From patients to policies in HIV:
toward epidemiologic methods for implementation science

Daniel Westreich

June 2017

Acknowledgements

I am primarily funded by NIH DP2 HD084070, as well as by the UNC WIHS U01 AI103390, R01 AI100654, and UNC CFAR P30 AI50410.

I work closely with Stephen R. Cole, Jessie K. Edwards, Alex Keil, Michael Hudgens, Ada Adimora and numerous other outstanding scientists.

Blame me for the errors in what follows.

Disclosures

Consulting with Sanofi-Pasteur related to influenza vaccines.

Disclosures

Consulting with Sanofi-Pasteur related to influenza vaccines.

I enjoy the music of ABBA.

Overview

1. HIV, smoking, and pregnancy loss: a motivating example
2. Discursion: on exposures and population interventions
3. HIV, smoking, and pregnancy loss: a motivated example
4. Discursion: internal validity and its discontents
5. Conclusions

Motivations

Smoking is extremely common among HIV-positive Americans: double (~42%) the prevalence compared with general US population (~21%).

Another 20% of HIV-positive Americans are former smokers.

Mdodo et al. *Annals of Internal Medicine* 2015

Motivations (2)

Smoking raises risks of miscarriage: Pineles et al. meta-analysis reports that smoking during pregnancy has a meta-analytic risk ratio of 1.32 (95% CI: 1.21, 1.44; n = 25 studies).

Pineles, Park, Samet *American Journal of Epidemiology* 2014

Motivations (3)

Almost no data on smoking and miscarriage among HIV-positive women.

Just for example, PubMed.gov (3 May 2017):

"smoking miscarriage"	→ 772 hits
"smoking miscarriage HIV"	→ 5 hits

Study question

What is the causal effect of smoking on risk of miscarriage, and does the effect vary between HIV-positive and HIV-negative women?

Sidebar: we are explicitly asking a causal question here. We are explicit about this only rarely, but it is frequently implicit.

Data source: the WIHS

The **Women's Interagency HIV Study** is a multicenter prospective observational cohort study of HIV-positive and sociodemographically matched un-infected women enrolled at ten cities throughout the United States. WIHS participants undergo a twice-yearly medical exam and interview; detailed procedures are described elsewhere.

Sites scattered across the US; UNC, UAB, and Emory are new sites of the WIHS.

Exposure

Cigarette smoking during or immediately prior to pregnancy, which we referred to as “current cigarette smoking.”

Outcome

Self-reported stillbirth or miscarriage (pregnancy loss before 20 weeks), compared with live birth.

Confounders

Identified from a DAG (not shown; it's a mess)

Included age, race, employment status, above-median income, body mass index, depression, and recent use of intravenous drugs, marijuana, and alcohol.

Variables captured at the same visit as the exposure; generally modeled as categorical variables or restricted quadratic splines as appropriate.

Statistical methods

Log-binomial regression; robust variances b/c some women had >1 pregnancy outcome in our analysis.

Results (AIDS 2017)

OPEN

Smoking, HIV, and risk of pregnancy loss

Daniel Westreich[a], Jordan Cates[a], Mardge Cohen[b], Kathleen M. Weber[c], Dominika Seidman[d], Karen Cropsey[e], Rodney Wright[f], Joel Milam[g], Mary A. Young[h], C. Christina Mehta[i], Deborah R. Gustafson[j], Elizabeth T. Golub[k], Margaret A. Fischl[l] and Adaora A. Adimora[a,m]

Objective: Cigarette smoking during pregnancy increases risks of poor pregnancy outcomes including miscarriage and stillbirth (pregnancy loss), but the effect of smoking on pregnancy loss among HIV-infected women has not been explored. Here, investigated the impact of smoking on risk of pregnancy loss among HIV-positive and HIV-negative women, and estimated the potential impact of realistic smoking cessation interventions on risk of pregnancy loss among HIV-positive women.

Design: We analyzed pregnancy outcomes in HIV-positive and HIV-negative participants in the Women's Interagency HIV Study between 1994 and 2014.

Methods: We estimated effects of current smoking at or immediately before pregnancy on pregnancy loss; we controlled for confounding using regression approaches, and estimated potential impact of realistic smoking cessation interventions using a semi-parametric g-formula approach.

Results: Analysis examined 1033 pregnancies among 659 women. The effect of

Smokers differed from nonsmokers

These data are for the pregnancies, not people

Characteristic	Current smoker n=377	Current nonsmoker n=656
Demographic		
Black race	253 (67%)	372 (57%)
Socioeconomic status		
Employed	74 (20%)	267 (41%)
Substance use		
Since last visit		
Alcohol consumption (any)	166 (44%)	173 (26%)
Intravenous drug use	8 (2%)	4 (1%)
Non-intravenous drug use	145 (38%)	78 (12%)
Marijuana use	114 (30%)	70 (11%)
Intravenous drug use at baseline‡	63 (17%)	34 (5%)
Clinical indicators		
HIV-positive	212 (56%)	380 (58%)
Body mass index†	27.4 (23.4, 32.9)	29.2 (25.4, 34.0)
Depression (CESD ≥ 16)	166 (44%)	159 (24%)

Smokers differed from nonsmokers

These data are for the pregnancies, not people

Characteristic	Current smoker n=377	Current nonsmoker n=656
Demographic		
Black race	253 (67%)	372 (57%)
Socioeconomic status		
Employed	74 (20%)	267 (41%)
Substance use		
Since last visit		
Alcohol consumption (any)	166 (44%)	173 (26%)
Intravenous drug use	8 (2%)	4 (1%)
Non-intravenous drug use	145 (38%)	78 (12%)
Marijuana use	114 (30%)	70 (11%)
Intravenous drug use at baseline‡	63 (17%)	34 (5%)
Clinical indicators		
HIV-positive	212 (56%)	380 (58%)
Body mass index†	27.4 (23.4, 32.9)	29.2 (25.4, 34.0)
Depression (CESD ≥ 16)	166 (44%)	159 (24%)

Smokers differed from nonsmokers

These data are for the pregnancies, not people

Characteristic	Current smoker n=377	Current nonsmoker n=656
Demographic		
Black race	253 (67%)	372 (57%)
Socioeconomic status		
Employed	74 (20%)	267 (41%)
Substance use		
Since last visit		
Alcohol consumption (any)	166 (44%)	173 (26%)
Intravenous drug use	8 (2%)	4 (1%)
Non-intravenous drug use	145 (38%)	78 (12%)
Marijuana use	114 (30%)	70 (11%)
Intravenous drug use at baseline‡	63 (17%)	34 (5%)
Clinical indicators		
HIV-positive	212 (56%)	380 (58%)
Body mass index†	27.4 (23.4, 32.9)	29.2 (25.4, 34.0)
Depression (CESD ≥ 16)	166 (44%)	159 (24%)

Main results, risk ratios

The risk ratio for current smoking vs. not-current smoking among women in the WIHS (controlling for possible confounding by HIV status):

Overall 1.55 (95% CL 1.28, 1.89)

The risk ratio for current smoking vs. not-current smoking among women in the WIHS by HIV status:

HIV-negative 1.31 (95% CL 0.99, 1.75)
HIV-positive 1.74 (95% CL 1.36, 2.23)

Interaction beta-coefficient is 0.28 (SE 0.18), p=0.123, suggesting interaction. Broadly, results supported by sensitivity analysis.

Conventional discussion (1)

Smoking has a stronger impact on risk of miscarriage among HIV-positive women than HIV-negative women.

Adjusted risk ratio among HIV-negative women is not statistically significant, but is delightfully coherent with meta-analytic result (1.32, 95% CI: 1.21, 1.44).

Conventional discussion (2)

Strengths: WIHS is well-understood and well-collected interval-cohort data.

Limitations: can't interpret as a causal effect because of possible uncontrolled confounding (non-exchangeability), possible measurement error, potential for meaningful treatment variation (is all current smoking created equal?), etc.

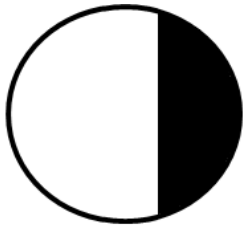
Unconventional discussion

But wait. What are we estimating here?

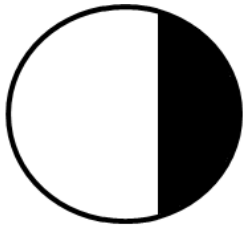
We were using a log-binomial regression model, which (under causal identifiability assumptions) estimates a (sample-)average causal effect.

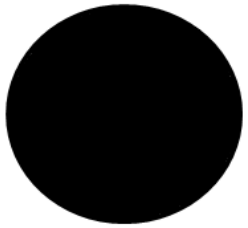
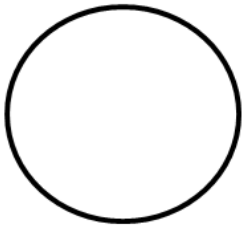
That is, we made a contrast between two counterfactual exposure distributions. Namely, all-exposed and none-exposed. Think about that for a second:

A. The observed population.

A. 

B. Sample average causal effect

A. 

B.  vs. 

Wait, what?

So we asked, “if all these women were smokers, what would their risk be?” and “if all these women were nonsmokers, what would their risk be?”

Neither of these proposed exposure distributions are observed: both are *counterfactual*.

And worse, neither is realistic. In what world are ALL these women smokers? What intervention do you propose to get them ALL to quit smoking?

What, then, does this contrast tell us?

Anyone?

In particular, does it have any bearing on policy-making (that is – on public health *per se*)?

What should an implementation scientist make of this number?

What, then, does this contrast tell us?

An opinion: it has some bearing on public health. It tells us that smoking is bad; and that smoking cessation may be higher priority among HIV-positive women than among HIV-negative women for purposes of preventing miscarriage.

I think it mainly tells us something about the effect of smoking on a typical woman in this population. This risk ratio is the best guess at the individual causal effect.

> *Though of course it cannot be assumed to actually apply to any individual in the cohort – so it might be better viewed as a prior?*

So I think of this as an “exposure” effect. One alternative:

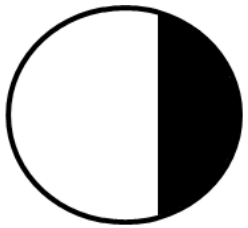
Population intervention effects

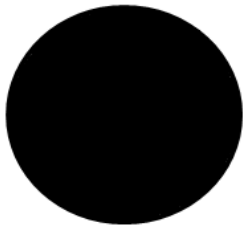
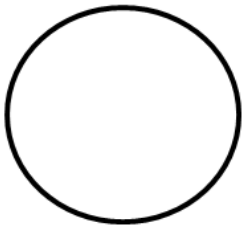
Name due to Hubbard and van der Laan 2005

Let us contrast the observed exposure distribution to a more-realistic counterfactual – ideally one that corresponds to a realistic intervention.

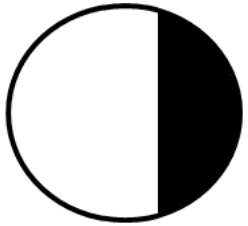
What does that look like?

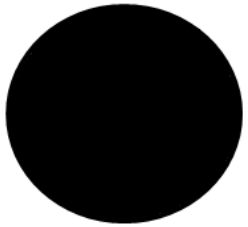
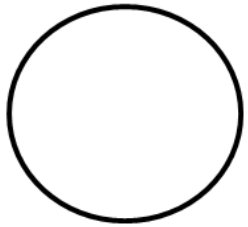
Where we left off:

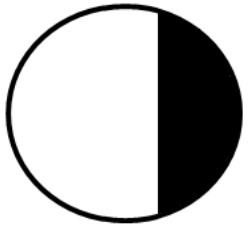
A. 

B.  vs. 

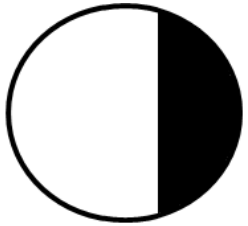
C. Population

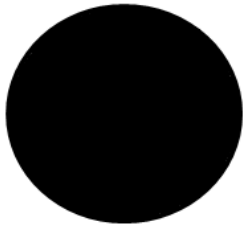
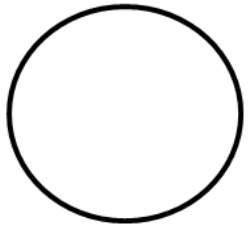
A. 

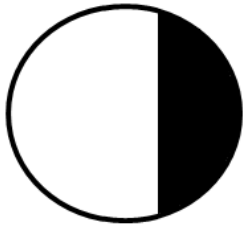
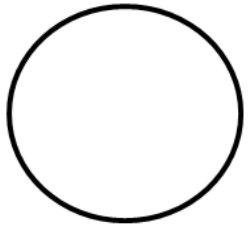
B.  vs. 

C. 

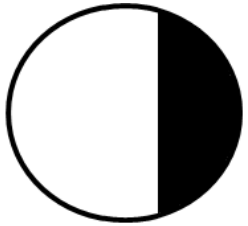
C. Population attributable fraction

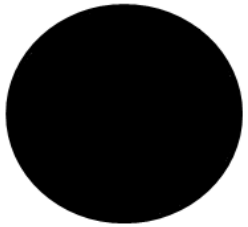
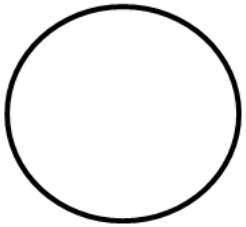
A. 

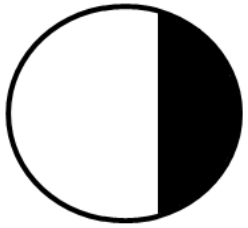
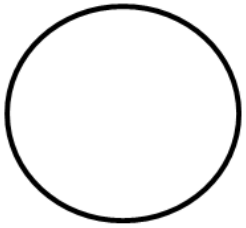
B.  vs. 

C.  vs. 

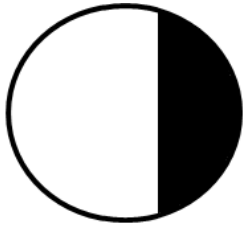
C. Population attributable ~~fraction~~

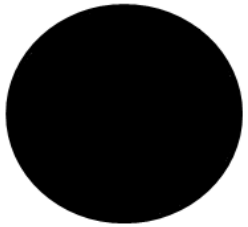
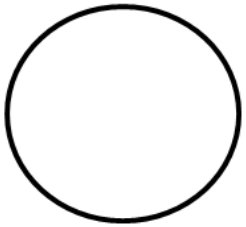
A. 

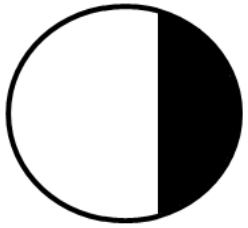
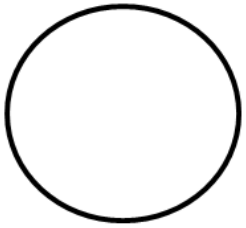
B.  vs. 

C.  vs. 

C. Population attributable effect

A. 

B.  vs. 

C.  vs. 

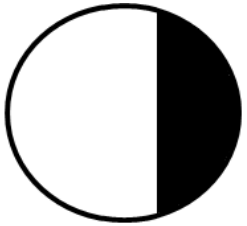
Population attributable effect

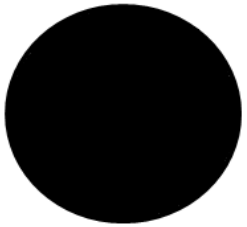
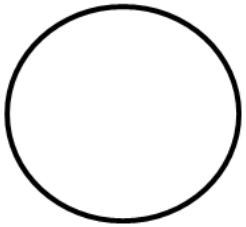
The risk under the observed exposure compared to the risk under all exposure removed (alternatively, under all exposure present).

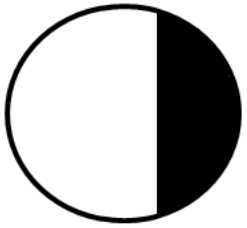
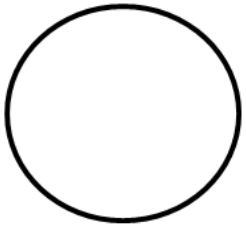
More realistic (not imagining all women are or become smokers), but still flawed: can't remove all smoking. So fails the "more realistic" test.

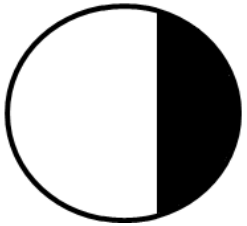
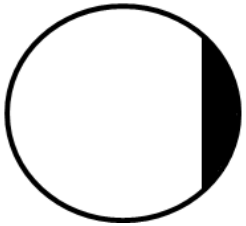
One possible interpretation: bounding condition for the perfect intervention.

D. Generalized intervention effect

A. 

B.  vs. 

C.  vs. 

D.  vs. 

Generalized intervention effect

The risk under the observed exposure compared to the risk under *some* exposure removed.

- Special case: all exposure removed (population attributable effect).

How much exposure removed? This can be answered by invoking a realistic (and ideally, empirically tested) intervention.

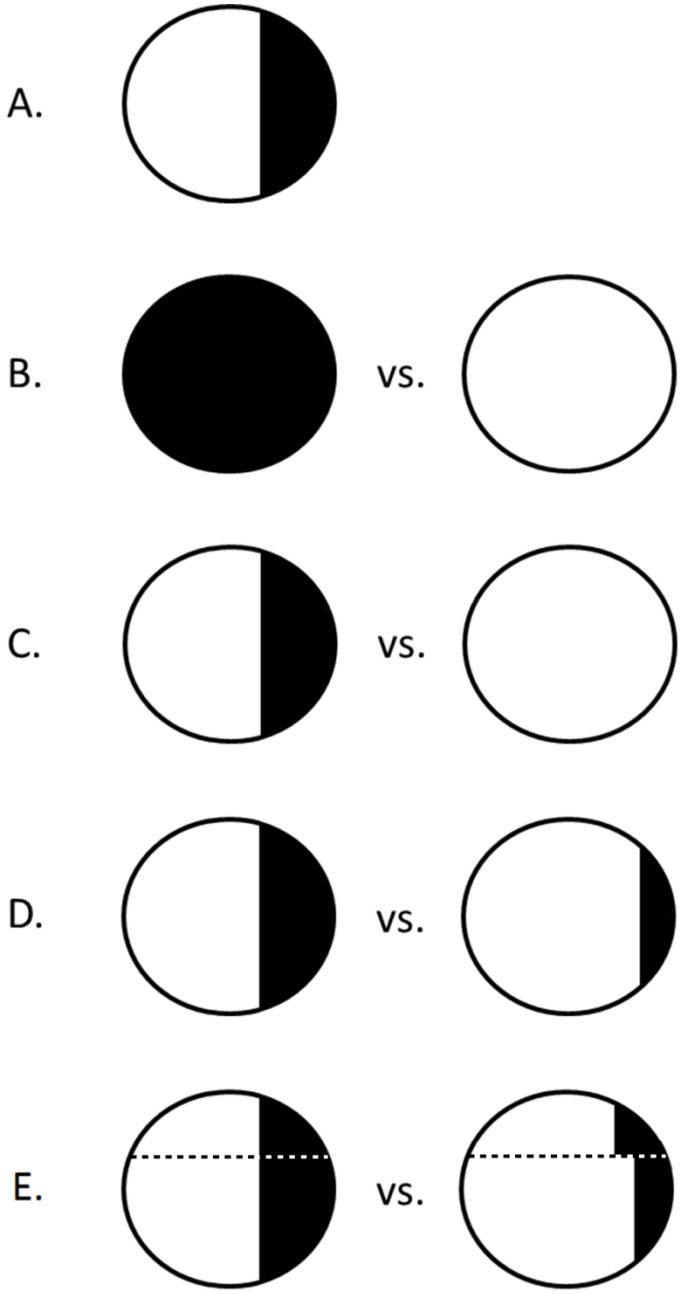
Interventional estimates

Always/never comparisons are exposure contrasts (B)

Interventional estimates are more immediately useful for policy. E.g.,

- population attributable effect (C)
- generalized intervention effect (D)
- dynamic intervention effect (E)

See Westreich Epidemiology 2017 for more discussion (and this figure).



Dynamic intervention effect

The risk under the observed exposure compared to the risk under *some* exposure removed conditional on characteristics **Z** of participants.

Dynamic and generalized effects will differ when there is (i) effect measure modification of the *exposure* by **Z**, and (ii) the *intervention* removes exposure differentially by **Z**.

E.g., smoking is worse for miscarriage when you're older than when you're younger; and older women quit smoking less readily than younger women.

A general rule for this schema

If an effect is invariant to the exposure prevalence, it is an exposure effect.

If an effect varies with exposure prevalence, it is a population intervention effect.

(Consider a randomized trial.)

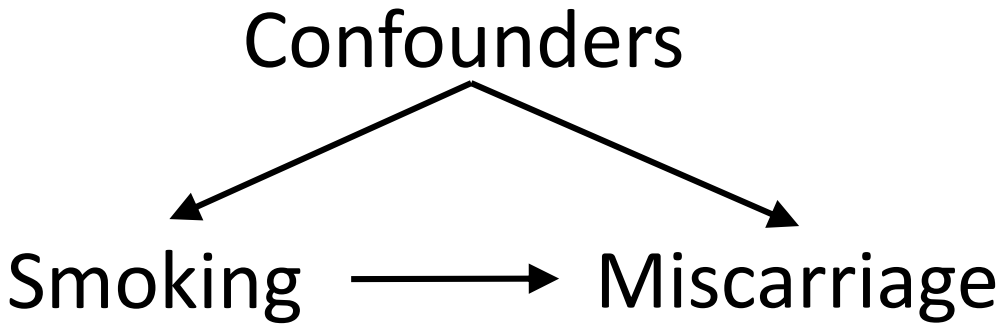
Implications for our study

We asked: what is the causal effect of smoking on risk of miscarriage, and does the effect vary between HIV-positive and HIV-negative women?

What if instead we asked a question more relevant to the WIHS itself. For example: what is the expected impact of a realistic intervention for smoking cessation on the total population risks of miscarriage in the HIV-positive women?

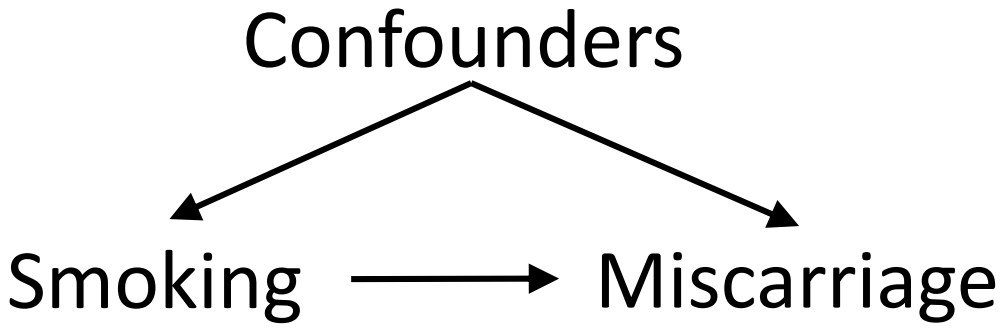
Causal question changes

From:

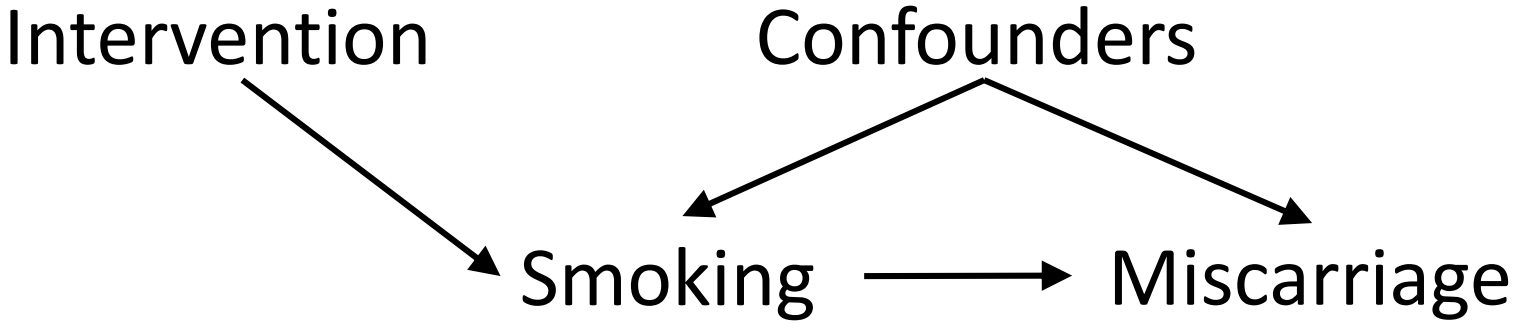


Causal question changes

From:



To:



Statistical methods

Risk differences rather than risk ratios. Since we are interested in potential interventions, we care not about relative risks, but about absolute changes in number of pregnancy losses.

Use the parametric g-formula to estimate this, because it is conceptually very clear way of simulating the impact of realistic interventions in observed data (and scales to highly complex data / interventions).

The g-formula in one (more) slide

For a time-fixed exposure, this is very easy.

- Fit a model for the outcome on exposure, confounders.
  - Easiest to think about a logistic model.
- Use beta-coefficients from fit model to predict outcomes for each individual under their observed exposure (risk under natural course).
- Then, go and alter everyone's exposure to what it would be under the proposed intervention (risk under intervention)
  - predict outcomes under THOSE exposures.
- Compare the two. Bootstrap for confidence intervals.

Much trickier to implement with a time-varying exposure, but similar in principle.

Note that simulating the intervention node in our amended DAG may require us to go outside the data we have at hand; and so perhaps this goes one single step beyond the g-formula toward microsimulation.

Robins J Chron Dis 1987; Petersen & van der Laan JAIDS 2006; Taubman et al. 2009; Westreich et al. 2012; others

Results (briefly, b/c time)

We woul have to offer a realistic smoking cessation intervention (which we have seen demonstrated elsewhere in the WIHS) to 36 HIV-positive women to prevent one miscarriage.

Full results: Westreich et al. AIDS 2017.

Discussion

Policy decisions usually involve at least some form of cost-effectiveness analysis. That last NNT is vastly more useful to a cost-effectiveness analysis than the risk ratio from the original analysis. But which do we usually report?

Implementation science needs to start implementing with a workable intervention. But typical epidemiologic analyses provide estimates of exposure effects, not estimates of intervention effects.

Estimating population intervention effects can help make your work more immediately applicable to public health policy and implementation science.

> *With the strong caveat that we should rarely be making policy based on a single study. We could adopt a semi-Bayesian approach to guard against errors in this realm, or rely on meta-analysis instead of my one study.*

Limitations

While more tied to the real world, estimating population intervention contrasts require additional assumptions.

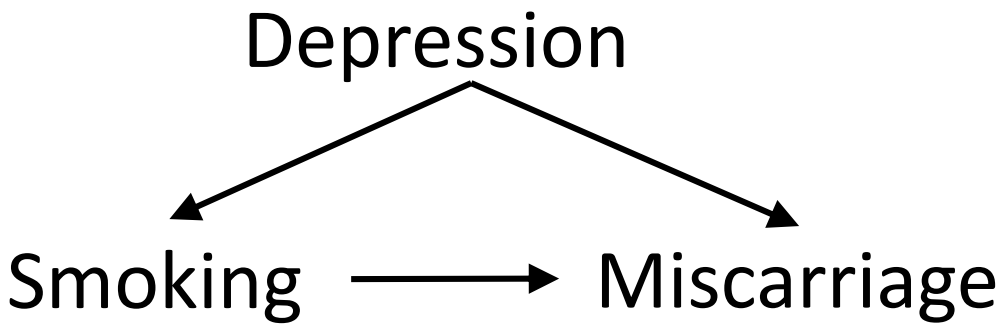
One of them is inherent in our definitions: that population intervention effects depend on exposure prevalence, so we have to consider generalizability closely (I'll return to this shortly).

Another: no meaningful side effects of the intervention.

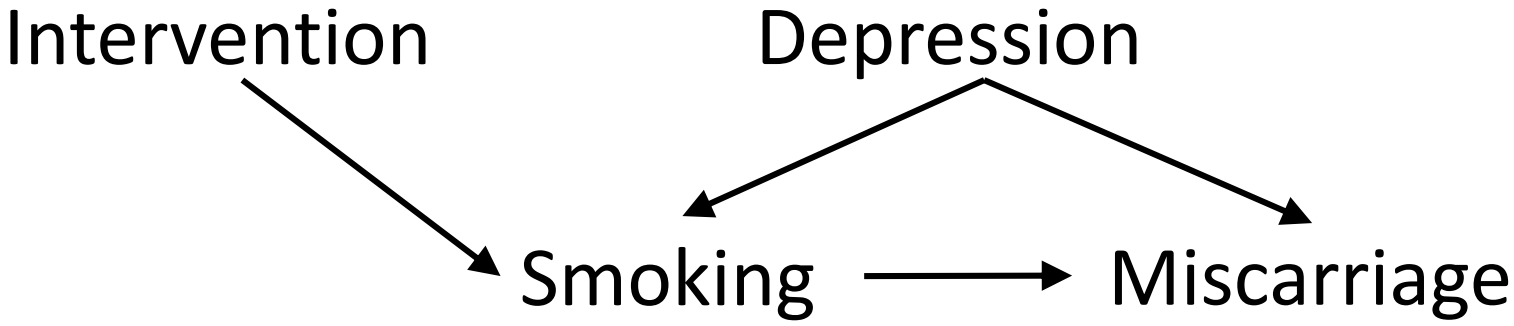
Sidebar: the sales price fallacy.

Causal question changes…

From:

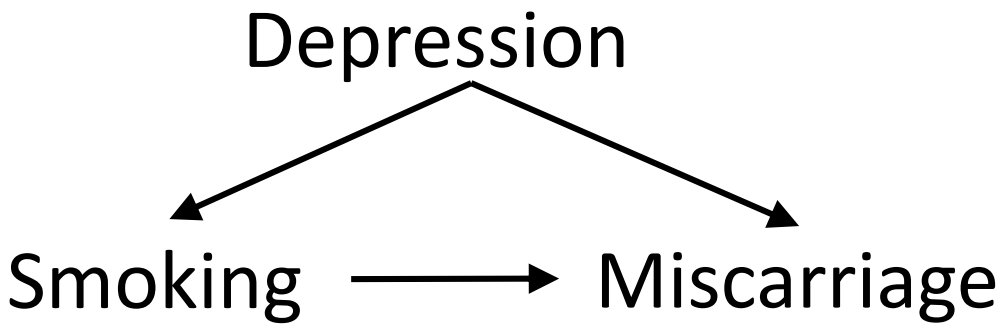


To:

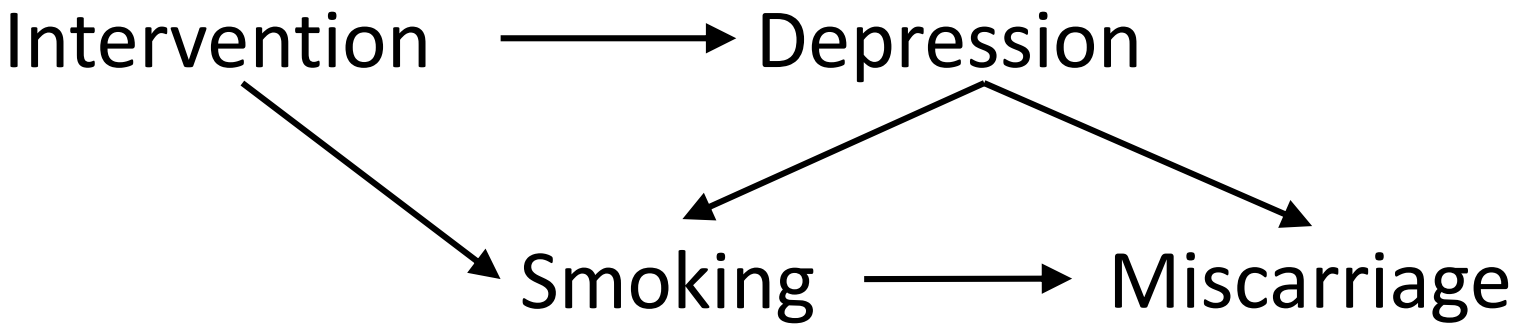


…and changes again.

From:



To:



Internal validity and its discontents

When epidemiologists and biostatisticians talk about “causal inference” they usually mean only internal validity – which basically comes down to (approximately!) “the stuff you get for free in a well-conducted doubled-blinded randomized trial with no loss to follow-up under an ITT analysis.”

In which case, “causal inference” is not the whole story. We proposed an alternative model:

Causal Impact: Epidemiological Approaches for a Public Health of Consequence

In particular, we argue for a framework that examines

1. Internal validity
2. External validity
3. Population intervention estimation

Westreich et al. American Journal of Public Health 2016

Briefly, on internal v. external validity

A trial (or well-conducted observational study) will usually identify an internally valid average causal effect (ACE) – one which is valid in the study sample.

But what is the target population? We are rarely interested in the study sample for its own sake, yet it is rare that we formally identify the target population. We almost never describe the target population in detail.

If the target population differs systematically from the study sample, the average causal effect estimated from the study sample may not generalize (or "transport") unconditionally.

Example

The causal effect of treatment on the outcome is a risk difference of 10% in HIV-negative women, and 20% in HIV-positive women.

Our study is 50% HIV-positive, because we oversampled HIV-positive women.

Our target population is only 10% HIV-positive women.

Same overall risk difference? (No: it's just about weighted averages.)

Example: every traditional randomized trial

1. We know that for a non-null effect, a change in the baseline risk of outcome necessitates effect measure modification on at least one scale (difference or ratio). Example:

Baseline risk	**Exposed risk**		**RD**		**RR**
5%	10%	*(10-5=)*	5%	*(10/5 =)*	2.0

Suppose baseline risk shifts to 10%: what is exposed risk?

10%	20%	10%	2.0
10%	15%	5%	1.5

RD and RR can't both stay constant here.

Example: every traditional randomized trial

2. Nearly all randomized trials misrepresent the baseline risk in the target population.

Typically, they will oversample individuals at high risk of the outcome to increase power.

In HIV contexts, for example: serodiscordant couples studies for HIV transmission outcomes; people with high-risk behaviors for HIV vaccine trials.

In CVD contexts: people with high risk of MI or stroke.

Example: every traditional randomized trial

Therefore:

Most randomized trials will not be generalizable to the intended target population (from which the study population was sampled) on at least one scale (ratio or difference).

Specifically, this will apply to any trial with a non-null result, where the baseline risk in the target population of interest differs from the baseline risk in the study sample.

Of course, if the target population IS the study sample, there is no issue; but this is rarely the case (indeed, elsewhere we argue never).

External validity: recent attention

Very little formal, quantitative attention was paid to external validity in the modern causal inference literature until recently.

Hernán et al. Epidemiology 2008; Weisberg et al. and Frangakis (comment) Clinical Trials 2009, Keiding and Louis JRSSA, 2016.

Cole & Stuart Am J Epid 2010 presented a method using inverse probability weights to standardize clinical trial results to an external target population. Extended by Westreich et al. AJE 2017 in press.

Pearl and Bareinboim (and others) have presented several papers presenting causal diagram-based methods for identification of externally valid causal effects.

Causal Impact: Epidemiological Approaches for a Public Health of Consequence

In particular, we argue for a framework that examines

1. Internal validity
2. External validity
3. Population intervention estimation

In conclusion

Smoking is bad.

…and over to Dr. Rogawski.