The Book of Why
"For Epidemiologists"

George Davey Smith
MRC Integrative Epidemiology Unit
University of Bristol

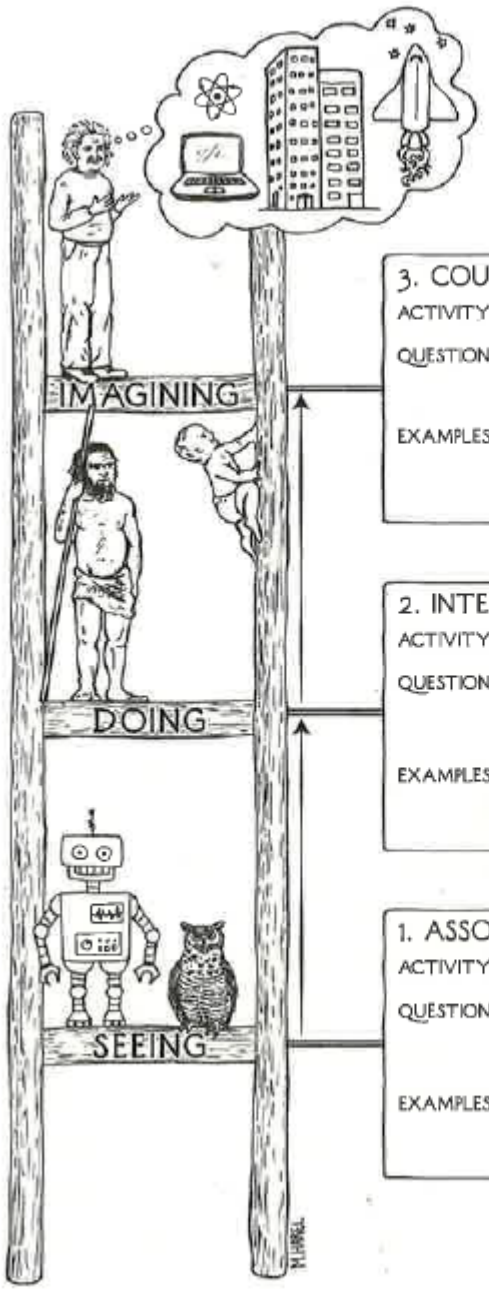
@mendel_random

University of BRISTOL

MRC Integrative Epidemiology Unit

Structure of talk

- Value of DAG theory to epidemiology
- The reality of use of DAGs in epidemiology
- Getting Wright wrong
- Where does “background knowledge” come from?
- Consequences of believing the DAGs



3. COUNTERFACTUALS

ACTIVITY: Imagining, Retrospection, Understanding

QUESTIONS: *What if I had done ...? Why?*
(Was it X that caused Y? What if X had not occurred? What if I had acted differently?)

EXAMPLES: Was it the aspirin that stopped my headache? Would Kennedy be alive if Oswald had not killed him? What if I had not smoked for the last 2 years?

2. INTERVENTION

ACTIVITY: Doing, Intervening

QUESTIONS: *What if I do ...? How?*
(What would Y be if I do X?
How can I make Y happen?)

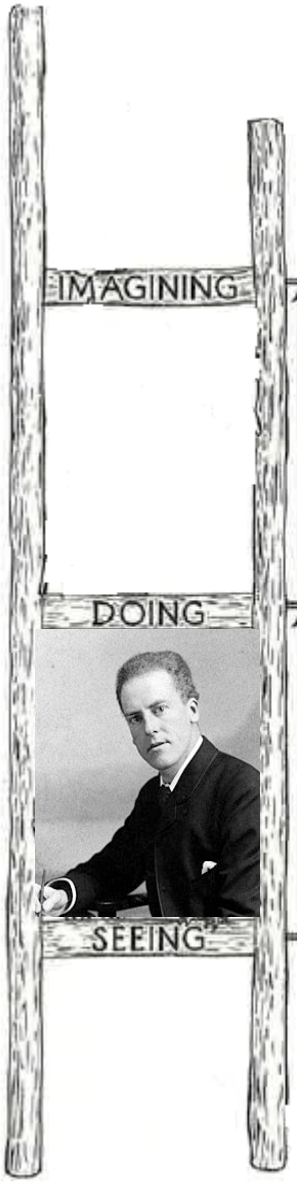
EXAMPLES: If I take aspirin, will my headache be cured?
What if we ban cigarettes?

1. ASSOCIATION

ACTIVITY: Seeing, Observing

QUESTIONS: *What if I see ...?*
(How are the variables related?
How would seeing X change my belief in Y?)

EXAMPLES: What does a symptom tell me about a disease?
What does a survey tell us about the election results?

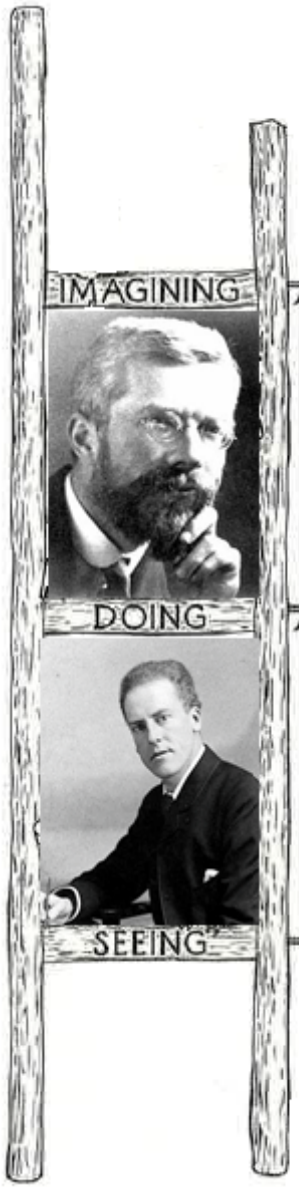


1. ASSOCIATION

ACTIVITY: Seeing, Observing

QUESTIONS: *What if I see ...?*
(How are the variables related?
How would seeing X change my belief in Y?)

EXAMPLES: What does a symptom tell me about a disease?
What does a survey tell us about the election results?



2. INTERVENTION

ACTIVITY: Doing, Intervening

QUESTIONS: *What if I do ...? How?*
(What would Y be if I do X?
How can I make Y happen?)

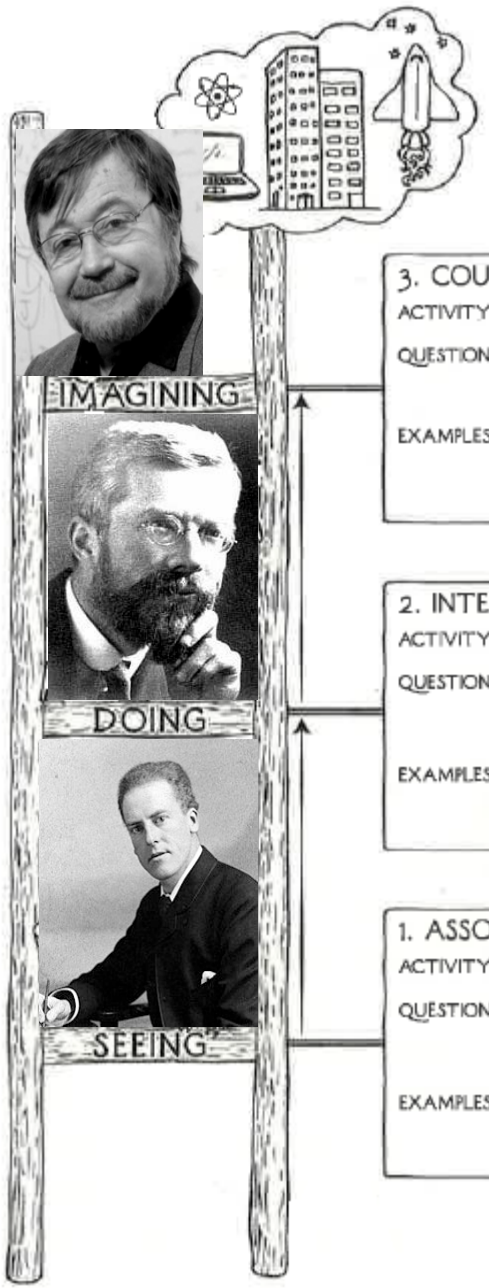
EXAMPLES: If I take aspirin, will my headache be cured?
What if we ban cigarettes?

1. ASSOCIATION

ACTIVITY: Seeing, Observing

QUESTIONS: *What if I see ...?*
(How are the variables related?
How would seeing X change my belief in Y?)

EXAMPLES: What does a symptom tell me about a disease?
What does a survey tell us about the election results?



3. COUNTERFACTUALS

ACTIVITY: Imagining, Retrospection, Understanding

QUESTIONS: *What if I had done ...? Why?*
(Was it X that caused Y? What if X had not occurred? What if I had acted differently?)

EXAMPLES: Was it the aspirin that stopped my headache? Would Kennedy be alive if Oswald had not killed him? What if I had not smoked for the last 2 years?

2. INTERVENTION

ACTIVITY: Doing, Intervening

QUESTIONS: *What if I do ...? How?*
(What would Y be if I do X?
How can I make Y happen?)

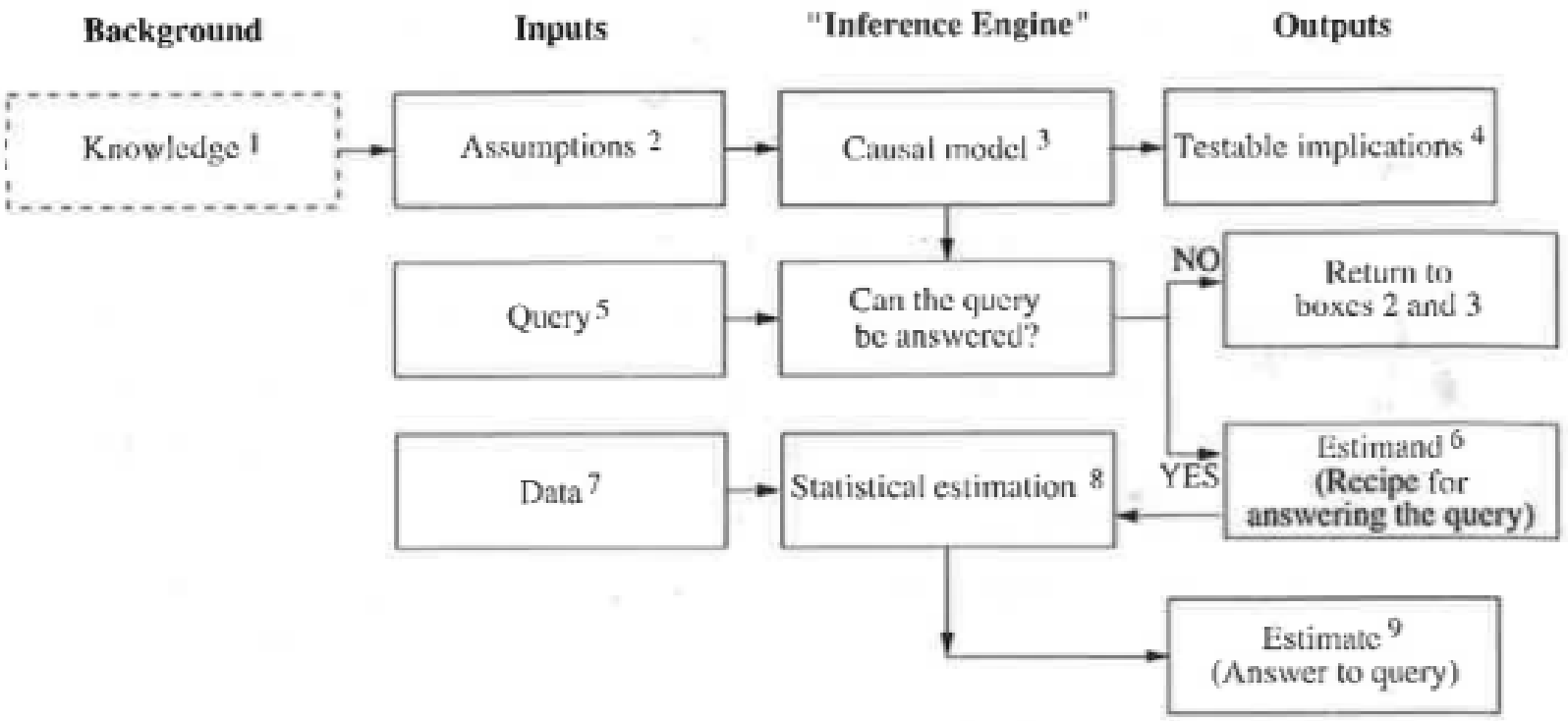
EXAMPLES: If I take aspirin, will my headache be cured?
What if we ban cigarettes?

1. ASSOCIATION

ACTIVITY: Seeing, Observing

QUESTIONS: *What if I see ...?*
(How are the variables related?
How would seeing X change my belief in Y?)

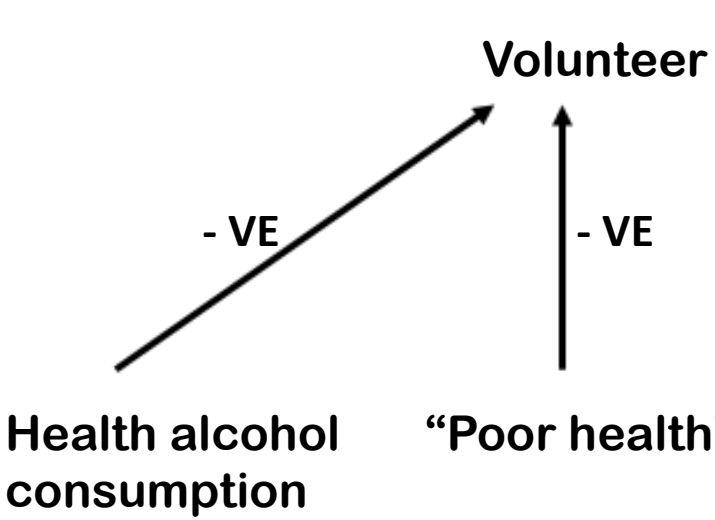
EXAMPLES: What does a symptom tell me about a disease?
What does a survey tell us about the election results?

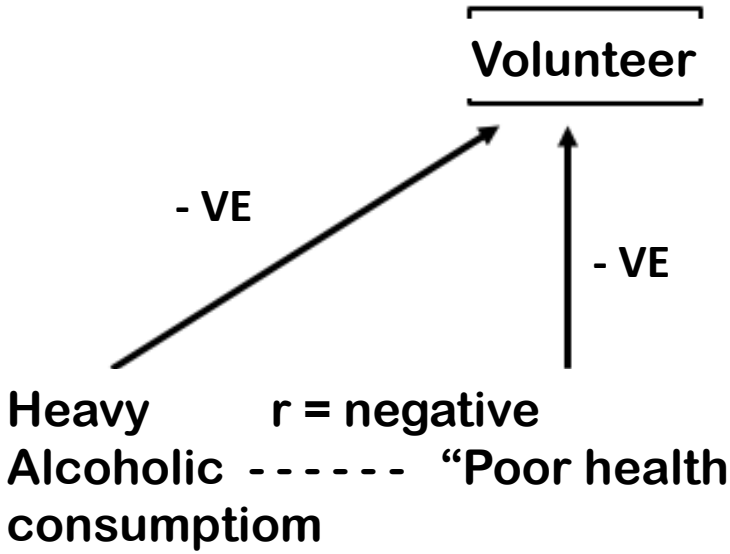

Background
Inputs
"Inference Engine"
Outputs
Knowledge 1
Assumptions 2
Causal model 3
Testable implications 4
Query 5
Can the query be answered?
NO
Return to boxes 2 and 3
YES
Estimand 6 (Recipe for answering the query)
Data 7
Statistical estimation 8
Estimate 9 (Answer to query)

Unequivocal gains to epidemiology from employing DAGs

- Structure of biases

Heavy alcohol consumption "protects" against stroke in the American Cancer Society volunteer cohort

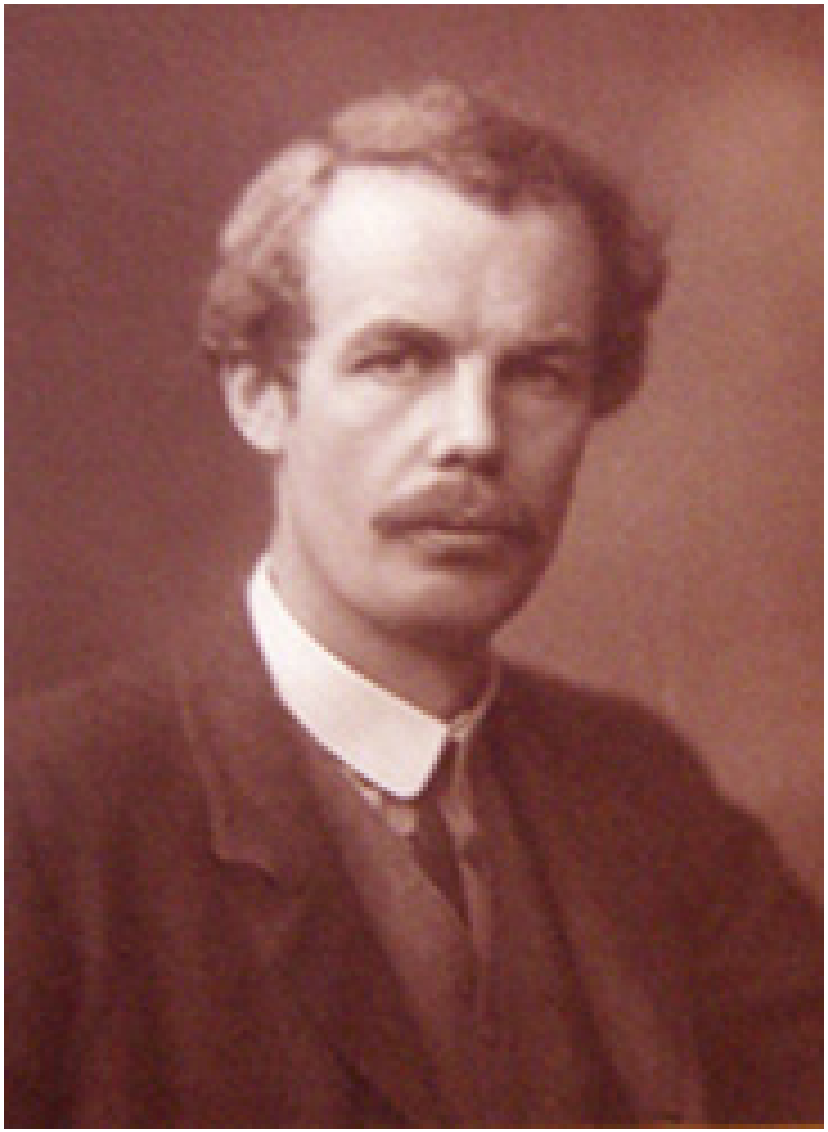




Ebrahim S, Davey Smith G. Should we always deliberately be non-representative? Int. J. Epidemiol. 2013;42:1022-1026.

In observational studies associations between an exposure and disease will generally be biased if there is selection according to an exposure–disease combination in case-control studies, or according to an exposure–disease risk combination in prospective studies. If, for example, people with an exposure and at low risk of disease for other reasons were differentially excluded from a study the exposure would appear to be positively related to disease outcome, even if there were no such association in the underlying population. This is a form of 'Berkson's bias', well known to epidemiologists.[14] A possible example of such associative selection bias relates to the finding in the large American Cancer Society volunteer cohort that high alcohol consumption was associated with a reduced risk of stroke.[15] This is somewhat counter-intuitive as the outcome category included haemorrhagic stroke (for which there is no obvious mechanism through which alcohol would reduce risk) and because alcohol is known to increase blood pressure[16,17]—a major causal factor for stroke.[18] Population-based studies have found that alcohol tends to increase stroke risk.[19–21] Heavy drinkers who volunteer for a study known to be about the health effects of their lifestyle are likely to be very unrepresentative of all heavy drinkers in the population, in ways that render them to be at low risk of stroke. Moderate and non-drinkers who volunteer may be more representative of moderate and non-drinkers in the underlying population. Thus the low risk of stroke in the heavy drinkers who volunteer for the study could erroneously make it appear that alcohol reduces the risk of stroke.

Davey Smith and Ebrahim IJE 2004



Arthur Cecil Pigou
1877 - 1959

Not the “fourth man” …

ALCOHOLISM AND HEREDITY.

By Professor A. C. Pigou.

In the middle of last year Professor Karl Pearson and Miss Elderton published a memoir entitled "A First Study of the Influence of Parental Alcoholism on the Physique and Ability of the Offspring." Since that time a violent controversy has taken place in the *Times*, the *British Medical Journal*, the *Journal of the Royal Statistical Society*, and elsewhere upon this matter. The controversy has covered so large an area and has been concerned to so great an extent with points of detail that the broad issue of chief interest to practical reformers has perhaps become somewhat obscured. That issue may, I conceive, be stated thus: "Would the introduction of a law leading to a diminution in the amount of parental alcoholism in any generation be likely to bring about an improvement in the physique and ability of the succeeding generation?" It is believed by some that this question can be answered in the negative as a result of statistical research into the comparative physique and ability of the descendants of alcoholic and non-alcoholic parents respectively in some carefully chosen district. This view is not, I should say, one which can fairly be imputed to Professor Pearson and Miss Elderton; and the paragraphs that follow are not intended as an attack on their work. My purpose is not to enter at this late stage into the controversy between these authors and their critics, but to examine directly the belief which I have formulated above.

Let us suppose, then, that a district has been chosen, and that statistical research of the kind contemplated has been undertaken. There can be no doubt that that research is competent to give immediately one very interesting piece of information. It can tell us whether or not there is any marked correlation between parental alcoholism and inferiority of offspring *in the selected district*; or, to put the point more broadly, whether on the whole the children of alcoholic parents in that district are inferior to the children of non-alcoholic parents there. Let us suppose that our local investigation has been so far extended as to take account of the condition of these children, not merely during their childhood, but throughout the whole of their life. Let us suppose, finally, that no appreciable correlation between parental alcoholism and inferiority of offspring is found, but that the children of the two groups of parents prove to be substantially indistinguishable. The problem is: Will this

Pigou AC. Alcoholism and Heredity. Westminster Gazette 2nd February 1911 reprinted in Int J Epidemiol.

International Journal of Epidemiology, 2017, 1–2
doi: 10.1093/ije/dyw340

Alcoholism and Heredity

By Professor A.C. Pigou[1]

[1]Pigou, A. C. 'Alcoholism and Heredity'. *Westminster Gazette*, 2nd February 1911. Permission to reproduce by Solo Syndication.

In the middle of last year Professor Karl Pearson and Miss Elderton published a memoir entitled "A First Study of the Influence of Parental Alcoholism on the Physique and Ability of the Offspring." Since that time a violent controversy has taken place in the *Times*, the *British Medical Journal*, the *Journal of the Royal Statistical Society*, and elsewhere upon this matter. The controversy has covered so large an area and has been concerned to so great an extent with points of detail that the broad issue of chief interest to practical reformers has perhaps become somewhat obscured. That issue may, I conceive, be stated thus: "Would the introduction of a law leading to a diminution in the amount of parental alcoholism in any generation be likely to bring about an improvement in the physique and ability of the succeeding generation?" It is believed by some between parental alcoholism and inferiority of offspring is found, but that the children of the two groups of parents prove to be substantially indistinguishable. The problem is: Will this result enable us to say that State action reducing alcoholism among parents in one generation is not likely to benefit the physique and ability of the succeeding generation? There are, I submit, two distinct sets of considerations, both of which compel us to answer this question in the negative.

The first consideration arises out of the fact that our statistical research has extended, not over the whole country, but only over a small part of it. Are we reasonably entitled to conclude that what is true of the part will probably also be true of the whole? I hold that we are not entitled to do this, for the following reason. It often happens that the

Unequivocal gains to epidemiology from employing DAGs

- Structure of biases .. and making these transportable

According to Roy Epstein [1987], Wright once gave a seminar on path coefficients to the Cowles Commission (the breeding ground for SEM) in the 1940s, but neither side saw particular merit in the other's methods. Why? After all, a diagram is nothing but a set of nonparametric structural equations in which, to avoid confustion, the equality signs are replaced with arrows.

My explanation is that early econometricians were extremely careful mathematicians; they thought they could keep the mathematics in purely equational-statistical form and just reason about structure in their heads. Indeed, they managed to do so surprisingly well, because they were truly remarkable individuals and *could* do it in their heads.

Pearl J. The new challenge, 1997

...for the purpose of predicting mortality from birthweight, specification in terms of absolute weight does not mean the same for male infants as for females; at any given weight the infant in the series with lower mean weight (females) will have, relative to males, a smaller proportion of members whose weight is reduced by those factors that are associated with increased mortality, and the group will consequently have a more favorable mortality rate. Similarly, for the offspring of smokers, if their weight is reduced but their over-all mortality unaffected, then at any given weight they will, relative to the offspring of nonsmokers, have a lower component of infants premature by gestation, and consequently, a more favorable mortality rate.

MacMahon B et al. Infant weight and parental smoking habits. Am J Epidemiol. 1965 Nov;82(3):247-61.

Unequivocal gains to epidemiology from employing DAGs

- Structure of biases .. and making these transportable
- Providing an explicit rationale for constructing adjustment sets

Unequivocal gains to epidemiology from employing DAGs

- Structure of biases .. and making these transportable
- Providing an explicit rationale for constructing adjustment sets
- Lead to an explicit presentation of some of the assumptions the researcher holds

Unequivocal gains to epidemiology from employing DAGs

- Structure of biases .. and making these transportable
- Providing an explicit rationale for constructing adjustment sets
- Lead to an explicit presentation of some of the assumptions the researcher holds
- Contributing to methodological developments with abstract DAGs

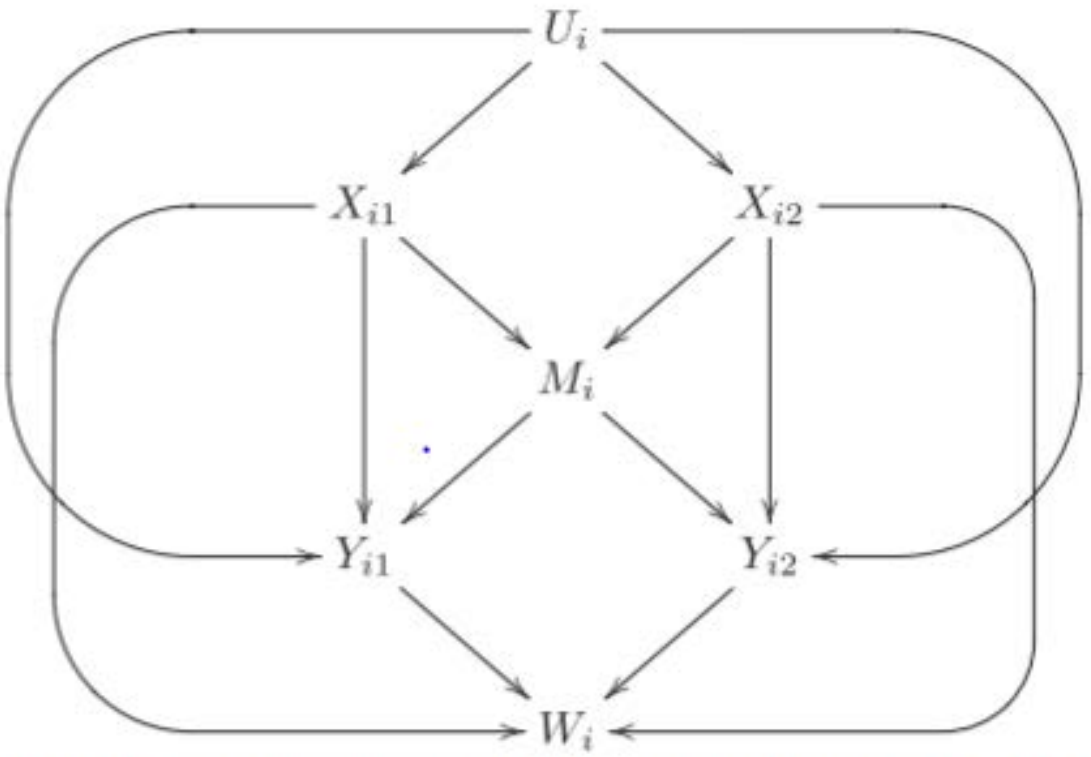


FIGURE. A causal diagram illustrating shared confounders (U_i), mediators (M_i), and colliders (W_i) in a sibling comparison design.

X = exposure
Y = outcome
U= eg parental genotype

Family environment may be M or W

Sjolander et al, Cofounders, mediators or colliders: what types of shared covariates does a sibling comparison design control for? Epidemiology 2017;28:540-7

The reality of the use of DAGs in Epidemiology

Of 388 individuals randomized to the Internet option, only 7 completed the survey online; the remaining 81 completed the survey on paper and mailed back the form. The small number completing the survey online prohibits us from making any conclusions about the online respondents.

The overall decrease in response rate when offering an Internet option in mailed surveys is consistent with other recent studies.[1,5] Finding no difference in response rates for adults aged 25–34 years is consistent with a Norwegian trial that saw no difference for persons aged 25–40.[4] For older participants, including an online option could cause some potential respondents to put aside the task to complete later online, and never reengage.[5] We concur with the recommendation by Dillman et al[1] that availability of multiple response modes is likely to bring about more harm than good.

Jeanette Y. Ziegenfuss
Timothy J. Beebe
Division of Health Care Policy and Research
Department of Health Sciences Research
Survey Research Center
ziegenfuss.jeanette@mayo.edu

Enrique Rey
Division of Digestive Diseases
Hospital Clinico San Carlos
Complutense University
Madrid, Spain

Cathy Schleck
Division of Biomedical Statistics and Informatics
Mayo Clinic
Rochester, MN

2. Friedman EM, Clausen NA, Hartzell M. The net effect: A comparison of internet and mail survey respondents. In: *Statistics as a Unified Discipline*. 2004:3519–3524. Available at: http://www.mathematica-mpr.com/publications/pdfs/internetmailsurvey.pdf.
3. Elliott MN, Edwards C, Angeles J, Hambarsoomians K, Hays RD. Patterns of unit and item nonresponse in the CAHPS® hospital survey. *Health Serv Res*. 2005;40(6 pt 2):2096–2119.
4. Brogger J, Nystad W, Cappelen I, Bakke P. No increase in response rate by adding a web response option to a postal population survey: A randomized trial. *J Med Internet Res*. 2007; 9:e40.
5. Griffin DH, Fischer DP, Morgan MT. Testing an internet response option for the American Community Survey at US Department of Commerce. Paper presented at: The 2001 AAPOR Conference; 2001; Montreal, Quebec, Canada.
6. Talley NJ, Phillips SF, Melton J III, Wiltgen C, Zinsmeister AR. A patient questionnaire to identify bowel disease. *Ann Intern Med*. 1989; 111:671–674.
7. Melton LJ III. The threat to medical-records research. *N Engl J Med*. 1997;337:1466–1470.

dagR

A Suite of R Functions for Directed Acyclic Graphs

To the Editors:

I have read with great interest the recent letter by Knüppel and Stang introducing a DOS program for assessing directed acyclic graphs (DAGs) with respect to minimal sufficient adjustment sets.[1] Others have elaborated on the value of DAGs for epidemiologists,[2]

The functions in essence implement the graphical rules outlined elsewhere.[2] After initializing a new DAG using a command line, the researcher can evaluate what associations are introduced by adjusting for covariables. Potentially biasing paths from exposure to outcome can be identified (see eFig. 1, http://links.lww.com/EDE/A395 demonstrating harmful adjustment using an example DAG from Fleischer and Diez Roux[3]). Functions to conveniently add or remove nodes and arcs are included, as is a function checking introduced associations and biasing paths for all possible adjustment sets (with options to force in adjustment or exclude unmeasured variables; see eFig. 2, http://links.lww.com/EDE/A395 for a re-evaluation of example 3 of the DAG program manual[1]), thereby allowing the identification of minimal sufficient adjustment sets. Note that the evaluation of all possible adjustment sets for more complex DAGs may be somewhat resource-intensive (see eFig. 3, http://links.lww.com/EDE/A395 for such a DAG motivated by Shrier and Platt[4]).

Some convenient features result from using R as the programming platform. dagR should run on all operating systems running R. DAGs generated and manipulated using dagR can be saved as an R object and transferred to another computer to continue manipulation and evaluation with dagR. The graphics capabilities of R allow fairly straightforward programming of basic DAG drawing routines, while also supporting the interactive repositioning of nodes and

Breitling LP. dagR: A Suite of R Functions for Directed Acyclic Graphs. Epidemiology 2010;21:586-587

DAGitty

A Graphical Tool for Analyzing Causal Diagrams

To the Editor:

Causal diagrams, also known as directed acyclic graphs,[1,2] provide an entirely graphical, yet mathematically rigorous methodology for minimizing bias in epidemiologic studies.[3,4] The analysis of causal diagrams can be cumbersome in practice, and lends itself well to automatization by a computer program. Important first steps in this regard include the development of the DAG program by Knüppel and Stang[5] and dagR by Breitling.[6] We announce the

to find MSA sets is to check each covariate set to see whether it is an MSA set. In a diagram with 50 covariates, this means that 2^{50} sets may have to be tested—a 16-digit number that is too large even for computers. To identify MSA sets more efficiently, we adapted an algorithm proposed recently for a related graph-theoretical problem.[8] This algorithm is guaranteed to output the list of MSA sets reasonably quickly (ie, in polynomial time per MSA set output). Note, however, that very large or very regularly structured diagrams could in theory have millions of different MSA sets. If such diagrams become practically relevant, further research will be necessary to develop appropriate computational methods for helping the user to choose appropriate MSA sets.

Sven Knüppel
Department of Epidemiology
German Institute of Human Nutrition
Potsdam-Rehbruecke
Nuthetal, Germany

REFERENCES

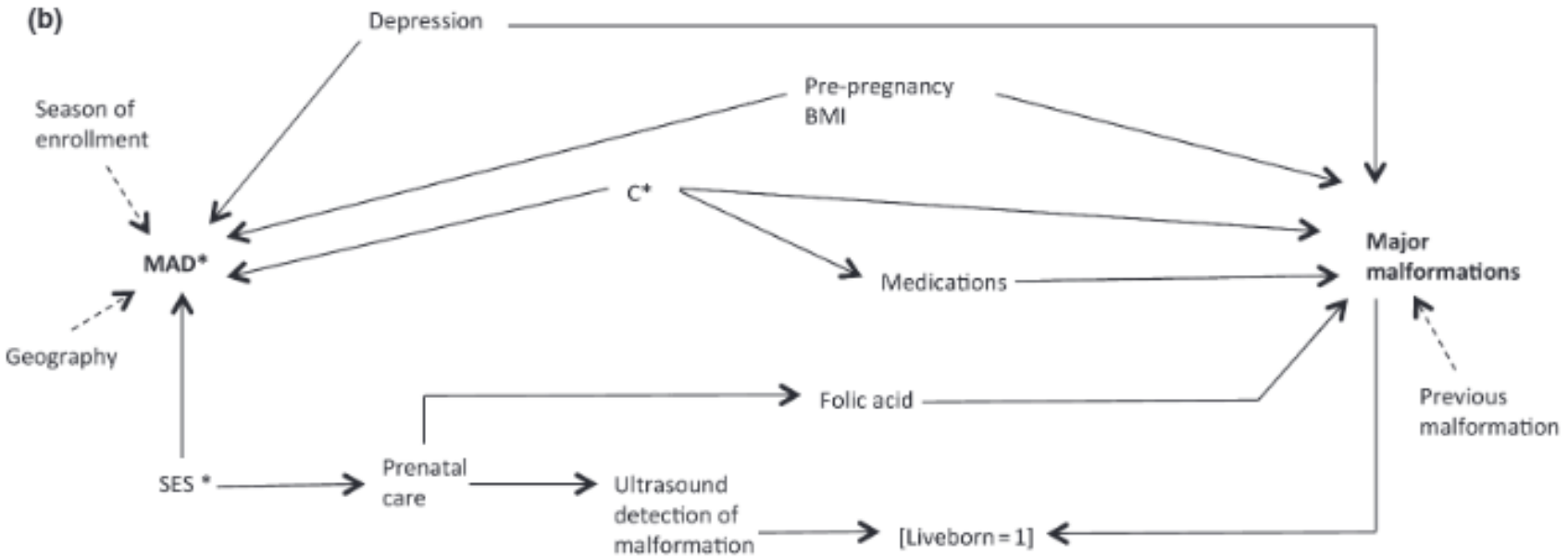
1. Pearl J. *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press; 2000.
2. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10:37–48.
3. Shrier I, Platt RW. Reducing bias through directed acyclic graphs. *BMC Med Res Methodol*. 2008;8:70.
4. Glymour MM, Greenland S. Causal diagrams. In: Rothman KJ, Greenland S, Lash TL, eds. *Modern Epidemiology*. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008:183–209.
5. Knüppel S, Stang A. DAG program: identifying minimal sufficient adjustment sets [Letter]. *Epidemiology*. 2010;21:159.
6. Breitling L. dagR: a suite of R functions for

Textor J et al. DAGitty: A graphical tool for analysing causal diagrams. Epidemiology 2011;22:745

"Directed Acyclic Graphs[1] and 10 percent change in estimate procedures were used to identify covariates for inclusion in multivariable models; these included age, education, living with a partner, parity, and history of preterm birth".

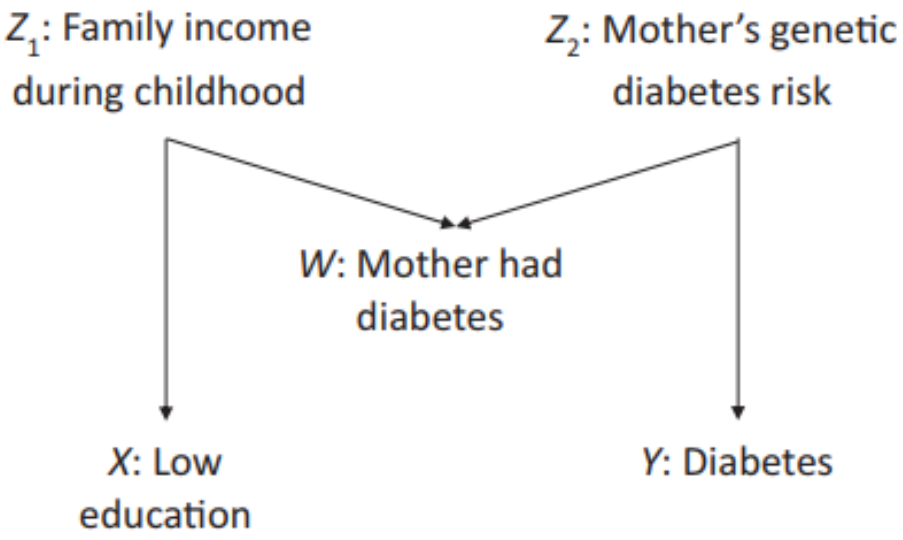
[1]. Textor J, Hardt J, Knuppel S. Dagitty: A graphical tool for analyzing causal diagrams. *Epidemiology* 2011;**22**(5):745.

Barcelona de Mendoza V et al. Acculturation and Intention to Breastfeed among a Population of Predominantly Puerto Rican Women. Birth 2016;43:78-85

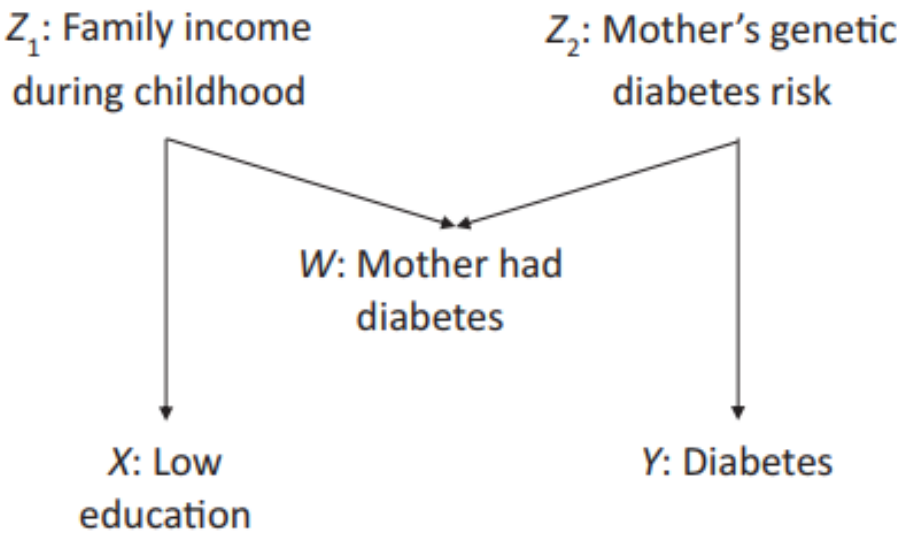


*MAD: maternal antidepressants; C: maternal age, race/ethnicity, alcohol or tobacco use, and chronic medical conditions; SES: socioeconomic status. Dashed lines denote paths that are not biasing paths; solid lines denote biasing paths. DAG assumes no direct effect of maternal antidepressants on major malformations.

Bandoli G et al. Constructing Causal Diagrams for Common Perinatal Outcomes: Benefits, Limitations and Motivating Examples with Maternal Antidepressant Use in Pregnancy. Paediatric and Perinatal Epidemiology 2016;30:521-528.



Glymour MM. Using causal diagrams to understand common problems in social epidemiology. In: Oakes JM, Kaufman JS (eds). Methods in Social Epidemiology. San Francisco, CA: Josey-Bass, 2006;393–428



"Under the graphical criteria, one should not include mother's diabetes status as a covariate"

Glymour MM. Using causal diagrams to understand common problems in social epidemiology. In: Oakes JM, Kaufman JS (eds). Methods in Social Epidemiology. San Francisco, CA: Josey-Bass, 2006;393–428

"A structural causal model provides a tool for understanding whether background knowledge, combined with the observed data, is sufficient to allow a causal question to be translated into a statistical estimand, and, if not, what additional data or assumptions are needed."

"In many cases, rigorous application of a formal causal framework forces us to conclude that existing knowledge and data are insufficient to claim identifiability—in itself a useful contribution."

Petersen ML et al. Causal Models and Learning from Data: Integrating Causal Modeling and Statistical Estimation. Epidemiology 2014;25:418-426.

But what of the assumptions of “causal DAGs” and causal modelling approaches?

No measurement error: the variables are measured without measurement error. This is a subtle assumption that is required to learn Causal Bayesian Networks (CBNs), often not realized by practitioners who apply these techniques.

Lagani V et al. Probabilistic Computational Causal Discovery for Systems Biology. In: Geris L, Gomez-Cabrero D (Eds). Uncertainty in Biology Volume 17 of the series. Studies in Mechanobiology, Tissue Engineering and Biomaterials pp 33-73; 2015

Oh yeah … and there's "no unmeasured confounding" too …

No measurement error

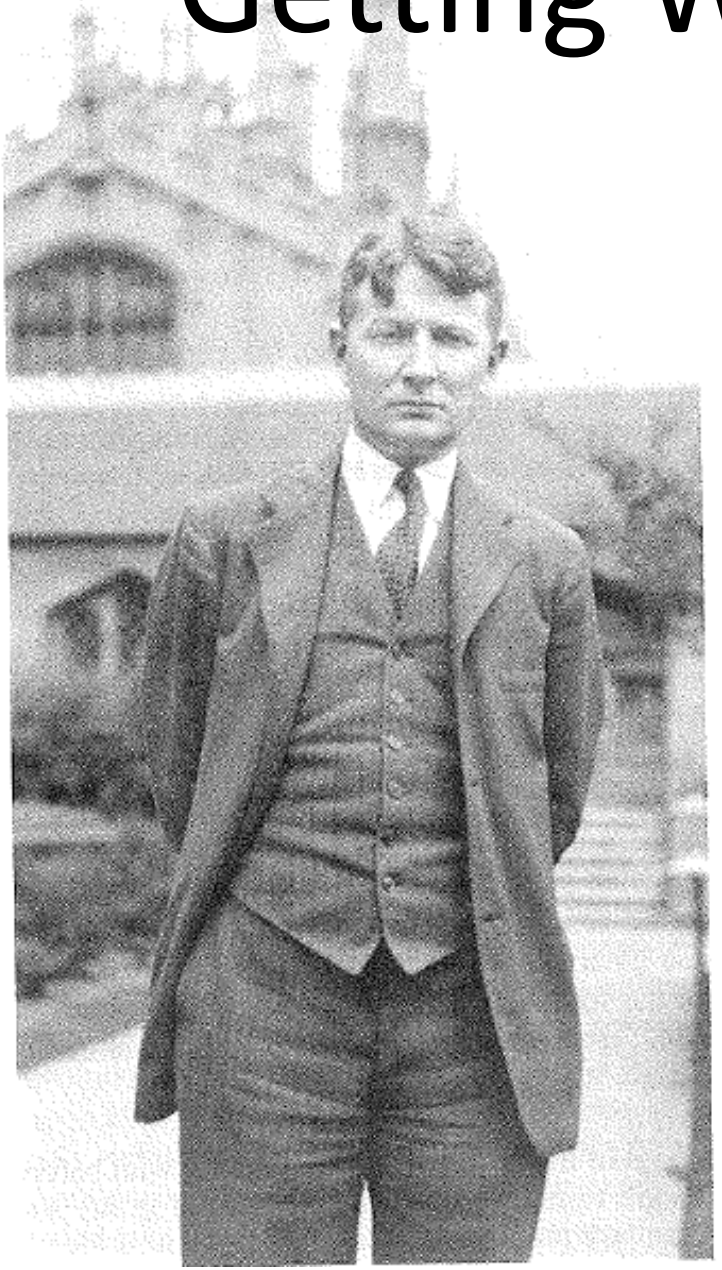
+

No unmeasured confounding

=

Not epidemiological data

Getting Wright Wrong



Sewall Wright on path analysis, causation and mediation

James Crow's NAS Biographical Memoir of Sewall Wright

"He read his father's math books and learned to extract cube roots before entering school, a skill that he said brought him instant, lasting unpopularity with the other students"

Wright's work was vigorously and explicitly discouraged by the thousand-pound gorillas of Pearson and Fisher. They stuck almost religiously to the positivistic ideology that all knowledge is sensory information, and sensory information can not encode causal connections.

Powell S. The Book of Why: The New Science of Cause and Effect. Journal of MultiDisciplinary Evaluation. 2018;14:47-54

Wright's work was vigorously and explicitly discouraged by the thousand-pound gorillas of Pearson and Fisher. They stuck almost religiously to the positivistic ideology that all knowledge is sensory information, and sensory information can not encode causal connections.

FROM "THE BOOK OF WHY"

".. A rebuttal published in 1921 by one Henry Niles, a student of American Statistician Raymond Pearl (no relation), who in turn was a student of Karl Pearson, the godfather of statistics"

Powell S. The Book of Why: The New Science of Cause and Effect. Journal of MultiDisciplinary Evaluation. 2018;14:47-54

THE RELATIVE IMPORTANCE OF HEREDITY AND ENVIRONMENT IN DETERMINING THE PIEBALD PATTERN OF GUINEA-PIGS

By Sewall Wright

Bureau of Animal Industry, United States Department of Agriculture

Communicated by R. Pearl, March 17, 1920

The authors also refer to the work of Barbara Burks (1926), who may have preceded Wright with the use of causal diagrams in particular in the study of mediation, but the uptake of her work suffered under the twin pressures of mainstream statistics and the prejudices against women in science in the early and mid-twentieth century.

Powell S. The Book of Why: The New Science of Cause and Effect. Journal of MultiDisciplinary Evaluation. 2018;14:47-54

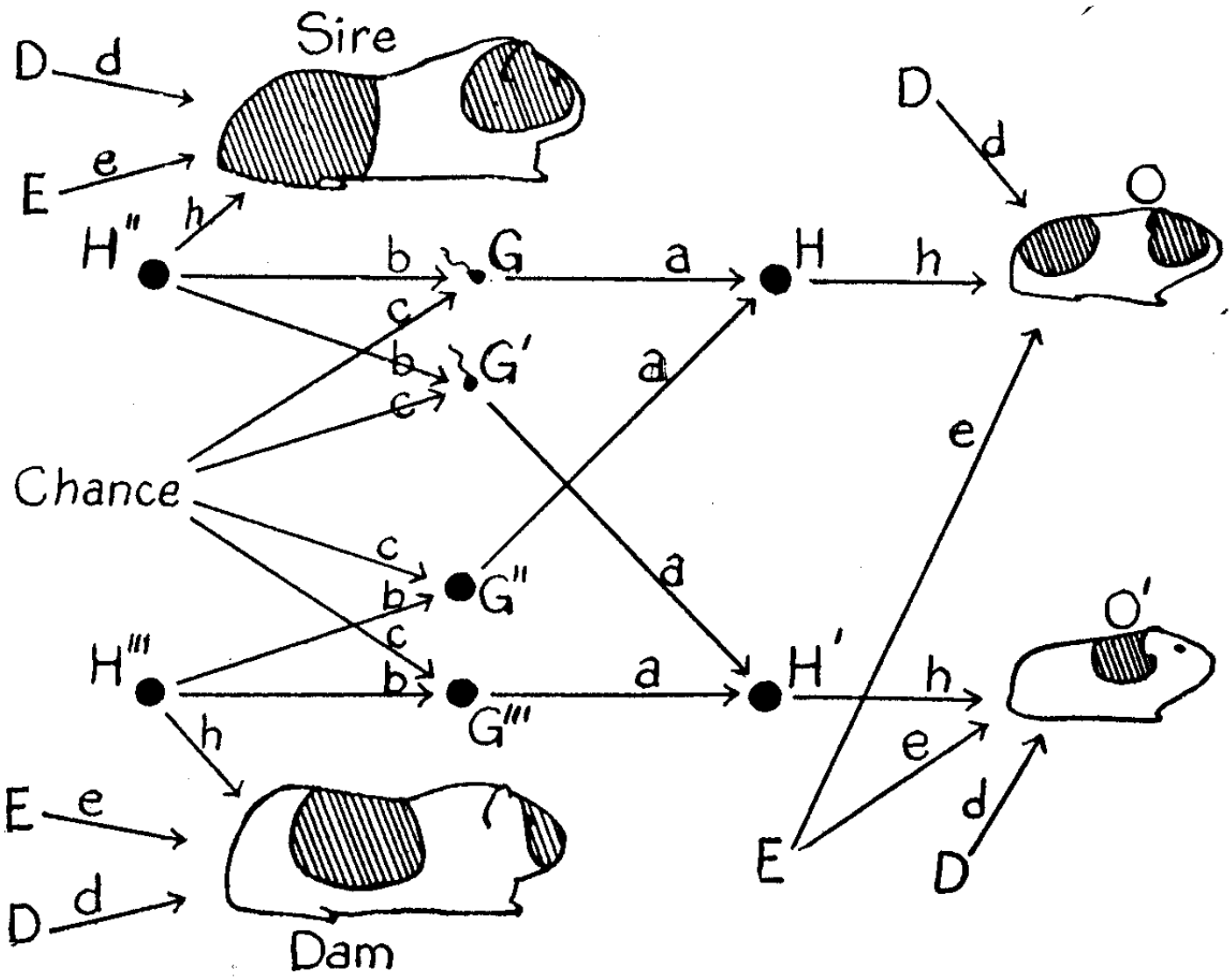
"A prominent SEM researcher once asked me, "Under what conditions can we give causal interpretation to identified structural coefficients?" I thought this colleague was joking. As a faithful reader of Wright (1921) and Haavelmo (1943), I had come to believe that the answer is simply, "Always!..."

Pearl J. TETRAD and SEM. Multivariate Behavioural Research 1998;33:119-128.

"The rate of decrease of heterozygosis in systems of mating more complicated than self-fertilization was first worked out from the recurrence relation between successive generations independently by Jennings (1914) and Fish (1914) for brother-sister mating and by Jennings (1916) for some others. The present writer, who had assisted Fish in his calculations, found a simpler way of finding this quantity, the method of path coefficients, based on the correlation between uniting gametes (Wright, 1921)."

Wright S. The Genetical Structure of Populations. Annals of Eugenics 1949;15:323-354.

Random phenotypic variance? Piebald pattern in guinea pigs



Sewall Wright 1921

It was of course realized that the "concrete, phenomenal actuality" of the results was not proved by the analysis by path coefficients. This rests on the validity of the premises, i.e., on the evidence for Mendelian heredity. The paper began with a quotation from EAST and JONES on the universality of Mendelian inheritance under sexual reproduction, as the justification for the analysis.

Wright S. The theory of path coefficients: A reply to Niles's criticism. Genetics 1923;8:239

And the same said in many, many other places

“The hypothesis that heredity is Mendelian may usually be used safely as information external to a system of correlations among relatives”

“.. external information of a most precise sort is provided by the pedigree and by the practical universality of Mendelian heredity”

It seems to the writer that what Wright was striving for, when he formulated path analysis, first, was progress up the ladder from descriptive to tangential to functional and that the fact that he halted at the tangential level was an accident – an accident of the temper of the times and of the problems which happened to concern him. It would seem appropriate to credit him with striving for a functional method and to classify the halt at the tangential level as temporary and of minor importance.

J Tukey – In: Statistics and Mathematics in Biology. The Iowa State College Press, Iowa 1954.

Path analysis does not analyse non-genetic paths

"Genetics has but one modest framework for paths. In contrast according to current journals sociologists keep discovering new fundamental path frameworks every month; and sociological graduate students are required routinely to hand in, as individual class exercises, new discoveries equalling Gregor Mendel's."

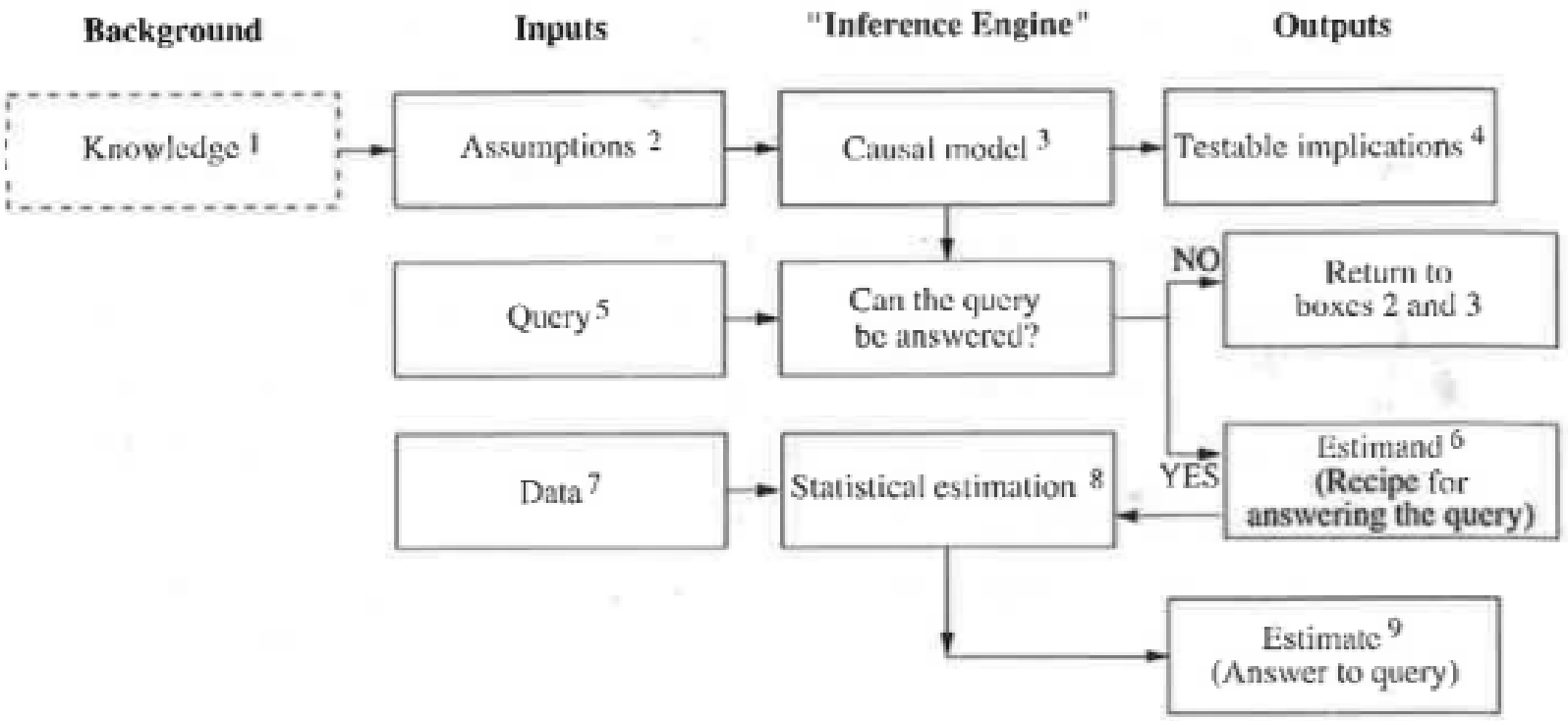
Guttman L. What is Not What in Statistics. Journal of the Royal Statistical Society. Series D. 1977;26:81-107

Letter from Egon Pearson to Jerzy Neyman

However, **by 1929 I had come down firmly to agree with Fisher that prior distributions should not be used**, except in cases where they were based on real knowledge, e.g., in some Mendelian problems. **You were disappointed, but accepted my decision**; after all, the whole mathematical development in the paper was yours." (42)

Lehmann EL. Fisher, Neyman, and the Creation of Classical Statistics. Springer 2011.

Where does background knowledge come from?


Background
Inputs
"Inference Engine"
Outputs
Knowledge 1
Assumptions 2
Causal model 3
Testable implications 4
Query 5
Can the query be answered?
NO
Return to boxes 2 and 3
YES
Estimand 6 (Recipe for answering the query)
Data 7
Statistical estimation 8
Estimate 9 (Answer to query)

I-1. A set A of qualitative causal *assumptions* that the investigator is prepared to defend on scientific grounds, and a model M_A that encodes these assumptions.

Pearl J. Trygve Haavelmo and the emergence of causal calculus. Econometric Theory 2015;31:152-179.

The basic limitation of the methods proposed in this paper is that the results must rest on the causal assumptions shown in the graph, and that these cannot usually be tested in observational studies. In related papers (Pearl, 1994a, 1995) we show that some of the assumptions, most notably those associated with instrumental variables, see Fig. 5(b), are subject to falsification tests.

Pearl J. Causal diagrams for empirical research. Biometrika 1995;82:669-688.

"As with regression models, causal models in observational health and social science (OHSS) are always false. Because we can never know we have a correct model (and in fact in OHSS we can't even know if we are very close), to say G is causal if unconfounded is a scientifically vacuous definition: It is saying the graph is causal if the causal model it represents is correct."

Greenland S. Overthrowing the Tyranny of Null Hypotheses Hidden in Causal Diagrams. In Dechter R et al (eds). Heuristic, Probabilities, and Causality: A Tribute to Judea Pearl. College Press 2010:365-382

Causality: it’s the new thing ..

SARA ACBOR

CAUSAL MODELING

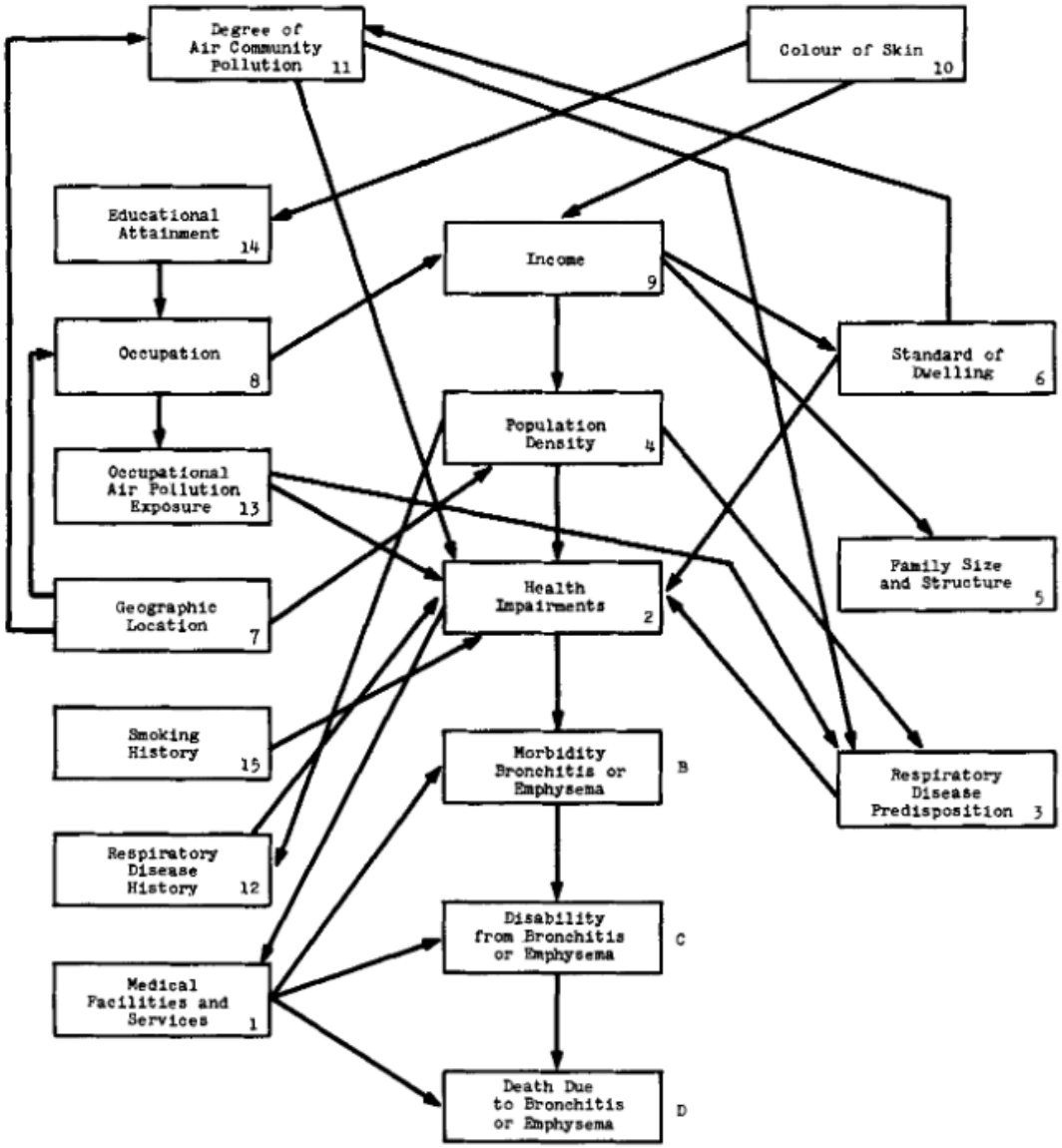
HERBERT B. ASHER

Series: Quantitative Applications in the Social Sciences

Editor: Eric M. Uslaner

a SAGE UNIVERSITY PAPER

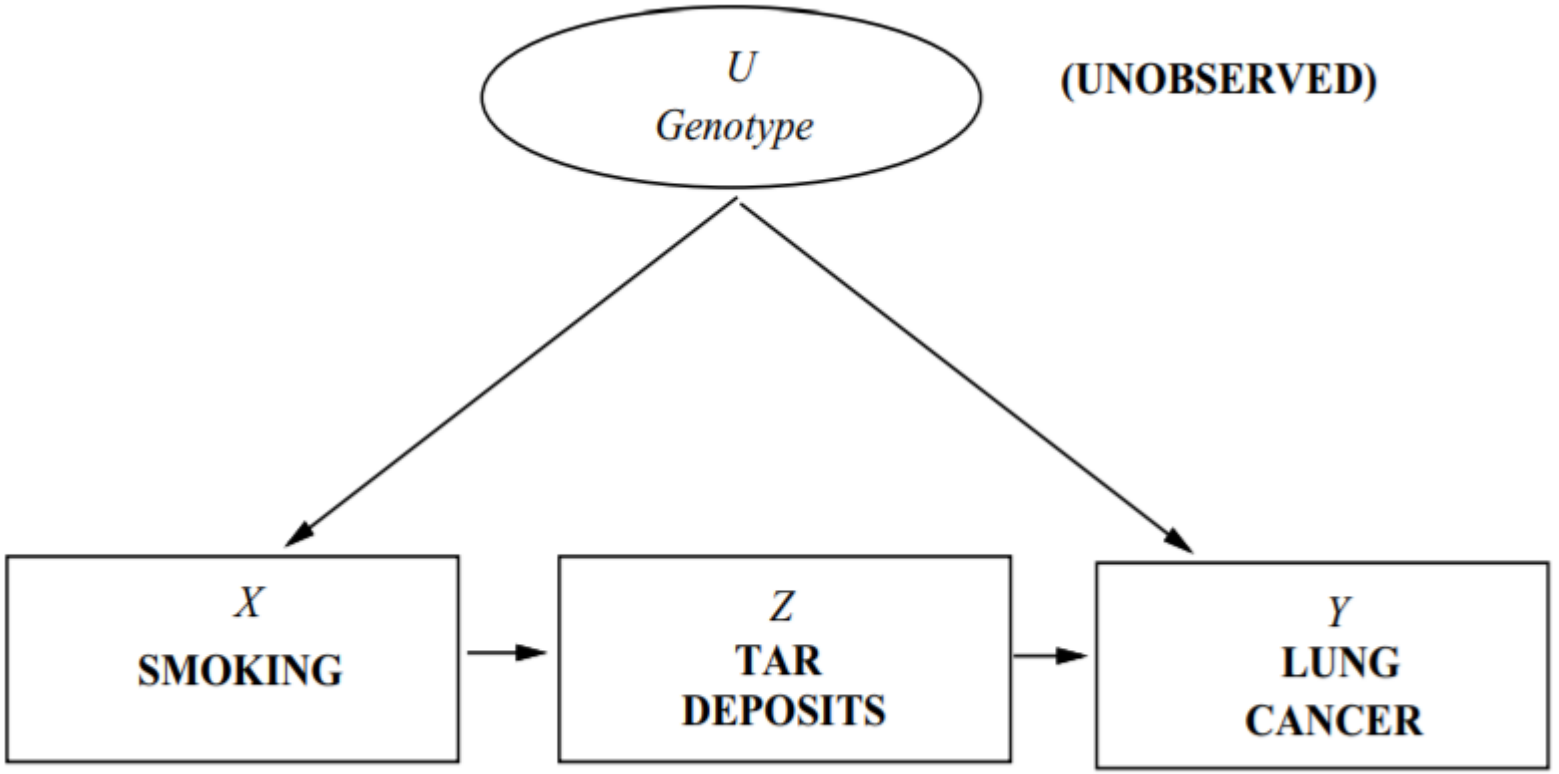
3



Goldsmith JR. Epidemiological approach to multiple factor interactions in pulmonary disease: the potential usefulness of path analysis. Annals of the New York Academy of Sciences 1974;221:361-375

Consequences of believing the DAGs

Introduction of front-door criteria



Pearl J. Mediating Instrumental Variables. Technical Report 1993.

RCTs are considered the "gold standard" of causal effect estimation for exactly the same reason. Because front-door estimates do the same thing, with the additional virtue of observing people's behavior in their own natural habitat instead of a laboratory, I would not be surprised if this method eventually becomes a serious competitor to randomized controlled trials.

Judea Pearl & Dana Mackenzie. The Book of Why: The New Science of Cause and Effect. Penguin, UK. 2018.

The most significant practical impact of the Causal Revolution would probably be a continuous erosion of the supremacy of randomized clinical trials (RCT) in the development and evaluation of drugs, therapeutical procedures, and social and educational policies.

Pearl J. Turing Award Winner, Longtime ASA Member Publishes *The Book of Why. AMSTAT News August 2018*

Under the 21st Century Cures Act, the Food and Drug Administration is tasked with developing a program to evaluate the use of RWE to support approval of new indications for approved drugs or to satisfy postapproval study requirements.[2] RWE can be defined as the clinical evidence regarding the usage and potential benefits or risks of a medical product derived from analysis of RWD. A framework for this program will be published by the end of 2018.

Corrigan-Curay J et al. Real-World Evidence and Real-World Data for Evaluating Drug Safety and Effectiveness. JAMA 2018;9:867-868.

Rosenbaum mistakingly perceives path analysis as a competitor to randomised experiments

Pearl J. Rejoinder to Discussions of “Causal diagrams for empirical research”. Biometrika 1995;82:702-710.

We establish two graphical conditions ensuring that causal effects can be estimated consistently from nonexperimental data. The first condition, named the back-door criterion, is equivalent to the ignorability condition of Rosenbaum & Rubin (1983). The second condition, named the front-door criterion, involves covariates that are affected by the treatment, and thus introduces new opportunities for causal inference.

Pearl J. Causal diagrams for empirical research" by J Pearl. Biometrika 1995;82:694-688

Our overall view of Pearl's framework is summarised by Hill's concluding sentence (1971, p. 296). 'Technical skills, like fire, can be an admirable servant and a dangerous master'. We feel that Pearl's methods, although formidable tools for manipulating directed acyclical graphs, can easily lull the researcher into a false sense of confidence in the resulting causal conclusions. Consequently, until we see convincing applications of Pearl's approach to substantive questions, we remain somewhat sceptical about its general applicability as a conceptual framework for causal inference in practice.

Imbens GW et al. Discussion of "Causal diagrams for empirical research" by J Pearl. Biometrika 1995;82:694-695

Social Science & Medicine 210 (2018) 60–62

Contents lists available at ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

Challenging the hegemony of randomized controlled trials: A commentary on Deaton and Cartwright

Judea Pearl

University of California, Los Angeles Computer Science Department, Los Angeles, CA 90095-1596, USA

I appreciate the opportunity to comment on the article by Angus Deaton and Nancy Cartwright (D&C) (Deaton and Cartwright, 2018), which touches on the foundations of causal inference.

My comments are a mixture of a welcome and a puzzle; I welcome D&C's stand on the status of randomized trials, and I am puzzled by how they choose to articulate the alternatives.

D&C's main theme is as follows: "We argue that any special status for RCTs is unwarranted. Which method is most likely to yield a good causal inference depends on what we are trying to discover as well as on what is already known."

As a veteran skeptic of the supremacy of the RCT, I welcome D&C's challenge wholeheartedly. Indeed, *The Book of Why* (Pearl and Mackenzie, 2018, http://bayes.cs.ucla.edu/WHY/) quotes me as saying: "If our conception of causal effects had anything to do with randomized experiments, the latter would have been invented 500 years before Fisher." In this, as well as in my other writings I go so far as claiming that the RCT earns its legitimacy by mimicking the *do*-operator,[1] not the other way around. In addition, considering the practical difficulties of conducting an ideal RCT, observational studies have a definite advantage: they interrogate populations at their natural habitats, not in artificial environments choreographed by experimental protocols.

Deaton and Cartwright's challenge of the supremacy of the RCT consists of two parts: The first (internal validity) deals with the curse of dimensionality and argues that, in any single trial, the outcome of the RCT can be quite distant from the target causal quantity, which is usually the average treatment effect (ATE). In other words, this part concerns imbalance due to finite samples, and reflects the traditional bias-precision tradeoff in statistical analysis and machine learning. The second part (external validity) deals with biases created by inevitable disparities between the conditions and populations under study versus those prevailing in the actual implementation of the treatment program or policy.

Here, Deaton and Cartwright propose alternatives to RCT, calling all out for integrating a web of multiple information sources, including observational, experimental, quasi-experimental, and theoretical inputs, all collaborating towards the goal of estimating "what we are trying to discover."

My only qualm with D&C's proposal is that, in their passion to advocate the integration strategy, they have failed to notice that, in the past decade, a formal theory of integration strategies has emerged from the brewery of causal inference and is currently ready and available for empirical researchers to use. I am referring of course to the theory of Data Fusion, which formalizes the integration scheme in the language of causal diagrams, and provides theoretical guarantees of feasibility and performance (see Bareinboim and Pearl (2016)).

Let us examine closely D&C's main motto: "Which method is most likely to yield a good causal inference depends on what we are trying to discover as well as on what is already known." Clearly, to cast this advice in practical settings, we must devise notation, vocabulary, and logic to represent "what we are trying to discover" as well as "what is already known" so that we can infer the former from the latter. To accomplish this nontrivial task we need tools, theorems and algorithms to assure us that what we conclude from our integrated study indeed follows from those precious pieces of knowledge that are "already known." D&C are notably silent about the language and methodology in which their proposal should be carried out. One is left wondering therefore whether they intend their proposal to remain an informal, heuristic guideline, similar to Bradford Hill's Criteria of the 1960's, or be explicated in some theoretical framework that can distinguish valid from invalid inference? If they aspire to embed their integration scheme within a coherent framework, then they should celebrate; such a framework has been worked out and is now fully developed.

To be more specific, the Data Fusion theory described in Bareinboim and Pearl (2016) provides us with notation to characterize the nature of each data source, the nature of the population interrogated, whether the source is an observational or experimental study, which variables are randomized and which are measured and, finally, the theory tells us how to fuse all these sources together to synthesize an estimand of the target causal quantity at the target population. Moreover, if we feel uncomfortable about the assumed structure of any given data source, the theory tells us whether an alternative source can furnish the needed information and whether we can weaken any of the model's assumptions.

Those familiar with Data Fusion theory will find it difficult to understand why D&C have not utilized it as a vehicle to demonstrate the

E-mail address: judea@cs.ucla.edu.

[1] For a gentle introduction to the do-operator and do-calculus, see Pearl and Bareinboim (2014) or Pearl et al. (2016).

From “Causal inference in statistics: a primer” Judea Pearl et al

“It proves the enormous, even revelatory, power that causal graphs have in not merely representing, but actually discovering causal information”

George Orwell wrote that language could be used to give the ". . . appearance of solidity to pure wind." It is disturbing that the language of "causal modeling" is being used to bestow the solidity of the complex process of causal inference upon mere statistical analysis of observational data.

Levine B. Causal Models. Epidemiology 2009;20:931.

COI: I am old and time-expired

Comments on: The tale wagged by the DAG

International Journal of Epidemiology, 2018, 1002–1004
doi: 10.1093/ije/dyy068
Advance Access Publication Date: 24 April 2018

Judea Pearl

Computer Science Department, University of California Los Angeles, Los Angeles, CA, USA. E-mail: judea@cs.ucla.edu

Accepted 12 April 2018

I am grateful to the editors for the opportunity to comment on Nancy Krieger and George Davey Smith's article, 'The tale wagged by the DAG', which appeared in the *IJE*'s special issue on causal analysis.[1]

Krieger and Davey Smith raise several objections to the direction taken by modern epidemiology, which they deem to be too narrowly wedded to a directed acyclic graph (DAG)–counterfactual framework. In this framework, graphical models (DAGs) are used to express scientific knowledge, and counterfactuals (or potential outcomes) are used to express queries of interest. As the article does not demonstrate concrete alternatives to current methodologies, I speculate that it is the dazzling speed with which epidemiology has modernized its tools which lies behind the authors' discomfort, and that it will subside as soon as researchers gain greater familiarity with the capabilities and flexibility of these new tools. Epidemiology, as I have written on several occasions, has been a pioneer in accepting the DAG–counterfactuals symbiosis as a ruling paradigm—way ahead of mainstream statistics and its other satellites.[2]

In examining the specific limitations that Krieger and Davey Smith perceive in DAGs, I must note that these limitations coincide precisely with the strengths for which DAGs are praised. For example, the article complains that DAGs provide no information about variables that investigators failed to include in the model. As noted in the introductory editorial in the same issue of the journal, by Davey Smith and other *IJE* editors, 'the DAG does not provide a comprehensive picture. For example, it does not include paternal factors, ethnicity, respiratory infections or socioeconomic position…'[3] This should not be taken as a limitation of DAGs or of any other scientific modelling. Quite the contrary. It would be a disaster if models were allowed to produce information unintended by the modeller. Instead, I have come to admire the ease with which DAGs enable researchers to incorporate new knowledge about new variables, or new mechanisms, when the need arises.

As has been suggested, the views I express here may well reflect the last spasms emitted by a redundant and diminishing group refusing to recognize its superfluousness (135).

Davey Smith G. Post-*Modern Epidemiology*: when methods meet matter, Am J Epidemiol 2019, in press