PATTER, Volume *1*

Supplemental Information

Using Machine Learning to Identify Adverse Drug Effects Posing Increased Risk to Women

Payal Chandak and Nicholas P. Tatonetti

Study	Drug	ADR	Expected Sex	Predicted Sex	logROR	95% CI
Tamargo et al. 2017	Paracetamol	Acute hepatic failure	F	No risk	-	-
Yu et al. 2016	Ibuprofen	Cholecystitis chronic	F	F	3.64	$(3.54, 3.75)$
Ofotokun et al. 2003	Ritonavir	Diarrhoea	M	M	0.27	$(0.29, 0.25)$
Makkar et al. 1993, Drici et al. 2001	Amiodarone	Electrocardiogram QT prolonged	F	F	0.68	$(0.60, 0.75)$
Makkar et al. 1993, Drici et al. 2001	Disopyramide	Electrocardiogram QT prolonged	F	Insufficient data	-	-
Seeman et al. 2020	Thioridazine	Electrocardiogram QT prolonged	F	Insufficient data	-	-
Makkar et al. 1993, Drici et al. 2001	Sotalol	Electrocardiogram QT prolonged	F	No risk	-	-
Parekh et al. 2011	Rosiglitazone	Fractures	F	F	0.80	$(0.76, 0.83)$
Tamargo et al. 2017	Heparin	Haemorrhage	F	No risk	-	-
Tharpe et al. 2011	Citalopram	Hyponatraemia	F	F	0.21	$(0.19, 0.23)$
Seeman et al. 2020	Clozapine	Metabolic syndrome	F	No risk	-	-
Ofotokun et al. 2003	Ritonavir	Nausea	F	F	0.29	$(0.28, 0.31)$
Schmetzer et al. 2012	Ifosfamide	Neurotoxicity	F	No risk	-	-
Schmetzer et al. 2012	Fluorouracil	Neutropenia	F	F	0.17	$(0.16, 0.19)$
Whitley et al. 2009, Tharpe et al. 2011	Aspirin	Platelet aggregation inhibition	M	No risk	-	-
Tamargo et al. 2017	Diazepam	Psychomotor skills impaired	F	No risk	-	-
Tharpe et al. 2011	Fentanyl	Respiratory depression	F	No risk	-	-
Tharpe et al. 2011	Oxycodone	Respiratory depression	F	No risk	-	-
Tamargo et al. 2017	Procainamide	Systemic lupus erythematosus	F	Insufficient data	-	-
Tamargo et al. 2017	Aspirin	Ulcers	M	No risk	-	-
Franconi et al. 2007, Whitley et al. 2009	Metoprolol	Vascular hypertensive disorders	F	F	0.33	$(0.32, 0.34)$
Franconi et al. 2007	Amlodipine	Vascular hypertensive disorders	F	F	0.17	$(0.17, 0.18)$

Table S4: Hypotheses and results for clinical validation.

Abbreviations: M - Male; F - Female; ADR - Adverse drug reaction; logROR - log $_e$ Reporting Odds Ratio; CI - Confidence Interval

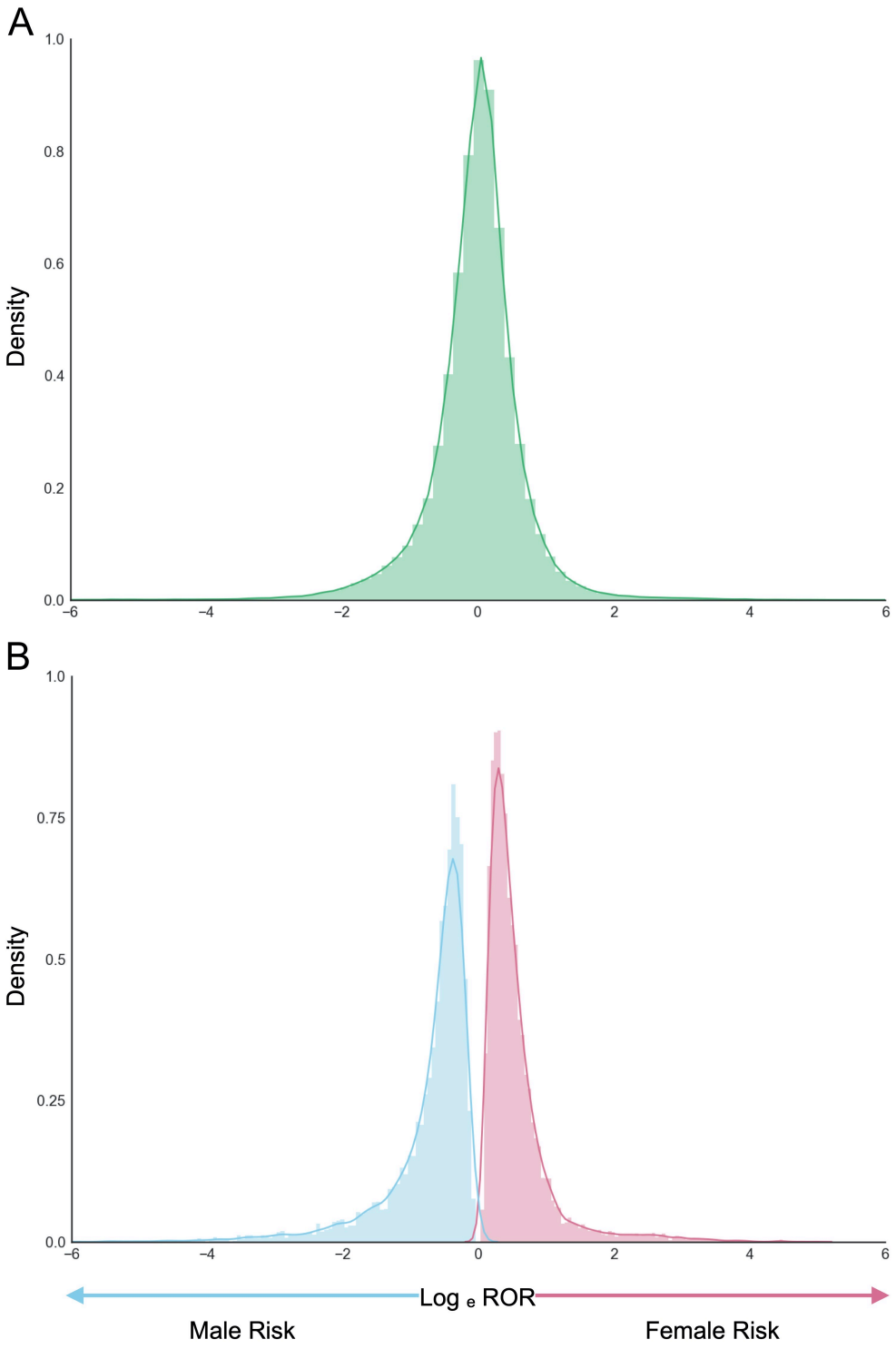

Figure S1: Distribution of sex risks. Both histograms visualize the magnitude of sex risks (log $_e$ ROR; x-axis) against normalized counts (density; y-axis). Panel A shows that the distribution of all sex risks follows a normal distribution. Panel B depicts the distribution of significant sex risks (adjusted P ≤ 0.05) grouped by sex. Pink indicates female risk and blue indicates male risk. Both distributions follow a lognormal distribution.

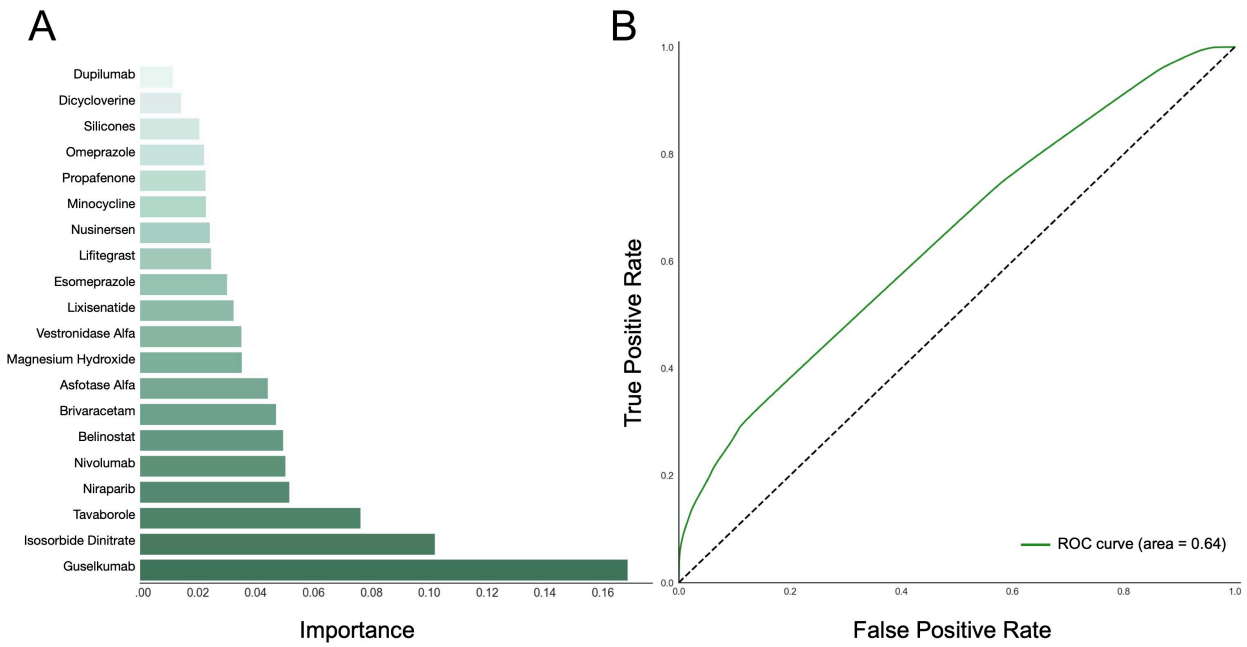

Figure S2: Characterizing the random forest model. Panel A shows features that had the highest importances. Panel B shows a receiver operating characteristic curve, where classification thresholds reflect different cut-offs in the estimated propensity score.

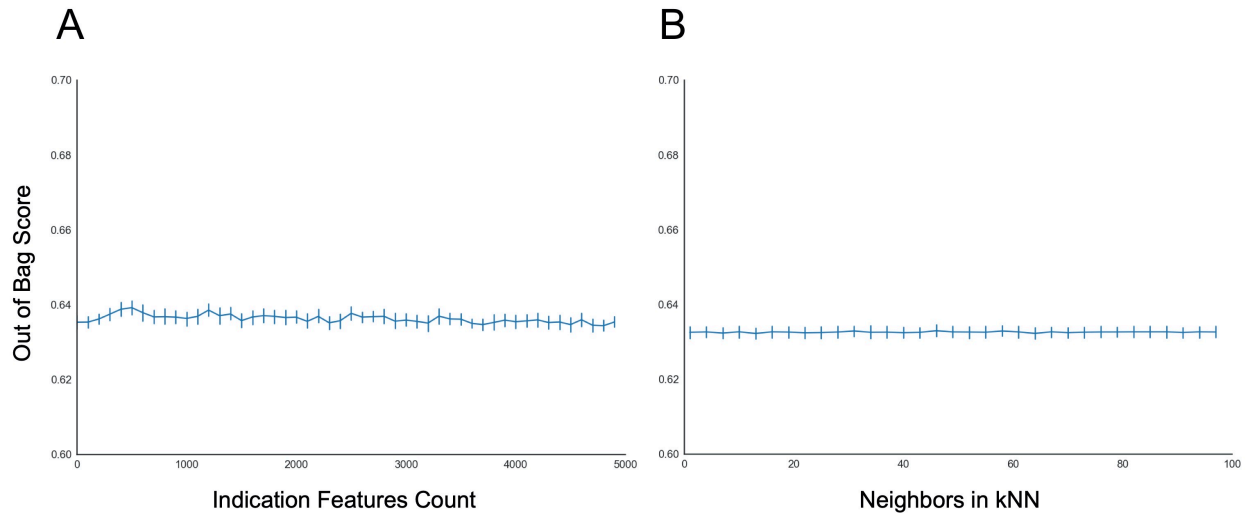

Figure S3: Random forest model is sufficiently robust. Making the random forest model more complex does not change performance. In both panels, each point indicates performance for the given experiment. Performance is represented as mean out of bag score $\pm$ 95% confidence intervals. Panel A shows the effect of adding upto 5000 indication features on performance. In each consecutive experiement, we add 100 indication features with the highest report counts. In Panel B, we test the outcome of using k-Nearest Neighbors to impute age for missing reports. Each successive experiment increases the number of neighbors. As shown by the flat lines in both panels, the model's performance remains largely unchanged by these modifications.

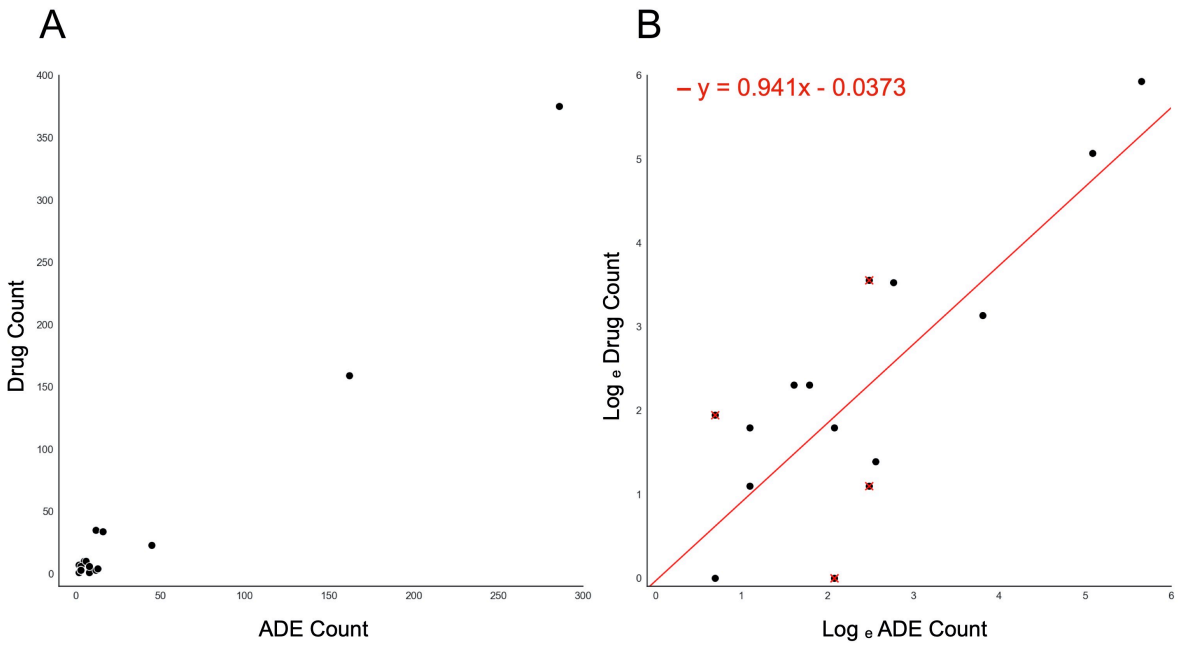

Figure S4: Prediction of pharmacogenes from sex risks. We leveraged sex risks identified by AwareDX to flag genes that could have variants with important, and possibly undiscovered, pharmacokinetic and pharmacodynamic effects. Both scatter plots visualize the count of drugs (y-axis) against the count of significant drug-event pairs identified by AwareDX (x-axis). Panel A shows a linear scale and Panel B shows a logarithmic scale. In B, we applied linear regression and used the residuals as a ranking mechanism to identify top sex-varying pharmcogene candidates (marked in red).

Supplemental References

Tamargo J., Rosano G., Walther T., Duarte J., Niessner A., Kaski J.C., et al. (2017). Gender differences in the effects of cardiovascular drugs. Eur. Heart J. Cardiovasc. Pharmacother. 3:163–182.

Yu Y., Chen J., Li D., Wang L., Wang W., Liu H. (2016). Systematic analysis of adverse event reports for sex differences in adverse drug events. Sci. Rep. 6. doi:10.1038/srep24955.

Ofotokun I., Pomeroy C. (2003). Sex differences in adverse reactions to antiretroviral drugs. Top. HIV Med.11:55–59.

Makkar R., Fromm B., Steinman R., Meissner M., Lehmann M. (1993). Female gender as a risk factor for torsades de pointes associated with cardiovascular drugs. JAMA 270:8.

Drici M.-D., Clement N. (2001). Is gender a risk factor for adverse drug reactions? The example of drug-induced long QT syndrome. Drug Saf. 24:575–585.

Seeman M.V. (2020). Men and women respond differently to antipsychotic drugs. Neuropharmacology 163:107.

Parekh A., Fadiran E.O., Uhl K., Throckmorton D.C. (2011). Adverse effects in women: Implications for drug development and regulatory policies. Expert Rev. Clin. Pharmacol. 4:453–466.

Tharpe N. (2011). Adverse drug reactions in women's health care. J. Midwifery Women's Health 56:205–213.

Schmetzer O., Flörcken A. (2012). Sex differences in the drug therapy for oncologic diseases. Handbook Exp. Pharmacol. 214:411–442.

Whitley H.P., Lindsey W. (2009). Sex-based differences in drug activity. Am. Fam. Phys. 80:1254–1258.

Franconi F., Brunelleschi S., Steardo L., Cuomo V. (2007). Gender differences in drug responses. Pharmacol. Res. 55:81– 95.