Data-driven methodology for discovery and response to pulmonary symptomology in hypertension through AI and machine learning: Application to COVID-19 related pharmacovigilance

Xuan Xu[1,2,3], Jessica Kawakami[1,4,5], Nuwan Indika Millagaha Gedara[1,3,6], Jim Riviere[1,7], Emma Meyer[1,4], Gerald J. Wyckoff[1,4,5], and Majid Jaberi-Douraki[1,2,3,*]

[1]1DATA Consortium, www.1DATA.life, USA

[2]Department of Mathematics, Kansas State University

[3]Kansas State University Olathe, Olathe, KS 66061-1304

[4]School of Pharmacy, Division of Pharmacology and Pharmaceutical Sciences, University of Missouri-Kansas City

[5]Molecular Biology and Biochemistry, School of Biological and Chemical Sciences, University of Missouri-Kansas City

[6]Department of Business Economics, University of Colombo, Sri Lanka

[7]Kansas State University and North Carolina State University

[*]**Correspondence:** Majid Jaberi-Douraki, jaberi@ksu.edu

Running title: AI and machine learning data-driven methods: COVID19 related pharmacovigilance

ABSTRACT

Background: Potential therapy and confounding factors including typical co-administered medications, patient's disease states, disease prevalence, patient demographics, medical histories, and reasons for prescribing a drug often are incomplete, conflicting, missing, or uncharacterized in spontaneous adverse drug event (ADE) reporting systems. These missing or incomplete features can affect and limit the application of quantitative methods in pharmacovigilance for meta-analyses of data during randomized clinical trials.

Methods: In this study, we implemented adaptive signal detection approaches to correct spurious association, hidden factors, and confounder misclassification when the covariates are unknown or unmeasured on medications affecting the renin-angiotensin system (RAS), potentially creating an increased risk of life-threatening outcomes in high-risk patients.

Results: Following multiple filtering stages to exclude insignificant and noise-driven reports, we found that drugs from antihypertensives agents, urologicals, and antithrombotic agents (macitentan, bosentan, epoprostenol, selexipag, sildenafil, tadalafil, and beraprost) form a similar class with a significantly higher incidence of pADEs. Macitentan and bosentan were associates with 64% and 56% of pADEs, respectively. Because these two medications are prescribed in diseases affecting pulmonary function and may be likely to emerge among the highest reported pADEs, in fact, they serve to validate the methods utilized here. Conversely, doxazosin and rilmenidine were found to have the least pADEs in selected drugs from hypertension patients. Nifedipine and candesartan were also found by our signal detection methods to form a drug cluster, shown by several studies an effective combination of these drugs on lowering blood pressure and appeared an improved side effect profile in comparison with single-agent monotherapy.

Conclusions: We consider pulmonary ADE (pADE) profiles in a long-standing group of therapeutics, RAS-acting agents, in patients with hypertension associated with high-risk for COVID-19. Using these techniques, we confirmed our hypothesis that drugs from the same drug class could have very different pADE profiles affecting outcomes in acute respiratory illness. We found that several indidvual drugs have significant differences between their drug classes and compared to other drug classes.

Funding: GJW and MJD accepted funding from BioNexus KC for funding on this project but BioNexus KC had no direct role in this article.

Clinical trial number: N/A

Author Summary

Underlying comorbidities continue to negatively affect COVID-19 patients. A recent focus has been on medications affecting RAS. Therefore, with the advent of COVID-19 acute respiratory distress syndrome (ARDS) in high-risk patients with hypertension, identifying specific RAS medications with the lowest incidence of pADEs would be beneficial. For this purpose, we curated the FDA ADE database to search for information related to human pADEs. As part of post-marketing drug safety surveillance, state/federal regulatory agencies and other institutions provide massive collections of ADE reports, these large data-sets present an opportunity to investigate ADEs to provide patient management based on comparative population data analysis. The abundance and prevalence of ADEs are not always detectable during randomized clinical trials and before a drug receives FDA approval for use in the clinic, which may appear with more widespread use. This is especially true for specific agents or diseases since there are simply too

68 few events to be assessed, even in a large clinical trial for side effect profiles of specific disease
69 states. For this purpose, we employed a novel method identifying extraneous causes of differential
70 reporting including sampling variance and selection biases by reducing the effect of covariates.
71

INTRODUCTION

The coronavirus disease 2019 (COVID-19) pandemic continues with 115,094,614 confirmed cases and over 2.6 million deaths as of March 5, 2021 (1, 2). Surprisingly, it is estimated that as high as 45% of infected individuals may remain asymptomatic, contributing to disease transmission and underlying the disparity in symptomology (3). A commonality of severe clinical course and mortality is comorbid conditions such as diabetes, heart disease, obesity, and hypertension (4). Hypertension was recognized early on as being a prevalent risk factor (5), possibly due to its pervasiveness. Hypertension affects 23% of adults in China, where the original study was conducted, but affects 45% of US adults. Moreover, specific antihypertensive medications, namely angiotensin-converting enzyme inhibitors (ACEIs) and angiotensin-II receptor blockers (ARBs), target proteins of the renin-angiotensin system (RAS) (6). The RAS is intricately linked to initial infection and possibly the progression of COVID-19 through a RAS receptor, angiotensin-converting enzyme 2 (ACE2), which acts as the viral entry point of coronavirus SARS-CoV-2 (7, 8).

In recent years, data science has emerged as a new and important discipline in medicine and healthcare. Different quantitative therapeutic efforts in drug repurposing or repositioning combined with adverse drug event (ADE) identification have led to more efficient therapies while improving the clinical course, lowering fatality, and decreasing cost burden (9). Our previous work focused on the incidence of pulmonary ADEs associated with ACEI and ARB use in patients with hypertension and other comorbidities (10, 11). Our findings indicate that specific drugs—rather than entire classes—have higher incidences of pulmonary ADEs, which may have implications for treating patients diagnosed with COVID-19. Most epidemiological studies are not this granular as

they do not analyze drug effects at the individual drug level but rather compare pharmacological classes. The current study examines additional drugs that more broadly target hypertension, including pulmonary hypertension, to describe methods used to identify clinically important patterns of ADE data. We utilized the Anatomical Therapeutic Chemical (ATC) classification system from the World Health Organization (WHO) Collaborating Center for Drug Statistics Methodology (https://www.whocc.no/). The ATC system classifies drugs based on site of action in addition to chemical, pharmacological, and therapeutic properties (12). Here we identify a clear signal distinct from different drugs in patients with hypertension as an underlying medical condition which helps to quantify the anomaly and unexpectedness of an ADE reported for a drug through disproportionality analysis. For this purpose, we proceeded with a specific pairwise analysis of individual drugs compared to the drug classes using a modified empirical Bayes method to identify any distinctions between drugs within a class and compared to other classes.

In our previous work, thirteen different pulmonary ADEs were selected based on clinical importance, and as they were prevalent among the top reported symptoms in patients with COVID-19, to assess the related variation due to adverse event differences (10, 11). In the present work, we include 25 pulmonary, infectious disease, or cardiac-associated ADEs. Our novel method identifies extraneous causes of differential reporting including sampling variance and selection biases by reducing the effect of covariates. This method is both adaptive (it removes different covariates for different drugs) and appropriate for the systematic application and routine analysis (13). We hypothesize that drugs from the same class based on the Anatomical Therapeutic Chemical (ATC) classification system could have different ADE profiles. For this purpose,

penalized regression method will be used to detect clusters of drugs, may differ from the ATC classification, and will be validated by the Friedman test (14-16). Safety signals for a specific drug and associated adverse events are then identified and evaluated through different methods, such as the proportional reporting ratio (PRR) (14), the relative reporting ratio (RR) (17), the information component (IC) (18), and the empirical Bayes geometric mean (EBGM) (17). These methods are utilized to calculate the ratio of an ADE compared to the same event occurring with other drugs, however, PRR or RR is more liberal when an event incidence is small (19).

RESULTS

Preprocessing and Data Cleaning

Here we briefly explain the data preprocessing and cleansing that will be used in different subsections. The focus of each subsection is given by the amount of data that will be used. A total of 480,236 spontaneous ADE reports for patients with hypertension were retrieved from our 1DATA databank of the FAERS database from the first quarter of 2004 until the first quarter of 2020. Alternatively, ADEs can be categorized by drug for a total of 612,733 reports (**Table 1**) arising from patients taking more than one drug. For example, a single ADE reported for a patient taking 2 different drugs, will generate one ADE report for each drug. This hypertension dataset was aggregated to 1520 ADEs in HLT codes corresponding to 1131 drugs with unique active substances. Next, drugs were excluded when the number of ADEs due to the fact that each drug was reported less than 500 times, accounting for approximately less than 0.1% of the data. Furthermore, 98.8% of the data corresponded to 134 of the 1131 drugs (**Table 1** with the column

header: # Drugs after initial filtering; this dataset will be exploited to calculate the relative risk for the disproportionality measures of a drug-ADE occurrence). This study focused on the 98.8% of the data remaining after the elimination of insignificant and noise-driven reports. The 134 drugs were grouped according to the following ATC drug classes (**Table 1**): ACEIs, ARBs, other RAS agents, other Antihypertensives Agents (AHAs), Antithrombotic Agents (ATAs), Beta blocking Agents (BBAs), Calcium channel blockers (CCBs), Diuretics, Lipid modifying agents, Urologicals (UAs), Vasoprotectives, and Combinations of antihypertensives (COMBs).

Since there were 5 unrelated pulmonary ADEs in the database (*coronavirus infections, conditions associated with abnormal gas exchange, neonatal hypoxic conditions, newborn respiratory disorders NEC, pulmonary hypertensions*), the hypertension dataset was further reduced to reports corresponding to the following 30 pulmonary ADEs: *bacterial lower respiratory tract infections, breathing abnormalities, bronchial conditions NEC, bronchospasm and obstruction, congenital lower respiratory tract disorders, coughing and associated symptoms, fungal lower respiratory tract infections, infectious disorders carrier, lower respiratory tract infections NEC, lower respiratory tract inflammatory and immunologic conditions, lower respiratory tract neoplasms, lower respiratory tract radiation disorders, lower respiratory tract signs and symptoms, occupational parenchymal lung disorders, parasitic lower respiratory tract infections, parenchymal lung disorders NEC, pleural conditions NEC, pleural infections and inflammations, pleural neoplasms, pneumothorax and pleural effusions NEC, pulmonary oedemas, pulmonary thrombotic and embolic conditions, respiratory failures (excl neonatal), respiratory signs and symptoms NEC, respiratory syncytial viral infections, respiratory tract disorders NEC, respiratory tract infections NEC, respiratory tract neoplasms NEC, vascular pulmonary disorders NEC, and viral lower respiratory tract infections*. Of the 30 pulmonary ADEs, 5 ADEs were additionally

163 excluded from the analysis since we did not have any reports for these ADEs: *congenital lower*
164 *respiratory tract disorders, lower respiratory tract radiation disorders, parasitic lower*
165 *respiratory tract infections, respiratory tract neoplasms NEC, and viral lower respiratory tract*
166 *infections*.
167

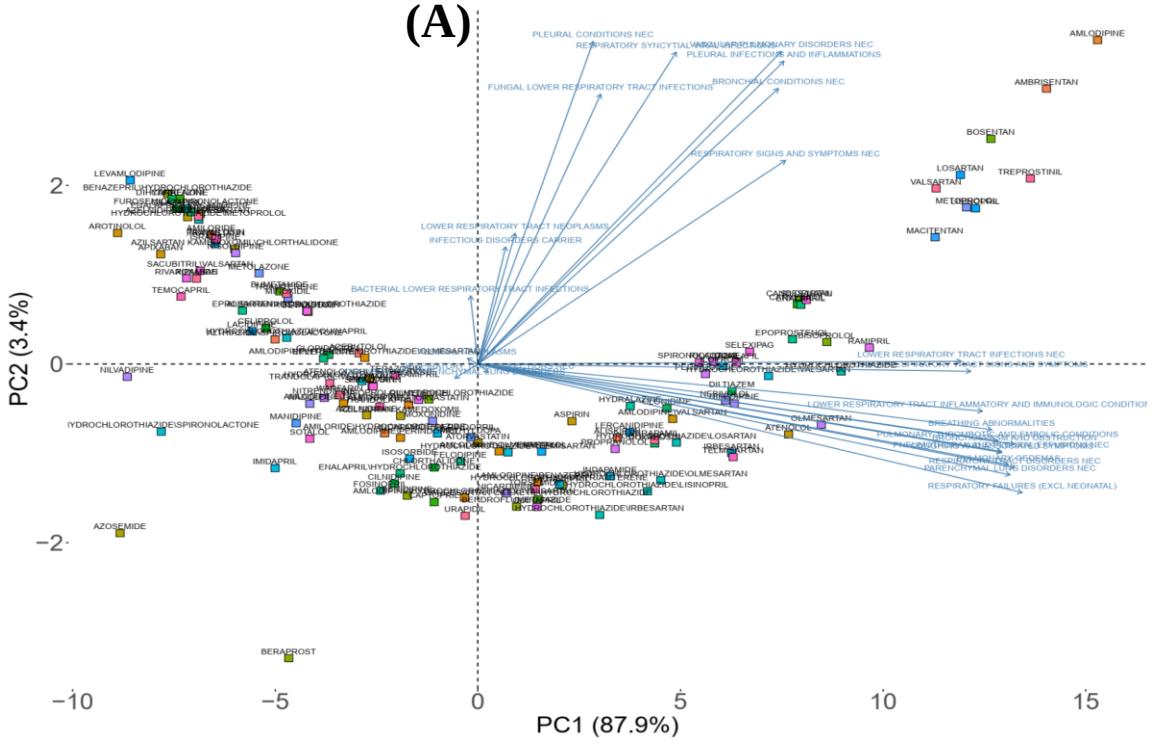
168 **Relative Risk (RR)**

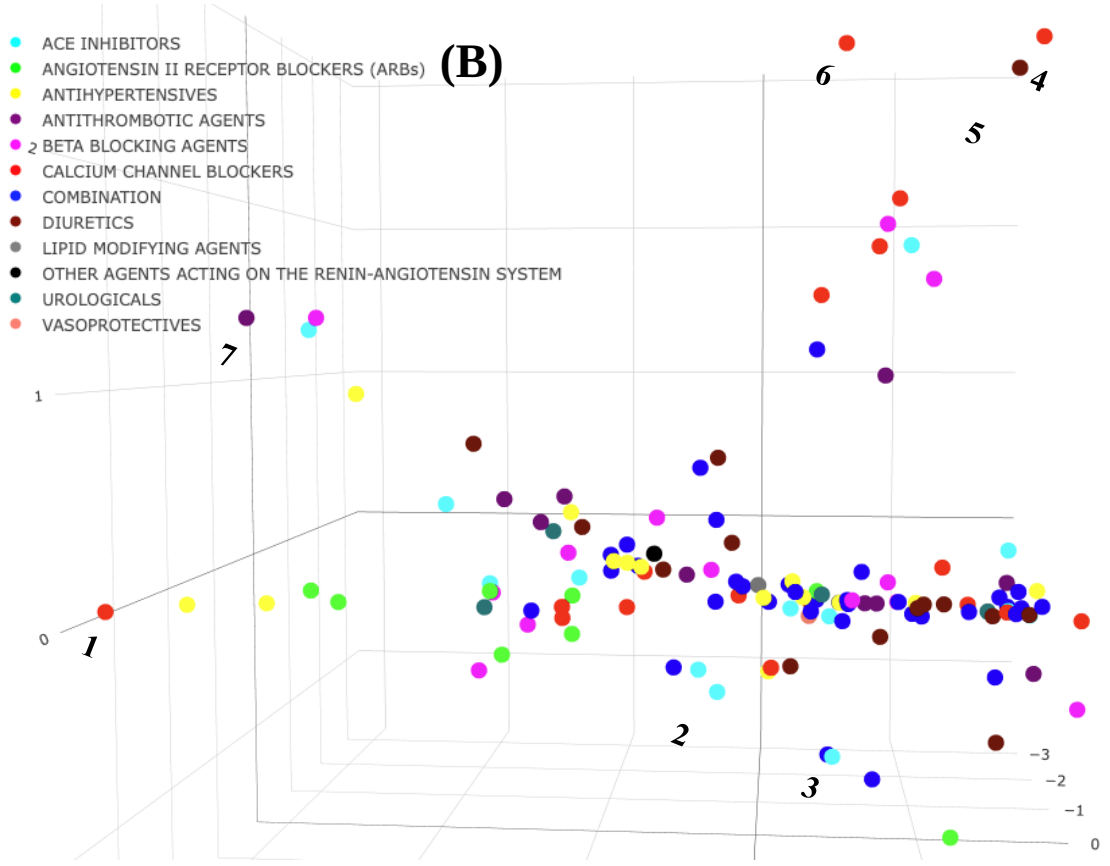
169 One of the frequentist methods, the relative reporting ratio (RR), based on the disproportionality
170 measures of a drug-ADE occurrence compared to other drug-event combinations was applied to
171 evaluate the weighting of drugs. To start our first analysis, we constructed a large contingency
172 table for the entire data from 134 selected drugs based on their frequencies with respect to all 1520
173 reported ADEs in HLT codes from MedDRA. We imposed the assumption that an ADE is selected
174 when RR >2 for a specific drug to assess the drug disproportionality in pharmacovigilance data by
175 observed-expected ratios prior to the EBGM analysis, a more conservative and accurate way of
176 disproportionality evaluation. Taking into account only 25 pulmonary ADEs in HLT codes, we
177 then obtain the results of **Table 2** displaying the top 22 drugs with their corresponding number of
178 pulmonary ADEs when RR >2. The order from the number of pulmonary ADEs is arranged based
179 on the EBGM results after GLASSO elimination and the clustering given in **Table 1** that will be
180 explained below. RR is also utilized to calculate the baseline frequency for EBGM and to construct
181 the PCA as explained below.
182

Principle Component Analysis (PCA)

RR calculated for the expected frequency of 25 pulmonary ADEs associated with 134 drugs prescribed to patients with hypertension was used to generate the matrix for the PCA plot. This helped illustrate how the loadings of pulmonary features could separate drugs in a 2D or 3D space. **Fig 1A** shows 134 drugs in a 2D PCA panel following a V shape scatter plot, no clear separation can be intuitively observed. ADEs (blue text) are also superimposed on the graph to obtain the corresponding loadings, direction, and weights with regards to the drugs. Generally, two clusters of pulmonary issues, one in the direction of the X-axis, and another in the Y-axis played an important role in separating these drugs in the space of PC1 and PC2. Twelve different pulmonary ADEs in HLTS codes (breathing abnormalities, bronchospasm and obstruction, coughing and associated symptoms, lower respiratory tract infections NEC, lower respiratory tract inflammatory and immunologic conditions, lower respiratory tract signs and symptoms, parenchymal lung disorders NEC, pneumothorax, and pleural effusions NEC, pulmonary oedemas, pulmonary thrombotic and embolic conditions, respiratory failures (excl neonatal), and respiratory tract disorders NEC) exhibited similar impact by differentiating these drugs when projected to PC1 (X-axis), and seven pulmonary ADEs in HLTS codes (bronchial conditions NEC, fungal lower respiratory tract infections, pleural conditions NEC, pleural infections and inflammations, respiratory signs and symptoms NEC, respiratory syncytial viral infections, and vascular pulmonary disorders NEC) contributed the most when projected to PC2 (Y-axis). A detailed contribution of all pulmonary variables is given in **Table S1** in **Supporting Information** and will be reviewed in the discussion.

Fig 1B illustrates how the pulmonary ADEs are separated in a 3D space. The first, second, and third principal components, PC1, PC2, and PC3, explain more than 90% of the variation. Drugs from different branches in the 3D plot represent distinctive effects of pulmonary ADEs on the separation. This figure shows the optimal representation of three active variables in biplots acquired by PCA by diminishing the effect of supplementary variables that have no or little influence on the pulmonary ADEs. Consistent with our previous finding (11), Quinapril and Trandolapril in hypertensive patients have a notably higher incidence of pulmonary ADEs compared with its drug class as well as other classes, **Fig 1B**.





213

Fig 1. Principal component analysis of the expected count for 134 drugs (from 12 ATC drug classes) in 2D (A) and 3D (B) spaces using the log expected value of RR, $\log E$. In Panel B, individual drugs are (significantly) separated on the extreme edges are marked by (1) amlodipine, (2) quinapril, (3) trandolapril, (4) nilvadipine, (5) azosemide, (6) azelnidipine, and (7) treprostinil. An interactive figure can be found on the 1DATA home page. Click the following URL to see the figure: https://1data.life/pages/publication/figure1B.html.

214

215 **Empirical Bayesian Geometric Mean (EBGM)**

216 While the RR method is widely utilized due to its simplicity and user-friendly processing, it is
217 difficult to dismiss high variability for infrequent occurrences. The assessment of drugs or ADEs
218 based on RR is variable because of information that the RR methodology does not include,
219 including underreported or overreported events. To assess the effect that the RR methodology has
220 when a small number of ADE occurrences are compared to the whole database, the 5th percentiles
221 from the lower confidence interval of EBGM (EB05) were used as a very conservative alternative,
222 and the results are compared to RR. This assessment was performed using EBGM, is reported
223 similar to the prevalence evaluation using RR values from above. The frequencies of a single drug
224 having multiple ADEs in HLT groups or a single HLT ADE occurrence in multiple drugs were
225 calculated. We then found that the top ten drugs with pulmonary ADEs consisted of AHAs, ATAs,
226 and UAs. Bosentan, tadalafil, treprostinil, and beraprost based on EBGM were ranked substantially
227 higher than their corresponding ranks when using RR, with respect to pulmonary ADEs. This
228 suggests that the conservative, EBGM method with a 5^{th} percentile cut-off will allow for the
229 examination of large datasets of ADEs when high variability is present in the number of ADEs

across drugs or drug classes, and still allow for a robust reporting methodology as compared to the RR methodology. This allows analysis of very large sets of drugs and ADEs (such as approximately 500,000x134 matrix here) without loss of sensitivity or imparting an over-emphasis on ADEs from infrequently prescribed drugs.

GLASSO

The total number of distinct drugs used by patients with hypertension was 134 after filtering out drugs with very low frequency (<0.001) in the PCA section. EBGM data were used to construct the new feature matrix for different drug classes. Then 44 drugs were selected based on two conditions: (1) the lower confidence interval of EBGM, EB05, of drugs was larger than one, and (2) a minimum of two different pulmonary ADEs is associated with each drug, **Table 1**. We found that few drugs in ACEIs, diuretics, and combinations tended to cause pulmonary issues. More than half of the drugs were in ARBs, AHAs, ATAs, and CCBs when considering two different pulmonary ADEs in the HLT level. After two filtering steps, 44 drugs were set as the input for the penalized regression GLASSO. To have an adequate number of correlated drugs, the tuning parameter λ of GLASSO was adjusted to shrink the less associated drugs to 0, which accounted for 50% of the selected drugs. The remaining 22 drugs selected by the GLASSO method based on Pearson correlation were classified using the therapeutic group Cardiovascular System (C01: Cardiac Therapy, C02: Antihypertensives, C03: Diuretics, C04: Peripheral Vasodilators, C05: Vasoprotectives, C07: Beta Blocking Agents, C08: Calcium Channel Blockers, and C10: Lipid Modifying Agents), except for agents acting on RAS, which are the pharmacological subgroup C09 (C09A: ACE Inhibitors, C09B: Ace Inhibitors, Combinations, C09C: Angiotensin II Receptor

Blockers (ARBS), and C09X: Other Agents Acting On the Renin-Angiotensin System,), the third level was applied to classify the RAS drugs since RAS drugs are the frontline agents in hypertension, **Table 1**.

Table 1. Drug class after applying first the two filtering rules to obtain 44 drugs and then the elimination process from the penalized regression GLASSO to obtain 22 drugs.

Drug class	# Reports (Total 612,733)	# Drugs after initial filtering (Total 134)	# Drugs correspond to ≥ 2 ADEs in HLT codes when EB05>1 (Total 44)	Drugs using GLASSO (Total 22)
ACEIs	69,327	13	3	1
ARBs	87,415	8	5	3
Other RAS agents	3,471	1	0	0
Other Antihypertensive	120,425	14	7	4
Antithrombotic Agents	67,767	10	7	3
Beta Blocking Agents	74,574	13	3	1
Calcium Channel Blockers	86,399	18	10	6
Diuretics	29,394	14	3	1
Lipid Modifying Agents	2,634	4	0	0
Urologicals	18,186	4	2	2
Vasoprotectives	909	1	0	0
Combinations	52,232	34	4	1

Circos plot

The drug-drug correlation matrix with shrinkage is displayed in a circular layout, depicting drug class and associations between drugs from different classes (**Fig 2**). For drugs in ACEIs, ARBs, AHAs, and BBAs, no association was observed between drugs within the same class. More within-class associations were depicted in AHAs, CCBs, and combinations. **Fig 2A** shows the association between the remaining 22 drugs after then the elimination process from the penalized regression GLASSO. After these stringent filtering methods, drug classes exhibit very low significant

correlations between drugs from the same class. This result is observed in **Fig 2A** by very few associations between drugs in the same class. Therefore, drug clustering using the RCM reordering method was employed in **Fig 2A,** with bridges connecting associated drugs. Without a doubt, this analysis corroborates our hypothesis that drugs from the same ATC class may have different pulmonary ADE profiles.

Given the 22 drugs selected by GLASSO, **Table 2** shows the assessment of drugs exclusively with respect to their pulmonary events. In the second column, # pulmonary ADEs defines the number of drug-ADE pairs from EBGM, which are depicted in the following section. Similarly, # pulmonary ADEs in the fourth column denotes the results when RR is larger than two. The order

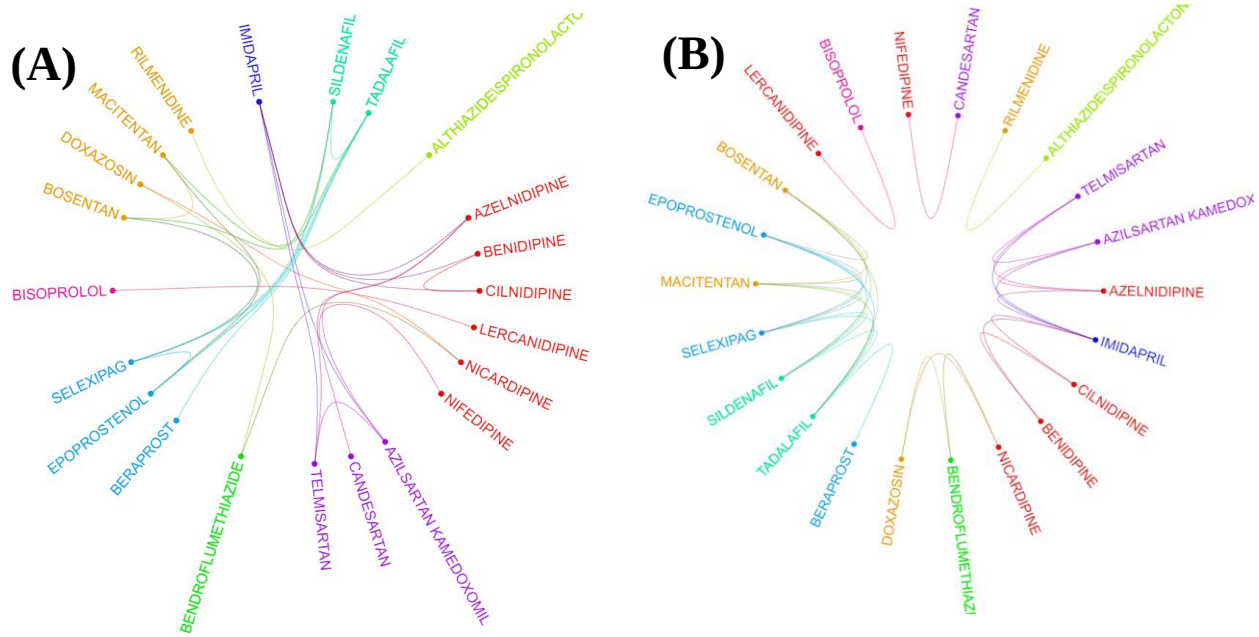


Figure 2. Two layouts of Circos plot for 22 hypertensive drugs selected by GLASSO. Circos plots of drugs were obtained based on the EBGM matrix after applying GLASSO. Edge bundling linkages for better visualization and drugs were selected by GLASSO with edge bundling. Grouped drugs based on their classes were assigned the same color based on their classes (A). Applying RCM reordering and edge bundling for grouping drugs based on the ATC class and edge bundling (B).

of drugs listed in **Table 2** is calculated based on the original 44 drugs from the EBGM scores and here we only show the arrangement for the remaining 22 drugs out of 44 drugs. Beraprost showed 13 pulmonary ADE profiles reported more commonly than other drugs used for patients with hypertension based on the estimated RR. Macitentan and Selexipag were equally located in the second most commonly reported drugs, each of which with 10 pulmonary ADEs. In contrast, beraprost was corrected from being the top drug with most pulmonary issues and then ranked down to the tenth location by EBGM. The assessment for bosentan and tadalafil also changed radically when the comparative analysis was done using RR or EBGM.

Table 2. The number of pulmonary ADEs when RR larger than two or the 5th quantile of EBGM, EB05, large than two after GLASSO filtering process implemented in **Table 1**.

Drug	# pulmonary ADEs	Order by EBGM	# pulmonary ADEs	Order by RR
Macitentan	16	1	10	2
Bosentan	14	2	5	11
Epoprostenol	11	4	9	4
Selexipag	10	5	10	2
Sildenafil	10	6	7	6
Tadalafil	10	7	3	44
Beraprost	7	10	13	1
Nifedipine	5	13	5	11
Candesartan	4	16	3	34
Althiazide\Spironolactone	3	20	4	18
Bisoprolol	3	21	#N/A	#N/A
Imidapril	3	24	5	11
Azelnidipine	2	30	4	23
Azilsartan Kamedoxomil	2	31	3	32
Bendroflumethiazide	2	32	3	33
Benidipine	2	33	5	11
Cilnidipine	2	34	5	11
Doxazosin	2	36	3	36
Lercanidipine	2	39	1	90
Nicardipine	2	40	5	11
Rilmenidine	2	42	#N/A	#N/A
Telmisartan	2	43	4	30

From GLASSO and **Table 2**, we can now obtain the ADE profiles in HLT groups for each drug in the newly identified group class, which we called GLASSO (GL) Clusters. The ADEs together with the drug classes from ATC and GL Clusters based on EB05>1 are arranged in **Table 3** and depicted by an arc diagram in **Fig S3**, **Supporting Information**. It is apparent from **Fig 2** and **Table 3** that GL Cluster 1 consists of most associated drugs with most pulmonary ADEs assessed by EBGM.

Table 3. Comparative analysis of each drug and associated pulmonary ADEs based on our new classification from different GLASSO (GL) Clusters

Drug	Drug Class	ADEs for EB05> 1 (n) *	GL Cluster
Macitentan	AHAs	1-15,17 (16)	1
Bosentan	AHAs	1,2,4-15 (14)	1
Epoprostenol	ATAs	1,2,4-9,11,12,15 (11)	1
Selexipag	ATAs	2,4-12 (10)	1
Sildenafil	UAs	1,2,4-12 (10)	1
Tadalafil	UAs	1,2,4-12 (10)	1
Beraprost	ATAs	1,2,5-9 (7)	1
Nifedipine	CCBs	1-3,15,16 (5)	2
Candesartan	ARBs	1,3,14,16 (4)	2
Althiazide\Spironolactone	COMBs	4,10,11 (3)	3
Rilmenidine	AHAs	4,10 (2)	3
Bisoprolol	BBAs	1,2,14 (3)	4
Lercanidipine	CCBs	1,14 (2)	4
Imidapril	ACEs	1-3 (3)	5
Azelnidipine	CCBs	1,3 (2)	5
Azilsartan Kamedoxomil	ARBs	1,3 (2)	5
Benidipine	CCBs	1,2 (2)	5
Cilnidipine	CCBs	1,2 (2)	5
Telmisartan	ARBs	1,3 (2)	5
Bendroflumethiazide	TDAs	3,13 (2)	6
Doxazosin	AHAs	3,13 (2)	6
Nicardipine	CCBs	3,13 (2)	6

***** Below we have ADEs found for each drug:

1. Parenchymal Lung Disorders NEC
2. Pneumothorax and Pleural Effusions NEC
3. Lower Respiratory Tract Inflammatory and Immunologic Conditions
4. Respiratory Tract Disorders NEC
5. Breathing Abnormalities
6. Lower Respiratory Tract Signs and Symptoms
7. Pulmonary Oedemas
8. Respiratory Failures (Excl Neonatal)
9. Vascular Pulmonary Disorders NEC
10. Bronchospasm and Obstruction
11. Coughing and Associated Symptoms
12. Respiratory Syncytial Viral Infections
13. Bronchial Conditions NEC
14. Pulmonary Thrombotic and Embolic Conditions
15. Lower Respiratory Tract Infections NEC
16. Fungal Lower Respiratory Tract Infections
17. Pleural Infections and Inflammations

297 **Friedman test and multiple pairwise comparisons**

298 To test the significant difference between drugs grouped by the original ATC classes and the GL
299 Clusters, which were from a shrinkage correlation matrix, a non-parametric Friedman test was
300 applied to compare separately the magnitude of difference when drugs in the same group for the
301 ATC classes or the GL Clusters. **Table 4** summarizes the results of the p-value for different
302 comparative analyses in the ATC classes or the GL Clusters. A p-value of 0.199 indicates that no
303 differences in EBGM of pulmonary ADEs for different drugs in GL Cluster 1 when excluding
304 Tadalafil. Similarly, GL Clusters 2, 3, 4, 5, and 6 showed no significant differences in EBGM
305 respectively (**Table 4**). However, given the original ATC class drugs belonging to, the Friedman
306 test did show significant differences in six of the ATC class before GLASSO. The same test was
307 applied to 22 drugs selected from GLASSO, only drugs in UAs showed no significant differences
308 in EBGM of pulmonary ADEs. This shows that instead of grouping drugs from the same ATC
309 class, isolated groups from GLASSO showed homogeneity.

310

311 **Table 4.** The Friedman test for drugs in ATC class and GLASSO class.

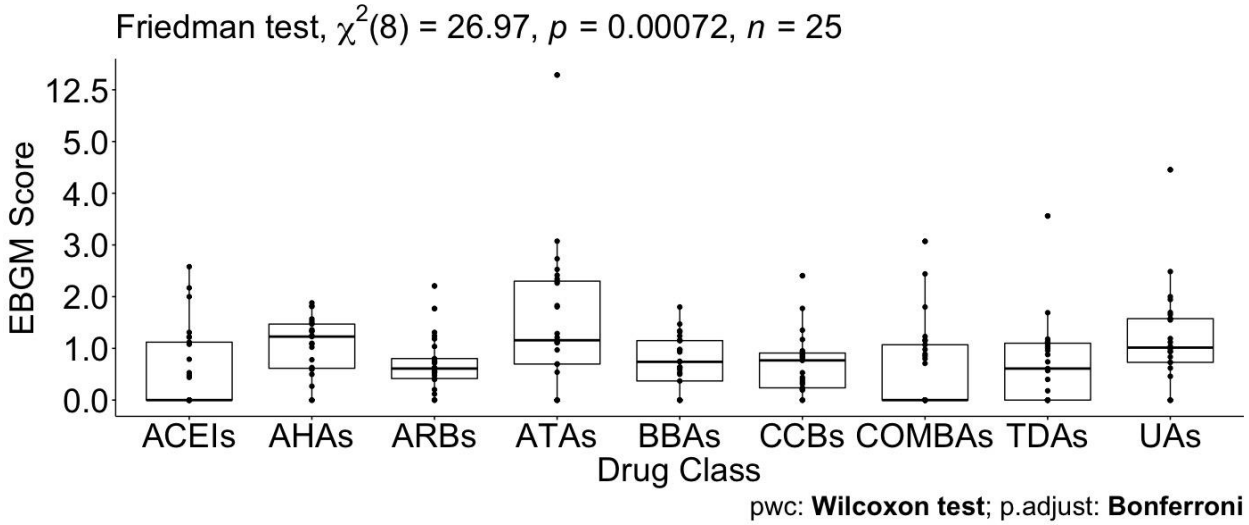
ATC Class	p-value (44 drugs)	p-value (22 drugs)
ACEIs	0.271	-
ARBs	**< 0.001**	**< 0.001**
AHAs	**< 0.001**	**< 0.001**
ATAs	**< 0.001**	**< 0.001**
BBAs	**0.0232**	-
CCBs	**0.001**	**0.001**
COMBs	0.236	-
TDAs	**0.0329**	-
UAs	0.127	0.127

GL Cluster	The p-value for 22 drugs
1	**< 0.001** (0.199, when excluding Tadalafil)
2	0.110
3	0.884
4	0.346
5	0.127
6	0.0522

312

313 Pairwise drug class comparisons based on ATC class are shown for all the pairs (nine drug classes:
314 ACEIs, ARBs, AHAs, ATAs, BBAs, CCBs, COMBs, TDAs, and UAs) in **Table S5-A** in

315 **Supporting Information**. The EBGM scores from the pulmonary ADE profiles were statistically
316 significant for the nine ATC classes using the Friedman test (*p-value = 0.0072*, **Fig 3**). Pairwise
317 comparisons showed no significant differences among any two ATC classes from the adjusted p-
318 value (**Table S5-A** in **Supporting Information**). However, using drug class determined by
319 GLASSO, Wilcoxon signed-rank test between groups revealed significant differences in EBGM
320 of pulmonary ADEs between GL Cluster 1 and GL Clusters 3, 4, and 5, respectively, compared to
321 the pairwise comparisons between ATC groups, **Table S6-A** in **Supporting Information** and **Fig**
322 **4**. Drugs in GL group 1 showed significantly higher EBGM regarding pulmonary events. Friedman
323 test confirming EBGM profile of selected drugs from GLASSO could be used for comparative
324 analysis of drugs regarding certain indications.
325



326

327 **Fig 3**. **Pairwise Wilcoxon signed-rank test between different ATC classes.** No pairwise
328 significant comparison was found similar to **Table S5-A** in **Supporting Information**. But the
329 group comparison was highly significant, *p-value = 0.00072*.
330

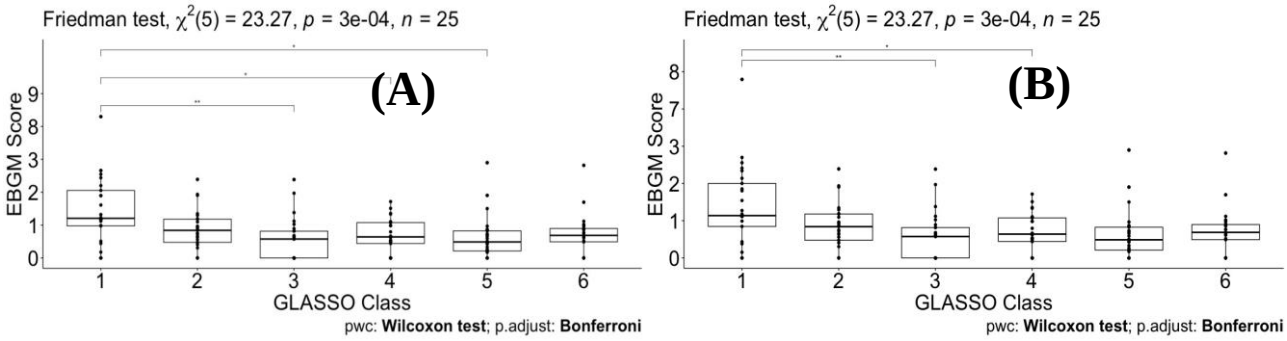


Fig 4. Pairwise Wilcoxon signed-rank test between different classes defined by GLASSO (A) and pairwise Wilcoxon signed-rank test between different classes defined by GLASSO excluding Tadalafil (B).

DISCUSSION

The future of large-scale biomedical science is data-driven decision-making and AI knowledge-based development and validation. AI-enabled technologies can help in better understanding disease indication occurrence and disease determinants or patterns. Quantitative methods have countlessly been applied in various medical fields of study, e.g. measurement of disease frequency, prevalence or incidence; evaluation of source of bias and variation of observational studies; multivariate data analysis of risk factors such as applied logistic regression analysis; machine learning for survival analysis or analysis of time at risk (survival) data; boosting power for clinical trials using AI-assisted analysis, etc. In our study, we aimed to apply AI-driven methodologies involving EBGM and GLASSO techniques in predicting SARS-Cov-2 comorbidity for high-risk populations with hypertension.

Quantitative methods, i.e., PPR, RR, ROR, EBGM have been used to detect signals for spontaneously reported data. After filtering data by quantitative methods, we proposed that selected drug-ADE based on drug association mechanism would be a valuable procedure for clinical review and comparison of similar drugs with similar ADE profiles. In this study, we demonstrated a systematic way of filtering and selecting data that addresses the noise inherent to such data. None of these methods are free from including false positive and false negative signals, however, EBGM and the Information Component (IC) are recommended over other quantitative methods when evaluating by mean average precision (16). This helped us to build a model to understand the bias-variance tradeoff to achieve a balance between the two desirable but incompatible features. Given the absence of a gold standard, no available method is overwhelmingly better than the others (18). The confirmatory methods proposed in this study (GLASSO and Friedman test) for assessing quantitative methods could reveal the strengths and drawbacks of the methods.

Drugs from different branches in the 3D plot represent distinctive effects of pulmonary ADEs on the separation. For example, PC3 is dominated by fungal, PC2 by more pleural and vascular, and PC1 by respiratory tract effects (see **Table S1** in **Supporting Information**). PCs were constructed using the expected counts of a drug and a pulmonary ADE through a linear combination. The spatial separation of drugs indicated that drugs at the perimeter of each branch (numbered) performed disparately regarding pulmonary ADE profiles, suggesting they may not best be managed as having ADE profiles defined by their class. This figure shows the optimal representation of three active variables in biplots acquired by PCA by diminishing the effect of supplementary variables that have no or little influence on the pulmonary ADEs. Using the

371 Friedman test, we found that these separated drugs have significant differences between their drug
372 classes and compared to other drug classes.
373
374 The consistency of the Friedman test and GLASSO to capture EBGM signals of drugs used in
375 small and large populations could be a beneficial tool for drug comparative analysis. Xu et al. (20)
376 and Stafford et al. (10) have already applied two methods in pharmacovigilance to animal and
377 human data separately. This study proposed and successfully combined penalized regression
378 together with the non-parametric Friedman test in considering to better visualization of drug-drug
379 and drug-ADE associations. The RR method is widely utilized due to its simplicity and user-
380 friendly processing. RR, however, may be highly variable for small occurrences of an event. Our
381 assessment of drugs or ADEs based on RR showed unstable performance, especially for hidden
382 information. The estimates of small occurrences compared to the whole database were also inflated
383 for events. To correct these issues, we introduce 5^{th} percentiles from the lower confidence interval
384 of EBGM (EB05) used as a conservative alternative compared to RR.
385
386 EBGM detected that 16 out of 25 pulmonary ADEs in MedDRA databases were associated with
387 macitentan, followed by bosentan with 14 pulmonary ADEs. Both of these drugs belong to the
388 endothelin receptor antagonist class of drugs and are utilized in pulmonary arterial hypertension
389 to prevent vasoconstriction, fibrosis, and inflammation on vascular endothelium and smooth
390 muscle (32). Both drugs are proposed to curb the pulmonary vascular resistance to prevent right
391 heart failure and death, however, pulmonary ADEs of both drugs can be of major concern
392 compared to the outcomes of several other antihypertensives agents we utilized in this study. At
393 the same time, because these two medications are used in a disease affecting pulmonary function

and commonly reported ADEs to include therapeutic failure, these drugs were not surprising to emerge among the highest with reported pulmonary ADEs and, in fact, they serve to validate the methods utilized in this paper. Conversely, doxazosin and rilmenidine were found to have the least pulmonary ADEs in selected drugs from hypertension patients since only two ADE signals were detected based on EBGM. Although it can be used in hypertension, doxazosin is primarily utilized for men with benign prostatic hyperplasia and works by blocking alpha-adrenergic receptors in the vascular smooth muscle, resulting in vasodilation (33). Additionally, studies in countries outside of the US suggest that rilmenidine, a sympatholytic, has a favorable ADE profile for patients with hypertension and diabetes, it is not approved in the US (34). After excluding GL Cluster 1, we did see almost the same results for the remaining GL clusters. It is also worth mentioning here that the results are shown in **Tables S2**, **S3**, **S4**, **SB-5**, and **S6-B** as well as **Figs S1** and **S2** in **Supporting Information**.

The second group found by EBGM and GL clustering consisted of two drugs from CCBs (nifedipine) and ARBs (candesartan) grouped in combination (**Fig 2**) and showed four similar pulmonary ADEs: parenchymal lung disorders NEC, pneumothorax and pleural effusions NEC, lower respiratory tract inflammatory and immunologic conditions, and fungal lower respiratory tract infections. Several studies based on these drugs showed effective combination and blood pressure lowering effects in patients with hypertension and appeared an improved side effect profile in comparison with single-agent monotherapy (35-38). This is undoubtedly an interesting finding resulted from our EBGM analysis and demonstrated how these two drugs can be combined and investigated for pharmacokinetic assessment in drug development including bioavailability

and bioequivalence, drug safety pharmacovigilance, and efficacy and comparative tolerability of the combination of nifedipine and candesartan (39, 40).

Our previous work showed that quinapril and trandolapril were significantly different from other ACEI and ARB drug classes (11). Separating from its drug class was initially observed in **Fig 1** when the PCA biplot was performed. However, these two drugs will not be present when more precautionary methods are applied for several reasons: (1) the dataset is no longer the same as before which contain only ACEIs or ARBs. (2) The methods are very different. (3) Several other drugs and ADEs are added to the study, 134 as opposed to only 16 drugs. (4) In our previous work, we only focused on analyzing 13 pulmonary ADEs at the PT level; however, in the current study, we obtained and compare 25 ADEs in HLT groups and each HLT contains several PT ADEs. To be more accurate, ADEs for the 25 ADEs in HLT groups contains approximately 200 different PT vs only 13 ADEs. (5) The whole purpose of this study was to use EBGM as a much more accurate method compared to RR and RR estimation is also better than the PRR method used before. (6) The implementation of the filtering process of penalized regression GLASSO helps eliminate the insignificant and noise-driven reports.

Two drugs, tadalafil and sildenafil, are also used for the modulation of dopaminergic pathways and modifying risk factors to prevent and treat erectile dysfunction. Using our database when curating the data for the medicinal products of these drugs and checking their active ingredients of tadalafil and sildenafil, the top products are found to be Adcirca (*n=32446*) and Revatio (*n=21358*) marketed for the treatment of pulmonary arterial hypertension, respectively, and Cialis (*n=15623*) and Viagra (*n=20820*) marketed to treat erectile dysfunction, respectively. We also assessed

whether these drugs only show up at high doses or not. This also confirmed that the dose has an insignificant effect on the outcome of ADEs, data are given in **Table S7** in **Supporting Information**.

As part of our future work, it is worth mentioning that this study aimed to reveal the potential risk of patients using hypertensive drugs in terms of pulmonary issues. Our database will be updated with MedDRA 24.0 that contains the new COVID-19 terms due to its outbreak. It has encouraged us to involve terms related to viral infections that facilitate the capture of ADEs caused by COVID-19 in patients with hypertension in the near future. In addition, the pulmonary ADEs of HLT codes in this study were filtered by setting the highest level, system organ class (SOC), with the focus on respiratory, thoracic, and mediastinal disorders (*n=28*), and infection class containing viral infection (*n=2*). We plan to include ADEs from the class of Blood and lymphatic system disorders such as thrombosis, coagulation, or platelet disorders. In the big data era, as the spontaneous reports from different data sources including the FDA FAERS database (21), the Vaccine Adverse Event Reporting System (VAERS) (41, 42), and the WHO International Database are increasing in size; drug profiles based ADEs can be established based on quantitative methods, retrieving the signals, or detecting new signals in large numbers of reports by different methods with the combination of clinical review is need for pharmacovigilance.

METHODS

To derive the desired information from datasets, there are a few main methodological steps in this study. In the following, we briefly illustrate procedures in our workflow integrated by machine

461 learning where some preprocessing points are first presented in **Fig 5**. This figure summarizes the

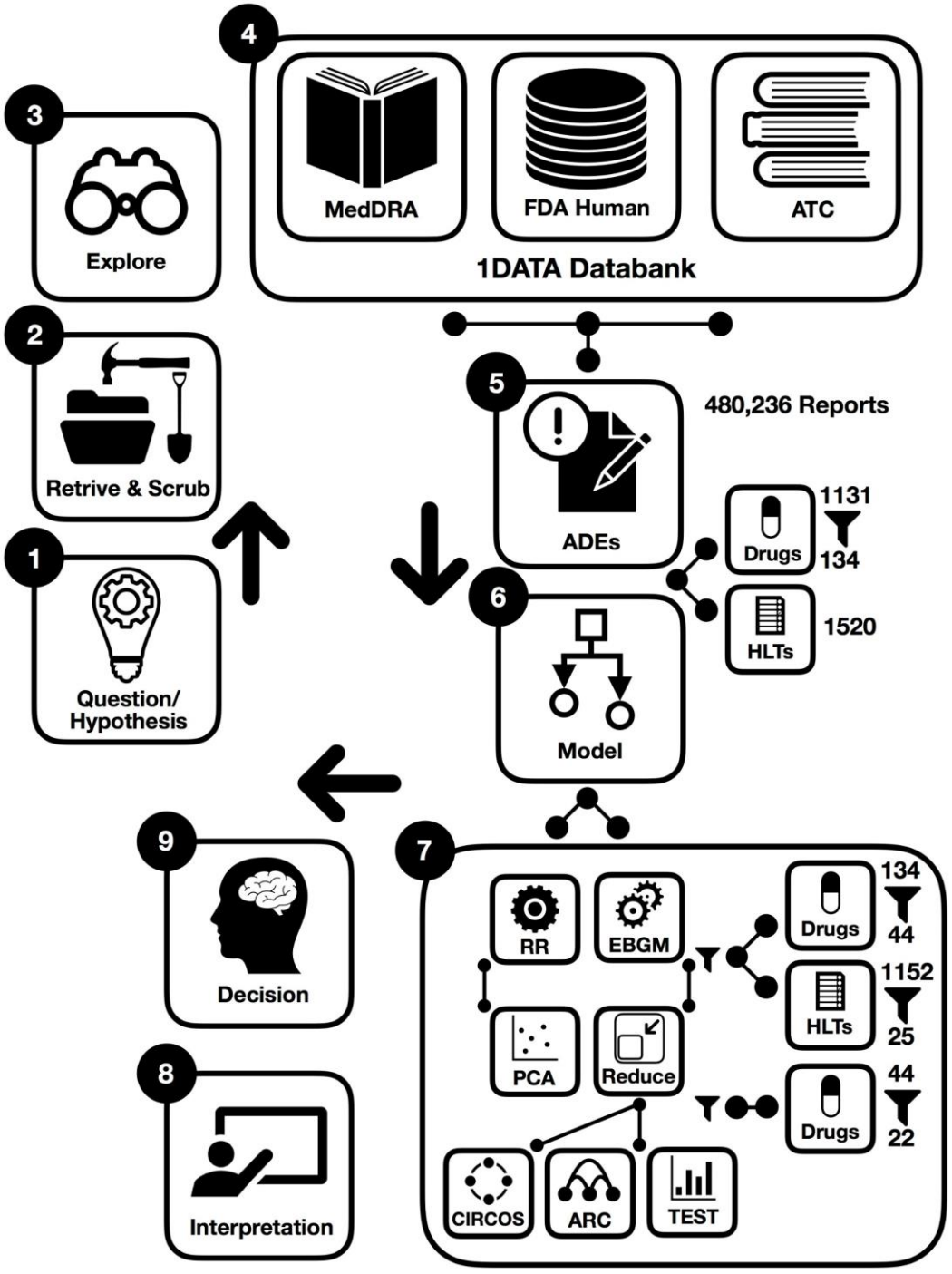


Fig 5. Workflow of data-driven methodology for pulmonary symptomology in hypertension using machine learning models from preprocessing and dictionary creation to storing tables in the database an analysis.

steps in the preparation and analysis of the ADE database to make a decision and interpret our results, each step is detailed in the following subsections:

1. Working hypothesis: drugs from the same drug class could have different pulmonary ADE profiles affecting outcomes in acute respiratory illness, with potential implication in SARS-CoV-2 infection.
2. Designing error correction techniques for data scrubbing and retrieval.
3. Implementing data exploration technique for initial data analysis to visually explore and understand the characteristics of the data from post-marketing drug safety surveillance.
4. Data curation and annotation to organize and integrate data collected from various sources from the FDA, MedDRA, and ATC classification. This phase entails annotation, organization, clustering, and presentation of the assorted data types from the 1DATA databank.
5. ADE-associated information retrieval for patients with hypertension provides massive collections of reports to investigate adverse drug events based on comparative population data analysis.
6. Integration of machine learning models.
7. Acquiring results after data preprocessing and cleansing that significantly reduces the size of data and eliminates insignificant and noise-driven reports.
8. and 9. Enhancing decision and interpretation via data-driven machine learning to help identify incidences of pulmonary ADEs for potential therapy and confounding factors that may have implications for treating patients diagnosed with COVID-19, respectively.

As a part of data cleaning, we were also challenged by multiple technical issues when combining drugs: (i) there were many drugs' names that did not track a specific standard. (ii) Formulations of

485 the same active ingredient with different generic or brand names for different routes of
486 administration created confusion in collecting data (for instance, Revatio, Viagra, sildenafil,
487 sildenafil citrate, APO sildenafil, sildenafil film-coated tablet, sildenafil citrate Aurobindo pharma,
488 sildenafil Amneal Pharmaceuticals, Teva sildenafil, sildenafil Pfizer, sildenafil Greenstone,
489 sildenafil Hormosan Filmtabletten, Revathio, sildenafil SUP, etc.). For this purpose, we combined
490 drugs with or without salt, alcohol, etc. from different generic names and brand names.
491

492 **Data Integration**

493 The data were integrated into the 1DATA databank (www.1DATA.life) (20) from multiple
494 sources, including the Food and Drug Administration (FDA) Adverse Drug Events Reporting
495 System (FAERS), the Medical Dictionary for Regulatory Activities (MedDRA), and the ATC
496 classification system. The FAERS database consists of voluntarily or mandatorily reported ADEs
497 from healthcare professionals, manufacturers, and consumers; encompassing drug-related adverse
498 occurrences pertaining to standard use, medical error, overdose, or product quality (21). ADE
499 reports from FAERS are typically coded in accordance with the Preferred Term (PT) level of
500 MedDRA. The MedDRA provides an internationally recognized hierarchical terminology [System
501 Organ Class (SOC), High-Level Group Term (HLGT), High-Level Term (HLT), PT, and Lowest
502 Level Term (LLT)] for coding ADE reports (22). This study aggregates raw ADE reports to terms
503 from the HLT and SOC levels. ATC classification is likewise an internationally applied
504 hierarchical system for active drug substances based on site of action (organ or system) and
505 mechanistic properties (therapeutic, pharmacological, and chemical). Drugs in this study were
506 grouped according to ATC classification. Data integration into 1DATA occurred through the

PostgreSQL 13.2 version (PostgreSQL Global Development Group), which allows concatenation of drug and ADE information (20, 23).

Adverse Drug Event (ADE)

ADEs cause approximately 30 billion dollars a year of added health care expenses, along with negative—including fatal—health outcomes (20). The practice of prescribing drugs based on information from drug preapproval labeling may misrepresent or deprecate the incidence and prevalence of specific ADEs. The FDA defines the term 'adverse event' as: "any untoward medical occurrence associated with the use of a drug in humans, whether or not considered drug related, including the following: an adverse event occurring in the course of the use of a drug product in professional practice; an adverse event occurring from drug overdose whether accidental or intentional; an adverse event occurring from drug abuse; an adverse event occurring from drug withdrawal; and any failure of expected pharmacological action" (24, 25).

Relative Risk (RR)

The main method used in this study, Bayesian shrinkage, is based on a baseline frequency, which is the relative risk or relative reporting ratio

$$\mathrm{RR}_{ij} = \frac{N_{ij}}{E_{ij}}.$$

It compares a drug-ADE count, N, to its expected count, *E*. For instance, when N_{ij}/E_{ij} is equal to 100, then $drug_i$ and ADE_j occurred 100 times as frequently as the baseline frequency represents. A huge difference of occurrences between two drug-ADE pairs might lead to similar RR due to *E* in

the denominator, even statistically the same, but the frequency illustrates sampling variation. When more events of ADE_j are caused by $drug_i$ higher than the same ADE in the database, $RR_{ij}>1$. Drug-ADE surveillance should be triggered when large RR scores show up for specific drug-ADE pairs. However, the variability of RR for small counts drug-ADE pairs is unreliable, the high value of RR might be accidental.

Principal Component Analysis (PCA)

Principal component analysis (PCA) was obtained based on the log expected value of RR, log(E), to analyze ADEs for different drugs, to reduce the features from the drug-ADE matrix. The distinct clusters from PCA plots were used to compare the similarities of drugs based on E. PCA was conducted using built-in function *PCA* in R (R 3.6.3 version, R Core Team, GNU GPL v2), and PCA biplots were produced using the R package *factoextra,* and 3D PC plots were produced using R package *plotly*.

Gamma-Poisson Shrinker (GPS)

DuMouchel (17) proposed an empirical Bayes approach based on the Gamma-Poisson Shrinker (GPS) algorithm to bring down the inflated value of RR due to small counts without impacting RR associated with large counts. Thus, the drug profile based on ADE could be reconstructed with reduced variation in RR. GPS redefines RR_{ij} as $\lambda_{ij}=\mu_{ij}/E_{ij}$ drawn from a prior distribution with a mixture of two gamma distributions, μ_{ij} is the mean of the Poisson distribution of counts for $drug_i$ and ADE_j

549 $$\text{prior: } \Pi(\lambda|\alpha_1, \beta_1, \alpha_2, \beta_2, P) = P \times \text{gamma}(\lambda|\alpha_1, \beta_1) + (1 - P) \times \text{gamma}(\lambda|\alpha_2, \beta_2)$$

550 which then gives the posterior probability from the components of the mixture model:

551 $$\text{posterior: } \lambda|N = n \sim \Pi(\lambda|\alpha_1 + n, \beta_1 + E, \alpha_2 + n, \beta_2 + E, Q_n)$$

552 GPS shrinks RR scores by using EBGM from

553 $$\text{EBGM} = e^{E(\log \lambda)}.$$

554 The shrinkage abates vagueness by reducing RR scores to a conservative level, which helps to
555 alleviate false-positive signals, avoiding arbitrary drug-ADE assessment. The R package
556 *openEBGM* was used to implement the GPS method (26).

557

558 **Correlation Matrix and GLASSO**

559 The profile of each drug comprises EBGM of all ADEs. The Pearson correlation matrix was
560 constructed based on the EBGM between pairs of drugs. The vector

561 $$EB_i = (EB_{i1}, EB_{i2}, \dots, EB_{ip})$$

562 for $i \in \{1,2,\cdots,n\}$ denotes the EBGM corresponding to $drug_i$. The Pearson correlation method
563 determines the associations between pairwise vectors of reported drugs, which are the elements in
564 the correlation matrix. This adjacency matrix was highly dense $(n \times n)$, and it is difficult to graph
565 the network when too many drugs (1131) are present. A penalized regression method, graphical
566 least absolute shrinkage and selection operator (GLASSO), was then introduced to encourage
567 sparsity in the adjacency matrix, in order to plot high dimensional graphs from the correlation
568 matrix (27). An R package called *huge* was utilized to perform GLASSO (28).

569

570 **Drug-ADE Correlation Diagram**

571 The MedDRA hierarchy is multi-axial, for example, "influenza" is from the PT level and is
572 encompassed within two SOC levels "Respiratory, thoracic and mediastinal disorders" and
573 "Infections and infestations". Therefore, the columns of EBGM calculations in the drug-ADE
574 matrix involve HLTs from the "Respiratory, thoracic and mediastinal disorders" and "Infections
575 and infestations" levels. For better visualization, ADE columns of one drug were put in a block
576 with other rows being zeros. The dimension of a drug-ADE matrix was expanded from $(m \times q)$ to
577 $(m \times mq)$ where $m(<n)$, and $m=22$ denotes the number of drugs selected by GLASSO from original
578 $n=44$ drugs, and $q=17$ denotes selected ADEs described above.
579

580 **Reverse Cuthill-Mckee Algorithm**

581 Reverse Cuthill-McKee (RCM) is a bandwidth and profile reduction method, which permutes a
582 sparse matrix into a band matrix with vertices reordered close to the diagonal (29). RCM in this
583 study implemented in MATLAB R2019b (MathWorks Inc., Natick, MA, USA) was applied to
584 arrange the connections between drugs and ADEs to encourage fewer crossings in Circos plots
585 and arc diagrams. Circos plots and arc diagrams were generated using the R packages *edgebundleR,*
586 *igraph, and ggrpah* (30).
587

588 **Friedman Test**

589 Using SAS (SAS University Edition version 9.4, North Carolina, U.S), sample differences among
590 antihypertensive drug groups according to therapeutic main group ATC (ACEIs, ARBs, BBAs,

591 CCBs, and TDs) were evaluated for a pairwise comparison analysis with the assumption that data
592 were not normally distributed using the non-parametric Friedman test for two independent
593 unequal-sized data. The Friedman test was also applied to perform multiple comparison tests (*P*-
594 value for statistical significance < 0.05). Pairwise comparison analysis was completed in SAS. The
595 significance level of comparing drug classes against each other was adjusted using a rigorous
596 paired Wilcoxon signed-rank test with Bonferroni correction to control family-wise type I error
597 (31).
598

599 ## Author contributions

600 **Xuan Xu:** Data Curation, Investigation, Methodology, Software, Validation, Visualization,
601 Writing – Original Draft Preparation; **Jessica Kawakami:** Investigation, Conceptualization,
602 Writing – Original Draft Preparation; **Nuwan Indika Millagaha Gedara:** Data Curation,
603 Investigation, Methodology, Software, Writing – Review & Editing; **Jim Riviere:** Investigation,
604 Project Administration, Conceptualization, Writing – Review & Editing; **Emma Meyer:**
605 Investigation, Conceptualization, Writing – Review & Editing; **Gerald J. Wyckoff:** Funding
606 Acquisition, Project Administration, Investigation, Conceptualization, Writing – Review &
607 Editing; and **Majid Jaberi-Douraki:** Conceptualization, Data Curation, Formal Analysis,
608 Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Software,
609 Supervision, Validation, Writing – Original Draft Preparation, Writing – Review & Editing.
610

Data Availability

The source code and data used to produce results and analyses presented in this manuscript are available at https://1data.life/pages/publication/data_driven_methodology_COVID19_related_pharmacovigilance/

References

1. Organization WH. Weekly Epidemiological Update—1 December 2020. 2020.

2. WHO Coronavirus (COVID-19) Dashboard [Available from: https://covid19.who.int/.

3. Oran DP, Topol EJ. Prevalence of asymptomatic SARS-CoV-2 infection: a narrative review. Annals of internal medicine. 2020;173(5):362-7.

4. CDC. Coronavirus Disease 2019 (COVID-19): People with Certain Medical Conditions [Available from: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html.

5. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. The lancet. 2020;395(10229):1054-62.

6. James PA, Oparil S, Carter BL, Cushman WC, Dennison-Himmelfarb C, Handler J, et al. 2014 evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the Eighth Joint National Committee (JNC 8). Jama. 2014;311(5):507-20.

7. Wiese O, Allwood B, Zemlin A. COVID-19 and the renin-angiotensin system (RAS): A spark that sets the forest alight? Medical Hypotheses. 2020;144:110231.

8. Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature. 2003;426(6965):450-4.

9. Smith MD, Smith JC. Repurposing Therapeutics for COVID-19: Supercomputer-Based Docking to theSARS-CoV-2 Viral Spike Protein and Viral Spike Protein-Human ACE2 Interface.

10. Stafford EG, Riviere JE, Xu X, Kawakami J, Wyckoff GJ, Jaberi-Douraki M. Pharmacovigilance in patients with diabetes: A data-driven analysis identifying specific RAS antagonists with adverse pulmonary safety profiles that have implications for COVID-19 morbidity and mortality. Journal of the American Pharmacists Association. 2020.

11. Stafford EG, Riviere J, Xu X, Gedara NIM, Kawakami J, Wyckoff GJ, et al. Pulmonary Adverse Event Data in Hypertension with Implications on COVID-19 Morbidity. 2020.

12. Rønning M. Coding and classification in drug statistics–From national to global application. Norsk epidemiologi. 2001;11(1).

13. Tatonetti NP, Patrick PY, Daneshjou R, Altman RB. Data-driven prediction of drug effects and interactions. Science translational medicine. 2012;4(125):125ra31-ra31.

14. Evans SJ, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. Pharmacoepidemiology and drug safety. 2001;10(6):483-6.

15. van Puijenbroek EP, Bate A, Leufkens HG, Lindquist M, Orre R, Egberts AC. A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. Pharmacoepidemiology and drug safety. 2002;11(1):3-10.

16. Zorych I, Madigan D, Ryan P, Bate A. Disproportionality methods for pharmacovigilance in longitudinal observational databases. Statistical methods in medical research. 2013;22(1):39-56.

17. DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. The American Statistician. 1999;53(3):177-90.

18. Bate A, Evans S. Quantitative signal detection using spontaneous ADR reporting. Pharmacoepidemiology and drug safety. 2009;18(6):427-36.

19. Duggirala H, Tonning J, Smith E, Bright R, Baker J, Ball R. Data mining at FDA—white paper. 2019.

20. Xu X, Mazloom R, Goligerdian A, Staley J, Amini M, Wyckoff GJ, et al. Making sense of Pharmacovigilance and drug adverse event reporting: comparative similarity association analysis using AI machine learning algorithms in dogs and cats. Topics in Companion Animal Medicine. 2019;37:100366.

21. FDA Adverse Event Reporting System [Available from: https://open.fda.gov/data/faers/.

22. Mozzicato P. MedDRA. Pharmaceutical Medicine. 2009;23(2):65-75.

23. PostgreSQL B. PostgreSQL. Web resource: http://www PostgreSQL org/about. 1996.

24. FDA. Code of Federal Regulations: 21CFR310.305 2020 [Available from: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=310.305.

25. FDA. Code of Federal Regulations: 21CFR314.80 2020 [Available from: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm?fr=314.80.

26. Canida T, Ihrie J. openEBGM: An R Implementation of the Gamma-Poisson Shrinker Data Mining Model. R J. 2017;9(2):499.

27. Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological). 1996;58(1):267-88.

28. Zhao T, Liu H, Roeder K, Lafferty J, Wasserman L. The huge package for high-dimensional undirected graph estimation in R. The Journal of Machine Learning Research. 2012;13(1):1059-62.

29. Gibbs NE, Poole J, William G, Stockmeyer PK. An algorithm for reducing the bandwidth and profile of a sparse matrix. SIAM Journal on Numerical Analysis. 1976;13(2):236-50.

30. Bostock MP, Ellis, Russell K, Tarr G. Package 'edgebundleR": Circle Plot with Bundled Edges August 29, 2016 [Available from: https://cran.r-project.org/web/packages/edgebundleR/index.html.

31. Eisinga R, Heskes T, Pelzer B, Te Grotenhuis M. Exact p-values for pairwise comparison of Friedman rank sums, with application to comparing classifiers. BMC bioinformatics. 2017;18(1):68.

32. Lexicomp. Specific Lexicomp Online Database [database on the Internet]. Hudson (OH): Lexicomp Inc. 2016 [Available from: http://online.lexi.com.

689 33. Lepor H, Kaplan SA, Klimberg I, Mobley DF, Fawzy A, Gaffney M, et al. Doxazosin for
690 benign prostatic hyperplasia: long-term efficacy and safety in hypertensive and normotensive
691 patients. The Journal of urology. 1997;157(2):525-30.

692 34. Meredith PA, Reid JL. Efficacy and Tolerability of Long-Term Rilmenidine Treatment in
693 Hypertensive Diabetic Patients. American Journal of Cardiovascular Drugs. 2004;4(3):195-200.

694 35. Hasebe N, Kikuchi K, Group NCS. Controlled-release nifedipine and candesartan low-
695 dose combination therapy in patients with essential hypertension: the NICE Combi (Nifedipine
696 and Candesartan Combination) Study. Journal of hypertension. 2005;23(2):445-53.

697 36. Kjeldsen SE, Sica D, Haller H, Cha G, Gil-Extremera B, Harvey P, et al. Nifedipine plus
698 candesartan combination increases blood pressure control regardless of race and improves the side
699 effect profile: DISTINCT randomized trial results. Journal of hypertension. 2014;32(12):2488.

700 37. Mancia G, Cha G, Gil-Extremera B, Harvey P, Lewin A, Villa G, et al. Blood pressure-
701 lowering effects of nifedipine/candesartan combinations in high-risk individuals: subgroup
702 analysis of the DISTINCT randomised trial. Journal of human hypertension. 2017;31(3):178-88.

703 38. Fujikawa K, Hasebe N, Kikuchi K. Cost-Effectiveness Analysis of Hypertension
704 Treatment: Controlled Release Nifedipine and Candesartan Low-Dose Combination Therapy in
705 Patients with Essential Hypertension—The Nifedipine and Candesartan Combination (NICE-
706 Combi) Study—. Hypertension research. 2005;28(7):585-91.

707 39. Patterson SD, Jones B. Bioequivalence and statistics in clinical pharmacology: CRC Press;
708 2017.

709 40. Midha KK, McKay G. Bioequivalence; its history, practice, and future. The AAPS journal.
710 2009;11(4):664-70.

711 41. Chen RT, Rastogi SC, Mullen JR, Hayes SW, Cochi SL, Donlon JA, et al. The vaccine
712 adverse event reporting system (VAERS). Vaccine. 1994;12(6):542-50.

713 42. Shimabukuro TT, Nguyen M, Martin D, DeStefano F. Safety monitoring in the vaccine
714 adverse event reporting system (VAERS). Vaccine. 2015;33(36):4398-405.

715