% of all Medicare expenditure,[29] and a 2018 survey by the American Society of Retina Specialists reports aflibercept being the first-line therapy for 13%, 16%, 37%, 41%, and 47% of ophthalmologists surveyed across Africa, the United States, Europe, Asia, and Central and South America.[14]

Unanswered questions and future research

With no existing trials comparing the efficacy of bevacizumab to aflibercept and with bevacizumab not being prescribed per ABC protocol during routine health care, it was not possible for us to triangulate our findings against a reference standard. We nevertheless acknowledge the importance of validating our findings, and aim to use the Medisoft EHRs to triangulate our methods relative to existing clinical trials. Furthermore, the estimation of per protocol effects, either by naive definition or by weighting on the conditional probability of treatment adherence,[30] would be an important contribution to the evidence base. This would require the delineation of nonadherence from pro re nata.

CONCLUSIONS

Our study highlights how real-world cohorts representing the counterfactual intervention could aid the interpretation of single-armed ophthalmological trials when undertaken in accord with the target trial framework.

CONFLICT OF INTEREST
Aaron Y. Lee is an employee of the FDA, has received research funding from Novartis, Santen, Carl Zeiss, and Meditec, and has acted as a consultant for Genentech, Topcon, and Verana Health. Praven J. Patel has received research funding from Bayer UK, and has acted as a consultant for Bayer UK, Genentech, Novartis UK, and Roche. Roy Schwartz has received travel expenses from Allergan (AbbVie) and is a board member of Eye-Wise diagnostics. Paul Taylor has received a grant from Novartis Pharmaceuticals. Adnan Tufail, Praven J. Patel, and Catherine Egan received a proportion of their funding from the Department of Health's NIHR Biomedical Research Centre for Ophthalmology at Moorfields Eye Hospital and UCL Institute of Ophthalmology. The views expressed in this publication are those of the authors and not necessarily those of the Department of Health. All other authors declared that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

DISCLAIMER
The opinions in this manuscript are those of the authors and should not be interpreted as the position of the US Food and Drug Administration.

AUTHOR CONTRIBUTIONS
All authors wrote the manuscript. A.Y.L., A.T., D.S.T., and P.T. designed the research. D.S.T. performed the research. All authors analyzed the data.

REFERENCES
1. Turner B, Rajeshuni N, Tran EM, et al. Characteristics of ophthalmology trials registered in ClinicalTrials.gov, 2007–2018. *Am J Ophthalmol.* 211, 132-141 (2020).

2. Lodi S, Phillips A, Lundgren J, et al. Effect estimates in randomized trials and observational studies: Comparing apples with apples. *Am J Epidemiol*. 2019;188:1569-1577.

3. Dickerman BA, García-Albéniz X, Logan RW, Denaxas S, Hernán MA. Avoidable flaws in observational analyses: An application to statins and cancer. *Nat Med*. 2019;25:1601-1606.

4. Hernán MA, Robins JM. *Causal Inference: What If*. Boca Raton, FL: Chapman & Hall/CRC; 2020.

5. Franklin JM, Pawar A, Martin D, et al. Nonrandomized real-world evidence to support regulatory decision making: Process for a randomized trial replication project. *Clin Pharmacol Ther*. 2020;107:817-826.

6. Carrigan G, Whipple S, Capra WB, et al. Using electronic health records to derive control arms for early phase single-arm lung cancer trials: Proof-of-concept in randomized controlled trials. *Clin Pharmacol Ther*. 2020;107:369-377.

7. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol*. 2016;183:758-764.

8. Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J Clin Epidemiol*. 2016;79:70-75.

9. Schmidli H, Häring DA, Thomas M, Cassidy A, Weber S, Bretz F. Beyond randomized clinical trials: Use of external controls. *Clin Pharmacol Ther*. 2020;107:806-816.

10. Gökbuget N, Kelsh M, Chia V, et al. Blinatumomab vs historical standard therapy of adult relapsed/refractory acute lymphoblastic leukemia. *Blood Cancer J*. 2016;6:e473.

11. Tan DS-W, Araújo A, Zhang J, et al. Comparative efficacy of ceritinib and crizotinib as initial ALK-targeted therapies in previously treated advanced NSCLC: An adjusted comparison with external controls. *J Thorac Oncol*. 2016;11:1550-1557.

12. Davies J, Martinec M, Delmar P, et al. Comparative effectiveness from a single-arm trial and real-world data: Alectinib versus ceritinib. *J Comp Eff Res*. 2018;7:855-865.

13. Tufail A, Patel PJ, Egan C, et al. Bevacizumab for neovascular age related macular degeneration (ABC Trial): Multicentre randomised double masked study. *BMJ*. 2010;340:c2459.

14. Singh RP, Stone TW. *2018 Global Trends in Retina Survey*. Chicago, IL: In American Society of Retina Specialists; 2018.

15. Patel PJ, Bunce C, Tufail A; ABC Trial Investigators. A randomised, double-masked phase III/IV study of the efficacy and safety of Avastin(R) (Bevacizumab) intravitreal injections compared to standard therapy in subjects with choroidal neovascularisation secondary to age-related macular degeneration: Clinical trial design. *Trials*. 2008;9:56.

16. Patel PJ, Tufail A, for the ABC Trial Investigators. Optimizing individualized therapy with bevacizumab for neovascular age-related macular degeneration[published online ahead of print January 31, 2012]. *Retina*. https://doi.org/10.1097/IAE.0b013e31823f0ba3.

17. Writing Committee for the UK Age-Related Macular Degeneration EMR Users Group. The neovascular age-related macular degeneration database: Multicenter study of 92 976 ranibizumab injections: report 1: visual acuity. *Ophthalmology* 2014;121:1092-1101.

18. Tufail A, Margaron P, Guerin T, Larsen M. Visual benefit versus visual gain: What is the effect of baseline covariants in the treatment arm relative to the control arm? A pooled analysis of ANCHOR and MARINA. *Br J Ophthalmol*. 2020;104:672-677.

19. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res*. 2011;46:399-424.

20. Dugel PU, Koh A, Ogura Y, et al. HAWK and HARRIER: phase 3, multicenter, randomized, double-masked trials of brolucizumab for neovascular age-related macular degeneration. *Ophthalmology*. 2020;127:72-84.

21. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019. https://www.R-project.org/

22. Wickham H, Averick M, Bryan J, et al. Welcome to the tidyverse. *J Open Source Soft*. 2019;4:1686.

23. Therneau TM.A Package for Survival Analysis in R, 2020. <https://CRAN.R-project.org/package=survival>.

24. Robinson D.fuzzyjoin: Join Tables Together on Inexact Matching, 2020. <https://CRAN.R-project.org/package=fuzzyjoin>.

25. CATT Research Group, Martin DF, Maguire MG, et al. Ranibizumab and bevacizumab for neovascular age-related macular degeneration. *N Engl J Med*. 2011;364:1897-1908.

26. Heier JS, Brown DM, Chong V, et al. Intravitreal aflibercept (VEGF trap-eye) in wet age-related macular degeneration. *Ophthalmology*. 2012;119:2537-2548.

27. U. S. Food and Drug Administration. Framework for FDA's Real-World Evidence Program, 2018.

28. van Asten F, Michels CTJ, Hoyng CB, et al. The cost-effectiveness of bevacizumab, ranibizumab and aflibercept for the treatment of age-related macular degeneration - a cost-effectiveness analysis from a societal perspective. *PLoS One*. 2018;13:e0197670.

29. Patel S. Medicare spending on anti-vascular endothelial growth factor medications. *Ophthalmol Retina*. 2018;2:785-791.

30. Murray EJ, Hernán MA. Adherence adjustment in the Coronary Drug Project: A call for better per-protocol effect estimates in randomized trials. *Clin Trials*. 2016;13:372-378.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.