OJPHI

ISDS 2014 Conference Abstracts

ISDS
INTERNATIONAL SOCIETY FOR DISEASE SURVEILLANCE

Towards a Framework for Data Quality Properties of Indicators used in Surveillance

Ian Painter*[1], Lauren Carroll[1], David Buckeridge[2] and Neil Abernethy[1]

[1]University of Washington, Seattle, WA, USA; [2]McGill University, Montreal, QC, Canada

Introduction

Effective use of data for disease surveillance depends critically on the ability to trust and quantify the quality of source data. The Scalable Data Integration for Disease Surveillance project is developing tools to integrate and present surveillance data from multiple sources, with an initial focus on malaria. Consideration of data quality is particularly important when integrating data from diverse clinical, population-based, and other sources. Several global initiatives to reduce the burden of malaria (Presidents Malaria Initiative, Roll Back Malaria Initiative and The Global Fund to Fight AIDS, Tuberculosis and Malaria[1]) have published lists of recommended indicators. Values for these indicators can be obtained from different data sources, with each source having different data quality properties as a consequence of the type of data collected and the method used to collect the data. Our goal is to develop a framework for organizing the data quality (DQ) properties of indicators used for disease surveillance in this setting.

Methods

We examined selected malaria indicators for the country of Uganda calculated from four sources: Uganda Health Management Information System (HMIS); Uganda Malaria Surveillance Project (UMSP) Sentinel Site Surveillance data, the Uganda 2010 Demographic Health Survey (DHS) and Uganda Indoor Residual Spraying (IRS) program data. Because different malaria indicators can be calculated from varied data sources, we organized the DQ properties according to the provenance of the data sources used to calculate each indicator. Using a hierarchical system, we grouped DQ properties at different levels of the process used to obtain an indicator: properties common to (or inherited from) the system generating the data, those inherited from the data source, those arising from data processing steps, those inherited from data fields used to calculate an indicator, and those specific to the calculation method.

Results

For any indicator, meta-data on the provenance of that indicator can be used to retrieve data quality properties from the appropriate level of the hierarchy. For example, the indicator "*Malaria test positivity rate*" calculated for inpatient data from the UMSP data source is calculated from the ratio of two other indicators: "*Any positive lab test for Malaria*" and "*Number of tested cases*". Each of these indicators is in turn calculated from multiple specific data fields as an aggregate of case summaries from the 6 UMSP sentinel sites. The data quality properties of the "*Malaria test positivity rate"* then consist of DQ properties specific to:

1. calculation of the indicator,
2. calculation of the two component indicators,
3. each field used in the calculations,
4. processing of the data used in the calculations,
5. sites providing the source data, and
6. the UMSP data source.

Examples of DQ properties at each level include standard errors for the ratio indicator and component indicators (1, 2), missing data rates for the each field (3), procedure for handling incomplete data forms (4), site specific sensitivity and specificity of microscopy detection of malaria, as measured at the start of the sentinel site program (5), and details on the training and quality assurance used in the program (6).

Conclusions

Our process captured meta-data elements relevant to provenance beyond those typically considered as DQ properties. Using a broad definition of DQ (e.g. "the totality of features and characteristics of an entity that bears on its ability to satisfy stated and implied needs" - ISO 849201986), we consider the meta-data elements relevant to provenance to be important DQ properties. For example, the DQ measures of source data elements might indicate that the numerator of an indicator is overestimated, while the denominator is underestimated. By propagating this knowledge to the calculated indicator, we can determine that there is a qualitative risk of overestimation of the given indicator.

Decision makers utilizing surveillance data need to be able to rely on the quality of data, to inspect that quality, and when possible, to quantify the quality. By developing a reusable framework for data quality and provenance meta-data, we hope to enable diverse decision makers to consistently and confidently interpret available surveillance data, indicators, and the analyses based on them.

Keywords

Data quality; Biosurveillance; Global health; Secondary data; Malaria

Acknowledgments

The SDIDS project is supported by the Bill and Melinda Gates Foundation.

References

Reithinger, R. (2014). Global malaria efforts. *Trans Royal Soc Trop Med & Hyg*,*108*, 247-248

*Ian Painter
E-mail: ipainter@uw.edu