Title

Sequential regression and simulation: a method for estimating causal effects from heterogeneous clinical trials without a common control group

Permalink

https://escholarship.org/uc/item/2gq4148t

Journal

BMC Medical Research Methodology, 23(1)

ISSN

1471-2288

Authors

Rudrapatna, Vivek A
Ravindranath, Vignesh G
Arneson, Douglas V
et al.

Publication Date

2023

DOI

10.1186/s12874-023-02020-5

Copyright Information

Peer reviewed

RESEARCH

Sequential regression and simulation: a method for estimating causal effects from heterogeneous clinical trials without a common control group

Vivek A. Rudrapatna[1,2*], Vignesh G. Ravindranath[2], Douglas V. Arneson[2], Arman Mosenia[3], Atul J. Butte[2] and Shan Wang[4*]

Abstract

Background The advent of clinical trial data sharing platforms has created opportunities for making new discoveries and answering important questions using already collected data. However, existing methods for meta-analyzing these data require the presence of shared control groups across studies, significantly limiting the number of questions that can be confidently addressed. We sought to develop a method for meta-analyzing potentially heterogeneous clinical trials even in the absence of a common control group.

Methods This work was conducted within the context of a broader effort to study comparative efficacy in Crohn's disease. Following a search of clnicaltrials.gov we obtained access to the individual participant data from nine trials of FDA-approved treatments in Crohn's Disease ($N = 3392$). We developed a method involving sequences of regression and simulation to separately model the placebo- and drug-attributable effects, and to simulate head-to-head trials against an appropriately normalized background. We validated this method by comparing the outcome of a simulated trial comparing the efficacies of adalimumab and ustekinumab against the recently published results of SEAVUE, an actual head-to-head trial of these drugs. This study was pre-registered on PROSPERO (#157,827) prior to the completion of SEAVUE.

Results Using our method of sequential regression and simulation, we compared the week eight outcomes of two virtual cohorts subject to the same patient selection criteria as SEAVUE and treated with adalimumab or ustekinumab. Our primary analysis replicated the corresponding published results from SEAVUE ($p = 0.9$). This finding proved stable under multiple sensitivity analyses.

Conclusions This new method may help reduce the bias of individual participant data meta-analyses, expand the scope of what can be learned from these already-collected data, and reduce the costs of obtaining high-quality evidence to guide patient care.

Keywords Individual participant data meta-analysis, Randomized clinical trials, Crohn's disease, Comparative effectiveness, Comparative efficacy, Evidence synthesis, Biostatistics

*Correspondence:
Vivek A. Rudrapatna
vivek.rudrapatna@ucsf.edu
Shan Wang
swang151@usfca.edu
Full list of author information is available at the end of the article

Rudrapatna *et al. BMC Medical Research Methodology*   (2023) 23:218

Page 2 of 12

Background

The individual participant data (IPD) meta-analysis of randomized trials is the gold-standard for clinical research [1, 2]. Access to the raw data from trials affords investigators the opportunity to verify published results, ask new questions of these data, and uncover findings with the potential to impact patient care.

Performing an IPD meta-analyses usually requires multiple trials with negligible heterogeneity across many dimensions: cohort definition, randomization, blinding, parallel study arms, interventions, and outcomes [1, 2]. Although this requirement ensures unbiased estimation, it substantially limits the number of meta-analyses that can be performed due to the rarity of replicate trials.

The method of mixed-effects regression is commonly used to address study heterogeneity when meta-analyzed trials include a shared control group (i.e. placebo). However, there is a paucity of methods for common situations where there is no shared control group across potential studies. The few methods that have been developed include naïve pooling [3], as well as the Bayesian method of power priors [4–7]. However, these methods fail to address the problem of cohort heterogeneity [8, 9]. Another major limitation is the lack of external validation against prospective studies. The result of this methodological gap is the common practice of excluding uncontrolled studies from potential meta-analyses, and ultimately fewer research questions that we are statistically powered to answer using already-collected data.

Here we report a new method for meta-analyzing clinical trials data in the absence of a common control group. We illustrate our method of sequential regression and simulation in the context of a comparative efficacy analysis in Crohn's disease, an immune disorder of the gastrointestinal tract. We use the data from six placebo-controlled trials ($N = 3153$) to develop a model of the placebo effect, then apply this to three placebo-less trials ($N = 239$) to normalize and separately model the drug-attributable response. Finally, we validate the method by predicting the results of SEAVUE (NCT03464136), a recent head-to-head trial of ustekinumab versus adalimumab [10].

Methods

This study was approved by the UCSF IRB (#18–24,588). It was pre-registered on PROSPERO [11] (#157,827), YODA [12], and Vivli [13] prior to the initiation of this work and the completion of SEAVUE.

Data access

In June 2019 we queried clinicaltrials.gov to identify studies for meta-analysis (Supplementary Fig. 1,

Additional File 2, Table 1). Our inclusion criteria were completed, phase 2–4, randomized, double-blinded, interventional trials of FDA-approved treatments for Crohn's disease that capture the Crohn's Disease Activity Index (CDAI) at week eight relative to treatment initiation. Our initial search led to 90 candidate studies. Following manual review, we confirmed 9 studies as meeting these criteria ($N = 3392$). We requested and were granted access to all IPD from these studies. We manually reviewed their major inclusion and exclusion criteria to ensure comparability, and confirmed that they were at low risk of bias using the Cochrane ROB2 tool.

Study design

We designed this study to emulate a hypothetical head-to-head, parallel-design, efficacy trial randomizing participants to two treatment arms (Fig. 1). Although a typical meta-analytic study design would have involved pooling cohorts from several internally controlled, parallel-arm trials, this was not possible in our case. Many of the included studies involved open-label induction followed by a randomization event to continue or discontinue the treatment (Fig. 2). We noted that three of the nine trials did not include a parallel arm placebo cohort randomized at week 0 and followed for eight weeks. Thus, for our study, they were considered uncontrolled. As a first step towards developing a method for handling this heterogeneity, we restricted our initial analyses to just the placebo-controlled trials (six trials; $N = 3153$; Table 1). Subsequent analyses used a second set of placebo-less trials of adalimumab, one of the drugs compared in SEAVUE.

Quality control, harmonization

We performed extensive tests of data quality (Supplementary Methods, Additional file 1). These included reproducing published results from each trial cohort (Supplementary Figs. 3, 4, Additional File 2). We used domain knowledge to select nine variables that were universally available across trials for subsequent modeling: Age, Sex, body mass index (BMI), baseline CDAI, c-reactive protein (CRP), history of tumor necrosis factor-alpha inhibitor (TNFi) use, steroid use, immunomodulator use, and ileal involvement.

3% of the participants had at least one missing covariate at baseline. Continuous variables were addressed by median imputation, and participants with missing categorical variables were dropped ($N = 86$). 11% of participants had a missing outcome at week eight. We used last-observation-carried-forward to impute these. This is the typical practice for the analysis of these
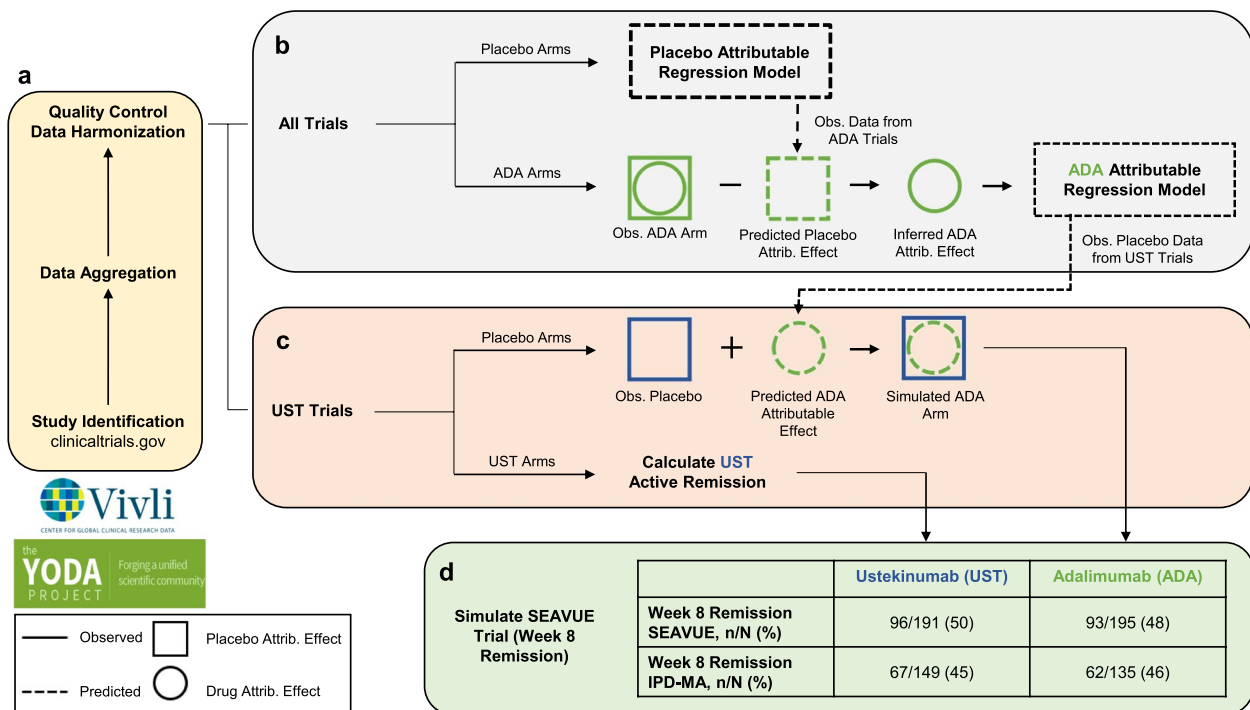
Fig. 1 Study overview. **A** Clinical trials were found using clinicaltrials.gov and sought for retrieval on the YODA and Vivli platforms. Individual participant data (IPD) from trials that collected CDAI scores at week 8 visits were then aggregated and harmonized. **B** Two linear mixed effect models—placebo-attributable and ADA-attributable—were developed from the harmonized data to partition the CDAI reduction based on baseline covariates (age, sex, BMI, etc.). Disease activity reduction was partitioned into placebo attributable (square) and drug-attributable (circle) effects. IPD (solid lines) were used to predict or simulate data (dashed lines). **C** Using the adalimumab (ADA) attributable model, we simulated the outcomes of the placebo group from the ustekinumab (UST) trials under a counterfactual scenario where they had instead been assigned to receive ADA. **D** Results from a simulated head-to-head trial were compared against a recently completed head-to-head trial, SEAVUE, to externally validate the proposed method
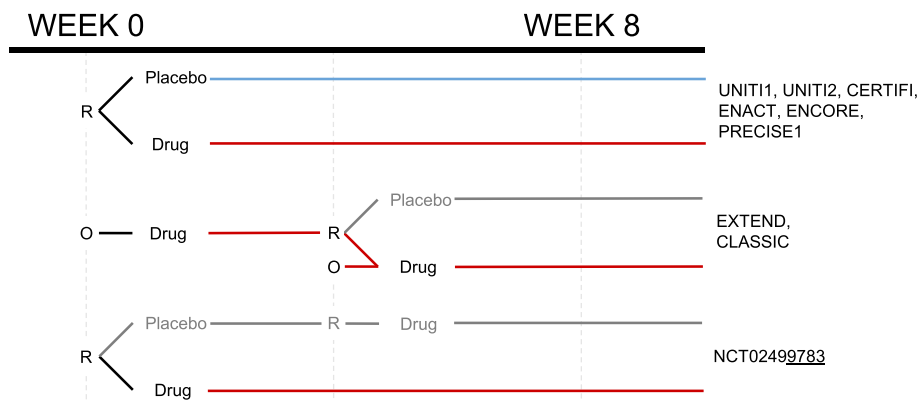
Fig. 2 Summary of randomized controlled trial study designs. Data harmonization required careful understanding of the study designs. All treatment arms that involved 8 weeks of consistent exposure to either placebo or (blue) or active treatment at the FDA-approved doses (red) were included. R = randomized and blinded; O = open label

trials in regulatory submissions and was the prespecified approach in the protocols of the included trials.

Some important variables could not be included in this study. Ethnicity was not collected in most trials. Race was missing in some trials, but when it was captured, it reflected significant imbalance (88% of participants were white). Other variables like disease behavior and duration were not uniformly captured across studies.

Rudrapatna *et al. BMC Medical Research Methodology* (2023) 23:218

Page 4 of 12

Table 1 Characterization of baseline covariates of included studies and simulated head-to-head trial. Placebo arms from the CLASSIC, EXTEND, and NCT02499783 studies were not included due to the absence of an 8-week parallel arm placebo group (see Fig. 2). CRP = c-reactive protein, TNFi = tumor necrosis factor inhibitor

Drug	Certolizumab-Pegol (CZP)	Natalizumab (NTZ)		Ustekinumab (UST)			Adalimumab (ADA)			Simulated Head-to-Head	
Trial Alias	PRECISE1	ENACT	ENCORE	CERTIFI	UNITI1	UNITI2	CLASSICI/II	EXTEND	NCT02499783	Observed Ustekinumab (UST)	Simulated Adalimumab (ADA)
NCT	NCT00152490	NCT00032786	NCT00078611	NCT00771667	NCT01369329	NCT01369342	NCT00055523 NCT00055497	NCT00348283	NCT02499783		
Year	2003	2001	2004	2008	2011	2011	2002	2006	2015		
Sample size	$N = 603$	$N = 879$	$N = 480$	$N = 252$	$N = 523$	$N = 416$	$N = 73$	$N = 64$	$N = 102$	$N = 149$	$N = 135$
Treatment Group											
Active	304 (50%)	701 (80%)	243 (51%)	126 (50%)	260 (50%)	209 (50%)	73 (100%)	64 (100%)	102 (100%)	149 (100%)	135 (100%)
Placebo	299 (50%)	178 (20%)	237 (49%)	126 (50%)	263 (50%)	207 (50%)	-	-	-	-	-
Age—Mean (SD)	37 (±12)	38 (±13)	38 (±13)	39 (±13)	38 (±12)	39 (±13)	38 (±11)	37 (±11)	33 (±10)	39 (±14)	39 (±12)
Sex: Female—N (%)	337 (56%)	502 (57%)	284 (59%)	145 (58%)	288 (55%)	227 (55%)	38 (52%)	40 (62%)	35 (34%)	82 (55%)	71 (53%)
BMI—Mean (SD)	24 (±5·3)	25 (±5·6)	25 (±6·4)	26 (±7·3)	22 (±0·58)	23 (±0·68)	26 (±6·0)	25 (±4·6)	19 (±2·7)	23 (±0·73)	22 (±0·50)
Baseline CDAI—Mean (SD)	300 (±61)	300 (±60)	300 (±61)	320 (±67)	320 (±60)	300 (±56)	290 (±52)	320 (±69)	270 (±48)	290 (±54)	290 (±53)
CRP (mg/L)—Mean (SD)	18 (±25)	20 (±29)	22 (±23)	21 (±28)	17 (±23)	16 (±20)	13 (±17)	20 (±21)	24 (±25)	18 (±24)	14 (±15)
History of TNFi Use—N (%)	163 (27%)	351 (40%)	223 (46%)	252 (100%)	520 (99%)	135 (32%)	2 (3%)	31 (48%)	0 (0%)	0 (0%)	0 (0%)
Steroid Use—N (%)	235 (39%)	348 (40%)	190 (40%)	134 (53%)	253 (48%)	172 (41%)	21 (29%)	6 (9%)	31 (30%)	68 (46%)	45 (33%)
Immunomodulator Use—N (%)	238 (39%)	302 (34%)	180 (38%)	66 (26%)	172 (33%)	155 (37%)	20 (27%)	26 (41%)	61 (60%)	60 (40%)	57 (42%)
Ileal Disease—N (%)	432 (72%)	678 (77%)	355 (74%)	182 (72%)	420 (80%)	335 (81%)	47 (64%)	48 (75%)	82 (80%)	120 (81%)	112 (83%)

Table 1 (continued)

Drug	Certolizumab-Pegol (CZP)	Natalizumab (NTZ)		Ustekinumab (UST)			Adalimumab (ADA)			Simulated Head-to-Head	
Trial Alias	PRECISE1	ENACT	ENCORE	CERTIFI	UNITI1	UNITI2	CLASSICI/II	EXTEND	NCT02499783	Observed Ustekinumab (UST)	Simulated Adalimumab (ADA)
NCT	NCT00152490	NCT00032786	NCT00078611	NCT00771667	NCT01369329	NCT01369342	NCT00055523 NCT00055497	NCT00348283	NCT02499783		
Year	2003	2001	2004	2008	2011	2011	2002	2006	2015		
Sample size	N=603	N=879	N=480	N=252	N=523	N=416	N=73	N=64	N=102	N=149	N=135
CDAI Reduction—Mean (SD)	67 (±93)	99 (±101)	92 (±98)	65 (±100)	51 (±97)	94 (±101)	114 (±104)	136 (±100)	123 (±82)	110 (±100)	140 (±100)

Modeling assumptions

We incorporated several assumptions when developing and interpreting candidate models. We assumed that the observed week eight reduction in CDAI reflected a combination of two distinct effects: a drug-independent (i.e., placebo) effect and drug-attributable effect. These effects were separately modeled as a function of the above predictors and study year. The justification for this is briefly summarized below and additionally presented graphically (Fig. 3).

The placebo effect was modeled as a function of the nine covariates as well as predictors of trial-specific heterogeneity. We assumed that much of the spontaneous improvement seen in placebo-assigned participants was related to regression to the mean, as study participation was limited to patients with currently active Crohn's disease. Conversely, we assumed that failure to spontaneously improve was likely to reflect chronic and cumulative disease burden with relative stability in symptoms. Thus, variables corresponding to concomitant and prior treatments were treated as proxies of chronic disease burden and included as predictors. Lastly, we considered other influences on overall heterogeneity, including differences in cohorts, data capture, outcome ascertainment, and study personnel. To account for these sources of variation, we included study year as well as trial identifier as additional covariates. In mixed-effect models, trial was included as a random effect. Other covariates were fixed effects.

The drug-attributable effect was separately modeled as a function of these same covariates, reflecting drug-specific (interaction) effects on the outcome. Many of these covariates are well-established as modifiers of treatment response, such as a history of TNFi and immunomodulators use [4]. Others (CRP, baseline CDAI) are proxies of bowel inflammation, the target of these medications. These variables were included to maximize the explained variation in the outcome.

Development and assessment of a placebo model

We fit a linear mixed effects model utilizing all nine predictors as well as study year as predictors of the placebo effect. To minimize the risks of residual bias due to model misspecification (e.g., non-linearities, unmodeled interactions), we compared the predictive performance of this model against other statistical and machine learning models. We further evaluated this model from the perspective of being used to impute unmeasured placebo effects, and thus normalize different trials to the same background. We performed a leave-one-trial-out analysis and inspected the trial-averaged residuals.

Estimation of the drug-attributable effect

To normalize the responses of drug-assigned cohorts that lacked a within-study, parallel-arm control group, we used the finalized placebo model to simulate and subsequently partition their overall response into drug-independent and drug-dependent (i.e., drug attributable) components (Fig. 1b). We performed this using the data from the adalimumab trials ($N = 239$) because they were all lacking an 8-week continuous placebo group and thus required normalization. From the outcomes of these patients, we subtracted the conditional mean outcomes associated with the placebo effect and used the residuals as the new outcome variable of a second mixed-effects model to estimate the adalimumab-attributable effect.

Validation, sensitivity analyses

Using the covariates associated with the placebo recipients of the ustekinumab trials, we used the adalimumab-attributable regression model to simulate their counterfactual week 8 outcomes had they received adalimumab instead (Fig. 1c). We identified the subset of these virtual patients who were naïve to TNFi (an additional inclusion criterion from SEAVUE) and did the same with the ustekinumab recipients. We compared their week 8 outcomes using the same definition of clinical remission as used in SEAVUE (CDAI < 150) and performed a Fisher's exact test to compare our results with SEAVUE's. We tested the robustness of our result using three sensitivity analyses: 1) removing ENACT and ENCORE from the dataset due to > 10% missingness of outcome data, 2) removing participants with missing outcomes, and 3) removing the ustekinumab trials from the placebo model training data, to address potential information leakage. Lastly, we compared our results with what we might have found had we not used our method to normalize cohorts.

Results

See Fig. 1 for an overview. This method was originally developed in the context of an existing effort to study comparative efficacy in Crohn's disease by reanalyzing the IPD of corresponding clinical trials. As the first step towards this goal, we sought to address the problem of meta-analyzing data from several potentially heterogeneous trials lacking a common control group.

Data access

We queried clinicaltrials.gov and performed manual review to confirm 16 trials as meeting these criteria: completed, phase 2–4, randomized, double-blinded, interventional trials of FDA-approved treatments for Crohn's disease as of June 2019 (Fig. 1a, Supplementary Fig. 1, Additional file 2). Included trials had common inclusion/exclusion criteria or had participant-level data available
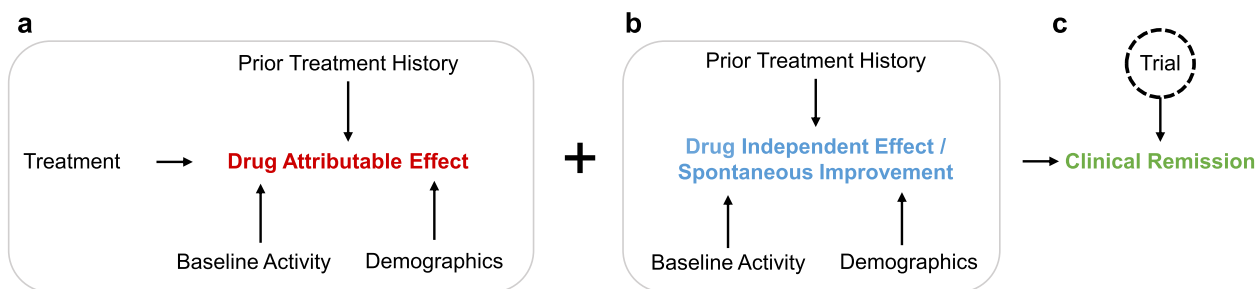
Fig. 3 A directed acyclic graph of the modeling strategy. **A** A directed acyclic graph (DAG) of the drug attributable effect. In addition to the active treatment itself, patient demographics (age, sex, BMI), baseline Crohn's disease activity (baseline CDAI, CRP, location), and treatment history (prior use of TNFis, current use of oral corticosteroids and immunomodulators) are all modelled as contributing to the drug attributable effect. The non-drug covariates are effect modifiers and are implicitly modeled as two-way interaction terms with the active drug. **B** A DAG of the drug independent effect (i.e., placebo effect). The same covariates except for the treatment term are modeled as effect modifiers and are implicitly represented as two-way interactions with the placebo effect. **C** Drug attributable and drug independent effects have additive effects on the overall clinical remission at week 8 (CDAI < 150), with any individual trial reflecting a noisy measurement of the true effect due to unmodeled heterogeneity in study design and execution (random effect)

to control for this heterogeneity (Supplementary Table 1, Additional File 2). They all measured the same endpoint (CDAI) at week eight and were at low risk of bias (Supplementary Fig. 2, Additional File 2). We obtained access to the IPD for 15 studies ($N=5703$), corresponding to trials of all six FDA-approved biologics as of 2019.

Development and assessment of a placebo model
We fit a linear mixed effects model utilizing nine clinical features and study year as predictors of the placebo effect (Fig. 1b, Table 2). To minimize the risk of residual bias due to model misspecification, we compared the predictive performance of this model against other machine learning models (Supplementary Table 2, Additional file 2). We found no significant differences in the root-mean-squared-error. Thus, we selected the mixed-effects model for downstream analyses.

We evaluated this model from the perspective of being used to impute unmeasured placebo effects, and thus normalize different trials to the same background placebo response. A leave-one-trial-out analysis suggested that the model predictions were robust and unbiased (Supplementary Figs. 4, 5, Additional file 2). The trial-averaged residuals were consistent with normality ($p=0.4$; Shapiro–Wilk test).

Table 2 Mixed effect linear regression outputs for the placebo attributable ($n=1310$) and ADA attributable ($n=239$) models. For training, Year was centered by subtracting 2000, Baseline CDAI was centered by subtracting 300, Age was centered by subtracting 35, BMI was centered by subtracting 20, and CRP (mg/L) was centered by subtracting 10. A. The placebo attributable model (intraclass correlation coefficient 0.02) trial random intercepts were found to be -12.808 (PRECISE1), -7.975 (UNITI1), -6.328 (CERTIFI), 6.077 (ENACT), 8.669 (ENCORE), and 12.366 (UNITI2). B. ADA attributable model (intraclass correlation coefficient 0.05) trial random intercepts were found to be -20.215 (CLASSIC), 9.439 (EXTEND), and 10.775 (NCT02499783)

Predictors	Placebo Attributable Model			ADA Attributable Model		
	Estimates	SE	*p*	Estimates	SE	*p*
Intercept	92.18	13.12	**< 0.001**	67.57	20.41	**0.001**
Year (Centered)	-1.89	1.55	0.223	-	-	-
Baseline CDAI (Centered)	0.37	0.04	**< 0.001**	0.12	0.11	0.276
Age (Centered)	0.13	0.21	0.549	-1.77	0.57	**0.002**
BMI (Centered)	0.52	0.54	0.341	-1.45	1.34	0.28
CRP (mg/L) (Centered)	-0.28	0.11	**0.012**	-0.35	0.28	0.213
Sex: Male	-0.72	5.17	0.889	7.14	12.13	0.556
History of TNFi Use	-37.56	6.32	**< 0.001**	7.51	20.38	0.712
Steroid Use	7.15	5.24	0.172	5.36	14.35	0.709
Immunomodulator Use	5.42	5.45	0.32	-17.43	12.54	0.164
Ileal Disease	-7.46	5.99	0.213	2.84	13.74	0.836

Rudrapatna *et al. BMC Medical Research Methodology*   (2023) 23:218

Page 8 of 12

We noted that the unmodeled variation in the placebo effect was relatively large and was independent of the choice of model (Supplementary Table 2, Additional file 2). These results explain the large placebo effects that have been seen in Crohn's disease randomized trials (regression to the mean) and suggest that more work will be needed to improve the measurement of Crohn's disease activity.

To study the placebo effect and identify potential opportunities to improve trial efficiency, we reviewed all significant predictors. A history of TNFi was associated with a 38-point reduction in the placebo effect. We interpreted this as reflecting a greater cumulative disease burden in patients who failed to improve with TNFis, with disease complications (e.g., minor intestinal strictures) that are unlikely to spontaneously regress over 8 weeks. Similarly, CRP was a negative predictor, suggesting that untreated acute inflammation is unlikely to improve over short time periods. The baseline CDAI was a positive predictor, likely reflecting regression to the mean effects. Age, sex, BMI, concomitant medications, and ileal involvement were not found significant, potentially due to multicollinearity.

Estimation of the drug-attributable effect

We sought to normalize the responses of drug-assigned cohorts that lacked a within-study, parallel-arm control group. Our strategy was to use the finalized placebo model to partition the overall response into drug-independent and drug-attributable components (Fig. 1b). We applied this approach to the data from three study cohorts assigned to receive adalimumab at the FDA-approved dose for treatment induction ($N=239$; Table 1). We selected this medication because it is one of the two treatments that were compared against each other in SEAVUE, the target of our emulation and validation efforts.

We used the coefficients of the fitted placebo model to predict and remove the placebo-attributable component from the observed outcomes of these participants. The residuals from this process were interpreted as reflecting the adalimumab-attributable effect (Fig. 1b). Across these patients the mean drug-attributable CDAI reduction was 68 points. We used these residuals to fit a second model for the adalimumab-attributable effect (Table 2).

As an exploratory analysis we reviewed the significant predictors of a response to adalimumab and compared these to the corresponding results from the placebo model. Although the sample size was relatively small, we noted a strong signal for age as a negative predictor: additional decades of life were associated with an 18-point reduction in the response to adalimumab. Interestingly, the direction of this effect was the opposite of that seen

in the placebo-only model, suggesting that this coefficient might not have been identified as significant had it not been handled as an interaction term as we did.

External validation

To validate our method, we designed an *in-silico* study to emulate SEAVUE, the only head-to-head study of FDA-approved biologics for Crohn's disease to date [3]. In SEAVUE, biologic-naive patients with active Crohn's disease were randomly to receive either adalimumab or ustekinumab as treatment. The primary endpoint was clinical remission at week 52, defined as a CDAI less than 150. Secondary endpoints included clinical remission at the time of all study visits, including week eight.

We identified all participants from the three ustekinumab-related trials who were biologic-naive. We identified 149 subjects who were assigned to ustekinumab and 135 participants assigned to placebo. We noted that the observed responses of the 135 placebo recipients reflected a combination of individual-specific variability and trial-specific variability (Fig. 3). We therefore reasoned that to simulate the effect of treatment assignment, we needed to 'add back' the conditional mean effect associated with adalimumab to the outcomes of the placebo recipients (Fig. 1c). Using the model coefficients identified in the adalimumab-attributable regression model (Table 2), we computed and added this extra reduction in the CDAI to the observed week eight outcomes of the placebo cohort.

Finally, we computed the proportion of patients who were in clinical remission at week eight, comparing the results of the observed ustekinumab recipients with that of the patients simulated to have received adalimumab and subject to the same background placebo effect (Fig. 1d). We found that ustekinumab and adalimumab appeared to be equally efficacious, with 45% and 46% of the cohorts in remission. This result closely matched that of SEAVUE ($p=0.9$), which found 50% and 48% of these corresponding cohorts in remission (Table 3). Our simulated trial was similar in sample size to SEAVUE, with 149 and 135 patients receiving ustekinumab and adalimumab in our study, compared to 191 and 195 in SEAVUE.

We tested the robustness of this result using three sensitivity analyses. In the first we removed two trials (PRECISE1, ENACT) associated with the greatest degree of outcome missing data (Supplementary Fig. 3, Additional File 2). In the second, we performed a complete case analysis (deleted patient data associated with missing outcomes) as an alternative to last-observation-carried-forward imputation. In the third we removed all participant data emanating from an ustekinumab trial from the placebo training data, to address a possibility of information leakage. Our results remained

Table 3 Comparison of clinical remission rates at week 8 for the TNF-naive ustekinumab (UST) cohort and TNF-naive adalimumab (ADA) cohort for the SEAVUE study, our primary analysis (simulation of SEAVUE), sensitivity analyses, and negative control. Because missing week 8 CDAI values were highest for trials PRECISE1 and ENACT, their participant-level data was removed (N = 1482) from the first sensitivity analysis to account for potential bias. In the complete case sensitivity analysis, all participants with missing week 8 CDAI values (N = 361) were removed. In the information leakage sensitivity analysis, participants from an ustekinumab study (N = 1191) were removed from training the placebo-attributable model to avoid potential information leakage when simulating the adalimumab (ADA) arm (Fig. 1c). The negative control summarizes the clinical remission rates at week 8 for TNF-naive participants from the adalimumab studies without applying our regression-based correction method. The final column corresponds to the results of null hypothesis testing, that of no statistically significant difference between each simulated result and the published SEAVUE results

| | Ustekinumab (UST) | Adalimumab (ADA) | *p*-value |
	n/N (%)	n/N (%)	
Published SEAVUE Results	96/191 (50.3)	93/195 (47.7)	-
Predicted SEAVUE Results, Primary Analysis	67/149 (44.9)	62/135 (45.9)	0.9
Sensitivity Analysis: Trials with High Capture of Outcomes	67/149 (44.9)	60/135 (44.4)	1
Sensitivity Analysis: Complete Cases	66/148 (44.6)	65/128 (50.3)	0.39
Sensitivity Analysis: Information Leakage	67/149 (44.9)	67/135 (49.6)	0.47
Negative Control, No Normalization	67/149 (44.9)	119/239 (49.8)	0.4

unchanged over all sensitivity analyses (Table 3), supporting the robustness of our primary findings as well as the validity of our overall methodology.

Finally, we sought to evaluate the value of using our modeling approach compared to a simpler approach using published trial results. One barrier we noted to the latter was that the aggregated response of the TNFi-naive subcohorts at week eight was only published in one out of the six trials that we included for this comparison of ustekinumab and adalimumab, making it impossible to emulate SEAVUE using this approach. Separate from this, and to specifically evaluate the value of normalizing disparate cohorts using placebo models, we simulated the potential results of our head-to-head assessment without a normalization step. Under this scenario, the unnormalized adalimumab cohort in clinical remission was 50% (Table 3). While this was not statistically significant compared to the observed ustekinumab arm (45%; p = 0.4), it reflects a trend towards a difference. We interpreted this as reflecting a degree of bias that could plausibly result in false positives in other similar studies, but one that is analytically controllable using our method.

Discussion
We developed a new method for meta-analyzing individual participant data (IPD) from heterogenous randomized trials lacking a shared control group. We validated our methodology by successfully reproducing a major endpoint of SEAVUE, a recent head-to-head trial of biologic therapies in Crohn's disease [3]. Our method involved several steps (Fig. 1):

- Identifying and isolating parallel arm cohorts from the available trials
- Harmonization and quality control
- Separately modeling the placebo effect from drug-attributable effects
- Sequentially partitioning and assembling different sources of variation to accurately simulate the outcomes of a suitably normalized comparator group.

After decades of calls for greater data sharing [14–16] we are now seeing many new platforms for accessing clinical trials data. The availability of these data has opened opportunities for researchers to verify published results as well as answer new questions using these data. This has never been more important, with the cost of new phase 3 clinical trials current at $20 M and climbing [17].

Although the growing availability of IPD portends well for the future of research, it has revealed new analytical challenges that require new methods. Existing methods for conducting IPD meta-analyses typically involve including trials with near-identical study designs, including fully parallel-design cohorts and shared placebo comparator arms. When these criteria are not met, problematic trials are often excluded from a given meta-analysis, sometimes in subtle ways. This substantially limits the numbers of questions that might already be answerable using existing clinical data. In some cases, this common practice might even introduce bias.

This work suggests that there may be better ways to handle this heterogeneity and discover new and trustworthy signals from these data. This method as well as extensions therein may substantially increase the numbers of studies that can be done, uncovering new evidence on

comparative efficacy, safety, and ultimately precision medicine. Taking the example of Crohn's disease, a major motivation for conducting the SEAVUE trial is the current level of uncertainty regarding the comparative efficacy of already approved treatments. While the use of causal inference methods to emulate randomized trials from observational data are receiving great attention, our approach leverages already randomized data and is much less susceptible to bias. Thus, methods such as what we propose here can address these gaps, particularly as more therapies are approved and thus the number of potential head-to-head comparisons grows exponentially.

While we have illustrated this methodology in a comparative efficacy analysis, this approach may have significant value in other contexts. Models for the placebo effect, such as we demonstrate here, may help improve the design and statistical power of clinical trials across diseases. Moreover, the use of cohort normalization methods may be useful to improve the robustness of external control arm studies. These are studies that typically utilize real-world data to draw indirect inferences against controlled cohorts, typically single-arm intervention studies. However, our analysis suggests that a major driver of the large placebo effects in Crohn's disease is the large unmodeled variation in the CDAI. Future work is needed to improve the measurement of Crohn's disease activity.

We acknowledge several limitations. First, although we undertook extensive efforts to harmonize the data, we could not perfectly reproduce all covariate statistics as published. It is likely that we could have overcome these issues with access to the original analytical code. Nonetheless, the degree of deviations from published results was small, and our primary results remained robust to many sensitivity analyses. Future efforts involving pre-harmonization to a common data model may improve the reproducibility and feasibility of these IPD meta-analyses. Second, we were unable to include many important covariates like race and ethnicity. Most included studies did not capture ethnicity. Some studies did capture race but showed evidence of significant skew towards white participants. This likely reflects the historical under-recognition of the importance of these factors.

Closely related to this point is potentially important role of variable selection and model selection. Our methodology rests on the ability to accurately model and neutralize interstudy heterogeneity with captured covariates. This itself is a function of what covariates were captured, which is subject to clinical knowledge about a given disease at the time that different trials were conducted. It is also a function of the model form. Future studies are needed to explore the sensitivity of this approach to unmodeled effect modifiers and model misspecification.

Lastly, we note that our validation was somewhat underpowered and was performed in the context of just one disease. This is largely a function of the relative rarity of clinical trials (the source of our data and sample size), and especially head-to-head trials like SEAVUE. This underscores the importance of methods for learning more from these small but high-quality data. Future studies are needed to confirm the robustness and generalizability of our methodology to other diseases.

Conclusion

In conclusion, we developed a new method for meta-analyzing data from heterogeneous trials lacking a common control group. We validated this method by reproducing the results of a recent comparative efficacy trial using pre-existing data. We are sharing our code for others to replicate and build upon these methods, and ultimately uncover new insights using the data we already have.

Author's contributors

VAR conceived the study and obtained access to the data. SW, VGR, DVA, and VAR designed the study, analyzed the data, and drafted the manuscript. AM performed the risk of bias assessment. SW performed an independent review of the analytical code. VAR, SW, VGR, DVA, AM, and AJB interpreted the data and critically edited the manuscript.

Abbreviations

ADA	Adalimumab
BMI	Body Mass Index
CDAI	Crohn's Disease Activity Index
CRP	C-Reactive Protein
CZP	Certolizumab Pegol
IPD	Individual Participant Data
NTZ	Natalizumab
TNFi	Tumor Necrosis Factor inhibitor
UST	Ustekinumab

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12874-023-02020-5.

Additional file 1: Supplementary methods.

Additional file 2: Figure S1. PRISMA-IPD flow diagram. **Figure S2.** Risk of bias. **Figure S3.** Reproducibility of published data. **Figure S4.** Leave-one-trial-out analysis. **Figure S5.** Checking model assumptions. **Table S1.** Major inclusion/exclusion criteria of included studies. **Table S2.** Model selection.

Authors' contributions
VAR conceived the study and obtained access to the data. SW, VGR, DVA, and VAR designed the study, analyzed the data, and drafted the manuscript. AM performed the risk of bias assessment. SW performed an independent review of the analytical code. VAR, SW, VGR, DVA, AM, and AJB interpreted the data and critically edited the manuscript.

Funding

Availability of data and materials
The data that support the findings of this study are available from YODA and Vivli, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of YODA and Vivli. The analytical R code for data processing and modeling can be found in the GitHub repository, https://github.com/ibd-ipd-ma/SequentialRegressionSimulation.

Declarations

Ethics approval and consent to participate
This study was approved in accordance with the Declaration of Helsinki and with the oversight of the UCSF IRB (#18–24588). This review board waived the need for participant consent given that the underlying data had already been deidentified by the owners of the raw data.

Consent for publication
Not applicable.

Competing interests
VAR received grant support from Janssen Inc and Alnylam Inc for unrelated work during this study. DVA is currently an employee at Bristol Myers Squibb. AJB is a co-founder and consultant to Personalis and NuMedii; consultant to Mango Tree Corporation, and in the recent past, Samsung, 10✕Genomics, Helix, Pathway Genomics, and Verinata (Illumina); has served on paid advisory panels or boards for Geisinger Health, Regenstrief Institute, Gerson Lehman Group, AlphaSights, Covance, Novartis, Genentech, and Merck, and Roche; is a shareholder in Personalis and NuMedii; is a minor shareholder in Apple, Meta (Facebook), Alphabet (Google), Microsoft, Amazon, Snap, 10✕Genomics, Illumina, Regeneron, Sanofi, Pfizer, Royalty Pharma, Moderna, Sutro, Doximity, BioNtech, Invitae, Pacific Biosciences, Editas Medicine, Nuna Health, Assay Depot, and Vet24seven, and several other non-health related companies and mutual funds; and has received honoraria and travel reimbursement for

invited talks from Johnson and Johnson, Roche, Genentech, Pfizer, Merck, Lilly, Takeda, Varian, Mars, Siemens, Optum, Abbott, Celgene, AstraZeneca, AbbVie, Westat, and many academic institutions, medical or disease specific foundations and associations, and health systems. Atul Butte receives royalty payments through Stanford University, for several patents and other disclosures licensed to NuMedii and Personalis. Atul Butte's research has been funded by NIH, Peraton (as the prime on an NIH contract), Genentech, Johnson and Johnson, FDA, Robert Wood Johnson Foundation, Leon Lowenstein Foundation, Intervalien Foundation, Priscilla Chan and Mark Zuckerberg, the Barbara and Gerson Bakar Foundation, and in the recent past, the March of Dimes, Juvenile Diabetes Research Foundation, California Governor's Office of Planning and Research, California Institute for Regenerative Medicine, L'Oreal, and Progenity. The authors have declared that no actual competing interests exist.

Author details
[1]Division of Gastroenterology, Department of Medicine, University of California, San Francisco, San Francisco, CA, USA. [2]Bakar Computational Health Sciences Institute, University of California, San Francisco, San Francisco, CA, USA. [3]School of Medicine, University of California, San Francisco, San Francisco, CA, USA. [4]Department of Mathematics and Statistics, University of San Francisco, San Francisco, CA, USA.

References
1. Chalmers I. The Cochrane collaboration: preparing, maintaining, and disseminating systematic reviews of the effects of health care. Ann N Y Acad Sci. 1993; 703: 156–63; discussion 163–5.
2. Tierney JF, Vale C, Riley R, et al. Individual Participant Data (IPD) Meta-analyses of randomised controlled trials: guidance on their use. PLoS Med. 2015;12: e1001855.
3. Schmitz S, Adams R, Walsh C. Incorporating data from various trial designs into a mixed treatment comparison model. Stat Med. 2013;32:2935–49.
4. Ibrahim JG, Chen M-H. Power prior distributions for regression models. Stat Sci. 2000;15:46–60.
5. Zhang J, Ko C-W, Nie L, Chen Y, Tiwari R. Bayesian hierarchical methods for meta-analysis combining randomized-controlled and single-arm studies. Stat Methods Med Res. 2019;28:1293–310.
6. Thompson L, Chu J, Xu J, Li X, Nair R, Tiwari R. Dynamic borrowing from a single prior data source using the conditional power prior. J Biopharm Stat. 2021;31:403–24.
7. Wang C, Li H, Chen W-C, et al. Propensity score-integrated power prior approach for incorporating real-world evidence in single-arm clinical studies. J Biopharm Stat. 2019;29:731–48.
8. Li X, Song Y. Target population statistical inference with data integration across multiple sources—an approach to mitigate information shortage in rare disease clinical trials. Stat Biopharm Res. 2020;12:322–33.
9. Li X, Miao W, Lu F, Zhou X. Improving efficiency of inference in clinical trials with external control data. Biometrics. 2021; biom.13583.
10. Sands BE, Irving PM, Hoops T, et al. Ustekinumab versus adalimumab for induction and maintenance therapy in biologic-naive patients with moderately to severely active Crohn's disease: a multicentre, randomised, double-blind, parallel-group, phase 3b trial. Lancet Lond Engl. 2022;399:2200–11.
11. PROSPERO | Comparative effectiveness of treatments for moderate-to-severe Crohn's disease: an individual participant data meta-analysis. https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=157827Accessed 28 Sept 2021.
12. The YODA Project | 2018–3476. Effic. Crohn's Dis. Treat. Stratif. Dis. Phenotype. https://yoda.yale.edu/data-request/2018-3476 . Accessed 28 Sept 2021.
13. Efficacy of Crohn's Disease Treatment Stratified by Disease Phenotype. Vivli. https://vivli.org/efficacy-of-crohns-disease-treatment-stratified-by-disease-phenotype-2/ . Accessed 28 Sept 2021.

14. Mello MM, Francer JK, Wilenzick M, Teden P, Bierer BE, Barnes M. Preparing for Responsible Sharing of Clinical Trial Data. N Engl J Med. 2013;369:1651–8.
15. Packer M. Data sharing in medical research. BMJ. 2018;360: k510.
16. Health Data Sharing Special Publication. Natl. Acad. Med. https://nam.edu/health-data-sharing-special-publication/ . Accessed 10 Aug 2022.
17. Examination of Clinical Trial Costs and Barriers for Drug Development. ASPE. https://aspe.hhs.gov/reports/examination-clinical-trial-costs-barriers-drug-development-0 . Accessed 12 June 2022.

Publisher's Note