Standardized Tools for Constructing DAGs - Advancing Causal Inference and Risk Assessment in Pharmaceutical Studies

Laura Watson, MS and Sherrine Eid, MPH
SAS Institute, Inc., Cary, NC, USA

MSR123

Abstract

Causal inference and directed acyclic graphs (DAGs) are powerful tools in pharmaceutical research, facilitating analysis of critical questions and informing decision-making. By visualizing causal relationships, DAGs help identify confounders, mediators, and colliders, thus guiding the selection of covariates to control confounding and improve causal estimates. They also play a pivotal role in detecting and mitigating biases, including selection, measurement, and collider biases, thereby improving study design and analysis robustness. For comparative effectiveness research, DAGs clarify causal pathways, supporting the evaluation of treatment effects and real-world evidence for drug efficacy and safety. Additionally, they inform and help optimize study designs for both randomized controlled trials and observational studies. DAGs support risk assessments by analyzing real-world data for adverse drug reactions and long-term safety. They also support data integration and evidence synthesis by identifying compatible datasets and combining findings while preserving causal interpretability. Beyond analysis, DAGs serve as effective educational tools for communicating complex causal relationships and informing regulatory decisions, drug labeling, and healthcare policies.

Specifically, with bias detection, DAGs provide a clear framework for visualizing relationships among variables to assess potential sources of bias in causal inference. Bias typically arises from three sources: the data source (e.g., systematic inclusion/exclusion of subjects or stakeholder influence), study design, and data analysis methods. While SAS procedures like PROC ASSESSBIAS address biases from analysis methods, they overlook biases from data sources and study design. DAGs bridge this gap by identifying biases from all sources, offering a wider understanding of causal inference.

Standardized tools for constructing DAGs would ensure consistency in analysis, enabling reproducible and comparable results across studies. SAS Viya, SAS 9, R, and Python offer strengths and limitations for conducting causal inference and leveraging DAGs in pharmaceutical research. We explore these comparative strength and limitations throughout this work.

Results

- There is a variety and extensive variability of tools currently available
- Tools vary greatly in features and functional capabilities
- Most are reputable within their respective user base
- They vary in ease of use from low code/no code non-statistician focused to heavily reliant on statistical programming skills
- Most had limited integration capabilities
- Some tools have broad application reach in a variety of industries
- Some focus on academic research
- Only SAS is documented to meet industry standards for regulatory compliance upon installation.
- All others either do not meet industry and regulatory standards or are dependent on implementation
- The majority of the tools had commercial costs for licensing
- Limitations of the tools also varied from scalability, steep learning curve, lacking advanced statistical capabilities to requiring programming proficiency and expertise

Key Applications of Causality and DAGs

- Confounding Identification and Control
- Bias Detection and Avoidance
- Comparative Effectiveness Research
- Study Design Optimization
- Risk Assessment
- Data Integration and Evidence Synthesis
- Educational and Communication Tool
- Policy and Decision-Making

Bias Detection

- DAGs provide a clear framework for visualizing relationships among variables.
- Assess potential sources of bias in causal inference.

- Three main sources of bias:
  - Data source
  - Study design
  - Data analysis methods.

- Biases from analysis methods are commonly addressed.
- Biases from data sources and study design are commonly overlooked.
- DAGs can illustrate bias from all three sources.

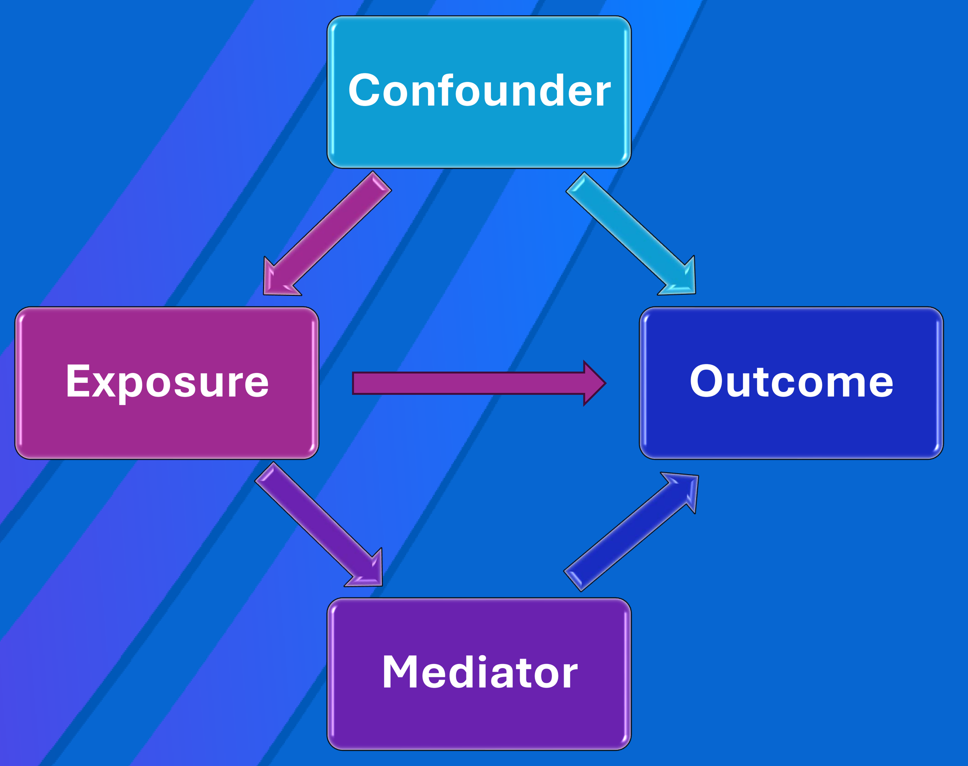
Conclusion

- Standardized tools for constructing DAGs would ensure consistency in analysis, enabling reproducible and comparable results across studies.
- SAS Viya, SAS 9, R, and Python offer strengths and limitations for conducting causal inference and leveraging DAGs in pharmaceutical research.
- There are a variety of tools currently available to researchers to develop DAGs.
- Depending on the desired result and application, researchers have an extensive list of tool options to assess bias and causality using DAGs.
- These tools allow researchers to conduct better research in a more mindful way and get medicines to patients faster

Table 1: Results of DAG Tool Comparison

Solution	Functional Capabilities	Reputation and Adoption	Ease of Use	Integration Capabilities	Application to Research	Collaboration and Sharing	Regulatory Compliance	Cost and Licensing	Limitations
Daggity.net	Graphical causal models (DAGs), model creation, simulation	Known in research, academic use	User-friendly web interface	Limited integration capabilities	Academic research, basic causal modeling	Model sharing online	N/A	Free to use	Limited scalability for large datasets
CausalFusion	Bayesian networks, model creation, learning, inference	Growing reputation, industry adoption	Requires Bayesian network expertise	API for integration with other tools	Healthcare, finance, engineering	Model sharing and collaboration	Depends on use case	Commercial with licensing fees	Complexity for beginners
CausalWizard	Structural equation modeling (SEM), causal inference	Popular in social sciences, psychology	User-friendly for SEM, non-statisticians	Limited API support	Social sciences, mediation analysis	Model sharing and collaboration	N/A	Commercial with licensing fees	May lack advanced statistical capabilities
TETRAD	Causal discovery algorithms, Bayesian networks	Established in academic research, wide usage	Requires statistical expertise	Limited integration capabilities	Genetics, social sciences, healthcare	Model and algorithm sharing	N/A	Free for academic use	Steep learning curve; algorithm understanding
STATA	Comprehensive statistical analysis, causal inference	Highly regarded in social sciences, economics	User-friendly with extensive documentation	Limited API support	Academic research, industry data analysis	Data and model sharing	Depends on data handling practices	Commercial with licensing fees	Additional tools for advanced causal modeling
SMILE by BayesFusion	Bayesian network modeling, model creation, learning	Highly regarded in probabilistic modeling, industry use	Intuitive graphical interface	API for integration with other tools	Decision support, risk assessment, predictive modeling	Model sharing and collaboration	Depends on data governance	Commercial with licensing fees	Additional modules may be required for advanced features
SAS	Analytics platform, statistical analysis, causal inference	Widely adopted in industry and research	Powerful with steep learning curve	Extensive integration capabilities	Statistical analysis, predictive modeling, machine learning, causal inference	Collaboration and sharing of models and results	Meets industry standards	Commercial with licensing fees	Cost may be prohibitive for smaller organizations
R	Open-source statistical programming language	Highly regarded in research and industry	Requires programming skills	Extensive integration with other tools and languages	Statistical modeling, causal inference	Strong community for sharing code and packages	Depends on implementation	Open-source; no licensing fees	Requires programming proficiency
Python	General-purpose programming language	Widely adopted in research and industry	Requires programming skills	Extensive integration with other tools and languages	Statistical modeling, machine learning, causal inference	Strong community for sharing code and packages	Depends on implementation	Open-source; no licensing fees	Requires programming expertise

Figure 1: What's in a DAG?



Let's Keep in Touch!