TECHNICAL BRIEF

Modeling & Analytics

What is at stake ?

The modeling & analytics service is,

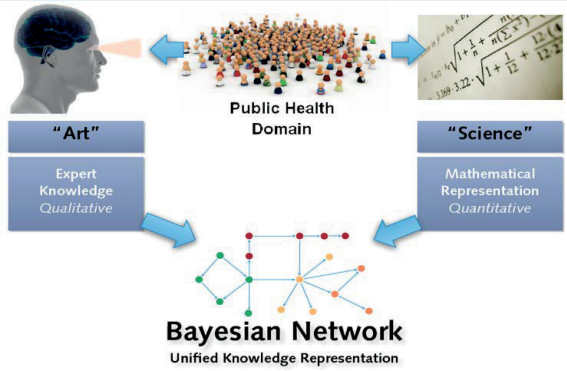
• crucial to improve monitoring and accountability for GAVI programs.
• enabling to better understand drivers of intervention vaccination coverage change and strengthening country analytical capacity

With the modeling approach, you can:

Take into account all the variables of a vaccination programme, including the six WHO's Health Systems Building Blocks, and more specifically the barriers to vaccination, gender and marginalized populations
The service aims at setting up a mathematical model, thanks to state-of-the-art Bayesian Network (BN) software which owns powerful statistical functionalities.

Bayesian Network

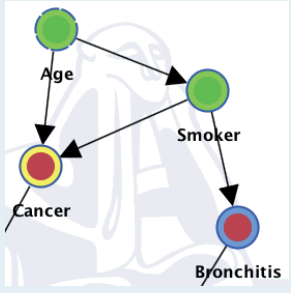
is a powerful approach for deep understanding of complex and high-dimensional problem domains, such as vaccination programs in the context of health system strengthening

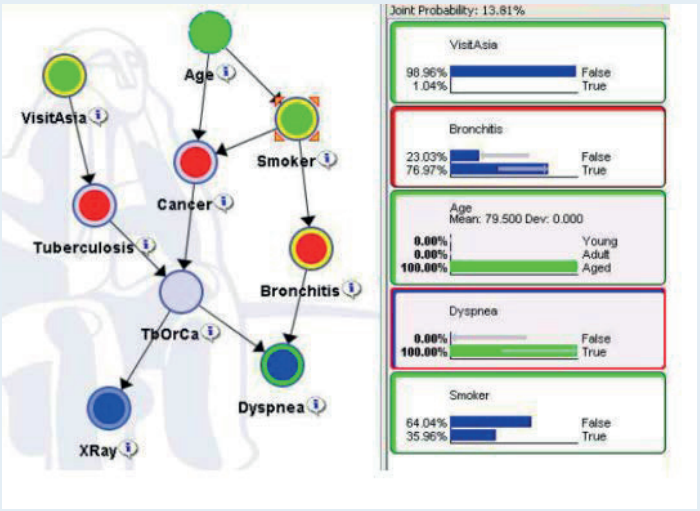


Bayesian networks have the ability of capturing both qualitative knowledge (through their network structure), and quantitative knowledge (through their parameters)

• Nodes represent the variables of the domain (e.g. the temperature of a device, a feature of an object, the occurrence of an event, the age of a patient)

• Links represent statistical (informational) or dependencies among the variable The dependencies are quantified by conditional probabilities for each node given its parents in the network.





The software will enable (i) to develop, communicate with and use readable illustrated model, (ii) to automatically find unknown relations (data mining) in the available data. Finally, it will be possible to take advantage of an original toolbox by using Bayesian network models in an interactive mode, testing scenarios, learning Public Health policies, performance of EPI programs.
*http://www.bayesia.com/BayesiaLab is a powerful desktop application based on the paradigm of Bayesian networks.

What does the term Analytic mean ?

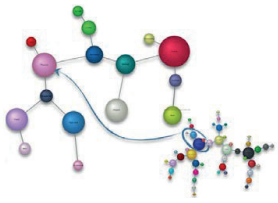
Data analytics per se imply letting powerful algorithms and computational tools dive into the dataset.This intensive mining of data, known as "data mining," can shed light on processes and interactions in the data that would not have appeared otherwise
Beyond correlations, analysing large quantities of data can help unveil facts—i.e. broadly recurring behaviours and patterns, which are instrumental for explaining immunisation coverage …

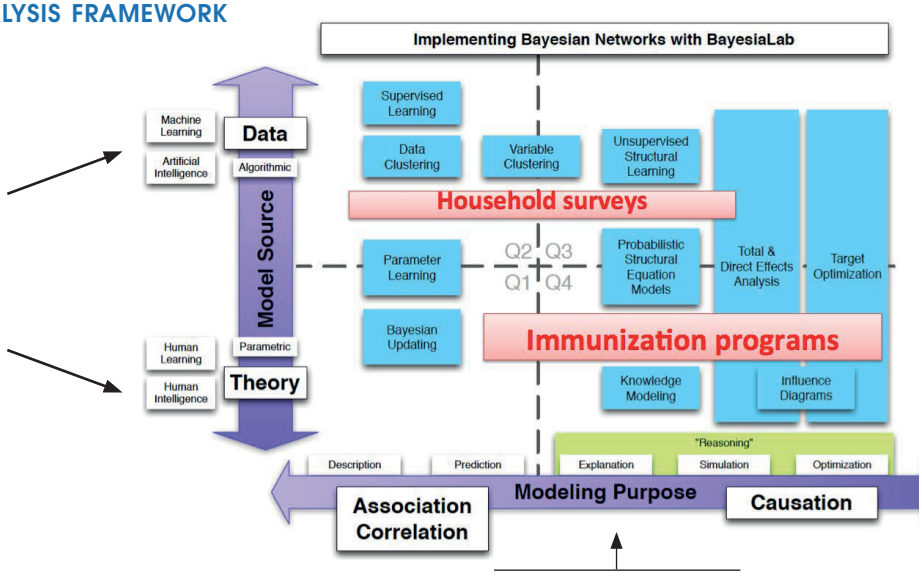
Bayesian Networks allows
a comprehensive data analysis framework

Bayesian Networks: Theory and Data

Bayesian networks can be machine-learned from Data
Specific data mining algorithm is used to deal with surveys…

…or they can be built from human knowledge, i.e. from Theory.
In the present case, it is possible to benefit from the Theory of change as laid down in the logical framework so to build up the model





Bayesian network models can cover the entire range from Association/Correlation to Causation. In practice, this means that we can add causal assumptions (Theory of change) to an existing-non-causal network and, thus, create a causal Bayesian network. It is of particular importance when we try to simulate EPI program aiming at appraising immunisation coverage or drop out children from immunization program for a particular district or region.

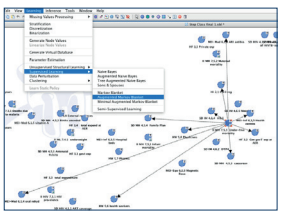
Probalistic Structural Equation Modeling (PSEM) within the framework of Immunisation program

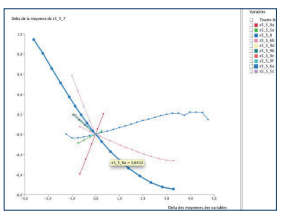
PSEM is a statistical technique for testing and estimating causal relations using a combination of statistical data and qualitative causal assumptions, it is much suitable to our complex context.

The following workflow will be systematically applied:

Unsupervised Learning, to discover the strongest relationships between the manifest variables.
Variable Clustering, to identify groups of variables thatare strongly connected.
Data Clustering for each Cluster of manifest variables
Connection of the target variable to the factors
Unsupervised Learning for discovering the path between the latent variables and their relationships **with the Target Node**.
Nota: The first targeted variable will be immunisation coverage

Example of Bayesia Lab's functionalities:





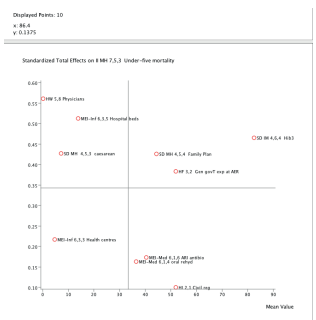
Supervised learning

BayesiaLab offers several Supervised Learning algorithms to search for the Bayesian network that best predicts the target variable while also taking into account the complexity of the resulting network

Total effects

The Total Effects axis shows the change of the mean value of the Target Node, given the observation of a one-unit change in each of the factors. This value is what we commonly interpret as slope.

Example of results:



The x-axis represents the "importance" with regard to the factors. The y-axis shows the Total Effect of each factor with regard to under 5-mortality rate:

RESULTS:
Physicians mark the top value on the y-axis of the plot, which means it is the most important contributing factor to reduce under 5 mortality.
Caesarean rate comes in second place.
Hospital beds…

Data source
WHO's Atlas of African Health Statistics series