Why Patient Patient Matching for PDMPs?

Carmen Smiley, IT Specialist, ONC
Why is patient matching important?

Without consistently and correctly matching individuals to their data to enable a complete record of a one’s health history, patients will continue to suffer consequences due to preventable medical errors and patient misidentification.

- Other reasons?
The Benefits of Accurate Patient Matching

• Accurately matching individuals or patients to their health information has a number of benefits:
  » improved patient outcomes
  » improved patient safety
  » greater efficiency
  » improved fraud detection
  » promoting data integrity
  » reduced inappropriate data exposure
(Some) Challenges to Patient Matching

- It is not possible to achieve perfection in patient matching
- Inaccuracy is not just a technology problem – there is a significant human component
- Poor data quality (both accuracy and completeness) significantly inhibits the ability to accurately match patients
- Limited published research, reported metrics, or best practices
- There is no one-size-fits-all solution
- Challenges increase as data gets further removed from the source, and when more sources of data are involved
- Universal identifiers for patients/individuals may be helpful, but are not a panacea
Lack of understanding and concurrence

• Makes communication, reporting, and cross-system or cross-organizational comparisons difficult or impossible

• Impedes a full and accurate assessment of the extent of the problem

• And therefore, the ability to devise sufficient solutions

• Prohibits evidence-based decision making

• Including decisions on which MPI or other technical solution to employ

• Limits research on complementary matching methods

• Inhibits progress and innovation
HIMSS developed a set of key performance indicators (KPIs) that allow an organization to evaluate its patient matching processes and technology and make continuous improvements.

- EMPI Database Activity Rate (EDAR)
- EMPI Database Duplicate Rate (EDDI)
- Duplicate Creation Rate (DCR)
- True Match Rate (TMR)
- False Positive Match Rate (FPMR)
- False Negative (Non-Match) Rate (FNR)
- Indeterminate Match Rate (IMR)
- Matching Accuracy Rate (MAR)
- Matching Error Rate (MER)
Precision and Recall

$$PPV = \text{Precision} = \frac{TP}{TP + FP}$$

The proportion of true matches that were found out of the total matches found. Precision is the number of correct results divided by the number of all returned results.

$$\text{Recall} = \frac{TP}{TP + FN}$$

How many from the matches found are real matches or the percent of all relevant documents that is returned by the search.

*The two measures are sometimes used together in the F-Score (or F1-score or f-measure) to provide a single measurement for a system.*
## Precision and Recall

<table>
<thead>
<tr>
<th>Algorithm result from matching</th>
<th>The records in reality belong to:</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Same person</td>
<td>True positive (TP)</td>
<td>False positive (FP)</td>
</tr>
<tr>
<td>Same person</td>
<td>Correct result</td>
<td>Type I Error</td>
<td></td>
</tr>
<tr>
<td>Different people</td>
<td>False negative (FN)</td>
<td>Type II Error</td>
<td></td>
</tr>
<tr>
<td>Correct result</td>
<td>Type II Error</td>
<td></td>
<td></td>
</tr>
<tr>
<td>True negative (TN)</td>
<td>Correct result</td>
<td>Type I Error</td>
<td></td>
</tr>
</tbody>
</table>

- **Precision**: The ratio of correct results among all results that were predicted as positive.
- **Recall**: The ratio of correct results among all actual positive cases.
F-Score

• a.k.a., F-1 score or F measure
• The harmonic mean between precision and recall

\[ F = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \]
## Type I and Type II Errors

<table>
<thead>
<tr>
<th>Algorithm result from matching</th>
<th>The records in reality belong to:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Same person</td>
</tr>
<tr>
<td></td>
<td>Different people</td>
</tr>
<tr>
<td>Same person</td>
<td>Correct result</td>
</tr>
<tr>
<td></td>
<td><strong>True positive (TP)</strong></td>
</tr>
<tr>
<td>Different people</td>
<td><strong>False positive (FP)</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Type I Error</strong></td>
</tr>
<tr>
<td></td>
<td><strong>False negative (FN)</strong></td>
</tr>
<tr>
<td></td>
<td>Correct result</td>
</tr>
<tr>
<td></td>
<td><strong>True negative (TN)</strong></td>
</tr>
</tbody>
</table>

Clinical information assigned to the wrong patient by creating an overlay if match meets or exceeds low threshold.

Clinical information not linked, patient has duplicate records.
### Sensitivity, Specificity, PPV & NPV

<table>
<thead>
<tr>
<th>Algorithm result from matching</th>
<th>The records in reality belong to:</th>
<th>PPV = ( \frac{TP}{TP + FP} )</th>
<th>NPV = ( \frac{TN}{TN + FN} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same person</td>
<td>Same person</td>
<td>Correct result (TP)</td>
<td>Type I Error (FP)</td>
</tr>
<tr>
<td>Different people</td>
<td>Different people</td>
<td>Type II Error (FN)</td>
<td>Correct result (TN)</td>
</tr>
<tr>
<td>Sensitivity = ( \frac{TP}{TP + FN} )</td>
<td>Recall</td>
<td>Specificity = ( \frac{TN}{TN + FP} )</td>
<td></td>
</tr>
</tbody>
</table>