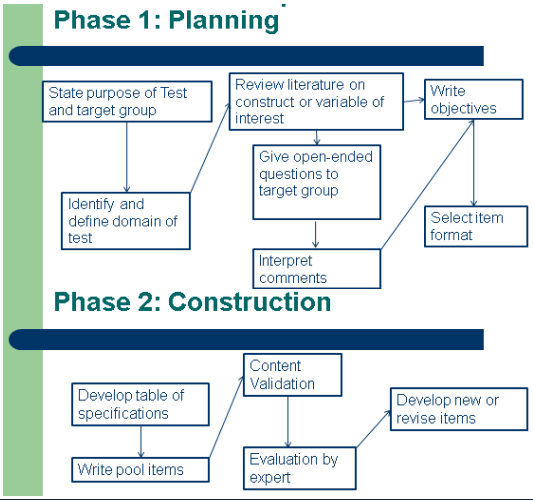
UCSF
University of California
San Francisco

Confounding

Lydia B. Zablotska, MD, PhD
Professor, Department of Epidemiology and Biostatistics

Highlights from Week #7

- Described principles of instrument development



UCSF

Highlights from Week #7

- Discussed and reviewed various methods of test development:
  - **Item analysis (based on CTT):**
    - Indices that describe the distribution of responses to a singular item (mean, SD)
    - Indices that describe the degree of relationship between response to the item and some other criterion (item difficulty)
    - Indices that are a function of both item variance and relationship to a criterion (index of discrimination)
  - **Item response theory (IRT): a model-based measurement in which trait level estimates (e.g., level of physical functioning or level of depression) depend on both persons' responses AND on the properties of the questions that were administered**
    - Compared IRT to CTT and GT
  - **ROC and AUC**
  - **Factor Analysis**

UCSF

Highlights from Week #7

- Described principles of instrument development
- Defined steps in designing new measures (norm- and criterion-referenced)
- Discussed instrument design, construction and interpretation in cross-cultural situations
- **Step 1-Defining the Concept in Emic Terms**
- **Step 2-Obtaining Cultural Consensus**
- **Step 3-Developing Measurement Items**
- **Step 4-Peer Validation**
- **Step 5-Pretesting Items with Prospective Study Participants**
- **Step 6-Seeking Additional Peer Input for Revising Items**
- **Step 7-Pilot Testing**
- **Step 8-Psychometric Evaluation**
- **Step 9-Evaluating Efficacy**

UCSF

Learning Objectives

- **Review and expand definition of confounding**
- **Revisit methods to control confounding in the design of the study:**
  - **Randomization, restriction, matching and analysis of matched data**
- **Examine statistical adjustment of confounding effects of known and measured confounders**
  - **Stratification, standardization**
  - **Multivariate adjustment models**
  - **Under- /over-adjustment and estimation of amount of confounding**
  - **Propensity scores**
  - **Structural measures and G-estimation**
- **Discuss new methods to control for residual and unmeasured confounders**
  - **Instrumental variable (IV) analysis**
  - **Probabilistic sensitivity analysis**
  - **Direct bias simulation**
  - **Negative control outcome calibration**

Epi 203 & 204 | New Material | Will be discussed in more detail in Epi 256, Biostats 215

UCSF

Learning Objectives

- **Review and expand definition of confounding**
- Revisit methods to control confounding in the design of the study:
  - Randomization, restriction , matching and analysis of matched data
- **Discuss statistical adjustment of confounding effects of known and measured confounders**
  - Stratification, standardization
  - **Multivariate adjustment models**
  - **Under- /over-adjustment and estimation of amount of confounding**
  - **Propensity scores**
  - **Structural measures and G-estimation**
- Examine new methods to control for residual and unmeasured confounders
  - **Instrumental variable (IV) analysis**
  - **Probabilistic sensitivity analysis**
  - **Direct bias simulation**
  - **Negative control outcome calibration**

UCSF

Confounding

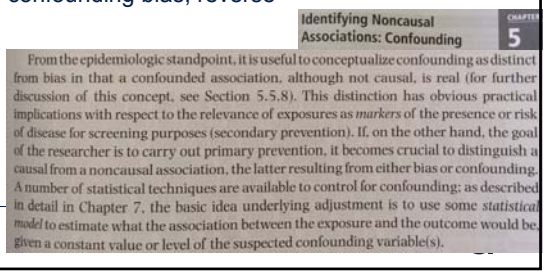
- **Importance of confounding in experimental research and observational studies**
- **Estimation of effects in observational studies**
  - Comparison of exposed and unexposed
  - Unexposed represent what the frequency of D would have been in the exposed cohort had exposure been absent (*counterfactual*)
  - Exposed cohort may differ from the unexposed cohort on many factors besides E, i.e. the use of unexposed as a referent for the exposed is *confounded*
  - *"Mixing of effects"* does not mean that exposure has to have an effect [Latin *confundere – to mix together*]

Is confounding a type of Bias?

- **Epi 203** – Yes, confounding is a type of bias (along with selection and measurement bias)

"I see confounding as a form of a bias, based on our definition that bias is non-random deviation from truth. What is "truth"? There can be truth in a descriptive study and truth in analytic study. We define both of these in our course. Most of the time, we are dealing with analytic studies and in that case, truth refers to a causal relationship. So, any systematic error that causes us to get something other than the true causal relationship is bias. Hence, if a causal relationship is the estimand, confounding is a form of bias. Defining bias in this way is particularly useful when describing the process of causal inference as one where we refute selection bias, measurement bias, confounding bias, reverse causality bias." - J. Martin

- Szklo and Nieto 2012


Identifying Noncausal Associations: Confounding CHAPTER 5

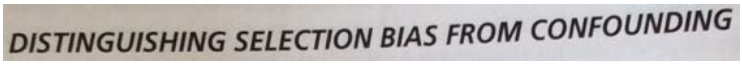
From the epidemiologic standpoint, it is useful to conceptualize confounding as distinct from bias in that a confounded association, although not causal, is real (for further discussion of this concept, see Section 5.5.8). This distinction has obvious practical implications with respect to the relevance of exposures as *markers* of the presence or risk of disease for screening purposes (secondary prevention). If, on the other hand, the goal of the researcher is to carry out primary prevention, it becomes crucial to distinguish a causal from a noncausal association, the latter resulting from either bias or confounding. A number of statistical techniques are available to control for confounding; as described in detail in Chapter 7, the basic idea underlying adjustment is to use some *statistical model* to estimate what the association between the exposure and the outcome would be given a constant value or level of the suspected confounding variable(s).

Is confounding a type of Bias?

- **Epi 207:** No, confounding is not a type of bias but selection bias could lead to some type of confounding

"If the association between E and D is confounded by C, then it is not real but spurious and due to the effect of C on E and D. This would have an effect on how I would interpret causality. In the case of selection or information bias, the association is biased by systematic error, but the association between E and D remains the same!" - L. Zablotska

- RGL2008 – pp. 136-137



UCSF

Natural vs. *artifactual* confounding

- Natural confounders: For a variety of reasons, including **biological factors, individual choices and social forces**, the exposed and unexposed tend to differ on other causes of the disease of interest (independent risk factors in addition to the main exposure of interest). E.g., race, gender
- Artifactual confounders: confounders associated with the exposure of interest only (or more strongly) within the context of the study, as a result of selective recruitment or other methodological artifacts. E.g., use of volunteers for cohort studies
- Note: Both still need to satisfy criteria for confounding!

 Susser et al. 2006. *Psychiatric Epidemiology*

UCSF

Recent developments on the types of confounding

Four notions of confounding

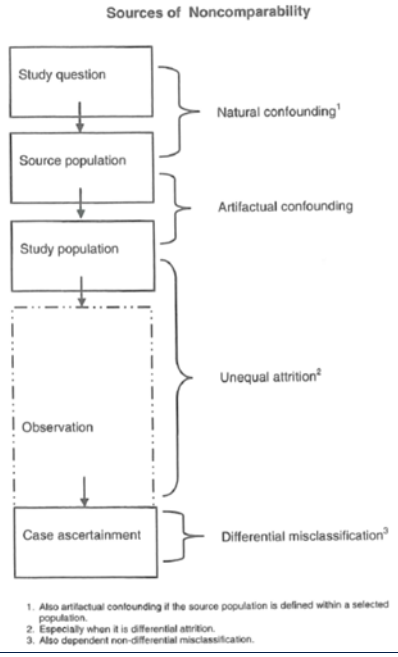
- **Confounding in distribution:** We say that there is no *confounding in distribution* of the effect of exposure on outcome if the group that actually had a particular exposure is representative of what would have occurred had the entire target population been exposed to the same level of exposure.
- **Confounding in measure:** We say that there is no *confounding in measure* of the effect of exposure on outcome if a particular measure of interest is equivalent to the corresponding causal measure in the target population.
- **Confounding in expectation:** We say that there is no *confounding in expectation* of the effect of exposure on outcome if the exposure assignment mechanism results in balance.
- **Realized confounding:** We say that there is no *realized confounding* of the effect of exposure on outcome if a particular exposure assignment results in balance, irrespective of the exposure assignment mechanism.

When we consider a binary exposure and a binary outcome, individuals can be classified into the following four different response types.[1]

- Type 1 or "doomed" persons: Exposure is irrelevant because outcome occurs with or without exposure
- Type 2 or "causal" persons: Outcome occurs if and only if they are exposed
- Type 3 or "preventive" persons: Outcome occurs if and only if they are unexposed
- Type 4 or "immune" persons: Exposure is irrelevant because outcome does not occur with or without exposure

11 Suzuki et al. J of Epidemiol 2017

UCSF



12 Susser et al. 2006. *Psychiatric Epidemiology*

UCSF

Criteria for confounding:

Well-known? ➢ use additional slides posted on the course CLE

1. Associated with disease
2. Associated with exposure
3. Not in the causal pathway from exposure to disease

Classic Examples:

http://epiville.ccnmtl.columbia.edu/interactive/confounding01.html

http://epiville.ccnmtl.columbia.edu/interactive/evaluation_of_confounding.html

Example: RCT of tolbutamide and insulin in the University Group Diabetes program (4 groups)

	Age, years				Total	
	<55		55+		<55	55+
	Tolbutamide	Placebo	Tolbutamide	Placebo	Tolbutamide	Placebo
CVD Deaths	8	5	22	16	30	21
Total at risk	106	120	98	85	204	205
Risk Ratio	1.81		1.19		1.44	

Study: Hypothesis: Lowering blood glucose to normal levels will reduce CVD deaths

Question: The crude value of the risk ratio is 1.44, which is between the values for the risk ratio in the two age strata. Could the crude risk ratio have been outside the range of the stratum-specific values, or must it always fall within the range of the stratum-specific values? Why or why not?

Example: RCT of tolbutamide and insulin in the University Group Diabetes program (4 groups)

	Age, years				Total	
	<55		55+		<55	55+
	Tolbutamide	Placebo	Tolbutamide	Placebo	Tolbutamide	Placebo
CVD Deaths	8	5	22	16	30	21
Total at risk	106	120	98	85	204	205
Risk Ratio	1.81		1.19		1.44	

Hypothesis: Lowering blood glucose to normal levels will reduce CVD deaths

Question: The crude value of the risk ratio is 1.44, which is between the values for the risk ratio in the two age strata. Could the crude risk ratio have been outside the range of the stratum-specific values, or must it always fall within the range of the stratum-specific values? Why or why not?

Answer: The crude risk ratio is not a weighted average of the stratum-specific risk ratios, and therefore it can be outside the range of the stratum-specific risk ratios. Even if the stratum-specific risk ratios are all identical in value, the crude can be equal to a different value.

Criteria for confounding: Some caveats

1. Associated with disease

- Predictive of disease occurrence ***apart*** from its association with exposure (*extraneous risk factor*)
- Should involve a mechanism ***other*** than the one under study (*a different causal pathway*)
- Associated with disease ***among unexposed*** *(referent group)*
- Does not have to actually cause the outcome, but must affect it in some way, predict who will develop disease

Criteria for confounding: Some caveats

2. Associated with exposure

- Associated with exposure ***among the source population*** for cases (*the control group of the case-control study*)
- Association between exposure and confounder among cases is not a valid estimate of the association in the source population
- Correlation ***not*** unidirectional association!

Criteria for confounding: Some caveats

3. Not in the causal pathway from exposure to disease

- Must not be affected by exposure ***or*** outcome (*mediator),* i.e., must not be in the causal pathway between exposure and outcome (see RGL2008, p. 194)
- Must not be affected by exposure ***and*** outcome (*collider*)
- Must not be a cause of the exposures of interest *(antecedent)*
- It is incorrect to simply state that a confounder cannot be a mediator or antecedent: a variable may act as a mediator or antecedent in one pathway, and as confounder of another pathway

Review of main types of relationships between variables

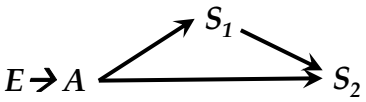
A. Independent risk factors
B. Confounders (alternatives to sole probability)
C. Antecedents
D. Mediators
E. Colliders
F. Causal partners

Notation:

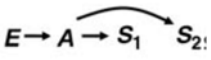
- **E-smoking**
- **A - conditional abnormality**
- **S_1 - 1st spontaneous abortion**
- **S_2 - 2nd spontaneous abortion**

Traditional depiction:

1. E is a risk factor for S_1
2. S_1 associated with S_2 among unexposed
3. S_1 is not an intermediate from E to S_2



DAG:



PROBLEMS?

Review of main types of relationships between variables

A. Independent risk factors

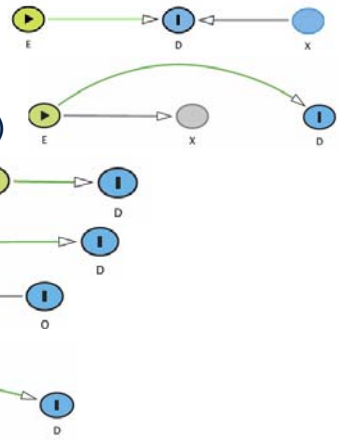
B. Confounders
(alternatives to sole probability)

C. Antecedents

D. Mediators

E. Colliders

F. Causal partners



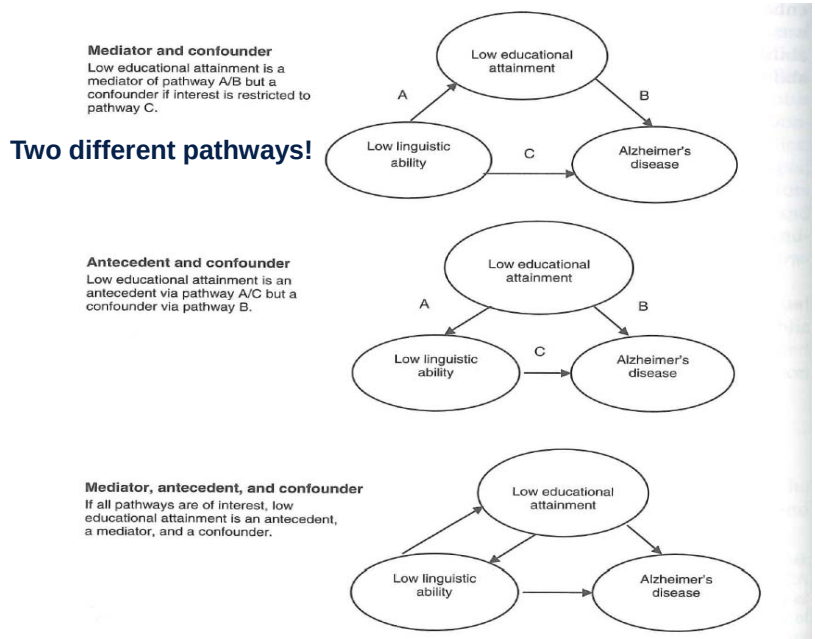
Notation:
E - exposure of interest;
D - disease of interest;
X - other variables (third variable)

UCSF

Example for confounding criterion #3: Linguistic ability in young adulthood and Alzheimer's disease

Snowdon et al. (1996)	*Study*: The Nun Study
See also: Snowdon et al. (1997)	*Cohort* ($N = 678$): Notre Dame nuns, at least 75 years old, who agreed to annual assessment of cognitive and physical function and to donate their brains after death. A subset of 93 participants with an archived writing sample from youth was used for a cohort study of the relation between linguistic ability in youth and cognitive ability in old age, and Alzheimer's disease. Cognitive ability was assessed at follow-up in late life; Alzheimer's disease, following death. Low linguistic ability in youth was a strong predictor of poor cognitive function and Alzheimer's disease in late life. *Potential confounding*: Educational attainment could account in part for an association of linguistic ability in youth with cognitive ability and Alzheimer's disease in late life.

UCSF



Two different pathways!

Figure 12.2 Variables with multiple roles: low educational attainment as confounder, mediator, and antecedent of the relationship between low linguistic ability and Alzheimer's disease in the Nun Study.

Confounding:
Some caveats

- Even if all three criteria are satisfied, the potential confounding factor may not produce any spurious excess or deficit of disease among exposed:
  - If there are multiple confounding variables whose effects are perfectly balanced
  - Always check for ***joint confounding*** *(this may become complicated if you have more than just a few confounders!)*
- The ***degree of confounding*** is of much greater concern than its mere presence or absence

Adjustment for confounding effects of known and measured confounders:

Magnitude of confounding

- "Simpson's Paradox" is rare
  - an extraneous factor, i.e. confounding factor, can change the direction of association between primary exposure of interest and outcome
  - In most studies a RR or OR of 2 or more is unlikely to be entirely explained by a single confounder

Examples:

1. Association between prenatal exposure to the Dutch Hunger Winter and adult schizophrenia could be confounded by social class (Stein et al. 1975)
2. Association between diabetes and death in the Poole Diabetes study could be confounded by age (Gatling et al. 1985)

25 Simpson *J R Statist Soc B 1951*

UCSF

EX1 – those conceived in the early months of the famine

EX2 – those conceived at the height of the famine

Example 1:

Dutch Hunger Winter and adult schizophrenia

Table 2. Schizophrenia* in Birth Cohorts Exposed† and Unexposed† to Early Prenatal Nutritional Deficiency During the Dutch Hunger Winter of 1944 Through 1945: Cumulative Incidence Data

	All			Men‡			Women‡		
	Exposed			Exposed			Exposed		
Incidence Data	EX2 (n=4190)	EX1 (n=5466)	Unexposed (n=136 691)	EX2 (n=2184)	EX1 (n=2800)	Unexposed (n=69 943)	EX2 (n=2006)	EX1 (n=2666)	Unexposed (n=66 748)
Cumulative incidence per 1000 persons‡ (n cases)	3.8 (16)	2.0 (11)	1.9 (257)	4.1 (9)	2.5 (7)	2.2 (151)	3.5 (7)	1.5 (4)	1.6 (106)
Relative risk at age 24-48 y (95% confidence interval)	2.0 (1.2-3.4)	1.1 (0.6-2.0)	. . .	1.9 (1.0-3.7)	1.2 (0.5-2.5)	. . .	2.2 (1.0-4.7)	0.9 (0.4-2.6)	. . .
P	<.01	.82	. . .	.05	.70	. . .	.04	.91	. . .

*Schizophrenia narrowly defined, ie, ICD-9 (International Classification of Diseases, Ninth Revision) 295.1,2,3,6.
†EX2 indicates exposed birth cohort of October 15 through December 31, 1945; EX1, exposed birth cohort of August 1 through October 14, 1945; and unexposed, all other births of 1944 through 1946 in the famine cities.
‡Denominators are live births minus deaths up to age 18 years. (Note that in men, denominators differ between Tables 1 and 2, being only live births in Table 1.)

- **Target population:** persons born in the cities of western Netherlands during 1944 through 1946
- **Exposure:** defined by conception during various periods of German occupation of Holland
- **Outcome:** the frequency of hospitalized patients with schizophrenia at age 24 to 48 years in the exposed and unexposed birth cohorts was ascertained from a national psychiatric registry.
- **Design:** cohort study comparing the risk for schizophrenia in those exposed and unexposed during early gestation to the Dutch Hunger Winter of 1944/1945.
- **Results**: those conceived at the height of the famine (EX2) had a two times higher risk of developing schizophrenia compared to unexposed.

26 Susser et al. 1996

UCSF

Example 1:

Dutch Hunger Winter and adult schizophrenia

Could it be due to confounding by SES?

- **SES as a confounder (C):**
  - C-E association: higher social class women were somewhat better nourished; over-represented among exposed (**ratio of upper SES to lower SES in exposed 3:2, in unexposed 1:1**)
  - C-O association: most studies show that lower SES is associated with higher risk of schizophrenia, but one study in Holland showed that higher parental social class was a significant risk factor for schizophrenia in offspring (university town based study of 34 psychiatric hospitalizations)
- **Assuming higher SES is associated with schizophrenia:**
  - 5,000 E+ and 100,000 E-
  - Assume that E-O not associated, i.e. risk in E+ and E- is the same
  - C-O associated, i.e. risk in C+ (high SES) = 0.5%, C- (lower SES) = 0.25%, RRcrude=2.0
  - Risk (O/ E+) = {(0.5%*3,000) + (0.25%*2,000)} /5,000 = 20 / 5,000 = 4 per 1,000
  - Risk (O/ E-)={(0.5%*50,000) + (0.25%*50,000)} /100,000 = 375 / 100,000 = 3.75 per 1,000
  - RRadjusted for SES = 4 / 3.75 = 1.07 vs. 1.00 expected
  - **Conclusion**: Even when assuming that E has no effect and C has a strong effect on O, the artifactual RR adjusted for C is barely detectable and cannot account for observed RR of 2.0.

van Os 1997
Susser et al. 1997

UCSF

Example 2:

Association between diabetes and death in the Poole Diabetes study

Survival by type of diabetes. Values are numbers (percentages)

Survival status	Type of diabetes: Non-insulin dependent	Type of diabetes: Insulin dependent	Total
All patients			
Censored	326 (60)	253 (71)	579
Dead	218 (40)	105 (29)	323
Total	544 (100)	358 (100)	902
Patients aged ≤40			
Censored	15 (100)	129 (99)	144
Dead	0	1 (1)	1
Total	15 (100)	130 (100)	145
Patients aged >40			
Censored	311 (59)	124 (54)	435
Dead	218 (41)	104 (46)	322
Total	529 (100)	228 (100)	757

RR from the Cox proportional hazards survival model with just type of diabetes indicates that insulin dependent diabetes gives a significantly better prognosis for survival than non-insulin dependent diabetes (0.69 (95%CI: 0.54, 0.87)). However, by entering age concurrently into the model with type of diabetes, RR=1.15 (0.91, 1.46).

Problem: dichotomizing continuous variable

UCSF

Adjustment for confounding effects of known and measured confounders:

Magnitude of confounding

- Confounding effects could be cumulative. Thus, several confounders with modest individual impact taken together, may account for an appreciable risk ratio distortion
  - Always check for joint confounding
- Magnitude of confounding is a result of the strength of the associations between the confounder AND BOTH exposure and disease

→**Question:** Do tests of statistical significance to evaluate presence of confounding work?

→ **Answer**: A significance test is only applied to the association between a confounder and the exposure or the disease. DAGs could help to untangle the complex relationship between E, D, and C

UCSF

Learning Objectives

- Review and expand definition of confounding
- **Revisit methods to control confounding in the design of the study:**
  - Randomization, restriction , matching and analysis of matched data
- **Discuss statistical adjustment of confounding effects of known and measured confounders**
  - Stratification, standardization
  - **Multivariate adjustment models**
  - **Under- /over-adjustment and estimation of amount of confounding**
  - **Propensity scores**
  - **Structural measures and G-estimation**
- Examine new methods to control for residual and unmeasured confounders
  - **Instrumental variable (IV) analysis**
  - **Probabilistic sensitivity analysis**
  - **Direct bias simulation**
  - **Negative control outcome calibration**

UCSF

Matching and analysis of matched data

- **Quick review:**
  - Methods to control confounding in the design stage:

Matching and analysis of matched data

- **Quick review:**
  - Methods to control confounding in the design stage:
    - Randomization
    - Restriction
    - Matching

Based on this, what is the purpose and effect of matching:

A. Control confounding

B. Improve precision of confounder-adjusted summary estimate for a given size

Matching and analysis of matched data

- **Quick review:**
  - Methods to control confounding in the design stage:
    - Randomization
    - Restriction
    - Matching

Based on this, what is the purpose and effect of matching:

B. Improve precision of confounder-adjusted summary estimate for a given size

Matching and analysis of matched data
Some caveats

- **In case-controls studies, matching introduces selection bias (towards the null) whether or not there is confounding by the matching factors in the source population:**
  - Matching selects controls who are more like cases with respect to exposure than would be controls selected at random from the source population
  - If controls are selected to match the cases on a factor that is correlated with the exposure, then the crude exposure frequency in controls would be distorted in the direction of similarity to that of the cases
- In case-control studies, it is *no longer possible to estimate the confounding effect of the matching factor* because matching distorts the relation of the factor to the disease. Is it still *possible to study the factor as a modifier of odds ratio* (by seeing how it varies across strata)

Example: Effect modification by race in the CAP dataset

Overmatching

1. **Matching that harms *statistical efficiency*:**
   - Matching on a non-confounder associated with exposure but not disease can cause a factor to behave like a confounder and control of the factor in the analysis will be necessary
2. **Matching that harms *validity*:**
   - Matching on the intermediate variable will shift the exposure prevalence among non-cases toward that of cases (crude biased toward the null)
3. **Matching that harms *cost efficiency*:**
   - Friend controls may result in similar exposures but not disease (see 1 above)

Learning Objectives

- Review and expand definition of confounding
- Revisit methods to control confounding in the design of the study:
  - Randomization, restriction , matching and analysis of matched data
- **Discuss statistical adjustment of confounding effects of known and measured confounders**
  - Stratification, standardization
  - **Multivariate adjustment models**
  - **Under- /over-adjustment and estimation of amount of confounding**
  - **Propensity scores**
  - **Structural measures and G-estimation**
- Examine new methods to control for residual and unmeasured confounders
  - **Instrumental variable (IV) analysis**
  - **Probabilistic sensitivity analysis**
  - **Direct bias simulation**
  - **Negative control outcome calibration**

Adjustment for confounding effects of known and measured confounders

- **Quick review**
  - Methods to adjust for confounding in the analysis stage

UCSF

Adjustment for confounding effects of known and measured confounders

- **Quick review**
  - **Methods to adjust for confounding in the analysis stage:**
    - Stratification
    - Standardization (SMRs and SIRs)
    - **Multivariate analysis**
      - **Selection of important confounders**
      - **Adjustment using scoring methods**
      - **G-estimation method to adjust for time-varying confounders**

UCSF

Adjustment for confounding effects of known and measured confounders:
Methods

- **Stratification shows distributions of key variables and patterns in the data that are less transparent when using other methods; it should be done *prior* to attempting regression methods**
  - Test of homogeneity of stratum-specific effect estimates
    - Comparison of stratum-specific estimates against a summary estimate obtained by using
      - Woolf method or weighted least squares (ample data)
      - Fisher exact method (sparse data)
      - Maximum likelihood (ML) method (at least 10 cases per stratum)
      - Mantel-Haenszel method (valid for sparse data but can have much higher variance than ML)
  - Comparison of observed cell counts against cell counts expected under the homogeneity hypothesis
  - Both methods have very low power
- **Multivariate analysis**
  - Confounding variables for the final model could be selected based on the
    - *change-in-estimate criterion (preferable)*
    - *statistical tests (collapsibility testing)*
    - *subject matter grounds ("known confounders")*

Adjustment for confounding effects of known and measured confounders:
"Comparability" vs. "Collapsibility"

- **Comparability** (Sander Greenland, James Robins, Hal Morgenstern, and Charles Poole)
  - is defined in relation to the counterfactual model for causal inference
  - confounding results from non-comparability, i.e., a difference between the distribution of outcomes for the unexposed group to what would have been observed in the exposed group if it had not been exposed. Since the latter value is hypothetical and unobservable, the comparability definition cannot be directly applied, though it has some theoretical advantages as well as practical implications.
- **Collapsibility** (D.A. Grayson and others)
  - confounding is present when the crude measure of association differs from the value of that measure when extraneous variables are controlled by stratification, adjustment, or mathematical modeling
  - readily applied in practice and is widely used
  - makes confounding specific to the measure of association used and the particular variables that are being controlled

Adjustment for confounding effects of known and measured confounders:

"Comparability" vs. "Collapsibility"

- ***Comparability*** explains the origins of confounding using the counterfactual model for causal inference
- ***Collapsibility*** has nothing to do with confounding
- **Non-collapsibility of the odds ratio**, which arises from the failure of group odds to equal a weighted average of subgroup odds, is a mathematical oddity, not a reflection of bias
- The two definitions ***generally agree*** on the presence or absence of confounding when the measure of effect is a ratio or difference of incidences (proportions, i.e. risks, or rates), but not the odds ratio (unless the situation is one where odds ratio closely estimates a risk or rate ratio, e.g., a rare outcome).

Question: But I always use comparison of crude and adjusted risk estimates in logistic regression to evaluate confounding!?

Answer: Be careful and see above!

Hernan et al. *Int J Epidemiol* 2011
Greenland et al. *Stat Science* 1999

UCSF

Adjustment for confounding effects of known and measured confounders:

Methods, *recommendations*

1. Forward selection step-wise regression method assesses individual effects of confounders, but ignores possible interaction effects between them (*joint confounding);* it is indicated when data are sparse but in all other situations a backwards deletion strategy should be used
   - Read more in *A Pocket Guide to Epidemiology, Ch. 11 "Confounding can be confounding – several risk factors."*
2. One way to ensure adequate power is to raise the alpha level for rejecting the null hypothesis of no confounding to 0.20 or higher.
3. To ensure proper CIs: increase alpha to 0.20 or more in confounder-selection tests (collapsibility testing) or make sure that the confidence intervals do not change to an important degree if a confounder is to be deleted from control
   - Use the confidence limits rather than the point estimate to monitor the change produced by adding or deleting control of a confounder

RGL Ch 15

UCSF

Adjustment for confounding effects of known and measured confounders:

Methods, *recommendations*

1. Adjustment for variables that violate any of the criteria for confounding could distort effect estimates (*over-adjustment*)
   - in stratified analysis it can increase the variance and reduce the efficiency of the estimation process
2. Computed 95% CIs assume that no selection of confounding variables was done. Because they do not reflect the uncertainty about the confounder effects, they may be too narrow

Sir Richard Doll

(American College of Epidemiology Newsletter for Fall 1992)

"*There have been many important steps along the way: larger scale studies, more powerful statistical techniques, and the development of computers that allow these techniques to be applied. I fear, however, that the ease of applying statistical packages is sometimes blinding people to what is really going on. You don't have a real close understanding of what the relationships are when you put environmental and all of the other components of the history together in a logistic regression that allows for fifteen different things. I am a great believer in simple stratification. You know what you are doing, and you really want to look at the intermediate steps and not have all of the data in the computer*".

Scoring methods

- **Confounder scores are treated as a single confounder in the model:**
  - A categorical compound variable with distinct values for every possible value of measured confounders
  - Problem: the strata of a compound variable rapidly become too sparse for analysis
- **Outcome scores are constructed to predict the outcome (obtained by regression outcome and confounder alone, without exposure of interest)**
- **Exposure scores also known as propensity scores (Rosenbaum and Rubin 1983)**
  - Criteria for selection of variables in the propensity score should be the same as those used for outcome regression

Propensity scores

- Propensity score e(x) is defined as conditional exposure probability given a set of observed covariates x
- In a cohort study, matching or stratifying treated and controlled subjects on a single variable, the propensity score, tends to balance all of the observed covariates; however, unlike random assignment of treatments, the propensity score may not also balance unobserved covariates.

Example: Study of the association between smoking and low birth weight

- **Exposure of interest: smoking**
- **Other factors:**
  - age of mother, weight at last menstrual period, history of premature labor, number of physicians visits during first trimester, hypertension, uterine irritability and race
- **Results from logistic regression:**
  - Positively associated with smoking: age of mother, history of premature labor, race (black or white vs. other)
  - Negatively associated with smoking: weight at last menstrual period, number of physicians visits during first trimester
  - Final model should include all of these to obtain an unbiased estimate of the effect: OR=2.45 (95% CI: 1.15, 5.21)

47 Hosmer and Lemeshow, 1989

UCSF

Example: Study of the association between smoking and low birth weight

- **Propensity score based on selected confounders:**
  - Continuous measure calculated for each study participants, categorized into 5 classes (quintiles)
  - Direct stratification on all confounders will result in at least 32 sub-classes if all confounders are dichotomized: OR=1.96 (95%CI: 0.75, 5.20)
  - Final model included categorical propensity score, other factors were not associated with smoking, OR=1.61 (95%CI: 0.70, 3.71)
  - Conclusions:
    - Some residual bias that has not been captured by propensity score (compare 1.61 and 2.45)
    - Interpretations of the results of logistic regression and regression with propensity scores are different:
      - Logistic regression: the odds for a smoker if smoking is ceased while other factors remain unchanged
      - Regression with propensity score: the odds due to smoking in a population of smokers when compared to a population of non-smokers with the same distribution of covariates
    - Logistic regression models individual effect while **propensity score analysis estimates population average**

48 Hosmer and Lemeshow, 1989

UCSF

Propensity scores

- Could be used for stratification, matching, or as a covariate in the multivariate regression
- Stratification or matching on a fitted score requires categorization of propensity score which may introduce residual confounding

Propensity scores

- Propensity scores are estimated in regression models and range from 0 to 1 and reflect the estimated probability, based on the subject's characteristics, that the subject will receive the treatment of interest
- Any two subjects with the same scores can have different covariate values, but the distributions of covariates for all treated subjects should be similar to those for untreated subjects with the same scores

Structural models and G-estimation

- **Confounders in the model could be:**
  - Endogenous (can be affected by other variables in the model)
  - Exogenous (cannot be affected by other variables)
  - Example:

    Y=y0 + b1x1 +b2x2 +b3x3 +b4x4

    Rate of asthma attacks= baseline (genetic)

    + endogenous factors (physical activity and medications)

    + exogenous factors (air pollution and weather)
- **Multiple causal relations could be modeled with multiple equations (structural-equations modeling)**

G-estimation in cohort studies

- Standard methods for analysis of cohort studies may give biased estimates of exposure effects in the presence of time-varying confounding
- Most easily fitted using a two-step procedure called G-estimation
- Use observational data to estimate parameters that would be obtained in a perfectly randomized controlled trial
- Outcomes that would have occurred under this alternate exposure scenario can be considered missing data
- Use the existence of counterfactuals to enable unbiased estimation of marginal causal effects
- Somewhat similar to inverse-probability-of-treatment weighting (IPTW), another method for causal inference

G-estimation in cohort studies, *assumptions*

- A covariate is a time-varying confounder for the effect of E on O if
  - 1) past covariate values predict current exposure
  - 2) past exposure predicts current covariate value
  - 3) current covariate value predicts outcome
- For each subject, Ui is defined as the time to failure if the subject was unexposed throughout follow-up
- Assume **NO** unmeasured confounders
- Conditional on measured history (past and present confounders and past exposure), present exposure is independent of Ui
- Exposure j* accelerates failure time by exp(–ψ). If ψ was known, the counterfactual survival time Uij*,ψ could be derived from the observed data
- G-estimation estimates the effect of exposure on survival by examining a range of values for ψ and choosing the value ψ0 for which current exposure is independent of *Ui*.
  - i.e., comparison of counterfactual failure times for two scenarios: exposure history=always exposed vs. exposure history=never exposed
- **Example**: conditional on past weight, smoking status, blood pressure, and cholesterol, a person's decision to quit smoking is independent of what his or her survival time would have been if he or she had never smoked

53 Witteman et al. 1998

UCSF

TABLE 4. The Weibull survival analysis and G-estimated relations between time-varying cardiovascular risk factors and survival for Atherosclerosis Risk in Communities participants with data from at least the first two visits (1987–1989 and 1990–1993)

Variable	Reference group	HR*	95% CI*	G-estimated HR	95% CI	G-estimated survival ratio†	95% CI
SBP* ≥140 mmHg‡	SBP <140 mmHg	1.72	1.23, 2.40	1.79	1.38, 2.24	0.62	0.51, 0.77
DBP* ≥90 mmHg‡	DBP <90 mmHg	1.91	1.02, 3.56	1.98	0.97, 28.56	0.57	0.06, 1.03
Diabetes present	No diabetes present	1.26	0.98, 1.62	1.62	1.06, 1.98	0.68	0.58, 0.95
Current smoker	Nonsmoker	1.15	0.88, 1.51	1.24	0.92, 1.61	0.84	0.69, 1.07
BMI* (kg/m²)	BMI 20–30 kg/m²						
≤20		3.09	2.03, 4.71	2.07	1.39, 3.64	0.56	0.36, 0.77
≥30		0.83	0.64, 1.07	0.71	0.51, 1.12	1.31	0.91, 1.69
HDL* cholesterol (mmol/liter)	HDL cholesterol 0.91–1.55 mmol/liter						
≤0.91		1.44	1.13, 1.82	1.46	1.11, 1.88	0.74	0.61, 0.92
≥1.55		1.08	0.84, 1.38	1.06	0.72, 1.45	0.96	0.74, 1.29
LDL* cholesterol (mmol/liter)	LDL cholesterol 3.36–4.14 mmol/liter						
≤3.36		1.12	0.92, 1.36	0.98	0.77, 1.22	1.01	0.86, 1.23
≥4.14		1.37	1.09, 1.71	1.26	0.92, 1.70	0.83	0.66, 1.06
Use of antihypertensives	No use of antihypertensives	0.87	0.69, 1.09	0.80	0.57, 1.03	1.20	0.98, 1.57

* HR, hazard ratio; CI, confidence interval; SBP, systolic blood pressure; DBP, diastolic blood pressure; BMI, body mass index; HDL, high density lipoprotein; LDL, low density lipoprotein.
† Ratio of time to event for a person with the exposure to that for a person without the exposure. A value of less than 1 indicates an adverse effect of exposure.
‡ Estimates are from a model including all of the same variables, but excluding participants on antihypertensive medication.

This study examined association between cardiovascular risk factors and all-cause mortality and risk of coronary heart disease (CHD), accounting for confounding between exposures over time which were ascertained through repeated visits. Results were compared with those from standard survival analyses (e.g., Weibull regression) with time-updated covariates.
G-estimate adjusted associations differed from those estimated using standard survival analysis. The G-estimated effect of low density lipoprotein shows a less U-shaped relation with mortality than did the Weibull model.

54 Tilling et al. 2002

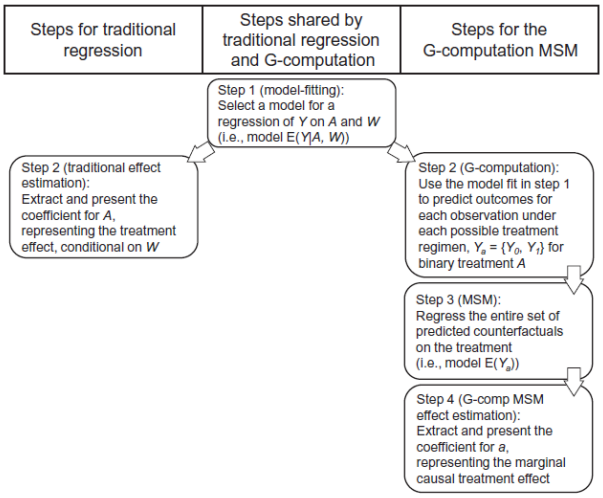


Figure 1. Contrasts between the implementation of traditional regression and the G-computation marginal structural model (G-comp MSM). *A* is a binary treatment, $W = \{W_1, W_2\}$ a vector of confounders, and *Y* a continuous outcome.

55 Snowden et al. 2011

UCSF

Adjustment for confounding effects of known and measured confounders:

Caveats

- Confounder category boundaries should be chosen in such a way that effect estimates are stable within categories
  - particularly important for strong confounders with uneven distributions (percentile categories vs. frequency categories)
- Some variables could be both confounders and effect modifiers of the exposure-disease association under study

56 RGL Ch 15

UCSF

Hypothetical Midlife Interventions in Women and Risk of Type 2 Diabetes

Goodarz Danaei,[a,b] An Pan,[c,d] Frank B. Hu,[b,c,e] and Miguel A. Hernán[b,f,g]

(*Epidemiology* 2013;24: 122–128)

Background: Randomized trials have examined short-term effects of lifestyle interventions for diabetes prevention only among high-risk individuals. Prospective studies have examined the associations between lifestyle factors and diabetes in healthy populations but have not characterized the intervention. We estimated the long-term effects of hypothetical lifestyle interventions on diabetes in a prospective study of healthy women, using the parametric g-formula.

Methods: Using data from the Nurses' Health Study, we followed 76,402 women from 1984 to 2008. We estimated the risk of type 2 diabetes under eight hypothetical interventions: quitting smoking, losing weight by 5% every 2 years if overweight/obese, exercising at least 30 minutes a day, eating less than three servings a week of red meat, eating at least two servings a day of whole grain, drinking two or more cups of coffee a day, drinking five or more grams of alcohol a day, and drinking less than one serving of soda a week.

Results: The 24-year risk of diabetes was 9.6% under no intervention and 4.3% when all interventions were imposed (55% lower risk [95% confidence interval = 47 to 63%]). The most effective interventions were weight loss (24% lower risk), physical activity (19%), and moderate alcohol use (19%). Overweight/obese women would benefit the most, with 10.8 percentage point reduction in 24-year risk of diabetes. The validity of these estimates relies on the absence of unmeasured confounding, measurement error, and model misspecification.

Conclusion: A combination of dietary and nondietary lifestyle modifications, begun in midlife or later in relatively healthy women, could have prevented at least half of the cases of type 2 diabetes in this cohort of U.S. women.

A simplified description of the process of estimating risks using parametric g-formula is as follows: we start by fitting regression models for all potential confounders and for the disease outcome using data on the entire study population. We then use these models to simulate the risk of the disease under various interventions in five steps: (1) take the observed joint distribution of covariates at baseline; (2) estimate the joint distribution of time-varying covariates at the next time-point using the parametric models; (3) "intervene" by setting the values of some covariates to values determined by the hypothetical interventions; (4) estimate the predicted probability of the outcome using these new values; (5) repeat steps 2 through 4 for the entire duration of follow-up to estimate the predicted risk of the disease under the selected interventions. See articles by Taubman et al[37] and Young et al[39] for a more detailed description.

More formally, the standardized cumulative risk estimated by the g-formula is a weighted average of the risks of type 2 diabetes conditional on the specified intervention values and the observed confounder history. The weights are the probability density functions of the time-varying confounders, which are estimated via parametric regression models. We approximated the weighted average by using a Monte Carlo simulation of 10,000 individuals with the baseline values of covariates sampled from their empirical distribution. The values of time-varying covariates for each 2-year interval were drawn from the distribution estimated via the regression models after setting the values of lifestyle factors to those specified by the interventions.

57

Learning Objectives

- Review and expand definition of confounding
- Revisit methods to control confounding in the design of the study:
  - Randomization, restriction , matching and analysis of matched data
- Discuss statistical adjustment of confounding effects of known and measured confounders
  - Stratification, standardization
  - Multivariate adjustment models
  - Under- /over-adjustment and estimation of amount of confounding
  - Propensity scores
  - Structural measures and G-estimation
- Examine new methods to control for residual and unmeasured confounders
  - Instrumental variable (IV) analysis
  - Probabilistic sensitivity analysis
  - Direct bias simulation
  - Negative control outcome calibration

58

UCSF

Residual confounding

- Various adjustment techniques control only for *between-stratum* confounding, not *within-stratum (residual)* confounding
- What to do:
  - More strata (categories) with narrower boundaries will control confounding more effectively than fewer strata (categories) with broader boundaries
  - Balance between better adjustment using a lot of strata vs. random error in imprecise estimates from thinly spread data
  - *Use instrumental variables analysis*
- Note: The term *residual confounding* is also used to describe confounding from factors that are not controlled at all or from factors that are controlled but are measured inaccurately

UCSF

Under- and over-adjustment for confounding effects

- **Under-adjustment:**
  - If confounders are not identified and not measured
  - If confounders are identified but not measured
  - If confounders are identified, but poorly measured (eg., SES, ethnicity)
- **Over-adjustment:**
  - Logistic regression can accommodate a lot of confounders, but the results are less transparent and more prone to undetected errors:
    - What do estimates mean?
    - Variance increases with the number of variables in the model
    - Reduction in precision of risk estimates will make it more difficult to detect a true association
  - Effects of some confounders may depend on absence or presence of other confounders (joint confounding)

UCSF

Unmeasured confounding

- Regardless of our best efforts, there is likely to be some residual confounding in analysis strata. Thus, *stratum-specific and summary estimates of associations of exposure with disease* can **differ** considerably from the *stratum-specific and summary effects of exposure on disease*. The latter could be estimated by allowing for residual bias.

Analysis of unmeasured confounding: Instrumental variables

Example: Assume that month and state of birth would have no impact on earnings later in life, except that state laws determine when one can start and end his or her K–12 education

IV model can be conceptualized as two separate equations estimated simultaneously: one specifies the relationship between the key independent variable and the outcome, the second specifies the relationship between the instrumental variables and the outcome. Thus,

1. $Y = X1B1 + KB2 + e1$
2. $K = X1B3 + IVB4 + e2$

where *Y* is the outcome; *K* is the key independent variable measuring the policy; *IV* is the instrumental variable; *X1* is a vector of other control variables; *B1*, *B2*, *B3*, and *B4* are parameters to be estimated; and *e1* and *e2* are error terms.

Example:

Comparison of conventional, PS, and IV adjustment methods to reduce confounding by indication

Treatment decisions are often related to disease severity, prognosis, and frailty. Therefore, evaluating treatment effects using observational data frequently encounters **confounding by indication**.

Design: 4,275 Iowa Medicaid patients aged 3-25 years with new or uncontrolled persistent asthma and at least 1 long-term control (LTC) therapy prescription in 1997-1999

Research question: What are the effects of LTC therapy on the occurrence of acute asthma exacerbation events?

Exposure: LTC, including inhaled corticosteroids, the combination of inhaled corticosteroids and long-acting beta agonists

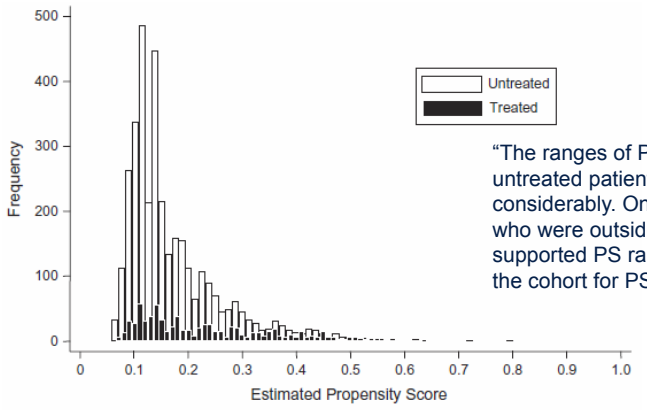
PS: Multivariate logistic regression was used to estimate the PS for receiving LTC therapy as a function of the measured covariates

IV: Area treatment ratios (ATRs) for each zip code calculated as the ratio of the observed treatment rate to the expected treatment rate (estimated from covariates in the regression model). ATR>1 indicates an area practice style that uses more LTC therapy for persistent asthma patients than would be expected on average. The larger the ATR, the stronger the preference in an area for use of LTC therapy.

63 Fang et al. *AJE* 2012

UCSF

Propensity score distribution for asthma long-term control therapies

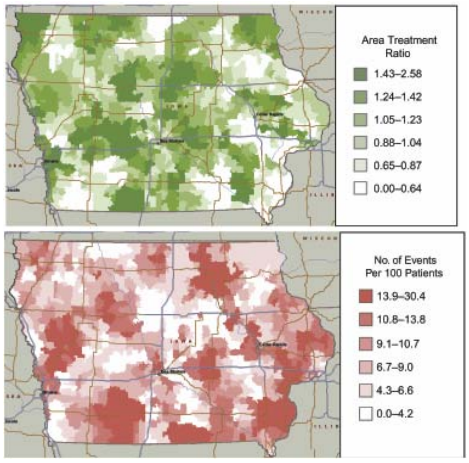


"The ranges of PS between treated and untreated patients overlapped considerably. Only 16 patients (0.37%) who were outside the commonly supported PS range were excluded from the cohort for PS adjustment analyses."

64 Fang et al. *AJE* 2012

UCSF

Geographic variation in use of LTC therapy and acute asthma exacerbation events during a 2-month follow-up



65 Fang et al. *AJE* 2012

UCSF

Fang et al. 2012, *Results*

	Actual LTC Therapy			Instrumental Variable					
				Median Groups of LTC Therapy ATR			Tertile Groups of LTC Therapy ATR		
	Yes (n = 787)	No (n = 3,488)	Absolute Difference	Upper Group (Above Median) (n = 2,138)	Lower Group (Below Median) (n = 2,137)	Absolute Difference	Third Tertile (n = 1,425)	First Tertile (n = 1,425)	Absolute Difference
Use of LTC therapy	100.0	0.0	100.0***	24.0	12.7	11.3***	25.9	10.3	15.6***

Model[a]	LTC Treatment Effectiveness	
	Estimate	95% Confidence Interval
Linear regression model		
Unadjusted	0.033	0.010, 0.056
Multivariate-adjusted	0.011	−0.013, 0.034
PS adjustment		
Continuous PS	0.010	−0.014, 0.033
Quintiles of PS	0.016	−0.008, 0.039
Nearest-neighbor matching	0.016	−0.022, 0.053[d]
Instrumental variable 2-stage least squares model		
Median groups of LTC therapy ATR as instrument	−0.161	−0.320, −0.001
Tertile groups of LTC therapy ATR as instrument	−0.166	−0.308, −0.024

- **Adjustment variables:**
- Patient demographic characteristics (age, gender, and poverty level)
- Patient clinical characteristics (total number of physician office visits with asthma, acute asthma exacerbation events within the 3 months prior to the index date; use of SABA (puffs/week) for the first SABA prescription; season)

66

UCSF

Fang et al. 2012, *Summary*

- RCT studies demonstrated that LTC therapies reduce acute asthma exacerbation in patients with persistent asthma
- patients who initiated LTC therapy had a more severe asthma status at baseline (confounding by indication)
- Risk adjustment (RA) and PS-adjusted measures of effect did not support RCT results (but data on severity of asthma was lacking)
- IV-adjusted measures of effect suggested significant protective effects of LTC therapy on acute asthma exacerbation
- RA methods assume that adjustment for measured covariates also adjusts for unmeasured covariates, while the IV approach requires specified instruments with a strong association with treatment choice and assumes that the specified instruments are unrelated to unmeasured confounders or outcomes directly.

UCSF

Analysis of unmeasured confounding: External adjustment (sensitivity analysis)

- Make assumptions about confounder – disease associations within exposure strata (are they constant?)
- Make assumptions about associations between exposure and confounder in the source population
- Consider the joint effects of measured and unmeasured confounders
- Compare estimated adjusted risk estimates with the unadjusted

UCSF

Analysis of unmeasured confounding: Probabilistic sensitivity analysis

(Monte-Carlo simulations)

- Extends simple sensitivity analysis by assigning probability distributions to the parameters rather than using a few fixed values for the parameters
- At each iteration of a Monte-Carlo analysis, values of the unknown confounder parameters are randomly selected from their assigned probability distributions and then used to produce a frequency distribution of adjusted estimates of the target parameter
- 2.5% and 97.5% limits of the distribution are the limits of an interval that contains 95% of the simulated estimates (Monte-Carlo simulation interval (MCSI))
- Could be additionally adjusted for random error

Example:

Sensitivity analyses in the study of risks of leukemia in Chernobyl cleanup workers

TABLE 4
Risk Estimates Based on the Linear ERR and Log-Linear Models: Impact of Different Analytic Strategies; All Studied Hematological Malignancies Combined (results based on the main data set[a])

		Linear ERR model				Log-linear model		
	Cases/ Controls	ERR/ 100 mGy	90%	CI	RR at 100 mGy	RR at 100 mGy	90%	CI
Different populations								
Subjects who worked in 30-km zone	54/180	**0.66**	0.00	2.65	**1.66**	**1.29**	0.95	1.77
Cases with confirmed diagnosis and their matched controls	34/140	**0.28**	−0.24	2.34	**1.28**	**1.26**	0.69	2.35
Cases with confirmed diagnosis and their matched controls (30-km zone)	27/83	**0.29**	−0.24	2.41	**1.29**	**1.28**	0.69	2.49
Entire population[b]	117/481	**0.00**	nd[c]	0.43	**0.99**	**0.99**	0.77	1.26
Entire population excluding SEAD doses[d]	105/429	**0.25**	−0.10	1.13	**1.25**	**1.15**	0.87	1.50
Dose without constraint	70/287	**0.14**	−0.06	0.89	**1.14**	**1.07**	0.91	1.23
Unconditional logistic regression[e]	70/429	**0.39**	−0.06	1.57	**1.39**	**1.24**	0.92	1.65
Dose uncertainties	70/287	**0.60**	−0.01	2.58	**1.60**	**1.29**	0.95	1.90

[a] Liquidators whose dose was estimated with RADRUE and with interview data with liquidators or colleagues.
[b] All liquidators included in the case-control study; their dose was estimated by the RADRUE method or by the SEAD method, and their interview data were completed by themselves or by a proxy (a colleague or a relative).
[c] Not defined.
[d] All liquidators included in the case-control study except those whose dose was estimated by the SEAD method.
[e] Stratified on region and age at the time of the accident in nine categories (<25, 25–30, 30–35, 35–40, 40–45, 45–50, 50–55, 55–60 and 60+).

Estimating effects of unmeasured confounders and random errors in doses

- dose–response models were fitted to each of the 10,000 data sets corresponding to the 10,000 realizations of the doses for each subject. An integrated profile likelihood was then generated by averaging the likelihoods at each of the 100 points over all of the 10,000 simulations, thus providing a MLE and a confidence interval that take into account both the statistical error of the model and the dosimetric uncertainties.
- dose-response models are fitted 10,000 times as parameter estimates are randomly sampled from their underlying distributions.

UCSF

Example:

Sensitivity analyses in the study of risks of leukemia in Chernobyl cleanup workers

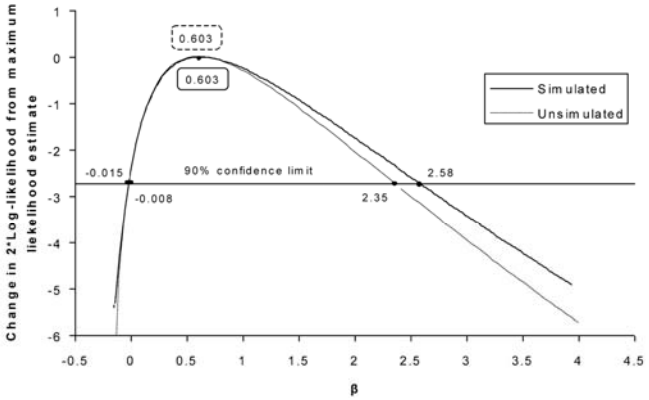


FIG. 4. Profile likelihood, MLE and 90% CI for β from both the unsimulated data and the 10,000 simulations, using the linear ERR model: all hematological malignancies combined (results based on the main data set[a]).

 Kesminiene et al. *Rad Res* 2008

UCSF

Example:
Analysis of uncertainties in dose estimation on the risk estimates

- "When dose errors are modest, a commonly used method of dealing with dose error is to replace the dose estimate in any regression with the expected true dose given the measured dose estimate, a process termed regression calibration (Carroll *et al*, 2006). When dose uncertainties are more substantial, full-likelihood methods may be indicated, in particular Monte Carlo likelihood integration (MCML) (Fearn *et al*, 2008; Stram & Kopecky, 2003), and Bayesian Markov Chain Monte Carlo (MCMC) (Kwon *et al*, 2014; Little *et al*, 2008)."

73 Little et al. *PLoS One* 2014

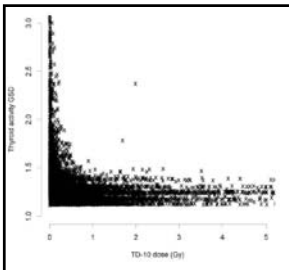


Figure 2. Distribution of the geometric standard deviation (GSD) of errors associated with assessments of thyroid activity GSD as a function of TD-10 thyroid dose. Low dose range.
doi:10.1371/journal.pone.0085723.g002

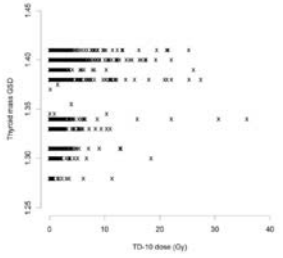


Figure 3. Distribution of the geometric standard deviation (GSD) of errors associated with assessments of thyroid mass GSD as a function of TD-10 thyroid dose.
doi:10.1371/journal.pone.0085723.g003

Sources of error:

- thyroid mass at the time of the Chernobyl accident from population-average measurements, supplemented with autopsy measurements
- Measurement instrument errors
- Both shared and unshared errors

Table 2. Analysis of curvature in fits of EOR model (2) with or without adjustment for dose errors using regression calibration, fo TD-10 doses.

Dose	Dose-response model	*p*-value[a]	Linear ERR (α) (Gy^{-1}) (+95% CI)	Exponential ERR (γ) (Gy^{-1}) (+95% CI)
Tronko *et al.* (TD-02) dose	αD	<0.001[b]	5.25 (1.70, 27.45)	
Tronko *et al.* (TD-02) dose	$\alpha D \exp[\gamma D]$	0.084	9.13 (2.46, 111.1)	−0.09 (−0.23, 0.01)
TD-10 unadjusted dose	αD	<0.001[b]	5.38 (1.86, 21.01)	
TD-10 unadjusted dose	$\alpha D \exp[\gamma D]$	0.104	8.85 (2.60, 54.58)	−0.11 (−0.29, 0.02)
1st regression calibration method (Kukush *et al.*) adjusted dose	αD	<0.001[b]	5.78 (1.92, 27.04)	
1st regression calibration method (Kukush *et al.*) adjusted dose	$\alpha D \exp[\gamma D]$	0.112	9.72 (2.67, 94.31)	−0.10 (−0.28, 0.02)
2nd regression-calibration method adjusted dose	αD	<0.001[b]	4.78 (1.64, 19.69)	
2nd regression-calibration method adjusted dose	$\alpha D \exp[\gamma D]$	0.101	8.19 (2.33, 60.87)	−0.09 (−0.25, 0.02)
Monte Carlo maximum likelihood	αD	<0.001[b]	4.93 (1.67, 19.90)	
	$\alpha D \exp[\gamma D]$	0.102	7.97 (2.32, 49.81)	−0.09 (−0.26, 0.01)

All models have underlying rates adjusted for age (treated categorically) and gender. Unless otherwise stated all CI are profile-likelihood based.
[a]unless otherwise stated all *p*-values refer to the improvement in fit of the current row in the Table with that of the model fitted in the row immediately above.
[b]*p*-value of improvement in fit compared with a model with no dose term.

Little et al. *PLoS One* 2014

74

Analysis of unmeasured confounding: Direct bias simulation

- Look for published studies of the same association which measured confounder in question and reported both unadjusted and adjusted estimates
- Calculate an estimate of the bias due to failing to adjust for the confounder (with CIs if possible)
- Use the estimate as a basis for a prior on the size of the unmeasured confounding

Analysis of unmeasured confounding: Negative Control Outcome Calibration

- A negative control outcome found to be empirically associated with the exposure after adjustment for observed confounders indicates that unobserved confounding may be present
- Counterfactual-based approach to correct causal effect estimates for bias due to unobserved confounding
- **Example**: Consumption of mercury-contaminated fish and amount of mercury in person's blood and percent of cells with chromosomal abnormalities
- Negative control outcome: asthma and influenza diagnoses

COCA, control outcome calibration approach

- A = mercury in fish
- Y = mercury rate in blood
- Z = influenza
- Assumption: no common cause of A and Z that does not also confound the relation between fish consumption and mercury in the blood

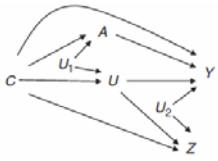


Figure 1. Causal diagram depicting unmeasured confounding of the $A - Y$ association and the negative control outcome Z, where A represents exposure; Y represents outcome; U, U_1, and U_2 represent unobserved confounders; and C represents observed confounders.

!!! relies on the non–empirically verifiable assumption that the negative control outcome is carefully chosen so that it is solely influenced by observed and unobserved confounders of the exposure-outcome relation in view

 Tchetgen Tchetgen *AJE* 2013

UCSF

Confounding, Conclusions

- Confounding is *a distortion or misattribution of effect to a particular study factor*. It results from noncomparability of a comparison group.
- A confounder should appear as an independent risk factor, i.e., not one whose association with disease results from its association with the study factor. There are multiple caveats to the "well-known" 3 criteria for confounding.
- Adequacy of control of confounding effects is compromised by *errors in the conceptualization, measurement, coding, and model specification* for potential confounders.
- Important to *remember unmeasured and residual confounding*, under- and over-adjustment for confounding effects.

UCSF

EXTRA SLIDES: CHECK YOUR KNOWLEDGE OF CONFOUNDING

Intellectually curious?

Question: The larger a randomized trial, the less the possibility for confounding. Why? Does this logic also apply to non-experimental studies? Why?

Intellectually curious?

Question: The larger a randomized trial, the less the possibility for confounding. Why? Does this logic also apply to non-experimental studies? Why?

Answer: The larger the trial, the smaller the baseline differences, because of the law of large numbers. Therefore the larger the trial the less the amount of confounding that can be expected. In contrast, in a non-experimental study, with no random assignment, larger observational study sizes do not lead to less confounding, because the differences between the exposed and unexposed groups with respect to potential confounding factors do not depend on the size of the study.

UCSF

Intellectually curious?

Question: Suppose that an investigator conducting an RCT of an old and a new treatment examines baseline characteristics of the subjects (such as age, sex, stage of disease, and so forth) that might be confounding factors and finds that the two groups are significantly different with respect to several characteristics. A significance test is a test of the null hypothesis, which is a hypothesis that chance alone can account for the observed difference. What is the explanation for baseline differences in a randomized trial? What implication does that explanation have for dealing with these differences?

UCSF

Intellectually curious?

Question: Suppose that an investigator conducting an RCT of an old and a new treatment examines baseline characteristics of the subjects (such as age, sex, stage of disease, and so forth) that might be confounding factors and finds that the two groups are significantly different with respect to several characteristics. A significance test is a test of the null hypothesis, which is a hypothesis that chance alone can account for the observed difference. What is the explanation for baseline differences in a randomized trial? What implication does that explanation have for dealing with these differences?

Answer: If the trial randomly assigns participants to the study groups, then any baseline differences are a consequence of the random assignment, so chance (the null hypothesis) accounts for all baseline differences. As a result, the significance test does not make any sense, because the null hypothesis is already known to be true. With respect to the control of confounding, it does not matter whether the baseline differences are caused by chance or not - any substantial confounding should be controlled.

UCSF