# What Does Statistics Have to Offer the Epidemiologist?:

## Very much, some of it very bad, some of it very good, most of it insufficient for epidemiologic inference

Sander Greenland

Departments of Epidemiology and Statistics, University of California Los Angeles

Example: Results from pooled analyses of magnetic fields & childhood leukemia

(Greenland, JRSS A 2005, p. 267-308)

Summary odds ratios comparing average field >3 mG to ≤3 mG from 14 studies:

$OR_{ML}$ = 1.7, 95% limits 1.3, 2.2,

**limit ratio = 1.7, P< 0.001**

(Same result from 15 studies, similar estimate with slightly wider limits from earlier analyses with fewer studies)

# A consequence of such results:

Claim by one California State health official (a physicist by training) in a report to California Public Utilities Commission:

“With near certainty, fields cause childhood leukemia” (his epidemiologist counterpart was more cautious).

When the same data are subject to analysis that assigns **nonidentified bias parameters reasonable priors, instead of setting those to zero**, one can get a variety of results…

# Consider a vaguely realistic model for a single exposure-disease analysis

X = Exposure, X*: measured X

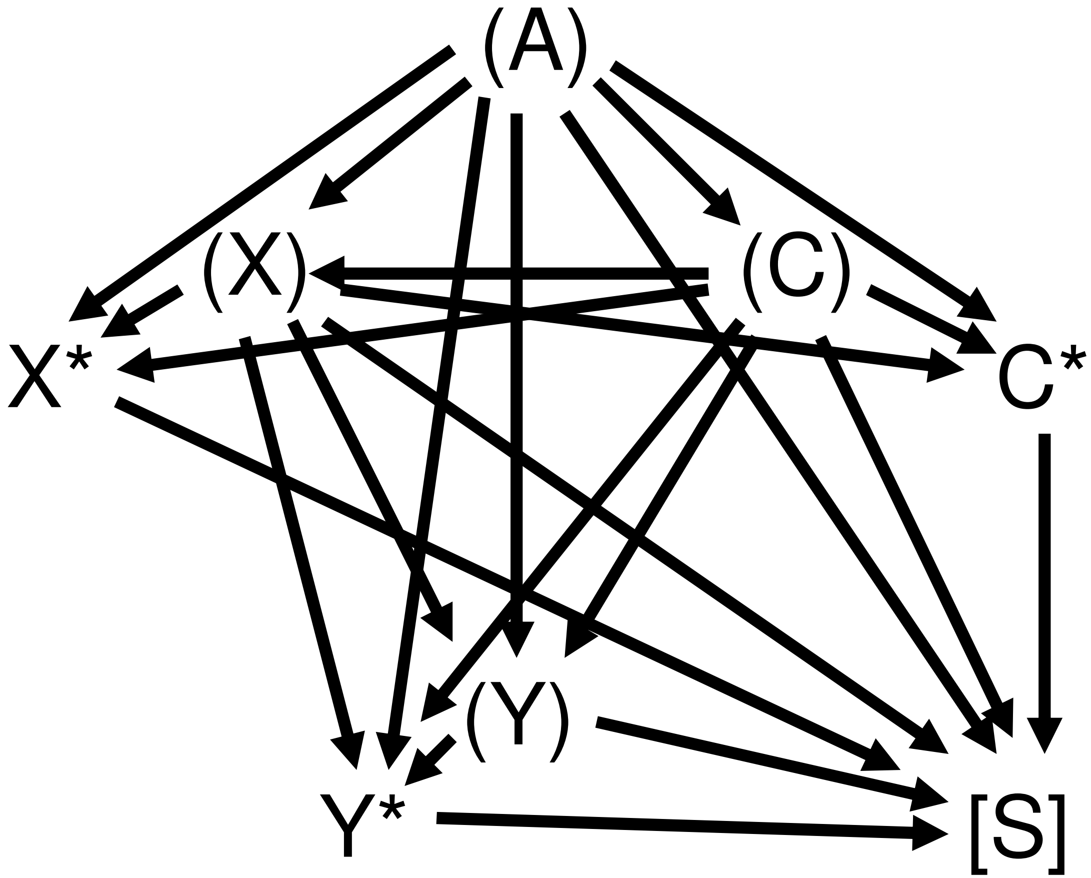
Y = Outcome, Y*: measured Y

C = Known antecedents, C*: measured C

A = Other antecedents (unmeasured or ignored)

S = Selection: by definition, always conditioned, so we always show [S]

Just to estimate the main effect of X on Y under this diagram you would need a parametric model for **all 23** arrows in:



Results from modeling bias sources in MF-leukemia example (assuming C=C*, Y=Y*):

- Percentiles from averaging Monte-Carlo Sensitivity Analyses of 14 studies :

  $OR_{MCSA}$ 50th percentile = 2.7

  (2.5th, 97.5th % = 0.99, 33, **limit ratio = 33**)

- A fully Bayesian analysis of 15 studies:

  $OR_{Bayes}$ 50th percentile = 2.8

  (2.5th, 97.5th % = 0.97, 8.4, **limit ratio = 8.7**)

  (Greenland & Kheifets, *Risk Analysis* 2006, 471-482)

# The Grandest Delusion: "Let the data speak for themselves"

**But** DATA SAY NOTHING AT ALL!

They are just markings on paper or bits that sit there and do nothing.

**If you hear them speaking seek psychiatric care immediately!**

# All our perceptions are filtered through models of the world…

…whether intuitive or mathematical models.

- Because of the profound complexity of realities we study, each of our conceptions must be a gross oversimplification of that reality, whether of carcinogenesis or societies.
- In nutritional and much other epi there is too little empirical evidence to scientifically support more than broad theorization.

# Progression of beliefs: One view

From

- Theory: A statement about proposed but not necessarily held as true.

To

- Factoid: A theory sold as true by many but not well supported by the facts…

To

- Fact: A theory held to be certainly true by a reference community used as the source of acceptable assumptions.

# The problem of excessive certainty

Even when experimental tests exist, we run a grave risk if we hold onto facts with nearly 100% certainty, for the closer we are to 100%, the harder it will be for us to correct our errors of fact.

(“True set of facts” means that there is a homorphism (functional mapping) from reality onto this set of facts. An “error of facts” means lack of such correspondence.)

# Among logical limits to knowledge: Nonidentification (Quine, 1948)

If a theory fits observed regularities well, that is a grossly inadequate basis for thinking it true.

- Many very distinct, perhaps even incompatible alternative theories will fit known facts just as well.
- These alternative theories always exist, even if we do not know of them

# Study design: also a key limiting factor

The value of statistical progress in epi will be limited as long as research remains restricted to a few standard designs.

- Alternative designs may be able to rule out some explanations in exchange for other vulnerabilities, e.g., the case-crossover design rules out confounding by fixed traits (e.g., genes) in exchange for enhanced risk of confounding by carry-over effects.

# Cognitive biases

Unfortunately, a well-documented and apparently innate cognitive bias of human brains is to ascribe more certainty to a deduction than one would assigned to the conjunction of its assumptions.

- This bias is the foundation of frequentist "inference" (as opposed to frequentist decision theory).
- It reflects a more general bias towards certainty (intolerance of uncertainty) in human nature.

# The perspectivalist view

Once we recognize how many reasonable alternatives theories there are, we may also recognize that

- each is a disposable perspective for trying to grasp some aspect of the truth and for making decisions
- it is dangerous it is to seize on and promote any of them as “the truth”

# What about “accepted facts” like “smoking causes lung cancer”?

It suffices to acknowledge that these facts appear true to us, in the sense that our community bet on them is so close to 100% that we may as well treat it as 100% (Cournot’s heuristic, usually called “Cournot’s principle” – but beware: people apply the word “principle” to an assumption when they have no justification for it!).

# What does systematic reasoning offer as a cure for “certainty disease”?

No methodology can take the place of imaginative theorization to develop alternative explanations for observed data… but (as illustrated by causal diagrams), some methodologies can help:

- in deducing empirical implications of proposed theories and accepted facts
- in suggesting alternative explanations
- in suggesting new study strategies

# There is no substitute for field experience

- No methodology take the place of field experience with data collection and study conduct, for that experience can reveal limits of our observation accuracy and point to key factors omitted by our models.
- Observation inaccuracy limits the practical importance of some (not all) methodologic subtleties -- a limit that is often ignored in statistical theory and methodologic studies.

# On the other hand, there is no substitute for quantitative reasoning

…a fact well-accepted for random error, but resisted for bias analysis. Yet nutritional epidemiology (beta-carotene, vitamin E, etc.) shows how qualitative discussion of bias encourages excessive certainty by

- failing to give adequate weight to reasonable alternative explanations ("we doubt that this problem was important")
- failing to accumulate uncertainties over all potential bias sources and interactions.

# Observational scientists at their best

- Pay only modest attention to conventional statistics. A P-value is worth only a glance, for it tests just one of many explanations (the deviation from null model is pure chance)
- Devote more effort to proposing alternative explanations (stories) for the data and testing **all** stories against **any** available observations.
- Recognize that the crucial part of inference is imagining alternative explanations for facts.

[Hill, 1965; Phillips and Goodman, EP&I 2004]

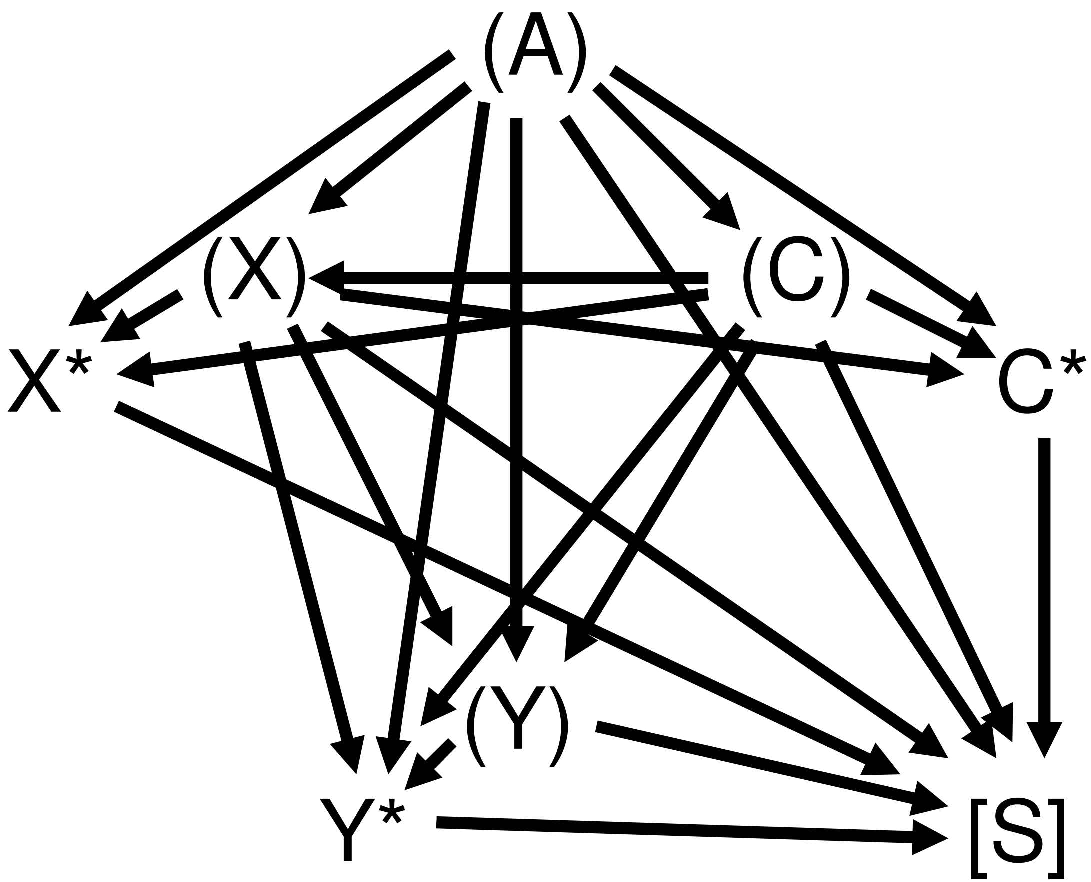
# Scientists at their worst

- Focus more on “significance” than on data, and even confuse the two (e.g., “we found no significant difference” as if “significance” were part of the natural world or the data).
- Focus on study design to prevent bias away from their preferred finding, thus leaving the largest bias toward their preferred finding.
- Search for ways to weave the data into their preferred story (e.g., rummage for results that support preferred explanations).

# Standard models = Arbitrary constraints

- Some arbitrary constraints may be unavoidable, some may even be harmless
- **But**, the excess certainties from what separately appear to be harmless constraints may combine in ways that seriously misguide conclusions
- Thus, **never forget the cost of arbitrary constraints: Excessive certainty in inferences**, whether those inferences are conventional (frequentist), Bayesian, or informal.

Just to estimate the main effect of X on Y under this diagram you would need a parametric model for **all 23** arrows in:



Fixed constraints are just prior distributions with zero variances. Examples of conventional (and arbitrary) constraints:

- No effects on (no arrows into) selection S, apart from those imposed by design.
- No uncontrolled confounding (no uncontrolled antecedent A connected to both X and Y; no open back-door path)
- No measurement error ($X^*=X$, $C^*=C$, $Y^*=Y$), or else some false error model.
- Linear models, multiplicative models.

# The tragedy of conventional statistics

- Assume only random uncontrolled errors.
- Base their derivation on certain **non**-universal value judgments like: "The null should be assumed until disproven"
- As programmed (brainwashed) into the research community, incorporate unjustified, arbitrary elements.
- Have badly warped the **experimental** literature by causing numerous forms of publication bias.

# The Tragedy of Epidemiologic Statistics

- Statistics for designed experiments have been applied to observational studies, with no compelling justification.
- Have badly warped the **non**experimental health literature via conclusion bias and publication bias, and via encouragement of grossly excessive certainty.
- Need to be supplemented (if not supplanted) by methods based on explicitly nonidentified models.

# Statistics at its worst

In the form that has predominated, statistics has forced epidemiologic analysis into a mold developed for sequences of perfect randomized experiments, e.g., measurement of uncertainty by P-values, "significance," and confidence intervals.

- This form is oblivious to key sources of error in epidemiologic studies, and has encouraged overconfident inferences.

# Statistics at its best

- Can supply informative, compact data summaries (undervalued in statistics)
- Can supply an antidote to overconfidence by incorporating **nonidentified** sources of uncertainty (biases) into its models.

When **nonidentification** is recognized, however, the information value of observations is seen to be greatly inflated by the unjustified assumptions of conventional statistics.

# Bayesian methods allow modeling of nonidentified biases, but…

- Tragically, most Bayesian applications are “objective” (ignorant) = make no use of these capabilities.
- Instead, they ape bad frequentist practice of imposing arbitrary constraints on some parameters, no constraint on others.
- To complicate matters (perhaps intentionally), Bayesians promote use of unnecessarily complex fitting procedures.

Bayes is inevitable: You will **have** to use priors to “make inferences”, because again

**The data alone say NOTHING about effects (Hume, 1748). They can only modify your judgments in concert with nonidentified assumptions.**

- You have priors: You used them to design the study! You picked what to measure!

For priors, you have a choice between

- Arbitrary statistical assumptions
- Judgments based on past experience

# Bias modeling may as yet be too complex for routine use; BUT…

It can be argued that every-day studies need not (indeed should not) attempt to make inferences (Greenland, Gago, Castellao, Epidemiology 2004).

- Any informed policy decisions should pool study data, not study conclusions

  -- often used as an argument against Bayesian stats for single studies, but it is an argument against **all** stats for single studies.

# What is needed now

- Reprogramming of the research community to think in **NO certain terms**
- Methods that interface researchers with nonidentified models
- Computer science may be able to make key contributions via visualization programs, allowing rapid and more thorough assessment of sensitivity to shifts in models and priors.

- Every statistical analysis should invite argumentation that challenges the model (set of assumptions) on which it is based.
- The assumptions used by most statistics are **harmful and unwarranted** by available data.
- Bias toward certain explanations may be unavoidable, but also may be more easily diagnosed when all recognized major sources of bias are parameterized and given explicit probability weights (because explicit weightings can be challenged easily).

# Shouldn’t we strive for “objectivity”?

- Yes, genuine self-critical objectivity, not delusions of grandeur encouraged by hijacking ordinary language to manufacture fake knowledge out of ignorance.
- The “subjectivity” in “subjective Bayes” simply acknowledges the fact that, upon a reasonably detailed analysis, we may find vast disagreements about models and assumptions across groups, and perhaps even across individuals within these groups.
- I am advocating **subjunctive** Bayes [Senn]

# Simple models can be useful

Model dependence is **not** cause for despair of useful knowledge. **On the contrary**, even recognizing a few basic predictive regularities in the behavior of societies could be extraordinarily powerful in helping avert social catastrophes like war and famine – even if there is no direct observation!

One of the greatest public-health triumphs since World War II: Prevention of thermo-nuclear war between the US and USSR.

Every argument for or against a particular decision is in part based on predictions of how the future will unfold with & without that action.

- If examined closely, decision arguments will always be found to have many hidden simplifying assumptions: dogmatic point prior distributions that control the dimensionality of the problem.

- Predictions get biased when the simplifications used to derive them are seriously violated.
- Predictions get imprecise without bound as simplifications are removed and the parameters once treated as known are now allowed to be of uncertain value.

(Any prediction (e.g., where stock values will be in a year or HIV rates will be in five) is a prior probability distribution.)

# Conventional models = forced oversimplifications

- Conventional models are bizarre extremes in which parameters are either known exactly or not at all.
- They allow no finer degrees of knowledge. That makes these models uniquely unsuitable for risk assessment when almost all knowledge is vague in form and imprecise (as in nutritional, occupational, and environmental epidemiology)

# Thus, a limited role for statistics

- Statistical methods should help us judge the compatibility of our existing theories (including our facts) with observations. But…
- Statistical theory bogs down in details when confronted with data having complex and partially uncontrolled origins (data from partially unknown missingness or coarsening mechanisms; “messy data”).

# Data: Everything and Nothing

Devaluation of data by bias modeling is at odds with the large stake in data collection held by many researchers, and has thus attracted distortive attacks on the methods.

- Why shoot the messenger? As nonidentified models demonstrate, feasible studies of lifestyle and health offer much less certainty and hence less value for money than their statistics indicate – a fact that has been known to many good epidemiologists for decades.

# Implications for Decisions

Nonidentified bias models will usually benefit one side to the cost of the opposition (but not always the way some think):

- Nonidentified models will show that much less certainty follows from available data than conventional statistics indicate – **but** possibly much more certainty than “expert” qualitative discussion indicates.
- Implication depends on who benefits from uncertainty about the epidemiology.

# Some Experiences to Date

- Pro-tobacco (e.g., R.A. Fisher) exploited qualitative uncertainty to defend tobacco
- Anti-tobacco (e.g., Cornfield, Bross) exploited nonidentified models to undermine these attacks.
- In modern US tort cases the chief beneficiaries of epidemiolgic uncertainty are often *plaintiffs*, because defense strategy has been to claim that certain studies demonstrate safety.

Because nonidentified models will be costly to those claiming high certainty and authority as epidemiologist or statisticians, they will attract controversy because they challenge that authority.

- As with all methods, this controversy can be healthy insofar as nonidentified models have serious abuse potential:

  Nonidentified models simplify working backward from any desired conclusion to an analysis that produces it.

# Bayesian analysis: A two-edged sword

The explicit prior distributions in bias analysis can reflect reality better than conventional models do. **But** they can be manipulated to produce any desired result.

- To deal with this weak point, we need contextual justifications for each prior.
- Derivation from past data is best but often unavailable.
- Regardless, it helps to exhibit an explicit thought experiment that generates the prior.

# Defending against statistics abuse

- Understand what each model assumes about the real world.
- Understand the knowledge claims of each prior distribution being used (whether the prior is implicit as in freqentist methods, or explicit as in Bayesian methods).
- Recognize that any claimed inference from an analysis logically cannot be more certain than the combination of its assumptions.

# Points to remember for any analysis

- Nothing can ever be deduced with certainty about the real world.
- At best we only reach social agreements to treat certain statements about the world as facts and certain methods as effective for for fact-finding.
- The spectacular divergence in these agreements should restrain one from indulging in excessive dogmatism or ardor about material facts (at least in a scientific venue). [material Jainism]

# Appendix: Interpretation of a data prior

The result from a perfect study that yields your prior interval as its confidence interval.

Example:

- Suppose your prior is normal and gives a 95% bet that RR is between ¼ and 4.
- This prior has the same information value as a perfect balanced RCT that observed 4 cases at X=1 and 4 cases at X=0 in a population in which the outcome is rare.

Under a normal prior,

- The same prior yields a 67% bet that RR is between is between ½ and 2;
- For a 95% bet that RR is between ½ and 2, the trial would have to have 16 cases at X=1 and 16 cases at X=0.

**Note how fast the trial size increases as the interval narrows** (quadratically) **or the prior percent increases** (approaches one)