Overview of multivariable analysis

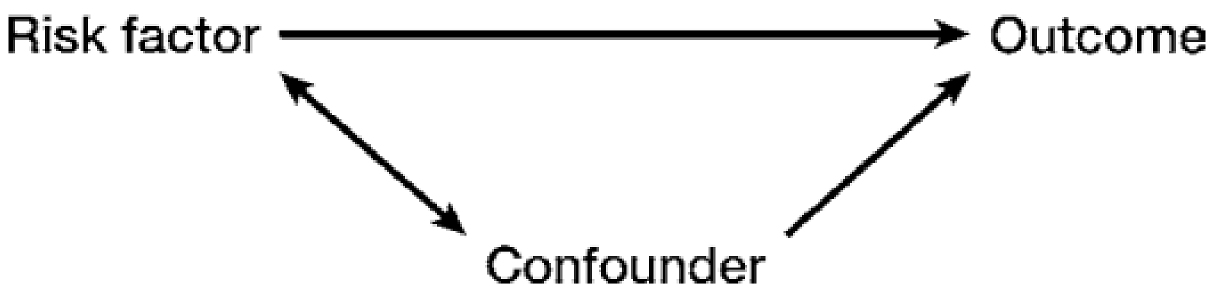
Dr Thida
MBBS, MMedSc (Pubic Health), PhD (Epidemiology)
Deputy Director/Head
Epidemiology Research Division
Department of Medical Research (POLB)

22nd -26th July 2019 Department of Medical Research (Pyin Oo Lwin branch)

- Important of multivariable analysis
  - Most diseases have multiple causes, and prognosis is usually determined by a large number of factors
  - Enable to sort out the multifaceted nature of risk factors and their relative contribution to outcome

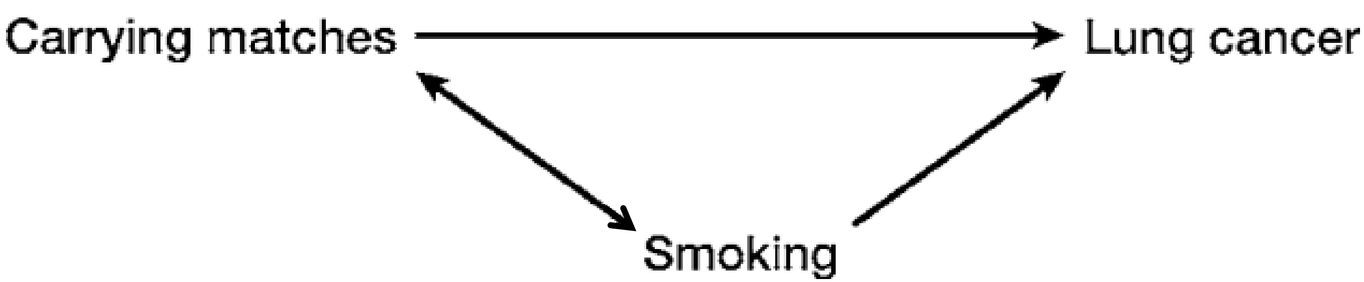
What are confounders and how does multivariable analysis help me to deal with them?

- Confounding occurs when the apparent association between a risk factor and an outcome is affected by the relationship of a third variable to the risk factor and the outcome; the third variable is called a confounder.



Relationships among risk factor, confounder, and outcome.

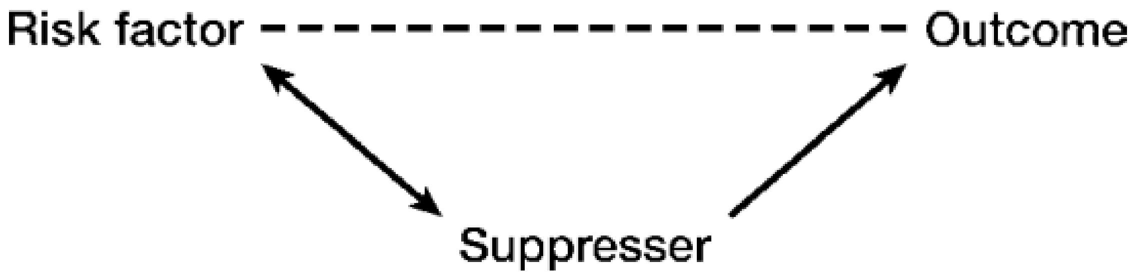
- A confounder is associated with the risk factor and causally related to the outcome.



Relationships among carrying matches, smoking, and lung cancer.

What are suppressers and how does multivariable analysis help me to deal with them?

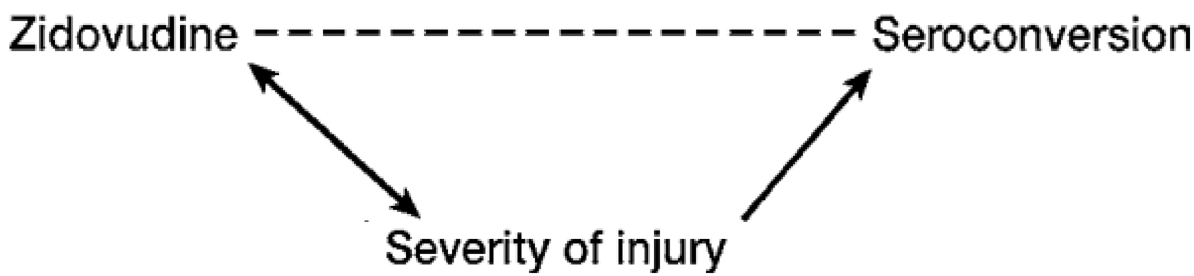
- A type of confounder
- As with confounders, a suppresser is associated with the risk factor and the outcome



Relationships among risk factor, suppresser, and outcome.

- On bivariate analysis there is no effect seen between the risk factor and the outcome
- When you adjust for the suppresser, the relationship between the risk factor and the outcome becomes significant.

- Unlike a typical confounder, when you have a suppresser you won't see any bivariate association between the risk factor and the outcome until you adjust for the suppresser.



Relationships among zidovudine, severity of injury, and seroconversion.

What are interactions and how does multivariable analysis help me to deal with them?

- An interaction occurs when the impact of a risk factor on outcome is changed by the value of a third variable.

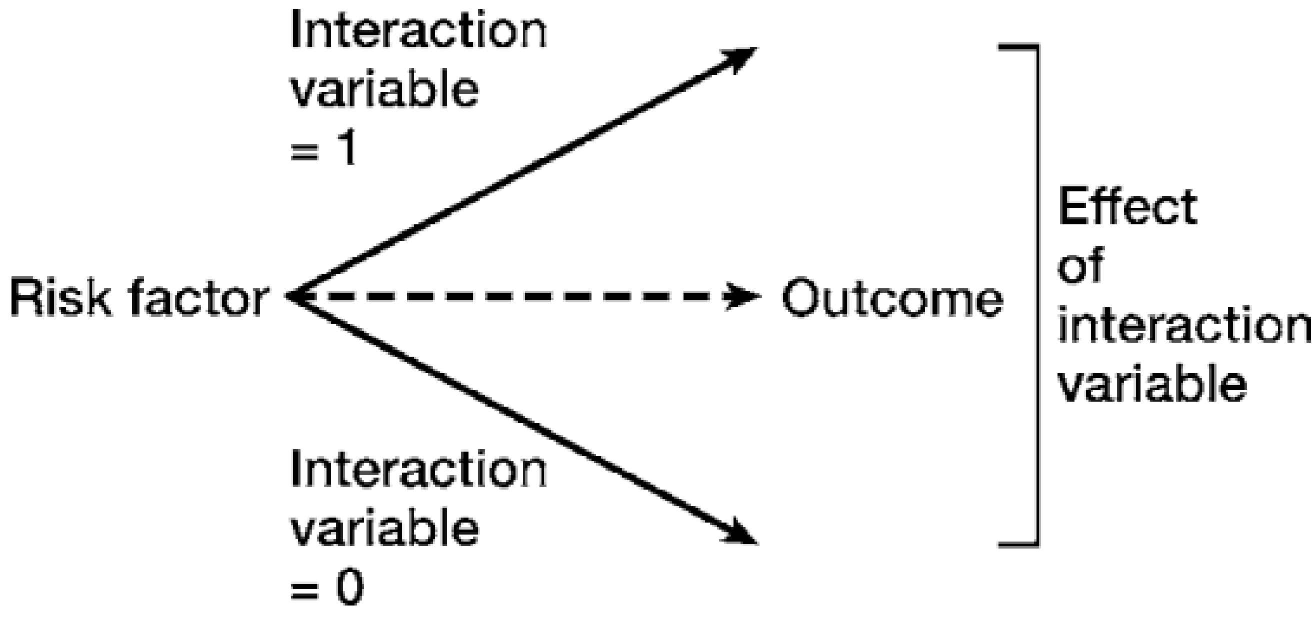


Illustration of an interaction effect.

Baseline characteristics of survivors and decedents, Aerobics Center Longitudinal study.

Characteristics	Men: Survivors (n = 24 740)	Men: Decedents (n = 601)
Age, y (SD)	42.7 (9.7)	52.1 (11.4)
Body mass index, kg/m^2 (SD)	26.0 (3.6)	26.3 (3.5)
Systolic blood pressure, mm Hg (SD)	121.1 (13.5)	130.4 (19.1)
Total cholesterol, mg/dL (SD)	213.1 (40.6)	228.9 (45.4)
Fasting glucose, mg/dL (SD)	100.4 (16.3)	108.1 (32.0)
Fitness, %		
Low	20.1	41.6
Moderate	42.0	39.1
High	37.9	19.3
Current or recent smoker, %	26.3	36.9
Family history of coronary heart disease, %	25.4	33.8
Abnormal electrocardiogram, %	6.9	26.3
Chronic illness, %	18.4	40.3

Was there any way to answer the multivariable association with the outcome?

- Performing stratified analysis.
- Stratified analysis assesses the effect of a risk factor on outcome while holding another variable constant.

	Deaths per 10 000 person-years	Stratum-specific relative risk (95% CI)
Smokers		
Low fitness	48.0	1.63 (1.26–2.13)
Moderate/high fitness	29.4	1.0 (ref.)
Nonsmokers		
Low fitness	44.0	2.19 (1.77–2.70)
Moderate/high fitness	20.1	1.0 (ref.)

Data supplied by Aerobics Center Longitudinal Study.

Inconveniences of performing stratified analysis

- Two variables (smoking and cholesterol), relationship between fitness and mortality in four groups
  - smokers with high cholesterol
  - smokers with low cholesterol
  - nonsmokers with high cholesterol
  - nonsmokers with low cholesterol
- Three variables: 8 groups ($2 \times 2 \times 2 = 8$)
- Four variables: 16 groups ($2 \times 2 \times 2 \times 2 = 16$)

Multivariable analysis is preferable to stratified analysis when you have multiple confounders.

Multivariable analysis of risk factors for all-cause mortality, Aerobics Center Longitudinal Study.

	Men	
Independent variable	Deaths per 10 000 person-years	Adjusted relative risk (95% CI)
Fitness		
Low	38.1	1.52 (1.28–1.82)
Moderate/High	25.0	1.0 (ref.)
Smoking status		
Current or recent smoker	39.4	1.65 (1.39–1.97)
Past or never smoked	23.9	1.0 (ref.)
Systolic blood pressure		
≥140 mm Hg	35.6	1.30 (1.08–1.58)
<140 mm Hg	27.3	1.0 (ref.)
Cholesterol		
≥240 mg/dL	35.1	1.34 (1.13–1.59)
<240 mg/dL	26.1	1.0 (ref.)
Family history of coronary heart disease		
Yes	29.9	1.07 (0.90–1.29)
No	27.8	1.0 (ref.)

Bivariate association between smoking status and risk of death

Bivariate	Nonsmokers	Former smokers	Recent quitters	Persistent smokers
Relative risk of death	1.0 (ref.)	1.08 (0.92–1.26)	0.56 (0.40–0.77)	0.74 (0.59–0.94)

Adapted from Hasdai, D., *et al*. "Effect of smoking status on the long-term outcome after successful percutaneous coronary revascularization." *N. Engl. J. Med*. **336** (1997): 755–61.

Association between demographic and clinical factors and smoking status

	Nonsmokers	Former smokers	Recent quitters	Persistent smokers
Age, year ± SD	67 ± 11	65 ± 10	56 ± 10	55 ± 11
Duration of angina, months ± SD	41 ± 66	51 ± 72	21 ± 46	29 ± 55
Diabetes, %	21%	18%	8%	10%
Hypertension, %	54%	48%	38%	39%
Extent of coronary artery disease, %				
One vessel	50%	51%	57%	55%
Two vessels	36%	36%	34%	36%
Three vessels	14%	13%	10%	9%

Adapted from Hasdai, D., *et al*. "Effect of smoking status on the long-term outcome after successful percutaneous coronary revascularization." *N. Engl. J. Med*. **336** (1997): 755–61.

Comparison of bivariate and multivariable association between smoking status and risk of death

	Nonsmokers	Former smokers	Recent quitters	Persistent smokers
Relative risk of death (bivariate)	1.0 (ref.)	1.08 (0.92–1.26)	0.56 (0.40–0.77)	0.74 (0.59–0.94)
Relative risk of death (multivariable)	1.0 (ref.)	1.34 (1.14–1.57)	1.21 (0.87–1.70)	1.76 (1.37–2.26)

Adapted from Hasdai, D., *et al*. "Effect of smoking status on the long-term outcome after successful percutaneous coronary revascularization." *N. Engl. J. Med*. **336** (1997): 755–61.

- The difference between the bivariate and multivariable analysis indicates that confounding is present

Association of independent variables with confirmed diagnosis of acute myocardial infarction based on multiple logistic regression model.

Independent variables	Coefficients	Odds ratio
Male gender	**0.4852**	**1.6**
Age <50	0.1432	1.2
Chest pain	0.8792	2.4
Chief complaint: chest pain	0.4399	1.6
Nausea/vomiting	0.5153	1.7
Congestive heart failure	0.6759	2.0
White race	0.0987	1.1
ST elevation	**2.0948**	**8.1**
ST depression	1.2632	3.5
Q waves	0.5311	1.7
History of diabetes mellitus	0.2781	1.3
History of hypertension	0.2032	1.2
History of angina	−0.2976	0.7
History of peptic ulcers	−0.3210	0.7
Dizziness	−0.4437	0.6
Interactions		
Male gender and congestive heart failure	−0.6899	0.5
Male gender and ST elevation	**−0.5187**	**0.6**
Male gender and white race	0.5206	1.7

Adapted with permission from Zucker, D. R., *et al.* "Presentations of acute myocardial infarction in men and women." *J. Gen. Intern. Med.* **12** (1997): 79–87.

- Odds ratio of male gender x ST elevations x men with ST elevations (1.6 × 8.1 × 0.6 = 7.8)
- Odds ratio of female gender x ST elevations x female with ST elevations (1.0 × 8.1× 1.0 = 8.1)

Common uses of multivariable models

1. Observational studies of etiology
2. Intervention studies (randomized and nonrandomized)
3. Studies of diagnosis
4. Studies of prognosis

Type of outcome variable determines choice of multivariable analysis.

Type of outcome	Example of outcome variable	Type of bivariate analysis	Type of multivariable analysis
Interval	Blood pressure, weight, temperature	Correlation coefficient, linear regression, *t* test, ANOVA	Multiple linear regression, analysis of variance (and related procedures)
Dichotomous	Death, cancer, intensive care unit admission	Chi-squared, Fisher's exact, *t* test, chi-squared for trend, Mann–Whitney test	Multiple logistic regression
Ordinal	Stage of disease, severity of symptoms	Chi-squared for trend, Mann–Whitney test, Spearman's rank correlation coefficient	Proportional odds regression
Nominal	Cause of death, site of cancer	Chi-squared, ANOVA, Kruskal–Wallis	Multinomial logistic regression
Time to outcome	Time to death, time to cancer	Log-rank	Proportional hazards analysis

Reference

- Mitchell H. Katz (2011). Multivariable analysis: a practical guide for clinicians and public health researchers (3rd Edition), the United States of America by Cambridge University Press, New York
- Bland JM, Altman DG. The log rank test. BMJ. 2004 May 1;328(7447):1073.

Thank you for your attention!