RWD175

A Data Quality Framework to Assess Healthcare Data in Saudi Arabia: An Automated Approach

Rami Al-Jafar, PhD[1,2], Ali Abuharb, MSc[1], Abdullah Al-Zeer, PhD[1], Fahad Alsaawi, MSc[2]

[1] Lean Business Services, Riyadh, Saudi Arabia; [2]Imperial College London, London, UK

INTRODUCTION

Despite the importance of high-quality healthcare data, to our knowledge, there is no scientific contribution to assess data quality in healthcare on a national level in Saudi Arabia. Moreover, there is a need for an automated data quality engine that assess data quality on regular basis.

OBJECTIVES

This study aims to assess the quality of healthcare data among Ministry of Health facilities in Saudi Arabia. Toward this end, a data quality engine will be built and validated, and a business rules list for data quality will be established.

METHODS

The study used a probability sampling technique to have a sample representing a large dataset gathered from three main health information systems in Saudi Arabia. Within this sample, we checked the quality of 25 data elements for outpatient data and 22 for inpatient data. The process consisted of three phases:

(i) columns identification and data cleansing phase
(ii) measurements and assessment phase
(iii) analysis and improvement phase.

The measurements and assessment phase was based on five dimensions: uniqueness (number of duplicates of a patient ID number), completeness (ratio between the completed values to the total number of values in the dataset), validity (if it conforms to the syntax [format, type, range] of its definition in Saudi Health Data Dictionary), consistency (the ratio of values matching the values of the source of truth) and timeliness (the degree to which data is updated from the specific point on time).

RESULT

The study data sample comprises 818,265,291 encounters . Our proposed approach has detected Most column types with high confidence (80%). Within the measurements and assessment phase, the completeness, consistency, validity, timeliness and uniqueness percentages were 79.4, 85.2, 78.8, 85.1 and 74.9, respectively (Figure 1)

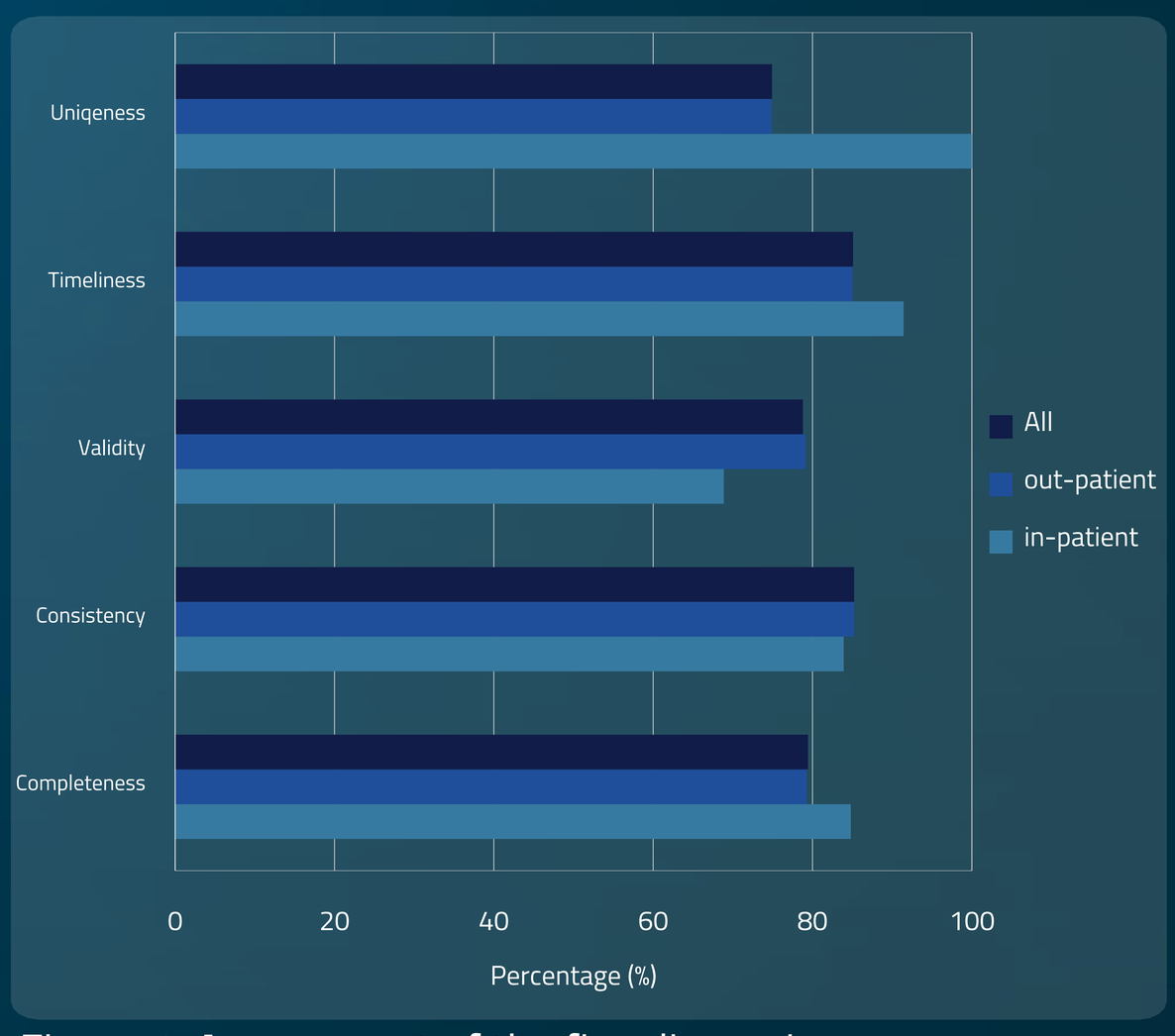

Figure 1: Assessment of the five dimensions.

DISCUSSION

The proposed data quality framework indicated that health data in Saudi Arabia could be improved and highlighted the areas to be targeted. Our automated approach can be applied to real-world data in other health systems to enhance the data quality assessment.

CONTACT INFOMRATION