

Online Appendix for Enhancing Safety Signaling: Integrating Clinical Trials and Post-Marketing Adverse Event Reports

Fernanda Bravo

Anderson School of Management, UCLA, fernanda.bravo@anderson.ucla.edu

Lawrence Yunliang Chen

Industrial Engineering and Operations Research, University of California, Berkeley, yunliang.chen@berkeley.edu

John Silberholz

Robert H. Smith School of Business, University of Maryland josilber@umd.edu

Endnotes

¹GI disorders (*e.g.*, nausea, vomiting, and diarrhea) are included in Bosutinib’s original label under “Warnings and Precautions” (DailyMed 2023).

²For instance, the log odds ratio effect size for ADR i defined in eq. (1) has variance $\frac{\sigma_i^2}{n} = \frac{2}{n} \left(\frac{1}{p_i^0} + \frac{1}{1-p_i^0} + \frac{1}{p_i^1} + \frac{1}{1-p_i^1} \right)$ in a trial with n patients evenly divided between the treatment (index 1) and control (index 0) arms. An ADR with a baseline rate of $p_i^0 = 0.001$ in the control arm will have $\sigma_i = 63.3$ when $\mu_i = 0$ (no effect of the treatment on the ADR rate), will have $\sigma_i = 61.8$ when $\mu_i = 0.1$ (treatment increases the odds of the ADR by roughly 10%), and will have $\sigma_i = 64.9$ when $\mu_i = -0.1$ (treatment reduces the odds of the ADR by roughly 10%).

³We assume b_i^O and b_i^C are uncorrelated, as clinical trial bias often results from narrow inclusion criteria, whereas observational bias typically arises from selection or reporting differences. Since these sources are distinct, a significant correlation is unlikely. Our numerical evaluation in Section 5 focuses on the setting where generalization bias is not large, limiting the impact of this modeling choice on the paper’s numerical results.

⁴Signal validation is a multi-step process involving detailed case reviews and supporting safety data. The FDA prioritizes signals based on seriousness, novelty, or unexpected frequency, and classifies individual reports as “probable,” “possible,” or “unlikely.” Probable cases trigger a case series review, which considers clinical and demographic details, exposure patterns, and reporting trends. If a signal is confirmed, sponsors are required to conduct follow-up studies (*e.g.*, observational studies, registries), and the FDA evaluates the results to determine causality and guide regulatory actions (Report of CIOMS Working Group VIII 2010).

⁵Despite not being the primary goal of safety surveillance systems, our model could be adapted to guide the removal of ADRs from drug labels. However, Type I and Type II error costs would need to be calibrated accordingly.

⁶Following Assumption 1, the off-diagonal terms of Σ^{sC} and Σ^{sO} are already assumed to be approximately 0.

⁷A bit more formally, it is easy to check that the difference $(\tilde{\sigma}_2^S)^2 - (\sigma_2^S)^2$ between the posterior variances of the single-ADR strategy and our strategy approaches a constant as $\sigma_2 \rightarrow \infty$, even though both posterior variances are growing large (scaling with σ_2^2). This implies that $\hat{\mu}_2^S/\sigma_2^S$, the z-score after collecting the observational data, will be

nearly identical for the two strategies when σ_2 is large, meaning that they will almost always make the same signaling decision (see [Lemma 2](#)).

⁸More precisely, the peak occurs where $\hat{\mu}_2^A/\sigma_2^S$, the clinical trial effect size normalized by the posterior variance after collecting observational data, is right at the decision boundary. Here we consider a clinical trial that precisely estimates the effect size, so the distinction is minor — with $\sigma_2 = 9.5$, we have a clinical trial z-score $\hat{\mu}_2^A/\sigma_2^A = 1.47$, while we have $\hat{\mu}_2^A/\sigma_2^S = 1.51$.

⁹Notice in the case of $k \leq 1$ that monotone increasing benefit requires $\hat{\mu}_2^A z^* \leq 0$. In fact, the “tiebreaker effect” we saw for σ_2 , n^C , and σ_2^{bC} can again play a role when $k \leq 1$ and $\hat{\mu}_2^A z^* > 0$, as can be seen in Figure EC.1 in [e-companion](#). Recall from [Section](#) that a spike in benefit can occur when $\hat{\mu}_2^A/\sigma_2^S$ coincides with the critical z-score for the signaling loss function. The increases in n^O correspond to decreases in σ_2^S , and the spike in Figure EC.1 in [e-companion](#) at $n^O \approx 10,000$ corresponds to the n^O value at which $\hat{\mu}_2^A/\sigma_2^S$ falls at the decision boundary. Interestingly, the “tiebreaker effect” does not play a role in the comparative statics for n^O when $k > 1$.

¹⁰We focus on new, extremely rare ADRs for which regulators, lacking trial evidence, initially take a wait-and-see approach and monitor post-market data. While the FDA also monitors ADRs seen in trials (*e.g.*, cardiac failure in bosutinib), we exclude these to avoid assumptions about regulator behavior. Our focal ADRs will not be labeled at approval, so approval-time loss functions T_I^A and T_{II}^A need not be estimated. Evaluating debiasing for more common ADRs is left for future work.

¹¹Formally: we perform random-effects meta-analysis to combine the $\hat{\mu}_i^C$ values across indications for the drug, and require that this second random-effect meta-analysis also not show signs of heterogeneity (we require heterogeneity parameter estimate $\tau = 0$).

¹²This system defines a hierarchy of drugs; we use level 4 of the hierarchy to obtain chemical subgroups of drugs. We chose to group drugs within a therapeutic class because they are chemically similar, are often studied in similar patient populations (subject to similar inclusion and exclusion criteria in clinical trials), and may exhibit similar patterns of confounding and reporting in observational data.

Appendix A: Notation and supporting material

Table A1 **Notation**

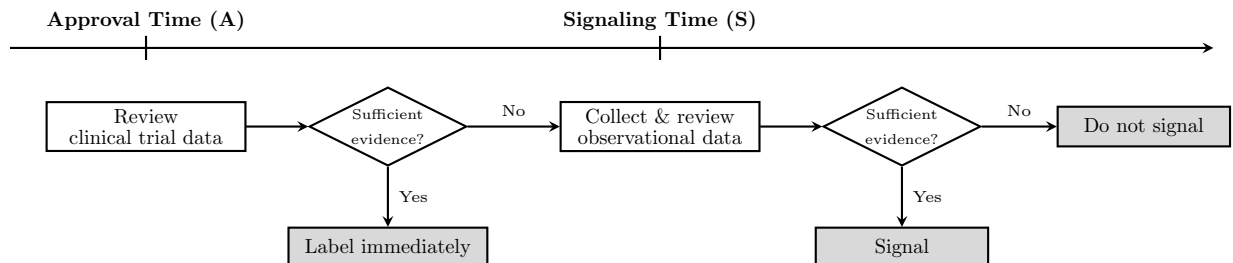
μ_i	true effect size of ADR i
$\hat{\mu}_i^C$ ($\hat{\mu}_i^O$)	effect size estimate in clinical trials (observational data) for ADR i
σ_i	population standard deviation for ADR i
n^C (n^O)	size of the clinical trial (observational data)
ϵ_i^C (ϵ_i^O)	sampling noise of clinical trial (observational data) for ADR i
ρ^{sC} (ρ^{sO})	clinical trial (observational data) sampling correlation between the common and rare ADRs
b_i^C	generalization bias in clinical trials for ADR i
μ_i^{bC}	mean clinical trial generalization bias for ADR i
ρ^{bC}	clinical trial generalization bias correlation between the common and rare ADRs
σ_i^{bC}	standard deviation of clinical trial generalization bias for ADR i
b_i^O	observational data bias for ADR i
μ_i^{bO}	mean observational bias for ADR i
ρ^{bO}	observational data bias correlation between the common and rare ADRs
σ_i^{bO}	standard deviation of observational data bias for ADR i
Σ^{sC} (Σ^{sO})	covariance matrix of clinical trials data (observational data) sampling errors
Σ^{bC} ($\tilde{\Sigma}^{bC}$)	covariance matrix of generalization bias (single-ADR approach)
Σ^{bO} ($\tilde{\Sigma}^{bO}$)	covariance matrix of observational data bias (single-ADR approach)
$\hat{\mu}_2^A$	ADR 2 <i>prior</i> mean effect size estimate at approval time
σ_2^A	ADR 2 <i>prior</i> effect size standard deviation at approval time
$\hat{\mu}_2^S$	ADR 2 <i>posterior</i> mean effect size estimate at signaling time
σ_2^S ($\tilde{\sigma}_2^S$)	ADR 2 <i>posterior</i> effect size standard deviation at signaling time (single-ADR approach)
$\nu(\tilde{\nu})$	<i>pre-posterior</i> standard deviation of $\hat{\mu}_2^S$ (uncertainty at approval time)
$T_I^l(\cdot)$	type I error cost function at time $l \in \{A, S\}$
$T_{II}^l(\cdot)$	type II error cost function at time $l \in \{A, S\}$
M_I^l	type I error cost of μ^k -loss function at time $l \in \{A, S\}$
M_{II}^l	type II error cost of μ^k -loss function at time $l \in \{A, S\}$
k	power of the μ^k -loss function
$\mathcal{L}_l(\cdot)$	expected loss function at time $l \in \{A, S\}$
$\mathcal{L}_S^*(\cdot)$	optimal expected loss function at signaling time
$V_l^*(\cdot)$	optimal expected cost at time $l \in \{A, S\}$
t_l^*	optimal mean effect size threshold for labeling (or signaling) at time $l \in \{A, S\}$
B	benefit or cost difference between the cross-ADR approach and the single-ADR approach

Appendix B: Numerical analysis

B.1. Testing of bias bivariate normality assumption

Recall from eq. (12) that we assume our bias estimates $\Delta \hat{\mathbf{b}} := \hat{\mu}^O - \hat{\mu}^C$ to be approximately bivariate normal under the filters defined in Section 5.2.2. We test this assumption via the Henze-

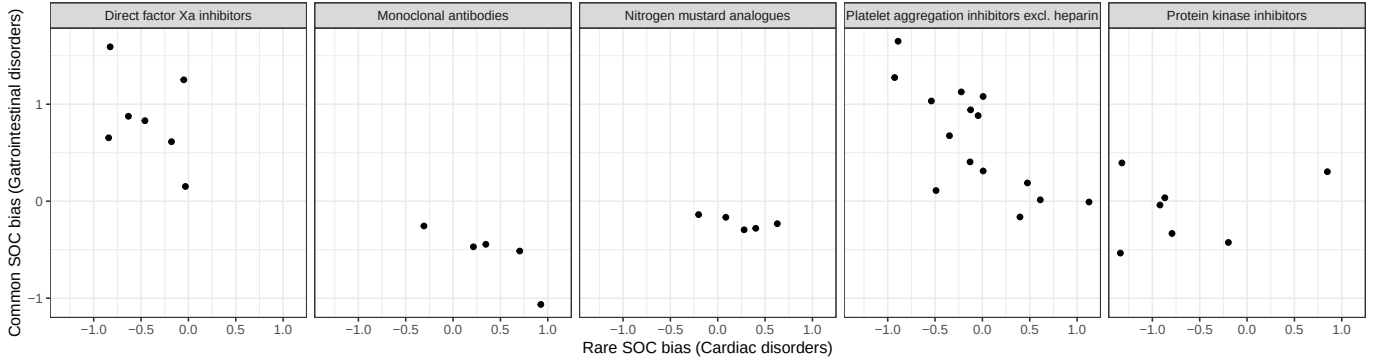
Figure A1 **Timeline of data collection and decision framework**



Zirkler test applied to bias difference estimates $\Delta \hat{\mathbf{b}}$. The test’s p-value can be used to reject the null hypothesis of bivariate normality for the observational bias. Given the small number of drug-indications in each drug class, we can only test bivariate normality for one of our drug classes (Platelet aggregation inhibitors excl. heparin), and we obtain a p-value > 0.05 for all 26 SOC’s.

We illustrate the bias data using the cardiac disorders and gastrointestinal SOC’s in **Figure B2**.

Figure B2 Example of SOC-level bias data used to test the bivariate normal assumption



B.2. Maximum-likelihood estimation of bias distribution parameters

As described in **Section 5.2.3**, the core data used to estimate the bias distribution are clinical trial and observational data for drug-indications of chemically similar drugs (drugs in the same level 4 ATC drug class), filtered according to a set of criteria laid out in **Section 5.2.2**. For each drug-indication pair (d, i) in our drug class $\mathcal{C}$, we obtain estimates for the SOC of the common ADR (which we will denote with index “1”) and of the SOC of the rare ADR (which we will denote with the index “2”). As discussed in **Section 5.2.3**, the MedDRA SOC term associated with an ADR is a higher-level categorization capturing not only that ADR but also similar ones. For each drug-indication, random-effects meta-analyses of clinical trial results (see **Section 5.2.1**) yield effect size estimates $\hat{\mu}_{(d,i),1}^C$ and $\hat{\mu}_{(d,i),2}^C$ of the effect of the drug on the incidence of the two SOC’s, along with standard errors $\sigma_{(d,i),1}^C$ and $\sigma_{(d,i),2}^C$. Further, the reporting odds ratios from FAERS for each drug-indication pair yields observational data effect size estimates $\hat{\mu}_{(d,i),1}^O$ and $\hat{\mu}_{(d,i),2}^O$ and standard errors $\sigma_{(d,i),1}^O$ and $\sigma_{(d,i),2}^O$.

Following equations (2) and (3), the differences between the observational and clinical trial effect size estimates can be re-expressed in terms of bias and sampling error terms:

$$\begin{aligned}\hat{\mu}_{(d,i),1}^O - \hat{\mu}_{(d,i),1}^C &= b_{(d,i),1}^O - b_{(d,i),1}^C + \epsilon_{(d,i),1}^O - \epsilon_{(d,i),1}^C, \text{ and} \\ \hat{\mu}_{(d,i),2}^O - \hat{\mu}_{(d,i),2}^C &= b_{(d,i),2}^O - b_{(d,i),2}^C + \epsilon_{(d,i),2}^O - \epsilon_{(d,i),2}^C.\end{aligned}$$

Before estimating the parameters of our system, we make two simplifying assumptions motivated by the filters applied to our drug-indication pairs (see **Section 5.2.2**, criterion 3):

1. Since we parameterize our model using only drug-indication pairs with limited concerns related to the generalizability of clinical trial results, we assume that the biases $b_{(d,i),1}^C, b_{(d,i),2}^C \approx 0$.
2. Since we parameterize our model using only drug-indication pairs where both the clinical trial and observational data effect sizes for the common ADR are well estimated from data (have small standard errors), we assume that the covariance term for our sampling errors are small and that our system can be well approximated by assuming that $\rho^{sO}, \rho^{sC} \approx 0$.¹³

Recall that the bias terms have means μ_1^{bO} and μ_2^{bO} , which we will estimate. Now, we can appeal to the distributions of bias $\mathbf{b}_{(d,i)}^O$ and sampling errors $\epsilon_{(d,i)}^C$ and $\epsilon_{(d,i)}^O$ to obtain the distribution of our effect size differences, which replicates [eq. \(12\)](#):

$$\begin{bmatrix} \hat{\mu}_{(d,i),1}^O - \hat{\mu}_{(d,i),1}^C \\ \hat{\mu}_{(d,i),2}^O - \hat{\mu}_{(d,i),2}^C \end{bmatrix} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N} \left(\begin{bmatrix} \mu_1^{bO} \\ \mu_2^{bO} \end{bmatrix}, \begin{bmatrix} (\sigma_1^{bO})^2 + (\sigma_{(d,i),1}^C)^2 + (\sigma_{(d,i),1}^O)^2 & \rho^{bO} \sigma_1^{bO} \sigma_2^{bO} \\ \rho^{bO} \sigma_1^{bO} \sigma_2^{bO} & (\sigma_2^{bO})^2 + (\sigma_{(d,i),2}^C)^2 + (\sigma_{(d,i),2}^O)^2 \end{bmatrix} \right). \quad (\text{B.1})$$

Within [eq. \(B.1\)](#), the data for each drug-indication pair $(d, i) \in \mathcal{C}$ is known, while we seek to estimate bias distribution parameters μ_1^{bO} , μ_2^{bO} , σ_1^{bO} , σ_2^{bO} , and ρ^{bO} . Using [eq. \(B.1\)](#), define $LL_{(d,i)}(\mu_1^{bO}, \mu_2^{bO}, \sigma_1^{bO}, \sigma_2^{bO}, \rho^{bO})$ to be the log-likelihood of observing our results for drug-indication pair (d, i) under a specified set of bias distribution parameters. We proceed to select these parameters using maximum likelihood estimation, with a penalty for large values of $|\rho^{bO}|$. In particular, we obtain the parameters for our SOC pair in drug class $\mathcal{C}$ by solving the following optimization problem under regularization parameter $\lambda > 0$:

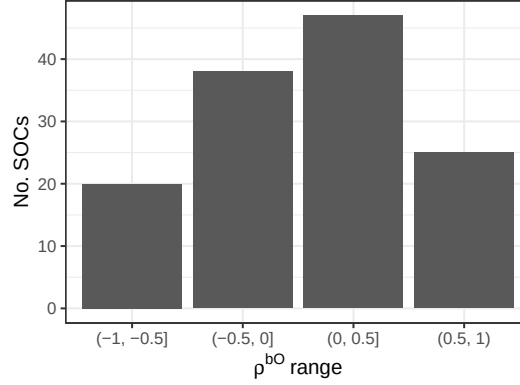
$$\max_{\mu_1^{bO}, \mu_2^{bO}, \sigma_1^{bO} \geq 0, \sigma_2^{bO} \geq 0, -1 < \rho^{bO} < 1} \frac{1}{|\mathcal{C}|} \sum_{(d,i) \in \mathcal{C}} LL_{(d,i)}(\mu_1^{bO}, \mu_2^{bO}, \sigma_1^{bO}, \sigma_2^{bO}, \rho^{bO}) - \lambda \cdot \log \frac{1 + |\rho^{bO}|}{1 - |\rho^{bO}|} \quad (\text{B.2})$$

Note that the regularization term takes value 0 when $\rho^{bO} = 0$, and otherwise increases with $|\rho^{bO}|$.

To select the λ to use in our regularized model, we proceed by 10-fold cross-validation. In particular, we randomly partition our drug-indication pairs for our drug class. For each partition, we optimize the bias distribution parameters using [eq. \(B.2\)](#), fitted to all drug-indications except those in the selected partition; we separately estimate parameters for each λ in a set of candidate values $\{0.001, 0.002, 0.005, 0.01, 0.02, 0.07, 0.1, 0.2, 0.3\}$. We then use those fitted parameters to obtain an out-of-sample log-likelihood for the drug-indication pairs within the selected partition. We repeat this process for each partition, obtaining a set of out-of-sample log-likelihoods for each drug-indication pair under each of our candidate λ values. We select the λ with the best out-of-sample performance and use that value to estimate the overall model using all the data from the drug class. The resulting estimates for our critical parameter ρ^{bO} are reported in [Figure B3](#). Recalling the discussion in [Section 3.1](#) of the mechanism behind the observational bias considered in this work, it is unsurprising that roughly half of the non-zero ρ^{bO} values are positive and the other half are negative—if a subgroup like diabetic patients have some side effects that constitute

a larger-than-average proportion of their reported ADRs (*e.g.*, renal and cardiovascular ADRs), then there must be a set of side effects that constitute a smaller-than-average proportion of their ADRs (*e.g.*, musculoskeletal and lymphatic ADRs).

Figure B3 ρ^{bO} estimates



Note. Each observation corresponds to a SOC paired with gastrointestinal disorders SOC. We estimate 130 values of ρ^{bO} across five drug classes.

B.3. Additional results

The final set of drug-PT pairs included in the analysis is reported in [Table B1](#). Throughout this appendix section, we present a series of results comparing the proposed approach’s performance to that of the two benchmark approaches.

Table B1 Summary of drug-PT pairs included in the numerical analysis.

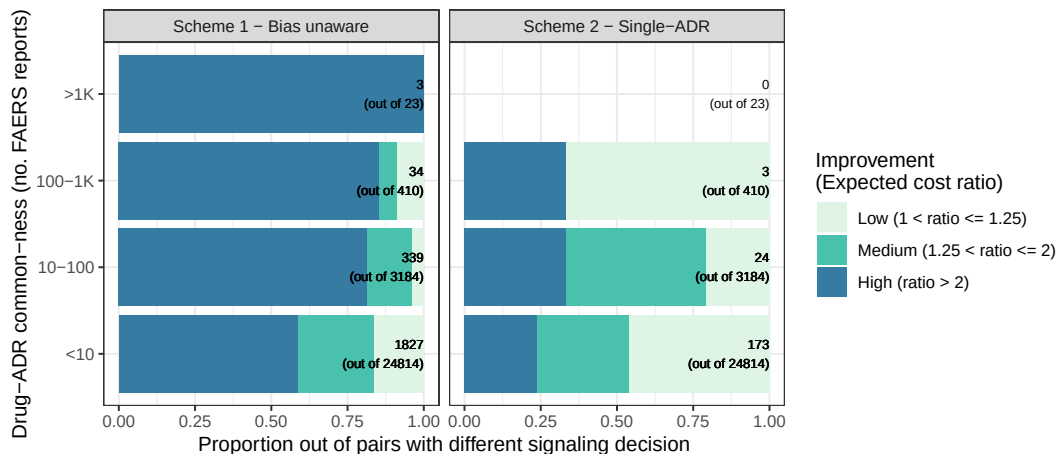
Drug Name	PTs	No. Signal (%)
Cyclophosphamide	2,560	930 (36%)
Rituximab	4,402	1,713 (39%)
Cetuximab	1,905	587 (31%)
Clopidogrel	1,973	553 (28%)
Cilostazol	1,349	561 (42%)
Dasatinib	1,949	427 (22%)
Sorafenib	2,650	967 (36%)
Chlorambucil	598	251 (42%)
Nilotinib	3,344	916 (27%)
Elotuzumab	520	224 (43%)
Vemurafenib	1,514	534 (35%)
Ruxolitinib	3,655	1,169 (32%)
Afatinib	1,310	461 (35%)
Obinutuzumab	702	272 (39%)
Total	28,431	9,565 (34%)

Note. Signals correspond to those made by our approach, assuming a z-score of 2 to calibrate M_1 .

Performance by commonness of the ADR. Our approach reports benefits across all commonness buckets, as shown in [Figure B4](#). Relative to Scheme 2, the cross-ADR improves about 0.7% of the pairs within each common-ness bucket except the largest ($> 1K$ reports), and about 7–13% when

compared to Scheme 1 across common-ness buckets. Of note, we find that the cross-ADR approach achieves superior performance for a large number of extremely rare ADRs (< 10 FAERS reports), which account for over 80% of all pairs with improvement. This is a key advantage of our approach; it has the potential to detect signals earlier in the post-marketing surveillance process (*i.e.*, with only a few reports).

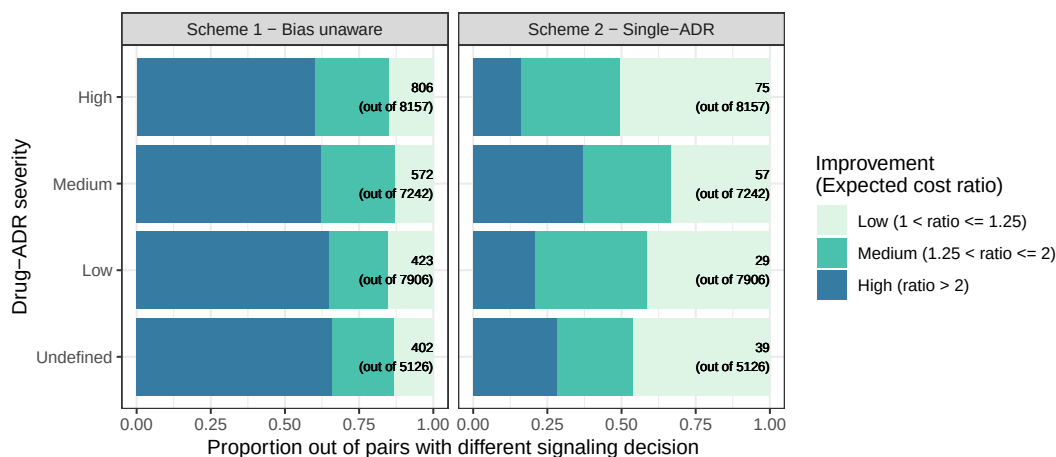
Figure B4 Expected cost ratio by ADR commonness



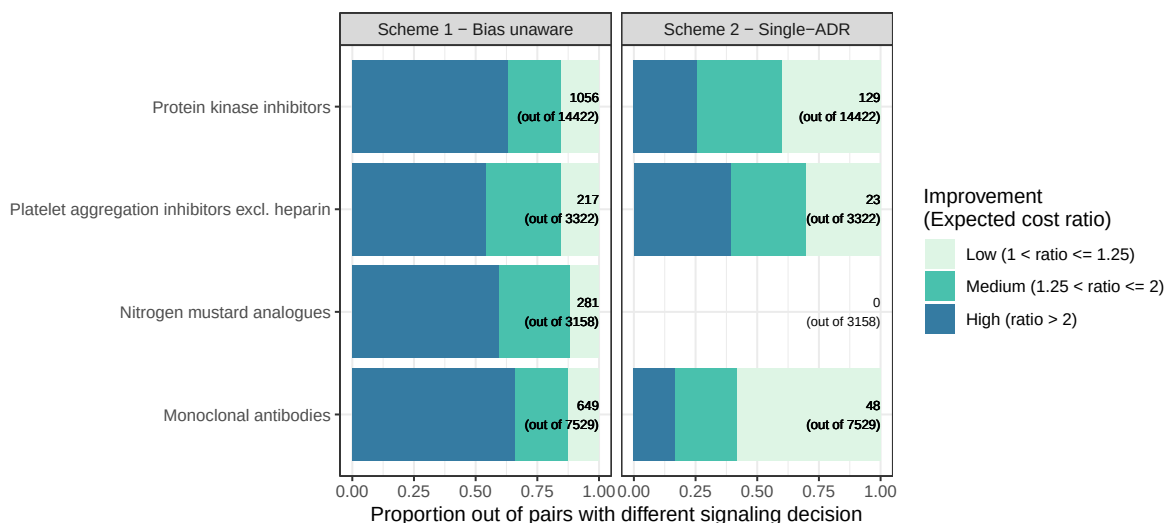
Note. The commonness bucket of a drug-ADR pair corresponds to the ADR raw count of FAERS reports. The number on the far right of each bucket corresponds to the number of drug-ADR pairs with different signaling decisions (compared to scheme 3), and the one in parentheses corresponds to the total number of drug-ADRs in that bucket.

Performance by ADR severity. Figure B5 shows the benefit of the cross-ADR approach in relation to Schemes 1 and 2, categorized by the severity of the ADR. We utilize the severity classification proposed by Lavertu et al. (2021). It is worth noting that the severity subgroups in our data tend to have a comparable number of pairs. Interestingly, the performance of the cross-ADR approach appears to be consistent across different ADR severities, as it shows comparable results for both high- and low-severity ADRs.

Performance by drug class. Figure B6 shows the improvement of the cross-ADR debiasing approach relative to Scheme 1 (left panel) and Scheme 2 (right panel) by drug class. The suboptimal signaling decisions made under the bias-unaware approach lead to expected costs that are twice as large as our approach for more than 50% of the pairs for all drug classes. Comparing the cross-ADR approach to Scheme 2, we observe comparable benefits across all but one drug class. In the case of nitrogen mustard analogues drug class, signaling differences are minimal due to the low correlation ρ^{bO} , which is near zero for all but three SOCs. Even in those SOCs (covering 715 PTs), the posterior mean difference between our method and the single-ADR approach remains low, explaining the alignment in signaling outcomes.

Figure B5 Expected cost ratio by ADR severity

Note. Severity classification is based on the PT of the pair. The number on the far right of each bucket corresponds to the number of drug-ADR pairs with different signaling decisions (compared to scheme 3), and the one in parentheses corresponds to the total number of drug-ADRs in that bucket.

Figure B6 Expected cost ratio by drug class

Note. The number on the far right of each drug class bucket corresponds to the number of drug-ADRs with different signaling decisions compared to scheme 3, and the one in parentheses corresponds to the total number of drug-ADR pairs in that bucket.

Appendix C: Model misspecification

In addition to the theoretical analysis of misspecification in the key correlation bias parameter ρ^{bO} (e-companion, Section E), we assess the impact of misspecification in all model parameters of the analytical model using estimates from the numerical analysis. Specifically, we focus on the same set of 28,431 drug-ADR pairs, and restrict to 26,307 pairs that are associated with SOC's other than

gastrointestinal disorders (our common ADR). Unlike our main numerical results, we compare the *expected cost at approval time* (noting that the ADRs have not been labeled at approval time since they have not been observed in clinical trials) (eq. (8)) rather than at signaling time to enable us to evaluate the impact of misspecification of all elements of our main model. While most parameter estimates are directly borrowed from Section 5.4, a few parameters are adjusted as follows:

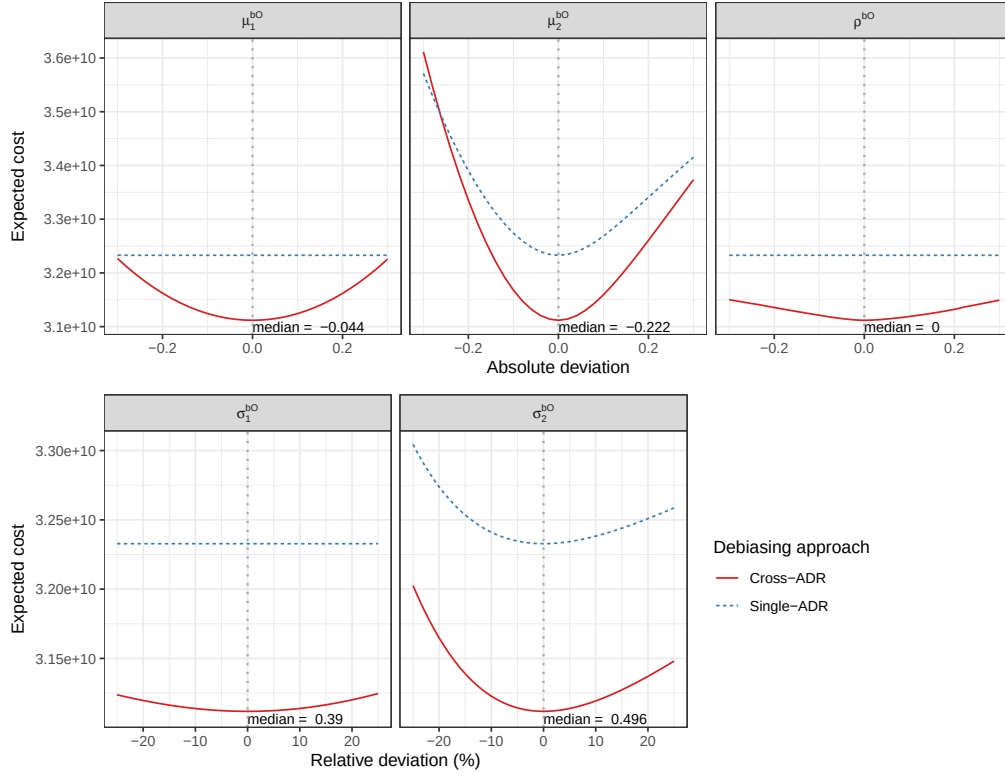
- *Loss function:* In contrast to our main results that consider a smoothed lump-sum loss function, we consider a standard lump-sum loss function ($k = 0$) for computation tractability. We use the cost parameters M_{II}^S and M_I^S (calibrated for a z-score = 2) from our main results for each drug-PT pair.
- *Bias distribution:* We use the bias distribution parameters estimated at the drug-class level for each pair of common-rare SOC.
- *Clinical trial distribution:* Unlike in the main analysis, we model clinical trial parameters for the rare ADR in the misspecification analysis, assuming a clinical trial effect size $\hat{\mu}_2^C = 0$ and estimating σ_2^C using a typical value of the clinical trial standard error at the SOC level (recall that the rare ADRs were not observed in the clinical trial for all our drug-ADR pairs, but similar ADRs within the same SOC have been observed).

We vary the level of misspecification, depending on the parameter, within a moderate range of values. For positive-valued parameters (*e.g.*, σ_i^{bO}), we consider relative deviations: $\text{param}^{new} := \text{deviation} \times \text{param}^{old}$. For all other parameters, we consider absolute deviations, *i.e.*, $\text{param}^{new} := \text{deviation} + \text{param}^{old}$, with both positive and negative deviation depending on the parameter.¹⁴ Note that if the parameter affects both the cross-ADR and the single-ADR benchmark, then both approaches make the signaling decision with the new, misspecified parameter. Note that all the parameters associated with the common ADR (*e.g.*, μ_1^{bO}) only affect our approach, not the single ADR benchmark.

C.1. Misspecification results

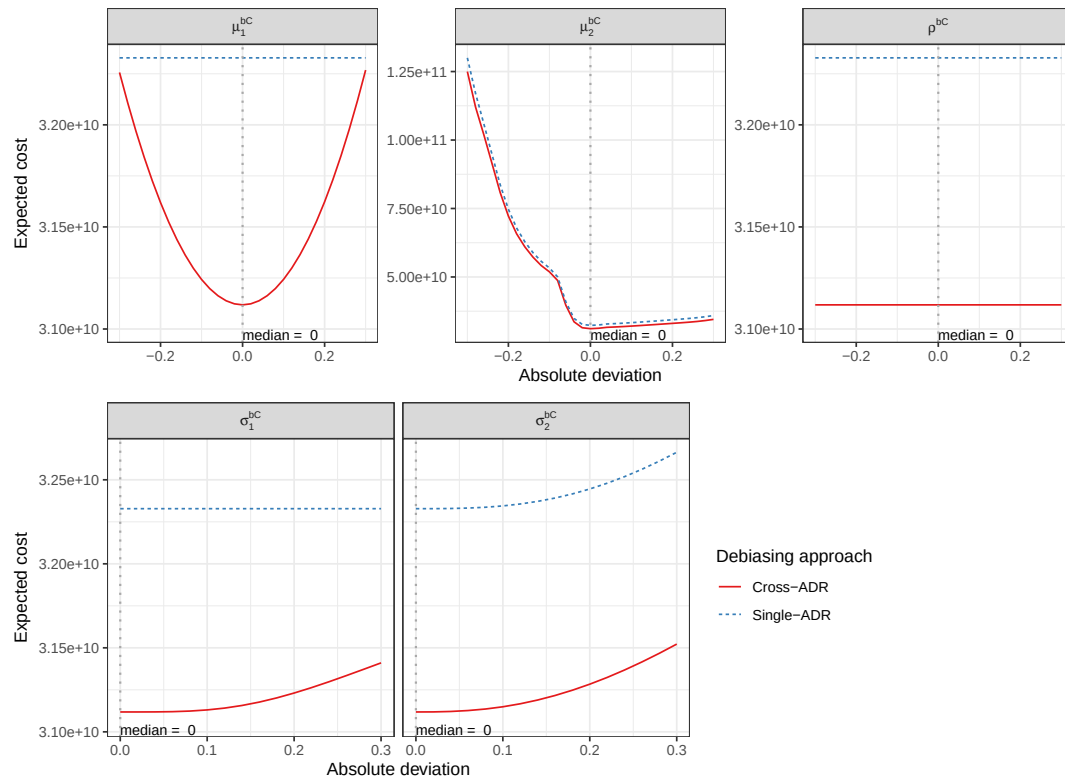
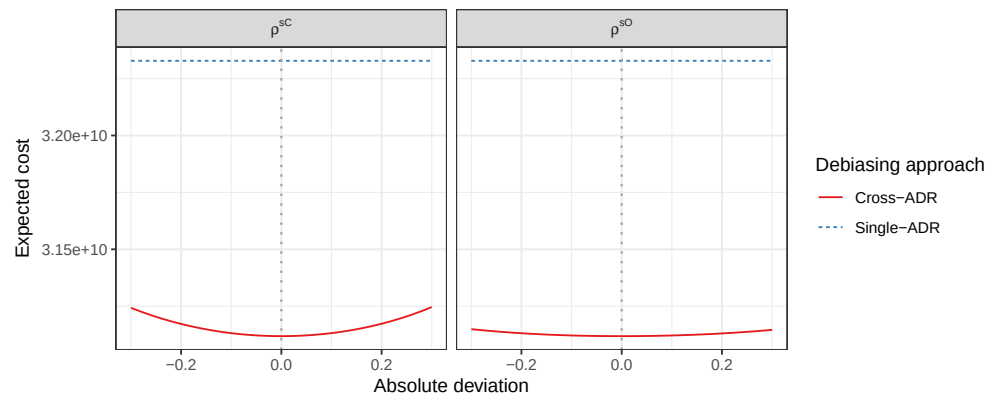
Our approach is generally robust to moderate misestimation.

Observational bias. Figure C7 shows that even with moderate misspecification, our approach consistently outperforms the single-ADR benchmark. Although the cost advantage decreases as misspecification increases, our method remains superior when the odds ratio parameter ($\exp(\mu_i^{bO})$) deviates within -26% to $+35\%$. In cases of severe underestimation of μ_2^{bO} , our method may perform worse. However, such large errors are unlikely because μ_2^{bO} is estimated from a large set of drug-indication pairs, reducing the risk of significant estimation error. Even with large deviations in the bias correlation parameter ρ^{bO} (up to ± 0.5), our approach still achieves a 1–2% reduction in expected costs. Our results also show strong robustness to estimation errors in the bias noise σ_i^{bO} . While errors in σ_2^{bO} affect both methods, our approach continues to maintain a performance advantage.

Figure C7 Misspecified observational bias parameters (median parameter across selected drug-PTs)

Generalization bias. Figure C8 illustrates the sensitivity of benefits to generalization bias parameter estimation errors. The correlation ρ^{bC} does not affect the performance under our selected subset of drug-ADRs (recall that we selected drug-ADRs with estimates $\sigma_1^{bC} = 0$). Similar to above, our approach performs better than the single-ADR approach for small-to-moderate estimation errors in μ_1^{bC} . The deviations in the rare ADR mean bias μ_2^{bC} affect both approaches in a similar way.

Sampling correlations. Figure C9 shows the sensitivity of the results to the sampling correlation parameters, ρ^{sC} and ρ^{sO} . Although we assume these parameters are zero, our approach continues to outperform the benchmark even with moderate deviations. The expected cost increases by less than 1.5% under the largest deviations considered. This is unsurprising, since we focus on drug-ADRs with well-estimated common ADRs.

Figure C8 Misspecified generalization bias parameters (median parameter across selected drug-PTs)**Figure C9** Misspecified sampling correlation parameters

Appendix D: Economic evaluation

D.1. Costs and loss functions

Define m_{ij} to be the expected number of future individuals taking drug i who will experience rare ADR j of interest. Some of these individuals would experience the ADR regardless of drug exposure; under a log risk ratio of $\beta \geq 0$ we would expect $e^{-\beta}m_{ij}$ individuals to experience the side effect without drug exposure, meaning $(1 - e^{-\beta})m_{ij}$ cases could possibly be avoided by raising a safety signal. While the logic above is stated in terms of log risk ratios, recall that the log risk ratio and log odds ratio are approximately equal for a rare event. Defining $M_{II} := c_{ij}m_{ij}$, where c_{ij} is the economic cost of one case of our ADR j , we obtain the expected type II error cost by integrating over effect sizes $\beta \sim \mathcal{N}(\hat{\mu}_2^S, (\sigma_2^S)^2)$:

$$\mathcal{L}_S(\hat{\mu}_2^S, \sigma_2^S, 0) = M_{II} \int_0^\infty (1 - \exp(-x)) \frac{1}{\sigma_2^S} \phi\left(\frac{x - \hat{\mu}_2^S}{\sigma_2^S}\right) dx. \quad (\text{D.3})$$

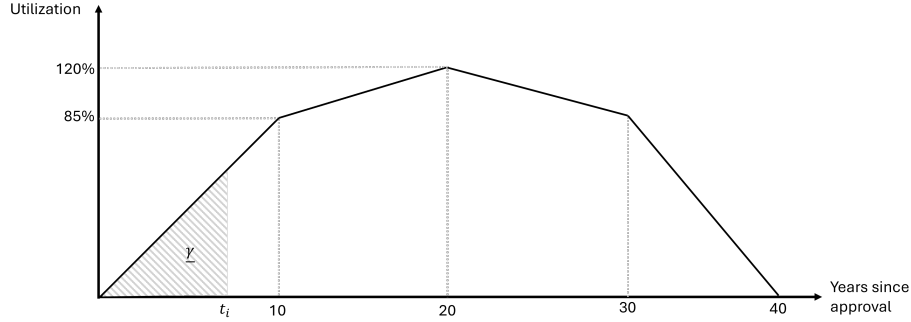
For symmetry, we consider a Type I error of the same functional form:

$$\mathcal{L}_S(\hat{\mu}_2^S, \sigma_2^S, 1) = M_I \int_{-\infty}^0 (1 - \exp(x)) \frac{1}{\sigma_2^S} \phi\left(\frac{x - \hat{\mu}_2^S}{\sigma_2^S}\right) dx. \quad (\text{D.4})$$

One interpretation is that these represent smoothed versions of the lump-sum loss function ($k = 0$), capturing the fact that raising a safety signal may avert some fraction of, but not all, ADRs. The two key parameters to estimate are the costs of mistakenly signaling M_I and missing a signal M_{II} . We focus on estimating M_{II} and perform sensitivity analyses with respect to M_I .

The cost of missing a sign M_{II} —the harm cost—depends on the number of patients who would experience the adverse reaction (m_{ij}) and the corresponding cost of experiencing the ADR (c_{ij}). We estimate the expected number of patients who would experience the adverse reaction (if no signal is raised) by projecting the number of ADR cases that are yet to occur based on the age of the drug (*i.e.*, time since entering the market), the number of reports so far observed (adjusting for FAERS underreporting), and a standard drug utilization pattern. We quantify the cost of an ADR based on its severity.

A drug life-cycle can span over 30 years (Danzon and Kim 2002), so we assume that a molecule stays in the market for 40 years on average. Danzon and Kim (2002) and Duflos and Lichtenberg (2012) identify an inverted U-shape molecule utilization pattern with a market protection period between 8 and 12 years. Examples can be found in Figure 5 and Figure 6 of Duflos and Lichtenberg (2012) and Figure 2a of Danzon and Kim (2002). Therefore, we assume that the utilization of a molecule, including both brand-name and generic use, follows the pattern in Figure D10. The area under the curve, denoted by A , represents the theoretical cumulative utilization of the molecule

Figure D10 Lifetime of molecule utilization (brand-name + generic)

Note: The y-axis values 85% and 120% are relative to year 14 since market introduction, extrapolated from [Duflos and Lichtenberg \(2012\)](#) and [Danzon and Kim \(2002\)](#).

over its life cycle. Let us consider a drug with time since approval t . Then, given t , we define $\underline{\gamma}(t)$ as the theoretical cumulative utilization up to time t , namely,

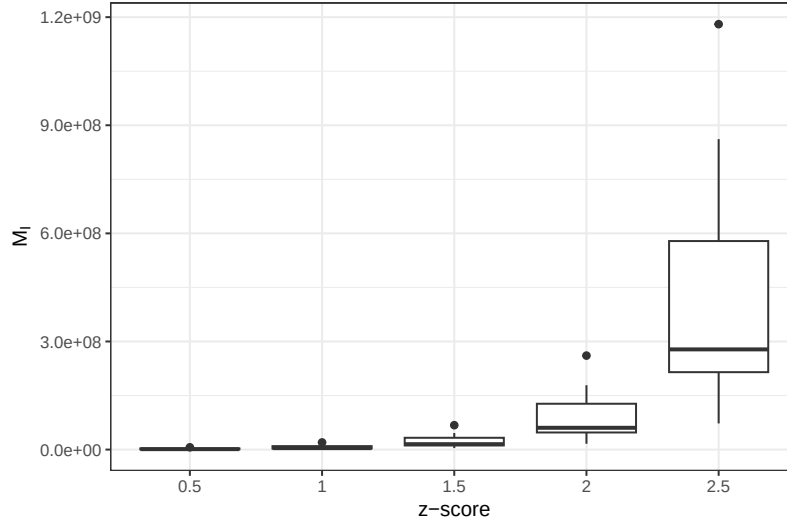
$$\underline{\gamma}(t) = \begin{cases} \frac{0.85 \times t^2}{20} & \text{if } t \in (0, 10] \\ \frac{0.85 \times 10}{2} + \frac{1}{2} (0.85 + 0.035t + 0.5) \times (t - 10) & \text{if } t \in (10, 20] \\ \frac{0.85 \times 10}{2} + \frac{1}{2} (0.85 + 1.2) \times 10 + \frac{1}{2} (1.2 - 0.035t + 1.9) \times (t - 20) & \text{if } t \in (20, 30] \\ \frac{0.85 \times 10}{2} + \frac{1}{2} (0.85 + 1.2) \times 20 + \frac{1}{2} (0.85 - 0.085t + 3.4) \times (t - 30) & \text{if } t \in (30, 40]. \end{cases}$$

Similarly, we define $\bar{\gamma}(t) := A - \underline{\gamma}(t)$ as the theoretical cumulative utilization from time t until the end of the molecule lifetime. Then, for a given molecule or drug i , with time since approval t_i , we project the number of patients who would experience ADR j as follows:

$$m_{ij} = \frac{\bar{\gamma}(t_i)}{\underline{\gamma}(t_i)} \times \frac{r_{ij}}{6\%}, \quad (\text{D.5})$$

where r_{ij} corresponds to the cumulative number of FAERS reports since approval, and the 6% adjustment captures the underreporting rate in FAERS ([Hazell and Shakir \(2006\)](#)). Note that r_{ij} and t_i can be computed from FAERS data. The fraction $\frac{\bar{\gamma}(t)}{\underline{\gamma}(t)}$ scales the future number of reports, so small fractions indicate that most reports have already been observed, while larger numbers indicate that most reports are yet to come, based on the drug's age and the theoretical utilization pattern in [Figure D10](#). In this calculation, we assume the ADR reporting rate only varies based on the theoretical drug utilization rate. However, in practice, it has been reported that ADR reporting increases in the early years since approval ([Hoffman et al. \(2014\)](#)).

Estimating the costs of experiencing an ADR is difficult as it involves non-medical, indirect, and opportunity costs for patients and society. Since we study many ADRs, it is infeasible to estimate each cost separately. Instead, we extrapolate costs based on the severity of the ADR. We assign a cost of US\$10,000, US\$5,000, and US\$1,000 depending on whether the ADR is high, medium, or low severity.¹⁵ The following literature informs this cost range: [Trumbo et al. \(2021\)](#) reports per-patient costs of ADRs associated with iron products to about US\$9,000 for high-severity ADRs and about

Figure D11 Summary of estimated drug-specific M_I for each z-score

Note. Each boxplot consists of one value for each unique drug.

US\$2,000 for low-severity ones. [Maity and Longo \(2021\)](#) reports average costs (all ADRs) for TNF- α inhibitor biologics on the order of CAD\$5,000, and [Tissot et al. \(2022\)](#) report costs on the order of €4,000 for ADRs that resulted in hospital admission. [Pirmohamed et al. \(2004\)](#) reports median ADR costs associated with length of stay, avoidability, and outcome on the order of US\$3,000. The costs reported in these studies expand to non-medical, indirect, and opportunity costs. Thus, the cost of missing a signal is $M_{II} := m_{ij} \times c_{ij}$, where c_{ij} represents the drug i per-patient cost of experiencing ADR j .

To obtain the cost of false positive signals, M_I , we consider various z-score values $z \in \{0.5, \dots, 2.5\}$ and solve for the value of $M_I(z)$ that makes the expected loss of signaling and not signaling equal under the specific z (*i.e.*, apply the optimality condition in [Lemma 1](#)). To minimize the impact of outliers, in our final evaluation, we assume that the cost of raising a false signal for a given drug does not vary based on the specific ADR for which we are raising the signal. Basically, raising a false signal can be more costly on a popular drug versus an unpopular one, but it does not depend on what we signal for. Then, for each drug, we use the median value, denoted by $\bar{M}_I(z)$, across all ADRs. For a z-score of 2, $\bar{M}_I(z) \in [1.6e+07, 2.6e+08]$ with a median of $6.0e+07$ across all PTs. [Figure D11](#) summarizes the estimated M_I values for each z-score.

D.2. Computing r_{ij} and t_i from FAERS data

The cumulative number of FAERS reports r_{ij} for each drug-ADR pair can be obtained directly from FAERS data by counting all reports up to 2020. The time since approval t_i can be inferred as follows: for each drug i , we identify the year of the first-percentile report as the year of the drug's approval and measure the drug's age as of 2020, *i.e.*, $t_i := 2020 - \text{approval year}$.

D.3. Limitations

Our procedure has several limitations. First, we estimate the cost of missing a safety signal, M_{II} , based on ADR severity, associated costs, and projected future cases. This projection relies on factors such as drug age, market lifespan (assumed to be 40 years), cumulative FAERS cases, and an assumed 6% FAERS reporting rate. However, this approach may inflate projections for certain drugs, particularly those recently approved or receiving heightened public attention due to lawsuits.

Additionally, our projections assume a 40-year drug lifespan, which may overstate M_{II} . To address this, we examine shorter lookahead periods (5 and 10 years), finding that while absolute cost estimates change, the relative performance of our approach remains consistent (results not shown).

Appendix References

- Danzon PM, Kim JD (2002) The life cycle of pharmaceuticals: A cross-national perspective. Technical report, Office of Health Economics, London UK.
- Duflos G, Lichtenberg FR (2012) Does competition stimulate drug utilization? the impact of changes in market structure on us drug prices, marketing and utilization. *International Review of Law and Economics* 32(1):95–109.
- Hazell L, Shakir SA (2006) Under-reporting of adverse drug reactions. *Drug safety* 29(5):385–396.
- Hoffman KB, Dimbil M, Erdman CB, Tatonetti NP, Overstreet BM (2014) The weber effect and the united states food and drug administration’s adverse event reporting system (faers): analysis of sixty-two drugs approved from 2006 to 2010. *Drug safety* 37:283–294.
- Maity T, Longo C (2021) Pragmatic pharmacoeconomic analyses by using post-market adverse drug reaction reports: an illustration using infliximab, adalimumab, and the canada vigilance adverse reaction database. *BMC Health Services Research* 21:1–12.
- Pirmohamed M, James S, Meakin S, Green C, Scott AK, Walley TJ, Farrar K, Park BK, Breckenridge AM (2004) Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients. *Bmj* 329(7456):15–19.
- Tissot M, Valnet-Rabier MB, Stalder T, Limat S, Davani S, Nerich V (2022) Epidemiology and economic burden of “serious” adverse drug reactions: Real-world evidence research based on pharmacovigilance data. *Therapies* 77(3):291–300.
- Trumbo H, Kaluza K, Numan S, Goodnough LT (2021) Frequency and associated costs of anaphylaxis-and hypersensitivity-related adverse events for intravenous iron products in the usa: an analysis using the us food and drug administration adverse event reporting system. *Drug safety* 44:107–119.

that debias the DA estimate of a drug’s effect on a rare ADR using the DA estimate for a more common ADR, as we propose. Our strategy could enhance any existing DA methods in practice.

There is substantial literature on using longitudinal data sources for postmarketing surveillance, such as electronic health record (EHR) data and medical claims data. These data are richer than spontaneous reporting databases, containing detailed patient covariates (EHR data) and longitudinal records of drug exposures and adverse events. Leveraging this richness, numerous methods incorporating temporal trends and time-varying covariates have been developed to investigate safety signals (Arnaud et al. 2017), with evidence from the U.S. (Ryan et al. 2013) and Europe (Schuemie et al. 2013) showing strong performance. In the operations literature, Ahuja et al. (2021) use logistic regression and matching designs to assess drug–ADR associations, demonstrating that the inaccurate black-box warning on rosiglitazone could have been avoided. Since such methods rely on careful cohort construction, covariate selection, and matching, they are mainly used to confirm rather than automatically generate new signals; for instance, the FDA has used EHR data accessed via its Sentinel system this way. Our work complements this stream by focusing on generating higher-quality signals for later evaluation with longitudinal databases.

Though much of the literature for longitudinal databases of health records relates to confirming safety signals, there is also significant literature on generating new safety signals (our task of “hypothesis generation”) from a variety of observational data sources, including EHR data. Sequential statistical testing (SST) can be used to monitor ADRs using longitudinal health record data, with classical approaches such as the sequential probability ratio test (SPRT) (Wald 1945) and maximized SPRT (Brown et al. 2007). An exemplary example of this type of approach within the operations literature can be found in Goh et al. (2015), where the authors propose an algorithm that can be used for continuous monitoring of multiple potentially correlated ADRs; their approach provides Type I error control, is robust to distributional assumptions, and takes into account the temporal sequence of using the drug and experiencing the ADR. Data sources beyond spontaneous reporting databases or longitudinal health records have also been considered to generate new safety signals—these include molecular fingerprints (Vilar et al. 2011), medical literature findings (Xu and Wang 2014, Shetty and Dalal 2011), and social media data (Liu and Chen 2015, Yeleswarapu et al. 2014, Li et al. 2020). In this literature, authors often propose to obtain signals from multiple data sources and combine them to obtain more confidence before raising a final safety signal (Li et al. 2020, Harpaz et al. 2013). Given its scale, spontaneous reporting data is likely to remain an important data source for safety signaling, either as a standalone source or in combination with others, motivating our focus on improving signaling quality by removing bias.

Methodologically, our work draws on causal inference research combining observational and experimental data, including reweighting RCT data (*e.g.*, Hartman et al. (2015)), extrapolating