National Institute of Environmental Health Sciences
Division of Translational Toxicology

An evaluation of the performance of survey sampling in systematic evidence mapping of cancer mechanistic evidence: Polycyclic aromatic hydrocarbons (PAHs) and key characteristics of carcinogens (KCCs) as a case study

Amy Wang[1], Grace A. Chappell[2], Arun Varghese[2], Brian Howard[3], Lena Schmidt[3], Raquel Silva[2], Whitney Arroyave[4], and Andrew A. Rooney[1]

[1]National Institute of Environmental Health Sciences (NIEHS)/Division of Translational Toxicology (DTT), Research Triangle Park (RTP), NC; [2]ICF Inc, Durham, NC; [3]Sciome LLC, RTP, NC; and [4]Inotiv, RTP, NC

Background

- For mechanistic evidence, an assessment might have anywhere from hundreds to hundreds-of-thousands potentially relevant references. When the number is large, screening all references to develop a systematic evidence map can be highly resource-intensive or even resource-prohibitive.
- We previously developed and conducted a survey sampling and statistical analysis approach to estimate the evidence distribution in each key characteristics of carcinogens (KCCs) to build an evidence map.
- Here we share the performance evaluation of survey sampling method and latest systematic evidence map.

Objectives

- To evaluate the performance of the survey sampling method by comparing the results from survey sampling to manual screening for all references (i.e., a gold-standard) for five selected PAHs.
- To estimate the number of references that must be screened to represent the distribution of evidence across KCCs.

Approach

- We selected five PAHs among previously survey sampled PAHs for manual screening (including tagging) of every reference.
- Within a PAH, to test if the distribution of references for each KCC tag changed by the number of relevant references (i.e., references included after title abstract screening), we analyzed the KCC tag distribution in sequentially increased numbers of relevant references.
- We estimated how many relevant studies need to be screened for the screening results to be no more than 5% different from the true results—the number of relevant studies when all references were manually screened.

Methods

1. Results from a previously-conducted PubMed, Scopus, and Web of Science literature search for mechanistic data for five PAHs were used in the analysis: benz[a]anthracene, dibenzo[a,l]pyrene, fluoranthene, phenanthrene, and pyrene.
   - Reviewers read the titles and abstracts of all references and labeled each reference with KCC tags for which information is mentioned. In other words, one reference can have multiple KCC tags.
   - We divided KCC8 into 8a: aryl hydrocarbon receptor (AHR)-mediated effects and 8b: all other receptor-mediated effects, and KCC10 into 10a, 10b, and 10c. This yielded up to 13 KCC tags per PAH x 5 PAHs = 65 PAH-KCC tag pairs.
2. To test if the distribution of KCC tags within a PAH is stable across numbers of screened references, we analyzed sequentially increased numbers of relevant references using T-tests with a Bonferroni correction to account for multiple comparisons. Fifty relevant references were used as the initial sample and 25 relevant references were added at a time until all were included.
3. From the sequentially increased numbers of reviewer-screened references, we estimated how many relevant references need to be screened for the screening results to be no more than 5% different from the true results—the number of relevant references when all references were manually screened.

Results

Fig 1: Number of manually-screened references

Reviewers manually screened over 5,800 references at the title and abstract level for five PAHs, and approximately 15% were determined to contain information relevant to one or more KCCs.

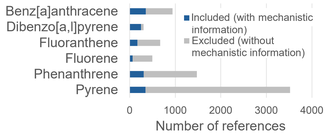


Fig 2: Based on manual screening, the true distribution of KCCs in each PAH varied

The width of each slice represent the portion of KCC reference to total included references. The number indicates number of reference.

KCC1: act as an electrophile
KCC2: be genotoxic
KCC3: alter DNA repair or cause genomic instability
KCC4: induce epigenetic alterations
KCC5: induce oxidative stress
KCC6: induce chronic inflammation
KCC7: induce Immuno-modulation
KCC8a: modulate receptor-mediated effect - aryl hydrocarbon receptor (AHR)
KCC8a: modulate receptor-mediated effect - other than AHR
KCC9: cause immortalization
KCC10a: increase cell proliferation
KCC10b: decrease cell death
KCC10c: alter nutrient supply

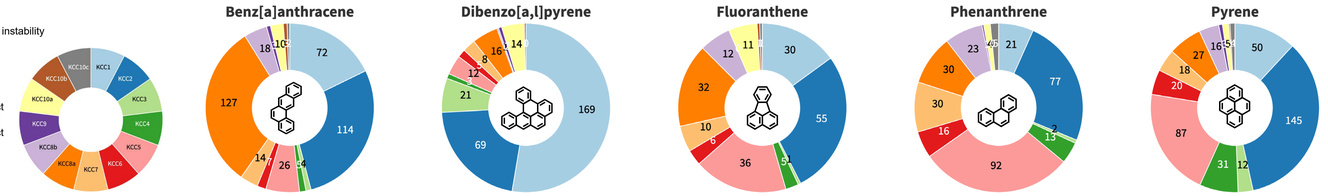


Fig 3: In increasing numbers of relevant references screened, comparisons of the KCC tag per PAH was not statistically different from the true distribution

- The consistency in the proportion of KCC references in each sample increments demonstrates that the distribution of KCC data remains stable for each PAH as additional references are screened.
- Survey sampling provides good indication of KCC reference relative distribution with fewer references screened.

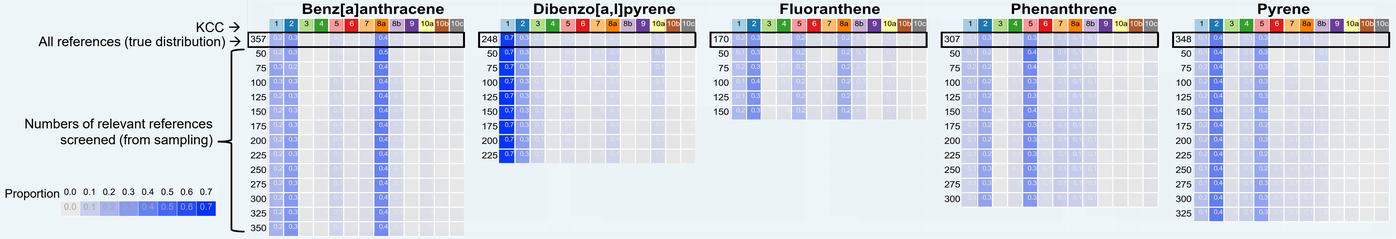
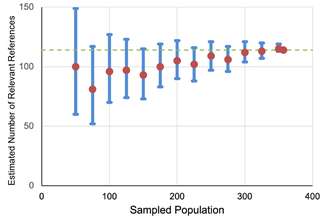


Fig 4: Estimating how many relevant studies need to be screened to be within 5% of the true number in a PAH-KCC tag pair

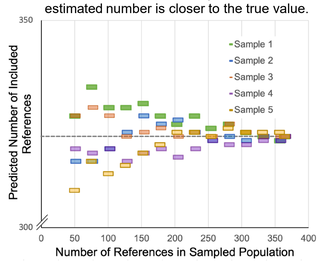
Benz[a]anthracene-KCC2 as an example.

4a: The true number of relevant reference (- - -) is within 95% confidence internals (|) of estimated number (●) in all sampled populations.



- As the sampled population grows, the 95% confidence interval shrinks.
- The 95% confidence interval is wide when sampled population is small→ We don't recommend using the estimation from small sampled populations

4b: As the sampled population grows, the estimated number is closer to the true value.



More references needed to be screened when the relevant proportion of a given PAH-KCC tag pair was low.

Example: Nearly all references are estimated need to be screened for all but three most abundant benz[a]anthracene-KCC tag pairs.

Summary and Conclusions

1. The number of included references for each PAH-KCC combination can be predicted early in the screening process (i.e., with a small initial sample).
   - For PAH-KCC tag pairs with little-to-no data, survey sampling and prediction has limited utility.
2. The distribution of references assigned to each KCC did not change significantly across samples of relevant references screened, compared to the full results.
3. Our results showed that screening a proportion of references can be a viable, less resource intensive option to obtain an accurate overview of evidence distribution of mechanistic research across multiple defined categories.

This research was supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences.