Achieving SGD 3 goals with genetics, internet technology and data driven improvement

Liesbeth Siderius, Elmas Citak, Rob Braamburg, Anton Heijs, Sanath Lamabadusuriya

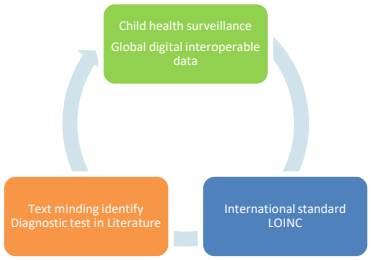
Introduction:

Some 93 million children under 15 years of age live with a moderate or severe disability (WHO, 2015). Disabilities, such as hearing deficit, cerebral palsy, developmental disabilities, turn out to have genetic causes.

In 2019 the WHO launched a classification of digital health interventions v1.01: Shared language to describe the uses of digital technology for health. In lower and middle-income countries the use of electronic data in medical care is still on the verge of development.

Global awareness on the advantages of the use of available genetic tests and eHealth should be improved.

Our concept towards digital global child health



Methods

Child health handbooks (WHO) containing essential information about pregnancy, birth, growth, development, physical examination and vaccination, to promote and maintain health. LOINC, the *Logical Observation Identifiers Names and Codes*, is an a universal standard and an electronic database for clinical care and management. We identified LOINC standards applicable to child health care as well as to medical guidelines on Shwachman Diamond Syndrome (SDS) and Thalassaemia. We wondered how text mining could to discover new information from text data, finding patterns across data-sets and/or separating signal from noise for processing PubMed document sets.

Results

A subset of LOINC standards was identified to enrich Child health handbooks with features of (rare) disabling conditions presenting in the first years of life (Table).
The query "(Shwachman) OR (Shwachman syndrome) OR (Shwachman disease)" resulted in 231 papers from 1975 to 2020 and "(Thalassemia) OR (Thalassemia syndrome) OR (Thalassemia disease)" 3385 papers from 1951 to 2020 in a PUBMED search (Figure 1).Text mining to discover new information from text data, finding patterns across data-sets and/or separating signal from noise was applied on publications based in LOINC gene codes to fasten the process of finding relevant literature. Spatial placement illustrated arrange the documents in such a way that the distance relations (similarities) are preserved since these distances are calculated in a high dimensional space and projected onto a two dimensional plane to generate a "landscape view" (Figure 2). On Thalassaemia papers processing steps to use multi-class classification were performed, which means that for each topic cluster in the document set a classifier was made. After training the classifiers they were run on all documents which provides a multiple probability scores for each paper. Using classification techniques, one can create high-performance text classifiers that can sort through the PubMed documents in seconds rather than days and potentially implement published data into (digital) practice.

Conclusion

The application of disease specific code-sets ensures the harmonization of data and the possibility of data exchange. As demonstrated with the LOINC, integration of interoperable data into child health record, including children with chronic illness and disabilities, can accelerate the use of new diagnostics into practice. Adjustment of data services and management is necessary to provide efficient novel digital solutions to support the SDG3.

Tabel: LOINC in Child Health data; SDS and Alpha Thalassaemia

	Clinic	Feature in LOINC	Diagnosis	Genes in LOINC
SDS	Delayed weight gain	LP36298-5 Failure to thrive	Pancreas insufficiency	41764-2 SBDS gene targeted mutation analysis in Blood or Tissue by Molecular genetics method
SDS	Fatty stool	16142-2 : Fat [Mass/time] in 24 hour Stool	Pancreas insufficiency	
SDS	Recurrent infections	751-8 Neutrophils [#/volume] in Blood by Automated count	Neutropenia	
Thal	Pale	718-7 Hemoglobine in blood	Anaemia	55234-9Alpha thalassaemia gene panel - Blood by Molecular genetics method
Thal	Screening Newborns	46740-7 Hemoglobine disorders newborn screen interpretation	Hemoglobino pathy	

Figure 2 A subset of the Alpha Thalassaemia papers after filtering on LOINC terms belonging to LOINC code 55234-9.

Figure 1 Distribution of published papers for Shwachman disease and Thalassemia disease from PubMed over time.