



# Prediction models for case review in PV - How It Started vs. How It's Going

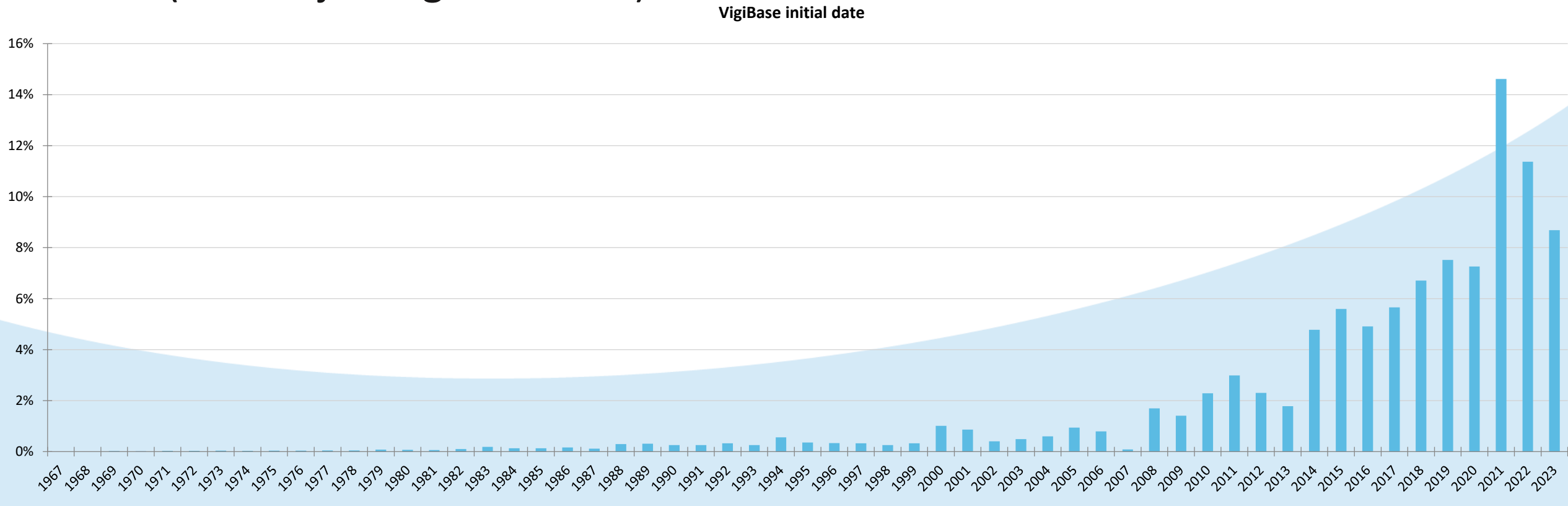
Florence van Hunsel, PharmD-Epidemiologist, PhD  
Lead Safety Innovation

# Conflicts of interest

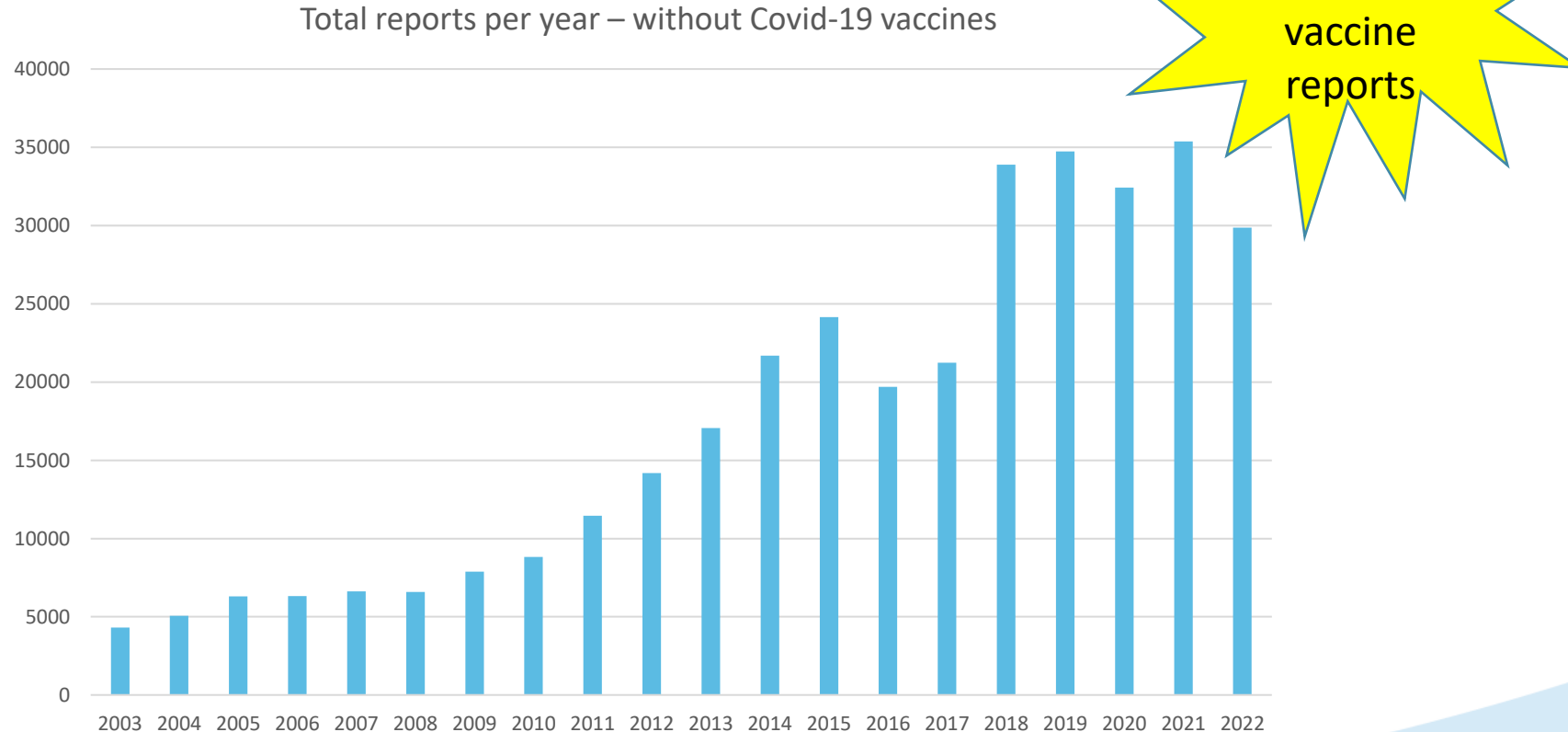
I have no conflicts of interest to declare

# Background

- The number of Individual Case Safety Reports (ICSRs) in pharmacovigilance databases are rapidly increasing world-wide (>34 milj in Vigibase now)

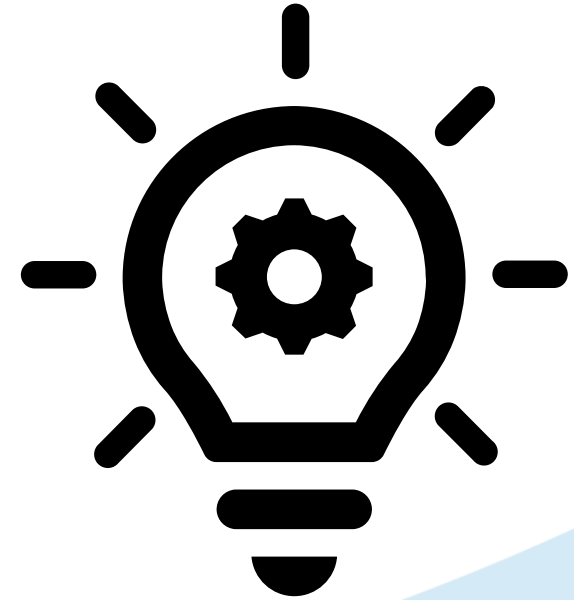


# Background



- The majority of ICSRs at the Netherlands Pharmacovigilance Centre Lareb gets a manual review to identify potential signal triggering reports (PSTR) or ICSRs which need further clinical assessment for other reasons

# Automation in review for signal detection



# Different approaches – Database screening

Drug Saf (2014) 37:617–628  
DOI 10.1007/s40264-014-0204-5

ORIGINAL RESEARCH ARTICLE

## Improved Statistical Signal Detection in Pharmacovigilance by Combining Multiple Strength-of-Evidence Aspects in *vigiRank*

Retrospective Evaluation against Emerging Safety Signals

Ola Caster · Kristina Juhlin ·  
Sarah Watson · G. Niklas Norén

Published online: 23 July 2014  
© The Author(s) 2014. This article is published with open access at Springerlink.com

### Abstract

**Background** Detection of unknown risks with marketed medicines is key to securing the optimal care of individual patients and to reducing the societal burden from adverse drug reactions. Large collections of individual case reports remain the primary source of information and require effective analytics to guide clinical assessors towards likely drug safety signals. Disproportionality analysis is based solely on aggregate numbers of reports and naively disregards report quality and content. However, these latter features are the very fundament of the ensuing clinical assessment.

**Objective** Our objective was to develop and evaluate a

from 2003 to 2007) and 5,280 negative controls (pairs of drugs and adverse events not listed in the Summary of Product Characteristics of that drug in 2012) was used for model fitting and evaluation; the latter used fivefold cross-validation to protect against over-fitting. All analyses were performed on a reconstructed version of *VigiBase*® as of 31 December 2004, at around which time most safety signals in our reference set were emerging.

**Results** The following aspects of strength of evidence were selected for inclusion into *vigiRank*: the numbers of informative and recent reports, respectively; disproportional



# Different approaches – Database screening

Drug Saf (2014) 37:617–628  
DOI 10.1007/s40264-014-0204-5

ORIGINAL RESEARCH ARTICLE

Received: 1 March 2017 | Revised: 10 October 2017 | Accepted: 2 November 2017

DOI: 10.1002/pds.4364

WILEY

## Improved Statistical Signal Detection by Combining Multiple Strength-of-Evidence Approaches: A Retrospective Evaluation against Emerging Safety Signals

Retrospective Evaluation against Emerging Safety Signals

Ola Caster · Kristina Juhlin ·  
Sarah Watson · G. Niklas Norén

Published online: 23 July 2014

© The Author(s) 2014. This article is published with open access at Springerlink.com




### Abstract

**Background** Detection of unknown risks with marketed medicines is key to securing the optimal care of individual patients and to reducing the societal burden from adverse drug reactions. Large collections of individual case reports remain the primary source of information and require effective analytics to guide clinical assessors towards likely drug safety signals. Disproportionality analysis is based solely on aggregate numbers of reports and naively disregards report quality and content. However, these latter features are the very fundament of the ensuing clinical assessment.

**Objective** Our objective was to develop and evaluate a

### ORIGINAL REPORT

## A prediction model-based algorithm for computer-assisted database screening of adverse drug reactions in the Netherlands

Joep H.G. Scholl<sup>1,2</sup>  | Florence P.A.M. van Hunsel<sup>1</sup>  | Eelko Hak<sup>2</sup> |  
Eugène P. van Puijenbroek<sup>1,2</sup> 

<sup>1</sup> Netherlands Pharmacovigilance Centre Lareb, 's-Hertogenbosch, The Netherlands

<sup>2</sup> PharmacoTherapy, -Epidemiology and -Economics, Groningen Research Institute of Pharmacy, University of Groningen, The Netherlands

### Correspondence

J. Scholl, Netherlands Pharmacovigilance Centre Lareb, Goudsbloemvallei 7, 5237 MH 's-Hertogenbosch, The Netherlands.

### Abstract

**Purpose:** The statistical screening of pharmacovigilance databases containing spontaneously reported adverse drug reactions (ADRs) is mainly based on disproportionality analysis. The aim of this study was to improve the efficiency of full database screening using a prediction model-based approach.

**Methods:** A logistic regression-based prediction model containing 5 candidate predictors was developed and internally validated using the Summary of Product Characteristics as the gold stan-

# Different approaches – Case utility

Drug Safety (2020) 43:329–338  
<https://doi.org/10.1007/s40264-019-00897-0>

ORIGINAL RESEARCH ARTICLE



## Towards Automating Adverse Event Review: A Prediction Model for Case Report Utility

Monica A. Muñoz<sup>1,2</sup>  · Gerald J. Dal Pan<sup>1</sup>  · Yu-Jung Jenny Wei<sup>2,4</sup>  · Chris Delcher<sup>3</sup>  · Hong Xiao<sup>2</sup>  ·  
Cindy M. Kortepeter<sup>1</sup>  · Almut G. Winterstein<sup>2,4,5</sup> 

Published online: 7 January 2020

© This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2020

### Abstract

**Introduction** The rapidly expanding size of the Food and Drug Administration's (FDA) Adverse Event Reporting System database requires modernized pharmacovigilance practices. Techniques to systematically identify high utility individual case safety reports (ICSRs) will support safety signal management.

**Objectives** The aim of this study was to develop and validate a model predictive of an ICSR's pharmacovigilance utility (PVU).

**Methods** PVU was operationalized as an ICSR's inclusion in an FDA-authored pharmacovigilance review's case series supporting a recommendation to modify product labeling. Multivariable logistic regression models were used to examine the association between PVU and ICSR features. The best performing model was selected for bootstrapping validation. As







# Different app

Drug Safety (2020) 43:329–338  
<https://doi.org/10.1007/s40264-019-00897-0>

## ORIGINAL RESEARCH ARTICLE

### Towards Automating Adverse Event Review for Case Report Utility

Monica A. Muñoz<sup>1,2</sup>  · Gerald J. Dal Pan<sup>1</sup>  · Yu-Jung Jenny W. Cindy M. Kortepeter<sup>1</sup>  · Almut G. Winterstein<sup>2,4,5</sup> 

Published online: 7 January 2020

© This is a U.S. government work and not under copyright protection in the U.S.;

#### Abstract

**Introduction** The rapidly expanding size of the Food and Drug database requires modernized pharmacovigilance practices. Ten case safety reports (ICSRs) will support safety signal management. **Objectives** The aim of this study was to develop and validate a (PVU).

**Methods** PVU was operationalized as an ICSR's inclusion in a supporting a recommendation to modify product labeling. Multi the association between PVU and ICSR features. The best perfor

Drug Safety (2023) 46:847–855  
<https://doi.org/10.1007/s40264-023-01327-y>

## ORIGINAL RESEARCH ARTICLE



### Finding Needles in the Haystack: Clinical Utility Score for Prioritisation (CUSP), an Automated Approach for Identifying Spontaneous Reports with the Highest Clinical Utility

Vijay Kara<sup>1</sup> · Greg Powell<sup>2</sup>  · Olivia Mahaux<sup>3</sup> · Aparna Jayachandra<sup>4</sup> · Naashika Nyako<sup>1</sup> · Christopher Golds<sup>1</sup> · Andrew Bate<sup>1,5</sup>

Accepted: 11 June 2023 / Published online: 3 August 2023

© GlaxoSmithKline Biologicals S.A 2023, corrected publication 2023

#### Abstract

**Introduction** Spontaneous reporting of adverse events has increased steadily over the past decades, and although this trend has contributed to improving post-marketing surveillance pharmacovigilance activities, the consequent amount of data generated is challenging to manually review during assessment, with each individual report requiring review by pharmacovigilance experts. This highlights a clear need for alternative or complementary methodologies to help prioritise review. **Objective** Here, we aimed to develop and test an automated methodology, the Clinical Utility Score for Prioritisation (CUSP), to assist pharmacovigilance experts in prioritising clinical assessment of safety data to improve the rapidity of case series review when case volumes are large.

**Methods** The CUSP method was tested on a reference dataset of individual case safety reports (ICSRs) associated to five drug-event pairs that led to labelling changes. The selected drug-event pairs were of varying characteristics across the portfolio of GSK's products.

**Results** The mean CUSP score for 'key cases' and 'cases of low utility' was 19.7 (median: 21; range: 7–27) and 17.3 (median: 10; range: 4–27), respectively. CUSP distribution for 'key cases' were skewed toward the higher range of scores compared



# How it started...

With a Question: **How can we better predict which cases need to be assessed by clinical reviewers at Lareb?**

- We aimed to develop a machine learning prediction model to identify ICSRs which require clinical review
- And determine the importance of the used predictors, referred to as features, in the model

ORIGINAL ARTICLE

# Development of a multivariate prediction model to identify individual case safety reports which require clinical review

Helen R. Gosselt  | Elizabeth A. Bazelmans | Thomas Lieber |  
Florence P. A. M. van Hunsel  | Linda Härmark

Netherlands Pharmacovigilance Centre Lareb,  
's-Hertogenbosch, The Netherlands

**Correspondence**

Helen R. Gosselt, Netherlands  
Pharmacovigilance Centre Lareb,  
Goudsbloemvallei 7, 's-Hertogenbosch, North  
Brabant, 5237MH, The Netherlands.  
Email: [h.gosselt@lareb.nl](mailto:h.gosselt@lareb.nl)

## Abstract

**Background:** The number of Individual Case Safety Reports (ICSRs) in pharmacovigilance databases are rapidly increasing world-wide. The majority of ICSRs at the Netherlands Pharmacovigilance Centre Lareb is reviewed manually to identify potential signal triggering reports (PSTR) or ICSRs which need further clinical assessment for other reasons.

**Objectives:** To develop a prediction model to identify ICSRs that require clinical review, including PSTRs. Secondly, to identify the most important features of these reports.

**Methods:** All ICSRs ( $n = 30\,424$ ) received by Lareb between October 1, 2017 and

# Methods

- Model based on potential signal triggering reports (PSTRs) which were discussed weekly in a Signal Detection Meeting (SDM)
- Exclusion criteria:
  - ICSRs originating from marketing authorisation holders
  - ICSRs with vaccines as a suspect drug
- Potential features were selected based on:
  - Expert opinion from scientific assessors at Lareb (e.g. *seriousness of an ADR, number of ADRs reported on the drug-event combination, number of suspected drugs, reporter type, positive dechallenge or rechallenge, latency of the ADR*)
  - Literature (*designated medical event, in this study defined as important medical event (IME) by European Medicines Agency*)
  - Availability in the ICSR database

# Methods

- Dataset used contained 30.424 ICSRs of which 1439 (4.7%) were PSTR=1 and 19 features
- Training (70%) and test set (30%) and correction for imbalanced dataset in training set
- Three models tested
  - Logistic regression
  - Elastic net logistic regression
  - Decision tree based model: eXtreme Gradient Boosting (XGBoost)
- Model performances were assessed using the area under the curve (AUC) of a receiver operating characteristic (ROC), together with sensitivity and specificity



# Methods

- For the application of the model in practice, no PSTRs should be missed
- We assessed model performances at a threshold where the sensitivity was 100% on the training data

# Results

- All three models performed equally with a highest AUC of 0.75 (0.73-0.77)
- At a cut-off where the sensitivity was 100%, which is required to not miss any PSTRs, false positive rates were also high
- ~5% of all reports were correctly classified as true negatives at a cut-off where sensitivity was 100%
- Features that were important in all 3 models were:
  - ‘absence of ADR in the SmPC’,
  - ‘ADR reported as serious’
  - ‘ADR labelled as an important medical event’,
  - ‘ADR reported by physician’
  - ‘rechallenge’

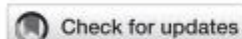
# Results

- All three models performed equally with a highest AUC of 0.75 (0.73-0.77)
- At a cut-off where the sensitivity was 100%, which is required to not miss any PSTRs, false positive rates were also high
- ~5% of all reports were correctly classified as true negatives at a cut-off where sensitivity was 100%
- Features that were important in all 3 models were:
  - ‘absence of ADR in the SmPC’,
  - ‘ADR reported as serious’
  - ‘ADR labelled as an important medical event’,
  - ‘ADR reported by physician’
  - ‘rechallenge’

# Is this good enough?

- The best model could distinguish PSTRs from non-triggering reports with an AUC of 0.75 (0.73 – 0.77), which is comparable to the AUC of 0.71 of the previously developed model by Munoz et al. on FDA reports\*
- However, many false positives with sensitivity set at 100%
- We defined PSTRs as reports that are discussed at a signal detection meetings (SDMs). Other reasons for taking ICSR to SMD than signal value
- Additional information in free text not taken in to account

\*Muñoz MA, Dal Pan GJ, Wei YJ, Delcher C, Xiao H, Kortepeter CM, Winterstein AG. Towards Automating Adverse Event Review: A Prediction Model for Case Report Utility. Drug Saf. 2020 Apr;43(4):329-338. doi: 10.1007/s40264-019-00897-0. PMID: 31912439.



#### OPEN ACCESS

EDITED BY  
Marco Tuccori,  
University of Pisa, Italy

REVIEWED BY  
Giulia Valdiserra,  
University of Pisa, Italy  
Michele Fusaroli,  
University of Bologna, Italy

\*CORRESPONDENCE  
Thomas Lieber,  
✉ t.lieber@lareb.nl

SPECIALTY SECTION  
This article was submitted to  
Advanced Methods in  
Pharmacovigilance and  
Pharmacoepidemiology,  
a section of the journal  
Frontiers in Drug Safety  
and Regulation

RECEIVED 09 December 2022  
ACCEPTED 24 January 2023  
PUBLISHED 07 February 2023

CITATION

# Natural language processing for automated triage and prioritization of individual case safety reports for case-by-case assessment

Thomas Lieber<sup>1\*</sup>, Helen R. Gosselt<sup>1</sup>, Pelle C. Kools<sup>2</sup>,  
Okko C. Kruijssen<sup>2</sup>, Stijn N. C. Van Lierop<sup>2</sup>, Linda Härmark<sup>1</sup> and  
Florence P. A. M. Van Hunsel<sup>1</sup>

<sup>1</sup>Netherlands Pharmacovigilance Centre Lareb, 's-Hertogenbosch, Netherlands, <sup>2</sup>Faculty of Social Sciences, Radboud Universiteit, Nijmegen, Netherlands

**Objective:** To improve a previously developed prediction model that could assist in the triage of individual case safety reports using the addition of features designed from free text fields using natural language processing.

**Methods:** Structured features and natural language processing (NLP) features were used to train a bagging classifier model. NLP features were extracted from free text



# Model II

- In addition to structured features, natural language processing (NLP) features were used to train a bagging classifier model
- NLP features were extracted from free text fields
- A bag-of-words model was applied. Stop words were deleted and words that were significantly differently distributed among the case and non-case reports were used for the training data.
- Besides NLP features from free-text fields, the data also consisted of a list of signal words deemed important by expert report assessors.
- Lastly, variables with multiple categories were transformed to numerical variables using the weight of evidence method.

# Results

- the model, a bagging classifier of decision trees, had an AUC of 0.921 (95% CI = 0.918–0.925)
- This time we did not go for 100% sensitivity, but for a prioritization list that could aid assessors in determining which case to look at

# Is this good enough?

- Model performance OK
- Dutch PV centre is working towards auto-coding/automated handling of majority of cases
- This requires a model which helps which triage as soon as a report comes is → before it is coded in MedDRA

## RESEARCH ARTICLE

# BERT based natural language processing for triage of adverse drug reaction reports shows close to human-level performance

Erik Bergman<sup>1</sup>, Luise Dürlich<sup>1,2,3</sup>, Veronica Arthurson<sup>1</sup>, Anders Sundström<sup>1</sup>, Maria Larsson<sup>1</sup>, Shamima Bhuiyan<sup>1</sup>, Andreas Jakobsson<sup>4</sup>, Gabriel Westman<sup>1,5\*</sup>

**1** Swedish Medical Products Agency, Uppsala, Sweden, **2** Department of Computer Science, RISE Research Institutes of Sweden, Kista, Sweden, **3** Department of Linguistics and Philology, Uppsala University, Uppsala, Sweden, **4** Centre for Mathematical Sciences, Lund University, Lund, Sweden, **5** Department of Medical Sciences, Uppsala University, Uppsala, Sweden

\* [gabriel.westman@lakemedelsverket.se](mailto:gabriel.westman@lakemedelsverket.se)



## Abstract

Post-marketing reports of suspected adverse drug reactions are important for establishing the safety profile of a medicinal product. However, a high influx of reports poses a challenge for regulatory authorities as a delay in identification of previously unknown adverse drug reactions can potentially be harmful to patients. In this study, we use natural language processing (NLP) to predict whether a report is of serious nature based solely on the free-text

## OPEN ACCESS

**Citation:** Bergman E, Dürlich L, Arthurson V,

# Model III

- Trained on 1st uncoded 'raw' version of ICSRs
- Use of a clinical BERT (Bidirectional Encoder Representations from Transformers) model
- Takes in to account vaccine reports as well
- Features in the model: ADR description in free text, seriousness, reporter type, etc



# What does BERT think?

No priority case (98%). Correct

#s Sent in by a patient . headache , fever : 37 . 5 to 38 degrees Cel ci us , chills ,  
nausea , feeling un well , joint pain , orbital pain , muscle ac hes , fatigue . Since  
August 8 th , still suffering from . #/s

Priority case (99%). Correct

#s Sent in by a doctor . Reaction was severe . arterial occlusion


No priority case (55%). Incorrect

#s Sent in by a patient . 4 years in the menopause , was suddenly menstr uating  
for 6 days . 4 years in the menopause , was suddenly menstr uating for 6 days .

# How it's going?

- Good model performance overall → but driven by high input of Covid-19 vaccine reports
- Sensitivity analysis for drugs vs vaccines
- Adding additional features to the model (for instance, did reporter upload a document etc)
- Importance of model evaluation!

# Black Swan Events and Intelligent Automation for Routine Safety Surveillance

Oeystein Kjoersvik<sup>1</sup> · Andrew Bate<sup>2,3,4</sup> 

Accepted: 27 February 2022  
© TransCelerate BioPharma, Inc. 2022

## Abstract

Effective identification of previously implausible safety signals is a core component of successful pharmacovigilance. Timely, reliable, and efficient data ingestion and related processing are critical to this. The term ‘black swan events’ was coined by Taleb to describe events with three attributes: unpredictability, severe and widespread consequences, and retrospective bias. These rare events are not well understood at their emergence but are often rationalized in retrospect as predictable. Pharmacovigilance strives to rapidly respond to potential black swan events associated with medicine or vaccine use. Machine learning (ML) is increasingly being explored in data ingestion tasks. In contrast to rule-based automation approaches, ML can use historical data (i.e., ‘training data’) to effectively predict emerging data patterns and support effective data intake, processing, and organisation. At first sight, this reliance on previous data might be considered a limitation when building ML models for effective data ingestion in systems that look to focus on the identification of potential black swan events. We argue that, first, some apparent black swan events—although unexpected medically—will exhibit data attributes similar to those of other safety data and not prove algorithmically unpredictable, and, second, standard and emerging ML approaches can still be robust to such data outliers with proper awareness and consideration in ML system design and with the incorporation of specific mitigatory and support strategies. We argue that effective approaches to managing data on potential black swan events are essential for trust and outline several strategies to address data on potential black swan events during data ingestion.





# Questions?



Email: [f.vanhunsel@lareb.nl](mailto:f.vanhunsel@lareb.nl)