Dependent Nondifferential Misclassification of Exposure

DISCLAIMER:

I am **REALLY** not an expert in data simulations …
or misclassification

Outline

- Relevant definitions
- Review of implications of dependent nondifferential misclassification of exposure and outcome
- Extension to dependent nondifferential misclassification of exposure and covariate
- Simulation Results - dependent misclassification of exposure and covariate

Review: Exposure Misclassification

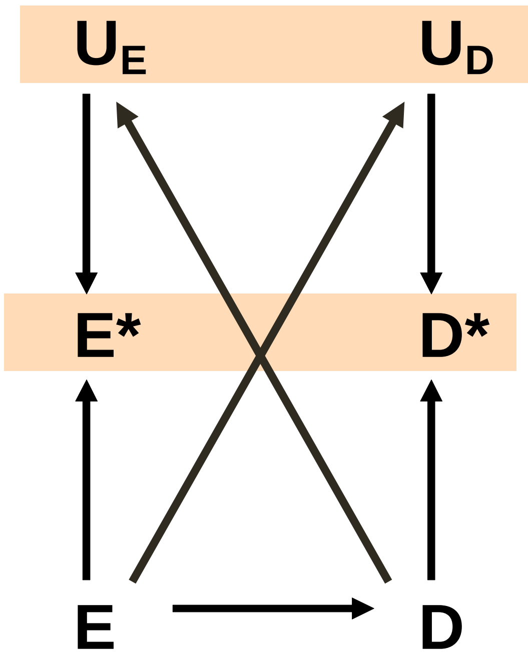
- **Nondifferential**
  - mechanisms that will on average produce equal exposure misclassification probabilities for cases and controls or diseased and nondiseased
  - Se and Sp of exposure classification is equal in each true category of outcome
- **Differential**
  - mechanisms that will on average produce unequal exposure misclassification probabilities for cases and controls or diseased and nondiseased
    - e.g., recall bias

Review: Exposure Misclassification

- **Independent**
  - likelihood of exposure misclassification is NOT predicted by likelihood of misclassification in another variable (e.g., outcome, confounder)

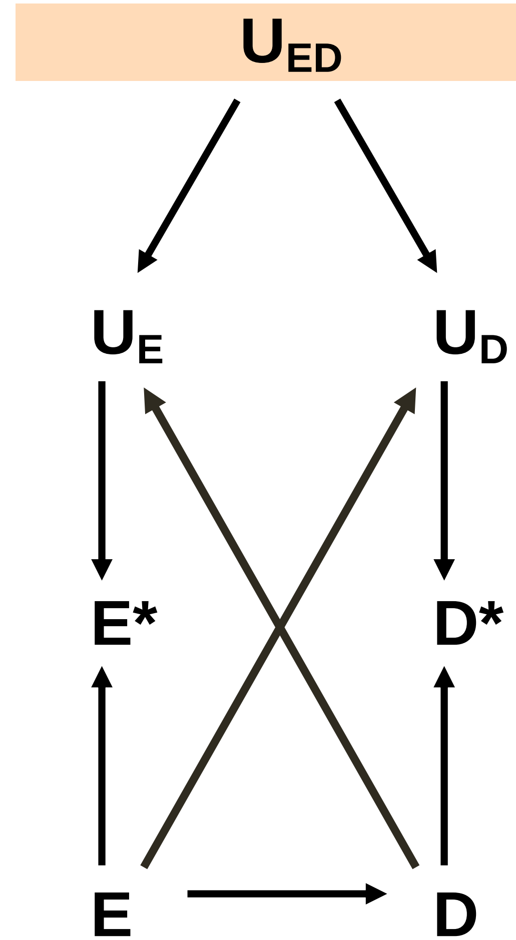
- **Dependent**
  - likelihood of exposure misclassification <u>IS</u> predicted by likelihood of misclassification in another variable (e.g., outcome, confounder)
  - shared mechanism

Causal diagrams for measurement bias[1]



Independent nondifferential joint misclassification

Independent differential joint misclassification



Dependent nondifferential joint misclassification

Dependent differential joint misclassification

1. Hernan and Cole. Am. J. Epidemiol. (2009) 170 (8)

Dependent Misclassification: Mechanisms

- Consistency motif
  - Ex. Saturated fat intake and child obesity
- Implicit theories
  - Ex. NSAID use and GI bleeding
- Social desirability
  - Ex. Flossing and dental caries
- Leniency
  - Ex. Sleep duration and BMI
- Positive/negative affectivity
  - Ex. Neighborhood safety and physical activity
- Transient mood state
  - Ex. EMS response time and recovery from head injury

1.Podsakoff et al. *J Appl Psychol* 2003;88(5).

Dependent Misclassification

- Most likely to occur when information on exposure and other variable come from same source
  - Perhaps worst in studies where exposure and outcome being reported at same time (i.e., cross-sectional)

- Usually self-report but could be other mechanisms
  - e.g., expert review - assignment of both exposure and outcome or covariate

Review: Misclassification

- 70% of original research articles do not evaluate the impact misclassification[1]

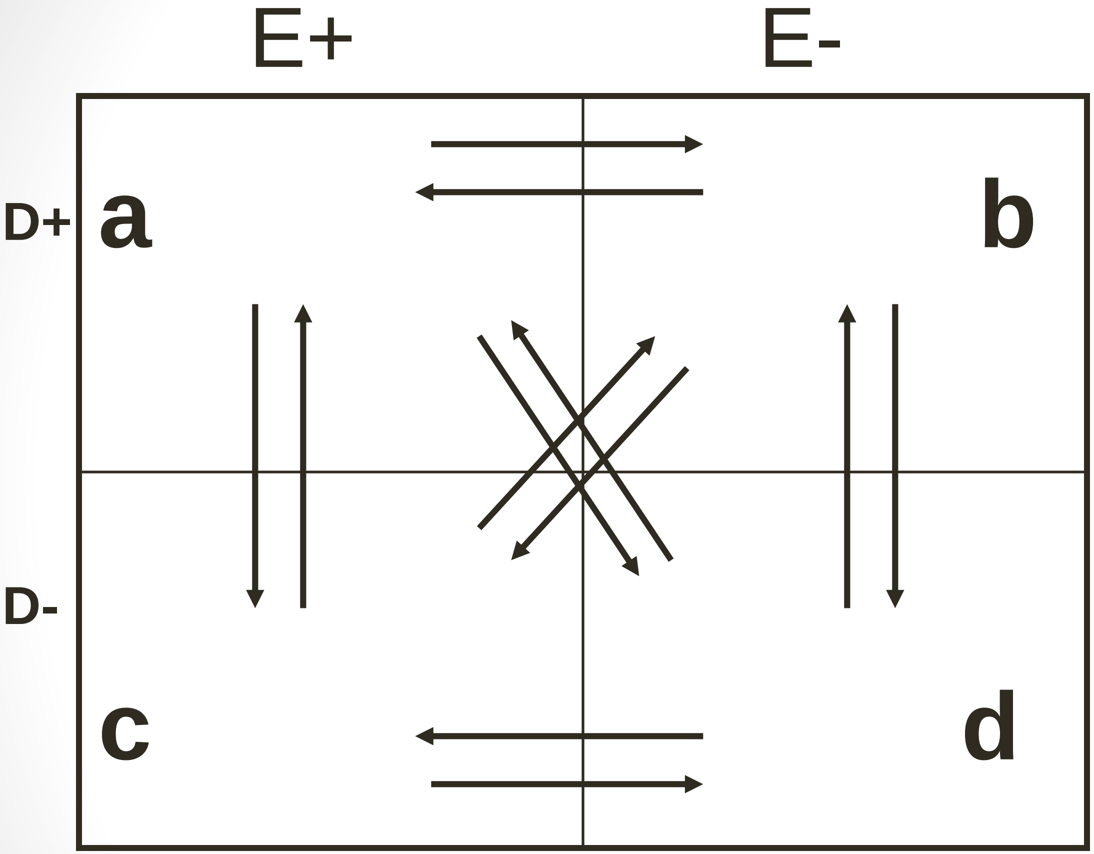
  - Most centered on independent, nondifferential scenario
    - A reliance on the mantra

    nondifferential misclassification = bias toward the null

- To date research into implications has focused on dependent misclassification of exposure and outcome

1. Jurek et al. *Eur J Epidemiol* 2006; 2. Brenner et al. *J Clin Epidemiol* 1993;46(10)

Implications- when “true” OR = 1



Rearrangement resulting from error in classification of E and O

Assumptions:

- **-“true” OR = 1**
- - positive correlation

Then:

- - INDM has no effect on OR
- -net effect = upward bias

Magnitude of bias depends on P(E+), P(D+) and Pr(DNDM)

Dependent Classification of E and D:
Implications- when “true” OR = 1

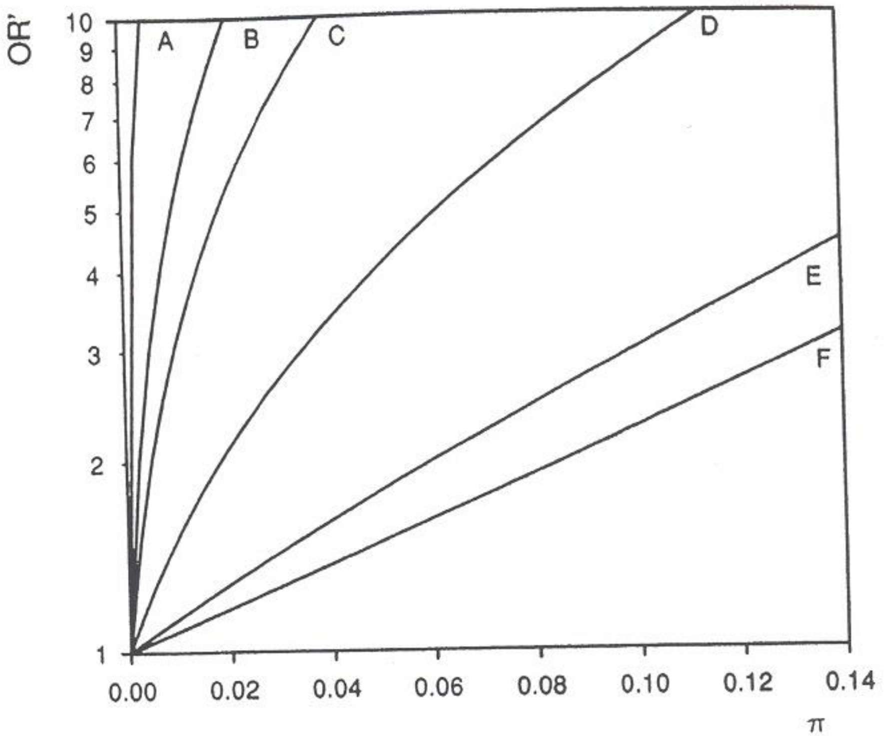


FIGURE 3. Observed odds ratio (OR′) (logarithmic scale) as a function of the probability of nonindependent misclassification of exposure and outcome (π), for different true prevalences of exposure and outcome (A–F), given that the true odds ratio = 1. True exposure and outcome prevalences: A = 0.05, 0.01; B = 0.05, 0.05; C = 0.1, 0.05; D = 0.2, 0.1; E = 0.3, 0.3; F = 0.5, 0.5.

Kristensen P. Epidemiology 1992;3:210-215.

Dependent Classification of E and D:
Effect of Under- vs Over-reporting

p_d	p_e	Pi	q	Over[a]	Under[b]	Both[c]
0.1	0.2	0.9	0.82	1.25	1.02	1.29
0.1	0.2	0.9	0.83	1.55	1.03	1.63
0.1	0.1	0.8	0.65	1.22	1.02	1.24
0.1	0.1	0.8	0.67	1.77	1.06	1.87
0.1	0.1	0.8	0.69	2.56	1.09	2.79
0.05	0.2	0.9	0.82	1.34	1.02	1.37
0.05	0.2	0.9	0.83	1.76	1.03	1.84
0.05	0.1	0.8	0.65	1.25	1.02	1.28
0.05	0.1	0.8	0.67	1.93	1.05	2.02
0.05	0.1	0.8	0.69	2.93	1.09	3.15
0.01	0.2	0.9	0.82	1.46	1.02	1.49
0.01	0.2	0.9	0.83	2.08	1.03	2.15
0.01	0.1	0.8	0.65	1.30	1.02	1.31
0.01	0.1	0.8	0.67	2.11	1.05	2.19
0.01	0.1	0.8	0.69	3.39	1.09	3.56

Pd=outcome prevalence; Pe=exposure prevalence; Pi=marginal sensitivity and/or specificity; q=probability of correct classification for subjects at risk of correlated misclassifications

[a] Overreporting with specificities = Pi

[b] Underreporting with sensitivities = Pi

[c] Over and under reporting with Se and Sp = Pi

Assumes “true” OR=1

Effect of overreporting
>> effect of underreporting

Cross-sectional studies - most likely to be healthy unexposed

Chavance M. Int J Epidemiol 1992;21(3):537-546.

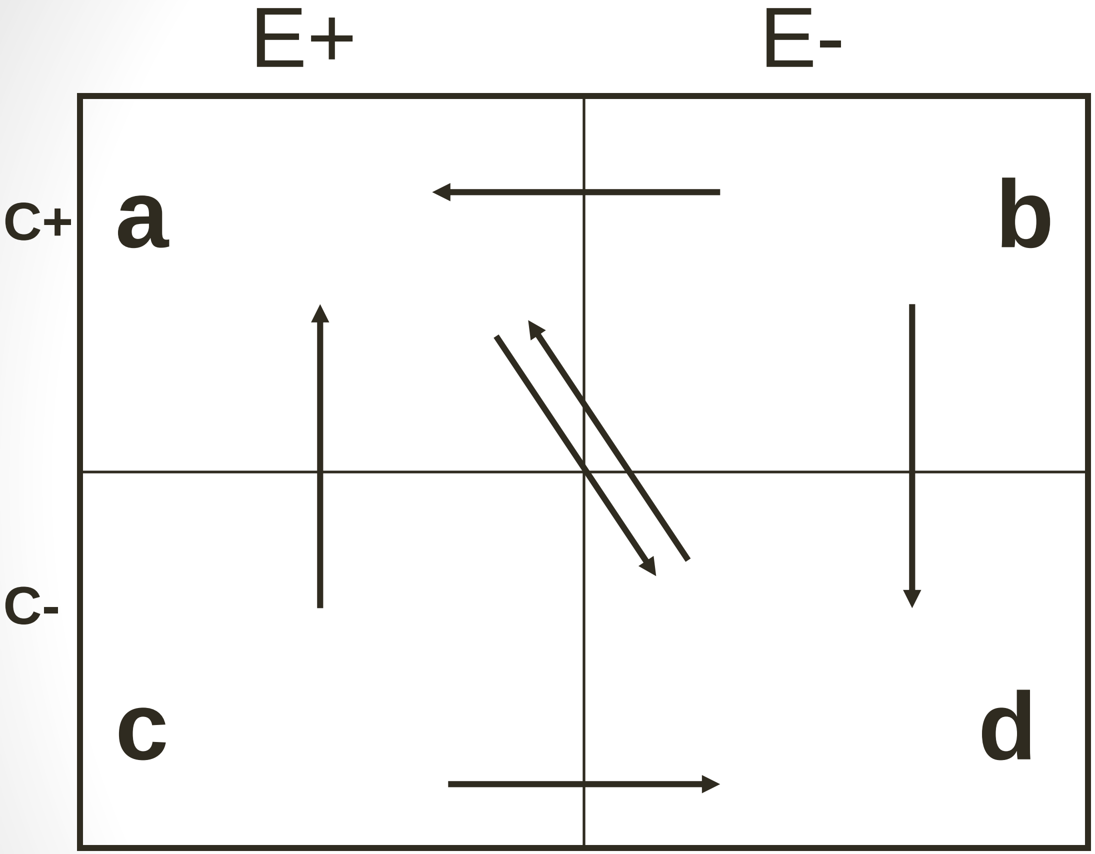
Dependent Classification of E and D:

- **Main solutions**
  - Separate sources of exposure and outcome information
    - Questionnaires, raters
    - Some areas of inquiry more likely to have same sources
  - Design of interview/questionnaire
    - Type of question: “objective” vs. “subjective”

PROBLEM: What about dependent misclassification of exposure and covariate?

- Even if exposure and outcome are determined from different sources, covariate and exposure information are typically measured by a common methods
  - What implications on observed measure of association?
  - Conventional wisdom: adjustment for a confounder that is nondifferentially misclassified results in partial control (some residual confounding)
    - Is this true even if errors in confounder classification are correlated with errors in exposure?

Dependent misclassification of exposure and covariate



Rearrangement resulting from error in classification of E and C

Assumptions:

-D is perfectly classified

-E and C captured via the same source

- positive correlation in errors

Result:

- dependent misclass of E and C will induce an assoc between E and C even if there is truly none

-Impact on observed crude and adjusted measures of association?

Simulation of Dependent Misclassification of Exposure and Covariate

- **Assumptions**
  - C is a risk factor for D
  - Misclassification of E and C is **positively correlated and nondifferential**
    - High threshold on both or low threshold on both
  - No other bias

Simulation of Dependent Misclassification of Exposure and Covariate

1. First simulate the 'truth'

```
data misclass_DEP;
    p_E   =0.1;            * Exposure Prevalence;
    p_D   =0.1;            * Disease Prevalence;
    p_C   =0.1;            * Covariate Prevalence;

    RR_ED=2;                    * RR of D for E+ vs E-;
    RR_CE=1.5;                  * RR of E for C+ vs C-;
    RR_CD=4;                    * RR of D for C+ vs C-;

    /* Create data set with N=100000 */
    do j = 1 to 100000;
        * simulate C independent of E and D;
        C=ranbin(-1,1,min(p_C,1));

        * simululate E dependent on C;
        if C = 1 then E = ranbin(-1,1,min(RR_CE*p_E,1));
        else if C = 0 then E = ranbin(-1,1,min(p_E,1));

        * simulate D based on the baseline risk of D and the C and E variables;
        if C = 1 and E = 1 then D = ranbin(-1,1, min(p_D * RR_CD * RR_ED,1));
        else if C = 1 then D = ranbin(-1,1,min(p_D * RR_CD,1));
        else if E = 1 then D = ranbin(-1,1,min(p_D * RR_ED,1));
        else D = ranbin(-1,1,min(p_D,1));
```

Simulation of Dependent Misclassification of Exposure and Covariate

2. Define misclassification probabilities

```
/* Dependendent Misclassification Prevalences */

joint = 0.05;

        ab=0.0;
        ac=0.0;
        ad= joint;
        ba= joint;
        bc=0.0;
        bd= joint;
        ca= joint;
        cb=0.0;
        cd= joint;
        da= joint;
        db=0.0;
        dc=0.0;

X = ranuni(-1);
```

Simulation of Dependent Misclassification of Exposure and Covariate

3. Create observed (misclassified) data

```
/*  DEPENDENT MISCLASSIFICATION of EXPOSURE and COVARIATE ----
        U = observed (misclassified) exposure
        Z = observed (misclassified) covariate                        */

        if E=1 and C=1 then do;
                if 1-ab lt X le 1 then do;
                U=0 ;
                Z=1;
                end;
                else if 1-(ab+ac) lt X le 1-ab then do;
                U=1 ;
                Z=0;
                end;
                else if 1-(ab+ac+ad) lt X le 1-(ab+ac) then do;
                U=0;
                Z=0;
                end;
                else do;
                U=1;
                Z=1;
                end;
        end;
```

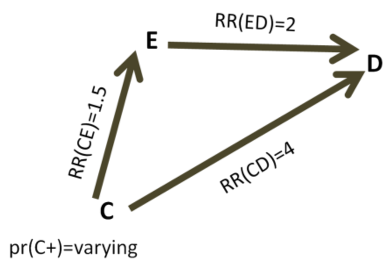
Repeat for all combinations of E,C

Simulation of Dependent Misclassification of Exposure and Covariate

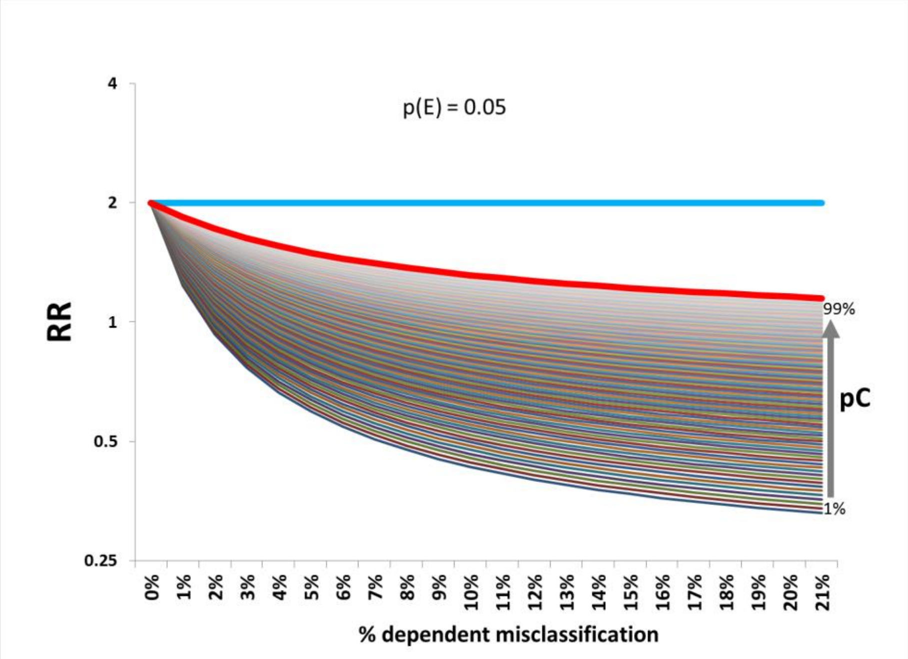
4. **Calculate observed crude and adjusted risk ratios**

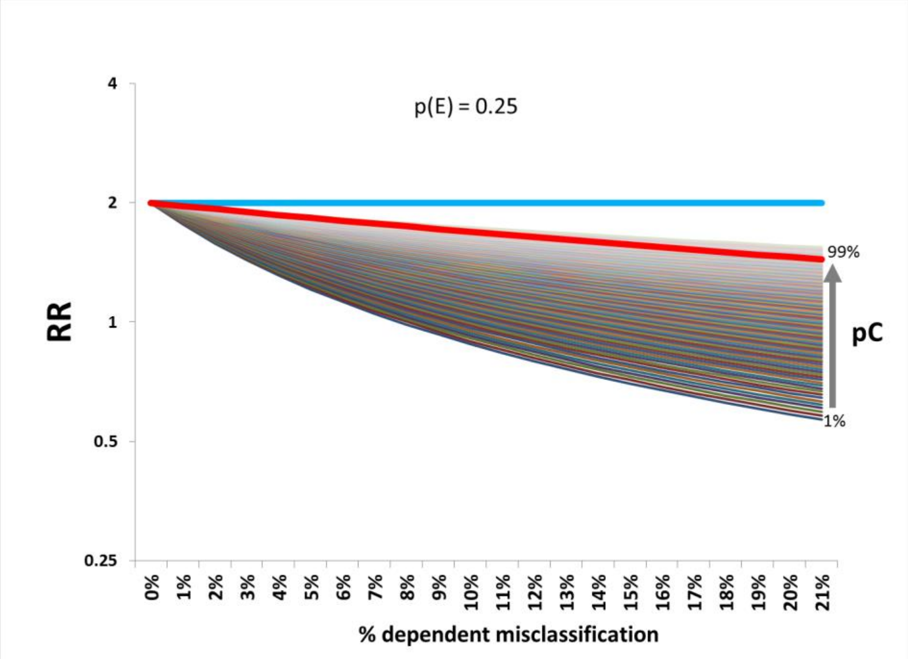
5. **Plot the findings**

Simulation of Dependent Misclassification of Exposure and Covariate



Truth (blue), Crude misclassified (red) and Adjusted RR (grey) for varying % dependent misclassification of E and C and varying prevalence of C





Simulation of Dependent Misclassification of Exposure and Covariate – a closer look

Simulation of Dependent EC misclassification

Input parameters in light blue, results in red

	Pr	+	-
prC+	0.05	0.05	0.95
prE+|C-	0.1	0.1	0.9
RR(CE)	1.5		
prE+|C+		0.15	0.85
prD+|E-C-	0.1	0.1	0.9
RR(CD)	4		
prD+|E-C+		0.4	0.6
RR(ED)	2		
prD+|E+C-		0.2	0.8
prD+|E+C+		0.8	0.2
N	250000		

Misclassification		0.03	
A->B	0	B->A	0.03
A->C	0	B->C	0
A->D	0.03	B->D	0.03
A	0.97	B	0.94
C->A	0.03	D->A	0.03
C->B	0	D->B	0
C->D	0.03	D->C	0
C	0.94	D	0.97

A	B	C	D
E+C+	E-C+	E+C-	E-C-

Simulation of Dependent Misclassification of Exposure and Covariate – a closer look

True data

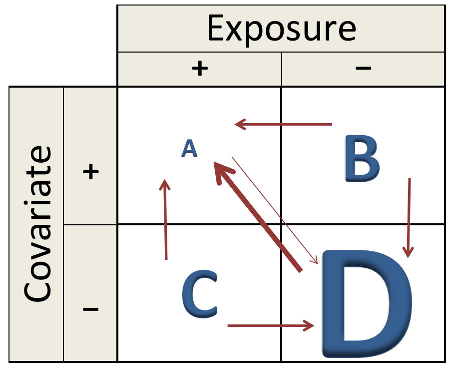
	C+			C-			Total		
	E+	**E-**		**E+**	**E-**		**E+**	**E-**	
D+	1500	4250	**D+**	4750	21375	**D+**	6250	25625	
D-	375	6375	**D-**	19000	192375	**D-**	19375	198750	
Total	1875	10625	**Total**	23750	213750	**Total**	25625	224375	
Risk	0.80	0.40	**Risk**	0.20	0.100	**Risk**	0.24	0.11	
RR	2.00		**RR**	2.00		**True cRR**	2.14		
						True SMR	2.00	RRc=	1.1

Misclassified data

	Z+			Z-			Total		
	U+	**U-**		**U+**	**U-**		**U+**	**U-**	
D+	2366	3995	**D+**	4465	21049	**D+**	6831	25044	
D-	6896	5993	**D-**	17860	187376	**D-**	24756	193369	
Total	9263	9988	**Total**	22325	208425	**Total**	31588	218413	
Risk	0.26	0.40	**Risk**	0.20	0.10099	**Risk**	0.22	0.11	
RR	0.64		**RR**	1.98		**Misc RR**	1.89		
Risk unchanged?	FALSE	TRUE		TRUE	FALSE	**Misc SMR**	1.15	RRc=	1.6

New relation between E (U) and C (Z)			**New relation between C (Z) and D**			**New relation between E (U) and D**		
	Z+	**Z-**		**Z+**	**Z-**		**U+**	**U-**
U+	9263	22325	**D+**	6361	25514	**D+**	6831	25044
U-	9988	208425	**D-**	12889	205236	**D-**	24756	193369
Total	19250	230750	**Total**	19250	230750	**Total**	31588	218413
Risk	0.48	0.10	**Risk**	0.33	0.11	**Risk**	0.22	0.11
RR	4.97	+	**RR**	2.99	+	**RR**	1.89	+

Simulation of Dependent Misclassification of Exposure and Covariate – a closer look



C+ Strata

	Exposure +	Exposure –
Disease+		
Total		
$Risk_{observed}$	<<< $R_{'true'}$	≈ $R_{'true'}$

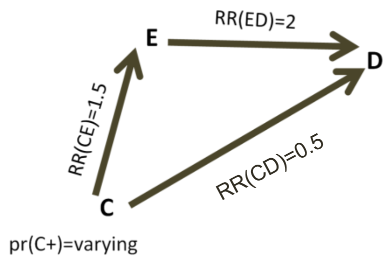
C- Strata

	Exposure +	Exposure –
Disease+		
Total		
$Risk_{observed}$	≈ $R_{'true'}$	≈ $R_{'true'}$

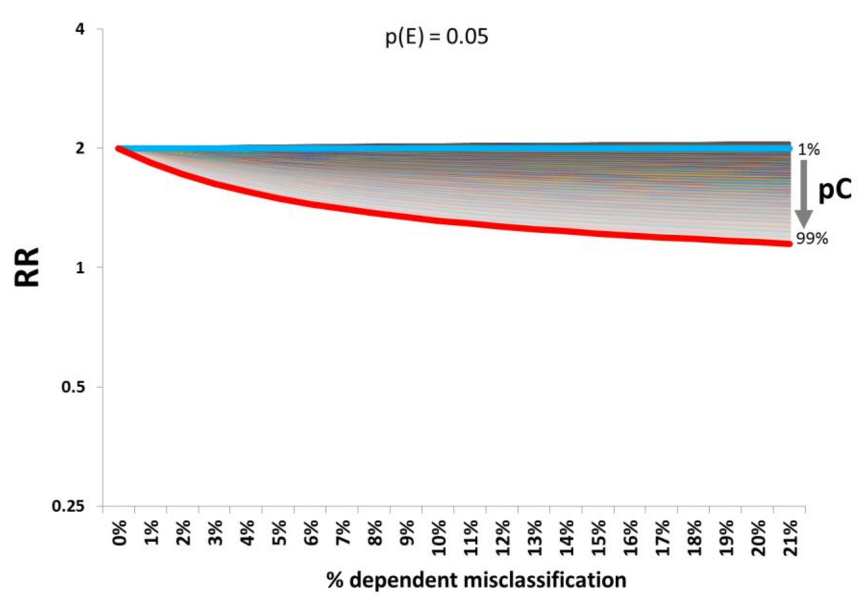
Total

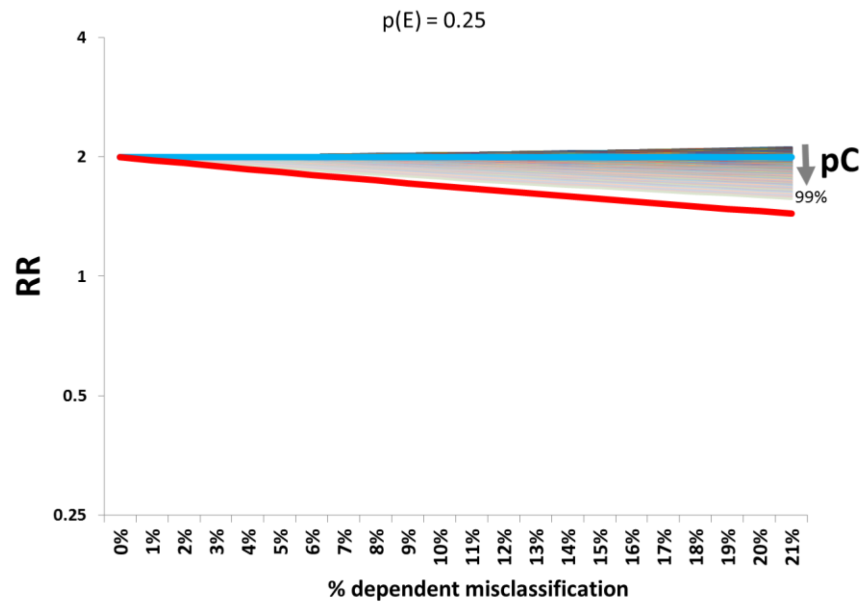
	Exposure +	Exposure –
Disease+		
Total		
$Risk_{observed}$	<< $R_{'true'}$	≈ $R_{'true'}$

Simulation of Dependent Misclassification of Exposure and Covariate



Truth (blue), Crude misclassified (red) and Adjusted RR (grey) for varying % dependent misclassification of E and C and varying prevalence of C





Findings: Dependent Misclassification of Exposure and Covariate

- Identify EMM when there is truly none

- Degree of bias and discrepancy between observed crude and observed adjusted RRs dependent upon pE, pC, RR_CD, RR_ED (and RR_CE), and % dependent misclassification

- By understanding the mechanism acting at the individual level, investigators can assess the likelihood of correlated errors for the variables under study and possible direction of bias

Acknowledgements

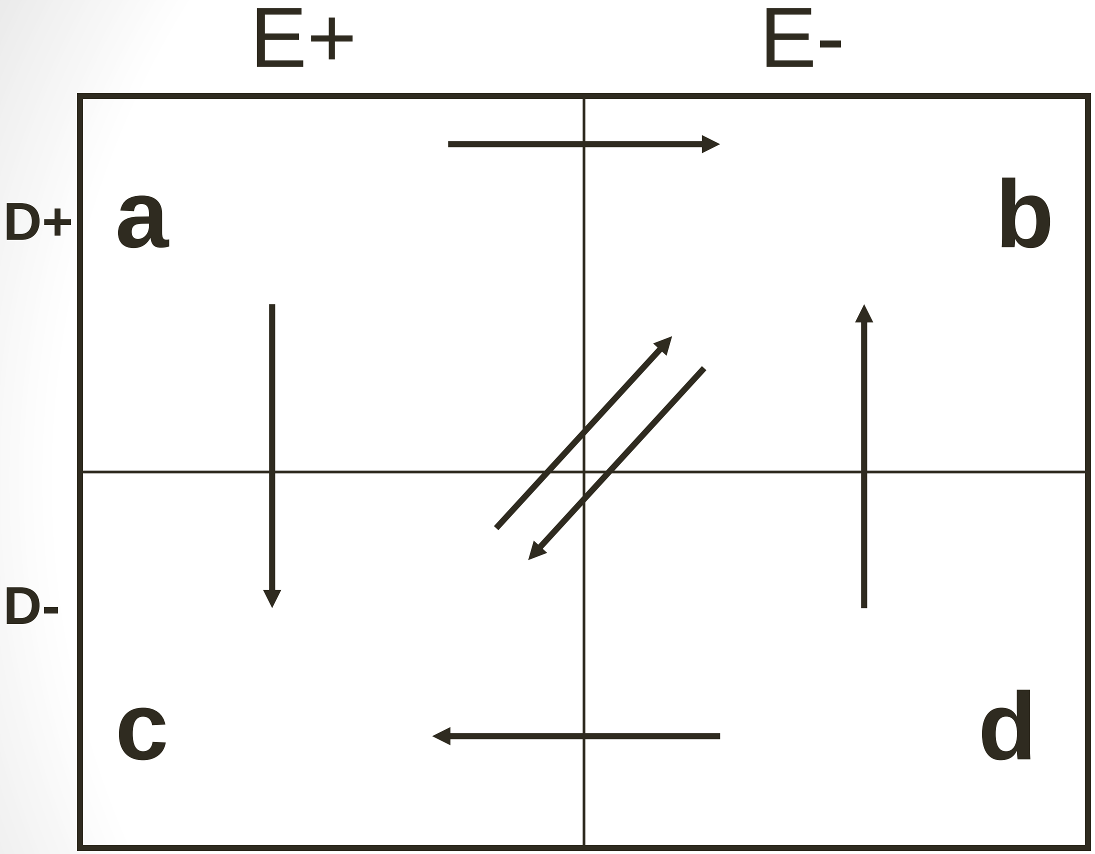
- Alana Brennan
- Matthew Fox
- Daniel Brooks

- My cohort of SimPLE -tons

Dependent Misclassification of Exposure and Outcome

LOGICAL EXTENSIONS OF CURRENT LITERATURE

Dependent Classification of E and D:
Implications- when "true" OR = 1



Rearrangement resulting from error in classification of E and O

Assumptions:

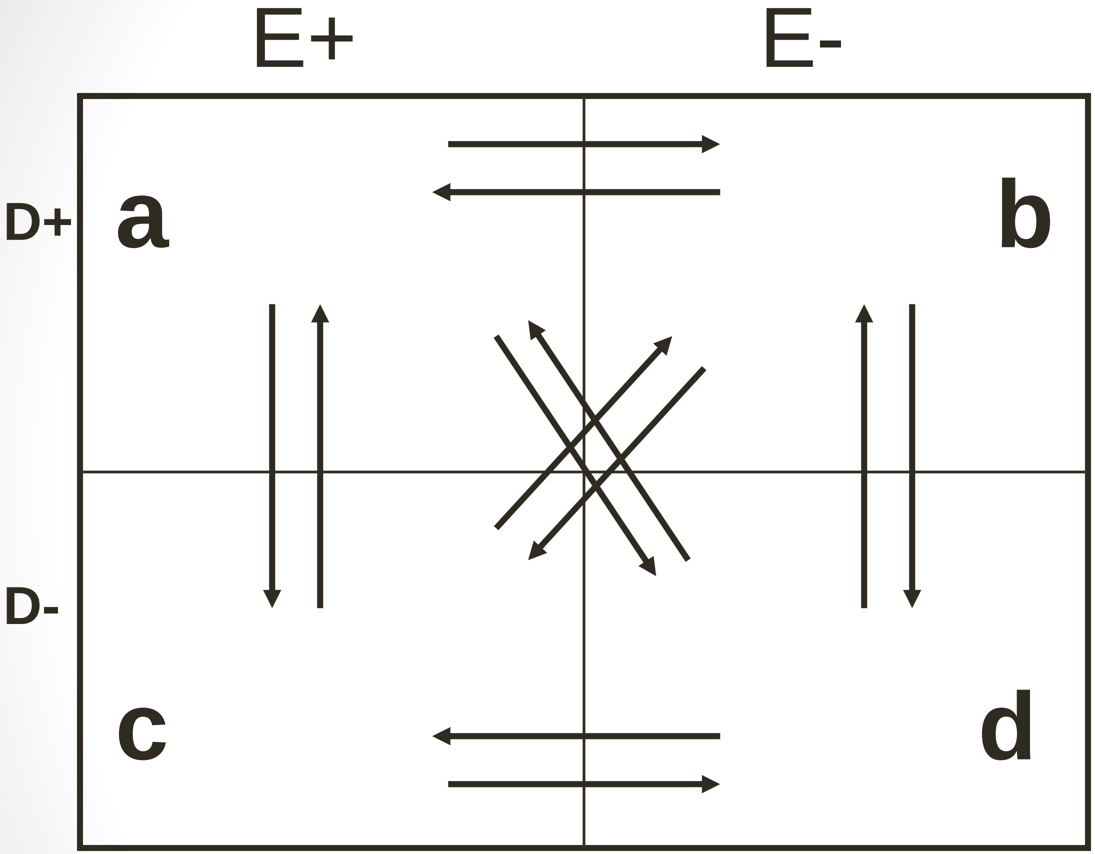
- **"true" OR = 1**
- negative correlation

Then:

- INDM has no effect on OR
- net effect = downward bias

Magnitude of bias depends on Pe, Pd and Pr(DNDM)

Implications – What if “true” OR ≠ 1 ?



Rearrangement resulting from error in classification of E and O

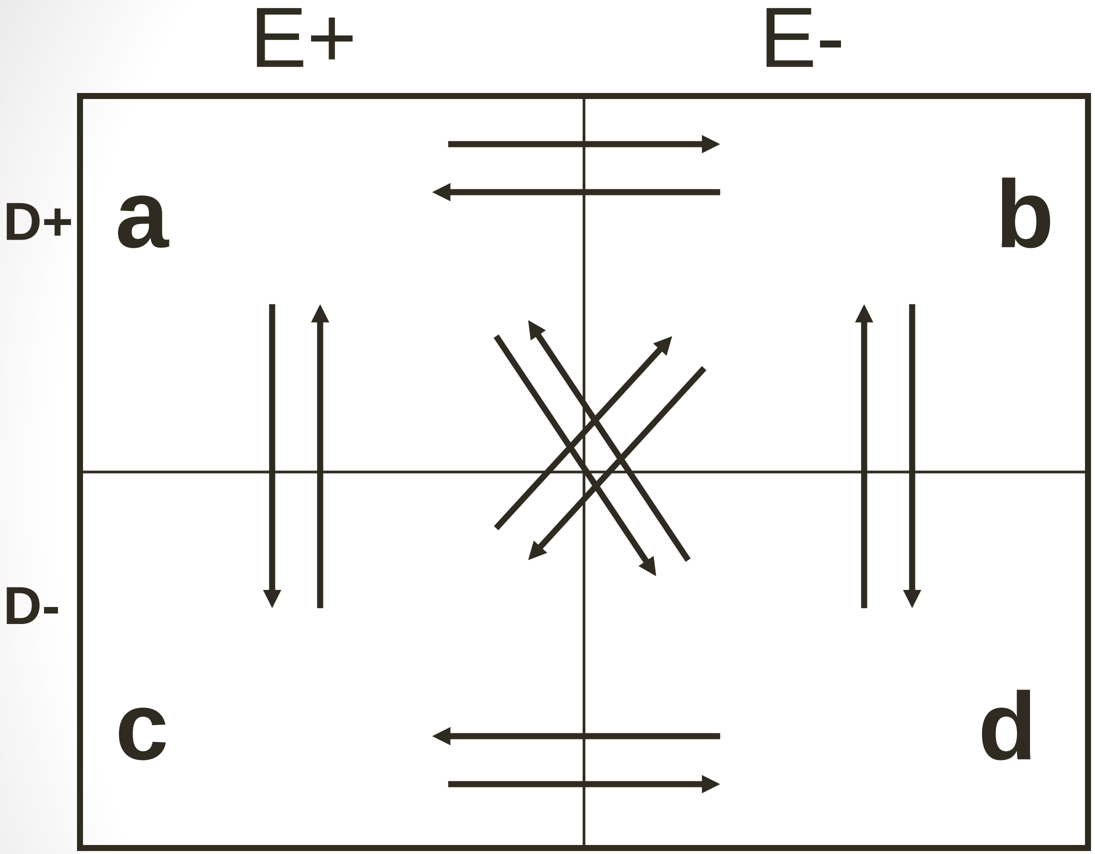
Assumptions:
- positive correlation

NOW:
- INDM will induce downward bias

- In order to bias upward DNDM must overcome

Direction and magnitude of bias depends on Pe, Pd, Pr(DNDM) and “true” OR

Implications – What if "true" OR ≠ 1 ?



Rearrangement resulting from error in classification of E and O

Assumptions:
- negative correlation

NOW:
- INDM will induce downward bias

- DNDM will bias downward even more may even reverse effect

Magnitude of bias depends on Pe, Pd, Pr(DNDM) and "true" OR