

Detecting adverse high-order drug combinations from individual case safety reports using computational statistics on disproportionality measures

Jules Bangard¹ * Einar Holsbø²  Kristian Svendsen³  Vittorio Perduca⁴  Étienne Birmelé¹ 

¹Institut de Recherche Mathématique Avancée, UMR 7501 Université de Strasbourg et CNRS 7 rue René-Descartes, 67000 Strasbourg, France ²Faculty of Science and Technology, UiT-The Arctic University of Norway, PO, Box 6050 Stakkevollan, N-9037 Tromsø, Norway ³Faculty of Health Sciences, UiT the Arctic University of Norway, Tromsø, Norway ⁴CNRS, MAP5, Université Paris Cité, F-75006 Paris, France

Date published: 2026-07-08 Last modified: 2026-07-08

Abstract

Adverse drug reactions linked to the intake of drug combinations are a critical concern in pharmacovigilance, particularly as the controlled environment of clinical trials often lacks the scale and diversity to detect rare events involving multiple medications. While spontaneous reporting systems provide the necessary breadth for post-market surveillance, identifying over-represented drug cocktails within such large-scale data remains a significant computational challenge. This study introduces a computational framework for the detection of drug cocktails associated with adverse events, leveraging disproportionality analysis on individual case safety reports. By integrating the Anatomical Therapeutic Chemical classification, the framework extends beyond individual drugs to capture hierarchical pharmacological relationships, enabling exploration of the space of drug combinations beyond pairwise analysis. To address biases inherent in existing disproportionality measures, we employ a hypergeometric risk metric, while a Markov Chain Monte Carlo algorithm provides robust empirical p-value estimation for the risk associated with cocktails. A genetic algorithm further facilitates efficient identification of high-risk drug cocktails. A post-treatment step based on penalized logistic regression allows distinguishing true pharmacological interactions from combined individual effects for cocktails of any size. Validation on synthetic and FDA Adverse Event Reporting System data demonstrates the method's efficacy in detecting established drugs and drug combinations associated with myopathy-related adverse events. Implemented as an R package, this framework offers a reproducible, scalable tool for post-market drug safety surveillance.

Keywords: Disproportionality Analysis, Drug-Combinations, Drug-Interactions, Genetic Algorithm, Markov Chain Monte Carlo, Pharmacovigilance, Spontaneous Reporting Systems

1 Contents

2 1 Introduction

2

*Corresponding author: jules.bangard@math.unistra.fr

3	2 Methods	4
4	2.1 ATC Tree and Cocktail Definition	4
5	2.2 Cocktail Risk Characterization	5
6	2.2.1 Scoring Functions in Pharmacovigilance	5
7	2.3 Distinguishing Combined Risk from True Interaction	5
8	2.4 Disproportionality Identification	6
9	2.4.1 High-risk Drug Cocktails Identification	6
10	2.4.2 Distance Between Drug Cocktails and Cocktail Penalization	7
11	2.4.3 Output Clustering	8
12	2.5 Approximate P-Value Assignment for Drug Cocktails	9
13	2.6 Datasets	10
14	2.6.1 Simulated Data	10
15	2.6.2 FAERS Data	11
16	3 Results	11
17	3.1 Score Comparison	11
18	3.2 Application to Dataset D_1	12
19	3.2.1 Estimation of Risk Distribution	12
20	3.2.2 Genetic Algorithm Output and Clustering	12
21	3.3 Application to the FAERS Spontaneous Reporting Data	15
22	3.3.1 Estimation of Risk Distribution	15
23	3.3.2 Genetic Algorithm Output and Clustering	15
24	3.3.3 Post-Treatment Interaction Analysis	17
25	4 Conclusion	18
26	Funding	19
27	5 Appendices	19
28	5.1 Appendix A : Distance Pseudo-code Algorithm	19
29	5.2 Appendix B : Calibration Under the Null	20
30	5.3 Appendix C : FAERS Clustering	20
31	5.4 Appendix D : Algorithmic Complexity	20
32	5.5 Appendix E: Illustrative Example of Package Usage	21
33	5.5.1 E.1 Setup and Toy Dataset Generation	21
34	5.5.2 E.2 Evaluating Cocktail Risk via the Hypergeometric Score	22
35	5.5.3 E.3 Estimating the Null Distribution via MCMC	22
36	5.5.4 E.4 Identifying High-Risk Cocktails with the Genetic Algorithm	23
37	5.5.5 E.5 Output Clustering	23
38	5.5.6 E.6 Post-Treatment with Firth’s Penalized Logistic Regression	24
39	6 Code and Results	25
40	References	25
41	Session information	28

42 1 Introduction

43 There are inherent limitations of clinical trials (RCTs) done before a medication is authorised in
44 terms of cost, surveillance time, size and a lack of diversity of patients included (e.g. that pregnant

45 women, children and elderly often are not included) (Sanson-Fisher et al. 2007). Another issue is
46 that RCTs are limited in their ability to detect rare adverse drug reactions (ADRs) and potential drug
47 interactions (DDIs) since they usually only include a few thousand participants at most. Moreover,
48 the concurrent use of multiple medications is often an exclusion criterion in RCTs, leaving gaps in
49 the assessment of polypharmacy risks. As a result, tested drug combinations tend to be limited to
50 those where physicians and pharmacologists have existing clinical experience or hypotheses about
51 potential interactions (Heijden et al. 2002). This means that monitoring of the safety of medications
52 and the potential for DDIs after marketing is essential. This monitoring is often referred to as
53 post-authorization pharmacovigilance.

54 Elderly individuals are commonly prescribed multiple medications due to the increasing prevalence
55 of comorbidities associated with aging. One study suggested that patients aged 75 or older in Austria
56 took on average 7.5 drugs (Schuler et al. 2008). Among this demographic, 10% of hospitalizations
57 were attributed to ADRs, and 18.7% of these ADRs were caused by a combination of more than one
58 drug, a drug-drug interaction. More broadly, a meta-analysis attributed approximately 22.2% of
59 hospitalizations caused by adverse drug reactions to DDIs (Dechanont et al. 2014).

60 Disproportionality analysis (DA) encompasses widely used methods in pharmacovigilance to detect
61 ADRs by assessing the frequency of adverse events (AE) associated with specific drug consumption
62 relative to what would be expected. Most established disproportionality methods like the proportional
63 reporting ratio (PRR) (Evans et al. 2001), reporting odds ratio (ROR) (Puijtenbroek et al. 2002), Bayesian
64 Confidence Propagation Neural Network (BCPNN) (Bate et al. 1998), and the Gamma Poisson Shrinker
65 (GPS) (DuMouchel 1999) allow for the identification of signals that warrant further investigation
66 in the context of single drug assessment. The tree-based scan statistic (Kulldorff et al. 2003; Heo et
67 al. 2024) takes advantage of the tree structure of the Anatomical Therapeutic Chemical (ATC) drug
68 classification in order to allow to select a drug family rather than a single drug, a family corresponding
69 to a subtree of the ATC tree. Likelihood-ratios tests then allow to choose the most relevant family.

70 In order not to overinterpret high proportions in case of rare drug cocktails, it is possible to derive
71 a score from an independence test on the contingency table counting the presence or absence of
72 the adverse event depending on the consumption of the drug cocktail. (Gosho et al. 2017) use a
73 chi-square statistic to do so, while (Ahmed et al. 2010) rather use the Fisher exact test which has the
74 advantage of being non-asymptotic.

75 Fewer DA methods are available to use for multiple drug consumption, such as the Ω shrinkage
76 method (Norén et al. 2008). This method allows the use of a DA measure in order to detect sets of
77 the type *Drug-Drug-Adverse Event* and stands as the standard measure when it comes to detecting
78 the interactions of two drugs. Adaptations of the PRR for a single drug has been proposed like the
79 Concomitant Signal Score (CSS) (Noguchi et al. 2020) and the PRR adaptation for DDIs (Wang et al.
80 2020).

81 Other methods than DA measures exist in order to detect DDIs. Among them, well known methods
82 encompass logistic models (Van Puijtenbroek et al. 2000, 1999) and association rules methods (Noguchi
83 et al. 2018; Ibrahim et al. 2016). One can refer to multiple reviews for a more detailed overview of
84 existing methods (Ibrahim et al. 2021; Hauben 2023).

85 Computational pharmacovigilance methods are typically applied to Spontaneous Reporting Systems
86 (SRS), which aggregate Individual Case Safety Reports (ICSRs)—real-world data submitted by health-
87 care professionals, manufacturers, and patients. Such reports contain information about intake of
88 medications for each patient as well as their experienced AE. While SRS databases, such as the FDA
89 Adverse Event Reporting System (FAERS), provide the scale necessary to detect rare events, they
90 possess inherent limitations. These include under-reporting, where only a fraction of adverse events
91 are captured (Bate and Evans 2009), and various reporting biases, such as the notoriety bias (Pariente

92 et al. 2007). Crucially, the spontaneity of these reports often results in an incomplete recording of all
93 drugs taken by a patient. Furthermore, because the exact timing of drug administration is frequently
94 omitted, it is difficult to distinguish between simultaneous intake and medications taken at different
95 times within the same reporting window. This temporal ambiguity potentially limits the precision of
96 interaction analyses, as the observed cocktail may represent a sequence of exposures rather than
97 a concurrent pharmacological event. Consequently, because the total number of patients exposed
98 to a drug is unknown, these databases cannot provide true incidence rates, necessitating the use of
99 disproportionality analysis to identify safety signals (Almenoff et al. 2005).

100 While previous work showed that further practical experiences should be of interest for high-order
101 drug interaction testing (Tekin et al. 2018), it is hard to know which exact combination one should test
102 and easy to miss an important combination in the overwhelming space of high-order drug cocktails.
103 The present article proposes a novel method to identify drug cocktails that are overrepresented in
104 adverse event reports, which constitute candidates for further interaction assessment. It requires the
105 choice of a scoring function able to measure, for a given AE, the importance of the disproportionality
106 for any medication set, including single medications. It then takes advantage of the tree structure
107 of the ATC classification to explore the space of medication sets in two ways: a genetic algorithm
108 looking for the sets of high score, and a MCMC algorithm learning the distribution of the scores
109 for a given cocktail size in order to evaluate how extreme a high score is. A post-treatment step
110 using penalized logistic regression then allows to assess whether an identified cocktail reflects a true
111 pharmacological interaction or merely the combined effect of its components. The approach is also
112 original with respect to the fact that a medication denotes either a drug or a drug family in the sense
113 of an internal node of the ATC classification.

114 The developed algorithms, available in the corresponding R package **emcAdr** (Bangard 2025), aim to
115 guide pharmacology researchers toward drug combinations that might correspond to drug interac-
116 tions and require more rigorous monitoring.

117 2 Methods

118 2.1 ATC Tree and Cocktail Definition

119 Multiple classifications of drug active ingredients exist, one being the ATC classification. This system
120 is structured hierarchically and can be represented as a tree with five levels containing, at the time
121 of this study, 6809 nodes. The leaves of this tree are the active ingredients, while the first level
122 consists of nodes representing organs or systems affected by the descendant active ingredients. The
123 remaining nodes correspond to therapeutic or pharmacological families.

124 By applying a Depth-First Search algorithm to enumerate each node of the ATC tree, we can define a
125 drug cocktail as a set of integers corresponding to specific nodes in the tree. More formally, a cocktail
126 C of size $k \geq 1$ is defined by $C = (x_1, \dots, x_k) \in \Delta^k$ where $k \in \mathbb{N}$, and Δ represents the set of numbered
127 nodes of the ATC tree T . Figure 1 shows examples in a simplified tree, the green nodes denoting the
128 considered cocktail. Note that for convenience, we also refer to single drugs as cocktails.

129 Considered cocktails can include internal nodes of the tree (thus representing families of active
130 ingredients), allowing for the detection of more general signals. For example, paracetamol might
131 send a weak signal, while if we move up the tree, analgesics as a whole could represent a stronger
132 and more general signal. Therefore, all patients taking at least one drug from this drug family will
133 be considered in the computation of the risk. That notion is equivalent to the notion of cut in the
134 tree-scan method (Kulldorff et al. 2003).

2.2 Cocktail Risk Characterization

In the field of pharmacology, accurately characterizing the risk associated with drug administration is a complex task. The aim of the developed method is to search the space of cocktails to maximize a score indicating if individuals taking a given drug combination have a higher risk of an AE. While the algorithm is designed to be flexible and can be run with any scoring function, we focus on a comparison of established metrics to justify our selection (see Section 3.1).

2.2.1 Scoring Functions in Pharmacovigilance

To evaluate the disproportionality of drug cocktails presence for a particular AE, several statistical measures have been proposed in the literature:

Proportional Reporting Ratio (RR) A standard measure defined as the ratio of the probability of an AE occurring in the group exposed to the cocktail versus the non-exposed group. Its application in signal generation from spontaneous reports was notably discussed by (Evans et al. 2001).

PRR for Drug-Drug Interactions (PRR) Unlike the single-drug PRR, this adaptation for drug combinations evaluates whether a combination represents a synergistic risk. Specifically, as discussed by (Wang et al. 2020), a signal is often defined by comparing the lower bound of the confidence interval of the combination’s risk to the maximum risk observed for the individual drug components.

Ω Shrinkage A Bayesian measure specifically developed for drug-drug-event combinations. As proposed by (Norén et al. 2008), it utilizes a “shrinkage” approach to reduce false positive signals in cases with limited data by pulling the observed association toward the value expected under independence.

Concomitant Signal Score (CSS) Introduced as an improved detection criterion, this metric aims to enhance the detection of DDI signals by building upon the PRR framework, as detailed by (Noguchi et al. 2020).

Hypergeometric Disproportionality Score While the methodology can accommodate various scores, we propose the use of a score derived from the Fisher exact test. As highlighted by (Ahmed et al. 2010), this approach has the advantage of being non-asymptotic. Consider a dataset of N patients, among which K experience the adverse event AE. Let n_C be the number of people taking a cocktail C and x the number of patients taking C and experiencing AE. We define the risk to be:

$$H(C) = -\log(\mathbb{P}(X \geq x))$$

where $X \sim \mathcal{H}(n_C, K, N)$. This measure effectively functions as the log p-value under the null hypothesis that the number of people with AE in a uniform sample of n_C out of N people follows the hypergeometric distribution. It has the distinct advantage of accounting for the total number of patients exposed to the cocktail; for instance, $H(C)$ will assign a higher risk to a cocktail taken by 100 patients with 50 AE cases than to a cocktail taken by 10 patients with 5 AE cases, whereas simple ratios like the RR might treat them identically. Such hypergeometric measures are well-established in other high-dimensional fields like bioinformatics for functional enrichment analysis (Grossmann et al. 2007).

2.3 Distinguishing Combined Risk from True Interaction

While the $H(C)$ score effectively identifies cocktails associated with a significant increase in the AE frequency, it does not inherently distinguish a synergistic pharmacological interaction from

177 situations corresponding to the independent addition of individual effects or an effect driven by a
178 subset of the cocktail (e.g., “innocent” medications co-prescribed with a high-risk drug).

179 To address this, we propose a post-treatment of detected signals using Firth’s penalized likelihood
180 logistic regression (Firth 1993). This method is particularly suited for spontaneous reporting systems
181 as it reduces bias in maximum likelihood estimates and provides a solution to the problem of
182 separation, which often occurs with rare adverse events. Using the *logistf* R package (Heinze et al.
183 2023), we estimate a model for a given cocktail C (e.g., d_1, d_2) as follows:

$$\text{logit}(P(AE = 1)) = \beta_0 + \beta_1 \mathbb{1}_{d_1} + \beta_2 \mathbb{1}_{d_2} + \beta_3 \mathbb{1}_{d_1 \cap d_2}$$

184 An interaction is confirmed if the interaction coefficient β_3 is positive and statistically significant.
185 This indicates that the risk associated with the co-prescription exceeds the additive effect of the
186 individual components on the logit scale. While the example above illustrates the case of a size-two
187 cocktail for simplicity, this post-treatment generalizes naturally to cocktails of any size k by including
188 all $2^k - 1$ interaction terms in the model, or by testing nested models comparing the full cocktail to
189 its sub-cocktails. Users may implement the post-treatment of their choice.

190 2.4 Disproportionality Identification

191 2.4.1 High-risk Drug Cocktails Identification

192 The number of possible cocktails is 2^L , where L denotes the number of nodes in the ATC tree, and
193 even for a given cocktail size k , the number of possibilities is still $\binom{L}{k}$. It is therefore not possible to
194 compute $H(C)$ for each possible cocktail C included in the dataset. Instead, to explore the space of
195 cocktails and locate those at high-risk of AE, we use a genetic algorithm (Pétrowski and Ben-Hamida
196 2017). It simulates the evolution of a population of cocktails according to the principle of natural
197 selection, in order to search for the most performing individuals with respect to an evaluation
198 criterion based on H .

199 The steps are the following, the algorithm repeating steps 2 and 3 until a user-defined number of
200 iterations is reached.

201 **Initialization.** The genetic algorithm’s population consists of m cocktails. These cocktails are
202 randomly initialized and can vary in size.

203 **Evaluation and selection.** At iteration n , the current population P_n undergoes an evaluation and
204 selection phase. The evaluation computes the score $H(C)$ for each cocktail C in the population.
205 A new population Q_{n+1} of the same size is drawn by sampling m times a pair of cocktails in
206 the original population and copying the one with highest score. Note that this step performs
207 selection as the expectation of the number of copies in Q_{n+1} of a cocktail from P_n is an
208 increasing function of its score. A penalization is however applied to avoid a uniformization
209 of the population, as further explained in Section 2.4.2.

210 **Stochastic modification.** Mimicking the genetic drift in nature, stochastic modifications occur in
211 the population in order to explore the large space of cocktails. First, a crossover operation allows
212 two cocktails to exchange information. Here, the crossover involves exchanging sub-trees
213 between two cocktails as follows:

- 214 • an internal node v of the ATC tree is randomly selected;
- 215 • the nodes of the subtree rooted at v are exchanged between the two sequences.

216 After performing the crossover, a mutation is applied to the resulting individuals, chosen from
 217 two types. The first type is a local mutation which changes a randomly selected node of the
 218 cocktail to one of its free neighboring nodes. This mutation is further explained in Section 2.5.
 219 The second type is an addition/deletion mutation which operates as follows, with k being the
 220 cocktail length and α a chosen hyperparameter:

- 221 • with probability $\min(1, \frac{\alpha}{k})$, a node uniformly drawn from Δ is added to the sequence;
- 222 • with probability $1 - \min(1, \frac{\alpha}{k})$, a uniformly drawn node from the cocktail is removed.

223 An example of crossover and mutations can be seen on Figure 1 (a), (c) and (d).

224 Applying those modifications to Q_{n+1} yields a population P_{n+1} which is used to loop at step
 225 $n + 1$.

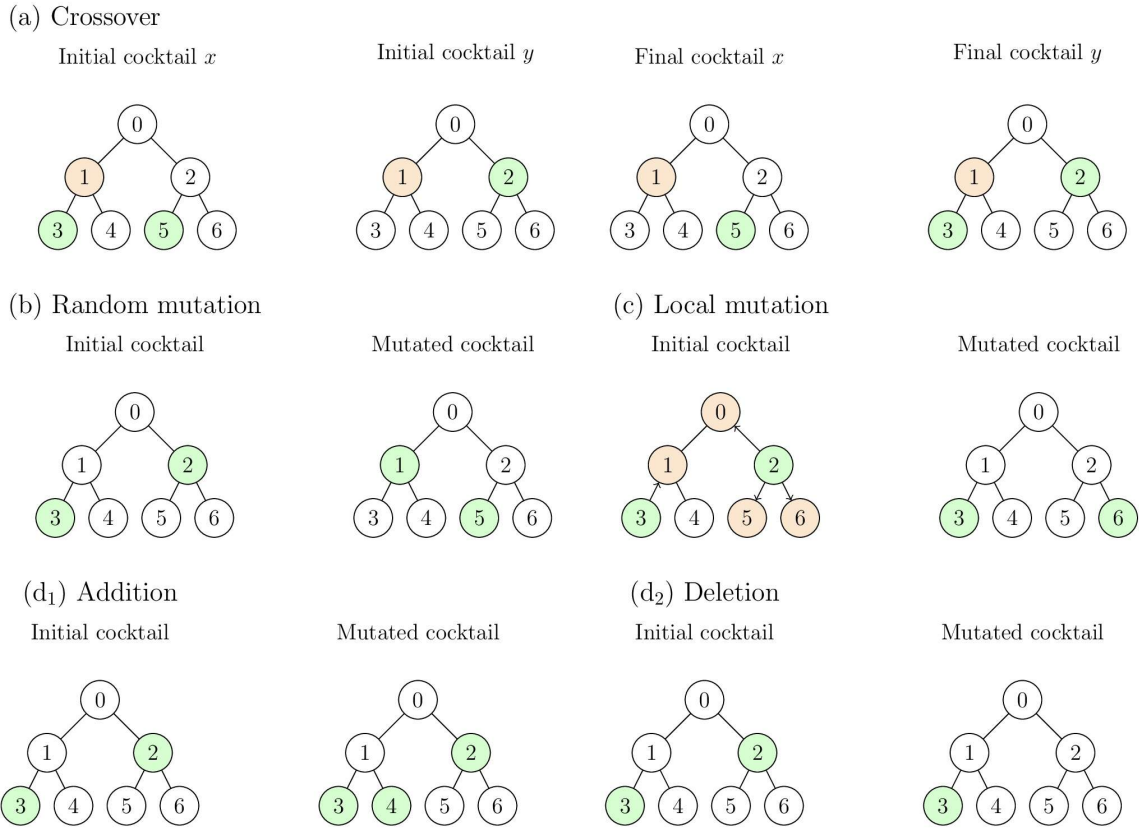


Figure 1: Cocktail modifications used for the genetic algorithm (a, c and d) and the MCMC algorithm (b and c). Green nodes are part of considered cocktails. In Crossover, the orange node represents the selected internal node whose subtrees are being swapped. In local mutations, orange nodes represent legal moves

226 2.4.2 Distance Between Drug Cocktails and Cocktail Penalization

227 If at some point of the algorithm, an individual of P_n has a high score compared to other individuals,
 228 the genetic algorithm may converge to a population consisting entirely of that cocktail. To avoid
 229 this uniformization phenomenon, similar cocktails are penalized in the evaluation phase as follows,

$$H_{pen}(C) = \frac{H(C)}{\sum_{C_i \in \mathcal{C}} \text{Sim}(C, C_i)}$$

230 The computation of the similarity $\text{Sim}(C, C')$ is based on a distance inspired by the Levenshtein
 231 distance (Levenshtein et al. 1966). However, unlike the traditional Levenshtein distance, sequences
 232 are treated as unordered sets. For two drug cocktails, C_1 and C_2 , of sizes n_1 and n_2 , the distance
 233 $d(C_1, C_2)$ is defined as the minimal cost required to transform C_1 into C_2 using three elementary
 234 operations.

- 235 • $\text{Ins}_a(C)$ consists of adding a to the cocktail C .
- 236 • $\text{Del}_a(C)$ consists of deleting a from the cocktail C .
- 237 • $\text{Sub}_{a,b}(C)$ consists of substituting $a \in C$ by b .

238 An associated cost is defined for each operation.

239 **Substitution.** The cost associated with the substitution operation is chosen to be consistent with
 240 the conceptual similarity of cocktails. If a is a drug belonging to C , the cost should increase as
 241 the drug b diverges further from drug a . For example, if a is a drug, and b a drug family that
 242 contains a , the cost should be moderate. Conversely, if b is a drug family not containing a the
 243 cost should be higher. This distance is thus defined by the maximal distance between a and b
 244 to their Lowest Common Ancestor.

245 **Insertion, Deletion.** The deletion and insertion cost are chosen as $\frac{\text{depth}(T)}{2}$. This choice implies
 246 that a substitution always costs less than a deletion followed by an insertion. The latter are
 247 used only when the two cocktails do not have the same length.

248 A transformation f from C_1 to C_2 is a composition of elementary operations that go from C_1 to C_2 .
 249 The associated cost $\text{cost}(f)$ is defined as the sum of the cost of the operations used in f . Finally,
 250 $d(C_1, C_2) = \min_f \text{cost}(f : f(C_1) = C_2)$.

251 Finally, the maximum distance between C_1 and C_2 being $(n_1 + n_2)\frac{\text{depth}(T)}{2}$, we define the similarity as

$$\text{Sim}(C_1, C_2) = 1 - \frac{2D(C_1, C_2)}{(n_1 + n_2)\text{depth}(T)}$$

252 The computation of the similarity is achievable in $\mathcal{O}(n_1 \times n_2 \times \text{depth}(T) + |\Delta|)$ operations in the worst
 253 case. The algorithm to compute the similarity is detailed in the Section 5.1 (see Algorithm 1).

254 2.4.3 Output Clustering

255 Despite the diversity mechanisms integrated into the genetic algorithm, not all drug cocktails
 256 retrieved within the genetic algorithm population are unique. It is moreover common to encounter
 257 solutions that are merely variations of others, differing only by transformations such as changing
 258 a node in the ATC tree to its parent or child. To streamline analysis and enhance the efficiency, a
 259 post-treatment clustering of similar drug cocktails is implemented. This method allows to focus on
 260 the most risky cocktails within each cluster or to interpret pharmaceutically clusters rather than
 261 individual cocktails.

262 To do so, drug cocktails are embedded into a two-dimensional space using the UMAP algorithm
 263 (McInnes et al. 2018), which aims to preserve similarity in the latent space. This representation
 264 enables the effective use of conventional machine learning clustering algorithms in \mathbb{R}^2 . Specifically,
 265 the DBSCAN algorithm (Ester et al. 1996) is applied to identify clusters of similar drug cocktails with
 266 an intuitive way of choosing hyperparameters.

2.5 Approximate P-Value Assignment for Drug Cocktails

Once a list of cocktails with high score has been established by the genetic algorithm, an important step of the analysis is to assign them p-values to decide if they are significant.

Consider a cocktail C of size k in the list and its score S_{obs} . Denote by H_0 the null hypothesis according to which a cocktail does not favor the AE, and by N_0 the number of cocktails of size k for which H_0 is the truth. Similarly, H_1 represents the alternative hypothesis of a cocktail favoring the AE and N_1 is the number of such cocktails of size k . The p-value corresponding to S_{obs} is then $\mathbb{P}_{H_0}(S > S_{obs})$, but cannot be estimated without a modelling hypothesis under H_0 .

However, if $\mathbb{P}_M(S)$ refers to the marginal distribution, that is the score distribution for the cocktails both in H_0 and H_1 ,

$$\mathbb{P}_M(S > S_{obs}) = \mathbb{P}_{H_0}(S > S_{obs}) \frac{N_0}{N_0 + N_1} + \mathbb{P}_{H_1}(S > S_{obs}) \frac{N_1}{N_0 + N_1}$$

so that

$$\frac{N_0 + N_1}{N_0} \mathbb{P}_M(S > S_{obs}) - \frac{N_1}{N_0} \mathbb{P}_{H_1}(S > S_{obs}) \leq \mathbb{P}_{H_0}(S > S_{obs}) \leq \frac{N_0 + N_1}{N_0} \mathbb{P}_M(S > S_{obs})$$

Under the reasonable assumption $N_1 \ll N_0$, the probability $\mathbb{P}_M(S > S_{obs})$ on the marginal distribution can thus be used as an approximation of the p-value. Note that the upper bound yields that, under the weak hypothesis that less than 10% of the cocktails of size k favor the AE, the real p-value is at most equal to 1.11 times its approximation. Moreover, the approximated p-value can be estimated by computing the score of cocktails sampled in the general cocktail population. The proposed method therefore focuses on a sampling scheme in the general population and uses the approximated p-value to declare significance.

A naive sampling of cocktails considers almost only cocktails taken by no patient in the dataset. A Metropolis-Hastings MCMC algorithm is thus considered, as it can be used by conditioning on the fact that all visited cocktails are present in the dataset (Au and Beck 2001).

To employ such an algorithm, it is necessary to define a state space $\mathcal{C} = \{C_1, \dots, C_p\}$, a computable target measure $f(C_i)$, and conditional laws $q(\cdot|C_i)$ under which simulation is possible and new states can be proposed.

State set. The state set is made of all cocktails of k drugs for a fixed k .

Proposal law. The proposal law is defined as a mixture of two mutation laws of the current cocktail. They operate as follows:

- Random mutation consists of a completely random movement in the cocktail space.
- Local mutation involves a local movement relative to the structure of the drug tree. Here, a node x_p of the state C_i is changed to one of its free neighboring nodes.

At each iteration, the random and local mutations have probability p_R and $1 - p_R$, where p_R is a hyperparameter.

Figure 1 (b, c) presents examples of a random and a local mutation.

State evaluation. The evaluation of a drug cocktail is based on the score $H(C)$. The chosen target measure is then:

$$f_T(C_i) = \frac{1}{Z(T)} \times e^{\frac{H(C_i)}{T}}$$

302 where $Z(T) = \sum_C e^{\frac{H(C)}{T}}$. T is a parameter known as temperature, which modulates space
303 exploration by more readily accepting cocktails with moderate scores (high T) or, conversely,
304 by strongly favoring combinations of drugs with high scores (low T).

305 The acceptance probability of cocktail C_{i+1} from cocktail C_i is given by:

$$\min\left(1, \frac{f_T(C_{i+1})}{f_T(C_i)} \times \frac{q(C_i|C_{i+1})}{q(C_{i+1}|C_i)}\right)$$

306 The theory related to the Metropolis-Hastings algorithm (Robert and Casella 2004) ensures that
307 the empirical distribution of $f_T(C_i)$ for the constructed cocktail chain converges to the distribution
308 of $f_T(C)$. A very long realization of such a walk, therefore, allows for the approximation of the
309 distribution that can be used to determine an empirical p-value for the score of a cocktail of interest.
310 This enables the possibility to say whether or not a high-risk is truly significant (e.g. among the top
311 5% of scores). It defines what a high-risk is in our case.

312 2.6 Datasets

313 2.6.1 Simulated Data

314 Multiple datasets were simulated to evaluate the method performance against known outcomes. The
315 datasets, designed to challenge the algorithm, simulate various patient scenarios. Each patient record
316 includes prescribed medications and the corresponding occurrence of an adverse event AE .

317 The first dataset (D_1) is composed of 200,000 patients, and has the following characteristics:

- 318 • 1% of the patients take a size-3 drug cocktail C_1 and have a $\frac{1}{100}$ chance of having AE .
- 319 • 1% take a size-3 drug cocktail C_2 and have a $\frac{1}{200}$ chance of having AE .
- 320 • 1% take a size-2 drug cocktail C_3 and have a $\frac{1}{100}$ chance of having AE .
- 321 • 1% take a size-2 drug cocktail C_4 and have a $\frac{1}{200}$ chance of having AE .

322 A small percentage of the dataset (1.5% per combination) are combinations of 2 out of the 3 drugs
323 from C_1 and C_2 , but with no risk of AE . This helps to mitigate the false identification of sub-cocktails
324 of C_1 and C_2 as high-risk cocktail because those who take two drugs of these cocktails will almost
325 surely take the remaining drug of the cocktail.

326 The remaining 87% of the dataset consists of patients assigned random cocktails drawn uniformly.
327 The size s of each cocktail is drawn according to a Poisson distribution with $\lambda = 4$ (mean size of drugs
328 cocktails taken by patients in the dataset). For each cocktail, s nodes of the ATC tree are selected
329 uniformly, with each combination assigned an adverse event with probability $\frac{1}{15000}$.

330 All simulations use the complete ATC classification tree comprising 6809 nodes across five hierarchical
331 levels at the time of this study. The cocktails C_1 through C_4 inducing an elevated AE probability
332 were chosen as nodes at level 4 of the ATC hierarchy (i.e., pharmacological subgroups, one level
333 above the leaf nodes representing individual active substances). This choice is deliberate: in the
334 simulated data, patients are assigned specific drugs at level 5 (leaves) as would occur in practice, but
335 any patient taking a drug that descends from one of the C_i nodes inherits the elevated AE probability
336 associated with that family. This design tests the method's ability to detect family-level signals – an
337 important feature, since the underlying pharmacological mechanism may apply to an entire drug
338 class rather than a single substance. The C_1 through C_4 nodes were selected to be non-overlapping
339 (i.e., belonging to distinct subtrees) and to have a sufficient number of descendant leaves to generate
340 realistic prescription patterns.

341 Three additional datasets were similarly constructed, varying the sizes of the cocktails inducing AE.
342 Dataset D_1 contains cocktails of sizes two and three as described above. Dataset D_2 contains only
343 size-two cocktails, Dataset D_3 only size-three cocktails, and Dataset D_4 contains size-two, three, and
344 four cocktails.

345 2.6.2 FAERS Data

346 As a proof-of-concept on real data, the method was also assessed using the FAERS dataset, which
347 consists of ICSRs submitted by healthcare professionals, consumers, and manufacturers. These
348 reports include details on patient drug intake and the side effects experienced. The methods were
349 deployed on FAERS data from the second quarter of 2013 to the second quarter of 2015, the restriction
350 to a two-year window being motivated by computational time constraints.

351 Significant refinement of the FAERS data was required. Duplicate reports were removed, retaining
352 only the report with the most recent ID. Subsequently, a link was established between the prescribed
353 drugs and the ATC codes of each active ingredient. This process involved matching drug names to
354 their respective active ingredients and converting these ingredients to their corresponding ATC codes.
355 The DiAna dictionary (Fusaroli et al. 2024) facilitated the standardization of FAERS drug names
356 to ATC codes. It is important to note that the DiAna dictionary handles combination products by
357 splitting them into their constituent active substances rather than assigning a single ATC combination
358 code. Reports with unmatchable drug names in the DiAna dictionary were excluded, reducing the
359 dataset from 2,043,231 to 1,612,931 patients.

360 For this study, we focused on myopathy as the selected adverse event outcome. It is a clinically
361 concerning condition with a sufficient number of reported cases in the dataset (536 cases). To validate
362 our results, we compared the identified drug-myopathy associations with known drugs already
363 established to cause myopathy (Miernik et al. 2024; Hall et al. 2011; Valiyil and Christopher-Stine
364 2010). Code for data refinement is available [on GitHub](#).

365 3 Results

366 3.1 Score Comparison

367 To support the use of the hypergeometric score $H(C)$ introduced in Section 2.2, we compared it to
368 other risk scores on Dataset D_2 , which comprises pairwise drug combinations. A set of cocktails was
369 generated under the alternative hypothesis H_1 , meaning they correspond to the drug combinations
370 specifically designed to increase the probability of AE in Dataset D_2 (see Section 2.6.1). A second set
371 of cocktails was sampled under the null hypothesis H_0 , representing a baseline risk without specific
372 interaction. For each cocktail in both sets, the risk scores were computed using the patient data in
373 Dataset D_2 . Figure 2 illustrates the performance of the risk scoring methods for detecting high-risk
374 drug combinations introduced in Section 2.2.1. Each subplot displays the score values for cocktails
375 representing true solutions, sampled under H_1 (green) and cocktails not representing true solutions,
376 sampled under H_0 (red). The bottom-right panel presents the Precision-Recall (PR) curve, comparing
377 the detection power of the scores in identifying high-risk cocktails. PRR is a special score because its
378 value is either one or zero, which allows the computation of only one value of precision and recall.
379 PRR is therefore represented by a single point rather than a PR curve.

380 PR curves of the different scoring methods are mainly comparable, confirming the conclusions of
381 (Candore et al. 2015). However, as illustrated by the jitter plots, the hypergeometric score and the Ω
382 Shrinkage measure better rank and isolate true solutions from other cocktails. It must be emphasized
383 that the Ω Shrinkage measure is difficult to compare using a threshold, as the original article (Norén
384 et al. 2008) suggests signaling a cocktail when the score exceeds zero. However, in the simulations,

385 the computed score never exceeded zero, as depicted by the x-axis of its jitter plot. Despite these
386 considerations, these two methods are the least biased toward cocktails taken by only a few patients.
387 The good behavior of the hypergeometric score and its easy generalization to cocktails of any size
388 justify the use of that scoring function for the exploration of the cocktail space.

389 **3.2 Application to Dataset D_1**

390 **3.2.1 Estimation of Risk Distribution**

391 Risk distribution was estimated for size-two drug cocktails on Dataset D_1 (Section 2.6.1). Estimation
392 of risk distribution for higher cocktail sizes is possible but it is nearly impossible to compare it to the
393 true distribution as it is computationally prohibitive to obtain. The distribution estimated by the
394 MCMC algorithm, is compared to the true risk distributions in Figure 3.

395 The left panels display the distributions of risk scores for both the estimated (top-left) and true
396 (bottom-left) risk values. Both distributions share a similar shape, with the majority of risk scores
397 concentrated at low values, as 95% of the scores fall below 11. However, some differences can
398 be observed, particularly in the tail of the distribution, where cocktails of higher risks are under-
399 represented in the estimated distribution.

400 The right panels of Figure 3 present a Probability-Probability (PP) plot (top) and a Quantile-Quantile
401 (QQ) plot (bottom), comparing the quantiles and probabilities of the estimated and true risk distri-
402 butions. While the right panel of the figure demonstrates good agreement at lower risks, where
403 most of the data lie, deviations at higher risk values suggest that the estimated distribution slightly
404 underrepresents the risk for more extreme values.

405 Results indicate that the method performs well in estimating risk scores for the majority of cocktails,
406 capturing the overall risk distribution with reasonable accuracy. The slight underestimations which
407 are present for high-risk cocktails are not a problem since the interest of the method is to assign
408 p-values. P-values are still reliable as shown in the PP-plot, Figure 3. Consequently, we can assign
409 robust empirical p-values to any size-two cocktails based on their risks. Furthermore the calibration
410 under the null hypothesis for the p-values attributed by the algorithm is assessed in Section 5.2.

411 **3.2.2 Genetic Algorithm Output and Clustering**

412 The genetic algorithm was applied to Dataset D_1 to identify high-risk drug cocktails. Multiple runs
413 of the algorithm were conducted using different hyperparameter sets (population size, number of
414 generations, parameter α) to ensure the visit of different regions of the space of possible cocktails.
415 The results were subsequently concatenated to create a comprehensive list of high-risk cocktails.

416 The genetic algorithm successfully identified nearly all high-risk size-two and size-three drug cocktails
417 in Dataset D_1 (Figure 4). For size-two cocktails, the algorithm consistently found the exact high-risk
418 combinations. However, for size-three cocktails, the algorithm sometimes identified cocktails that
419 were very close to the true high-risk combinations, missing only one drug from the correct cocktail
420 in a few cases (oftentimes, choosing parent nodes instead of the actual nodes).

421 To streamline the analysis of the large set of results, significant cocktails were filtered using the
422 empirical p-value by setting a threshold of 5%. Clustering techniques were then applied to group
423 similar cocktails together. As discussed in Section 2.4.3, the UMAP algorithm has been used for
424 dimensionality reduction, followed by the DBSCAN clustering method. This post-processing step
425 allows to reduce redundancy by grouping cocktails that differed only slightly, such as by substituting
426 a drug for another within the same pharmacological family. Such cocktails would have similar
427 medical interpretations.

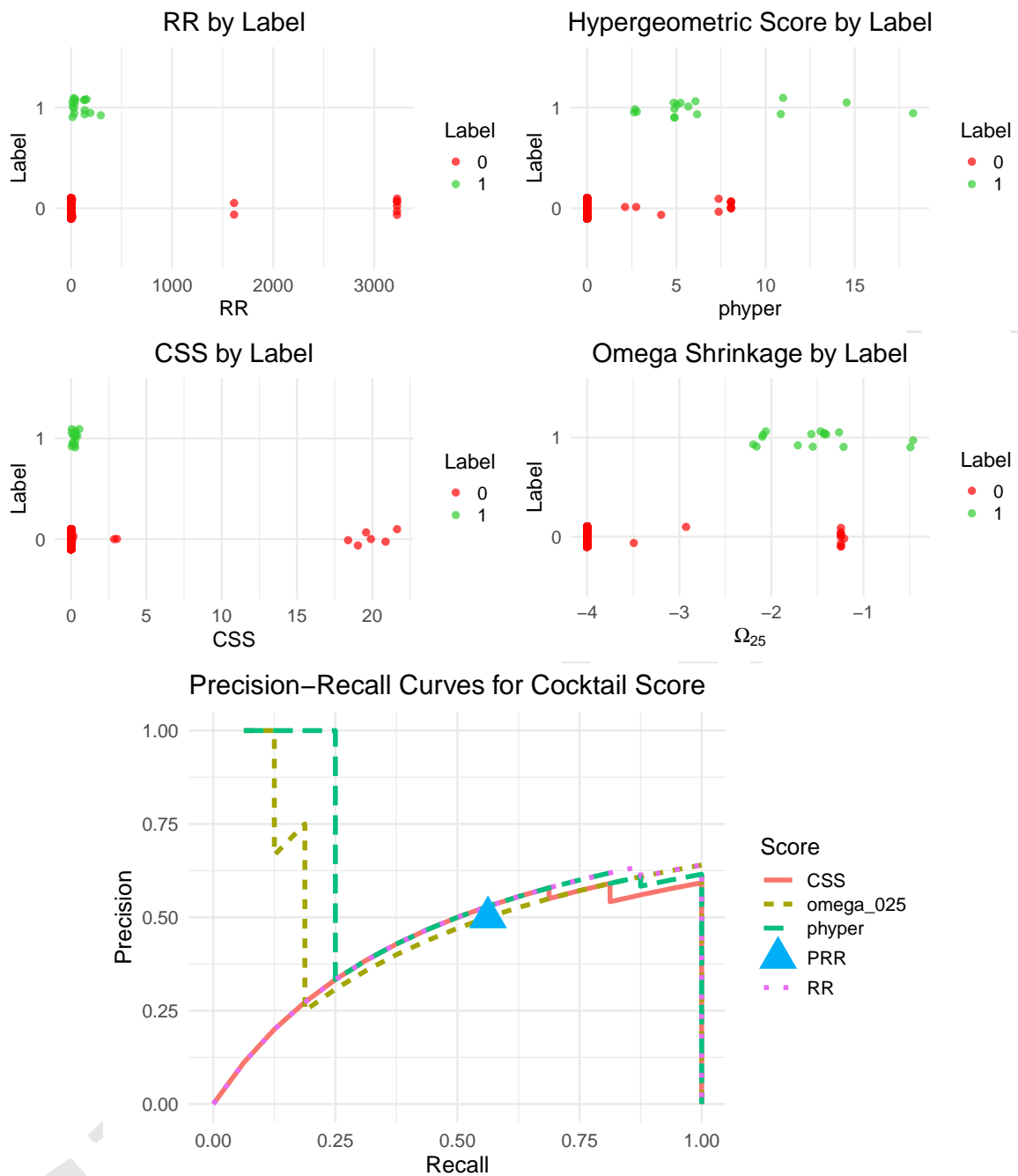


Figure 2: Comparison of scores for cocktail of size-two. Each dot denotes a cocktail, and its risk computed on Dataset D_2 (Section 2.6.1) is shown on the x-axis. Red dots represent cocktails that do not induce adverse event (negative controls), green ones represent cocktail inducing adverse event (positive controls). The bottom-right corner shows the Precision-Recall curves for each score. The perfect classification corresponds to the upper right corner. The areas under the curves allow to compare different methods. RR, PRR, Omega shrinkage, phyper and the CSS are defined in Section 2.2.1.

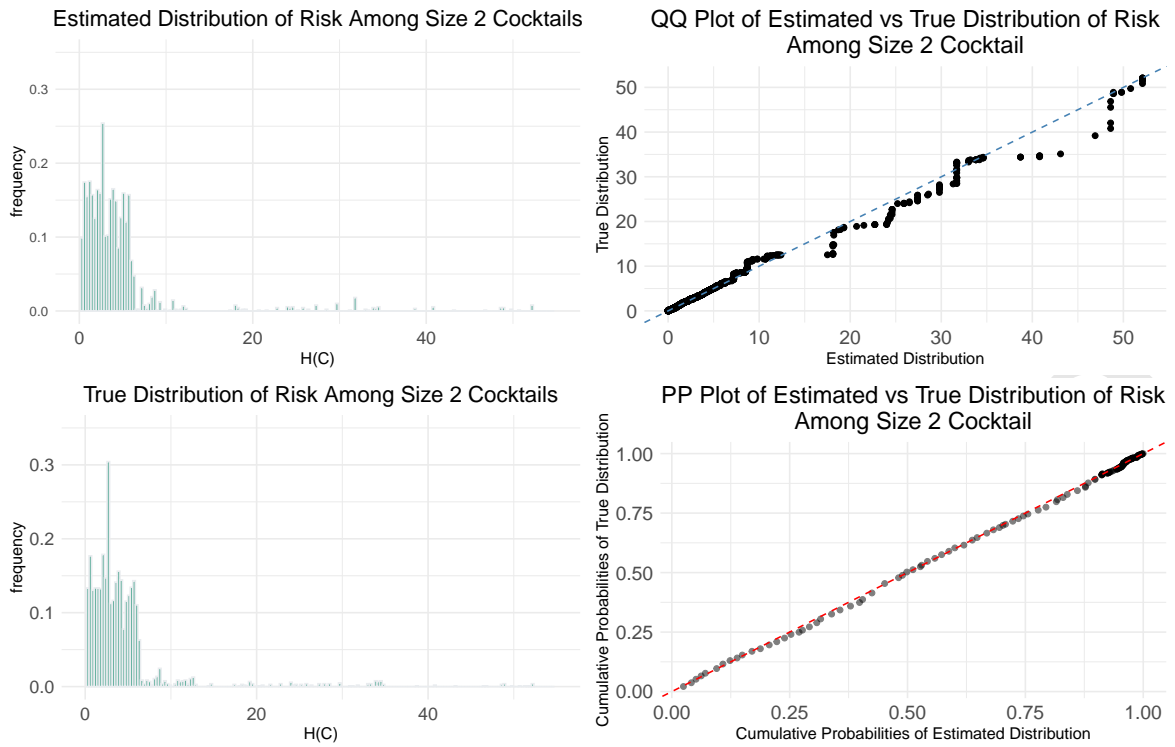


Figure 3: Comparison of estimated and true risk distributions for size-two drug cocktails on Dataset D_1 . Left panels show comparison of risk distribution among size-two cocktails, right panels allow comparison of probabilities and quantiles of both distributions

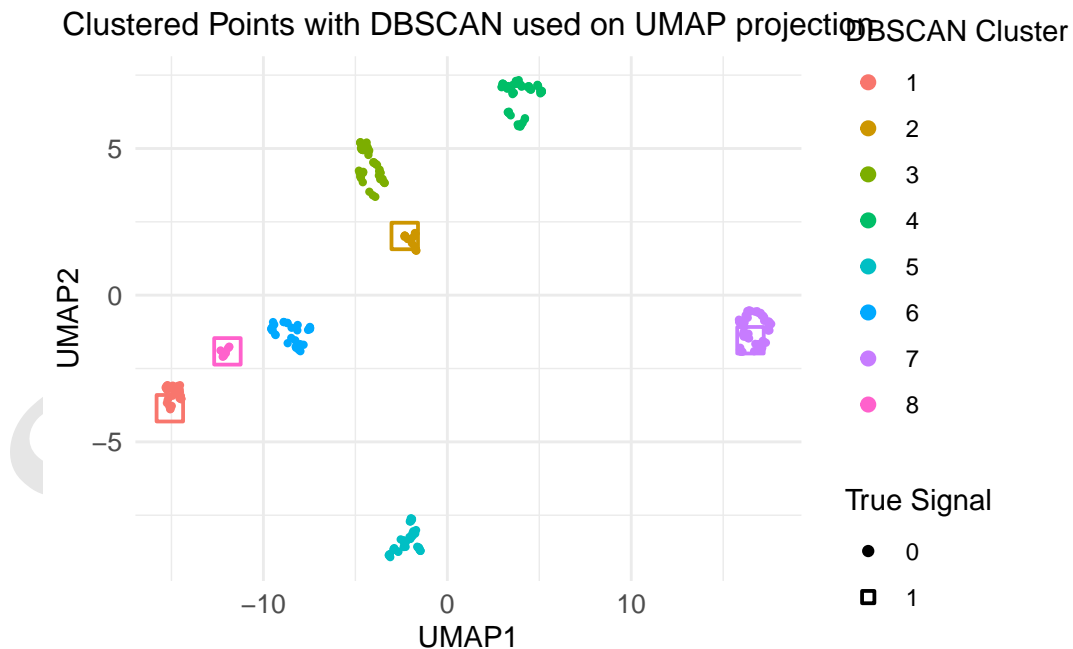


Figure 4: Clustering of high-risk drug cocktails identified by the genetic algorithm on the simulated dataset. Each dot represents a drug cocktail, true cocktail solutions are the centers of the squares

428 Figure 4 illustrates the results of the clustering process after mapping the selected cocktails in
429 the plane. Each point represents a drug cocktail, and the 7 automatically determined clusters are
430 represented by the colors.

431 The first conclusion is that the clusters are well separated in terms of the editing distance defined
432 in Section 2.4.2. The genetic algorithm thus identified seven regions of the cocktails space of high
433 score that are different in terms of interpretation as their cocktails are far away between classes.
434 On the other hand, cocktails in the same clusters are close, indicating a possible pharmacological
435 interpretation of the cluster.

436 The second conclusion is that the algorithm effectively found the true solutions or at least very similar
437 solutions. Indeed, the embedding of the true solutions yields the four squares, that clearly belong
438 to four different clusters among the seven. The method therefore identifies and separates regions
439 of the cocktail space containing the true solutions. That simulation thus enhances the confidence
440 that investigating the high-score cocktails returned by the methods may allow to detect relevant
441 phenomena.

442 3.3 Application to the FAERS Spontaneous Reporting Data

443 3.3.1 Estimation of Risk Distribution

444 The risk estimation method was applied to the FAERS spontaneous reporting dataset presented in
445 Section 2.6.2. Figure 5 presents a comparison between the estimated risk distribution and the true
446 risk distribution for size-two drug cocktails. The left panel of the figure shows the distributions of
447 risk scores, while the right panel presents the QQ plot and the PP plot, comparing the quantiles and
448 probabilities of both distributions.

449 The histogram reveals that the estimated distribution aligns well with the true distribution for the
450 majority of cocktails with lower risk scores. However, deviations begin to emerge in the tail of the
451 distribution. Specifically, 15 of the riskiest cocktails in the true distribution were not captured by
452 the estimated distribution as we see in the QQ plot. This explains the observed shift in the QQ plot
453 at the highest quantiles since higher risk cocktails have not been found by the MCMC algorithm,
454 highlighting the need of a complementary method like the genetic algorithm in order to find riskiest
455 cocktails.

456 Despite this slight deviation, the empirical p-values remain robust for both lower and higher-risk
457 cocktails as shown by the PP-plot in the Figure 5.

458 3.3.2 Genetic Algorithm Output and Clustering

459 The genetic algorithm was applied to the FAERS data focusing on the myopathy AE by running 180
460 parallel executions of the genetic algorithm on varying population sizes (from 100 to 1000 cocktails
461 per generation). The whole procedure ran in less than 8 hours on a 24-core server.

462 Cocktail sizes present in the merged final populations vary from 1 to 6. The MCMC algorithm was
463 run for each cocktail size in this range to assign an empirical p-value to each solution. All solutions
464 having a p-value lower than 0.05 were kept. No multiple testing correction was made in order to
465 avoid false negatives, that is disregarding interesting cocktails, even if this may inflate the number of
466 false positives. 682 drug cocktails compose the final list. Applying Benjamini-Hochberg corrections
467 to control the FDR at level 5% and 10% respectively kept the first 107 and 564 elements of the list.

468 The clustering procedure was applied to the 682 cocktails, leading to 15 clusters. To validate the
469 results, the first 150 solutions were tagged with the families they correspond to, if any, without
470 knowledge of the clustering's result. To do so, some drugs or drug families known to be linked to

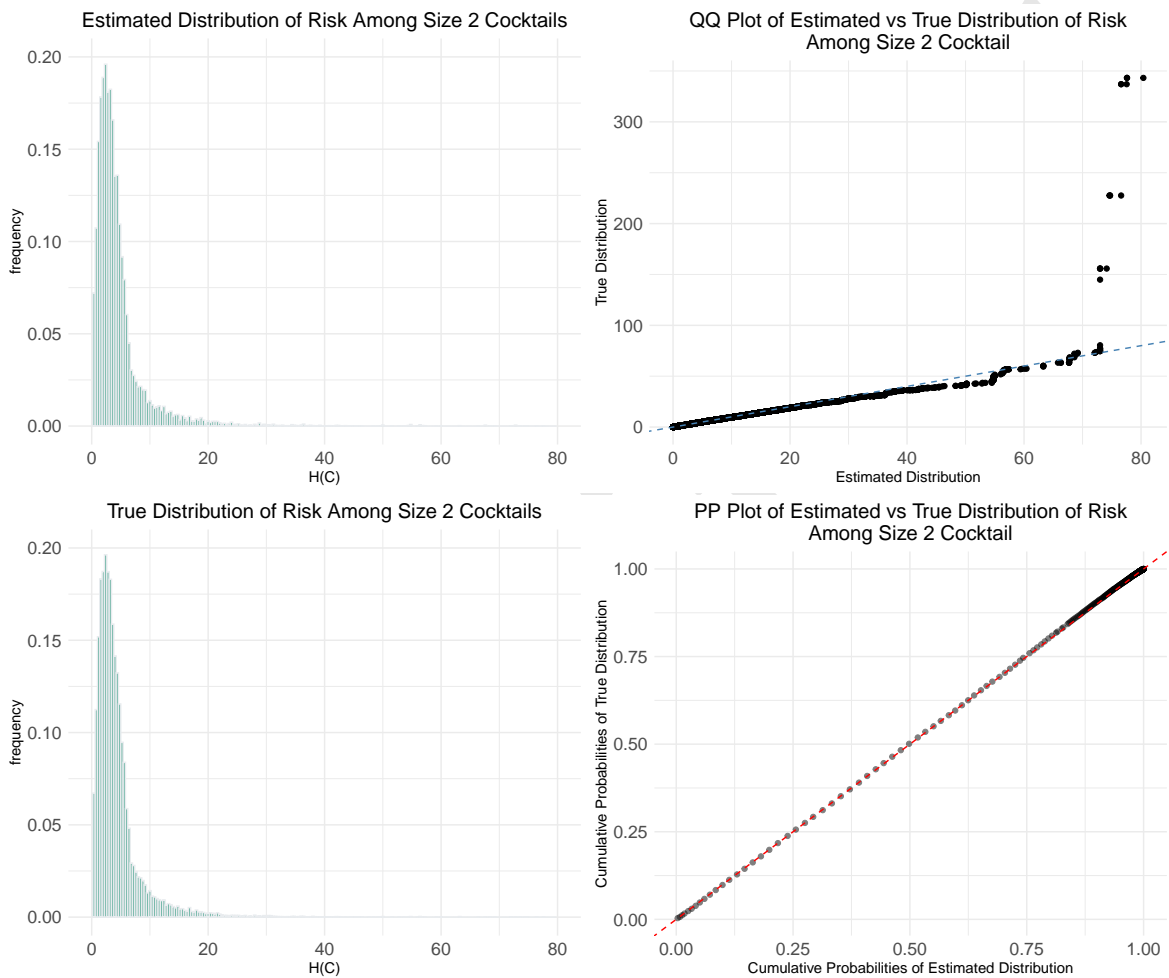


Figure 5: Comparison of estimated and true risk distributions for size-two drug cocktails on the FAERS dataset presented Section 2.6. Left panels show comparison of risk distribution among size-two cocktails, right panels allow comparison probabilities and quantiles of both distributions. Left distributions are truncated to 80 for visualization purposes. A few cocktails have a risk around 300 as shown on the Y-axis of the QQ plot.

471 myopathy adverse events were considered, that is hypolipemic drugs, Colchicine, corticosteroids,
 472 Ciclosporine, Beta blocking agents, Fluoroquinolones, and anti-malaria drugs (Miernik et al. 2024;
 473 Hall et al. 2011; Valiyil and Christopher-Stine 2010).

474 The final result is a table of 682 rows, one per selected cocktail, indicating its composition, the number
 475 of patients taking it, the number of patients taking it and facing myopathy, hypergeometric and RR
 476 score. It also contains the cluster it belongs to and the tagged families. The entire table is available
 477 [on this link](#). Table 1 summarizes the cluster assignments for the top 150 cocktails identified in the
 478 genetic algorithm. Among these, eleven cocktails could not be assigned to any drug or drug family
 479 listed in the table headers. Note that a cocktail can be associated with more than one drug or drug
 480 family.

Table 1: Summary of clustering applied to the identified solutions. Each row corresponds to a cluster. The 150 cocktails with the highest risk were analyzed. Box (i, j) in the table represents the number of cocktails in cluster i that include the drug or drug family j

Cluster	Hy- polipemic Drugs	Colchicine	Steroids	Cy- closporine	Beta blocking agents	Domperi- done	Fluoro- quinolones
1	31	0	3	0	0	0	4
2	0	11	0	4	0	0	0
3	7	0	0	0	19	0	0
4	0	0	0	0	0	50	0
5	0	0	0	0	0	19	0
6	0	0	0	0	0	0	4

481 This table shows the method’s capability to detect known ADRs from the FAERS dataset and highlights
 482 the ability of the clustering method to group cocktails with similar pharmacological interpretations.
 483 Specifically, cluster 1 corresponds to cocktails containing Hypolipemic Drugs, cluster 2 to Colchicine
 484 and Cyclosporine, cluster 3 to beta blocking agents and cluster 6 to Fluoroquinolones. These clusters
 485 align with previously reported pharmacological associations (Miernik et al. 2024). Similarly, clusters
 486 4 and 5 correspond to Domperidone-containing cocktails which have been associated with adverse
 487 effects, including potential cardiac complications, as reported by the British Medicines and Healthcare
 488 products Regulatory Agency (Medicines and Healthcare products Regulatory Agency 2014).

489 Interestingly, additional drug families reported in (Miernik et al. 2024), such as anti-malarial drugs,
 490 were also identified but at later ranks. For instance, anti-malarial drugs are primarily grouped in
 491 cluster 9, as shown in the complete solution table.

492 3.3.3 Post-Treatment Interaction Analysis

493 The penalized logistic regression described in Section 2.3 was applied to the strongest signals identi-
 494 fied by the genetic algorithm to assess whether the detected cocktails reflect true pharmacological
 495 interactions. For the cyclosporine–colchicine combination, the interaction coefficient remained posi-
 496 tive and highly significant ($p = 1.02 \times 10^{-6}$) after accounting for individual drug effects, corroborating
 497 existing medical literature indicating a potentiation of muscle toxicity (Ducloux et al. 1997).

498 Conversely, the analysis of the identified size-four cocktail (metformin, prasugrel, bisoprolol, sim-
 499 vastatin) revealed that the risk signal was almost entirely explained by a size-three sub-cocktail
 500 (metformin, bisoprolol, simvastatin). In this instance, the addition of prasugrel did not yield a signifi-
 501 cant additional interaction effect, illustrating the importance of the post-treatment step in identifying
 502 the core of a detected signal.

4 Conclusion

As co-medication becomes increasingly common, there is a growing need for methods capable of detecting signals of harmful drug combinations from the available large databases for further assessment. The proposed method addresses this need by identifying signals and assigning them a p-value using a hypergeometric disproportionality analysis measure. Additionally, the method enables the identification of broader signals within the ATC hierarchy by proposing “cocktails” of not only active substances but also chemical, therapeutic, and anatomical families, leveraging the hierarchical classification of active substances.

Application on synthetic datasets demonstrated that using the hypergeometric score reduces the false positives from cocktails taken by a small number of patients, enhancing the robustness of the measure. The results on these datasets of our MCMC algorithm to estimate the distributions of cocktail risks were encouraging, as the estimated distributions closely aligned with the true distributions, indicating reliable p-value assignment. Furthermore, the genetic algorithm effectively identified the majority of the harmful cocktails, with a high success rate, highlighting its efficiency in navigating the large solution space.

Applying this method to previous FAERS data for the myopathy adverse event yielded promising results. A literature review confirmed the intersection between the identified signals and drugs known to have a higher likelihood of causing myopathy, demonstrating the effectiveness of the proposed methodology. Results also indicate that certain drug combinations are more strongly associated with myopathy than individual medications. Notably, the cyclosporine/colchicine combination exhibits a higher hypergeometric score (73.1 vs. 63.4) and RR (879.4 vs. 57.1) compared to colchicine alone, reinforcing the importance of analyzing drug interactions in pharmacovigilance. This combination is known to be more likely to induce myopathy (Ducloux et al. 1997).

Furthermore, our approach identified a size-four drug cocktail (metformin, prasugrel, bisoprolol, simvastatin) associated with an increased risk signal (hypergeometric score = 72.2, RR = 3060.6). This combination was observed in nine patients, all of whom experienced the adverse event. While this finding demonstrates the feasibility of detecting higher-order drug interactions, we emphasize that no clinical validation is currently available for this specific combination. This underscores both the potential of the method in identifying complex drug interactions and the need for further validation through complementary studies.

As demonstrated in Section 3.3.3, the post-treatment logistic regression step provides clinical refinement by distinguishing true interactions from combined individual effects.

This approach can also be extended to other settings where adverse events are explored. For instance, the method could be applied using the ICD diagnosis classification or the MedDRA system, both of which are hierarchical classifications. Such an application would facilitate the identification of symptoms associated with the consumption of drug combinations.

The proposed method is implemented as an R package *emcAdr*, available on GitHub and on the CRAN (Bangard 2025). A tool allowing the processing of quarterly FAERS xml files to csv file directly usable by our method is also available on [GitHub](#).

For researchers and stakeholders it is crucial to remember that our method is hypothesis generating as are other signal detection methods in single medications. These methods aim to generate as few false negative results as possible but with the trade-off of more false positives. This means that any results should firstly be assessed for biological plausibility as well as validated in a more formal causal framework such as RCTs and target trial emulations. Furthermore, while spontaneous reporting systems have a key role in pharmacovigilance, they have known drawbacks as underlined in the

548 introduction. They however provide important knowledge about the safety of drugs, and with our
 549 method also drug cocktails.

550 Funding

551 This work was partially supported by the “PHC AURORA” programme (project number: 49704QC),
 552 funded by the French Ministry for Europe and Foreign Affairs, the French Ministry for Higher
 553 Education and Research and the Norwegian Council for Research.

554 5 Appendices

555 5.1 Appendix A : Distance Pseudo-code Algorithm

Algorithm 1 CalculateDistance

Require: Matrix M , Lists $idxC_1$, $idxC_2$

Ensure: Cost of the distance calculation

```

1:  $height \leftarrow 0$ 
2:  $cost \leftarrow 0$ 
3: while  $idxC_1 \neq \emptyset$  and  $idxC_2 \neq \emptyset$  do
4:   for all  $iC_1 \in idxC_1$  do
5:     for all  $iC_2 \in idxC_2$  do
6:       if  $M[iC_1][height] = M[iC_2][height]$  then
7:          $cost \leftarrow cost + height - \min(M[iC_1].count(M[iC_1][0]), M[iC_2].count(M[iC_2][0])) + 1$ 
8:         remove  $iC_1$  from  $idxC_1$ 
9:         remove  $iC_2$  from  $idxC_2$ 
10:        break
11:       end if
12:     end for
13:   end for
14:    $height \leftarrow height + 1$ 
15: end while
16:  $insertion\_cost \leftarrow ATC\_HEIGHT / 2$ 
17:  $cost \leftarrow cost + (|idxC_1| + |idxC_2|) \times insertion\_cost$ 
18: return  $cost$ 

```

556 Algorithm 1 computes the distance between two cocktails, C_1 and C_2 , based on three inputs.

557 The first input is an integer matrix M . This matrix represents the nodes of C_1 and C_2 , as well as
 558 their corresponding parent nodes. The matrix has dimensions (PN, ATC_HEIGHT) , where P and N
 559 denote the sizes of C_1 and C_2 , respectively. For example, consider the simplified tree shown in Figure
 560 1 of the article. In this tree, C_1 could be the cocktail containing only the node [3], and C_2 could be
 561 the cocktail containing only the node [2] ($P = 1$ and $N = 1$).

562 The corresponding matrix M would be:

$$M = \begin{bmatrix} 3 & 1 & 0 \\ 2 & 2 & 0 \end{bmatrix}$$

563 The second input consists of $idxC_1$ and $idxC_2$, which represent the indices of the rows in M that
 564 contain the nodes of C_1 and C_2 , respectively.

5.2 Appendix B : Calibration Under the Null

To assess the validity of the approximation of the p-value, we plotted the empirical distribution of p-values for cocktails presumed not to increase the adverse-event rate on the Dataset D_1 . The theoretical p-values following a Uniform(0, 1) distribution, the approximated ones should exhibit a distribution close to uniform if the approximation is valid. Figure 6 shows a histogram of these p-values, which appears approximately flat.

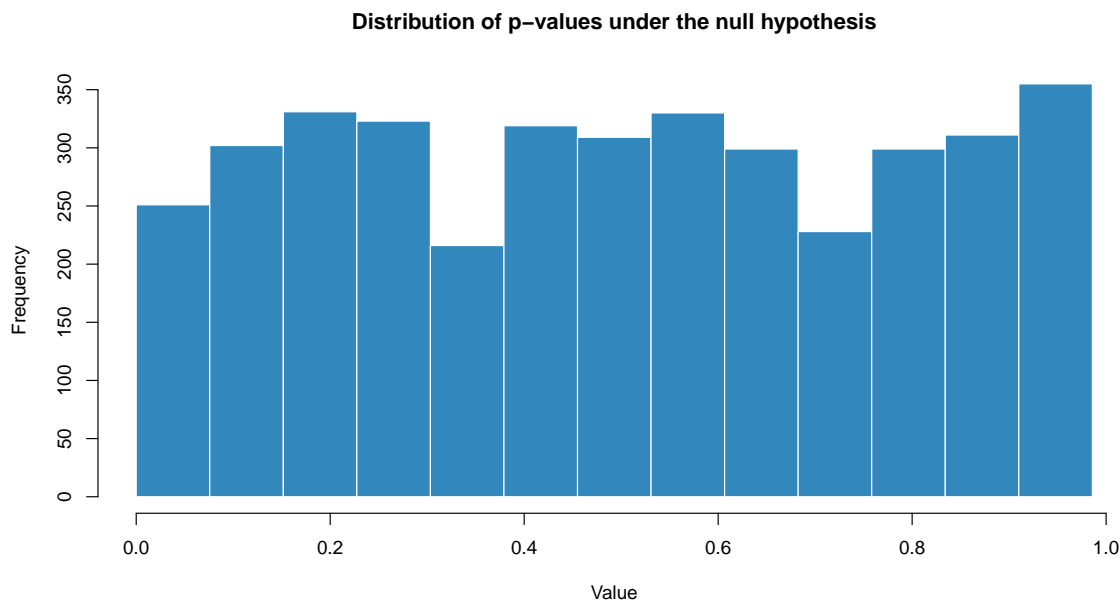


Figure 6: Histogram of p-values for cocktails presumed to satisfy the null hypothesis on Dataset D_1 (no elevated adverse-event risk)

5.3 Appendix C : FAERS Clustering

Figure 7 shows the clustering applied to the results of the genetic algorithm run on the FAERS datasets.

5.4 Appendix D : Algorithmic Complexity

The total time complexity of the high-risk cocktail identification is given by:

$$\mathcal{O}(G \cdot P \cdot T_{eval})$$

Where G represents the number of generations (or MCMC iterations) and P represents the population size of the GA (this equals one for the MCMC algorithm). The evaluation time for a single cocktail, T_{eval} , is defined by the matching logic required to count occurrences across the database:

$$T_{eval} = \mathcal{O}(N \cdot K \cdot R)$$

- N (Number of Reports): The algorithm iterates once through the database to compute the hypergeometric score for a given cocktail. The runtime scales linearly with the number of patients.
- K (Cocktail Size): The number of drugs in the candidate cocktail. In this study, K is typically small since there is a limit to the number of drugs a patient takes.

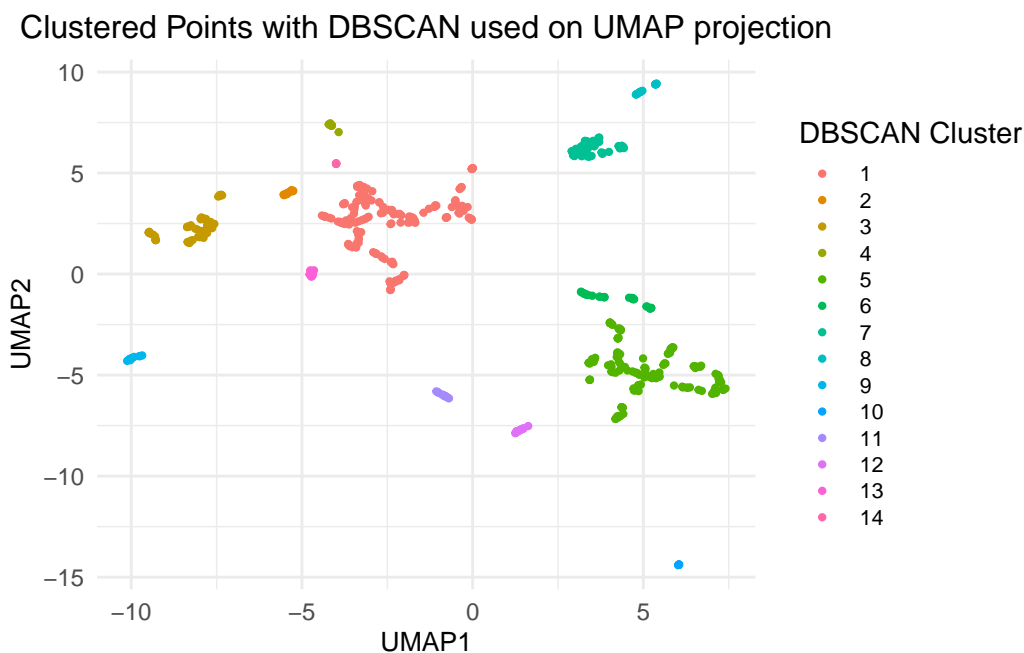


Figure 7: Clustering of High-Risk Drug Cocktails Identified by the Genetic Algorithm on the FAERS dataset. This Figure allows to see proximity between multiple clusters and better understand why some cluster exhibits similarities.

- R (Drugs per Report): The number of medications listed in a single patient report. On our data the mean size of drug cocktails per patient is approximately 2.2. Because K and R are bounded, the complexity simplifies to $O(N)$ per candidate evaluation.

5.5 Appendix E: Illustrative Example of Package Usage

This appendix demonstrates the end-to-end usage of the `emcAdr` package on a small simulated dataset that compiles in under a minute. It is intended as a hands-on illustration of the API: full reproduction of the results in Section 3.2 and Section 3.3 requires the longer-running scripts available in the article's GitHub repository. An accompanying `README.md` file in the article's GitHub repository documents the full reproduction workflow.

The toy dataset follows the same simulation design as Dataset D_1 (Section 2.6.1), at a reduced scale and with deliberately inflated adverse-event probabilities so that the signals remain detectable in a small sample. Patients taking only one of the two component drugs are added as decoys to break the spurious equivalence between the size-two cocktail and either of its components in a small dataset.

5.5.1 E.1 Setup and Toy Dataset Generation

```
# A tibble: 7 x 3
  group    AE_FALSE AE_TRUE
  <chr>    <int>    <int>
1 C1_decoy_d1    132      0
2 C1_decoy_d2    134      2
3 C1_full        80      16
4 C2_decoy_d1    174      0
5 C2_decoy_d2    177      0
```

```

606 6 C2_full          84      10
607 7 random          4174    17

```

608 The dataset contains 5000 patients, of whom approximately 2% take the full cocktail C_1 , 2% the full
609 cocktail C_2 , and 3% take a single drug from each cocktail as a decoy. The remaining patients receive
610 a random combination of active substances drawn from the ATC tree leaves.

611 $C1_full$ patients show a 20% AE rate vs 0.4% in the random group. Decoys show baseline AE rates as
612 designed.

613 Please note that the column of the patient dataframe has to be PatientATC and PatientADR.

614 5.5.2 E.2 Evaluating Cocktail Risk via the Hypergeometric Score

615 We start by computing the hypergeometric score $H(C)$ (Section 2.2.1) for a solutions cocktail and for
616 a random combination, to confirm that C_1 is clearly separated on the score scale.

```

617 [1] 37.79204 0.00000

```

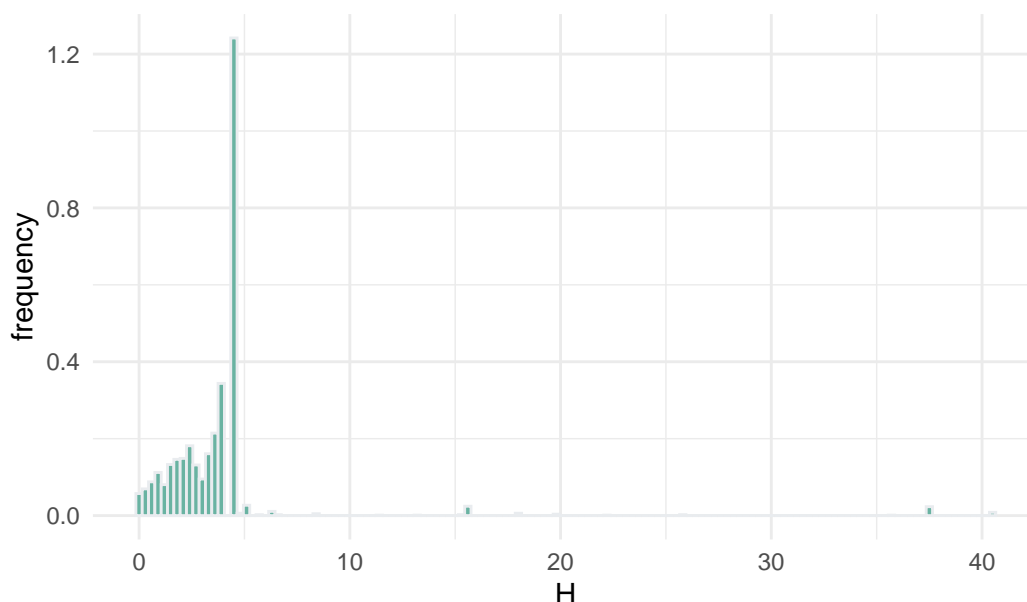
618 In datasets of this scale, the score function is sparse: most randomly drawn size-2 pairs are taken
619 by zero or one patient. A random pair therefore has near 100% probability of scoring exactly zero,
620 as illustrated by random_C above. The solution family C_1 scores $H(C_1) \approx 38$, but whether this
621 magnitude is extreme among cocktails actually present in the data is the question answered by the
622 next section.

623 5.5.3 E.3 Estimating the Null Distribution via MCMC

624 A short Metropolis-Hastings run (Section 2.5) approximates the distribution of $H(C)$ for size-two
625 cocktails. The full-scale analysis uses substantially longer chains.

626 openMP available

The approximation of the distribution histogram



627

628 p-value of solution C1 (47 1393) : 0.002559415

629 The bulk of the distribution sits near zero. The visible spike around $H(C) \approx 4 - 5$ corresponds to a
630 small-sample artifact, cocktails taken by only one patient where the patient happens to experience
631 the AE. With only ~ 45 AE cases over 5000 patients, such configurations are common enough to
632 form a visible noise peak. Furthermore, with only 5000 data points, each cocktail tend to be taken by
633 few patient. The solution families generate $H(C)$ values in the 18-40 range, above this peak, which is
634 why the empirical p-value of C_1 comes out at $\sim 2.5 \times 10^{-3}$.

635 For size-two cocktails, computing the *exact* score distribution is tractable and offered by the package
636 via `trueDistributionSizeTwoCocktail()`. This allows assessing the MCMC approximation directly
637 (see also Figure 3 in the body):

638 The PP/QQ-plot comparison would confirm that the MCMC tracks the null faithfully on the regions
639 of H that actually occur in the data.

640 5.5.4 E.4 Identifying High-Risk Cocktails with the Genetic Algorithm

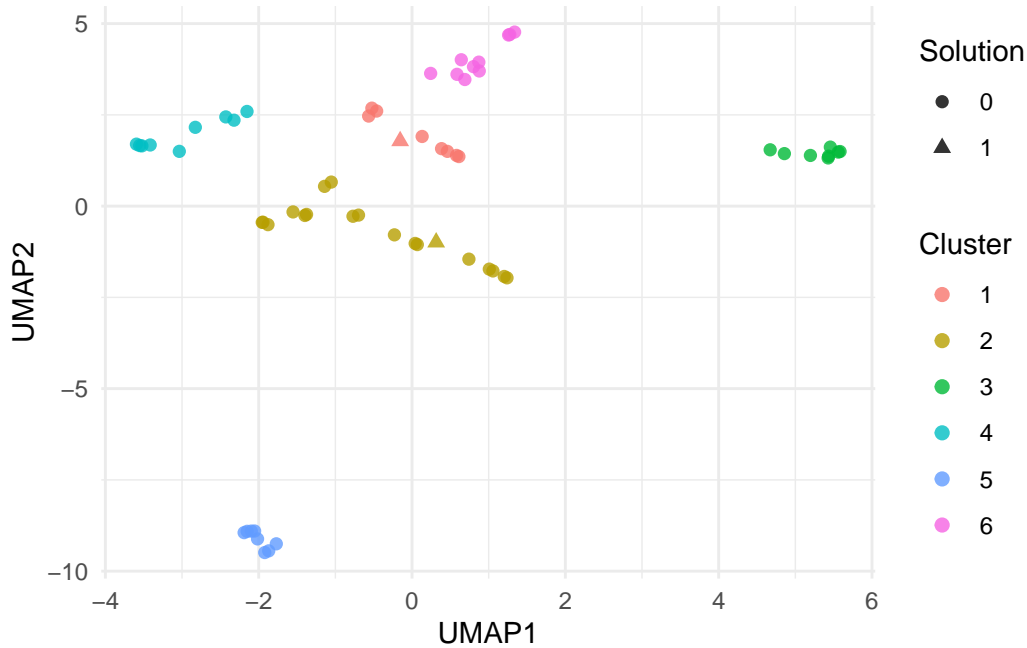
641 To mirror the multi-run aggregation workflow of Section 3.2, we run the genetic algorithm five times
642 with different seeds and pool the final populations. A single short run typically converges on one of
643 the two solutions. Aggregation across runs increases the chance of recovering both. For full-scale
644 runs on D_1 , parallel execution scripts compatible with SLURM and similar workflow managers are
645 available in the article's GitHub repository under `figures/scripts/`.

```
646 # A tibble: 10 x 3
647   cocktails score cocktails_display
648   <list>     <dbl> <chr>
649 1 <int [2]> 40.6 0, 1393
650 2 <int [2]> 37.8 47, 1391
651 3 <int [2]> 37.6 45, 1393
652 4 <int [2]> 36.0 47, 1023
653 5 <int [1]> 33.2 1393
654 6 <int [2]> 18.9 630, 888
655 7 <int [2]> 18.9 658, 888
656 8 <int [2]> 18.7 659, 863
657 9 <int [2]> 18.6 659, 738
658 10 <int [2]> 18.2 0, 888
```

659 The top outputs correspond to the solution families. Cocktails involving the parent ATC nodes for
660 C_1 (nodes 0, 45, 47 paired with 1391, 1393) score in the 36-40 range. Those for C_2 (nodes 630, 658,
661 659 paired with 888, 863) score around 18-19.

662 5.5.5 E.5 Output Clustering

663 For didactic purposes we apply the same UMAP + DBSCAN pipeline as described in Section 2.4.3 on
664 the recovered cocktails. We expect only a handful of distinct signals at this scale.



665

666 The two triangles (true solutions) embed in clusters 1 and 2 alongside related GA cocktails, demon-
 667 strating that the dissimilarity-based clustering groups variants of the same underlying signal in a
 668 reasonable way.

669 5.5.6 E.6 Post-Treatment with Firth’s Penalized Logistic Regression

670 For the C_1 cocktails, we apply the post-treatment of Section 2.3 to test whether the signal reflects a
 671 true synergistic interaction or merely the additive effect of its components. A function in emcAdr has
 672 been added in order to test a cocktail.

```
673 logistf::logistf(formula = Y ~ ., data = df_model)
```

674

675 Model fitted by Penalized ML

676 Coefficients:

	coef	se(coef)	lower 0.95	upper 0.95	Chisq	p
(Intercept)	-5.188918	0.1985810	-5.6048123	-4.823127	Inf	0.000000000
X47.TRUE	-0.557285	1.4303230	-5.4020854	1.417979	0.1824861	0.669245382
X1393.TRUE	1.692411	0.5180643	0.5357799	2.615262	7.3118976	0.006849955
X47.1393.TRUE	2.468896	1.5322560	0.1735472	7.390559	4.6094408	0.031796401

682

	method
(Intercept)	2
X47.TRUE	2
X1393.TRUE	2
X47.1393.TRUE	2

687

688 Method: 1-Wald, 2-Profile penalized log-likelihood, 3-None

689

690 Likelihood ratio test=80.73281 on 3 df, p=0, n=5000

691 Wald test = 787.0257 on 3 df, p = 0

692 The p-value of 0.032 on the interaction term confirms that this synergistic effect is statistically
 693 distinguishable, even with only 16 AE cases in the $C1_full$ group. It confirms that the genetic

694 algorithm's discovery reflects a true drug–drug interaction rather than the marginal contribution of
695 either family alone.

696 Genetic algorithm output also contains cocktail 1393 which contributes significantly to the AE on
697 his own as depicted in the summary.

698 **6 Code and Results**

699 See the complete results of application of methodology on the FAERS dataset [on this link](#).

700 Code to reproduce the experiments conducted in this article can be found here :

- 701 • GitHub link of the data refinement code : [JulesBa-Git/FAERS-xml-parser](#).
- 702 • GitHub link to the latest release of the R package of the complete methodology : [JulesBa-Git/emcAdr](#).
- 703 • The code of the methodology is also available on the [CRAN](#).
- 704 • The code to obtain figures is available on the `figures/scripts` folder of this repository.
- 705

706 **References**

707 Ahmed, Ismaïl, Cyril Dalmasso, Françoise Haramburu, Frantz Thiessard, Philippe Broët, and Pascale
708 Tubert-Bitter. 2010. “False Discovery Rate Estimation for Frequentist Pharmacovigilance Signal
709 Detection Methods.” *Biometrics* 66 (1): 301–9.

710 Almenoff, June, Joseph M Tonnig, A Lawrence Gould, et al. 2005. “Perspectives on the Use of Data
711 Mining in Pharmacovigilance.” *Drug Safety* 28 (11): 981–1007.

712 Au, Siu-Kui, and James L Beck. 2001. “Estimation of Small Failure Probabilities in High Dimensions
713 by Subset Simulation.” *Probabilistic Engineering Mechanics* 16 (4): 263–77.

714 Bangard, Jules. 2025. *The emcAdr Package*. <https://cran.r-project.org/web/packages/emcAdr/>.

715 Bate, Andrew, and Stephen JW Evans. 2009. “Quantitative Signal Detection Using Spontaneous ADR
716 Reporting.” *Pharmacoepidemiology and Drug Safety* 18 (6): 427–36.

717 Bate, Andrew, Marie Lindquist, I Ralph Edwards, et al. 1998. “A Bayesian Neural Network Method
718 for Adverse Drug Reaction Signal Generation.” *European Journal of Clinical Pharmacology* 54:
719 315–21.

720 Candore, Gianmario, Kristina Juhlin, Katrin Manlik, et al. 2015. “Comparison of Statistical Signal
721 Detection Methods Within and Across Spontaneous Reporting Databases.” *Drug Safety* 38 (6):
722 577–87.

723 Dechanont, Supinya, Sirada Maphanta, Bodin Butthum, and Chuenjid Kongkaew. 2014. “Hospital
724 Admissions/Visits Associated with Drug–Drug Interactions: A Systematic Review and Meta-
725 Analysis.” *Pharmacoepidemiology and Drug Safety* 23 (5): 489–97.

726 Ducloux, D, V Schuller, C Bresson-Vautrin, and JM Chalopin. 1997. “Colchicine Myopathy in Renal
727 Transplant Recipients on Cyclosporin.” *Nephrology, Dialysis, Transplantation: Official Publication
728 of the European Dialysis and Transplant Association-European Renal Association* 12 (11): 2389–92.

- 729 DuMouchel, William. 1999. "Bayesian Data Mining in Large Frequency Tables, with an Application
730 to the FDA Spontaneous Reporting System." *The American Statistician* 53 (3): 177–90.
- 731 Ester, Martin, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. 1996. "A Density-Based Algorithm
732 for Discovering Clusters in Large Spatial Databases with Noise." *Kdd* 96: 226–31.
- 733 Evans, Stephen JW, Patrick C Waller, and S Davis. 2001. "Use of Proportional Reporting Ratios (PRRs)
734 for Signal Generation from Spontaneous Adverse Drug Reaction Reports." *Pharmacoepidemiology
735 and Drug Safety* 10 (6): 483–86.
- 736 Firth, David. 1993. "Bias Reduction of Maximum Likelihood Estimates." *Biometrika* 80 (1): 27–38.
- 737 Fusaroli, Michele, Valentina Giunchi, Vera Battini, et al. 2024. "Enhancing Transparency in Defining
738 Studied Drugs: The Open-Source Living DiAna Dictionary for Standardizing Drug Names in the
739 FAERS." *Drug Safety*, 1–14.
- 740 Gosho, Masahiko, Kazushi Maruo, Keisuke Tada, and Akihiro Hirakawa. 2017. "Utilization of
741 Chi-Square Statistics for Screening Adverse Drug-Drug Interactions in Spontaneous Reporting
742 Systems." *European Journal of Clinical Pharmacology* 73: 779–86.
- 743 Grossmann, Steffen, Sebastian Bauer, Peter N Robinson, and Martin Vingron. 2007. "Improved
744 Detection of Overrepresentation of Gene-Ontology Annotations with Parent–Child Analysis."
745 *Bioinformatics* 23 (22): 3024–31.
- 746 Hall, Mederic M, Jonathan T Finnoff, and Jay Smith. 2011. "Musculoskeletal Complications of
747 Fluoroquinolones: Guidelines and Precautions for Usage in the Athletic Population." *PM&R* 3 (2):
748 132–42.
- 749 Hauben, Manfred. 2023. "Artificial Intelligence and Data Mining for the Pharmacovigilance of
750 Drug–Drug Interactions." *Clinical Therapeutics* 45 (2): 117–33.
- 751 Heijden, Peter GM van der, Eugène P van Puijenbroek, Stef van Buuren, and Jacques W Van der
752 Hofstede. 2002. "On the Assessment of Adverse Drug Reactions from Spontaneous Reporting
753 Systems: The Influence of Under-Reporting on Odds Ratios." *Statistics in Medicine* 21 (14):
754 2027–44.
- 755 Heinze, Georg, Meinhard Ploner, and Meinhard Dunkler. 2023. *Logistf: Firth's Penalized Likelihood
756 Logistic Regression*. <https://CRAN.R-project.org/package=logistf>.
- 757 Heo, Seok-Jae, Sohee Jeong, Dageom Jung, and Inkyung Jung. 2024. "Signal Detection Statistics of
758 Adverse Drug Events in Hierarchical Structure for Matched Case–Control Data." *Biostatistics* 25
759 (4): 1112–21.
- 760 Ibrahim, Heba, A Abdo, Ahmed M El Kerdawy, and A Sharaf Eldin. 2021. "Signal Detection in
761 Pharmacovigilance: A Review of Informatics-Driven Approaches for the Discovery of Drug-Drug
762 Interaction Signals in Different Data Sources." *Artificial Intelligence in the Life Sciences* 1: 100005.
- 763 Ibrahim, Heba, Amr Saad, Amany Abdo, and A Sharaf Eldin. 2016. "Mining Association Patterns
764 of Drug-Interactions Using Post Marketing FDA's Spontaneous Reporting Data." *Journal of
765 Biomedical Informatics* 60: 294–308.

- 766 Kulldorff, Martin, Zixing Fang, and Stephen J Walsh. 2003. "A Tree-Based Scan Statistic for Database
767 Disease Surveillance." *Biometrics* 59 (2): 323–31.
- 768 Levenshtein, Vladimir I et al. 1966. "Binary Codes Capable of Correcting Deletions, Insertions, and
769 Reversals." *Soviet Physics Doklady* 10: 707–10.
- 770 McInnes, Leland, John Healy, and James Melville. 2018. "Umap: Uniform Manifold Approximation
771 and Projection for Dimension Reduction." *arXiv Preprint arXiv:1802.03426*.
- 772 Medicines and Healthcare products Regulatory Agency. 2014. "Domperidone: Risks of Cardiac Side
773 Effects." *Drug Safety Update* 7 (10): A1. [https://www.gov.uk/drug-safety-update/domperidone-
774 risks-of-cardiac-side-effects](https://www.gov.uk/drug-safety-update/domperidone-risks-of-cardiac-side-effects).
- 775 Miernik, Sebastian, Agata Matusiewicz, and Marzena Olesińska. 2024. "Drug-Induced Myopathies: A
776 Comprehensive Review and Update." *Biomedicines* 12 (5): 987.
- 777 Noguchi, Yoshihiro, Keisuke Aoyama, Satoaki Kubo, Tomoya Tachi, and Hitomi Teramachi. 2020.
778 "Improved Detection Criteria for Detecting Drug-Drug Interaction Signals Using the Proportional
779 Reporting Ratio." *Pharmaceuticals* 14 (1): 4.
- 780 Noguchi, Yoshihiro, Anri Ueno, Manami Otsubo, et al. 2018. "A New Search Method Using Association
781 Rule Mining for Drug-Drug Interaction Based on Spontaneous Report System." *Frontiers in
782 Pharmacology* 9: 197.
- 783 Norén, G Niklas, Rolf Sundberg, Andrew Bate, and I Ralph Edwards. 2008. "A Statistical Methodology
784 for Drug–Drug Interaction Surveillance." *Statistics in Medicine* 27 (16): 3057–70.
- 785 Pariente, Antoine, Fleur Gregoire, Annie Fourrier-Reglat, Françoise Haramburu, and Nicholas Moore.
786 2007. "Impact of Safety Alerts on Measures of Disproportionality in Spontaneous Reporting
787 Databases the Notoriety Bias." *Drug Safety* 30 (10): 891–98.
- 788 Pétrowski, Alain, and Sana Ben-Hamida. 2017. *Evolutionary Algorithms*. John Wiley & Sons.
- 789 Puijtenbroek, Eugene P. van, Andrew Bate, Hubert G. M. Leufkens, Marie Lindquist, Roland Orre,
790 and Antoine C. G. Egberts. 2002. "A Comparison of Measures of Disproportionality for Signal
791 Detection in Spontaneous Reporting Systems for Adverse Drug Reactions." *Pharmacoepidemiology
792 and Drug Safety* 11. <https://api.semanticscholar.org/CorpusID:21212215>.
- 793 Robert, Christian P., and George Casella. 2004. "The Metropolis–Hastings Algorithm." In *Monte
794 Carlo Statistical Methods*. Springer New York. https://doi.org/10.1007/978-1-4757-4145-2_7.
- 795 Sanson-Fisher, Robert William, Billie Bonevski, Lawrence W Green, and Cate D'Este. 2007. "Limita-
796 tions of the Randomized Controlled Trial in Evaluating Population-Based Health Interventions."
797 *American Journal of Preventive Medicine* 33 (2): 155–61.
- 798 Schuler, Jochen, Christina Dückelmann, Wolfgang Beindl, Erika Prinz, Thomas Michalski, and Max
799 Pichler. 2008. "Polypharmacy and Inappropriate Prescribing in Elderly Internal-Medicine Patients
800 in Austria." *Wiener Klinische Wochenschrift* 120.
- 801 Tekin, Elif, Cynthia White, Tina Manzhu Kang, et al. 2018. "Prevalence and Patterns of Higher-

802 Order Drug Interactions in Escherichia Coli.” *Npj Systems Biology and Applications* 4 (1): 31.
803 <https://doi.org/10.1038/s41540-018-0069-9>.

804 Valiyil, Ritu, and Lisa Christopher-Stine. 2010. “Drug-Related Myopathies of Which the Clinician
805 Should Be Aware.” *Current Rheumatology Reports* 12: 213–20.

806 Van Puijenbroek, Eugène P, Antoine CG Egberts, Eibert R Heerdink, and Hubert GM Leufkens. 2000.
807 “Detecting Drug–Drug Interactions Using a Database for Spontaneous Adverse Drug Reactions:
808 An Example with Diuretics and Non-Steroidal Anti-Inflammatory Drugs.” *European Journal of*
809 *Clinical Pharmacology* 56: 733–38.

810 Van Puijenbroek, Eugène P, Antoine CG Egberts, Ronald HB Meyboom, and Hubert GM Leufkens.
811 1999. “Signalling Possible Drug–Drug Interactions in a Spontaneous Reporting System: Delay of
812 Withdrawal Bleeding During Concomitant Use of Oral Contraceptives and Itraconazole.” *British*
813 *Journal of Clinical Pharmacology* 47 (6): 689–93.

814 Wang, Xueying, Lang Li, Lei Wang, Weixing Feng, and Pengyue Zhang. 2020. “Propensity Score-
815 Adjusted Three-Component Mixture Model for Drug-Drug Interaction Data Mining in FDA
816 Adverse Event Reporting System.” *Statistics in Medicine* 39 (7): 996–1010.

817 Session information

818 R version 4.5.1 (2025-06-13)
819 Platform: x86_64-pc-linux-gnu
820 Running under: Ubuntu 24.04.4 LTS
821
822 Matrix products: default
823 BLAS: /usr/lib/x86_64-linux-gnu/blas/libblas.so.3.12.0
824 LAPACK: /usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.12.0 LAPACK version 3.12.0
825
826 locale:
827 [1] LC_CTYPE=C.UTF-8 LC_NUMERIC=C LC_TIME=C.UTF-8
828 [4] LC_COLLATE=C.UTF-8 LC_MONETARY=C.UTF-8 LC_MESSAGES=C.UTF-8
829 [7] LC_PAPER=C.UTF-8 LC_NAME=C LC_ADDRESS=C
830 [10] LC_TELEPHONE=C LC_MEASUREMENT=C.UTF-8 LC_IDENTIFICATION=C
831
832 time zone: Etc/UTC
833 tzcode source: system (glibc)
834
835 attached base packages:
836 [1] stats graphics grDevices datasets utils methods base
837
838 other attached packages:
839 [1] tidyr_1.3.2 purrr_1.2.2 stringr_1.6.0 umap_0.2.10.0 ggplot2_4.0.3
840 [6] dbscan_1.2.4 dplyr_1.2.1 emcAdr_1.3
841
842 loaded via a namespace (and not attached):
843 [1] gtable_0.3.6 shape_1.4.6.1 xfun_0.57
844 [4] formula.tools_1.7.1 lattice_0.22-9 vctrs_0.7.3
845 [7] tools_4.5.1 Rdpack_2.6.6 generics_0.1.4

846 [10] tibble_3.3.1
847 [13] jomo_2.7-6
848 [16] S7_0.2.2
849 [19] farver_2.1.2
850 [22] yaml_2.3.12
851 [25] nloptr_2.2.1
852 [28] openssl_2.4.0
853 [31] rpart_4.1.24
854 [34] mitml_0.4-5
855 [37] tidysselect_1.2.1
856 [40] labeling_0.4.3
857 [43] fastmap_1.2.0
858 [46] magrittr_2.0.5
859 [49] broom_1.0.12
860 [52] backports_1.5.1
861 [55] nnet_7.3-20
862 [58] askpass_1.2.1
863 [61] knitr_1.51
864 [64] rlang_1.2.0
865 [67] renv_1.2.2
866 [70] R6_2.6.1

pan_1.9
Matrix_1.7-5
lifecycle_1.0.5
codetools_0.2-20
glmnet_4.1-10
pillar_1.11.1
reformulas_0.4.4
boot_1.3-31
nlme_3.1-169
digest_0.6.39
splines_4.5.1
grid_4.5.1
survival_3.8-3
withr_3.0.2
logistf_1.26.1
lme4_2.0-1
png_0.1-9
rbibutils_2.4.1
Rcpp_1.1.1-1.1
minqa_1.2.8

pkgconfig_2.0.3
RColorBrewer_1.1-3
compiler_4.5.1
htmltools_0.5.9
mice_3.19.0
MASS_7.3-65
iterators_1.0.14
foreach_1.5.2
RSpectra_0.16-2
stringi_1.8.7
operator.tools_1.6.3.1
cli_3.6.6
utf8_1.2.6
scales_1.4.0
rmarkdown_2.31
reticulate_1.46.0
evaluate_1.0.5
mgcv_1.9-4
glue_1.8.1
jsonlite_2.0.0