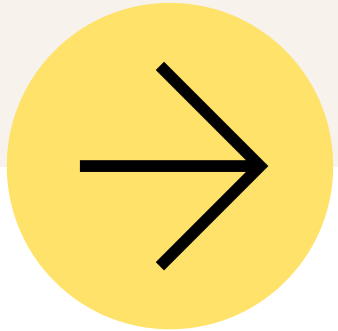
Graph-indexed sensitivity analysis

for residual non-comparability of external controls in Bayesian dynamic borrowing

worked throughout on ataluren in Duchenne muscular dystrophy



esylos

Since the abstract

The question is unchanged: what dynamic borrowing cannot see. So is the debt to Bayesian multiple-bias modelling. Three things have moved, deliberately.

01 Elicit arcs, not a bias per graph

The abstract gave each DAG a scalar bias prior. Here each graph carries two arcs a clinician can contradict, and B is derived from them.

02 One posterior per world view, not a mixture

The abstract proposed DAG-weighted, worst-case and uniform summaries. The data cannot rank the stories, so each is reported on its own.

03 One real case, not the simulation grid

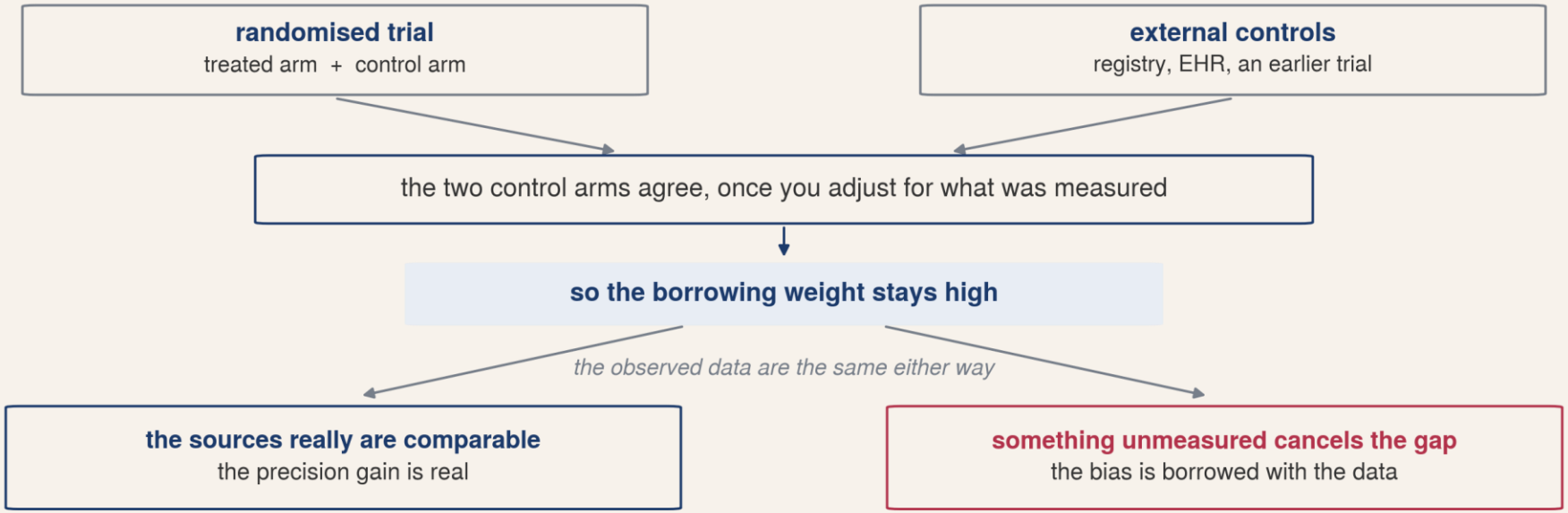
A single case, worked end to end, conveys the idea better than a grid of scenarios. The simulations are there for the discussion.

The five graphs are those of the abstract: no disagreement, confounding, drift, compounding and cancellation.

Bayesian multiple-bias modelling: Greenland 2005; McCandless, Gustafson & Levy 2007

Borrowing decides by agreement

A randomised trial with a small, noisy control arm, and an external control cohort that is available and looks similar. Dynamic borrowing compares the two control arms and borrows in proportion to how well they agree: a discrepancy pulls the weight down, agreement leaves it high and the external control arm is then treated as exchangeable.



Agreement is observable. Exchangeability is not.

esylos

Viele et al. 2014 · Galwey 2017

The case: ataluren in nonsense-mutation DMD

	Trial (Study 007, decline phase)	External control (ACT DMD)
treated	n = 32 −12.3 (SD 69.4)	—
control	n = 31 −62.2 (SD 84.5)	n = 114 −60.7 (SE 9.3)
entry criteria	steroids, 7–16 y, 6MWD ≥150 m, ≤80% predicted	the same
period	2008–10	2013–14

Study 007's decline-phase analysis (n = 63) was post hoc; here it is analysed as if pre-specified, to ask what borrowing adds.

Analysis	Effect (m)	95% interval
Randomised, trial only (reference)	49.9	[11.7, 88.1] CI
External control	48.4	[18.2, 78.6] CrI
Hybrid borrowing, E[w] = 0.56	49.1	[17.9, 80.2] CrI

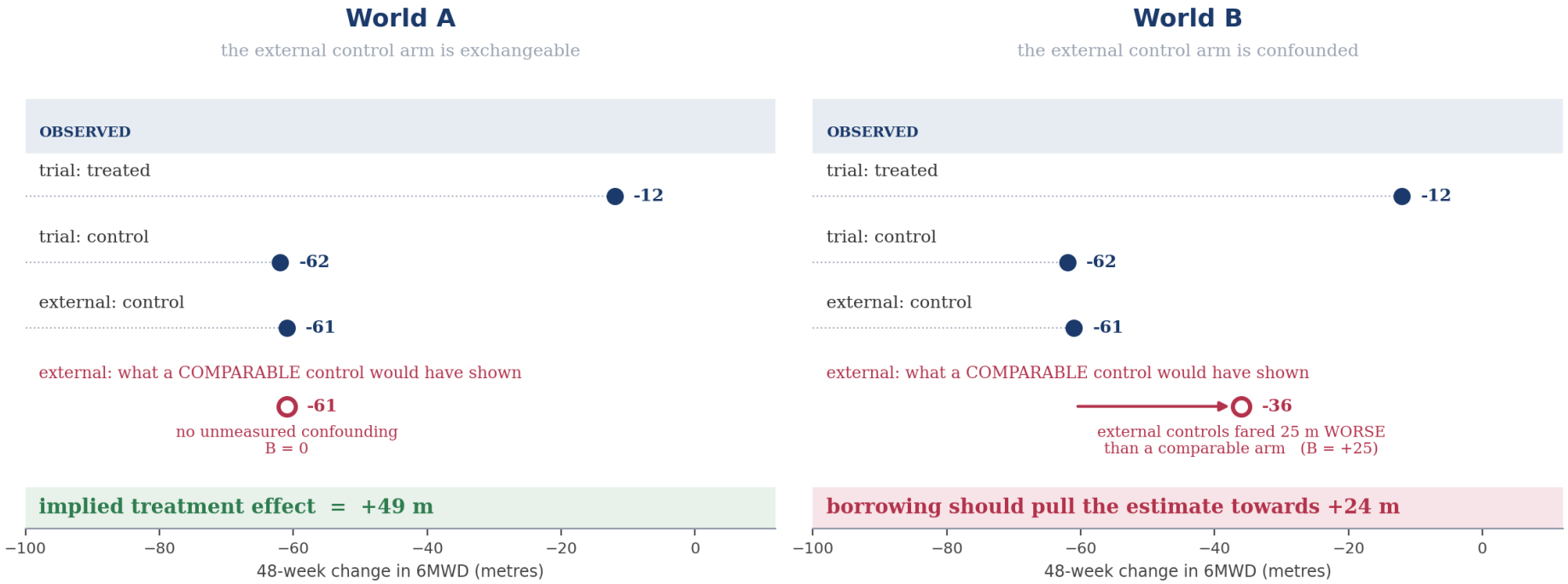
The randomised interval already excludes zero. Borrowing buys precision: the lower bound moves from 11.7 to 17.9 m, against a 30 m MCID, if the external control arm is exchangeable. Everything that follows is the last row.

esylos

Sources: Study 007 decline phase, EMA/CHMP/369266/2014 (Fig. 12); external control arm, McDonald et al. 2017

Agreement is not exchangeability

Both panels are the case you just saw: the same three observed means, the same 1.5 m agreement, the same E[w] = 0.56.



Nothing observable separates these two pictures. The external control arm either is comparable, or is confounded by roughly 25 m.

One reading gives 49 m; in the other, borrowing should pull the estimate towards 24 m. B absorbs any residual non-comparability of the two control arms, not confounding alone.

esylos

Why the data cannot choose between them

The external source contributes CONTROL patients only, adjusted for the same covariates: baseline 6MWD, age and steroid use.

$$Y_i = \alpha + \theta T_i + X_i^\top \beta - B\,\mathbf{1}\{i \in E\} + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2)$$

θ = E[Y(1) − Y(0) | randomised population], the treatment effect in the trial. B is the residual displacement of the external control mean after adjustment for X.

B is not estimable.

It enters only through the external controls, whose mean is $\alpha + X^\top\beta - B$. Admit a free external-arm intercept and the two are perfectly confounded, so whatever the data appear to say about B comes from the trial's own intercept — and estimating it removes exactly the precision gain borrowing exists to provide.

esylos

What we do instead

No new borrowing model: the machinery is a standard power prior. What changes is the object we elicit, and what the likelihood is allowed to see.

01 Elicit the graph, not the bias

How strongly an unmeasured cause moves membership of the external control arm, and how many metres of 6MWD it moves. B is derived from the two arcs, never asserted.

02 Cut the displacement

The data update the borrowing weight. They never update B, because they cannot identify it: a free external-arm intercept absorbs it exactly.

03 Report, do not average

One posterior per world view, and the arc strength each conclusion would need in order to fail. No posterior model probabilities, no mixture.

Power priors, MAP and commensurate priors decide how much to borrow; tipping-point analyses ask how large a bias would have to be. Neither asks by what mechanism — and the likelihood cannot rank mechanisms.

esylos

Power prior: Ibrahim & Chen 2000 · MAP: Schmidli et al. 2014 · commensurate: Hobbs et al. 2011 · tipping point / E-value: VanderWeele & Ding 2017

How we would elicit the arcs

Take one of the five: standard of care drifted between the 2008–10 trial and the 2013–14 external control arm. We would put two questions to the clinicians. Both name that mechanism, and neither asks about bias.

QUESTION 1 *"Boys on the care pathway that only arrived after 2010 — how much more likely are they to turn up in the external control arm than in the trial?"*

We assume: strongly. specified as $\gamma \sim N(2.20, 0.50^2)$: about nine times the odds per SD of the unrecorded factor

---

QUESTION 2 *"Two boys alike in baseline 6MWD, age and steroid use, one on that pathway and one not. Over 48 weeks, how many metres separate them?"*

We assume: about 10 m, in the external arm's favour. specified as $\lambda \sim N(-10, 4^2)$, the only direction we assert

---

NOT ASKED *"By how many metres is the external control arm out?"*

B = −12.7 m (sd 5.1). derived from the two arcs above, never named by anyone

The same two arcs imply B over a four-fold range depending on the latent law assumed. Elicitation invariance is an open problem.

esylos

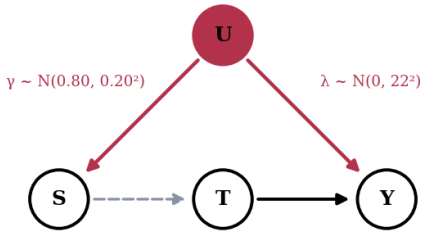
Five stories about the external controls we cannot rule out

V1 exchangeable
no unmeasured path



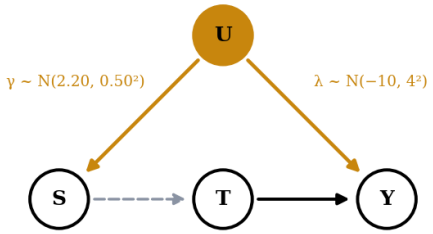
$B \equiv 0$

V2 severity / genotype
one latent severity axis



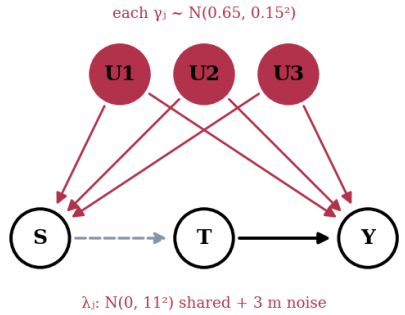
$B = \lambda\Delta(\gamma)$, sign of λ unknown

V3 era and care drift
2008–10 versus 2013–14



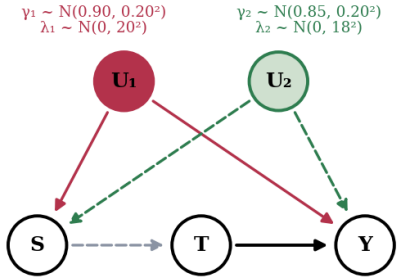
$B = \lambda\Delta(\gamma)$, $\lambda < 0$ asserted

V4 three compounding
referral, adherence, genotype



$B = \Sigma_j \lambda_j\Delta(\gamma_j)$

V5 two opposing
severity versus supportive care



$B = \lambda_1\Delta(\gamma_1) - \lambda_2\Delta(\gamma_2) + \eta$

Specification	Membership arc γ	Outcome arc λ (m lost per SD)	Implied B: mean (sd)
V2 severity / genotype	N(0.80, 0.20^2)	N(0, 22^2) — direction unknown	0.0 (16.0)
V3 era and care drift	N(2.20, 0.50^2)	N(−10, 4^2) — direction asserted	−12.7 (5.1)
V4 three compounding	N(0.65, 0.15^2) × 3	N(0, 11^2) shared + 3 m noise	0.0 (20.1)
V5 two opposing	N(0.90/0.85, 0.20^2)	N(0, 20^2) − N(0, 18^2)	+0.3 (21.0)

Only the era story asserts a direction, and it is the only one that moves the estimate rather than widening it. Provenance: Δ by quadrature with $a_0 = \log(114/31)$; the 22 m per SD is the published SPP1 genotype contrast. Sign: λ and B positive = external controls fare worse.

esylos

Source node S: Bareinboim & Pearl 2016 · bias from confounder arcs: Lin, Psaty & Kronmal 1998; Greenland 2005; McCandless, Gustafson & Levy 2007; Turner et al. 2009 · SPP1: Bello et al. 2012

How much we borrow, and what we refuse to update

The external control data enter through a NORMALISED power prior: their likelihood is raised to a random power w, with a Beta(1,1) prior, integrated out, and its posterior carries the normalising constant C(w,B). That constant is what makes the borrowing adaptive rather than a fixed discount — the observed discrepancy enters the weight posterior through its square. Here E[w | data] = 0.56 when the control arms agree to 1.5 m.

That is the part the data are allowed to decide. The displacement is not: under world view k it is drawn from the elicited distribution, held fixed, and withheld from the likelihood

$$p_k(\theta \mid D) = \iint \left[\int p(\theta \mid D, w, B)\, p(w \mid D, B)\, \mathrm{d}w \right] p_k(B)\, \mathrm{d}B$$

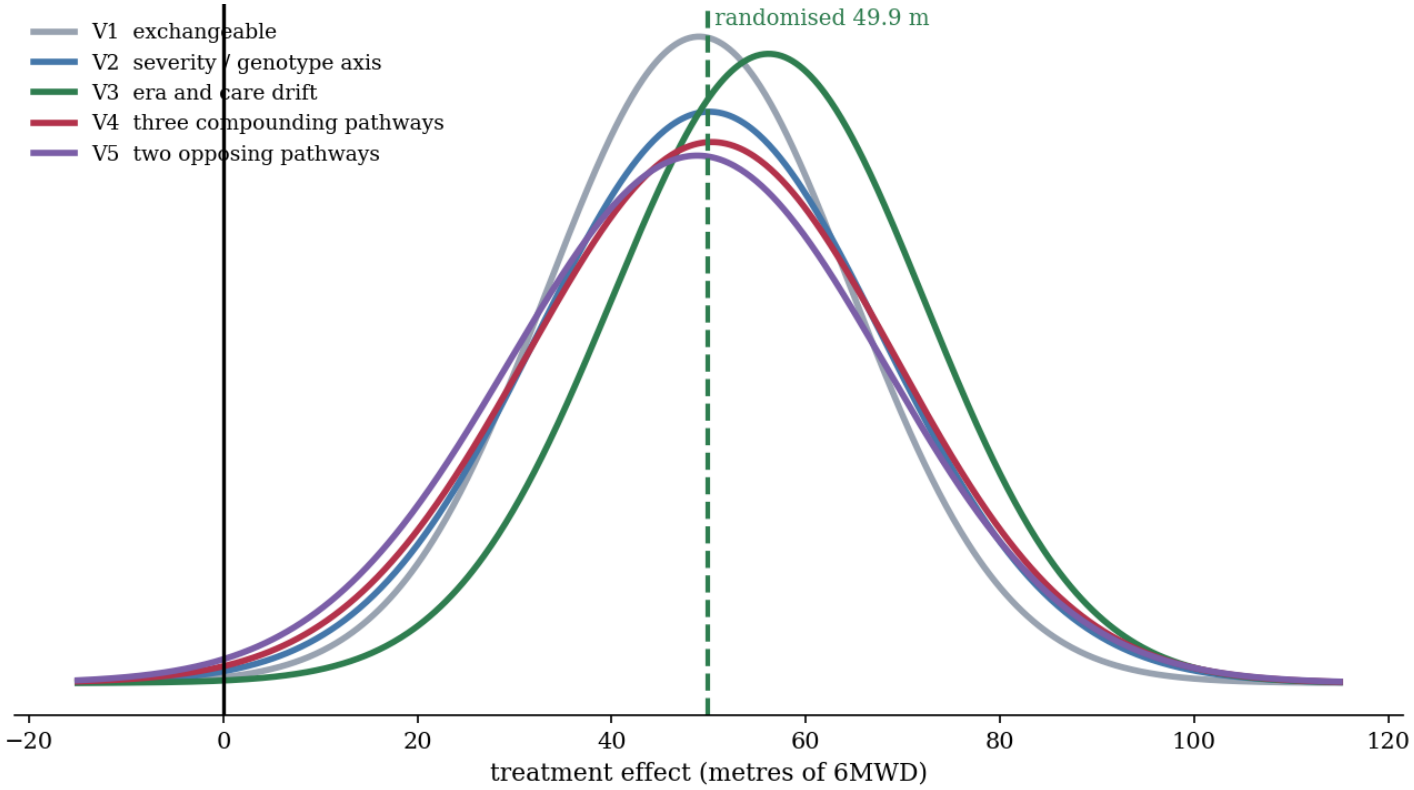
We refuse that update because B is an elicited causal discrepancy that the external control outcomes cannot identify. Let the likelihood have it and it estimates the confounding away.

Following the logic of probabilistic bias analysis (Greenland 2005), we have a cut distribution or prior weighted mixture of conditional posteriors.

esylos

Normalised power prior: Duan, Ye & Smith 2006; Neuenschwander, Branson & Spiegelhalter 2009; Shen et al. 2026 · cut / modular inference: Liu, Bayarri & Berger 2009; Plummer 2015

What each story does to the answer



Borrowing weight

E[w | data], per world view

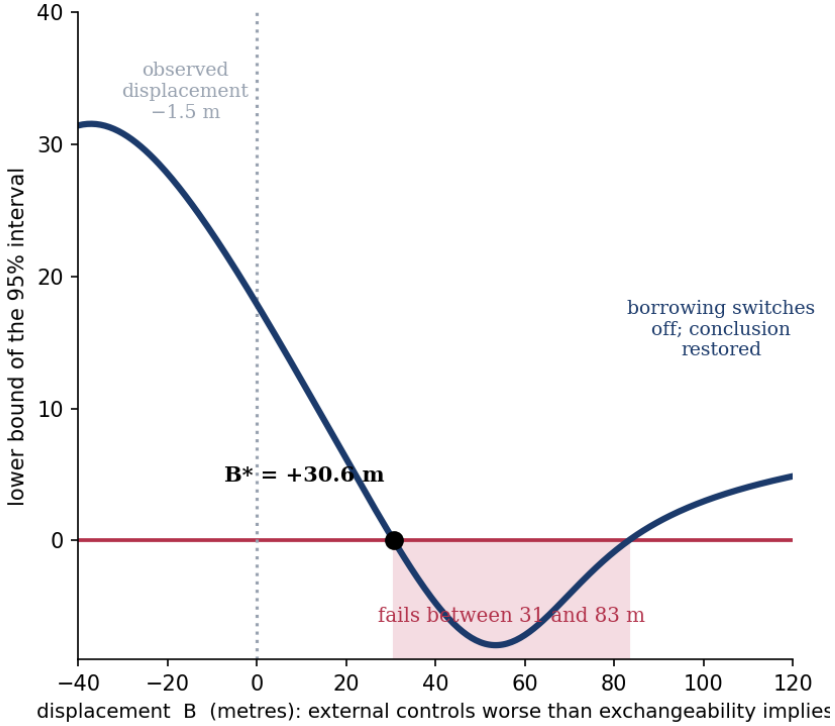
V1 exchangeable	0.56
V2 severity	0.54
V3 era drift	0.55
V4 compounding	0.52
V5 opposing	0.52

Weights barely move: agreement, not uniform discounting.

Five posteriors for the treatment effect, one per world view, displacement cut — this is the output of the analysis. Four specifications leave the estimate at 49.1–49.2 m; era drift is the exception, its direction assertable and negative, so correcting for it RAISES the estimate to 56.0 m.

esylos

What displacement would overturn the conclusion



The question: could any assumed source of unmeasured confounding imply an effect that overturns the conclusion?

The conclusion is that the 95% interval excludes zero. It fails at a displacement of B* = +30.6 m, the external controls faring worse than a comparable arm by that much, and stays failed up to +83 m; beyond that borrowing switches off and the trial-only result returns.

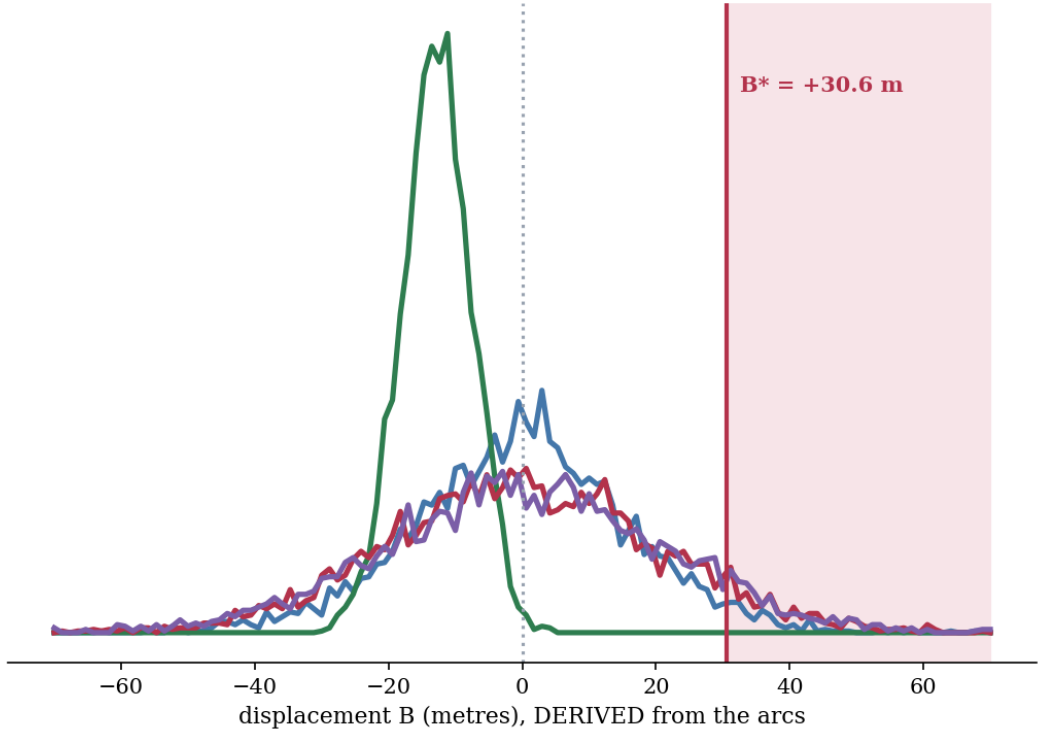
Borrowing buys precision, and the cost shows in the curve: once the displacement exceeds about +11 m, accounting for it leaves the lower bound below the trial-only 11.7 m.

Specification	Arc needed (m/SD)	Elicited scale	Ratio	Mass beyond B*
V2 severity / genotype	42.5	22	1.9×	0.029
V3 era and care drift	23.8	10	2.4×	0.000
V4 three compounding	16.9	11	1.5×	0.066
V5 two opposing	38.8	20	1.9×	0.073

At the strength elicited, no story reaches B*: each would need an arc 1.5 to 2.4 times stronger, and each places at most 7.3% of its prior beyond it. Among the stories we could state, none overturns the conclusion.

esylos

Why we do not weight the stories



■ V2 severity / genotype ■ V3 era and care drift ■ V4 three compounding ■ V5 two opposing

The displacement each story implies before any outcome is seen: the distributions its arcs imply, not estimates. Shaded: beyond B* = +30.6 m.

$$m_k(D) = \int p_k(B)\,\ell(B)\,\mathrm{d}B$$

and ℓ(B) contains no reference to the graph.

Two world views implying the same displacement receive the same posterior probability, whatever each claims about mechanism.

So we report each world view separately — no weights, no mixture.

esylos

Penalty for a diffuse prior: the Occam effect of the marginal likelihood (Lindley 1957)

What to put in the report

01 **One posterior per world view**

Five of them for this analysis, not a mixture.

02 **The range within each world view**

A narrow interval with a wide range across views means the elicitation is holding the answer up.

03 **The borrowing weight, per world view**

Five near-identical estimates can mean agreement or uniform discounting. Here all five sit near 0.56.

04 **The tipping point, on the arc scale**

What the unmeasured arc would have to be, set against what was elicited. Here 1.5 to 2.4 times the elicited scale.

esylos

Where this sits in the guidance

No new regulatory machinery is needed. The framework answers questions the guidance already asks, organised by mechanism.

01 FDA draft guidance (2026): dynamic discounting

Valued because it borrows less when the data are less similar. That protects against detectable conflict, not against similarity created by confounding.

02 The same guidance already asks for the ingredients

Changes in standard of care as a relevance factor; a bias parameter when bias is suspected; sensitivity analyses that build uncertainty about assumptions into the prior.

03 ICH E9(R1): a sensitivity analysis of the same estimand

θ, the effect in the trial population, is unchanged; only the assumption about the external controls varies. Pre-specified, with excluded sources recorded.

Sensitivity analyses usually vary how much is borrowed. This one varies why the borrowed data might mislead, on a scale a clinician can contest.

esylos

FDA draft guidance, Use of Bayesian Methodology in Clinical Trials of Drug and Biological Products (Jan 2026), §V.D–F, §VIII · ICH E9(R1) (2019)

Dynamic borrowing protects against detectable incompatibility, not against confounding that creates compatibility.

Elicit as widely as you can justify . Pre-specify the mechanisms, the arc distributions, the exclusions and the decision rule before you look at outcomes.

esylos

References

Bareinboim E, Pearl J. Causal inference and the data-fusion problem. PNAS 2016;113:7345–52.

Bello L, et al. Importance of SPP1 genotype as a covariate in clinical trials in DMD. Neurology 2012;79:159–62.

Duan Y, Ye K, Smith EP. Evaluating water quality using power priors to incorporate historical information. Environmetrics 2006;17:95–106.

EMA. Translarna (ataluren) assessment report, EMA/CHMP/369266/2014.

Galwey NW. Supplementation of a clinical trial by historical control data: is the prospect of dynamic borrowing an illusion? Stat Med 2017;36:899–916.

Greenland S. Multiple-bias modelling for analysis of observational data. J R Stat Soc A 2005;168:267–306.

Hobbs BP, Carlin BP, Mandrekar SJ, Sargent DJ. Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. Biometrics 2011;67:1047–56.

Ibrahim JG, Chen M-H. Power prior distributions for regression models. Stat Sci 2000;15:46–60.

Kopp-Schneider A, Calderazzo S, Wiesenfarth M. Power gains by using external information in clinical trials are typically not possible when requiring strict type I error control. Biom J 2020;62:361–74.

Lin DY, Psaty BM, Kronmal RA. Assessing the sensitivity of regression results to unmeasured confounders in observational studies. Biometrics 1998;54:948–63.

Lindley DV. A statistical paradox. Biometrika 1957;44:187–92.

Liu F, Bayarri MJ, Berger JO. Modularization in Bayesian analysis, with emphasis on analysis of computer models. Bayesian Anal 2009;4:119–50.

McCandless LC, Gustafson P, Levy A. Bayesian sensitivity analysis for unmeasured confounding in observational studies. Stat Med 2007;26:2331–47.

McDonald CM, et al. Ataluren in patients with nonsense mutation Duchenne muscular dystrophy (ACT DMD). Lancet 2017;390:1489–98.

Neuenschwander B, Branson M, Spiegelhalter DJ. A note on the power prior. Stat Med 2009;28:3562–6.

Plummer M. Cuts in Bayesian graphical models. Stat Comput 2015;25:37–43.

Pocock SJ. The combination of randomized and historical controls in clinical trials. J Chronic Dis 1976;29:175–88.

Schmidli H, Gsteiger S, Roychoudhury S, O'Hagan A, Spiegelhalter D, Neuenschwander B. Robust meta-analytic-predictive priors in clinical trials with historical control information. Biometrics 2014;70:1023–32.

Shen Y, Carvalho LM, Psioda MA, Ibrahim JG. Optimal priors for the discounting parameter of the normalized power prior. Stat Sinica 2026;36:843–66.

Turner RM, Spiegelhalter DJ, Smith GCS, Thompson SG. Bias modelling in evidence synthesis. J R Stat Soc A 2009;172:21–47.

VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. Ann Intern Med 2017;167:268–74.

Viele K, et al. Use of historical control data for assessing treatment effects in clinical trials. Pharm Stat 2014;13:41–54.

esylos