ASSOCIATION STUDY OF DRUGS AND ADVERSE EVENTS

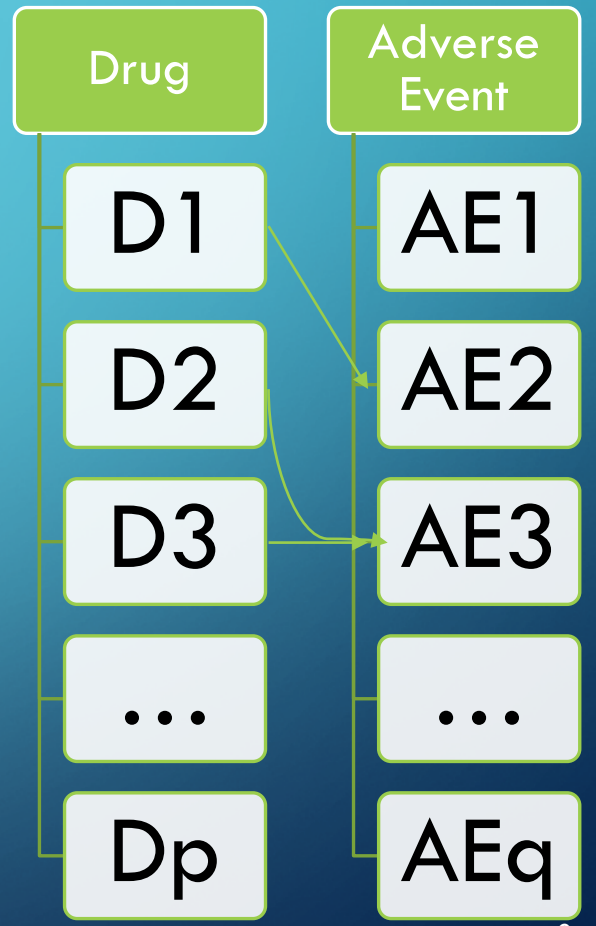
MINH PHAM, DEPARTMENT OF MATHEMATICS & STATISTICS, UNIVERSITY OF SOUTH FLORIDA

FENG CHENG, DEPT OF PHARMACEUTICAL SCIENCE & DEPT OF EPIDEMIOLOGY & BIOSTATISTICS, UNIVERSITY OF SOUTH FLORIDA

KANDETHODY RAMACHANDRAN, DEPARTMENT OF MATHEMATICS & STATISTICS, UNIVERSITY OF SOUTH FLORIDA

WHAT IS THE PROBLEM AND WHY

- Preliminary detection of drugs' side effects
  - For regulators
  - For the public
  - Automate signal detection from the FAERS database
    - Which statistical method can drive this process?
- Challenges:
  - Too many drugs and AEs
  - Sparseness of the data



WHAT IS THE PROBLEM AND WHY

primaryid	D1	D2	D3	…	Dp	AE1	AE2	AE3	…	Aeq
88438052	0	0	1	…	0	0	0	0	…	0
53841413	0	0	0	…	0	0	0	1	…	0
77175049	0	0	1	…	0	1	0	0	…	0
62598515	0	0	0	…	0	0	0	0	…	0
37699682	0	0	0	…	0	0	0	0	…	0
65866064	1	0	0	…	0	0	1	0	…	0
18502693	0	0	0	…	0	0	0	0	…	0
59014545	0	0	0	…	0	0	0	0	…	0

OUTLINE

- Methods in the literature
- A gold standard for comparison study
- Results of comparison study
- Future consideration

METHODS IN THE LITERATURE

- Disproportionality (Univariate) methods:

	Drug X_i	Other drugs
Effect Y_j	a	b
Other effects	c	d

- Proportional Reporting Ratio $PRR = \frac{a/(a+c)}{b/(b+d)}$
- Reporting Odds Ratio $ROR = \frac{a/c}{b/d}$

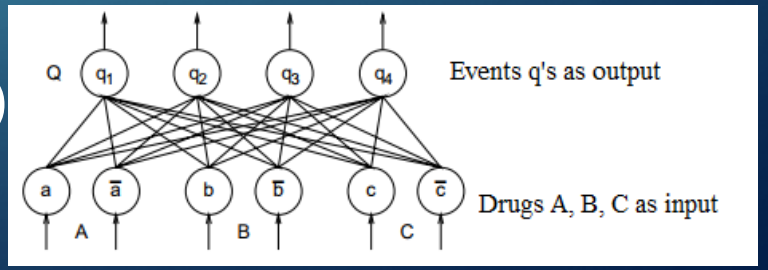
METHODS IN THE LITERATURE

- Bayesian Confidence Propagation Neural Network (BCPNN)

	Drug X_i	Other drugs
Effect Y_j	n_{11}	n_{10}
Other effects	n_{01}	n_{00}

- $P(n_{11}, n_{10}, n_{01}, n_{00} | n, p_{11}, p_{10}, p_{01}, p_{00}) = \frac{n!}{n_{11}! n_{10}! n_{01}! n_{00}!} p_{11}^{n_{11}} p_{10}^{n_{10}} p_{01}^{n_{01}} p_{00}^{n_{00}}$
- $p_{11}, p_{10}, p_{01}, p_{00} \sim Dir(\alpha_{11}, \alpha_{10}, \alpha_{01}, \alpha_{00})$

$$E(weight) = \log_2\left(\frac{E(p_{11})}{E(p_{1.})E(p_{.1})}\right)$$



METHODS IN THE LITERATURE

- Gamma-Poisson Shrinkage model (GPS)
  - $N_{ij} = \sum_{l=1}^{n} X_{il} Y_{jl}$: count of drug *i* and event *j*
  - $N_{ij} \sim Poisson(\lambda_{ij} * E_{ij})$
  - $\lambda \sim mixture\ of\ 2\ Gamma\ distribution$
  - Use posterior distribution of λ to raise signal

	Drug X_i	Other drugs
Effect Y_j	a	b
Other effects	c	d

METHODS IN THE LITERATURE

- Disproportionality analysis:
  - Frequentist: PRR and ROR
  - Bayesian: BCPNN and GPS
- Going multivariate:
  - Reason: multi-pharmarcy

primaryid	D1	D2	D3	…	Dp	AE1	AE2	AE3	…	Aeq
88438052	1	0	1	…	0	1	0	0	…	0

METHODS IN THE LITERATURE

- Logistic Regression (LR):

$$logit\left(P\left(Y_j = 1\right)\right) = \log\left(\frac{P\left(Y_j = 1\right)}{1 - P\left(Y_j = 1\right)}\right) = \sum \beta_i X_i$$

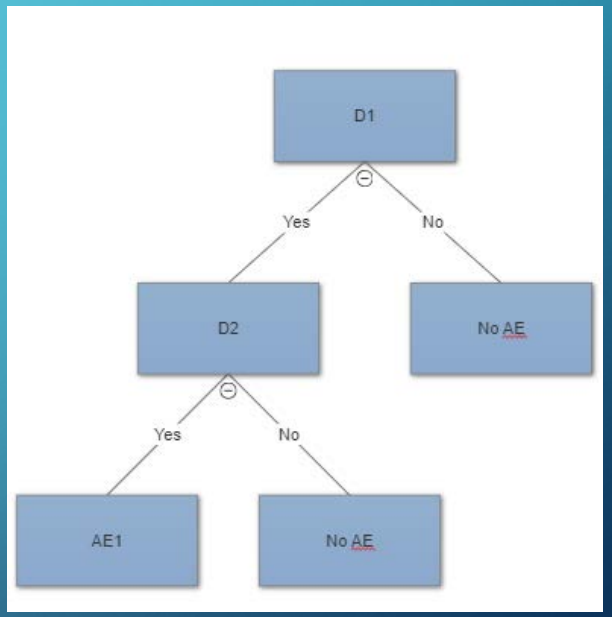
- Regression-adjusted Gamma-Poisson Shrinkage Model (RGPS):

$$E_{ij} = \frac{e^{\sum \beta_l X_l - \sum \beta_i X_i}}{1 + e^{\sum \beta_l X_l - \sum \beta_i X_i}}$$

Then regular GPS

METHODS IN (OTHER) LITERATURE

- Need improvement: drug interaction
- Borrowed ideas from Genome-wide Association Study:
  - Random Forests: rank drugs by variable important



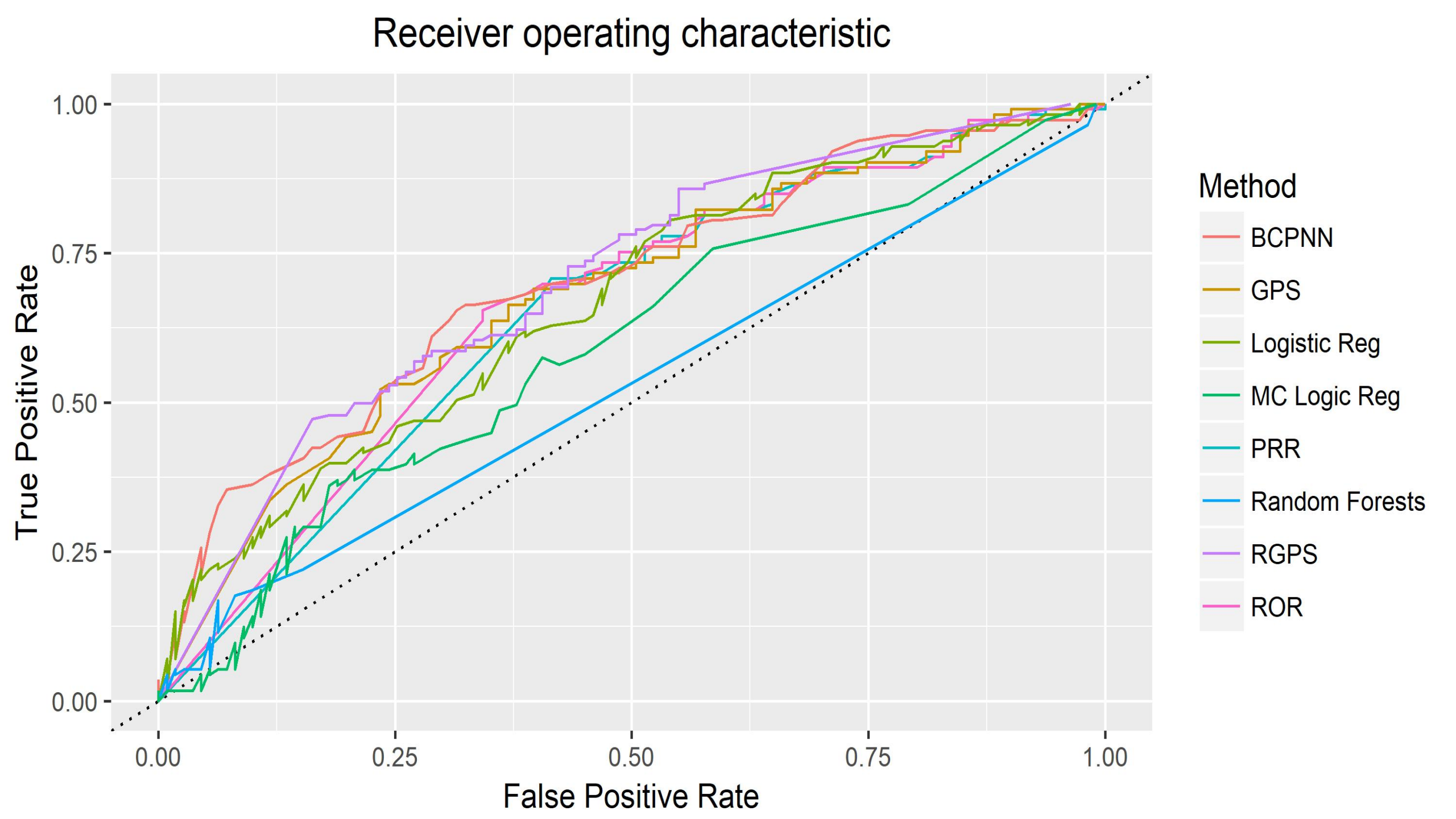
  - Monte Carlo Logic Regression:

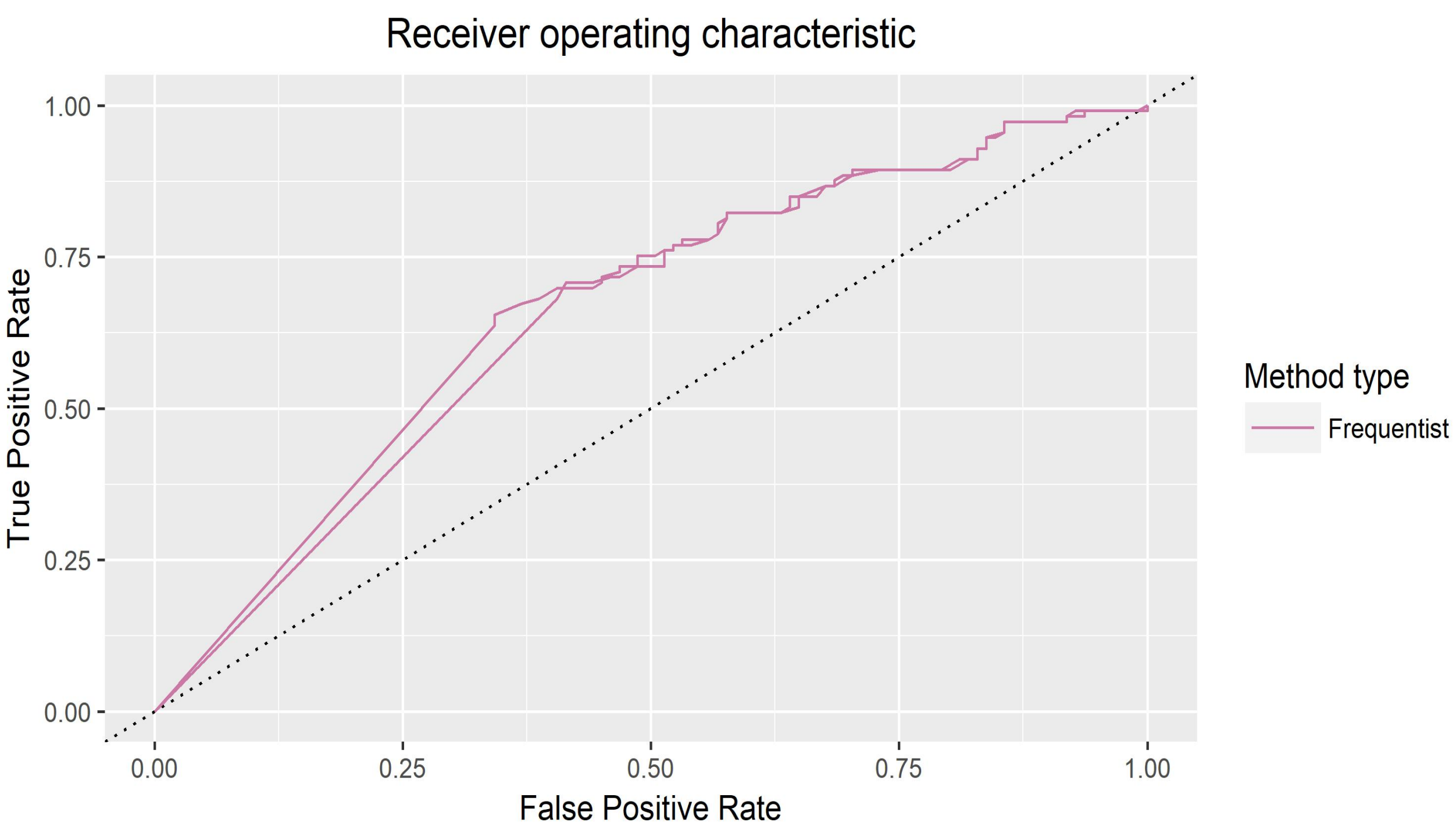
$$logit\left(P(Y_j = 1)\right) = \beta_0 + \sum_{i=1}^{K} \beta_i L_i$$

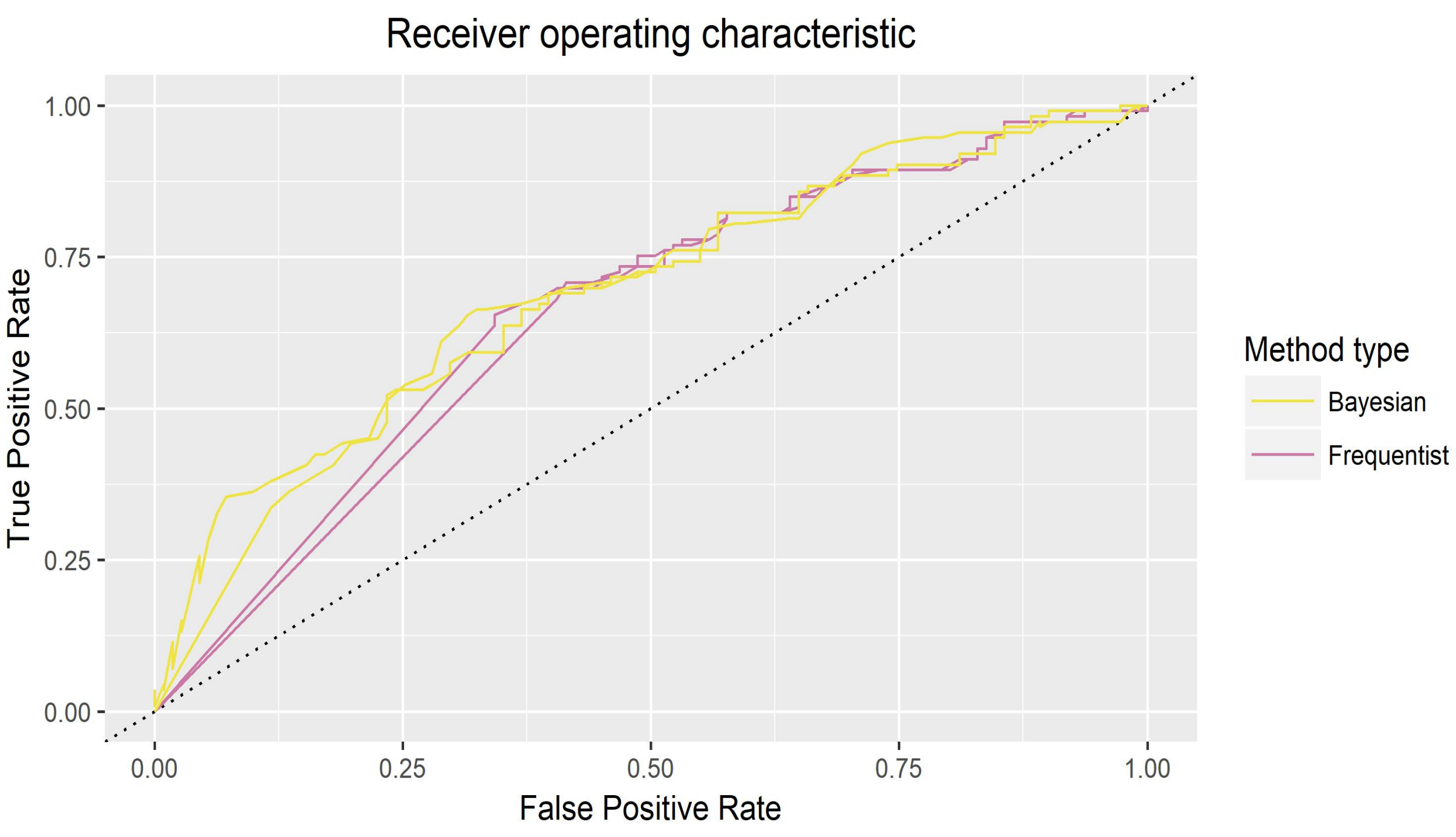
$L_i = I(X_1 = 1\ and\ X_2 = 1)$ or $L_i = I((X_1 \neq 1\ and\ X_2 = 1)\ or\ X_3 \neq 1)$

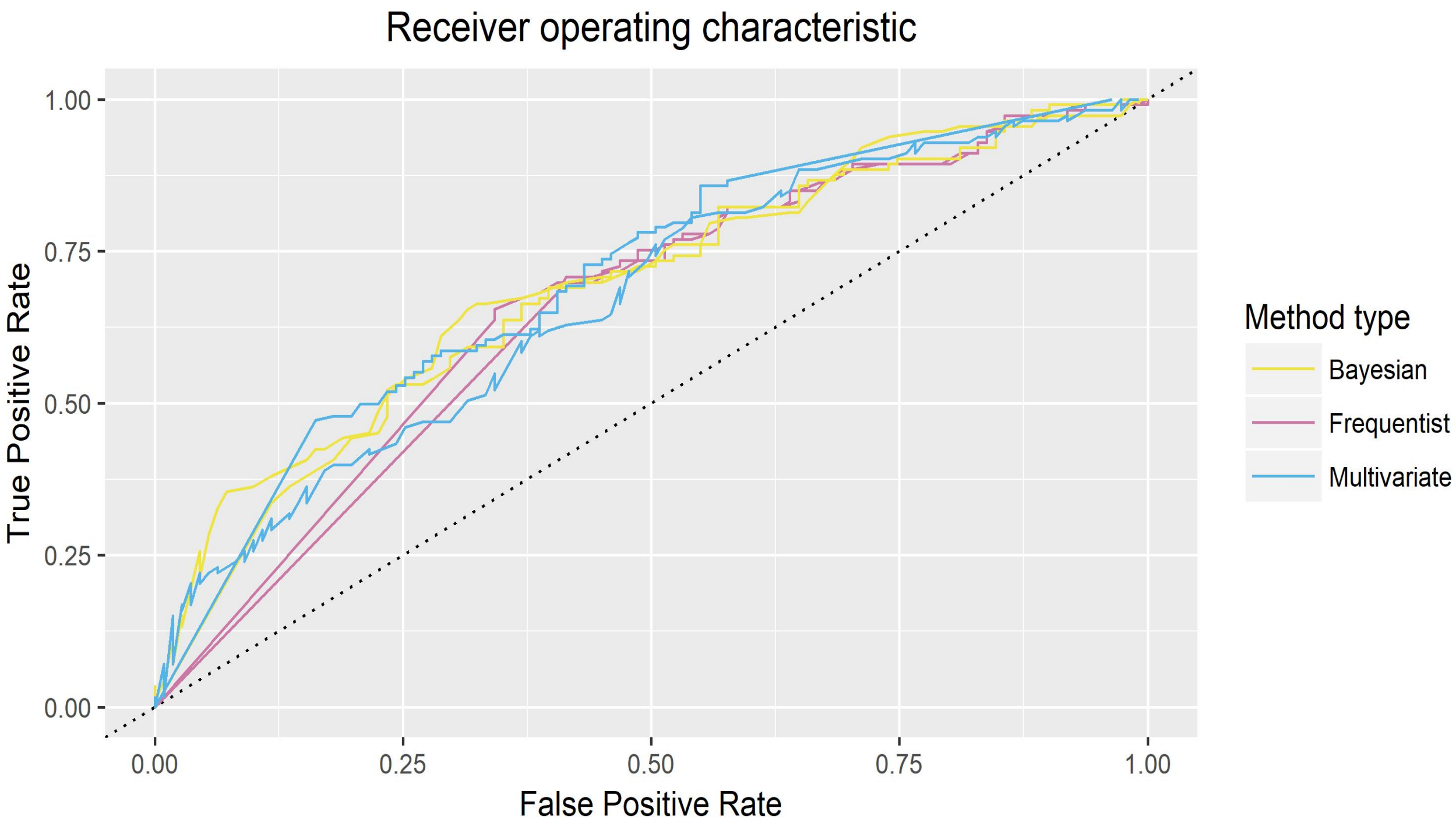
THE TESTING BED

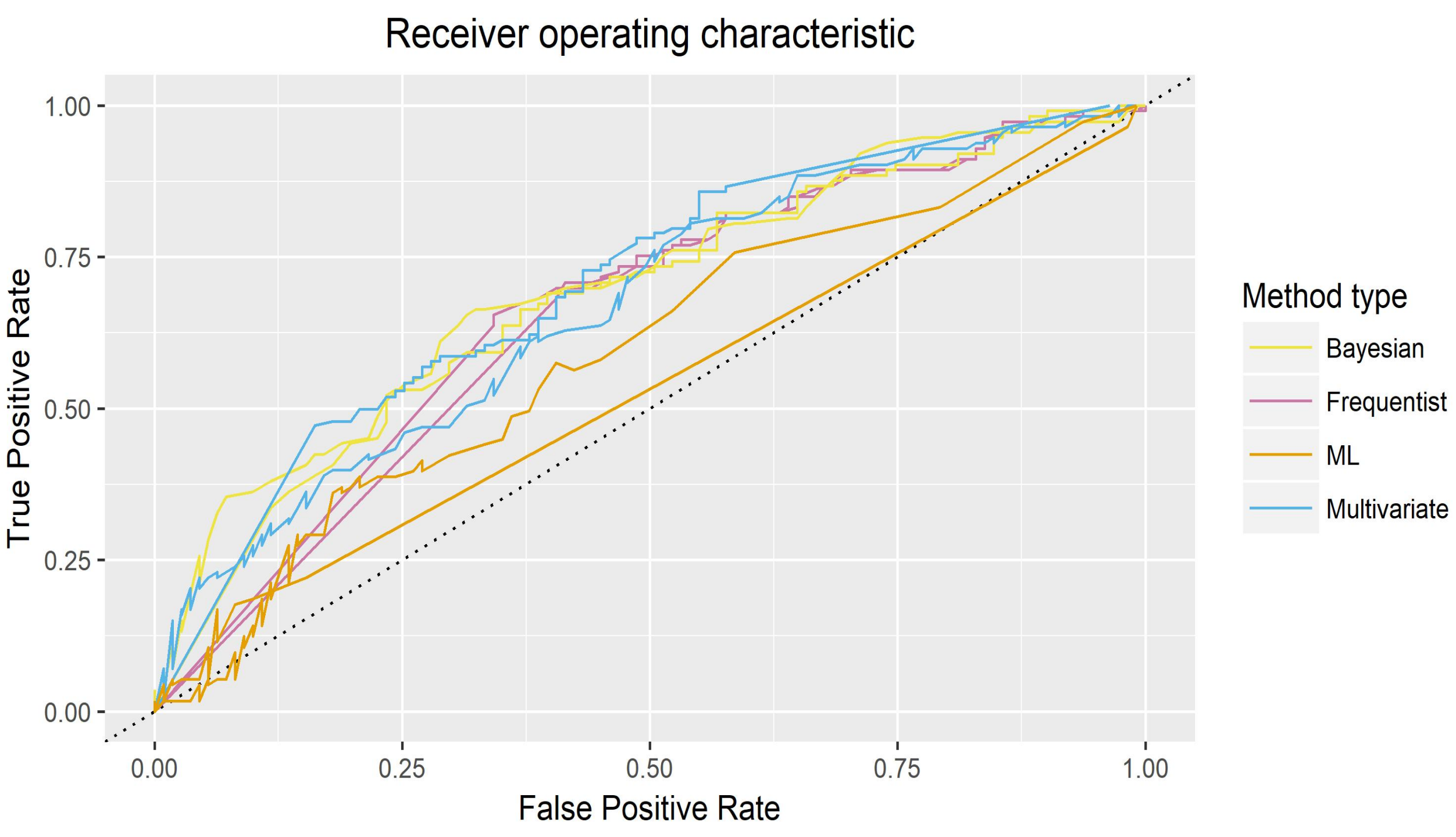
- Observational Medical Outcomes Partnership (OMOP) Gold Standard
  - Made by expert consensus
  - 165 positive controls
  - 233 negative controls


Receiver operating characteristic
True Positive Rate
False Positive Rate
0.00
0.25
0.50
0.75
1.00
Method
BCPNN
GPS
Logistic Reg
MC Logic Reg
PRR
Random Forests
RGPS
ROR


Receiver operating characteristic
True Positive Rate
False Positive Rate
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
Method type
Frequentist


Receiver operating characteristic
True Positive Rate
False Positive Rate
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
Method type
Bayesian
Frequentist


Receiver operating characteristic
True Positive Rate
False Positive Rate
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
Method type
Bayesian
Frequentist
Multivariate


Receiver operating characteristic
True Positive Rate
False Positive Rate
0.00
0.25
0.50
0.75
1.00
Method type
Bayesian
Frequentist
ML
Multivariate

AREA UNDER CURVE & SPEED

Method	AUC
RGPS	0.7091
BCPNN	0.6939
GPS	0.6803
ROR	0.6653
Logistic Reg	0.6604
PRR	0.6514
MC Logic Reg	0.5850
Random Forests	0.5208

Method	Computing Time
Logistic Reg	8.04 minutes
PRR	12.73 minutes
ROR	12.73 minutes
GPS	12.73 minutes
BCPNN	12.81 minutes
RGPS	14.13 minutes
MC Logic Reg	14.21 minutes
Random Forests	8.17 hours

FUTURE DEVELOPMENT

- Methods to detect drug interactions without specifying
- GPS and RGPS: can be improved by using EM algorithm to estimate parameters instead of Newton-type algorithm
- Automate the screening process and make public

REFERENCES

- Evans, S. J. W., Waller, P. C., & Davis, S. (2001). Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. Pharmacoepidemiology and drug safety, 10(6), 483-486.
- Rothman, K. J., Lanes, S., & Sacks, S. T. (2004). The reporting odds ratio and its advantages over the proportional reporting ratio. Pharmacoepidemiology and drug safety, 13(8), 519-523.
- Waller, P., Van Puijenbroek, E. P., Egberts, A. C. G., & Evans, S. (2004). The reporting odds ratio versus the proportional reporting ratio:'deuce'. Pharmacoepidemiology and drug safety, 13(8), 525-526.
- DuMouchel, W., & Pregibon, D. (2001, August). Empirical bayes screening for multi-item associations. In Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 67-76). ACM.
- DuMouchel, W. (1999). Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. The American Statistician, 53(3), 177-190.
- DuMouchel, W., Fram, D., Yang, X., Mahmoud, R. A., Grogg, A. L., Engelhart, L., & Ramaswamy, K. (2008). Antipsychotics, glycemic disorders, and life-threatening diabetic events: a Bayesian data-mining analysis of the FDA adverse event reporting system (1968–2004). Annals of Clinical Psychiatry, 20(1), 21-31.
- Ryan, P. B., Madigan, D., Stang, P. E., Marc Overhage, J., Racoosin, J. A., & Hartzema, A. G. (2012). Empirical assessment of methods for risk identification in healthcare data: results from the experiments of the Observational Medical Outcomes Partnership. Statistics in medicine, 31(30), 4401-4415.
- A. Liaw and M. Wiener (2002). Classification and Regression by randomForest. R News 2(3), 18--22.
- Kooperberg, C., & Ruczinski, I. (2011). LogicReg: Logic Regression. R package version, 1(10).
- Bate, A., Lindquist, M., Edwards, I. R., Olsson, S., Orre, R., Lansner, A., & De Freitas, R. M. (1998). A Bayesian neural network method for adverse drug reaction signal generation. European journal of clinical pharmacology, 54(4), 315-321.