Bayesian Network Models to combine and update models using incomplete data

Author: Tanja Krone[1], Xavier Sa Castro Pinho[1], Susan Dekkers [1], Eugene van Someren [1]

[1] Unit Healthy Living and Work, Netherlands Organization for Applied Scientific research (TNO), Leiden, the Netherlands

Safe-and-Sustainable-by-Design is a new approach where companies adjust the design of new innovative products to avoid safety and sustainability concerns with their products later on. The main aim of the HARMLESS Decision Support System (DSS) is to support companies in this, by providing insight in the current status with respect to the various safety, benefit and sustainability aspects. An important element in the DSS is a Bayesian Network that predicts occupational exposure of workers to nanomaterials. This network combines several mechanistic models and updated them with available data.

Bayesian Networks consist of nodes (variables) and edges (relations) that depict the (inter)dependencies of the variables in a system where each node has a distribution of possible values or states. An advantage of the Bayesian Network over the mechanistic models is the ability to quantify uncertainty and allow for missing data when predicting the exposure. Recreating mechanistic models means adjusting the point estimates to distributions, quantifying uncertainty in the parameters and the outcome variables. The variables are lognormally distributed, and the underlying mechanistic models are multiplicative, thus log-transforming the data allows the Bayesian network to work with normal distributions and to add the distributions instead of multiplying them.

Bayesian networks are a versatile way of combining related models and accompanying data and can be used to model complex mechanisms and make predictions based on these models. The models can work with incomplete and partly overlapping datasets, such as often found in fields where several distinct but related models are used to predict states or risks, e.g., exposure modelling, cardiovascular or diabetes research. Furthermore, the model can be updated over time by adding newly collected data as evidence.

Keywords: Bayesian networks, Exposure studies, Missing data, Model updating, Risk analyses

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 953183 (HARMLESS).