The Evidence Base® In Focus

PicnicHealth

Resolving challenges in data completeness with machines and humans

Different quality measures to evaluate **real-world data (RWD)** have been proposed, but **data completeness** is consistently described as essential to RWD quality.

Greater completeness contributes to **higher quality**, **more longitudinal data**, improving the **accuracy**, **reliability** and potential **scope of research**.

The PicnicHealth Approach

PicnicHealth builds RWD by **working with consenting patients** and employing **human-in-the-loop** machine learning to tackle **three barriers to completeness**:

1. Poor and incompatible coverage across **multiple providers**
2. **Lack of critical data abstraction** from structured and unstructured portions of electronic medical records
3. **Inaccuracy** of natural language processing (NLP) when processing data at scale

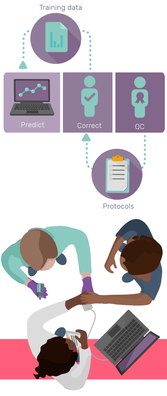
Human-in-the-Loop Machine Learning

What is Human-in-the-Loop (HIL)?

A **machine-learning** (**ML**) process combining machine and human intelligence. ML-generated predictions are reviewed by one human annotator, creating a feedback loop. Another trained annotator provides final revisions, improving prediction accuracy.

Why HIL?

- Ensures **high quality** data
- Captures **additional information** beyond the immediate context of the predictions
- Expert **human annotators** leverage the context to provide **critical metadata**



Summary of Analysis

- **4,646** — 4,646 patients with **Multiple Sclerosis (MS)**
- **350,000** — 350,000 inpatient and outpatient **visits**
- **62 years** of data
- **76 visits** — Average **76 visits** (s.d. 91) per patient
- **8.3 years** — Average of **8.3 yrs of data** (s.d. 6.6) per patient

3 Simulations to Assess Completeness or Missingness in a real-world data set

1. **What is missed when analyses are restricted to a single provider ?**
2. **What concepts are missed when looking only at structured data (e.g., medications lists and ICD-coded diagnoses)?**
3. **To what extent can ML prediction accuracy be improved by human intervention?**

Key Clinical Concepts in MS

1. **First disease-modifying therapy (DMT) taken by a patient**
2. **Most recent MS subtype record for each patient**

Precision and recall results for simulation studies

		Simulation 1	Simulation 2	Simulation 3	Full PicnicHealth dataset
	Provider completeness	One provider network	All providers	All providers	All providers
	Abstraction depth	Structured and unstructured data	Structured data only	Structured and unstructured data	Structured and unstructured data
	Structuring process	NLP + HIL	NLP + HIL	NLP Only	NLP + HIL
First DMT	Precision	80.5%	78.7%**	66.0%	100%
First DMT	Recall	74.5%	71.4%	71.0%	100%
MS Subtype	Precision	90.2%	0%	30.2%	100%
MS Subtype	Recall	71.6%	0%	52.8%	100%

Overall Learnings from Simulations

25% of predictions are missed when limited to a single provider, **unable to capture the whole picture of disease and treatment.**

Without unstructured data, no MS subtype information is available (0% precision and recall) and nearly 25% of patients would have had their first DMT regimen captured incorrectly.

30% of total predicted disease subtypes are correct when using only NLP and **47%** of patients had an inaccurate subtype recently listed.

Summary

Maximization of RWD completeness supports its use to generate real world evidence. This complex process requires a multifaceted approach. This analysis focused on electronic medical records but additional sources, such as healthcare claims, enrich RWD further. PicnicHealth is currently developing methods to integrate claims data into this process and improving RWD completeness will contribute to the richness of this data to inform healthcare research and practice.

This infographic was created as part of The Evidence Base's In Focus on data completeness, in association with PicnicHealth.

The Evidence Base® In Focus

PicnicHealth