Wisdom of Stroke Experts translated into Directed Acyclic Graphs (WISE-DAG): Developing a user-friendly interface to present aggregated knowledge

What is the motivation?

In 1907, Francis Galton, the famous statistician and cousin of Darwin, published an article on Nature entitled “Vox Populi”. In this article, Galton shows, using empirical data from a weight-judging game, that the median guess of the game participants, who were a heterogeneous crowd, is very close to the real weight that the participants were supposed to guess. The author, who held anti-democratic convictions, is surprised by his own results, and concludes that probably the “democratic judgment” deserves more trust.

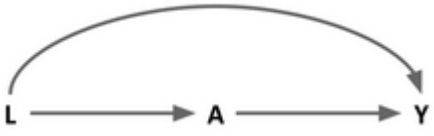
Almost one hundred years later, in 2005, James Surowiecki published a book called “The Wisdom of Crowds” about human collective intelligence, openly inspired by Galton’s surprising results. The basic idea of this concept is that combining several different independent judgements, each developed based on personal knowledge and experience, provides a good estimate of the truth because the aggregation “cancels out” the “noise” in each single personal judgment.

Stroke is the second-leading cause of death and the third-leading cause of death and disability combined worldwide. Therefore, research in this field is urgently needed to advance public health. However, much of stroke research is observational and is threatened by a lack of causal inference methods. That is, researchers often fail to properly incorporate knowledge about the causal processes related to stroke into their data analysis. This knowledge is needed to determine which data are needed when setting up a study or which variables should be considered in the analysis phase to prevent biases.

By applying the principle of the Wisdom of the Crowd, we aim to summarize independent, decentralized opinions and knowledge about stroke-related processes from a crowd of experts in the topic from various adjacent research disciplines.

What are directed acyclic graphs?

In order to collect and synthesize information about the causal knowledge from the different experts, we will use directed acyclic graphs (DAGs):



DAGs are compact, intuitive, visual tools that depict causal relationships between variables (here L represents a common cause of A and Y, a so-called “confounder”) in the form of assumptions about the data generation process. DAGs encode mathematical relationships among the variables and therefore are useful to inform study design and analytic methods to answer causal questions.

Progress thus far: WISE-DAG Workshops

The WISE-DAG (Wisdom of Stroke Experts translated into Directed Acyclic Graphs) project was originally conceptualized at the Center for Stroke Research Berlin at the Charité in 2020.

In 2021, a series of workshops led by Dr. Jessica L. Rohmann from the Charité - Universitätsmedizin Berlin, in which basic concepts of causality and an introduction to DAGs were presented to clinicians and researchers willing to participate in the project. After the workshop, clinicians and researchers were asked to independently draw their DAG about stroke-related processes using the Causalify webapp (https://causalify.github.io). After the workshop, the participants submitted their own DAGs. These stroke expert-generated DAGs will serve as the data source for this spin-off project.

Aim of your Master project

Using the DAGs submitted by the WISE-DAG participants, you will help create a consensus DAG that can be queried by stroke researchers to inform their study designs and data analyses in the future. You will develop an algorithm to aggregate the collected expert knowledge about stroke-related processes and develop an online application to make this knowledge readily available to stroke researchers.

This online application can guide future research in stroke worldwide and may also prove important for clinical research. Furthermore, a generalization of the online platforms may inform building consensus causal graphs across disciplines, and can be the basis for ongoing work in our lab on causal inference and tools for designing studies.

Details of your Master project

This Master project will contain 2 main work packages, which combine aspects of causal inference, database management and app development. Follow-up optional work may generalize the platform beyond stroke.

1. **Develop an algorithm to synthesize information about stroke-related processes from expert-generated DAGs.** You will need to generate a series of rules to aggregate a large amount of unstructured data about causal relationships. Stroke researchers generated distinct causal diagrams considering different time horizons and time scale granularity. They also considered a different granularity regarding the included variables in their DAGs (e.g., some researchers decided to break a causal process down into several steps while others condensed information, providing a “bigger picture” view). Given these differences, you will need to find creative solutions to challenges arising from concept hierarchies and time sequences. This work package will conclude with the creation of a database containing the aggregated knowledge in a harmonized way.

2. **Develop an app designed to support stroke researchers aiming to answer causal questions.** We envision an online application that can be used to query the database

generated from work package #1. The presentation of the data can be visual, for example, presenting a so-called consensus DAG, or tabular. In either case, you will be tasked with designing an interface that allows for a personalized user experience. By facilitating access to this consolidated knowledge database, the results of this work package will provide important insights for the study design phase as well as help researchers anticipate (and mitigate) potential biases that could arise during data analysis.

What should you bring to the project?

The project is open to students from all programs at HPI. To carry out this project successfully, you will need expertise in ***at least one*** of the following areas to contribute to the team:

1. Programming skills (e.g., mobile app development, web development, Python/R)
2. Database management
3. Causal inference methods including directed acyclic graphs (DAGs)
4. Clinical/experimental stroke research

Further references

- Galton, Francis. Vox Populi. Nature 75, 450–451 (1907).
- Surowiecki, James. 2005. The Wisdom of Crowds. Anchor.
- Hernán, Miguel A, and Robins, James M. 2020. "Causal Inference: What If." Boca Raton: Chapman & Hill/CRC 2020.

Contact

Get in touch with questions and ideas. Our offices are on the 1st floor of the Digital Health Center on Campus III, Building G2, Rudolf-Breitscheid-Str. 187, 14482 Potsdam, and you can always reach us by email or phone. See also our webpage (https://hpi.de/lippert/health-intervention-analytics-group.html) for more information.

Dr. Stefan Konigorski
Stefan.Konigorski@hpi.de

Dr. Marco Piccininni
Marco.Piccininni@hpi.de

Partner from Charité – Universitätsmedizin Berlin:

Dr. Jessica L. Rohmann
jessica.rohmann@charite.de