Triangulation of evidence based on a formal causal model: An example using a directed acyclic graph (DAG) in assessing benzene as a potential cause of chronic lymphocytic leukemia (CLL)

Anthony J. Russell[1], Daniel Lauer[2], William J. Thompson[3], Kenneth A. Mundt[3]
Office Locations: [1]Cincinnati, OH | [2]Boulder, CO | [3]Boston, MA

RATIONALE

- Benzene is causally associated with acute myeloid leukemia (AML) and myelodysplastic syndrome (MDS) at sufficient exposure levels (IARC 2018).
- However, inconsistent findings from observational epidemiological studies and methodological limitations present challenges for determining whether a causal association exists for benzene and other malignancies such as chronic lymphocytic leukemia (CLL).
- A formal structural causal model such as a directed acyclic graphic (DAG) provides a framework for visually depicting complex exposure-outcome pathways in which multiple potentially causal relationships (i.e., "confounding") simultaneously may be considered– and whether the epidemiological evidence aligns with the causal model.

APPROACH

- A quality-based systematic review of benzene and CLL was conducted in parallel with the development of a formal causal DAG.
- A formal causal DAG or "I-DAG" was initially generated based on a robust evaluation of the key risk factors and confounding variables underlying the association between benzene and AML and CLL, respectively. Subsequently, toxicological and mechanistic streams of evidence were integrated with human data to inform the plausibility of various associations observed in the DAG. Arcs in DAGs are assumed to represent true causal relationships (the "faithfulness assumption") but are challenged when associations are not biologically plausible or contradicted by toxicological and epidemiological evidence, allowing the causal model to be refined.
- As such, incorporating causal DAGs within systematic review frameworks provides a complementary approach for triangulating findings from human, animal, and in vitro data streams.

RESULTS

Benzene and AML

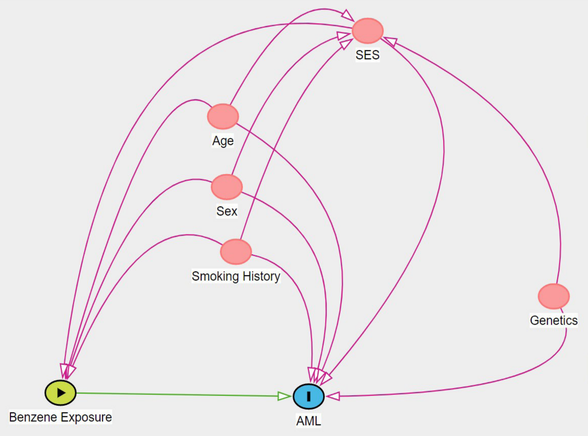

Figure 1. A DAG depicting the relationship between benzene exposure and potential causal paths of AML.

Established risk factors for AML include sex, smoking, age, a genetic predisposition for lymphoid malignancies and socio-economic status (SES). Further, occupational cohort data consistently demonstrates that risk of AML is elevated at high levels of benzene exposure.

- Benzene is the only industrial chemical classified by IARC as a Group 1 carcinogen, i.e., 'known' to cause AML (IARC 2018).
- The risk of AML increases following sustained occupational exposure to high concentrations of benzene, especially in the two to ten years following such exposure (Linet et al. 2018, North et al. 2021)
- The biological plausibility of benzene causing AML at sufficient levels is also supported by toxicological and mechanistic data. For example, benzene exposure has been associated with higher levels of chromosomal changes commonly observed in AML (Arenas et al. 2021)

Benzene and CLL

Chronic lymphocytic leukemia (CLL) is a mature B-cell lymphoid malignancy affecting the immune system (Swerdlow et al. 2016; Leukemia & Lymphoma Society 2019). Among seventeen studies assessing occupational benzene exposure and CLL mortality and seven studies assessing CLL incidence, only one studied showed a statistically significant association (deficit) (Satin et al. 1996). Most risk estimates were close to or below unity. In contrast to AML, the molecular etiology of CLL is still largely unknown (Arenas et al. 2021). A similar review to above yielded the following risk factors for CLL:

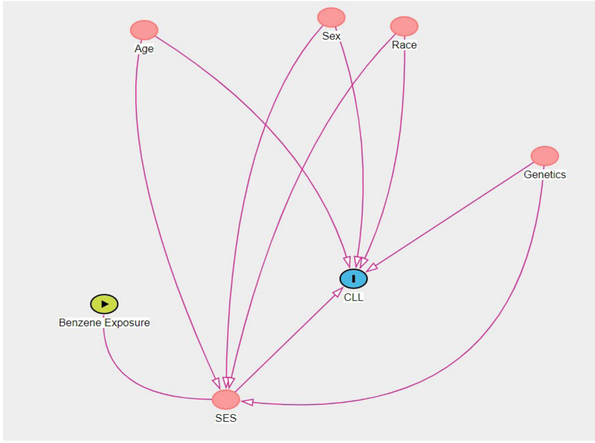

Figure 2. DAG depicting Risk factors for CLL and associations with Benzene

- **AGE:** The review clearly showed that CLL incidence increases with increasing age. More than 98% of cases occurring among those over the age of 45 and over 89% occurring among those 55 and over (SEER Cancer Stat Facts). The median age at diagnosis was 70 (SEER Cancer Stat Facts). Among white men, CLL risk increases approximately 130-fold from age 30 to age 65 and 300-fold from age 30 to age 85 or older: rates rise from about 0.2 per 100,000 at ages 30-34 to 26.9 per 100,000 at ages 65-69 and 60.4 per 100,000 among those 85 and over (Howlader et al. 2020).
- **RACE:** CLL incidence in the US is highest among white men and lowest among black women, with age-adjusted rates of 7.3 per 100,000 among white men, 4.8 per 100,000 among black men, 3.9 per 100,000 among white women and 2.4 per 100,000 among black women, respectively (Howlader et al. 2020)
- **GENETICS and CLL:** Family history of CLL or other lymphohematopoietic cancers consistently demonstrate an increased risk for CLL (Goldin et al. 2009; Goldin et al. 2010; Linet et al. 2007; Wang et al. 2007; Slager et al. 2014).

DISCUSSION

- A causal DAG provided a framework for evaluating the possible roles of environmental stressors, including benzene, on AML and CLL within the context of all possible risk factors (i.e. confounders and co-variates).
- Beginning with, and refining, a causal model may help prevent erroneously labeling observed associations "causal" that arise due to chance, bias and confounding.
- We encourage an adoption of this proposed framework (a biologically grounded model, combined with observational data, and statistical models) to assist further investigation in determining the causes of CLL.
- In conclusion: Benzene appears not to be a clear risk factor of CLL, a disease of elusive environmental causes; further investigation triangulating biological, toxicological, and mechanistic evidence will assist in EPA classification for Benzene in regards to a causal association with CLL.
- This methodology provides a unique opportunity to integrate epidemiological evidence within a triangulation framework of epidemiological, toxicological, and mechanistic evidence;
  - The DAG allows evaluation of the epidemiological evidence based on a biologically grounded causal model as well as observational data. This can then inform a statistical model (Rather than attempting to guess at a causal model solely from observational data and their statistical manipulation).
  - From our analysis and subsequent review of the current epidemiological literature risk factors for CLL remain age, sex, and race along with a genetic predisposition and SES.

Cardno ChemRisk now Stantec

*References are available upon request
**All figures generated on www.dagitty.net/dags.html

www.cardnochemrisk.com
www.stantec.com