Cover letter: Making, Updating, and Querying Causal Models with `CausalQueries`

Berlin, July 2025

Dear Editors:

Thank you and the reviewers for these useful comments on our manuscript.

We address these in turn:

Reviewers 1's comments were mostly formal. We address these in turn:

1. `\section`, `\subsection`, etc. should be in sentence style

   We have implemented this change.

2. The code presented in the manuscript should not contain comments within the verbatim code. Instead the comments should be made in the normal LaTeX text.

   We have removed all comments from verbatim code.

3. For the code layout in R publications, we typically distinguish input/output using Sinput/Soutput (or equivalently CodeInput/CodeOutput). Unless there are special reasons to format it differently, the input should use the text width (up to 76 or 77 characters) and be indented by two spaces,

   We relied on the `.qmd` template for jss; we have now modified this for `CodeInput`; for `CodeOutput` we adjusted the `.tex` files after compiling.

4. Code should have enough spaces to facilitate reading. Please include spaces before and after operators and after commas (unless spaces have syntactical meaning).

   We have checked compliance with this provision.

5. For R-related manuscripts: The first argument of data() and library() should always be quoted, e.g., library("foo").

We have checked compliance with this provision.

6. If using "e.g." and "i.e." add a comma after the period to keep LaTeX from interpreting them as the end of a sentence, i.e.: "e.g.," and "i.e.,".

   We have checked compliance with this provision.

7. The rule for capitalizing the starting letters of Figure, Section and Table is as follows: If you are referring to a particular figure/section/table then capitalize the first letter, otherwise use a lower-case first letter.

   We have checked compliance with this provision.

8. Figures, tables and equations should be marked with a `\label` and referred to by `\ref`

   We have checked compliance with this provision.

9. All captions should appear below the corresponding figure/table. The captions should be in sentence style and end with a period. No additional formatting should be used for the caption.

   We have checked compliance with this provision.

10. All table row/column headers should also be in sentence style. There should not be further footnote-style annotations in tables; these should all be placed in the caption.

    We have checked compliance with this provision.

11. Equations should be marked with a `\label` and referred to by either `Equation~\ref{...}` (with capitalization, without parentheses) or (`\ref({...}`) with the former being preferred if the number of equation references is not too large.

    We have checked compliance with this provision.

12. In all cases, code input/output must fit within the normal text width of the manuscript.

    We have checked compliance with this provision.

13. For bullet lists/itemized lists please use either a comma, semi-colon, or period at the end of each item.

    We have checked compliance with this provision.

14. Abbreviations should be spelled in upper-case letters without additional formatting (i.e., without periods, without small caps, italics, etc.). All abbreviations should be introduced with their expansion where the expansion should not be capitalized.

    We have checked compliance with this provision.

15. Do not use additional formatting for specific words unless explicitly required by the JSS style guide

    We have checked compliance with this provision.

16. As a reminder, please make sure that: `\proglang`, `\pkg` and `\code` have been used for highlighting throughout the paper (including titles and references), except where explicitly escaped.

    We have checked compliance with this provision.

17. Please make sure that all software packages are [**?**]'d properly.

    We have checked compliance with this provision.

18. All references should be in title style.

    We have checked compliance with this provision.

Reviewer 2

1. Page 6, paragraph 4, line 1: "…define arbitrary DAG…" should be "…define an arbitrary DAG…".

   Corrected, with thanks.

2. $2^{(4^2)}$ with $2^{(2^4)}$

   Corrected, with thanks.

3. Page 7, Definition 1, item 3: For consistency with items 1 and 2, the authors may want to replace "a collection of functions…" with "an ordered collection of functions…".

   Edited, with thanks.

4. Also, since the paper's introduction refers to "…causal models consistent with the structural model…" it would be beneficial to add a note on how these terms are distinguished in this paper somewhere.

   We add a paragraph distinguishing between these types of model beside the introduction of the definition.

5. Page 7, first paragraph following Definition 1: For readers familiar with Pearl's work, the authors may want to add a note that the nodal type variables in this paper often are referred to as response function variables in his terminology.

Edited, with thanks.

6. Page 7, third paragraph following Definition 1: It is unclear what "…align the domain…" means. Using "…partition the domain…" is more clear and standard terminology. Pearl calls this a canonical partition.

   Edited, with thanks.

7. $f^j \rightarrow f_{Y_j}$, and similar

   Edited, with thanks.

8. Page 8, line 14: Following common terminology on distributions over variables, "…Dirichlet distribution over λ_j…" should be either "…Dirichlet distribution over θ_j…" or "…Dirichlet distribution for λ…". Likewise, the displayed Bayesian updating equation following the paragraph should use θ in place of λ, since λ is itself the probability of the causal type θ.

   Please note: We have maintained the existing language here. The reason is that the `CausalQueries` *package in fact updates over the λs. The reviewer correctly notes that the λ terms characterize the distribution over the θs; they are thus part of the definition of a causal model. However* `CausalQueries` *works on distributions over the λs, that is it lets users update over causal models. Intuitively one might think of λ_j as capturing the share of units of type θ^j; if a unit is drawn at random then the probability that a unit has type θ^j is λ_j. We assume however that there is uncertainty not just over θ^j but over the shares, λ_j and so we update over these shares.*

9. Page 8, bottom: In the displayed multinomial, the parameter N has not been defined.

   Edited, with thanks.

10. Page 9, first displayed formula: Here it becomes relevant what kinds of responses the subscripts actually encode.

    This is correct; we have now added text providing the encoding we use explicitly: "As a practical matter we need to label response types. In CausalQueries this is done using subscripts that indicate the response given different combinations of parents. A node, Y, with one binary parent, X, has a nodal types subscripted with two values indicating the two possible values of Y's parent (0 or 1): $(\theta_{00}, \theta_{01}, \theta_{10}, \theta_{11})$, where θ_{ab} labels nodal type $(Y(X = 0) = a, Y(X = 1) = b)$. The same approach is used for nodes with more (or fewer) nodal types, where the ith digit in the subscript corresponds to the value the node takes when the parents take on the ith combination of possible parent values (listed in colexicographic binary order given the specified ordering of parents)."

11. Page 16–17: Regarding caution and sensitivity to priors, the authors may also want to refer to Richardson et al. [2011].

    Added, with thanks.

12. Page 17, line 1: In "…when models are not identified…", the word models should rather be replaced by estimands or queries.

    Edited, with thanks.

13. Page 27, second code block should use `"Y[X=0]==1"` rather than `"Y[X=1]==1"` to be consistent with the preceding paragraph.

    Corrected, with thanks.

14. Page 28, first line of paragraph following second code block: "…in addition…" should be "… and in addition,…" or "…in addition to…".

    Corrected, with thanks.

15. Page 29, example under "Nested queries": The authors may want to add a note that the query `Y[M=M[X=0],X=1]==1` is commonly referred to as a natural direct effect in mediation analysis by Pearl.

    Added, with thanks.

16. Page 32, sentence before code block: "…limits model…" should be "…lipids model…".

    Corrected, with thanks.

17. Page 34: The authors note that risk ratios are currently not implemented but trying `lipids_model |> get_query_types("Y[X=1] / Y[X=0]")` yields output without warnings.

    The reviewer is correct. The output in this case indicates possible problems but no warning was provided:

```
>lipids_model |> get_query_types("Y[X=1] / Y[X=0]")
    Z0.X00.Y00 Z1.X00.Y00 Z0.X10.Y00 Z1.X10.Y00 Z0.X01.Y00 Z1.X01.Y00
           NaN        NaN        NaN        NaN        NaN        NaN
    Z0.X01.Y10 Z1.X01.Y10 Z0.X11.Y10 Z1.X11.Y10 Z0.X00.Y01 Z1.X00.Y01
             0          0          0          0        Inf        Inf
```

    The output from the updated package now includes a warning:

```
Warning message:
In check_query(query) :
  Non-linear transformations (such as / or ^) are not supported in querying.
```

18. In `?make_model`, the possibility for confounded `X<->Y` should be mentioned.

    Added, with thanks.

19. Also nice to have: links to documentations of related functions mentioned, e.g., to set priors and set parameters in `?make_model`.

    Implemented, with thanks.

20. In `summary.causal_model`, the meaning of the display column for a node is not immediately clear and might benefit from an example in the documentation.

    Implemented, with thanks.

This careful set of reviews and the close reads have greatly strengthened this manuscript.

We note only one point of disagreement with reviewer B but hope to have improved the text at this point also to make the logic clearer.

With thanks,

Macartan Humphreys