What evidence do we want to generate?

Patrick Ryan, Daniel Prieto Alhambra, Niklas Norén
A caricature of the patient journey
Questions asked across the patient journey

- Which treatment did patients choose after diagnosis?
- Which patients chose which treatments?
- How many patients experienced the outcome after treatment?
- Does one treatment cause the outcome more than an alternative?
- Does treatment cause outcome?
- What is the probability I will develop the disease?
- What is the probability I will experience the outcome?
Complementary evidence to inform the patient journey

Clinical characterization: What happened to them?

Patient-level prediction: What will happen to me?

Population-level effect estimation: What are the causal effects?

Observation

Inference

Causal inference
What evidence do we want to generate?

Treatment utilization studies

D Prieto-Alhambra
Prof of Pharmaco- and Device Epidemiology
Centre for Statistics in Medicine, NDORMS, University of Oxford
Complementary evidence to inform the patient journey

Clinical characterization: What happened to them?

Patient-level prediction: What will happen to me?

Population-level effect estimation: What are the causal effects?

Observation

Inference

Causal inference
AGENDA

• Real World Evidence: Why do I bother?
• Why are treatment utilization studies relevant?
• Multi-country DUS
• Future directions and opportunities
AGENDA

• Real World Evidence: Why do I bother?
  • Why are treatment utilization studies relevant?
  • Multi-country DUS
• Future directions and opportunities
Why should I bother?
Why ‘real world’ evidence?
1. When RCTs are not possible ...

Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials

BMJ 2003

What this study adds

No randomised controlled trials of parachute use have been undertaken

The basis for parachute use is purely observational, and its apparent efficacy could potentially be explained by a “healthy cohort” effect

The medicalisation of free fall

It is often said that doctors are interfering monsters obsessed with disease and power, who will not be satisfied until they control every aspect of our lives (Journal

# OF RCTs = 0
Why ‘real world’ evidence?
2. The data is out there ...

- Over 20,000 TKA patients
- Over 3y median follow-up

Study completed in just under 1 year ...

How long would it take you to recruit and follow-up?
Why ‘real world’ evidence?
2. The data is out there ...
Why ‘real world’ evidence?
3. External validity

ELIGIBILITY CRITERIA
• >70 year old
• Woman
• ...
Why ‘real world’ evidence?
3. External validity
AGENDA

• Real World Evidence: Why do I bother?
• Why are treatment utilization studies relevant?
  • Multi-country DUS
  • Future directions and opportunities
• Lack of evidence on certain population subgroups
• Adherence and persistence in actual practice
• On vs off-label use
• Risk Minimisation Measure/s Effectiveness

None of these can be assessed in RCTs or observational studies other than using ‘real world’ (routinely collected) data
Actual Drug Users vs RCT participants
How big is the gap?

Table 1  Comparison of the exclusion criteria in the FIT trial with the incident users of alendronate in the SIDIAP and DHR database

<table>
<thead>
<tr>
<th>FIT exclusion criteria</th>
<th>Operational definition/ICD-10 Codes</th>
<th>Incident users of Alendronate ( N = )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>SIDIAP ( N = 14,316 ) (% )</td>
</tr>
<tr>
<td>Men</td>
<td>Sex according to administrative data</td>
<td>3818 (26.7 %)</td>
</tr>
<tr>
<td>Age (&lt;55 \text{ years old}</td>
<td>Age at first ALD dispensation</td>
<td>1844 (12.9 %)</td>
</tr>
<tr>
<td>Age (&gt;80 \text{ years old}</td>
<td>Age at first ALD dispensation</td>
<td>2347 (16.4 %)</td>
</tr>
</tbody>
</table>

C Reyes et al. Osteoporos Int 2014
Adherence in the real world ... vs RCT

Adherence in RCT (Vigor study) vs “real life”
Rofecoxib users in CPRD
[TV Staa PLoS One ‘09]
Persistence in actual practice conditions ...

N = 127,076 SIDIAP participants, Cat, Spain

AGENDA

• Real World Evidence: Why do I bother?
• Why are treatment utilization studies relevant?
• Multi-country DUS
• Future directions and opportunities
THE NEED - The European context

• Different healthcare settings
• Different funding systems
• Different conditions of use for medicines
• ...

• And hence the importance of EU-wide DUS for the EMA
A previous example – EU-ADR Alliance Multi-country Population Level DUS

Monthly incidence rates (10,000 PY) of use

K Berencsi et al. ICPE 2018
A previous example – EU-ADR Alliance Multi-country Population Level DUS

Monthly incidence rates (10,000 PY) of use

Between-country differences

K Berencsi et al. ICPE 2018
A previous example – EU-ADR Alliance
Multi-country Population Level DUS

Monthly incidence rates (10,000 PY) of use

Incidence Rate

Calendar Years

Overall
Denmark
Italy
The Netherlands
United Kingdom
Spain

RMM Effectiveness

K Berencsi et al. ICPE 2018
AGENDA

• Real World Evidence: Why do I bother?
• Why are treatment utilization studies relevant?
• Multi-country DUS
• Future directions and opportunities
The EU DUS challenge
Interoperability
What can (and should) be improved?

• Automation of common processes
  – Data curation, analyses
• Speed -> from 1-2 years to 1-2 days
• Reproducibility
• More complex analytics
• Visualisation/s
<table>
<thead>
<tr>
<th>ANALYSES/ TOOLS</th>
<th>UC1 – DUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Population level (incidence, prevalence)</td>
<td></td>
</tr>
<tr>
<td>- Patient level (adherence, % with contra/indications)</td>
<td></td>
</tr>
<tr>
<td>- Treatment pathways</td>
<td></td>
</tr>
<tr>
<td>- Dose / Indication: NLP</td>
<td></td>
</tr>
<tr>
<td>- Secular trends (RMM)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DASHBOARD/S VISUALISATION</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>- Sunburst / Sankey plots</td>
<td></td>
</tr>
<tr>
<td>- Interactive filters/zoom</td>
<td></td>
</tr>
<tr>
<td>- Subgroup stratification</td>
<td></td>
</tr>
<tr>
<td>- Patterns over time</td>
<td></td>
</tr>
</tbody>
</table>
What evidence do we want to generate?
Pharmacovigilance

Niklas Norén
Chief Science Officer
Uppsala Monitoring Centre
Complementary evidence to inform the patient journey

Clinical characterization: What happened to them?

Patient-level prediction: What will happen to me?

Population-level effect estimation: What are the causal effects?

Observation

Inference

Causal inference
On the lookout for the unexpected?
Broad scope

Generic analysis strategies

Humility
Individual case reports

Report of Suspected Adverse Drug Reaction including Birth Defects

Note: Identifiers of Reporter, Patient and Institution will remain Confidential

Patient: [Redacted]
Date of Birth: 05 DEC 1965
Age: 80
Sex: M
Weight: 160
Height: [Redacted]

Adverse Reaction Description: In the case of a Patient with a Non ST Elevation MI who diagnosed hypertension some days before, the same day had PCI to LAD which experienced profound and sustained hypotension not believed to be related to angioplasty, a patient who experienced profound hypotension not believed to be related to angioplasty, and a patient who experienced profound hypotension not believed to be related to angioplasty, and a patient who experienced profound hypotension not believed to be related to angioplasty.

Date of Onset of Reaction: 29/11/06

All Drug Therapy Prior to Reaction

Daily Dosage and Route Date Begun Date Stopped Reason for Use

Aspirin 300 mg PO 11/14 NS MAN
Captopril 50 mg PO 11/14 NS MAN
Labetalol 30 mg PO 11/14 NS MAN
Diltiazem 20 mg PO 11/14 NS MAN
Hydroxyzine 50 mg PO 11/14 NS MAN

Daily Dosage and Route Date Begun Date Stopped Reason for Use

Livingstone 10 mg PO 11/14 NS MAN
Cimetidine 300 mg PO 11/14 NS MAN

Daily Dosage and Route Date Begun Date Stopped Reason for Use

Clonidine 0.1 mg PO 11/14 NS MAN
Procainamide 200 mg PO 11/14 NS MAN

Treatment: [Redacted]

Outcome: Recovered

Sequela: Myocardial Infarction

Comments: [Redacted]

Signature: [Redacted]

[Redacted]
By the millions
Hard to spot
Hard to assess
Whose risk?

228 vs 267
Observational health data

+ Denominators
  Longitudinal

- Secondary use
  No causality assessment
Signal detection in electronic health records
Signal strengthening in electronic health records

Longitudinal medical records as a complement to routine drug safety signal analysis

Kristina Star, Samih Watan, Lovisa Sandberg, Jeanette Jehanno and I. Ralph Edwards

Updata Monitoring Centre, WHO Collaborating Centre for International Drug Monitoring, Uppsala, Sweden

ABSTRACT

Objective: To explore whether and how longitudinal medical records could be used as a source of evidence in the early phases of signal detection and analysis of novel adverse drug reactions (ADRs) in a global pharmacovigilance database.

Methods: Drug and ADR combinations from the routine signal detection process at Uppsala in 2011 were matched to combinations in the Global Drug Safety Interchange Network (GDSIN). The number and type of drugs and ADRs from the data sets were investigated. For unmatched combinations, additional display of temporal overlap patterns in Longitudinal Medical Records (LMRs) was performed to determine if the pattern supported signal strengthening.

Results: 109 OR combinations in the Uppsala data set, 109 matched to corresponding combinations in GDSIN after excluding drugs with less than 100 prescriptions in GDSIN. Exploration of the Uppsala and GDSIN data sets identified 15 drug and ADR combinations with observed temporal overlap patterns in LMRs. For all of these combinations, differences in prescribing patterns between specific drugs and their temporal overlap patterns in LMRs were explored. The clinical relevance of drug use was assessed using the clinical context around the drug and supported ADR combination.

Conclusion: Although the study is exploratory, administration of an electronic health record system to support signal strengthening for signal detection and signal strengthening for signal strengthening in electronic health record systems, the clinical context around a drug and supported ADR combination. The study is exploratory, administration of an electronic health record system to support signal strengthening for signal detection and signal strengthening for signal strengthening in electronic health record systems, the clinical context around a drug and supported ADR combination.

INTRODUCTION

To evaluate the implications of suspected harm from a drug, it is useful to signal analysis. The definitive description of the clinical setting is essential. Too often, this information is incomplete in individual case safety reports (ICSRs) of suspected adverse drug reactions (ADRs). The identification of signals in large collections of ICSR data, however, may conceal important signals that are reported disproportionately more frequently than expected. Reasons need to be made whether the statistical signal should be subject to in-depth investigation and whether a signal should be communicated. Sometimes decisions in this process must be based on detailed or very limited information be available. Randomized clinical trials, on which a drug’s marketing approval in Europe, are not always publicly available, and supportive published case reports might not yet exist.

Electronic medical record (EMR) databases contain clinical data—diagnoses, observations, laboratory results and procedures. They provide the potential to capture information from a patient’s entire medical history throughout the course of a patient’s life. Longitudinal data might help to understand the clinical relevance of the drug and the relationship between adverse events. They present opportunities for further insights into drug safety problems and have been used primarily in in-depth analyses. Recently, several initiatives have investigated whether the statistical signal should be subject to in-depth investigation and whether a signal should be communicated. Sometimes decisions in this process must be based on detailed or very limited information be available. Randomized clinical trials, on which a drug’s marketing approval in Europe, are not always publicly available, and supportive published case reports might not yet exist.

Electronic medical record (EMR) databases contain clinical data—diagnoses, observations, laboratory results and procedures. They provide the potential to capture information from a patient’s entire medical history throughout the course of a patient’s life. Longitudinal data might help to understand the clinical relevance of the drug and the relationship between adverse events. They present opportunities for further insights into drug safety problems and have been used primarily in in-depth analyses. Recently, several initiatives have investigated whether the statistical signal should be subject to in-depth investigation and whether a signal should be communicated. Sometimes decisions in this process must be based on detailed or very limited information be available. Randomized clinical trials, on which a drug’s marketing approval in Europe, are not always publicly available, and supportive published case reports might not yet exist.

Electronic medical record (EMR) databases contain clinical data—diagnoses, observations, laboratory results and procedures. They provide the potential to capture information from a patient’s entire medical history throughout the course of a patient’s life. Longitudinal data might help to understand the clinical relevance of the drug and the relationship between adverse events. They present opportunities for further insights into drug safety problems and have been used primarily in in-depth analyses. Recently, several initiatives have investigated whether the statistical signal should be subject to in-depth investigation and whether a signal should be communicated. Sometimes decisions in this process must be based on detailed or very limited information be available. Randomized clinical trials, on which a drug’s marketing approval in Europe, are not always publicly available, and supportive published case reports might not yet exist.

Electronic medical record (EMR) databases contain clinical data—diagnoses, observations, laboratory results and procedures. They provide the potential to capture information from a patient’s entire medical history throughout the course of a patient’s life. Longitudinal data might help to understand the clinical relevance of the drug and the relationship between adverse events. They present opportunities for further insights into drug safety problems and have been used primarily in in-depth analyses. Recently, several initiatives have investigated whether the statistical signal should be subject to in-depth investigation and whether a signal should be communicated. Sometimes decisions in this process must be based on detailed or very limited information be available. Randomized clinical trials, on which a drug’s marketing approval in Europe, are not always publicly available, and supportive published case reports might not yet exist.

Electronic medical record (EMR) databases contain clinical data—diagnoses, observations, laboratory results and procedures. They provide the potential to capture information from a patient’s entire medical history throughout the course of a patient’s life. Longitudinal data might help to understand the clinical relevance of the drug and the relationship between adverse events. They present opportunities for further insights into drug safety problems and have been used primarily in in-depth analyses. Recently, several initiatives have investigated whether the statistical signal should be subject to in-depth investigation and whether a signal should be communicated. Sometimes decisions in this process must be based on detailed or very limited information be available. Randomized clinical trials, on which a drug’s marketing approval in Europe, are not always publicly available, and supportive published case reports might not yet exist.

Electronic medical record (EMR) databases contain clinical data—diagnoses, observations, laboratory results and procedures. They provide the potential to capture information from a patient’s entire medical history throughout the course of a patient’s life. Longitudinal data might help to understand the clinical relevance of the drug and the relationship between adverse events. They present opportunities for further insights into drug safety problems and have been used primarily in in-depth analyses. Recently, several initiatives have investigated whether the statistical signal should be subject to in-depth investigation and whether a signal should be communicated. Sometimes decisions in this process must be based on detailed or very limited information be available. Randomized clinical trials, on which a drug’s marketing approval in Europe, are not always publicly available, and supportive published case reports might not yet exist.

Electronic medical record (EMR) databases contain clinical data—diagnoses, observations, laboratory results and procedures. They provide the potential to capture information from a patient’s entire medical history throughout the course of a patient’s life. Longitudinal data might help to understand the clinical relevance of the drug and the relationship between adverse events. They present opportunities for further insights into drug safety problems and have been used primarily in in-depth analyses. Recently, several initiatives have investigated whether the statistical signal should be subject to in-depth investigation and whether a signal should be communicated. Sometimes decisions in this process must be based on detailed or very limited information be available. Randomized clinical trials, on which a drug’s marketing approval in Europe, are not always publicly available, and supportive published case reports might not yet exist.

Electronic medical record (EMR) databases contain clinical data—diagnoses, observations, laboratory results and procedures. They provide the potential to capture information from a patient’s entire medical history throughout the course of a patient’s life. Longitudinal data might help to understand the clinical relevance of the drug and the relationship between adverse events. They present opportunities for further insights into drug safety problems and have been used primarily in in-depth analyses. Recently, several initiatives have investigated whether the statistical signal should be subject to in-depth investigation and whether a signal should be communicated. Sometimes decisions in this process must be based on detailed or very limited information be available. Randomized clinical trials, on which a drug’s marketing approval in Europe, are not always publicly available, and supportive published case reports might not yet exist.

Electronic medical record (EMR) databases contain clinical data—diagnoses, observations, laboratory results and procedures. They provide the potential to capture information from a patient’s entire medical history throughout the course of a patient’s life. Longitudinal data might help to understand the clinical relevance of the drug and the relationship between adverse events. They present opportunities for further insights into drug safety problems and have been used primarily in in-depth analyses. Recently, several initiatives have investigated whether the statistical signal should be subject to in-depth investigation and whether a signal should be communicated. Sometimes decisions in this process must be based on detailed or very limited information be available. Randomized clinical trials, on which a drug’s marketing approval in Europe, are not always publicly available, and supportive published case reports might not yet exist.

Electronic medical record (EMR) databases contain clinical data—diagnoses, observations, laboratory results and procedures. They provide the potential to capture information from a patient’s entire medical history throughout the course of a patient’s life. Longitudinal data might help to understand the clinical relevance of the drug and the relationship between adverse events. They present opportunities for further insights into drug safety problems and have been used primarily in in-depth analyses. Recently, several initiatives have investigated whether the statistical signal should be subject to in-depth investigation and whether a signal should be communicated. Sometimes decisions in this process must be based on detailed or very limited information be available. Randomized clinical trials, on which a drug’s marketing approval in Europe, are not always publicly available, and supportive published case reports might not yet exist.

Electronic medical record (EMR) databases contain clinical data—diagnoses, observations, laboratory results and procedures. They provide the potential to capture information from a patient’s entire medical history throughout the course of a patient’s life. Longitudinal data might help to understand the clinical relevance of the drug and the relationship between adverse events. They present opportunities for further insights into drug safety problems and have been used primarily in in-depth analyses. Recently, several initiatives have investigated whether the statistical signal should be subject to in-depth investigation and whether a signal should be communicated. Sometimes decisions in this process must be based on detailed or very limited information be available. Randomized clinical trials, on which a drug’s marketing approval in Europe, are not always publicly available, and supportive published case reports might not yet exist.

Electronic medical record (EMR) databases contain clinical data—diagnoses, observations, laboratory results and procedures. They provide the potential to capture information from a patient’s entire medical history throughout the course of a patient’s life. Longitudinal data might help to understand the clinical relevance of the drug and the relationship between adverse events. They present opportunities for further insights into drug safety problems and have been used primarily in in-depth analyses. Recently, several initiatives have investigated whether the statistical signal should be subject to in-depth investigation and whether a signal should be communicated. Sometimes decisions in this process must be based on detailed or very limited information be available. Randomized clinical trials, on which a drug’s marketing approval in Europe, are not always publicly available, and supportive published case reports might not yet exist.

Electronic medical record (EMR) databases contain clinical data—diagnoses, observations, laboratory results and procedures. They provide the potential to capture information from a patient’s entire medical history throughout the course of a patient’s life. Longitudinal data might help to understand the clinical relevance of the drug and the relationship between adverse events. They present opportunities for further insights into drug safety problems and have been used primarily in in-depth analyses. Recently, several initiatives have investigated whether the statistical signal should be subject to in-depth investigation and whether a signal should be communicated. Sometimes decisions in this process must be based on detailed or very limited information be available. Randomized clinical trials, on which a drug’s marketing approval in Europe, are not always publicly available, and supportive published case reports might not yet exist.
Risk group identification in electronic health records
On the market?

Primary vs secondary care

Recorded?

SCOPE

Over the counter
Enough data

on patients at risk
Speed & Cost efficiency?
Iteration and exploration
What evidence do we want to generate?

Patient-level prediction

Patrick Ryan
Janssen Research and Development
Columbia University Medical Center
Complementary evidence to inform the patient journey

Clinical characterization: What happened to them?

Patient-level prediction: What will happen to me?

Observation

Population-level effect estimation: What are the causal effects?

Inference

Causal inference
A candidate for predictive modeling?

• Fairly common disease (~80-110 cases per 100,000 person-years; lifetime risk ~5-13%)
• Serious prognosis following initial diagnosis
  – Though technical advances have greatly decreased mortality and improved quality of life
• The earlier the diagnosis, the higher likelihood of an effective treatment
Would you use this model in clinical practice?

Opportunity:
• Patients predicted to have disease who are confirmed and can obtain more timely treatment, reducing disease-related mortality

Challenges:
• Prediction is imperfect
  • 10% False positives – induces unnecessary worry and results in extra tests to refute model finding
  • 20% False negatives – provides inappropriate reassurance, potentially delaying timely treatment
• Model is not interpretable by a non-expert
• Results can be uncertain

Person ‘at risk’ for disease

Model

Predicted to have disease

Predicted to not have disease
Screening for breast cancer in 2018—what should we be doing today?

J.M. Seely MD* and T. Alhassan MD*

ABSTRACT

Although screening mammography has delivered many benefits since its introduction in Canada in 1988, questions about perceived harms warrant an up-to-date review. To help oncologists and physicians provide optimal patient recommendations, the literature was reviewed to find the latest guidelines for screening mammography, including benefits and perceived harms of overdiagnosis, false positives, false negatives, and technologic advances.

For women 40–74 years of age who actually participate in screening every 1–2 years, breast cancer mortality is reduced by 40%. With appropriate corrections, overdiagnosis accounts for 10% or fewer breast cancers. False positives occur in about 10% of screened women, 80% of which are resolved with additional imaging, and 10%, with breast biopsy. An important limitation of screening is the false negatives (15%–20%). The technologic advances of digital breast tomosynthesis, breast ultrasonography, and magnetic resonance imaging counter the false negatives of screening mammography, particularly in women with dense breast tissue.

Key Words  Breast cancer, screening mammography, digital breast tomosynthesis, overdiagnosis

Curr Oncol. 2018 Jun;25(S1):S115-S124

www.current-oncology.com
FDA advances landmark policy changes to modernize mammography services and improve their quality

Proposed rule would require breast density reporting, enhance the FDA’s ability to enforce mammography facilities’ compliance with standards

For Immediate Release
March 27, 2019
Amongst a target population of 1000 patients, 10% of the patients experience the outcome during the time-at-risk.
Without predictive modeling, EVERYONE has a 10% risk and outcomes are randomly distributed across the population.
A prediction model can rank all 1000 patients by their predicted probability of experiencing the outcome. The largest probability is $p=0.82$ for a patient. The second largest probability is $p=0.81$ for another patient. The 10th largest probability is $p=0.65$. The 100th largest probability is $p=0.42$. The 250th largest probability is $p=0.27$. The smallest probability is $p=0.0001$ for the last patient in the list.
Choosing a threshold on the predicted probability can operationalize the model results into a decision-making criteria.

If threshold set at $p=0.65$, 10 would be predicted to have the outcome and 990 would be predicted to not have the outcome...

Of the 10 predicted positive, 6 did have the outcome...

<table>
<thead>
<tr>
<th>True condition positive</th>
<th>True condition negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted positive</td>
<td></td>
</tr>
<tr>
<td>6 (TP)</td>
<td>4 (FP)</td>
</tr>
<tr>
<td>Predicted negative</td>
<td></td>
</tr>
<tr>
<td>94 (FN)</td>
<td>896 (TN)</td>
</tr>
</tbody>
</table>

**Precision = Positive predictive value**

\[
= \frac{6}{10} = 60\%
\]

**Negative predictive value**

\[
= \frac{896}{990} = 91\%
\]

**Sensitivity = Recall**

\[
= \frac{6}{100} = 6\%
\]

**Specificity**

\[
= \frac{896}{900} = 99.6\%
\]

**Accuracy**

\[
= \frac{6+896}{1000} = 90\%
\]

**F1 score**

\[
= \frac{2}{\left(\frac{1}{6\%}\right)+\left(\frac{1}{60\%}\right)} = 11\%
\]
Different thresholds offer different tradeoffs in operating characteristics

If threshold set at $p=0.27$, 250 would be predicted to have the outcome and 750 would be predicted to not have the outcome... Of the 250 predicted positive, 62 did have the outcome...

<table>
<thead>
<tr>
<th>Predicted positive</th>
<th>True condition positive</th>
<th>True condition negative</th>
<th>Precision = Positive predictive value = $\frac{62}{250} = 25%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>62</td>
<td>188</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predicted negative</td>
<td></td>
<td></td>
<td>Negative predictive value = $\frac{712}{750} = 95%$</td>
</tr>
<tr>
<td>38</td>
<td>712</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity = Recall = $\frac{62}{100} = 62%$</td>
<td>Specificity = $\frac{712}{900} = 79%$</td>
<td>Accuracy = $\frac{(62+712)}{1000} = 90%$ F1 score = $\frac{2}{(1/62%)+(1/25%)} = 35%$</td>
<td></td>
</tr>
</tbody>
</table>
What are the key inputs to a patient-level prediction study?

<table>
<thead>
<tr>
<th>Input parameter</th>
<th>Design choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target cohort (T)</td>
<td></td>
</tr>
<tr>
<td>Outcome cohort (O)</td>
<td></td>
</tr>
<tr>
<td>Time-at-risk</td>
<td></td>
</tr>
<tr>
<td>Model specification</td>
<td></td>
</tr>
<tr>
<td>-which model(s)?</td>
<td></td>
</tr>
<tr>
<td>-which parameters?</td>
<td></td>
</tr>
<tr>
<td>-which covariates?</td>
<td></td>
</tr>
</tbody>
</table>

Research and Applications

Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data

Jenna M Reis,1 Martijn J Schuemie,1 Marc A Suchard,2 Patrick B Ryan,1 and Peter R Rijnbeek2

1Janssen Research and Development, Raritan, NJ, USA; 2Department of Biomathematics, UCLA School of Medicine, CA, USA, and 3Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands

Corresponding Author: Dr. Jenna M Reis, Janssen Research and Development, Raritan, New Jersey, USA; jreis@its.jnj.com

Received 30 May 2017; Revised 8 December 2017; Editorial Decision 23 February 2018; Accepted 15 March 2018
### Types of prediction problems in healthcare

<table>
<thead>
<tr>
<th>Type</th>
<th>Structure</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease onset and progression</td>
<td>Amongst patients who are newly diagnosed with <em>&lt;insert your favorite disease&gt;</em>, which patients will go on to have <em>&lt;another disease or related complication&gt;</em> within <em>&lt;time horizon from diagnosis&gt;</em>?</td>
<td>Among newly diagnosed AFib patients, which will go onto to have ischemic stroke in next 3 years?</td>
</tr>
<tr>
<td>Treatment choice</td>
<td>Amongst patients with <em>&lt;indicated disease&gt;</em> who are treated with either <em>&lt;treatment 1&gt;</em> or <em>&lt;treatment 2&gt;</em>, which patients were treated with <em>&lt;treatment 1&gt;</em> (on day 0)?</td>
<td>Among AFib patients who took either warfarin or dabigatran, which patients got warfarin? (as defined for propensity score model)</td>
</tr>
<tr>
<td>Treatment response</td>
<td>Amongst patients who are new users of <em>&lt;insert your favorite chronically-used drug&gt;</em> , which patients will <em>&lt;insert desired effect&gt;</em> in <em>&lt;time window&gt;</em>?</td>
<td>Which patients with T2DM who start on metformin stay on metformin after 3 years?</td>
</tr>
<tr>
<td>Treatment safety</td>
<td>Amongst patients who are new users of <em>&lt;insert your favorite drug&gt;</em> , which patients will experience <em>&lt;insert your favorite known adverse event from the drug profile&gt;</em> within <em>&lt;time horizon following exposure start&gt;</em>?</td>
<td>Among new users of warfarin, which patients will have GI bleed in 1 year?</td>
</tr>
<tr>
<td>Treatment adherence</td>
<td>Amongst patients who are new users of <em>&lt;insert your favorite chronically-used drug&gt;</em> , which patients will achieve <em>&lt;adherence metric threshold&gt;</em> at <em>&lt;time horizon&gt;</em>?</td>
<td>Which patients with T2DM who start on metformin achieve &gt;=80% proportion of days covered at 1 year?</td>
</tr>
</tbody>
</table>
The Journey from Data to Evidence

Erasmus MC Rotterdam The Netherlands

www.ohdsi-europe.org