Case-control simulation meeting 12/12/17

Genesis of this idea
Straightforward ways to combine data sources, weights

Basic - sim made up pop
outcome / disease incidence
covars
show that we can recover CIR, IDR, OR
↓complexity
1. all cases, SRS controls - show this works
2. " " , probability-based sample of controls
3. " " # stratified/complex sampling
4. diff source pops - assumptions around W, modifiers, transportability

NSF grant app - potential - due feb 28
→ creation of data resources - need to be public
→ advancement of methodology
social, econ, behav. sciences
↳ all prob using big, admin data, even if not case-control

- Jen to think about whether or not she wants to do this
→ will talk to Will, psychologists, sociology

Start w/ several categorical, binary, cont W's - age, race, sex, county, zip, household, urban/rural
- related
- exposure - binary
- outcome - binary

Catherine, Chris, Ellie - we can meet + work on basic structure

Sanitz - interpreting epi evidence
case-control studies - characteristics of control group

after 1st week of Jan

Would be good to check what's been done on this - Catherine to do a lit search first

Patrick - thinking
start w/ a known case-control study
what if we didn't have these controls
Build simulations off of real data
→ simulate outcome only - use real covariate data

Jen wants to start from cohort + show that all the case-control designs

Would be good to have an applied example - use cohort data as base

BioLink - pubicly available datasets
ERIC
Framingham

Start [redacted] basic - given that we don't have a lot of time left on grant

More complexity
- pair matching
- freq. "

Start w/ generic simulation
n = 1 mil or 100,000
OF = 5% ish
How many controls - 4:1 - diminishing returns
Start w/ cols from blown up ACS + then shrink it down to something manageable
from 1 year

Computing

Variability
Treat population as truth
Repeat sampling many times to characterize the distribution of point estimates + SE ests from each run

Patrick will ask Burke for Grizzlybear access for Catherine

Timeline - as soon as we can.

Parameters? I assume conditional not marginal? OR/RR?

Report design ~~effect~~ factor / relative efficiency

Possible extensions: measurement error in the number of controls (needed for case-control weights)