

A Universal Prioritized Set of Pharmacovigilance Outcomes Derived From Nine Complementary Sources Using Large Language Models

Frank J. DeFalco^{1,2,3}, Mitchell M. Conover^{1,2,3}

¹ Coordinating Center, Observational Health Data Science and Informatics, New York City, NY 10032, United States, ² Observational Health Data Analytics, Johnson & Johnson, Titusville, NJ 08560, United States, ³ The two authors contributed equally

Background

Global pharmacovigilance (PV) efforts have produced multiple lists of outcomes for safety surveillance, but these resources were developed for different purposes and use inconsistent terminology, vocabularies, and levels of granularity. This heterogeneity limits their direct use as a unified foundation for prioritizing outcomes for PV and phenotype development. Large language models (LLMs) offer an efficient way to harmonize these disparate sources by resolving synonymy, subsumption, and mapping challenges across disease vocabularies, taxonomies, and ontologies. We used LLMs to integrate nine prioritized PV outcome lists, each derived from orthogonal and complementary approaches, to generate a universal set of outcomes likely to be highly relevant for pharmacovigilance and phenotype development.

Methods

We used nine different available resources and analytic approaches to produce lists of candidate PV outcomes (**Table 1**).

Table 1. Nine different sources of candidate outcomes for pharmacovigilance

#	Label	Description
1	FDA Labeled Events	Adverse events which appear on ≥ 20 FDA drug labels ¹
2	Hospitalized Events	Adverse events appearing in U.S. Food & Drug Administration (FDA) drug labels where $> 15\%$ condition records are in the inpatient setting when cross referenced with Optum's de-identified Clinformatics® Extended Data Mart Database, Date of Death (DOD) ¹
3	OCMQs	FDA's Office of New Drugs Custom Medical Queries for Safety Analyses of Clinical Trial Data ²
4	Common Events	Diagnoses that occur > 1 million times in the Optum Clinformatics® database in the time window 1 to 30 days after patients initiate any drug for the first time
5	Tisdale	Chapter/section headings from Tisdale & Miller's "Drug Induced Diseases" 3 rd ed. ³
6	FAERS Events	Outcomes flagged for ≥ 100 distinct drugs in the FDA Adverse Event Reporting System (FAERS) Database
7	MEDLINE	Adverse events appearing in > 150 case reports documented in MEDLINE

8	FAERS PRR	Adverse events reported for > 50 drug ingredients using the OHDSI CommonEvidenceModel ⁴
9	Sentinel	Outcomes evaluated by Sentinel Safety System between 2022-2024 ⁵

We applied LLM-based techniques for categorization, synonym resolution, and subsumption analysis, learning across the nine lists, to produce a set of outcomes which prioritizes those consistently identified by the various approaches (GPT-5.0). We used LLM to categorize outcomes into categories classified by therapeutic area, then within categories, we used LLM for subsumption to produce a modified list of normalized outcomes. After normalizing the outcome set, we performed a set-intersection analysis of distinct adverse event concepts across the nine lists. We counted events per unique intersection pattern and devised a rule to select high priority outcomes based on their inclusion in the different lists.

Results

There were 1,952 outcomes which mapped to at least one list. The number of outcomes included in each list were as follows: FDA Labeled Events (85), Hospitalized Events (1,359), OCMQs (86), Common Events (378), Tisdale (109), FAERS Events (231), MEDLINE (186), FAERS PRR (317), Sentinel (55). Most mapped events were source-specific (1,494, 76.5%), while 458 (23.5%) appeared in 2 or more sources and 206 (10.6%) appeared in 3 or more sources. The large number of outcomes that mapped to only one list is due to one of the lists (Hospitalized Events) including a substantially larger number of candidate outcomes than any of the other lists. Below, **Figure 1** presents the distribution of the cumulative number of outcomes that appear across increasing number of lists, stratified by whether they were included in the OCMQs list. The percentage shown in green reflects the proportion of the OCMQs list that meets each threshold.

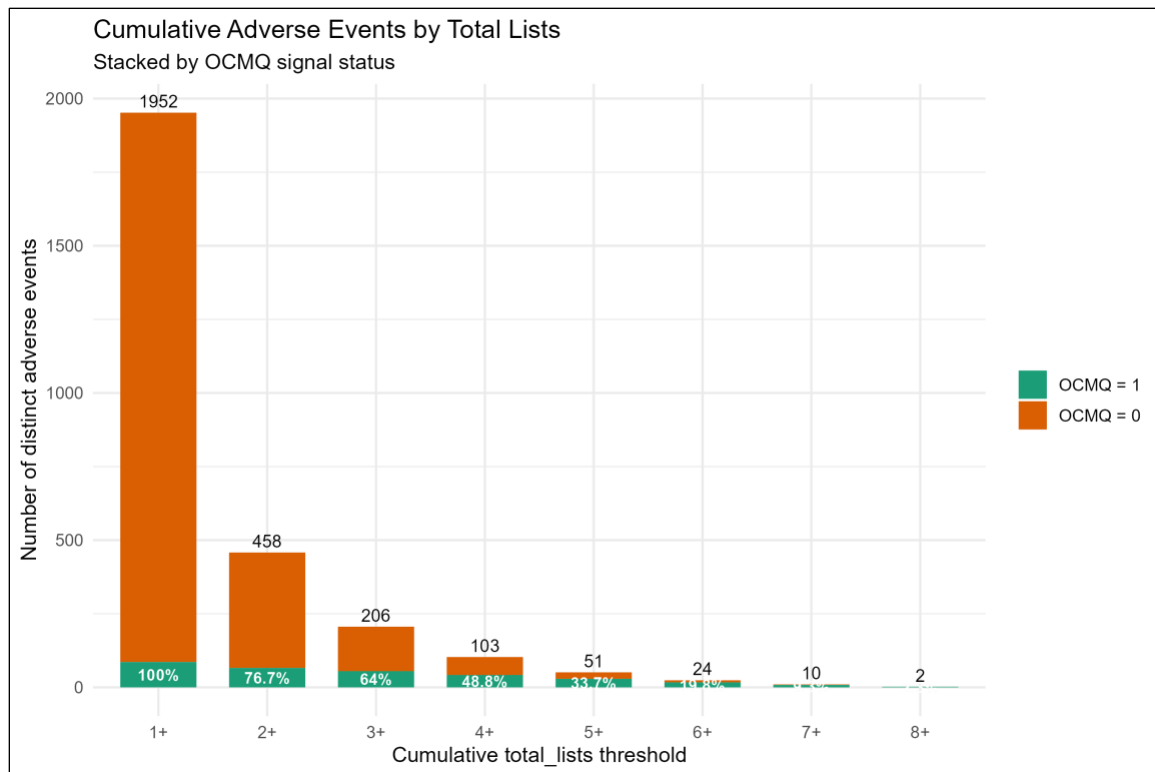


Figure 1. Cumulative number of adverse events appearing across multiple lists, stratified by inclusion in the OCMQs list.

The final list of outcomes prioritized for inclusion were outcomes that appeared on ≥ 3 lists, all outcomes included in the OCMQ list, and outcomes included in the Tisdale list. Outcomes selected for inclusion in the final list are highlighted in orange in the UpSet plot below (**Figure 2**), which restricts to intersections with at least two outcomes. As shown in the plot, a substantial number of outcomes that were included in the OCMQs list and the Tisdale list were not selected by the rule requiring that an outcome be observed across three or more lists. As a result, we selected all outcomes included in the OCMQ or Tisdale lists, regardless of whether they were observed in other lists, given that they already reflect a strong assertion that included outcomes are highly relevant for pharmacovigilance.

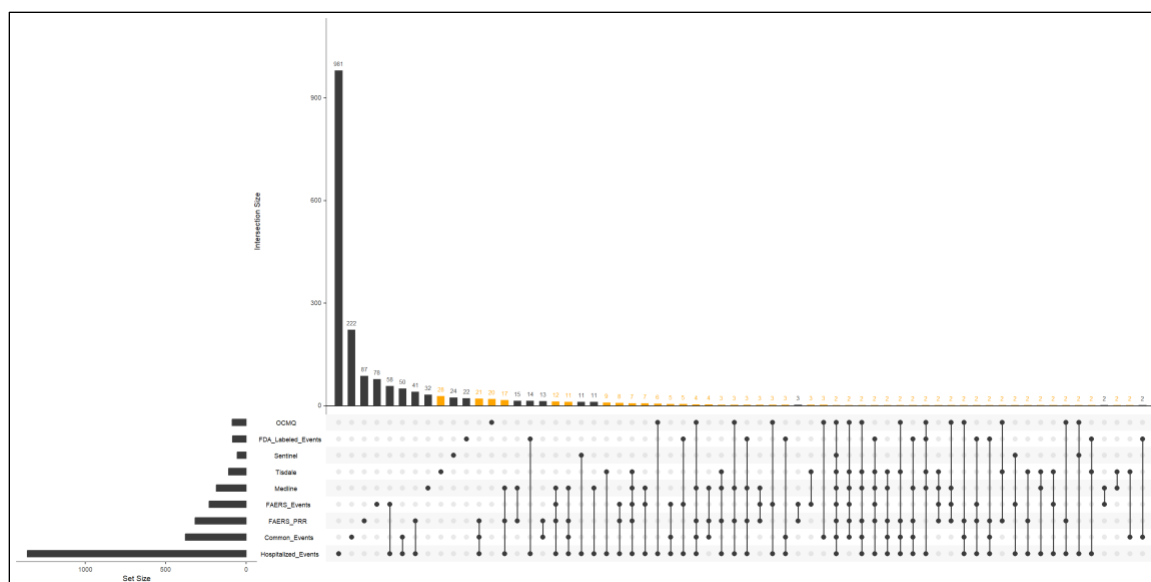


Figure 2. UpSet plot showing the number of events per unique intersection pattern and highlighting intersections that were selected for inclusion in the final list.

The final set of outcomes selected are presented in **Table 2**.

Conclusions

We identified 282 outcomes that should be included in the pharmacovigilance systems. This list draws from insights including the outcome’s frequency, prior causal assessments from RCTs and observational data, and history of being attributed to drug exposures. We recommend that work to develop computable phenotypes in support of drug safety should consider these outcomes to be of high priority.

References

1. Tanaka Y, Chen HY, Belloni P, et al. OnSIDES database: Extracting adverse drug events from drug labels using natural language processing models. *Med.* 2025;6(7). doi:10.1016/j.medj.2025.100642
2. Proestel S, Popat V, Unger EF, Jeng LJB. The Development and Use of Office of New Drugs Custom Medical Queries for Safety Analyses of Clinical Trial Data. *Drug Saf.* 2025;48(12):1331-1337. doi:10.1007/s40264-025-01582-1
3. Tisdale JE, Miller DA. Drug-Induced Diseases: Prevention, Detection, and Management. ASHP; 2018.

4. Voss, Boyce, Ryan, van der Lei, Rijnbeek, Schuemie. Accuracy of an automated knowledge base for identifying drug adverse reactions. *Journal of Biomedical Informatics*. 2017;66:72-81. doi:10.1016/j.jbi.2016.12.005
5. PDUFA VII Commitment: An Assessment of the Sentinel System (2022 to 2024). U.S. Food & Drug Administration. <https://www.fda.gov/industry/prescription-drug-user-fee-amendments/pdufa-vii-commitment-assessment-sentinel-system-2022-2024>. Published September 30, 2025. Accessed February 3, 2026.

Table 2. Selected high-priority pharmacovigilance outcomes

1. Thrombocytopenia	78. DVT	155. Encephalitis	224. Hyperthyroidism
2. heart failure	79. Death	156. Encephalopathy	225. Hyperuricemia
3. Arrhythmia	80. Dementia	157. Endometrial cancer	226. Ischemic stroke
4. Fever	81. Depressive Disorder	158. Eosinophilia	227. Leg edema
5. Neutropenia	82. Diabetes mellitus	159. Epilepsy	228. Malignant hyperthermia
6. Seizure	83. Dysphagia	160. Esophageal disorder	229. Metabolic alkalosis
7. anaphylaxis	84. Dystonia	161. Esophagitis	230. Myalgia
8. myocardial infarction	85. Edema	162. Glomerulonephritis	231. Photosensitivity Disorders
9. Angioedema	86. Erythema Multiforme	163. Gynecomastia	232. Renal & urinary tract infection
10. Atrial Fibrillation	87. Gastrointestinal hemorrhage	164. Hematoma	233. Respiratory acidosis
11. Bradycardia	88. Hallucination	165. Hemolytic uremic syndrome	234. Serum sickness or serum sickness-like reaction
12. Diarrhea	89. Heart Diseases	166. Hepatic Encephalopathy	235. Sexual dysfunction
13. Hyperglycemia	90. Hemorrhage	167. Hyperbilirubinemia	236. Skin cancer
14. Interstitial lung disease	91. Hepatitis	168. Hyperlipidemia	237. Viral infection
15. Pancreatitis	92. Hepatotoxicity	169. Hyperpigmentation of skin	238. Visual disturbances
16. Stevens-Johnson syndrome	93. Hypercalcemia	170. Hypersensitivity	239. Volume depletion
17. Syncope	94. Hyperprolactinemia	171. Hypoxemia	240. Weight increased
18. Tachycardia	95. Hypokalemia	172. Intestinal perforation	241. Abnormal uterine bleeding
19. thrombosis	96. Hypothermia	173. Intracerebral hemorrhage	242. Acute and constrictive pericarditis
20. Abdominal Pain	97. Jaundice	174. Ischemia	243. Acute interstitial nephritis
21. Acute kidney injury	98. Lactic acidosis	175. Kidney stone	244. Allergic-mediated blood disorders
22. Agranulocytosis	99. Lung disease	176. Leukocytosis	245. Amenorrhea
23. Anxiety	100. Lymphadenopathy	177. Liver disorder	246. Androgenic Alopecia
24. Constipation	101. Malignancy	178. Liver injury	247. Atrial tachycardia
25. Delirium	102. Myocarditis	179. Long QT syndrome	248. Atrioventricular nodal reciprocating tachycardia
26. Depression	103. Myoclonus	180. Low blood pressure	249. Bacterial vaginosis
27. Dyspnea	104. Nephrotic Syndrome	181. Lower gastrointestinal bleeding	250. Cardiac Conduction Disturbance
28. Erythema	105. Paralysis	182. Lymphoma	251. Decreased menstrual bleeding
29. Hyperkalemia	106. Paresthesia	183. Mania	252. Drug fever
30. Hypoglycemia	107. Pneumonitis	184. Mental Disorders	253. Dry mouth
31. Hyponatremia	108. Pruritus	185. Metabolic acidosis	254. Dyspepsia
32. Hypotension	109. Pulmonary edema	186. Methemoglobinemia	255. Ejaculation disorder
33. Hypothyroidism	110. Pulmonary hypertension	187. Multiple Organ Failure	256. Fatigue
34. Liver failure	111. Rash	188. Myocardial ischemia	257. Glaucoma
35. Peripheral neuropathy	112. Respiratory depression	189. Neuroleptic malignant syndrome	258. Growth hormone deficiency
36. Pneumonia	113. Respiratory failure	190. Neurotoxicity	259. Hepatic and cholestatic disease
37. Pulmonary embolism	114. Shock	191. Pain	260. Hirsutism
38. Urticaria	115. Skin ulcer	192. Palpitations	261. Hyperphosphatemia
39. Ventricular tachycardia	116. Status Epilepticus	193. Pemphigus	262. Hypertrichosis
40. ARDS	117. Toxic epidermal necrolysis	194. Pneumonia bacterial	263. Infertility
41. Acidosis	118. Ulcer	195. Posterior reversible encephalopathy syndrome	264. Investigational agent abuse potential
42. Anagen Effluvium Alopecia	119. Urinary retention	196. Postoperative Complications	265. Irritability
43. Asthma	120. Urinary tract infection	197. Proteinuria	266. Lipid disorder
44. Cardiomyopathy	121. Venous Thrombosis	198. Pseudomembranous colitis	267. Local administration reactions
45. Confusion	122. Ventricular Fibrillation	199. Psoriasis	
46. Diabetic Ketoacidosis	123. Vision Disorders	200. Purpura	
47. Disseminated intravascular coagulation	124. drug induced liver injury	201. Renal disease	
48. Dizziness and giddiness	125. Acute coronary syndrome	202. Respiratory Insufficiency	
49. Erectile dysfunction	126. Acute myeloid leukemia	203. Respiratory system disorder	
50. Hypertension	127. Acute tubular necrosis		
51. Hypocalcemia	128. Adrenal Insufficiency		
52. Leukopenia	129. Akathisia		
53. Orthostatic hypotension	130. Amnesia		
54. Pancytopenia	131. Angina pectoris		
55. Pleural effusion	132. Anxiety disorder		
	133. Arthritis		

56. Psychosis	134. Ataxia	204. Skin Diseases	268. Nasopharyngitis
57. Renal impairment	135. Autoimmune Diseases	205. Steatosis of liver	269. Osteomalacia
58. SIADH	136. Bacteremia	206. Supraventricular tachycardia	270. Ototoxicity
59. Secondary parkinsonism	137. Brain Edema	207. Thromboembolism	271. Ovarian hyperstimulation
60. Serotonin syndrome	138. Breast cancer	208. Thrombotic microangiopathy	272. Pericardial tamponade
61. Systemic lupus erythematosus	139. Bundle branch block	209. Tongue disorder	273. Purulent material
62. Torsade de pointes	140. Cardiogenic shock	210. Type 1 diabetes mellitus	274. Respiratory alkalosis
63. Tremor	141. Chest pain	211. fractures	275. Stroke-TIA
64. Vasculitis	142. Cholestasis	212. Acne	276. Tardive dyskinesia
65. Apnoea	143. Chorea	213. Anorexia	277. Telogen Effluvium Alopecia
66. Arthralgia	144. Chronic kidney disease	214. Bladder cancer	278. Tendinopathy
67. Ascites	145. Coronary arteriosclerosis	215. Cachexia	279. Upper gastrointestinal ulceration
68. Atrioventricular block	146. Coronary vasospasm	216. Cholecystitis	280. Valvular heart disease
69. Bipolar disorder	147. Cough	217. DRESS syndrome	281. Warfarin-Induced Skin Necrosis
70. Bleeding	148. Cushing Syndrome	218. Diabetes insipidus	282. Water and sodium imbalance
71. Bronchospasm	149. Cyanosis	219. Dysgeusia	
72. Cardio-respiratory arrest	150. Cystitis	220. Excessive menstrual bleeding	
73. Cellulitis	151. Cytokine release syndrome	221. Fall	
74. Cerebral hemorrhage	152. Delusion	222. Fixed drug eruption	
75. Cerebral infarction	153. Dermatitis exfoliative generalized	223. Hypermagnesemia	
76. Cognitive disorder			
77. Coma	154. Drug hypersensitivity		