Validation Rules as a First Step for Data Quality: Pharmacovigilance Application in Portugal

Bárbara COSTAa,b, Bruno FIGUEIREDOb,1 and Vera PINHEIROa,b,c
a CINTESIS, Faculty of Medicine, University of Porto, Portugal
b MEDCIDS, Faculty of Medicine, University of Porto, Portugal
c Public Health Unit, Matosinhos Local Health Unit, Matosinhos, Portugal

Abstract: Tracking and reporting Adverse Drug Reactions (ADRs) is crucial for patient safety. This work aims to improve the data quality of the SIRAI application in Portugal by developing data validation rules and a scoring system for each record and the overall dataset. The goal is to enhance the effectiveness of the SIRAI application in monitoring adverse drug reactions.

Keywords. Data Quality, Pharmacovigilance, Adverse Drug Reaction.

1. Introduction
Adverse drug reactions (ADRs) are potentially preventable responses to medication caused by factors like incorrect administration and communication breakdowns in the healthcare system. [1]. The SIRAI application (Information System for Adverse Reactions and Incidents), at the Matosinhos Local Health Unit (MLHU) is part of the ADR surveillance system in Portugal, but it currently lacks a structured data quality system, essential to ensure accurate reporting of adverse drug reactions [2]. This involves data editing, which detects and corrects errors or inconsistencies in data, usually through data checking validation rules [3]. Thus, the main goal of this work is to contribute to the development of a data quality system for the SIRAI application by establishing checking rules for data validation.

2. Methods
We performed (1) an analysis of the application’s guidelines/training materials/interface, the variable list, and the dataset, then (2) we developed checking rules for data validation and (3) we defined a scoring system. It was analyzed a sample of SIRAI dataset, with ADRs reported by MLHU between 2018 and 2022. The checking rules were grouped according to EUROSTAT guidelines [3]: Structural Validation rules, to ensure the technical integrity of the data file, and Content Validation rules, for logical/statistical consistency validation. Finally, we defined a scoring system in accordance with the BANFF methodology [3].

1 Corresponding Author: Bruno FIGUEIREDO, E-mail: brunofigueiredo@msn.com.
3. Results

From the analysis of the dataset we found that out of the 5250 analyzed entries 35% of non-mandatory variables were missing. Additionally, we identified reporting errors/inconsistencies, which were addressed in the next phase. Based on the dataset structure and issues identified, we established data validation rules - Table 1.

### Table 1. Validation rules for the SIRAI Application (some examples)

<table>
<thead>
<tr>
<th>Rule name</th>
<th>Condition</th>
<th>Error message</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consistency in drug administration dates</td>
<td>Date.end &gt;= Date.beginning</td>
<td>The end of the drug administration must be after the beginning of the drug administration</td>
</tr>
<tr>
<td>Consistency in number of ADR-severity</td>
<td>N ADR = N Severity</td>
<td>For N adverse events recorded, there must be exactly the same N records for severity</td>
</tr>
<tr>
<td>Completeness of administration route record consistency</td>
<td>VIA_ADMIN AND DESIGNATION != N/A</td>
<td>Drug administration route cannot be missing</td>
</tr>
<tr>
<td>Consistency between drug reintroduction and repeated reaction</td>
<td>IF Reintroduction_same_medication=&quot;No&quot; THEN Similar_Reaction_Reintroduction=&quot;No&quot;</td>
<td>When the suspected drug is not reintroduced, there cannot be a reaction upon reintroduction</td>
</tr>
<tr>
<td>Range check for Height</td>
<td>30 cm &lt;= Height &lt;= 250 cm</td>
<td>Height must be between 30 and 250 cm</td>
</tr>
<tr>
<td>Range check for Weight</td>
<td>1000 g &lt;= Weight &lt;= 600000 g OR 1 kg &lt;= Weight &lt;= 600 kgs</td>
<td>Weight must be between 1000g and 600000 g (or 1kg and 600 kgs)</td>
</tr>
</tbody>
</table>

ADR - Adverse Drug Reaction; N - Number

We then defined a scoring system. Each record is assigned a status code (PASS, MISS of FAIL) based on the validation rule. Then, an overall record status is derived. Using this method, we can score each recorded ADR and the overall dataset according to the defined validation rules.

4. Discussion and Conclusions

Our goal was to develop a data quality system for the SIRAI application by establishing checking rules for data validation. Data editing systems are essential for the overall quality of any pharmacovigilance system. The defined validation rules and scoring system will be accessible by both managers and reporters of the application, for analyzing data quality issues. In conclusion, implementing a data quality system can improve SIRAI’s effectiveness in monitoring ADRs. Future work includes studying the impact of the enhanced data quality system on ADR reporting accuracy and completeness.

References