

# Mapping Laboratory Data to LOINC Codes Using Distributions of Values

Laura A. Evans<sup>1</sup>, Matvey B. Palchuk, MD, MS<sup>1,2</sup>

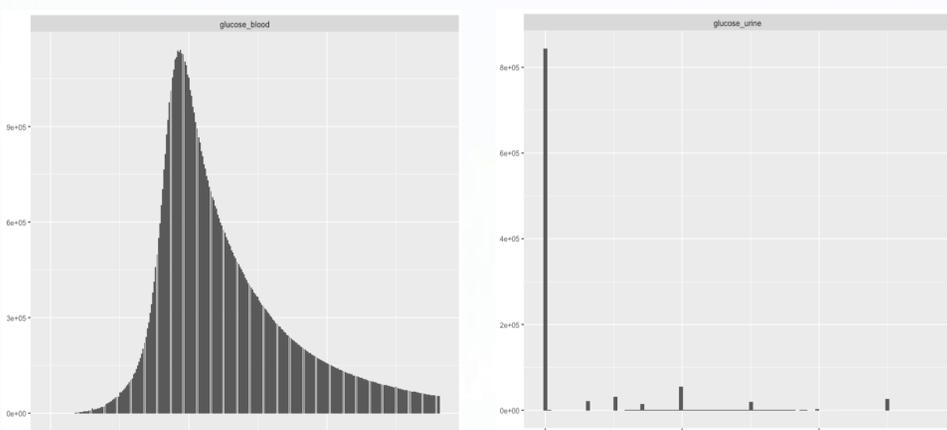
<sup>1</sup>TriNetX, Cambridge, MA; <sup>2</sup>Harvard Medical School, Boston, MA

## INTRODUCTION

- Semantic mapping from local lab codes to LOINC is necessary to achieve meaningful interoperability
- Typically such mapping relies primarily on description of a lab test
- We propose considering the shape of a value distribution curve (see Figure 1) to aid the mapping process

## METHODS

- Note that lab value distribution curves have distinct shapes
- Shapes can be quantified using the four “moments” or quantitative measures of a shape: mean, standard deviation, skew and kurtosis (see Figure 2)
- Quantified shapes can be matched to gold standard (see Figure 3)



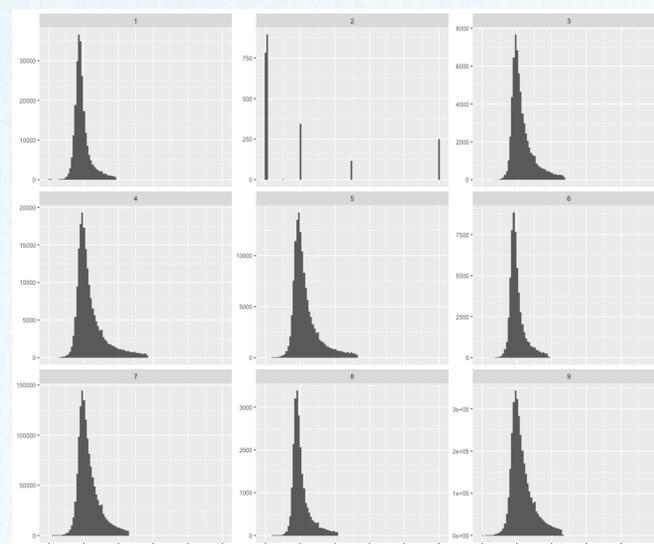
**Figure 3.** Gold standard distributions for glucose in blood / serum / plasma (left) and urine (right).

**Table 1.** Performance of the matching algorithm.

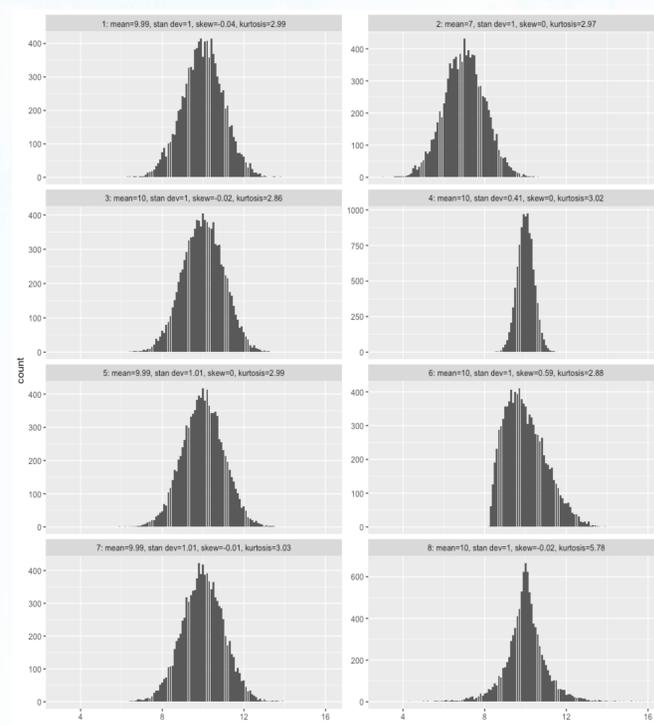
Lab	Precision	Recall	F-Score
<b>Glucose</b>			
Glucose in Blood	0.529	1	0.692
Glucose in CSF	0.125	1	0.222
Glucose in Urine	1	1	1
<b>Sodium</b>			
Sodium in Urine	1	1	1
Sodium in Blood	0.4	0.286	0.333
<b>Creatinine</b>			
Creatinine in Urine	0.227	0.833	0.357
Creatinine in Blood	0.429	0.375	0.4
<b>All</b>			
Creatinine in Urine	0.6	0.6	0.6
Sodium in Urine	0.333	1	0.5
Creatinine in Blood	0.833	0.625	0.714
Glucose in Blood	0.533	1	0.696
Sodium in Blood	0.5	0.143	0.222
Glucose in CSF	0	0	0

## CONCLUSIONS

- Examining error cases suggest that modest manual clean-up based on descriptions will have a considerable impact on performance
- Matching on shape isn't perfect, but coupled with traditional semantic mapping techniques it can provide valuable assistance to a human reviewer



**Figure 1.** Sample of lab value distributions where description contains “glucose.” X-axis is test result and y – number of tests.



**Figure 2.** Illustration of four “moments” or quantitative measures of a shape. In each row, the graph on the right has one of the “moments” altered. Top to bottom, they are: mean, standard deviation, skew & kurtosis.

## RESULTS

- Using moments that describe the shape of the of lab result value distribution, clustered all unknown labs using k-means clustering
- Labs in the same cluster as the gold standard were assumed to be “mapped”
- We performed four experiments: (1) “glucose” in description, (2) “sodium,” (3) “creatinine,” and (4) match irrespective of the description (see Table 1)
- Algorithm demonstrates reasonably good match performance when taking test description into account