

1      **The Name of the Title Is Hope**

2  
3      ANONYMOUS AUTHOR(S)

4  
5      bla blabla

6  
7      CCS Concepts: • **Applied computing** → *Document analysis; Environmental sciences;* • **Human-centered computing** → **HCI**  
8      **theory, concepts and models.**

9      Additional Key Words and Phrases: Large language models

10  
11     **ACM Reference Format:**

12     Anonymous Author(s). 2025. The Name of the Title Is Hope. In *Proceedings of CHI Conference on Human Factors in Computing Systems (CHI'26)*. ACM, New York, NY, USA, 11 pages. <https://doi.org/XXXXXX.XXXXXXX>

- 13  
14  
15     • Something about “analysis review” - Roger thinks it’s a better to have a new word for this.  
16     • provide a baseline understand - place to start  
17     • demonstrate - analytically homogeneous - the table won’t look like that

18  
19     **1 Introduction**

20  
21     Decisions are everywhere in data analysis, from the initial data collection, data pre-processing to the modelling  
22     choices. These decisions will impact the final output of the data analysis, which may lead to different conclusions  
23     and policy recommendations. When such flexibility can be misused—through practices such as p-hacking, selective  
24     reporting, or unjustified analytical adjustments—it can inflate effect sizes or produce misleading results that meet  
25     conventional thresholds for statistical significance. They have been demonstrated through many-analysts experiments,  
26     where independent teams analyzing the same dataset to answer a pre-defined research question often arrive at markedly  
27     different conclusions. These practices not only compromise the validity of individual studies but also threaten the  
28     broader credibility of statistical analysis and scientific research as a whole.

29  
30     Multiple recommendations have been proposed to improve data analysis practices, such as pre-registration and  
31     multiverse analysis. Bayesian methods also offer a different paradigm to p-value driven inference for interpreting  
32     statistical evidence. Most empirical studies of data analysis practices focus on specially designed and simplified analysis  
33     scenarios. While informative, these setups may not adequately capture the complexity of the data analysis with  
34     significant policy implications. [In practice, studying the data analysis decisions with actual applications is challenging.]  
35     Analysts may no longer be available for interviews due to job changes, and even when they are, recalling the full set  
36     of decisions and thinking process made during the analysis is often infeasible. Moreover, only until the last decades,  
37     analysis scripts and reproducible materials were not commonly required by journals for publishing. [As a result, it  
38     remains challenging to study how analytical decisions are made.]

39  
40     In this work, we focus on a specific class of air pollution modelling studies that estimate the effect size of particulate  
41     matter (PM2.5 or PM10) on mortality, typically using Poisson regression or generalized additive models (GAMs).

---

42     Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not  
43     made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components  
44     of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to  
45     redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

46     © 2025 ACM.

47     Manuscript submitted to ACM

48  
49     Manuscript submitted to ACM

53 While individual modelling choices vary, these studies often share a common structure: they adjust for meteorological  
 54 covariates such as temperature and humidity, apply temporal or spatial treatments, like including lagged variables and  
 55 may estimate the effect by city or region before combining results. Because these studies investigate similar scientific  
 56 questions using a shared modelling framework, they form a natural many-analyst setting. This allows us to examine, in  
 57 a real-world context, the range of analytical decisions made by different researchers addressing the same underlying  
 58 question.  
 59

60 In this work, we develop a structured tabular format to record the analytical decisions made by researchers in the air  
 61 pollution modelling literature. Using large language models (LLMs), we automate the extraction of these decisions from  
 62 published papers. This allows us to treat decisions as data – allowing us to track them over time, compare methodology  
 63 across papers, and query commonly used approaches. We further introduce a workflow to cluster studies based on  
 64 decision similarity, revealing three distinct groups of papers that reflect the modelling strategies differ in the European  
 65 and U.S. studies, which offers a new way to visualize the field in the air pollution mortality modelling.  
 66

67 The rest of the paper is organized as follows. In Section 2, we review the background on data analysis decisions.  
 68 Section 3 describes the data structure for recording decisions, the use of large language models to process research  
 69 papers, and the validation of LLM outputs. In Section 4, we present the method for calculating paper similarity based  
 70 on decision similarities. Section 5 reports the finding of our analysis, including the clustering of papers according to  
 71 similarity scores and sensitivity analyses related to LLM providers, prompt engineering, and LLM parameters. Finally,  
 72 Section 6 discusses the implications of our study.  
 73

## 74 2 Background

75 Data analysis as an complicated, iterative process to make sense [ref] of the data collected. The iterative process of  
 76 formulating hypothesis Jun et al. [11].  
 77

78 Choices are made at nearly every stage of data analysis, ranging from variable pre-processing variables, variable  
 79 and lag selection in model formulation, to the specification of smoothing parameter during model construction. These  
 80 possible choices contribute to what Gelman and Loken [8] describe as the “garden of forking paths”. These choices  
 81 can introduce substantial variability in results, which has been demonstrated in many-analyst experiments, where  
 82 independent teams analyzing the same dataset to answer a pre-defined research question often arrive at markedly  
 83 different conclusions. A prominent example is Silberzahn et al. [19] where researchers reported a wide range of point  
 84 estimates and 95% confidence intervals for the effect of soccer players’ skin tone on the number of red cards awarded  
 85 by referees (odds ratio from 0.89 to 2.93). Similar findings have emerged in other domains, including structural equation  
 86 modeling [18], applied microeconomics [10], neuroimaging [5], and ecology and evolutionary biology [9].  
 87

88 Another line of work focuses on developing software tools to support analysts in making more informed decisions.  
 89 For example, the Tisane package [11] integrates conceptual ideas, such as DAGs, and modelling structure (group/  
 90 cluster/ hierarchical structure), to assist junior researchers in specifying GLM and GLMM model. The DeclareDesign  
 91 package [4] introduces the MIDA framework for researchers to declare, diagnose, and redesign their analyses to produce  
 92 a distribution of the statistic of interest. This approach has been applied in randomized controlled trial [3].  
 93

### 94 The multiverse package

- 95 • facilitates the specification and execution of multiple parallel choices for sensitivity analysis, allowing researchers  
 96 to systematically explore how different choices affect results and to report the range of plausible outcomes that  
 97 arise from alternative analytic paths.

### 98 Study decisions in data analysis:

- interview analysts and researchers to provide recommendation for data analysis practices [1, 12, 15].
- Liu et al. [15] provides visualization to communicate the decision processes through the Analytic Decision Graphs (ADG)
- Simson et al. [20] conducts a participatory AI study to demonstrate the “garden of forking paths” of decisions in data analysis and how it affects ML fairness

### 3 Extracting decisions from data analysis

#### 3.1 Decisions in data analysis

To record these decisions in a data structure, the tidy data principle Wickham [21] provides the guidance on how to structure the data where each variable should be in a column, each observation in a row. In the context of decisions made during data analysis, we adapt this principle to mean that each row represents a decision made by the authors of a paper and a paper often spans multiple decisions.

- what constitute a decision in data analysis
- some decisions are related to how the variable is estimated spatially and temporally
- model level decisions on how the model is estimated spatially (for multi-site analyses) and/or temporally (different treatments for years or seasons)
- extract the exact text from the paper

An example decisions may look as follows:

Paper	ID	Model	variable	method	parameter	type	reason	decision
ostro	1	Poisson regression	temperature	smoothing spline	degree of freedom	parameter	NA	3 degree of freedom
ostro	2	Poisson regression	temperature	smoothing spline	degree of freedom	temporal	NA	1-day lag
ostro	3	Poisson regression	relative humidity	LOESS	smoothing parameter	parameter	to minimize Akaike's Information Criterion	NA
ostro	4	Poisson regression	model	NA	NA	spatial	to account for variation among cities	separate regression models fit in each city

However, decisions statements are often implicit, and the justifications may not directly align with the decisions themselves. We identify four common anomalies:

<sup>157</sup> 1. **Authors may combine multiple decisions into a single sentence** for coherence and conciseness of the  
<sup>158</sup> writing. Consider the following excerpt from Ostro et al. [17]:  
<sup>159</sup>

<sup>160</sup> Other covariates, such as day of the week and smoothing splines of 1-day lags of average temperature  
<sup>161</sup> and humidity (each with 3 df), were also included in the model because they may be associated with  
<sup>162</sup> daily mortality and are likely to vary over time in concert with air pollution levels.

<sup>163</sup> This sentence contains four decisions: two for temperature (the temporal lag and the smoothing spline parameter)  
<sup>164</sup> and two for humidity. These decisions should be structured as separate entries.  
<sup>165</sup>

<sup>166</sup> 2. **The justification does not directly address the decision choice.** In the example above, the stated rationale  
<sup>167</sup> (“and are likely to vary over time in concert with air pollution levels”) supports the general inclusion of temporal  
<sup>168</sup> lags but does not justify the specific choice of 1-day lag over alternatives, such as 2-day average of lags 0 and 1  
<sup>169</sup> (lag01) and single-day lag of 2 days (lag2). As such, the reason field should be recorded as NA.  
<sup>170</sup>

<sup>171</sup> 3. **Some decisions may be omitted because they are data-driven.** For instance, Katsouyanni et al. [13] states:  
<sup>172</sup> The inclusion of lagged weather variables and the choice of smoothing parameters for all of the weather  
<sup>173</sup> variables were done by minimizing Akaike’s information criterion.  
<sup>174</sup>

<sup>175</sup> In this case, while the method of selection (minimizing AIC) is specified, the actual degree of freedom used is not.  
<sup>176</sup> Such data-driven decisions may be recorded with “NA” in the decision field, but the reason field should still be recorded  
<sup>177</sup> as “by minimizing Akaike’s information criterion”  
<sup>178</sup>

<sup>179</sup> 4. **Information required to interpret the decision may be distributed across multiple sections.** In the  
<sup>180</sup> previous example, “weather variables” refers to mean temperature and relative humidity, as defined earlier in  
<sup>181</sup> the text. This requires cross-referencing across sections to identify the correct variables associated with each  
<sup>182</sup> modeling choice.  
<sup>183</sup>

### <sup>184</sup> 3.2 Automatic reading of literature with LLMs

<sup>185</sup> While decisions can be extracted manually from the literature, this process is labor-intensive and time-consuming.  
<sup>186</sup> Recent advances in Large Language Models (LLMs) have demonstrated potential for automating the extraction of  
<sup>187</sup> structured information from unstructured text [ref]. In this work, we use LLMs to automatically identify decisions  
<sup>188</sup> made by authors during their data analysis processes.  
<sup>189</sup>

<sup>190</sup> Text recognition from PDF document relies on Optical Character Recognition (OCR) to convert scanned images into  
<sup>191</sup> machine-readable text – capability currently offered by Anthropic Claude and Google Gemini. We instruct the LLM  
<sup>192</sup> to generate a markdown file containing a JSON block that records extracted decisions, which can then be read into  
<sup>193</sup> statistical software for further analysis. The exact prompt feed to the LLM is provided in the Appendix. The ellmer  
<sup>194</sup> package [22] in R is used to connect to the Gemini and Claude API, providing the PDF attachment and the prompt in a  
<sup>195</sup> markdown file as inputs. `?@fig-llm` shows the overall workflow for decision extract using LLMs.  
<sup>196</sup>

### <sup>197</sup> 3.3 Review the LLM output

- <sup>198</sup> • something about result validation of LLM output
- <sup>199</sup> • The sensitivity of the two models to the prompt and the model parameters, such as temperature and seed, is  
<sup>200</sup> discussed in Section 5.2.

<sup>201</sup> The shiny app is designed to provide users a visual interface to review and edit the decisions extracted by the LLM  
<sup>202</sup> from the literature. The app allows three actions from the users: 1) *overwrite* – modify the content of a particular  
<sup>203</sup> cell, equivalently `dplyr::mutate(xxx = ifelse(CONDITION, "yyy" , xxx))`, 2) *delete* – remove a particular cell,  
<sup>204</sup> Manuscript submitted to ACM  
<sup>205</sup>

209 `dplyr::filter(!(CONDITION))`, and 3) *add* – manually enter a decision, `dplyr::bind_rows()`. Figure 1 illustrates  
 210 the *overwrite* action in the Shiny application, where users interactively filter the data and preview the rows affected by  
 211 their edits—in this case, changing the model entry from “generalized additive Poisson time series regression” to the  
 212 less verbose “Poisson regression”. Upon confirmation, the corresponding tidyverse code is generated, and users can  
 213 download the edited table and incorporate the code into their R script.

242 Fig. 1. The Shiny application interface for editing Large Language Model (LLM)-generated decisions (overwrite, delete, and add). (1)  
 243 the default interface after loading the input CSV file. (2) The table view will update interactively upon the user-defined filter condition  
 244 – expressed using `dplyr::filter()` syntax (e.g., `paper == anderson2008size`), (3) The user edits the model column to “Poisson  
 245 regression” and applies the change by clicking the Apply changes button. The table view updates to reflect the changes (4) After  
 246 clicking the Confirm button, the corresponding tidyverse code is generated, and the table view returns to its original unfiltered view.  
 247 The edited data can be downloaded by clicking the Download CSV button.

249 Source: Article Notebook

## 251 4 Calculating paper similarity

- 253 • pre-processing
  - 254 – standardize statistical methods its corresponding parameters (LOESS, smoothing spline, etc)
  - 255 – group variables into broader categories: time, temperature, humidity, PM
- 256 • identify the most frequent analysis decisions across papers
- 257 • retain only papers that report more than x such decisions
- 258 • measure similarity between decisions and their justification using NLP

- 261           – word embedding with attention mechanism, instead of bag of word,  
 262           – specific NLP models (default to bert-base-uncased), aggregation methods from word to text  
 263       • compute paper similarity score for each paper pair by aggregating decision-level comparisons  
 264           – check/ report on the number of decisions compared in each paper pair  
 265       • similarity score can serve as the distance matrix to cluster papers by their similarity on decision choices  
 266  
 267

## 268       5 Results

### 269       5.1 Air pollution mortality modelling

271       Decision quality summary

- 273           • look at for one type of decision (time) - what are the choices made by different papers
- 274           • look at whether decisions changes across time
- 275           • Visualize the decision database: apply clustering algorithm and visualize the database through sigma.js
- 276

### 277       5.2 Sensitivity analysis

279       sensitivity of the pipeline: 1) LLM, 2) text model, 3) prompt, 4) LLM parameters

- 280           • standard BERT [7], Roberta [16]: trained on a much larger dataset (160GB v.s. BERT's 15GB), transformer-xl  
 281           [6], xlnet by Google Brain [23], and two domain-trained BERT models: sciBert [2] and bioBert[14], trained  
 282           on PubMed and PMC data.
- 283           • A section on reproducibility of LLM outputs: prompt experiment (see if there are papers discussing this:  
 284           <https://arxiv.org/pdf/2406.06608.pdf>)
- 285
- 286

## 287       6 Discussion

- 289           • Only prompting engineering is used to extract decisions from the literature. We expect that fine-tuning the  
 290           model on statistical or domain-specific literature to yield more robust performance on the same document,  
 291           though it would require substantially more training effort.
- 293           • people from the NYU-LMU workshop are interested to have code script attached as well because people can do  
 294           one thing in the script but report another in the paper - it would be interesting to compare the paper and the  
 295           script with some syntax extraction.
- 296           • Validation of the output:

297       the nature of the task: Our task involve a reasoning component in that it requires causal reasoning to identify  
 299       the decisions made by the authors, and its justification/ rationale, rather than purely summarizing the text through  
 300       pattern-matching.

## 302       References

- 304       [1] Sara Alspaugh, Nava Zokaei, Andrea Liu, Cindy Jin, and Marti A. Hearst. Futzling and moseying: Interviews with professional data analysts on  
 305           exploration practices. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):22–31, 01 2019. doi: 10.1109/TVCG.2018.2865040. URL  
 306           <https://ieeexplore.ieee.org/document/8440815>.
- 307       [2] Iz Beltagy, Kyle Lo, and Arman Cohan. Proceedings of the 2019 conference on empirical methods in natural language processing and the  
 308           9th international joint conference on natural language processing (emnlp-ijcnlp). pages 3613–3618, Hong Kong, China, 2019. Association for  
 309           Computational Linguistics. doi: 10.18653/v1/D19-1371. URL <https://www.aclweb.org/anthology/D19-1371>.
- 310       [3] Dorothy V. M. Bishop and Charles Hulme. When alternative analyses of the same data come to different conclusions: A tutorial using declaredesign  
 311           with a worked real-world example. *Advances in Methods and Practices in Psychological Science*, 7(3):25152459241267904, 07 2024. doi: 10.1177/  
 312           25152459241267904. URL <https://doi.org/10.1177/25152459241267904>. Publisher: SAGE Publications Inc.

- [4] Graeme Blair, Jasper Cooper, Alexander Coppock, and Macartan Humphreys. Declaring and diagnosing research designs. *American Political Science Review*, 113(3):838–859, 08 2019. doi: 10.1017/S0003055419000194. URL [https://www.cambridge.org/core/product/identifier/S0003055419000194/type/journal\\_article](https://www.cambridge.org/core/product/identifier/S0003055419000194/type/journal_article).
- [5] Rotem Botvinik-Nezer, Felix Holzmeister, Colin F. Camerer, Anna Dreber, Juergen Huber, Magnus Johannesson, Michael Kirchler, Roni Iwanir, Jeanette A. Mumford, R. Alison Adcock, Paolo Avesani, Blazej M. Baczkowski, Aahana Bajracharya, Leah Bakst, Sheryl Ball, Marco Barilari, Nadège Bault, Derek Beaton, Julia Beitner, Roland G. Benoit, Ruud M. W. J. Berkers, Jamil P. Bhanji, Bharat B. Biswal, Sebastian Bobadilla-Suarez, Tiago Bortolini, Katherine L. Bottenhorn, Alexander Bowring, Senne Braem, Hayley R. Brooks, Emily G. Brudner, Cristian B. Calderon, Julia A. Camilleri, Jaime J. Castrellon, Luca Cecchetti, Edna C. Cieslik, Zachary J. Cole, Olivier Collignon, Robert W. Cox, William A. Cunningham, Stefan Czoschke, Kamalaker Dadi, Charles P. Davis, Alberto De Luca, Mauricio R. Delgado, Lysia Demetriou, Jeffrey B. Dennison, Xin Di, Erin W. Dickie, Ekaterina Dobryakova, Claire L. Donnat, Juergen Dukart, Niall W. Duncan, Joke Durnez, Amr Eed, Simon B. Eickhoff, Andrew Erhart, Laura Fontanesi, G. Matthew Fricke, Shiguang Fu, Adriana Galván, Remi Gau, Sarah Genon, Tristan Glatard, Enrico Gleerean, Jelle J. Goeman, Sergej A. E. Golowin, Carlos González-García, Krzysztof J. Gorgolewski, Cheryl L. Grady, Mikella A. Green, João F. Guassi Moreira, Olivia Guest, Shabnam Hakimi, J. Paul Hamilton, Roeland Hancock, Giacomo Handjaras, Bronson B. Harry, Colin Hawco, Peer Herholz, Gabrielle Herman, Stephan Heunis, Felix Hoffstaedter, Jeremy Hogeveen, Susan Holmes, Chuan-Peng Hu, Scott A. Huettel, Matthew E. Hughes, Vittorio Iacobella, Alexandru D. Iordan, Peder M. Isager, Ayse I. Isik, Andrew Jahn, Matthew R. Johnson, Tom Johnstone, Michael J. E. Joseph, Anthony C. Juliano, Joseph W. Kable, Michalis Kassinopoulos, Cemal Koba, Xiang-Zhen Kong, Timothy R. Kosciuk, Nur Erkut Kucukboyaci, Brice A. Kuhl, Sebastian Kupek, Angela R. Laird, Claus Lamm, Robert Langner, Nina Lauharatanahirun, Hongmi Lee, Sangil Lee, Alexander Leemans, Andrea Leo, Elise Lesage, Flora Li, Monica Y. C. Li, Phui Cheng Lim, Evan N. Lintz, Schuyler W. Liphardt, Annabel B. Losecaat Vermeer, Bradley C. Love, Michael L. Mack, Norberto Malpica, Theo Marins, Camilla Maumet, Kelsey McDonald, Joseph T. McGuire, Helena Melero, Adriana S. Méndez Leal, Benjamin Meyer, Kristin N. Meyer, Glad Mihai, Georgios D. Mitsis, Jorge Moll, Dylan M. Nielson, Gustav Nilsonne, Michael P. Notter, Emanuele Olivetti, Adrian I. Onicas, Paolo Papale, Kaustubh R. Patil, Jonathan E. Peelle, Alexandre Pérez, Doris Pischedda, Jean-Baptiste Poline, Yanina Prystauka, Shruti Ray, Patricia A. Reuter-Lorenz, Richard C. Reynolds, Emiliano Ricciardi, Jenny R. Rieck, Anais M. Rodriguez-Thompson, Anthony Romyn, Taylor Salo, Gregory R. Samanez-Larkin, Emilio Sanz-Morales, Margaret L. Schlichting, Douglas H. Schultz, Qiang Shen, Margaret A. Sheridan, Jennifer A. Silvers, Kenny Skagerlund, Alec Smith, David V. Smith, Peter Sokol-Hessner, Simon R. Steinkamp, Sarah M. Tashjian, Bertrand Thirion, John N. Thorp, Gustav Tinghög, Loreen Tisdall, Steven H. Tompson, Claudio Toro-Serey, Juan Jesus Torre Tresols, Leonardo Tozzi, Vuong Truong, Luca Turella, Anna E. van 't Veer, Tom Verguts, Jean M. Vettel, Sagana Vijayarajah, Khoi Vo, Matthew B. Wall, Wouter D. Weeda, Susanne Weis, David J. White, David Wisniewski, Alba Xifra-Poxas, Emily A. Yearling, Sangsuk Yoon, Rui Yuan, Kenneth S. L. Yuen, Lei Zhang, Xu Zhang, Joshua E. Zosky, Thomas E. Nichols, Russell A. Poldrack, and Tom Schonberg. Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, 582(7810):84–88, 06 2020. doi: 10.1038/s41586-020-2314-9. URL <https://www.nature.com/articles/s41586-020-2314-9>. Publisher: Nature Publishing Group.
- [6] Zihang Dai, Zhilin Yang, Yiming Yang, Jaime Carbonell, Quoc V. Le, and Ruslan Salakhutdinov. Transformer-xl: Attentive language models beyond a fixed-length context. doi: 10.48550/arXiv.1901.02860.
- [7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Naacl-hlt 2019. page 4171–4186, Minneapolis, Minnesota, 06 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-1423. URL <https://aclanthology.org/N19-1423>.
- [8] Andrew Gelman and Eric Loken. The statistical crisis in science. *American Scientist*, 102(6):460–465, 12 2014. URL <https://www.proquest.com/docview/1616141998/abstract/5E050DCE82414037PQ/1>. Num Pages: 6 Place: Research Triangle Park, United States Publisher: Sigma XI-The Scientific Research Society.
- [9] Elliot Gould, Hannah S. Fraser, Timothy H. Parker, Shinichi Nakagawa, Simon C. Griffith, Peter A. Vesk, Fiona Fidler, Daniel G. Hamilton, Robin N. Abbey-Lee, Jessica K. Abbott, Luis A. Aguirre, Carles Alcaraz, Irith Aloni, Drew Altschul, Kunal Arekar, Jeff W. Atkins, Joe Atkinson, Christopher M. Baker, Meghan Barrett, Kristian Bell, Suleiman Kehinde Bello, Iván Beltrán, Bernd J. Berauer, Michael Grant Bertram, Peter D. Billman, Charlie K. Blake, Shannon Blake, Louis Bliard, Andrea Bonisolí-Alquati, Timothée Bonnet, Camille Nina Marion Bordes, Aneesh P. H. Bose, Thomas Botterill-James, Melissa Anna Boyd, Sarah A. Boyle, Tom Bradfer-Lawrence, Jennifer Bradham, Jack A. Brand, Martin I. Brengdahl, Martin Bulla, Luc Bussière, Ettore Camerlenghi, Sara E. Campbell, Leonardo L. F. Campos, Anthony Caravaggi, Pedro Cardoso, Charles J. W. Carroll, Therese A. Catanach, Xuan Chen, Heung Ying Janet Chik, Emily Sarah Choy, Alec Philip Christie, Angela Chuang, Amanda J. Chunco, Bethany L. Clark, Andrea Contina, Garth A. Covernton, Murray P. Cox, Kimberly A. Cressman, Marco Crotti, Connor Davidson Crouch, Pietro B. D'Amelio, Alexandra Allison de Sousa, Timm Fabian Döbert, Ralph Dobler, Adam J. Dobson, Tim S. Doherty, Szymon Marian Drobniak, Alexandra Grace Duffy, Alison B. Duncan, Robert P. Dunn, Jamie Dunning, Trishna Dutta, Luke Eberhart-Hertel, Jared Alan Elmore, Mahmoud Medhat Elsherif, Holly M. English, David C. Ensminger, Ulrich Rainer Ernst, Stephen M. Ferguson, Esteban Fernandez-Juricic, Thalita Ferreira-Arruda, John Fieberg, Elizabeth A. Finch, Evan A. Fiorenza, David N. Fisher, Amélie Fontaine, Wolfgang Forstmeier, Yoan Fourcade, Graham S. Frank, Kathryn A. Freund, Eduardo Fuentes-Lillo, Sara L. Gandy, Dustin G. Gannon, Ana I. García-Cervigón, Alexis C. Garretson, Xuezhen Ge, William L. Geary, Charly Géron, Marc Gilles, Antje Girndt, Daniel Glikzman, Harrison B. Goldspiel, Dylan G. E. Gomes, Megan Kate Good, Sarah C. Goslee, J. Stephen Gosnell, Eliza M. Grames, Paolo Gratton, Nicholas M. Grebe, Skye M. Greenler, Maaike Griffioen, Daniel M. Griffith, Frances J. Griffith, Jake J. Grossman, Ali Güncan, Stef Haesen, James G. Hagan, Heather A. Hager, Jonathan Philo Harris, Natasha Dean Harrison, Sarah Syedia Hasnain, Justin Chase Havird, Andrew J. Heaton, María Laura Herrera-Chastrue, Tanner J. Howard, Bin-Yan Hsu, Fabiola Iannarilli, Esperanza C. Iranzo, Erik N. K. Iverson, Saheed Olaide Jimoh, Douglas H. Johnson, Martin Johnsson, Jesse Jorna, Tommaso Jucker, Martin Jung, Ineta Kačergytė, Oliver Kaltz, Alison Ke, Clint D. Kelly, Katharine Keegan, Friedrich Wolfgang Keppeler, Alexander K. Killion, Dongmin Kim, David P. Kochan, Peter Korsten, Shan Kothari, Jonas Kuppler, Jillian M. Kusch, Małgorzata Lagisz, Kristen Marianne Lalla, Daniel J. Larkin, Courtney L. Larson, Katherine S. Lauck, M. Elise Lauterbur, Alan Law, Don-Jean

- 365 Léandri-Breton, Jonas J. Lembrechts, Kiara L'Herpinere, Eva J. P. Lievens, Daniela Oliveira de Lima, Shane Lindsay, Martin Luquet, Ross MacLeod,  
 366 Kirsty H. Macphie, Kit Magellan, Magdalena M. Mair, Lisa E. Malm, Stefano Mammola, Caitlin P. Mandeville, Michael Manhart, Laura Milena  
 367 Manrique-Garzon, Elina Mäntylä, Philippe Marchand, Benjamin Michael Marshall, Charles A. Martin, Dominic Andreas Martin, Jake Mitchell  
 368 Martin, April Robin Martinig, Erin S. McCallum, Mark McCauley, Sabrina M. McNew, Scott J. Meiners, Thomas Merkling, Marcus Michelangeli,  
 369 Maria Moiron, Bruno Moreira, Jennifer Mortensen, Benjamin Mos, Taofeek Olatunbosun Muraina, Penelope Wrenn Murphy, Luca Nelli, Petri  
 370 Niemelä, Josh Nightingale, Gustav Nilsonne, Sergio Nolazco, Sabine S. Nooten, Jessie Lanterman Novotny, Agnes Birgitta Olin, Chris L. Organ,  
 371 Kate L. Ostevik, Facundo Xavier Palacio, Matthieu Paquet, Darren James Parker, David J. Pascall, Valerie J. Pasquarella, John Harold Paterson, Ana  
 372 Payo-Payo, Karen Marie Pedersen, Grégoire Perez, Kayla I. Perry, Patrice Pottier, Michael J. Proulx, Raphaël Proulx, Jessica L. Pruitt, Veronarindra  
 373 Ramananjato, Finaritra Tolotra Randimbiarison, Onja H. Razafindratsima, Diana J. Rennison, Federico Riva, Sepand Riyahi, Michael James Roast,  
 374 Felipe Pereira Rocha, Dominique G. Roche, Cristian Román-Palacios, Michael S. Rosenberg, Jessica Ross, Freya E. Rowland, Deusdedith Rugemalila,  
 375 Avery L. Russell, Suvi Ruuskanen, Patrick Saccone, Asaf Sadeh, Stephen M. Salazar, Kris Sales, Pablo Salmón, Alfredo Sánchez-Tójar, Leticia Pereira  
 376 Santos, Francesca Santostefano, Hayden T. Schilling, Marcus Schmidt, Tim Schmoll, Adam C. Schneider, Allie E. Schroeder, Nicolas  
 377 Schtitzelle, Nick L. Schultz, Drew A. Scott, Michael Peter Scroggie, Julie Teresa Shapiro, Nitika Sharma, Caroline L. Shearer, Diego Simón, Michael I.  
 378 Sitvarin, Fabrício Luiz Skupien, Heather Lea Slinn, Grania Polly Smith, Jeremy A. Smith, Rahel Sollmann, Kaitlin Stack Whitney, Shannon Michael  
 379 Still, Erica F. Stuber, Guy F. Sutton, Ben Swallow, Conor Claverie Taff, Elina Takola, Andrew J. Tanentzap, Rocío Tarjuelo, Richard J. Telford,  
 380 Christopher J. Thawley, Hugo Thierry, Jacqueline Thomson, Svenja Tidau, Emily M. Tompkins, Claire Marie Tortorelli, Andrew Trlica, Biz R.  
 381 Turnell, Lara Urban, Stijn Van de Vondel, Jessica Eva Megan van der Wal, Jens Van Eckhoven, Francis van Oordt, K. Michelle Vanderwel, Mark C.  
 382 Vanderwel, Karen J. Vanderwolf, Juliana Vélez, Diana Carolina Vergara-Florez, Brian C. Verrelli, Marcus Vinícius Vieira, Nora Villamil, Valerio  
 383 Vitali, Julien Vollering, Jeffrey Walker, Xanthe J. Walker, Jonathan A. Walter, Paweł Waryszak, Ryan J. Weaver, Ronja E. M. Wedegärtner, Daniel L.  
 384 Weller, and Shannon Whelan. Same data, different analysts: variation in effect sizes due to analytical decisions in ecology and evolutionary biology.  
*BMC Biology*, 23(1):35, 02 2025. doi: 10.1186/s12915-024-02101-x. URL <https://doi.org/10.1186/s12915-024-02101-x>.
- [10] Nick Huntington-Klein, Andreu Arenas, Emily Beam, Marco Bertoni, Jeffrey R. Bloem, Pralhad Burl, Naibin Chen, Paul Grieco, Godwin Ekpe, Todd Pugatch, Martin Saavedra, and Yaniv Stopnitzky. The influence of hidden researcher decisions in applied microeconomics. *Economic Inquiry*, 59(3):944–960, 2021. doi: 10.1111/ecin.12992. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/ecin.12992>. \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/ecin.12992>.
- [11] Eunice Jun, Melissa Birchfield, Nicole De Moura, Jeffrey Heer, and René Just. Hypothesis formalization: Empirical findings, software limitations, and design implications. *ACM Transactions on Computer-Human Interaction*, 29(1):1–28, 02 2022. doi: 10.1145/3476980. URL <https://dl.acm.org/doi/10.1145/3476980>.
- [12] Alex Kale, Matthew Kay, and Jessica Hullman. Decision-making under uncertainty in research synthesis: Designing for the garden of forking paths. CHI '19, page 1–14, New York, NY, USA, 05 2019. Association for Computing Machinery. doi: 10.1145/3290605.3300432. URL <https://dl.acm.org/doi/10.1145/3290605.3300432>.
- [13] Klea Katsouyanni, Giota Touloumi, Evangelia Samoli, Alexandros Gryparis, Alain Le Tertre, Yannis Monopolis, Giuseppe Rossi, Denis Zmirou, Ferran Ballester, Azedine Boumghar, Hugh Ross Anderson, Bogdan Wojtyniak, Anna Paldy, Rony Braunestein, Juha Pekkanen, Christian Schindler, and Joel Schwartz. Confounding and effect modification in the short-term effects of ambient particles on total mortality: Results from 29 european cities within the aphea2 project. *Epidemiology*, 12(5):521, 09 2001. URL [https://journals.lww.com/epidem/fulltext/2001/09000/confounding\\_and\\_effect\\_modification\\_in\\_the\\_11.aspx](https://journals.lww.com/epidem/fulltext/2001/09000/confounding_and_effect_modification_in_the_11.aspx).
- [14] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240, 02 2020. doi: 10.1093/bioinformatics/btz682. URL <https://academic.oup.com/bioinformatics/article/36/4/1234/5566506>.
- [15] Yang Liu, Tim Althoff, and Jeffrey Heer. Paths explored, paths omitted, paths obscured: Decision points & selective reporting in end-to-end data analysis. CHI '20, page 1–14, New York, NY, USA, 04 2020. Association for Computing Machinery. doi: 10.1145/3313831.3376533. URL <https://dl.acm.org/doi/10.1145/3313831.3376533>.
- [16] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. doi: 10.48550/arXiv.1907.11692.
- [17] Bart Ostro, Rachel Broadwin, Shelley Green, Wen-Ying Feng, and Michael Lipsett. Fine particulate air pollution and mortality in nine california counties: Results from calfine. *Environmental Health Perspectives*, 114(1):29–33, 01 2006. doi: 10.1289/ehp.8335. URL <https://ehp.niehs.nih.gov/doi/10.1289/ehp.8335>. Publisher: Environmental Health Perspectives.
- [18] Marko Sarstedt, Susanne J. Adler, Christian M. Ringle, Gyeongcheol Cho, Adamantios Diamantopoulos, Heungsun Hwang, and Benjamin D. Liengaard. Same model, same data, but different outcomes: Evaluating the impact of method choices in structural equation modeling. *Journal of Product Innovation Management*, 41(6):1100–1117, 2024. doi: 10.1111/jipm.12738. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/jipm.12738>. \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/jipm.12738>.
- [19] R. Silberzahn, E. L. Uhlmann, D. P. Martin, P. Anselmi, F. Aust, E. Awtry, Š. Bahnik, F. Bai, C. Bannard, E. Bonnier, R. Carlsson, F. Cheung, G. Christensen, R. Clay, M. A. Craig, A. Dalla Rosa, L. Dam, M. H. Evans, I. Flores Cervantes, N. Fong, M. Gamez-Djokic, A. Glenz, S. Gordon-McKeon, T. J. Heaton, K. Hederos, M. Heene, A. J. Hofelich Mohr, F. Höglund, K. Hui, M. Johannesson, J. Kalodimos, E. Kaszubowski, D. M. Kennedy, R. Lei, T. A. Lindsay, S. Liverani, C. R. Madan, D. Molden, E. Molleman, R. D. Morey, L. B. Mulder, B. R. Nijstad, N. G. Pope, B. Pope, J. M. Prenoveau, F. Rink, E. Robusto, H. Roderique, A. Sandberg, E. Schlüter, F. D. Schönbrodt, M. F. Sherman, S. A. Sommer, K. Sotak, S. Spain, C. Spörlein, T. Stafford,

- 417 L. Stefanutti, S. Tauber, J. Ullrich, M. Vianello, E.-J. Wagenmakers, M. Witkowiak, S. Yoon, and B. A. Nosek. Many analysts, one data set: Making  
 418 transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science*, 1(3):337–356, 09 2018. doi:  
 419 10.1177/2515245917747646. URL <https://doi.org/10.1177/2515245917747646>. Publisher: SAGE Publications Inc.
- 420 [20] Jan Simson, Fiona Draxler, Samuel Mehr, and Christoph Kern. Preventing harmful data practices by using participatory input to navigate the  
 421 machine learning multiverse. CHI ’25, page 1–30, New York, NY, USA, 04 2025. Association for Computing Machinery. doi: 10.1145/3706598.3713482.  
 422 URL <https://dl.acm.org/doi/10.1145/3706598.3713482>.
- 423 [21] Hadley Wickham. Tidy data. *Journal of Statistical Software*, 59:1–23, 09 2014. doi: 10.18637/jss.v059.i10. URL <https://doi.org/10.18637/jss.v059.i10>.
- 424 [22] Hadley Wickham, Joe Cheng, and Aaron Jacobs. *ellmer: Chat with Large Language Models*, 2025. URL <https://CRAN.R-project.org/package=ellmer>.  
 425 R package version 0.1.1.
- 426 [23] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, and Quoc V. Le. Xlnet: Generalized autoregressive pretraining for  
 427 language understanding. doi: 10.48550/arXiv.1906.08237.
- 428 [1] Sara Alspaugh, Nava Zokaei, Andrea Liu, Cindy Jin, and Marti A. Hearst. Futzng and moseying: Interviews with professional data analysts on  
 429 exploration practices. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):22–31, 01 2019. doi: 10.1109/TVCG.2018.2865040. URL  
 430 <https://ieeexplore.ieee.org/document/8440815>.
- 431 [2] Iz Beltagy, Kyle Lo, and Arman Cohan. Proceedings of the 2019 conference on empirical methods in natural language processing and the  
 432 9th international joint conference on natural language processing (emnlp-ijcnlp). pages 3613–3618, Hong Kong, China, 2019. Association for  
 433 Computational Linguistics. doi: 10.18653/v1/D19-1371. URL <https://www.aclweb.org/anthology/D19-1371>.
- 434 [3] Dorothy V. M. Bishop and Charles Hulme. When alternative analyses of the same data come to different conclusions: A tutorial using declaredesign  
 435 with a worked real-world example. *Advances in Methods and Practices in Psychological Science*, 7(3):25152459241267904, 07 2024. doi: 10.1177/  
 436 25152459241267904. URL <https://doi.org/10.1177/25152459241267904>. Publisher: SAGE Publications Inc.
- 437 [4] Graeme Blair, Jasper Cooper, Alexander Coppock, and Macartan Humphreys. Declaring and diagnosing research designs. *American Political Science  
 438 Review*, 113(3):838–859, 08 2019. doi: 10.1017/S0003055419000194. URL [https://www.cambridge.org/core/product/identifier/S0003055419000194/type/journal\\_article](https://www.cambridge.org/core/product/identifier/S0003055419000194/type/journal_article).
- 439 [5] Rotem Botvinik-Nezer, Felix Holzmeister, Colin F. Camerer, Anna Dreber, Juergen Huber, Magnus Johannesson, Michael Kirchler, Roni Iwanir,  
 440 Jeanette A. Mumford, R. Alison Adcock, Paolo Avesani, Blazej M. Baczkowski, Aahana Bajracharya, Leah Bakst, Sheryl Ball, Marco Barilaro, Nadège  
 441 Bault, Derek Beaton, Julia Beitner, Roland G. Benoit, Ruud M. W. J. Berkers, Jamil P. Bhanji, Bharat B. Biswal, Sebastian Bobadilla-Suarez, Tiago  
 442 Bortolini, Katherine L. Bottenhorn, Alexander Bowring, Senne Braem, Hayley R. Brooks, Emily G. Brudner, Cristian B. Calderon, Julia A. Camilleri,  
 443 Jaime J. Castrellon, Luca Ceccetti, Edna C. Cieslik, Zachary J. Cole, Olivier Collignon, Robert W. Cox, William A. Cunningham, Stefan Czoschke,  
 444 Kamalaker Dadi, Charles P. Davis, Alberto De Luca, Mauricio R. Delgado, Lysia Demetriou, Jeffrey B. Dennison, Xin Di, Erin W. Dickie, Ekaterina  
 445 Dobryakova, Claire L. Donnat, Juergen Dukart, Niall W. Duncan, Joke Durnez, Amr Eed, Simon B. Eickhoff, Andrew Erhart, Laura Fontanesi,  
 446 G. Matthew Fricke, Shiguang Fu, Adriana Galván, Remi Gau, Sarah Genon, Tristan Glatard, Enrico Gleerean, Jelle J. Goeman, Sergej A. E. Golowin,  
 447 Carlos González-García, Krzysztof J. Gorgolewski, Cheryl L. Grady, Mikella A. Green, João F. Guassi Moreira, Olivia Guest, Shabnam Hakimi,  
 448 J. Paul Hamilton, Roeland Hancock, Giacomo Handjaras, Bronson B. Harry, Colin Hawco, Peer Herholz, Gabrielle Herman, Stephan Heunis, Felix  
 449 Hoffstaedter, Jeremy Hogeveen, Susan Holmes, Chuan-Peng Hu, Scott A. Huettel, Matthew E. Hughes, Vittorio Iacovella, Alexandru D. Iordan,  
 450 Peder M. Isager, Ayse I. Isik, Andrew Jahn, Matthew R. Johnson, Tom Johnstone, Michael J. E. Joseph, Anthony C. Juliano, Joseph W. Kable, Michalis  
 451 Kassinopoulos, Cemal Koba, Xiang-Zhen Kong, Timothy R. Kosciuk, Nuri Erkut Kucukboyaci, Brice A. Kuhl, Sebastian Kupek, Angela R. Laird,  
 452 Claus Lamm, Robert Langner, Nina Lauharatanahirun, Hongmi Lee, Sangil Lee, Alexander Leemans, Andrea Leo, Elise Lesage, Flora Li, Monica  
 453 Y. C. Li, Phui Cheng Lim, Evan N. Lintz, Schuyler W. Liphardt, Annabel B. Losecaat Vermeer, Bradley C. Love, Michael L. Mack, Norberto Malpica,  
 454 Theo Marins, Camille Maumet, Kelsey McDonald, Joseph T. McGuire, Helena Melero, Adriana S. Méndez Leal, Benjamin Meyer, Kristin N. Meyer,  
 455 Glad Mihai, Georgios D. Mitsis, Jorge Moll, Dylan M. Nelson, Gustav Nilsson, Michael P. Notter, Emanuele Olivetti, Adrian I. Onicas, Paolo  
 456 Papale, Kaustubh R. Patil, Jonathan E. Peelle, Alexandre Pérez, Doris Pischedda, Jean-Baptiste Poline, Yanina Prystauka, Shruti Ray, Patricia A.  
 457 Reuter-Lorenz, Richard C. Reynolds, Emiliano Ricciardi, Jenny R. Rieck, Anais M. Rodriguez-Thompson, Anthony Romyn, Taylor Salo, Gregory R.  
 458 Samanez-Larkin, Emilio Sanz-Morales, Margaret L. Schlichting, Douglas H. Schultz, Qiang Shen, Margaret A. Sheridan, Jennifer A. Silvers, Kenny  
 459 Skagerlund, Alec Smith, David V. Smith, Peter Sokol-Hessner, Simon R. Steinkamp, Sarah M. Tashjian, Bertrand Thirion, John N. Thorp, Gustav  
 460 Tinghög, Loreen Tisdall, Steven H. Tompson, Claudio Toro-Serey, Juan Jesus Torre Tresols, Leonardo Tozzi, Vuong Truong, Luca Turella, Anna E.  
 461 van ‘t Veer, Tom Verguts, Jean M. Vettel, Sagana Vijayarajah, Khoi Vo, Matthew B. Wall, Wouter D. Weeda, Susanne Weis, David J. White, David  
 462 Wisniewski, Alba Xifra-Porras, Emily A. Yearling, Sangsuk Yoon, Rui Yuan, Kenneth S. L. Yuen, Lei Zhang, Xu Zhang, Joshua E. Zosky, Thomas E.  
 463 Nichols, Russell A. Poldrack, and Tom Schonberg. Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, 582(7810):  
 464 84–88, 06 2020. doi: 10.1038/s41586-020-2314-9. URL <https://www.nature.com/articles/s41586-020-2314-9>. Publisher: Nature Publishing Group.
- 465 [6] Zihang Dai, Zhilin Yang, Yiming Yang, Jaime Carbonell, Quoc V. Le, and Ruslan Salakhutdinov. Transformer-xl: Attentive language models beyond  
 466 a fixed-length context. doi: 10.48550/arXiv.1901.02860.
- 467 [7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Naacl-hlt 2019. page 4171–4186, Minneapolis, Minnesota, 06 2019. Association  
 468 for Computational Linguistics. doi: 10.18653/v1/N19-1423. URL <https://aclanthology.org/N19-1423>.
- 469 [8] Andrew Gelman and Eric Loken. The statistical crisis in science. *American Scientist*, 102(6):460–465, 12 2014. URL <https://www.proquest.com/docview/1616141998/abstract/5E050DCE82414037PQ/1>. Num Pages: 6 Place: Research Triangle Park, United States Publisher: Sigma XI-The Scientific

- 469 Research Society.  
 470 [9] Elliot Gould, Hannah S. Fraser, Timothy H. Parker, Shinichi Nakagawa, Simon C. Griffith, Peter A. Vesk, Fiona Fidler, Daniel G. Hamilton, Robin N.  
 471 Abbey-Lee, Jessica K. Abbott, Luis A. Aguirre, Carles Alcaraz, Irith Aloni, Drew Altschul, Kunal Arekar, Jeff W. Atkins, Joe Atkinson, Christopher M.  
 472 Baker, Meghan Barrett, Kristian Bell, Suleiman Kehinde Bello, Iván Beltrán, Bernd J. Berauer, Michael Grant Bertram, Peter D. Billman, Charlie K.  
 473 Blake, Shannon Blake, Louis Blairst, Andrea Bonisoli-Alquati, Timothée Bonnet, Camille Nina Marion Bordes, Aneesh P. H. Bose, Thomas Botterill-  
 474 James, Melissa Anna Boyd, Sarah A. Boyle, Tom Bradfer-Lawrence, Jennifer Bradham, Jack A. Brand, Martin I. Brengdahl, Martin Bulla, Luc Bussière,  
 475 Ettore Camerlenghi, Sara E. Campbell, Leonardo L. F. Campos, Anthony Caravaggi, Pedro Cardoso, Charles J. W. Carroll, Therese A. Catanach,  
 476 Xuan Chen, Heung Ying Janet Chik, Emily Sarah Choy, Alec Philip Christie, Angela Chuang, Amanda J. Chunco, Bethany L. Clark, Andrea Contina,  
 477 Garth A. Covernton, Murray P. Cox, Kimberly A. Cressman, Marco Crotti, Connor Davidson Crouch, Pietro B. D'Amelio, Alexandra Allison  
 478 de Sousa, Timm Fabian Döbert, Ralph Dobler, Adam J. Dobson, Tim S. Doherty, Szymon Marian Drobniak, Alexandra Grace Duffy, Alison B. Duncan,  
 479 Robert P. Dunn, Jamie Dunning, Trishna Dutta, Luke Eberhart-Hertel, Jared Alan Elmore, Mahmoud Medhat Elsherif, Holly M. English, David C.  
 480 Ensminger, Ulrich Rainer Ernst, Stephen M. Ferguson, Esteban Fernandez-Juricic, Thalita Ferreira-Arruda, John Fieberg, Elizabeth A. Finch, Evan A.  
 481 Fiorenza, David N. Fisher, Amélie Fontaine, Wolfgang Forstmeier, Yoan Fourcade, Graham S. Frank, Cathryn A. Freund, Eduardo Fuentes-Lillo,  
 482 Sara L. Gandy, Dustin G. Gannon, Ana I. Garcia-Cervigón, Alexis C. Garretson, Xuezhen Ge, William L. Geary, Charly Géron, Marc Gilles, Antje  
 483 Girndt, Daniel Glikman, Harrison B. Goldspiel, Dylan G. E. Gomes, Megan Kate Good, Sarah C. Goslee, J. Stephen Gosnell, Eliza M. Grames, Paolo  
 484 Grattan, Nicholas M. Grebe, Skye M. Greener, Maaike Griffioen, Daniel M. Griffith, Frances J. Griffith, Jake J. Grossman, Ali Güncan, Stef Haesen,  
 485 James G. Hagan, Heather A. Hager, Jonathan Philo Harris, Natasha Dean Harrison, Sarah Syedia Hasnain, Justin Chase Havird, Andrew J. Heaton,  
 486 María Laura Herrera-Chaustré, Tanner J. Howard, Bin-Yan Hsu, Fabiola Iannarilli, Esperanza C. Iranzo, Erik N. K. Iverson, Saheed Olaide Jimoh,  
 487 Douglas H. Johnson, Martin Johnson, Jesse Jorna, Tommaso Jucker, Martin Jung, Ineta Kačergytė, Oliver Kaltz, Alison Ke, Clint D. Kelly, Katharine  
 488 Keegan, Friedrich Wolfgang Keppeler, Alexander K. Killion, Dongmin Kim, David P. Kochan, Peter Korsten, Shan Kothari, Jonas Kuppler, Jillian M.  
 489 Kusch, Małgorzata Lagisz, Kristen Marianne Lalla, Daniel J. Larkin, Courtney L. Larson, Katherine S. Lauck, M. Elise Lauterbur, Alan Law, Don-Jean  
 490 Léandri-Breton, Jonas J. Lembrechts, Kiara L'Herpinere, Eva J. P. Lievens, Daniela Oliveira de Lima, Shane Lindsay, Martin Luquet, Ross MacLeod,  
 491 Kirsty H. Macphie, Kit Magellan, Magdalena M. Mair, Lisa E. Malm, Stefano Mammola, Caitlin P. Mandeville, Michael Manhart, Laura Milena  
 492 Manrique-Garzon, Elina Mäntylä, Philippe Marchand, Benjamin Michael Marshall, Charles A. Martin, Dominic Andreas Martin, Jake Mitchell  
 493 Martin, April Robin Martinig, Erin S. McCallum, Mark McCauley, Sabrina M. McNew, Scott J. Meiners, Thomas Merkling, Marcus Michelangeli,  
 494 Maria Moiron, Bruno Moreira, Jennifer Mortensen, Benjamin Mos, Taofeek Olatunbosun Muraina, Penelope Wren Murphy, Luca Nelli, Petri  
 495 Niemelä, Josh Nightingale, Gustav Nilsson, Sergio Nolazco, Sabine S. Nooten, Jessie Lanterman Novotny, Agnes Birgitta Olin, Chris L. Organ,  
 496 Kate L. Ostevik, Facundo Xavier Palacio, Matthieu Paquet, Darren James Parker, David J. Pascall, Valerie J. Pasquarella, John Harold Paterson, Ana  
 497 Payo-Payo, Karen Marie Pedersen, Grégoire Perez, Kayla I. Perry, Patrice Pottier, Michael J. Proulx, Raphaël Proulx, Jessica L. Pruitt, Veronarindra  
 498 Ramananjato, Finaritra Tolotra Randimbirison, Onja H. Razafindratsima, Diana J. Rennison, Federico Riva, Sepand Riyahi, Michael James Roast,  
 499 Felipe Pereira Rocha, Dominique G. Roche, Cristian Román-Palacios, Michael S. Rosenberg, Jessica Ross, Freya E. Rowland, Deusdedith Rugemalila,  
 500 Avery L. Russell, Suvi Ruuskanen, Patrick Saccone, Asaf Sadeh, Stephen M. Salazar, Kris Sales, Pablo Salmón, Alfredo Sánchez-Tójar, Leticia Pereira  
 501 Santos, Francesca Santostefano, Hayden T. Schilling, Marcus Schmidt, Tim Schmoll, Adam C. Schneider, Allie E. Schrock, Julia Schroeder, Nicolas  
 502 Schtickzelle, Nick L. Schultz, Drew A. Scott, Michael Peter Scroggie, Julie Teresa Shapiro, Nitika Sharma, Caroline L. Shearer, Diego Simón, Michael I.  
 503 Sitvarin, Fabricio Luiz Skupien, Heather Lea Slinn, Grania Polly Smith, Jeremy A. Smith, Rahel Sollmann, Kaitlin Stack Whitney, Shannon Michael  
 504 Still, Erica F. Stuber, Guy F. Sutton, Ben Swallow, Conor Claverie Taff, Elina Takola, Andrew J. Tanentzap, Rocío Tarjuelo, Richard J. Telford,  
 505 Christopher J. Thawley, Hugo Thierry, Jacqueline Thomson, Svenja Tidau, Emily M. Tompkins, Claire Marie Tortorelli, Andrew Trlica, Biz R.  
 506 Turnell, Lara Urban, Stijn Van de Vondel, Jessica Eva Megan van der Wal, Jens Van Eeckhoven, Francis van Oordt, K. Michelle Vanderwel, Mark C.  
 507 Vanderwel, Karen J. Vanderwolf, Juliana Vélez, Diana Carolina Vergara-Florez, Brian C. Verrelli, Marcus Vinícius Vieira, Nora Villamil, Valerio  
 508 Vitali, Julien Vollering, Jeffrey Walker, Xanthe J. Walker, Jonathan A. Walter, Paweł Waryszak, Ryan J. Weaver, Ronja E. M. Wedegärtner, Daniel L.  
 509 Weller, and Shannon Whelan. Same data, different analysts: variation in effect sizes due to analytical decisions in ecology and evolutionary biology.  
 510 *BMC Biology*, 23(1):35, 02 2025. doi: 10.1186/s12915-024-02101-x. URL <https://doi.org/10.1186/s12915-024-02101-x>.
- 511 [10] Nick Huntington-Klein, Andreu Arenas, Emily Beam, Marco Bertoni, Jeffrey R. Bloem, Pralhad Burli, Naibin Chen, Paul Grieco, Godwin  
 512 Ekpe, Todd Pugatch, Martin Saavedra, and Yaniv Stopnitzky. The influence of hidden researcher decisions in applied microeconomics.  
 513 *Economic Inquiry*, 59(3):944–960, 2021. doi: 10.1111/ecin.12992. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/ecin.12992>. \_eprint:  
 514 <https://onlinelibrary.wiley.com/doi/pdf/10.1111/ecin.12992>.
- 515 [11] Eunice Jun, Melissa Birchfield, Nicole De Moura, Jeffrey Heer, and René Just. Hypothesis formalization: Empirical findings, software limitations,  
 516 and design implications. *ACM Transactions on Computer-Human Interaction*, 29(1):1–28, 02 2022. doi: 10.1145/3476980. URL <https://dl.acm.org/doi/10.1145/3476980>.
- 517 [12] Alex Kale, Matthew Kay, and Jessica Hullman. Decision-making under uncertainty in research synthesis: Designing for the garden of forking  
 518 paths. *CHI '19*, page 1–14, New York, NY, USA, 05 2019. Association for Computing Machinery. doi: 10.1145/3290605.3300432. URL <https://dl.acm.org/doi/10.1145/3290605.3300432>.
- 519 [13] Klea Katsouyanni, Giota Touloumi, Evangelia Samoli, Alexandros Gryparis, Alain Le Tertre, Yannis Monopolis, Giuseppe Rossi, Denis Zmirou,  
 520 Ferran Ballester, Azedine Boumghar, Hugh Ross Anderson, Bogdan Wojtyniak, Anna Paldy, Rony Braunstein, Juha Pekkanen, Christian Schindler,  
 521 and Joel Schwartz. Confounding and effect modification in the short-term effects of ambient particles on total mortality: Results from 29 european  
 522 cities within the aphea2 project. *Epidemiology*, 12(5):521, 09 2001. URL [https://journals.lww.com/epidem/fulltext/2001/09000/confounding\\_and\\_](https://journals.lww.com/epidem/fulltext/2001/09000/confounding_and_)

- 521           effect\_modification\_in\_the.11.aspx.
- 522 [14] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. Biobert: a pre-trained biomedical  
523 language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240, 02 2020. doi: 10.1093/bioinformatics/btz682. URL  
524 <https://academic.oup.com/bioinformatics/article/36/4/1234/5566506>.
- 525 [15] Yang Liu, Tim Althoff, and Jeffrey Heer. Paths explored, paths omitted, paths obscured: Decision points & selective reporting in end-to-end  
526 data analysis. CHI '20, page 1–14, New York, NY, USA, 04 2020. Association for Computing Machinery. doi: 10.1145/3313831.3376533. URL  
527 <https://dl.acm.org/doi/10.1145/3313831.3376533>.
- 528 [16] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov.  
529 Roberta: A robustly optimized bert pretraining approach. doi: 10.48550/arXiv.1907.11692.
- 530 [17] Bart Ostro, Rachel Broadwin, Shelley Green, Wen-Ying Feng, and Michael Lipsett. Fine particulate air pollution and mortality in nine California  
531 counties: Results from calfine. *Environmental Health Perspectives*, 114(1):29–33, 01 2006. doi: 10.1289/ehp.8335. URL <https://ehp.niehs.nih.gov/doi/10.1289/ehp.8335>. Publisher: Environmental Health Perspectives.
- 532 [18] Marko Sarstedt, Susanne J. Adler, Christian M. Ringle, Gyeongcheol Cho, Adamantios Diamantopoulos, Heungsun Hwang, and Benjamin D.  
533 Liengaard. Same model, same data, but different outcomes: Evaluating the impact of method choices in structural equation modeling. *Journal of  
534 Product Innovation Management*, 41(6):1100–1117, 2024. doi: 10.1111/jipim.12738. URL [https://onlinelibrary.wiley.com/doi/abs/10.1111/jipim.12738.\\_eprint](https://onlinelibrary.wiley.com/doi/abs/10.1111/jipim.12738._eprint): <https://onlinelibrary.wiley.com/doi/pdf/10.1111/jipim.12738>.
- 535 [19] R. Silberzahn, E. L. Uhlmann, D. P. Martin, P. Anselmi, F. Aust, E. Awtry, Š. Bahník, F. Bai, C. Bannard, E. Bonnier, R. Carlsson, F. Cheung,  
536 G. Christensen, R. Clay, M. A. Craig, A. Dalla Rosa, L. Dam, M. H. Evans, I. Flores Cervantes, N. Fong, M. Gamez-Djokic, A. Glenz, S. Gordon-McKeon,  
537 T. J. Heaton, K. Hederos, M. Heene, A. J. Hofelich Mohr, F. Högden, K. Hui, M. Johannesson, J. Kalodimos, E. Kaszubowski, D. M. Kennedy, R. Lei,  
538 T. A. Lindsay, S. Liverani, C. R. Madan, D. Molden, E. Molleman, R. D. Morey, L. B. Mulder, B. R. Nijstad, N. G. Pope, B. Pope, J. M. Prenoveau, F. Rink,  
539 E. Robusto, H. Roderique, A. Sandberg, E. Schlüter, F. D. Schönbrodt, M. F. Sherman, S. A. Sommer, K. Sotak, S. Spain, C. Spörlein, T. Stafford,  
540 L. Stefanutti, S. Tauber, J. Ullrich, M. Vianello, E.-J. Wagenmakers, M. Witkowiak, S. Yoon, and B. A. Nosek. Many analysts, one data set: Making  
541 transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science*, 1(3):337–356, 09 2018. doi:  
542 10.1177/2515245917747646. URL <https://doi.org/10.1177/2515245917747646>. Publisher: SAGE Publications Inc.
- 543 [20] Jan Simson, Fiona Draxler, Samuel Mehr, and Christoph Kern. Preventing harmful data practices by using participatory input to navigate the  
544 machine learning multiverse. CHI '25, page 1–30, New York, NY, USA, 04 2025. Association for Computing Machinery. doi: 10.1145/3706598.3713482.  
545 URL <https://dl.acm.org/doi/10.1145/3706598.3713482>.
- 546 [21] Hadley Wickham. Tidy data. *Journal of Statistical Software*, 59:1–23, 09 2014. doi: 10.18637/jss.v059.i10. URL <https://doi.org/10.18637/jss.v059.i10>.
- 547 [22] Hadley Wickham, Joe Cheng, and Aaron Jacobs. *ellmer: Chat with Large Language Models*, 2025. URL <https://CRAN.R-project.org/package=ellmer>.  
R package version 0.1.1.
- 548 [23] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, and Quoc V. Le. Xlnet: Generalized autoregressive pretraining for  
549 language understanding. doi: 10.48550/arXiv.1906.08237.
- 550
- 551
- 552
- 553
- 554
- 555
- 556
- 557
- 558
- 559
- 560
- 561
- 562
- 563
- 564
- 565
- 566
- 567
- 568
- 569
- 570
- 571
- 572