Review of methods for sensitivity analysis for unmeasured confounding

Nicola Orsini

October 12, 2018

Biostatistics Team
Department of Public Health Sciences
Karolinska Institutet

Outline

- The problem
- Ordinary sensitivity analysis
- Probabilistic sensitivity analysis
- E-Value
- Final comments

Background

Sensitivity analysis is the study of how the variation in the output of a model can be attributed to different sources of variation.

Methods dealing with uncertainty in model outputs are well known in

- Decision modeling
- Risk analysis

and applied in a variety of industries and applications

Engineering	Health Care
Financial Planning	Pharmaceuticals
Project Management	Consulting
Government	Insurance

Application to epidemiology

The collection of observational data is subject to many sources of uncertainty including errors of measurement, absence of information, and poor or partial understanding of the driving forces and data generation mechanisms.

The two steps of a conventional analysis

Step 1) Use standard statistical methods based on the following not testable assumptions:

1. No unmeasured confounders

2. Random selection, participation, and missing

3. No mismeasurement

Step 2) address possible violations of assumption 1-3 with speculative discussions.

In practice, the assumptions of **Step 1)** may be grossly violated, and the **Step 2)** is often skipped (Greenland 2005).

Various approaches to bias

1. Ignore biases (or hope that they cancel out)
2. Mention something about potential biases
3. Address qualitatively the effect of bias
4. Address quantitatively the effect of bias

Survey of 57 papers published in Epi, AJE, IJE (Jurek, et al. 2006) between Dec 2000 and Oct 2001 about exposure measurement error.

n = 22 (39%) approach Nr 1
n = 34 (60%) approach Nr 2 or Nr 3
n = 1 (1%) approach Nr 4

Why quantitative methods are rarely used?

1. Lack of training in epidemiology and biostatistics courses

2. Limited requests from editors/reviewers

3. Subjectivity in quantifying the values of the parameters controlling the magnitude and direction of bias

4. Not appealing to present additional sources of uncertainty

Hypothetical example

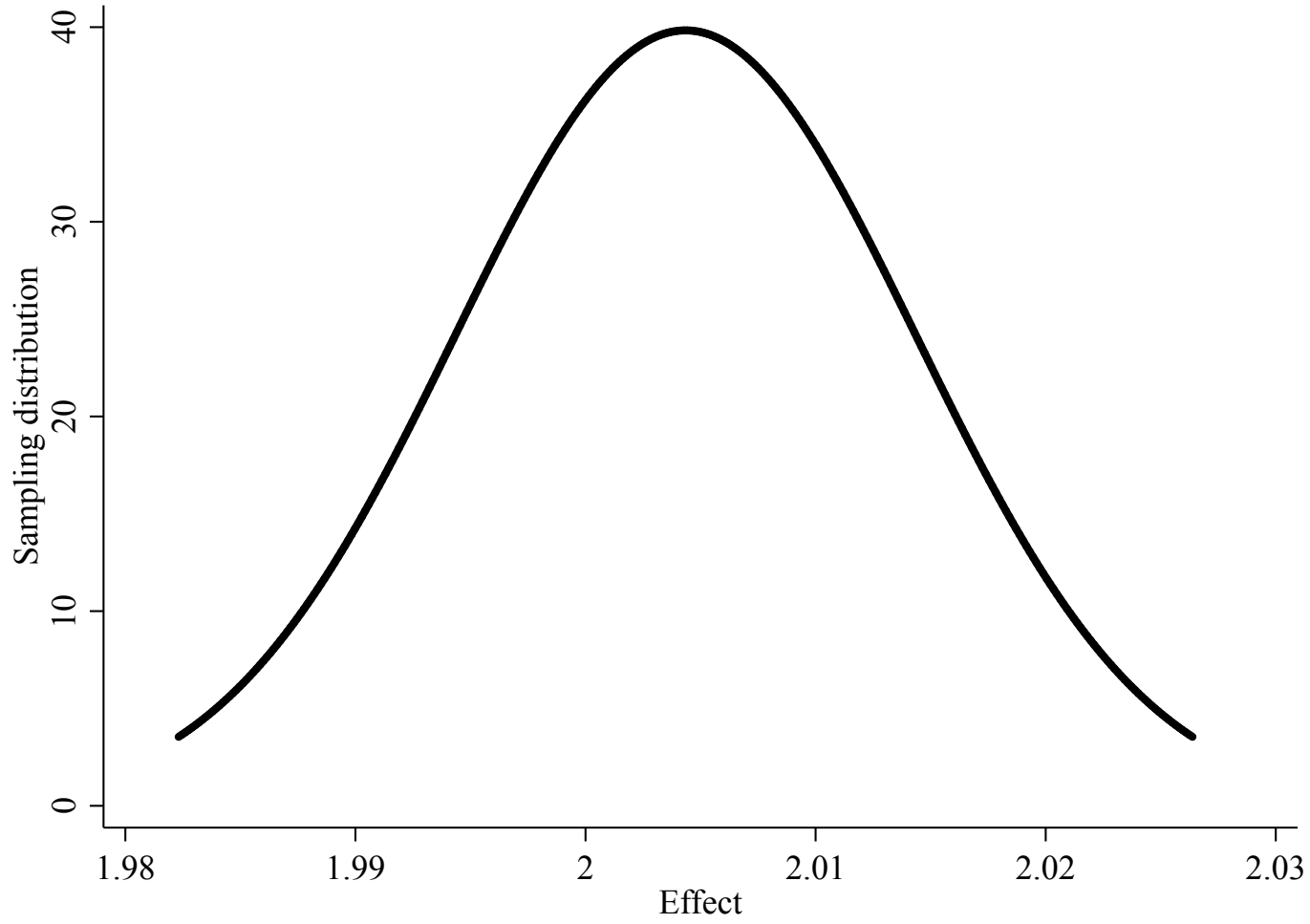
D = lifespan (years)
E = red wine consumption (glasses per day)

Study design: observational based on a big sample

Result: Every additional glass per day of red wine confers, on average, 2 additional years of life.

Question: Is this a *real* effect or is it a *spurious* effect? In other words, is this an effect produced by the variable under study or does it come from uncontrolled or unknown variables?

Repeating the same large study over and over again would not help to answer the question about the nature of the effect.



Questions concerning spurious or confounding effects are often hard to answer

- Experienced investigators tend to develop an intuitive '*feel*' for such matters.

- One important guide is the *size of the effect* in question (Bross, 1966). If it is a slight effect that is barely detectable, then most investigators would feel that there is a distinct possibility that the effect is spurious.

- As the effect becomes increasingly large, most investigators would feel intuitively that the chance that the effect is spurious is thereby reduced.

What other factors can explain the observed findings?

For example, family income (proxy of socio-economic status) can be associated with red wine consumption and life span. Unfortunately, we don't have it.

U = family income (MSek per year)

What is the relationship between income (U) and length of life (D)?

What is the relationship between income (U) and red wine consumption (E)?

Answering these two questions is important to understand the strength of confounding due to income when investigating the relationship between red wine and life span.

How to adjust for a confounder using a standard linear regression model?

$$\text{Mean}(D|E,U) = \beta_0 + \beta_1 E + \beta_2 U$$

1) Remove the confounding effect from the disease

$$\text{Mean}(D|U) = \gamma_0 + \gamma_1 U$$

$$D_r = D - \text{Mean}(D|U)$$

2) Remove the confounding effect from the exposure

$$\text{Mean}(E|U) = \alpha_0 + \alpha_1 U$$

$$E_r = E - \text{Mean}(E|U)$$

3) Examine the residual exposure-disease relationship

$$\text{Mean}(D_r|\, E_r) = \delta_0 + \delta_1 E_r$$

The confounder-adjusted exposure-disease association is given by either β_1 or δ_1.

To judge the chance that an effect is spurious or due to the unmeasured factor U, one need to have an idea of the confounding associations (γ_1 and α_1).

The ratio between observed MD_{ED}^{obs} and adjusted $MD_{ED|U}^{adj}$ effect is unlikely to be 1 in the presence of confounding effects.

$\gamma_1 = 3$ Every 1 additional MSek in family income confers 3 additional years of life.

$\alpha_1 = 0.5$ Every 1 additional MSek in family income is associated with an additional half glass of red wine per day.

The concern is that the beneficial effect of red wine on life span can be explained away by the effect of family income.

```
. regress d e
------------------------------------------------------------------------------
           d |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           e |    2.004348   .0100152   200.13   0.000     1.984719    2.023978
       _cons |    70.00157   .0100038  6997.47   0.000     69.98196    70.02118
------------------------------------------------------------------------------
```

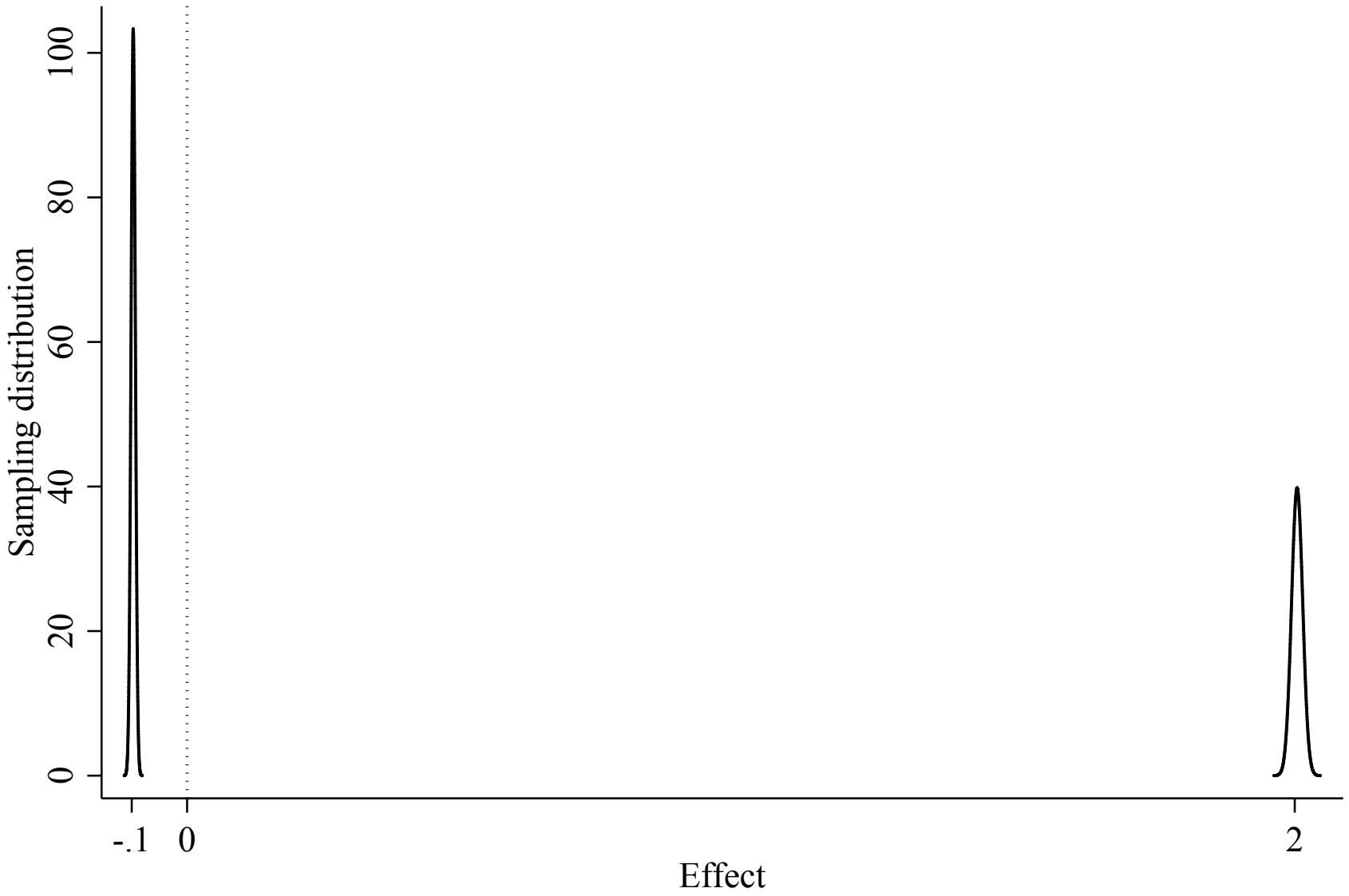
$MD_{ED}^{obs} = 2$ Every additional glass per day of red wine confers 2 additional years of life.

```
. regress d e u
------------------------------------------------------------------------------
           d |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           e |   -.0972068    .003861   -25.18   0.000    -.1047743   -.0896392
           u |    2.995126   .0031537   949.71   0.000     2.988945    3.001307
       _cons |    70.00098   .0031604  2.2e+04   0.000     69.99478    70.00717
------------------------------------------------------------------------------
```

$MD_{ED|U}^{adj} = -0.1$ For any income level, every additional glass per day of red wine reduces life span by approximately 1 month (1/12).

The unadjusted beneficial effect of red wine on life span became a harmful effect when adjusting for family income.

Adjusted vs unadjusted



Lessons learned from this hypothetical example

- Ignorance about the data generating mechanism lead to confounding bias
- A central concern with observational data
- Large sample size is not mitigating the problem of confounding
- The ability to adjust the effect of the exposure for a confounder depends on knowing the confounder-disease (UD) and confounder-exposure (EU) associations as well as the distribution of the confounder (U).

Ordinary sensitivity analysis

- It estimates what the measure of effect would be in light of the observed data and some hypothetical level of bias.

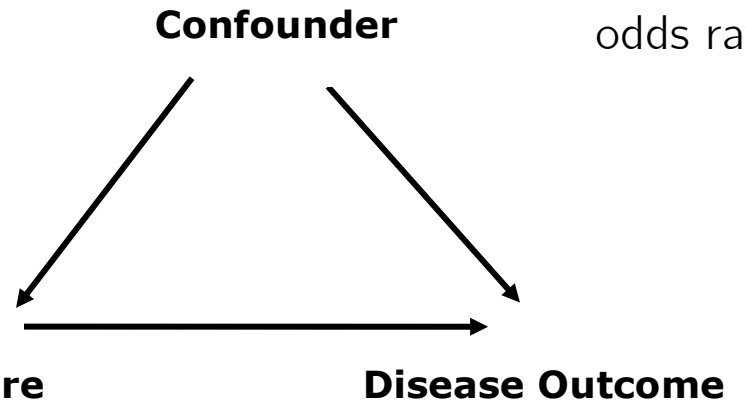
- The idea is to back-calculate the data that would have been observed without bias, assuming particular values for the bias parameters.

- Ordinary sensitivity analysis can be seen as a series of educated guesses about the bias parameters (Greenland 1996, Greenland & Lash, 2008).

Unmeasured or uncontrolled confounder

P_{U1} = Prevalence of the confounder among the exposed

P_{U0} = Prevalence of the confounder among the unexposed



OR_{UD}= confounder-disease odds ratio

A confounder is associated with the exposure and is also an independent risk factor of the disease outcome.

If either association is non-existent, there is no confounding.

The bias parameters are P_{U1} , P_{U0}, and OR_{UD}

Unadjusted exposure-disease association

Let E denote the binary exposure, D denote a binary outcome, and U denote one unmeasured confounder.

	E=1	E=0	Total
$D = 1$	A_{1+}	A_{0+}	M_{1+}
$D = 0$	B_{1+}	B_{0+}	M_{0+}

$$OR_{ED}^{obs} = \frac{\text{Odds}(D = 1|E = 1)}{\text{Odds}(D = 1|E = 0)} = \frac{\left(\frac{A_{1+}}{B_{1+}}\right)}{\left(\frac{A_{0+}}{B_{0+}}\right)}$$

OR_{ED}^{obs} is the observed or apparent odds ratio relating exposure to disease.

Expected table of counts stratified by confounder

General layout (expected data) for sensitivity analysis and external adjustment for a dichotomous confounder U

	$U = 1$			$U = 0$		
	$E = 1$	$E = 0$	Total	$E = 1$	$E = 0$	Total
$D = 1$	A_{11}	A_{01}	M_{11}	$A_{1+} - A_{11}$	$A_{0+} - A_{01}$	$M_{1+} - M_{11}$
$D = 0$	B_{11}	B_{01}	M_{01}	$B_{1+} - B_{11}$	$B_{0+} - B_{01}$	$M_{0+} - M_{01}$

$$OR_{ED|U=1} = \frac{\text{Odds}(D = 1|E = 1|U = 1)}{\text{Odds}(D = 1|E = 0|U = 1)} = \frac{\left(\frac{A_{11}}{B_{11}}\right)}{\left(\frac{A_{01}}{B_{01}}\right)}$$

$$OR_{ED|U=0} = \frac{\text{Odds}(D = 1|E = 1|U = 0)}{\text{Odds}(D = 1|E = 0|U = 0)} = \frac{\left(\frac{A_{1+} - A_{11}}{B_{1+} - B_{11}}\right)}{\left(\frac{A_{0+} - A_{01}}{B_{0+} - B_{01}}\right)}$$

Suppose the smoking prevalences among the exposed and unexposed populations are estimated or assumed to be P_{U1} and P_{U0} and the odds ratio relating the confounder and disease within levels of exposure is OR_{UD}.

Assuming the control group is representative of the source population, we set
$B_{11} = B_{1+}P_{U1}$
$B_{01} = B_{0+}P_{U0}$

Next, to find A_{11} and A_{01}, we solve the pair of equations

$$OR_{UD|E=1} = \frac{\text{Odds}(D = 1|E = 1, \text{U} = 1)}{\text{Odds}(D = 1|E = 1, \text{U} = 0)} = \frac{\left(\frac{A_{11}}{A_{1+} - A_{11}}\right)}{\left(\frac{B_{11}}{B_{1+} - B_{11}}\right)}$$

$$OR_{UD|E=0} = \frac{\text{Odds}(D = 1|E = 0, \text{U} = 1)}{\text{Odds}(D = 1|E = 0, \text{U} = 0)} = \frac{\left(\frac{A_{01}}{A_{0+} - A_{01}}\right)}{\left(\frac{B_{01}}{B_{0+} - B_{01}}\right)}$$

Assuming that the confounder-disease association is similar (homogeneous) within levels of the exposure ($OR_{UD|E=1} \approx OR_{UD|E=0}$) we have the solutions

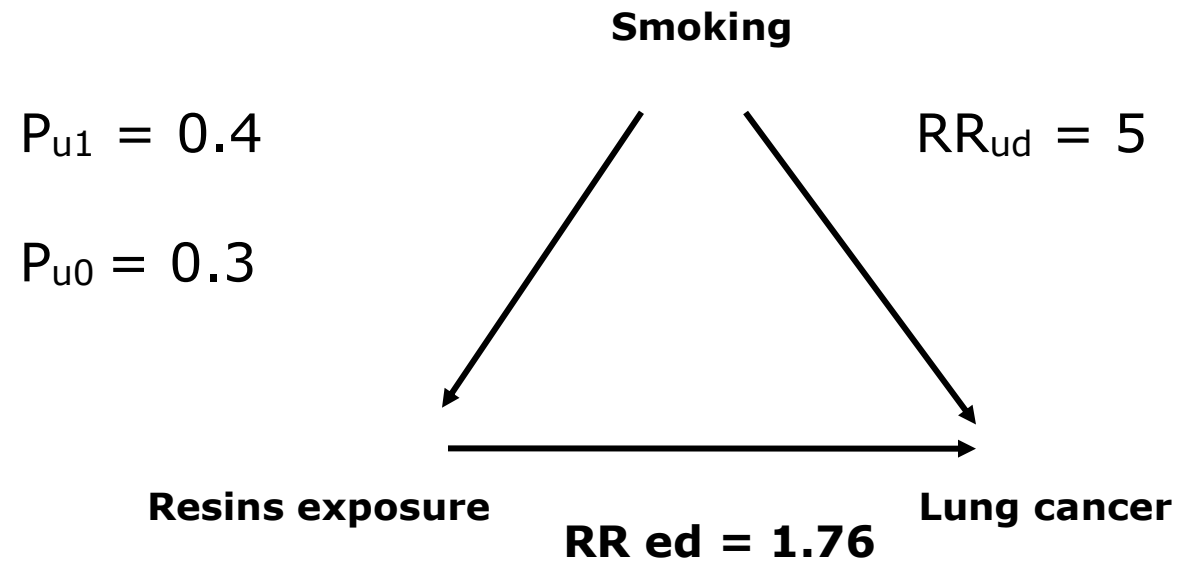
$$A_{11} = OR_{UD|E=1} A_{1+} B_{11} / (OR_{UD|E=1} B_{11} + B_{1+} - B_{11})$$

$$A_{01} = OR_{UD|E=0} A_{0+} B_{01} / (OR_{UD|E=0} B_{01} + B_{0+} - B_{01})$$

Having obtained data counts corresponding to A_{11}, A_{01}, B_{11}, and B_{01}, we can put these numbers in the table reported in the previous slide and directly compute a U-adjusted estimate of the exposure-disease odds ratios, OR_{ED}^{adj}.

We could estimate an OR_{ED}^{adj} using Mantel-Haenszel method or maximum likelihood method based on summarized data.

Example – Occupational resins and lung cancer



Binary outcome: Lung cancer

Binary exposure: Resins exposure, yes vs no

Binary unmeasured confounder: Smoking, yes vs no

Case-control data

TABLE. *Crude data for case-control study of occupational resins exposure (E) and lung cancer mortality (Greenland et al., 1994).*

	E=1	E=0	Total
Cases (D = 1)	$A_{1+} = 45$	$A_{0+} = 94$	$M_{1+} = 139$
Controls (D = 0)	$B_{1+} = 257$	$B_{0+} = 945$	$M_{0+} = 1202$

Odd ratio after adjustment for age and death year: 1.77
Age-year adjusted conventional 95% confidence limits for OR_{ED}: 1.18, 2.64

$OR_{ED}^{obs} = 1.76$ Occupational exposure to resins was associated with 76% higher mortality odds (95% CI = 1.20, 2.58).

P_{U1}=0.4 the fraction of smokers among those exposed to resins
P_{U0}=0.3 the fraction of smokers among those not exposed to resins
RR_{UD} = 5 the association between smoking and lung cancer

Data on occupational resins exposure and lung cancer

	E=1	E=0
D=1	45	94
D=0	257	945

Hypothetical data on resins exposure and lung cancer according to smoking

	U = 1		U = 0	
	E=1	E=0	E=1	E=0
D=1	35	64	10	30
D=0	103	284	154	661

```
B11 = 257*.4 = 103
B01 = 945*.3 = 284
B10 = 257-103 = 154
B10 = 945-284 = 661
A11 = 5*45*103/(5*103+257-103) = 35
A01 = 5*94*284/(5*284+945-284) = 64
A10 = 45-35 = 10
A00 = 94-64 = 30
```

Mantel-Haenszel adjusted odds ratio

OR_{ED}^{adj} = 1.49 (95% CI = 1.00, 2.21, p=0.05)

Logistic regression adjusted odds ratio

OR_{ED}^{adj} = 1.49 (95% CI = 1.00, 2.21, p=0.05)

Smoking-adjusted effect of occupational exposure to resins was associated with 49% higher mortality odds.

Table. Deterministic sensitivity analysis of the resins-cancer odds ratios to choice of different values for the bias parameters: smoking prevalences among exposed (P_{u1}) and unexposed (P_{u0}), and the smoking-lung cancer relative risk (RR_{ud}).

P_{u1}	P_{u0}	OR_{ce}	RR_{cd}		
			5	10	15
0.40	0.30	1.56	1.49	1.42	1.39
0.55	0.45	1.49	1.54	1.49	1.48
0.70	0.60	1.56	1.57	1.54	1.53
0.45	0.25	2.45	1.26	1.13	1.09
0.60	0.40	2.25	1.35	1.27	1.24
0.75	0.55	2.45	1.41	1.35	1.33

The observed unadjusted resins-lung cancer odds ratio is 1.8 (95% CI, 1.2-2.6).

OR_{ue} is the confounder-exposure OR, calculated from the prevalences P_{u1} and P_{u0}.

Relation of the unadjusted to adjusted odds ratios

An equivalent approach to that just given uses the following formulas for the ratio of the unadjusted to *U*-adjusted odds ratios:

$$\frac{OR_{ED}^{obs}}{OR_{ED}^{adj}} = \frac{OR_{UD}P_{U1} + 1 - P_{U1}}{OR_{UD}P_{U0} + 1 - P_{U0}}$$

Assuming *U* is the sole uncontrolled confounder, this ratio can be interpreted as the degree of bias due to failure to adjust for *U*.

This series of equations show that when *U* is not associated with the disease ($OR_{UD} = 1$) or is not associated with exposure ($P_{U1} = P_{U0}$), the ratio of the unadjusted and adjusted odds ratios is 1, and so there is no confounding by *U*.

In other words, a confounder must be associated with the exposure and the disease in the source population.

PHARMACOEPIDEMIOLOGY AND DRUG SAFETY 2006; **15**: 291–303
Published online 31 January 2006 in Wiley InterScience (www.interscience.wiley.com). **DOI**: 10.1002/pds.1200

ORIGINAL REPORT

Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics†

Sebastian Schneeweiss MD, ScD*

Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

http://www.drugepi.org

Array approach

The confounded relative risk (RR), which we call apparent RR (ARR), can be expressed as the 'true' or fully adjusted RR times Bias ($\mathrm{ARR} = \mathrm{RR} \times \mathrm{Bias_M}$), which is an expression of the imbalance of a binary confounding factor among exposed (P_{C1}) and unexposed (P_{C0}, using the notation in Table 2):[21]

$$\mathrm{ARR} = \mathrm{RR} \times \frac{P_{\mathrm{C1}}(\mathrm{RR_{CD}} - 1) + 1}{P_{\mathrm{C0}}(\mathrm{RR_{CD}} - 1) + 1} \qquad (1a)$$

bias = [(0.4*(5-1)+1)/(.3*(5-1)+1)]

1.76 / bias = 1.49

Results are equivalent of the analysis of the stratified table of expected data.

Historical note on sensitivity analysis for confounding

The idea of sensitivity analysis for the simple 2×2 table with an unmeasured binary confounder was originally introduced by Cornfield (Cornfield, et al. 1959, Bross 1966) and further explained by Schlesselman, Greenland and others (Greenland 1996, Lin, et al. 1998, Rosenbaum and Rubin 1983, Schlesselman 1978)

Lin (Lin, et al. 1998) have shown that above formula holds for different statistical regression models (for instance logistic and Cox regression) applied to both binary response (case-control) and censored survival time (cohort) data.

Recently Vanderwheele generalized methods for sensitivity analysis *without* assumptions (Ding & Vanderwheele, 2016).

Limitation of ordinary sensitivity analysis

- Lack probability structure for the bias parameters
- Fail to discriminate among the different scenarios in terms of their likelihood
- It is not easy to summarize results

Probabilistic sensitivity analysis

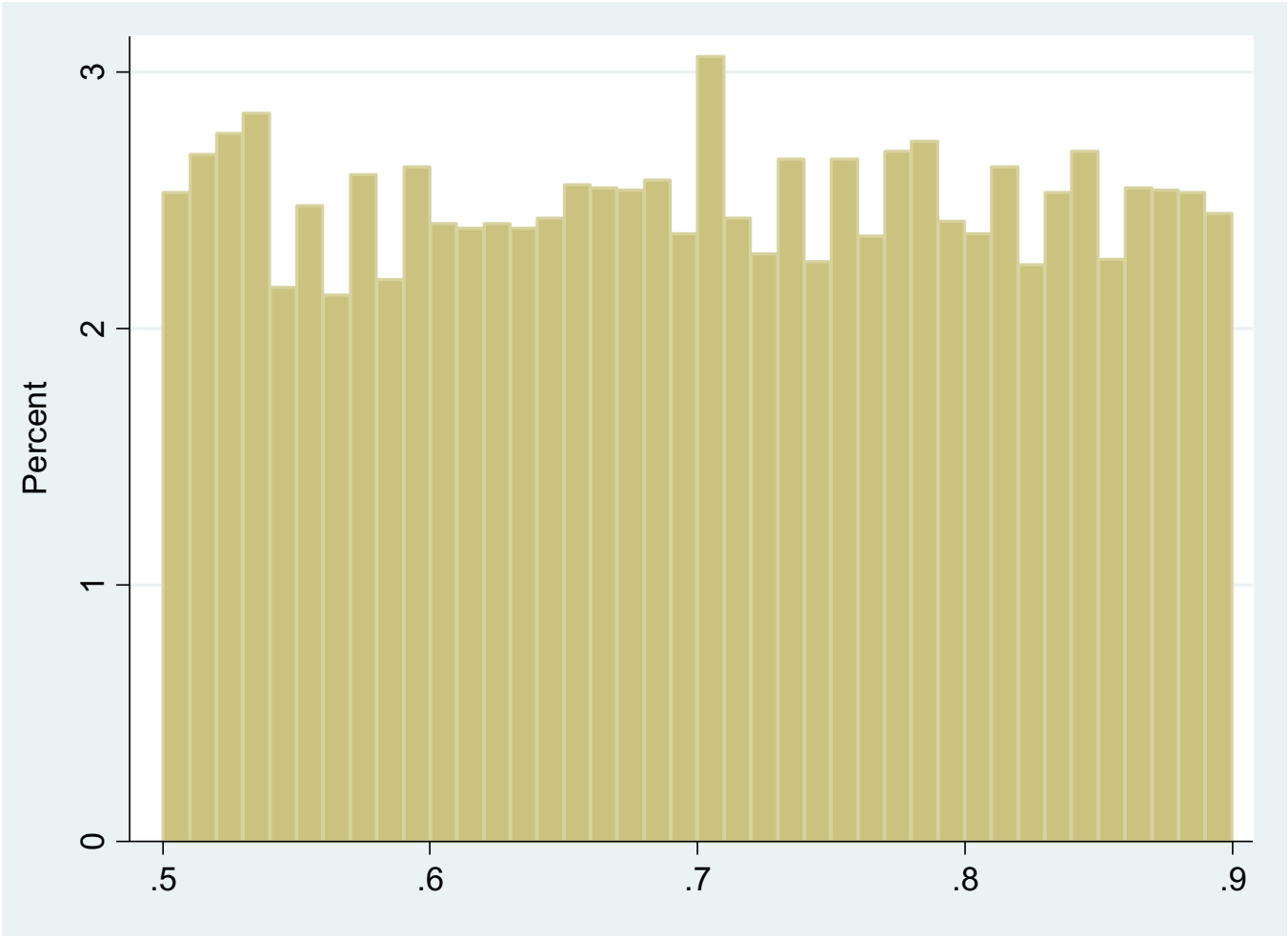
A more realistic approach allows for uncertainty in the bias parameters (Steenland and Greenland 2004).

By specifying a probability distribution for the bias parameters, the bias-adjusted relative risk reflects the uncertainty in the bias parameters.

For example, the statistical software **episens** (Stata) allows the user to specify a variety of probability densities for the bias parameters, and use these densities to obtain simulation limits for the bias adjusted exposure-disease measure of effect.

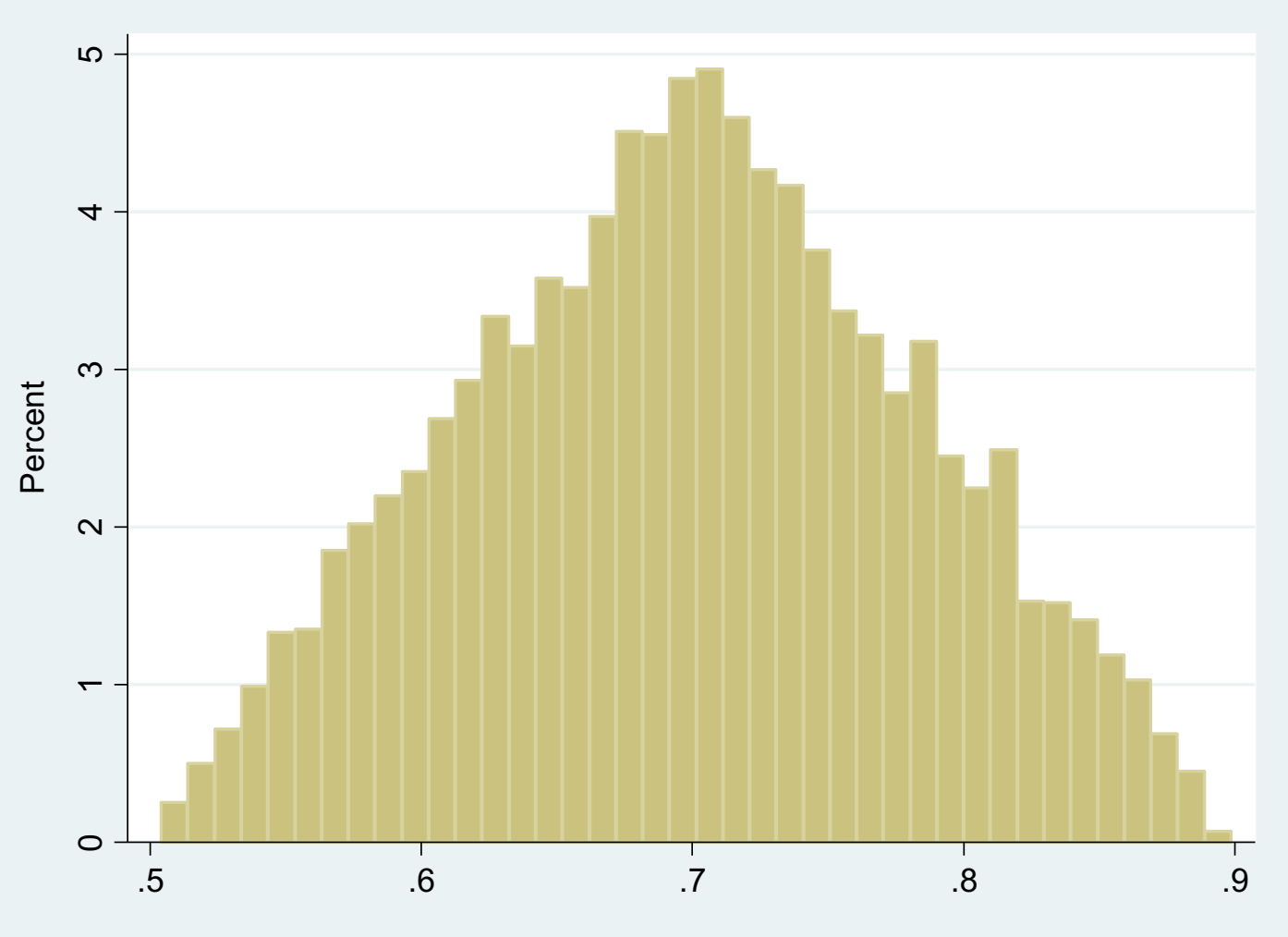
Type of systematic error and bias parameters	Description	Probability density functions
Misclassification of the exposure		
dseca	Sensitivity cases	constant(k)
dspca	Specificity cases	uniform(a b)
dsenc	Sensitivity non-cases	triangular(a b c)
dspnc	Specificity non-cases	trapezoidal(a b c d)
		logit-logistic(m s [lb ub])
		logit-normal(m s [lb ub])
Selection bias		
dpscex	Pr selection cases exposed	constant(k)
dpscun	Pr selection cases unexposed	uniform(a b)
dpsnex	Pr selection non cases exposed	triangular(a b c)
dpsnun	Pr selection non case sunexposed	trapezoidal(a b c d)
		logit-logistic(m s [lb ub])
		logit-normal(m s [lb ub])
dsbfactor	Selection bias factor	constant(k)
		log-normal(m s)
		log-logistic(m s)
Unmeasured confounding		
dpexp	Pr confounder exposed	constant(k)
dpunexp	Pr confounder unexposed	uniform(a b)
		triangular(a b c)
		trapezoidal(a b c d)
		logit-logistic(m s [lb ub])
		logit-normal(m s [lb ub])
drrcd	RR confounder-disease	constant(k)
dorce	OR confounder-exposure	log-normal(m s)
		log-logistic(m s)

Uniform distribution



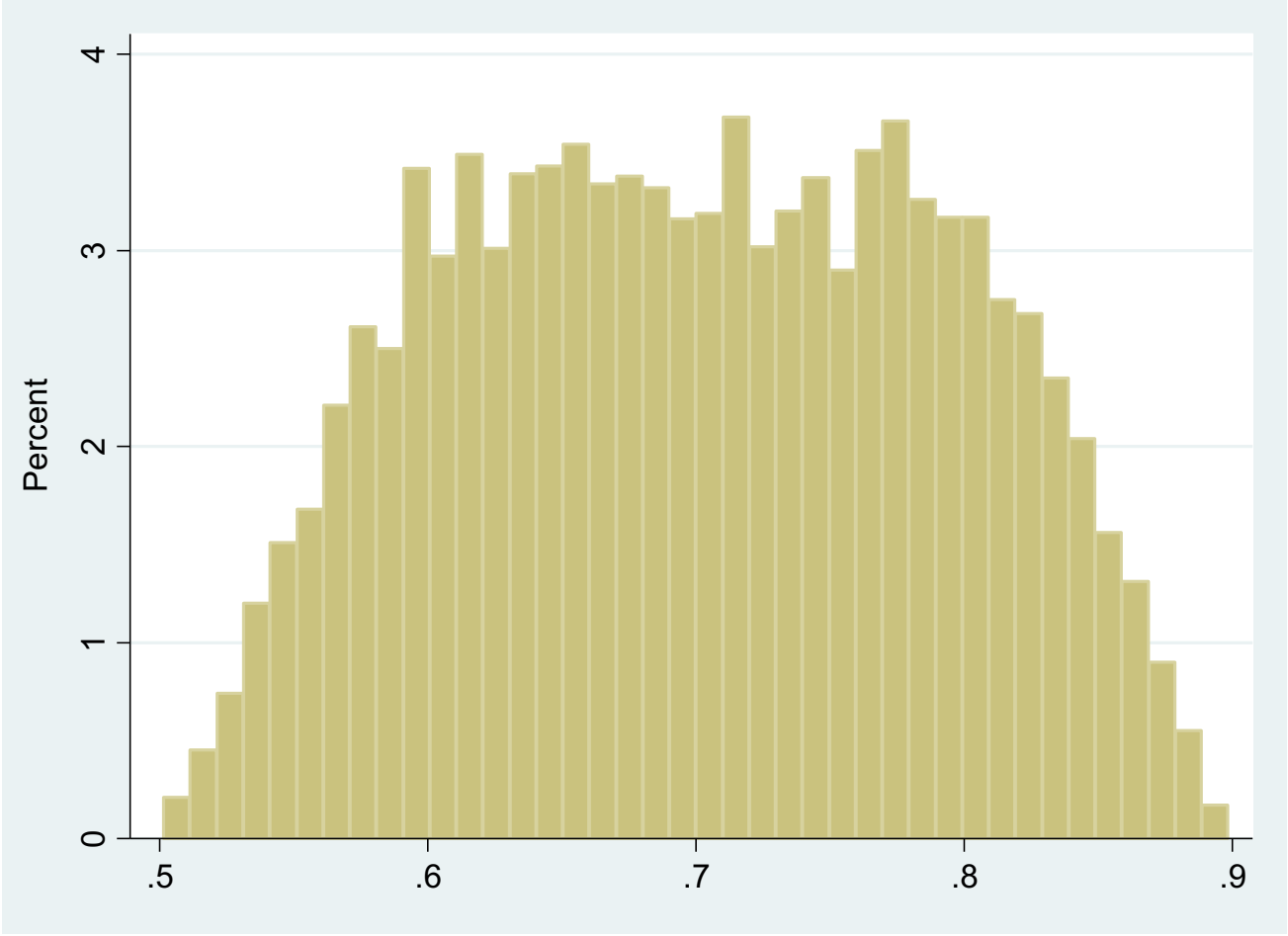
All the values within the specified bounds (a=.5, b=.9) are equally probable

Triangular distribution



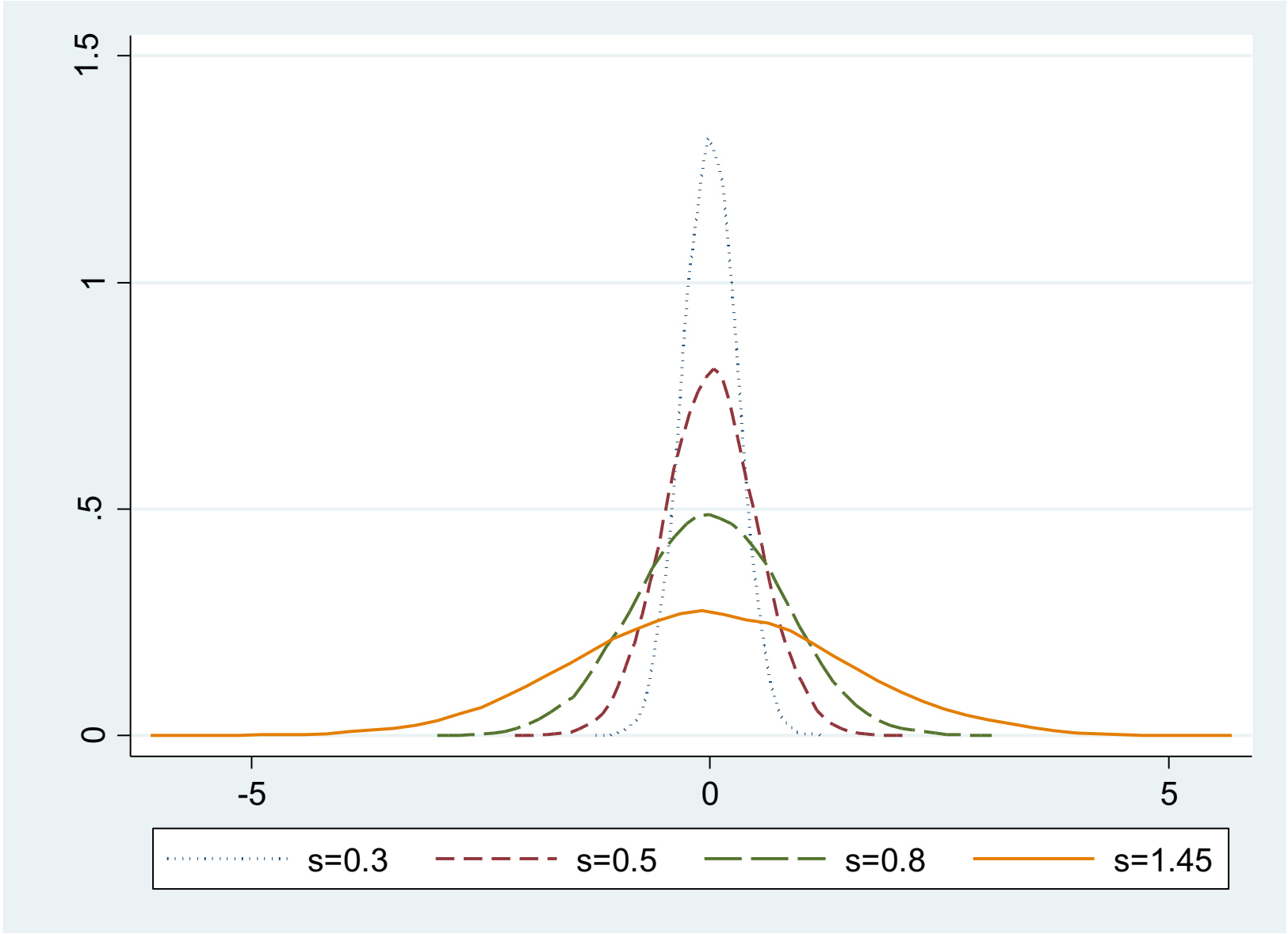
There is a mode (most likely value, b=.7) within the specified bounds (a=.5, c=.9)

Trapezoidal distribution

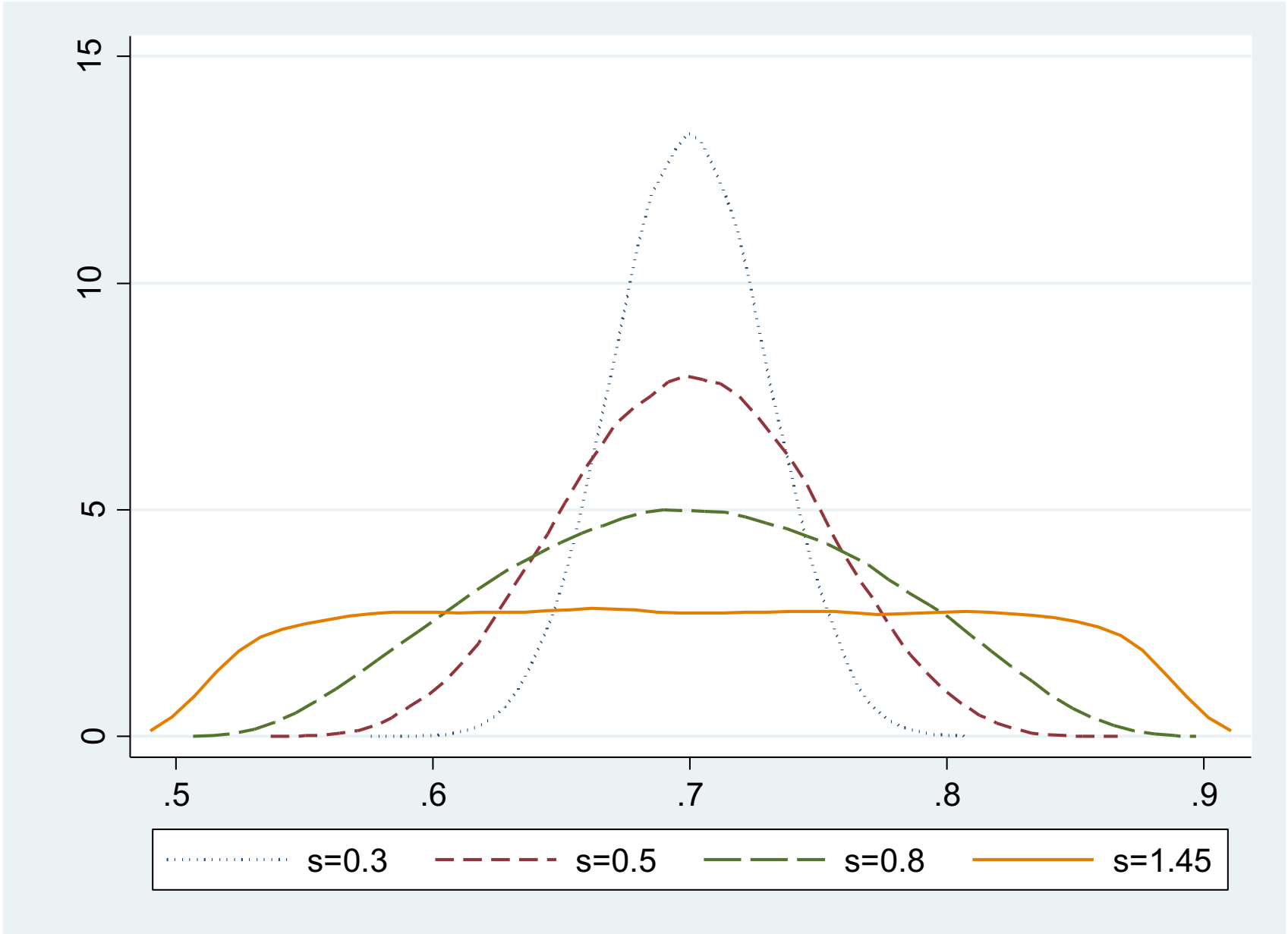


There is an interval of equally probable values between .6 and .8, within specified bounds (.5, .9).

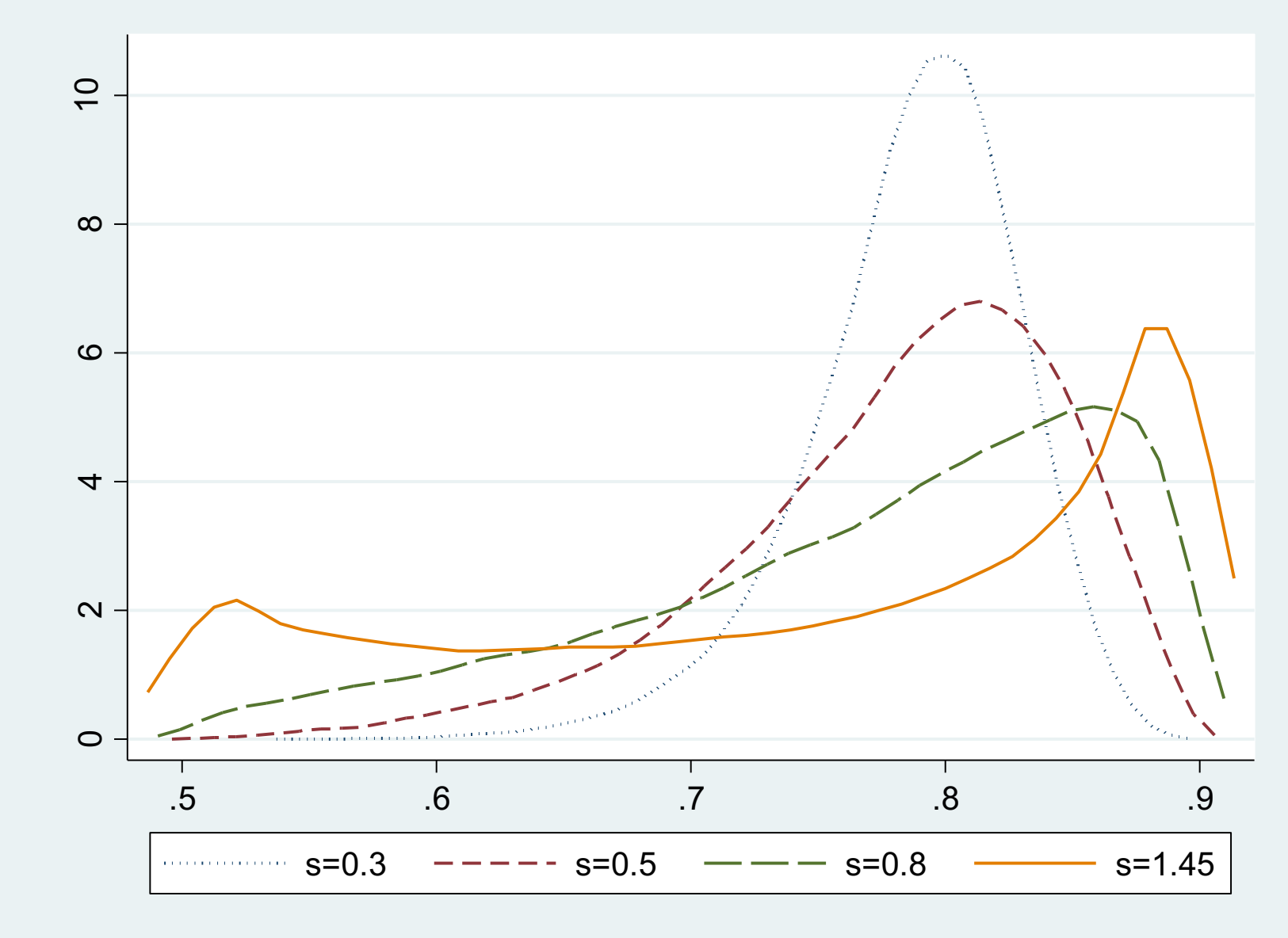
Log-normal distribution (m=0)



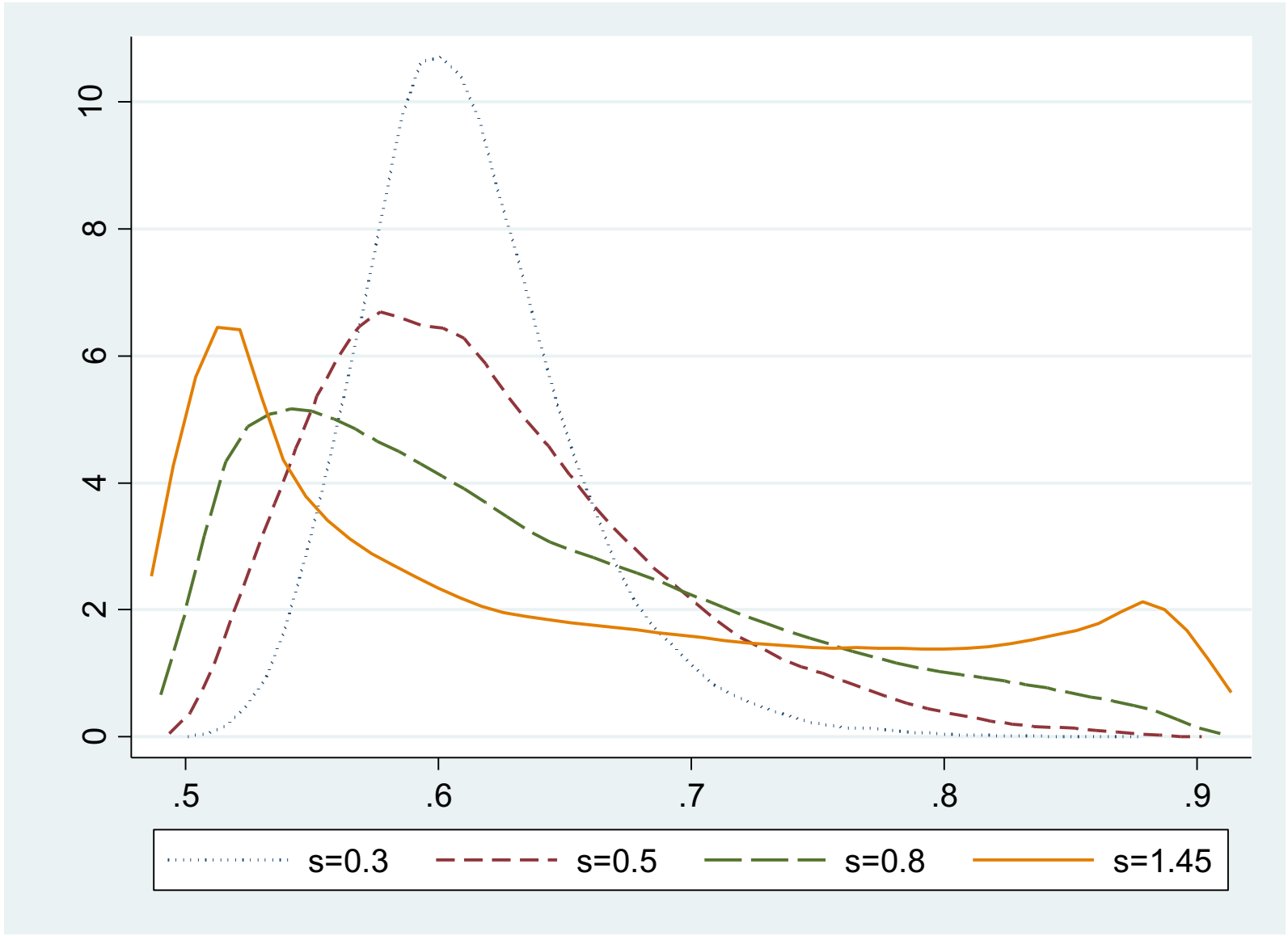
Logit-normal distribution (m=0, lb=.5, ub=.9)



Logit-logistic distribution (m=1, lb=.5, ub=.9)



Logit-logistic distribution (m=-1, lb=.5, ub=.9)



Monte Carlo-type simulations

Monte Carlo (random number-based) simulations involve two steps:

step 1) generate a dataset containing observations from the user specified probability density functions of the bias parameters

step 2) draw a random sample (one set of likely bias parameters) from this dataset to back-calculate the relative risk

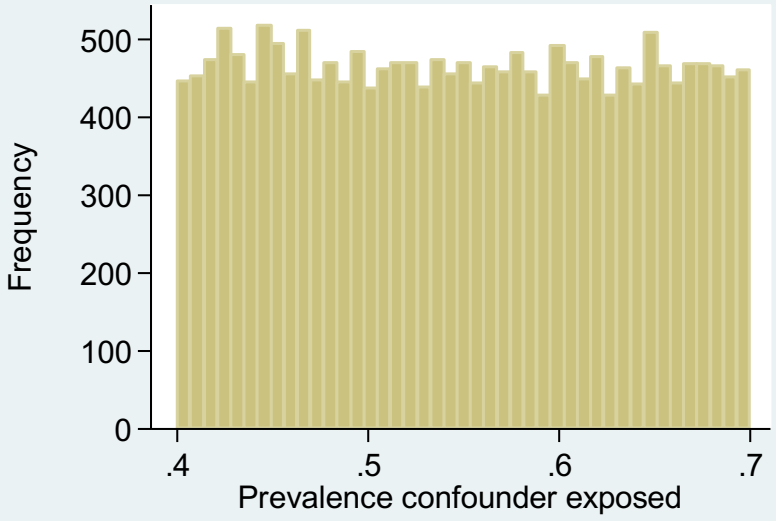
We repeat steps 1 and 2 a large number of times to obtain a distribution of bias-corrected estimates.

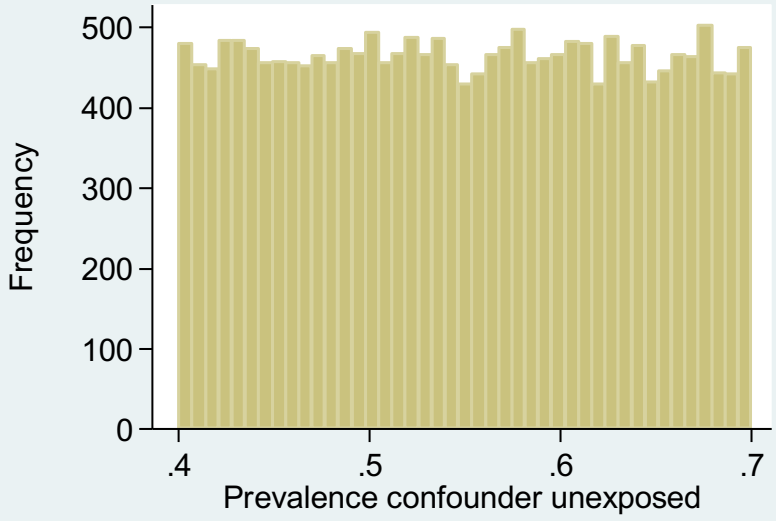
Example: Exposure to resins and lung cancer

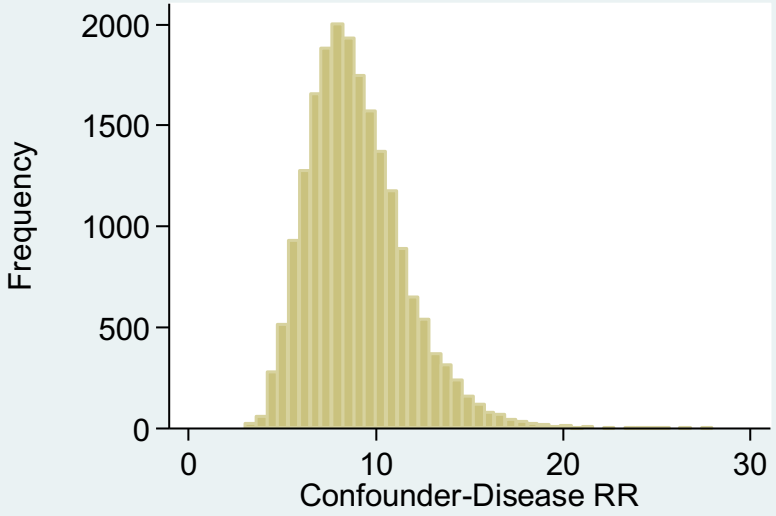
Two uniform distributions for the smoking prevalences among exposed and unexposed between 0.4 and 0.7.

The probability density function of the smoking-lung cancer mortality RR is assumed to be log-normal with 95% confidence limits of log(5) and log(15).

The limits imply that the mean of this distribution is [(log(15)-log(5)]/2=2.159 with standard deviation [log(15)-log(5)]/2*1.96=0.280.


Frequency
0
100
200
300
400
500
.4
.5
.6
.7
Prevalence confounder exposed


Frequency
0
100
200
300
400
500
.4
.5
.6
.7
Prevalence confounder unexposed


Frequency
0
500
1000
1500
2000
0
10
20
30
Confounder-Disease RR

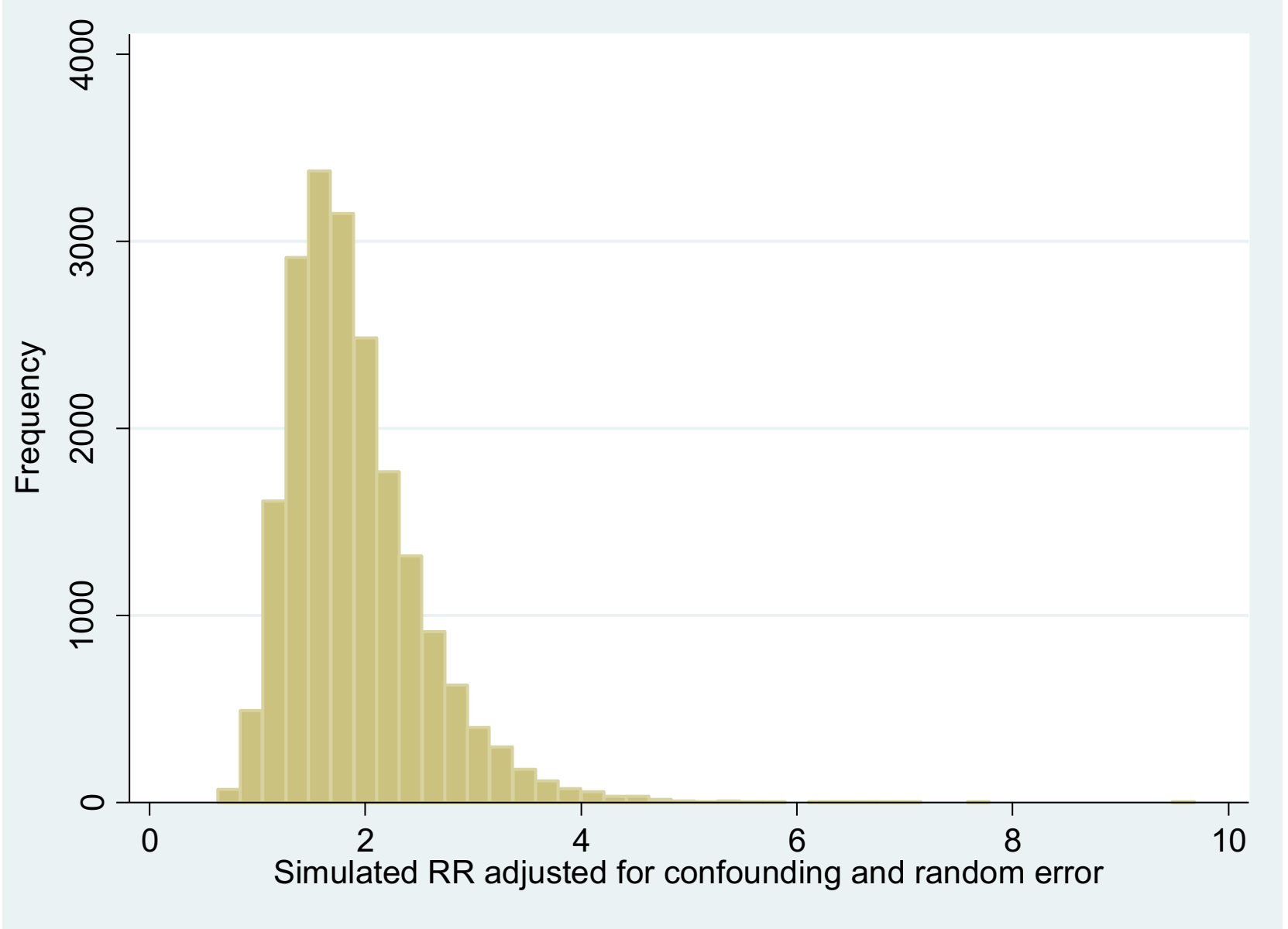
```
. episensi  45 94 257 945 , st(cc) reps(20000) nodots
dpexp(uni(.4 .7)) dpunexp(uni(.4 .7)) drrcd(log-n(2.159 .280))
grarrsys grarrtot grprior

Pr(c=1|e=1): Uniform(.4,.7)
Pr(c=1|e=0): Uniform(.4,.7)
RR_cd      : Log-Normal(2.16,0.28)

Probabilistic sensitivity analysis for unmeasured confounding
```

	Percentiles			Ratio
	2.5	50	97.5	97.5/2.5
Conventional	1.17	1.76	2.61	2.23
Systematic error	**1.24**	**1.76**	**2.49**	**2.00**
Systematic and random error	**1.04**	**1.76**	**3.01**	**2.90**

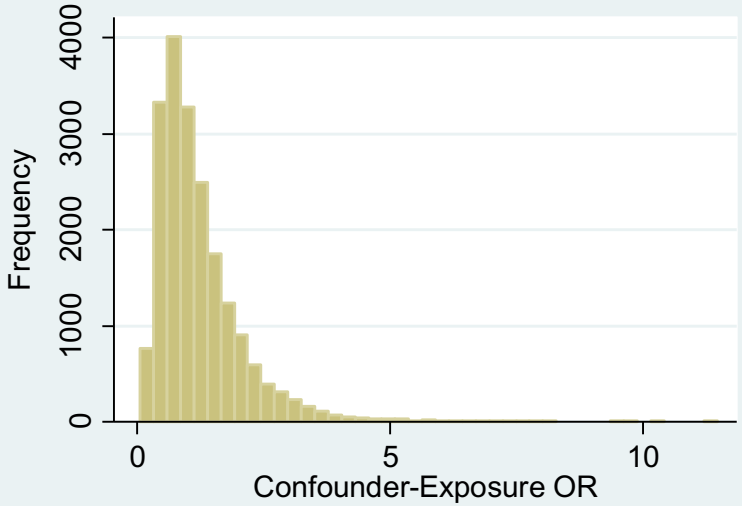
The median smoking-adjusted resins-lung cancer OR is 1.76 with 95% simulation limits of 1.04 and 3.01. As expected, the ratio of the smoking-adjusted simulation limits (2.9) is higher than the ratio of the conventional limits (2.2).

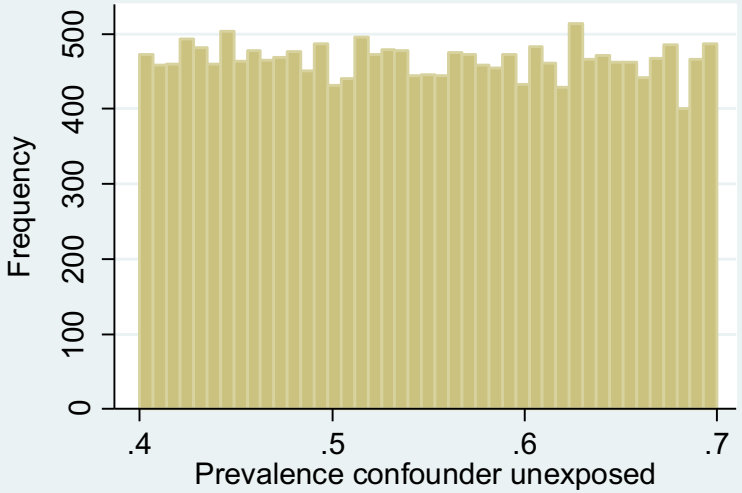

Frequency
0
1000
2000
3000
4000
0
2
4
6
8
10
Simulated RR adjusted for confounding and random error

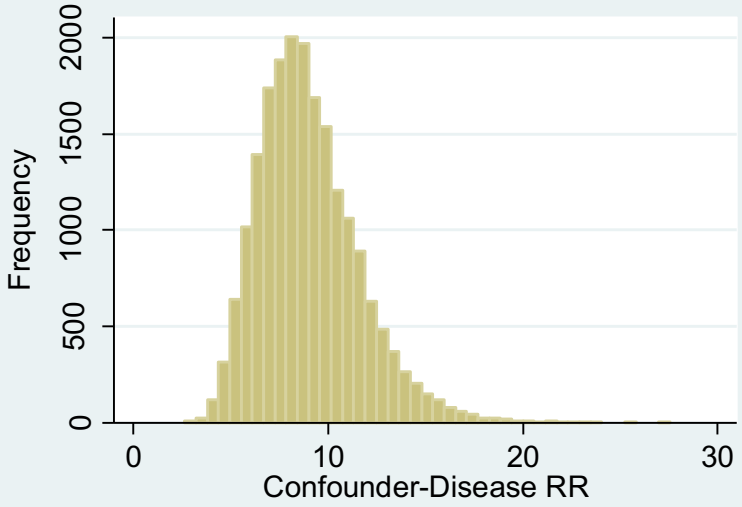
More reasonable priors

Given that there is no reason to expect great differences in the prevalence of smoking among resins exposed and unexposed, small differences are more likely than large ones.

One way to address non independent distributions of the confounder-exposure specific prevalences is to specify a probability density function for the confounder-exposure OR (option **dorce**) instead of the prevalence of the confounder among the exposed (option **dpexp**).


Frequency
0
1000
2000
3000
4000
0
5
10
Confounder-Exposure OR


Frequency
0
100
200
300
400
500
.4
.5
.6
.7
Prevalence confounder unexposed


Frequency
0
500
1000
1500
2000
0
10
20
30
Confounder-Disease RR

```
. episensi  45 94 257 945 , st(cc) reps(20000) nodots    ///
dpunexp(uni(.4 .7)) drrcd(log-n(2.159 .280))
dorce(log-normal(0 .639))

Pr(c=1|e=0): Uniform(.4,.7)
RR_cd      : Log-Normal(2.16,0.28)
OR_ce      : Log-Normal(0.00,0.64)

Probabilistic sensitivity analysis for unmeasured confounding

                                    Percentiles          Ratio
                              2.5       50        97.5      97.5/2.5
                              ------------------------------------
Conventional                  1.20      1.76      2.58      2.14
Systematic error              1.24      1.76      3.04      2.45
Systematic and random error   1.04      1.77      3.44      3.30
```

Table. Percentiles of Monte Carlo simulated distribution of the smoking-adjusted resins-lung cancer odds ratio.

	Percentiles		
Type of analysis	2.5^{th}	Median	97.5^{th}
Conventional	1.2	1.8	2.6
Systematic error			
Adjusted Odds Ratio	1.2	1.8	3.0
Systematic and random-sampling error			
Adjusted Odds Ratio	1.0	1.8	3.4

Set up of traditional sensitivity methods

- Binary outcome and binary exposure
- A single binary confounder
- Absence of interaction (homogeneity assumption) between the effect of the exposure and the confounder on the disease
- Only sensitivity analysis for the null hypothesis of no association

Epidemiology • Volume 27, Number 3, May 2016

ORIGINAL ARTICLE

OPEN

Sensitivity Analysis Without Assumptions

Peng Ding[a] *and Tyler J. VanderWeele*[b]

This article is moving beyond simplifying assumptions of previous methods

The authors propose a bounding factor to measure the strength of confounding

RR_{UD} = Maximum risk ratio for the disease comparing any 2 categories of the confounder

RR_{EU} = Maximum risk ratio for the confounder comparing any 2 categories of the confounder

The maximum relative amount the confounder (or set of confounders) could alter an observed risk ratio is obtained as follows

$$RR_{ED}^{adj} = \frac{RR_{ED}^{obs}}{\frac{RR_{UD}RR_{EU}}{RR_{UD}+RR_{EU}-1}}$$

One may also divide the confidence limits (CI) by the bias factor to obtain the maximum the unmeasured confounder could move the CI toward the null.

Illustration using resins exposure and lung cancer

P_{U1}=0.4 the fraction of smokers among those exposed to resins
P_{U0}=0.3 the fraction of smokers among those not exposed to resins
RR_{UD} = 5 the maximum association between smoking and lung cancer
$RR_{EU} = \frac{0.4}{0.3}$ = 1.3 the maximum association between smoking and exposure to resins

The most that unmeasured confounding could alter the effect estimate and its CI is obtained as follows

$$RR_{ED}^{adj} = \frac{1.76}{\frac{5 \times 1.3}{5 + 1.3 - 1}} = 1.41$$

$$95\%\ \text{CI for}\ RR_{ED}^{adj} = (0.96, 1.27)$$

Unmeasured confounding of this strength would suffice to explain away the effect estimate.

E-value

The E-value is the *minimum strength of association*, on the risk ratio scale, that an unmeasured confounder would need to have with both the exposure and the disease, conditional on measured covariates, to explain away the observed exposure-disease association (Mathematical details are in Ding & Vanderwheele, 2016).

$$\text{E-value} = RR_{ED}^{obs} + \sqrt{RR_{ED}^{obs}(RR_{ED}^{obs} - 1)}$$

If the $RR_{ED}^{obs} < 1$ one first takes the inverse of the observed risk ratio and then apply the formula above.

The higher the E-value, the stronger the confounder associations must be to explain away the effect.

Illustration using resins exposure and lung cancer

$$\text{E-value} = 1.76 + \sqrt{1.76(1.76-1)} = 2.92$$

The observed risk ratio of 1.76 could be explained away by an unmeasured confounder that was associated with both the exposure and the disease by a risk ratio of 2.9-fold each, above and beyond the measured confounders, but weaker confounding could not do so.

Sensitivity Analysis in Observational Research: Introducing the E-Value

Tyler J. VanderWeele, PhD, and Peng Ding, PhD

Sensitivity analysis is useful in assessing how robust an association is to potential unmeasured or uncontrolled confounding. This article introduces a new measure called the "E-value," which is related to the evidence for causality in observational studies that are potentially subject to confounding. The E-value is defined as the minimum strength of association, on the risk ratio scale, that an unmeasured confounder would need to have with both the treatment and the outcome to fully explain away a specific treatment–outcome association, conditional on the measured covariates. A large E-value implies that considerable unmeasured confounding would be needed to explain away an effect estimate. A small E-value implies little unmeasured confounding would be needed to explain away an effect estimate. The authors propose that in all observational studies intended to produce evidence for causality, the E-value be reported or some other sensitivity analysis be used. They suggest calculating the E-value for both the observed association estimate (after adjustments for measured confounders) and the limit of the confidence interval closest to the null. If this were to become standard practice, the ability of the scientific community to assess evidence from observational studies would improve considerably, and ultimately, science would be strengthened.

Ann Intern Med. 2017;167:268-274. doi:10.7326/M16-2607 **Annals.org**
For author affiliations, see end of text.
This article was published at Annals.org on 11 July 2017.

Remarks about the E-value

- It is a continuous measure. There are no cut-offs.
- Less extreme RR_{UD} will require a more extreme RR_{EU}
- The E-value depends on the magnitude of the RR_{ED}^{obs} (large effect size will give a large E-value).
- Small E-value means that robustness to unmeasured confounding may be weak.
- Small E-value does not mean that there is evidence of no effect.
- Weak evidence for an effect does not imply evidence that the effect is absent.

Strengths

- Sensitivity analysis helps the investigator to make explicit the location and shape of the distribution of the bias parameters.

- The distributions of the bias parameters reflect the knowledge and judgment of the investigator about the potential systematic errors that may affect the observed findings.

- Probabilistic sensitivity analysis provides a wider confidence interval that includes both systematic and random error, which conventional analysis fails to consider.

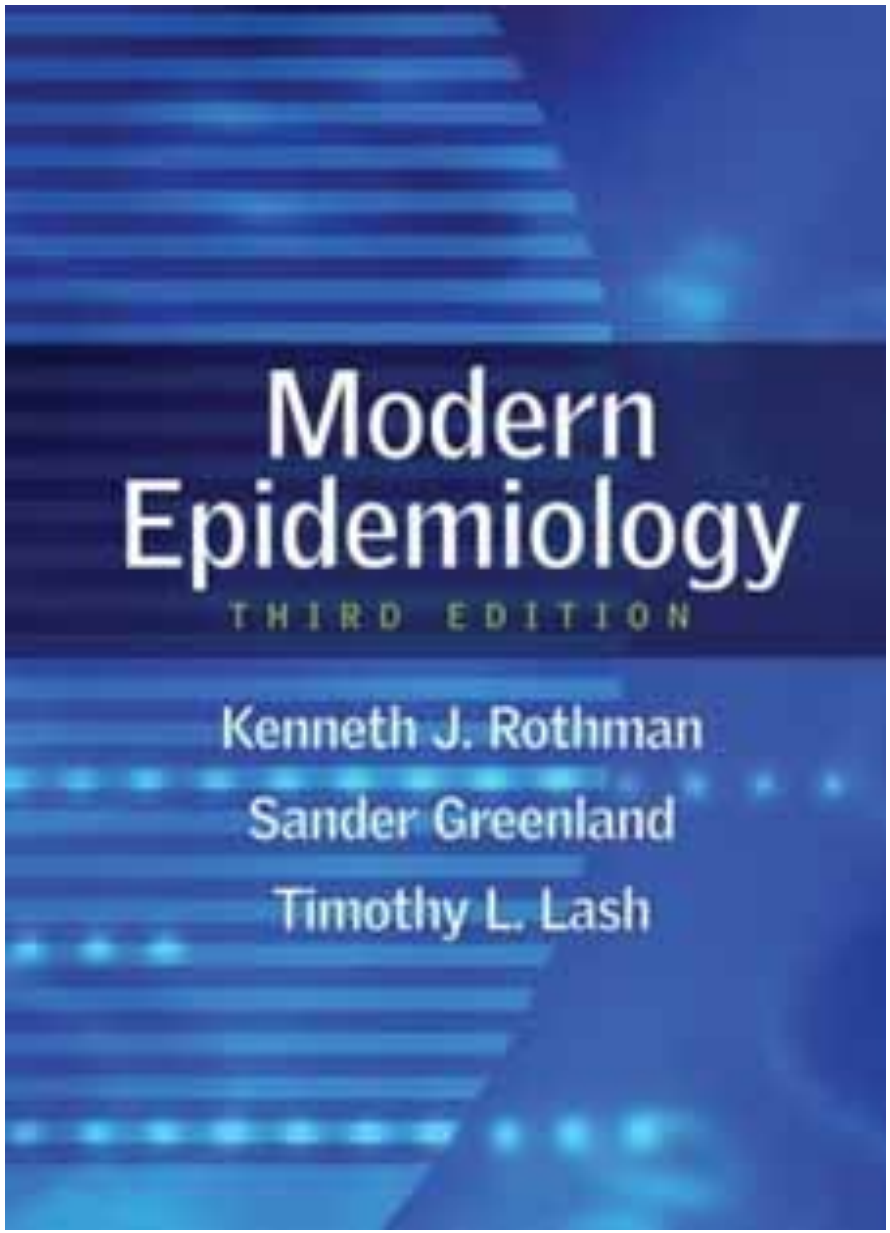
Limitations

- Concerns have been raised by some about the arbitrariness in the particular distributions assumed for the bias parameters, which can lead to different distributions of the adjusted exposure-disease RR.

- However, it should be emphasized that in order to make a shared and meaningful bias correction of the exposure-disease RR, the distributions of the bias parameters should be based on the best available evidence and by careful judgment.

- Informed sensitivity analysis is therefore limited by lack of data and/or scientific knowledge about the role of bias in a specific exposure-disease association.


Statistics for Biology and Health
Timothy L. Lash
Matthew P. Fox
Aliza K. Fink
Applying Quantitative Bias Analysis to Epidemiologic Data
Springer


Modern Epidemiology
THIRD EDITION
Kenneth J. Rothman
Sander Greenland
Timothy L. Lash

Main references

Greenland S. & Lash T. (2008) Chapter 19. *Bias analysis*. Modern Epidemiology.

Ding, P., & VanderWeele, T.J. (2016). Sensitivity analysis without assumptions. *Epidemiology*. 27(3), 368.

References

Cornfield J, Haenszel W, Hammond EC, Lilienfeld AM, Shimkin MB, Wynder EL. Smoking and lung cancer: recent evidence and a discussion of some questions. J Natl Cancer Inst 1959;22:173-203.

Friberg E, Mantzoros CS, Wolk A. Physical activity and risk of endometrial cancer: a population-based prospective cohort study. Cancer Epidemiol Biomarkers Prev 2006;15:2136-40.

Greenland S. Basic methods for sensitivity analysis of biases. Int J Epidemiol 1996;25:1107-16.

Greenland S. Multiple-bias modelling for analysis of observational data. Journal of the Royal Statistical Society Series a-Statistics in Society 2005;168:267-291.

Jurek AM, Maldonado G, Greenland S, Church TR. Exposure-measurement error is frequently ignored when interpreting epidemiologic study results. Eur J Epidemiol 2006.

Lin DY, Psaty BM, Kronmal RA. Assessing the sensitivity of regression results to unmeasured confounders in observational studies. Biometrics 1998;54:948-63.

Phillips CV. Quantifying and reporting uncertainty from systematic errors. Epidemiology 2003;14:459-66.

Rosenbaum PR, Rubin DB. Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome. Journal of the Royal Statistical Society Series B-Methodological 1983;45:212-218.

Schlesselman JJ. Assessing effects of confounding variables. Am J Epidemiol 1978;108:3-8.

Steenland K, Greenland S. Monte Carlo sensitivity analysis and Bayesian analysis of smoking as an unmeasured confounder in a study of silica and lung cancer. Am J Epidemiol 2004;160:384-92.