Matching

Heather J. Baer EPI 208 July 27, 2016

Design Options to Avoid Confounding

- Randomization
- Restriction
- Matching: restricts enrollment within comparison group
 - Subjects in comparison group are chosen so that group has similar or identical distribution of the matching factor(s) as the index group
 - Index group = exposed in cohort study, cases in case-control study

Implications of Matching

- · Cohort study
 - Choose <u>unexposed</u> subjects so that they have same distribution of matching factor(s) as <u>exposed</u> subjects
 - Matching factor will no longer be a confounder
- Case-control study
 - Choose <u>controls</u> so that they have same distribution of matching factor(s) as <u>cases</u>
 - But matching alone does <u>not</u> get rid of confounding; still need to account for matching in the analysis

Individual vs. Frequency Matching

- Individual matching: identify individual subjects for comparison, each resembling a study subject on matched variable(s)
 - Example: cohort study of exercise and colon cancer, where age and sex are potential confounders
 - If exposed subject (exerciser) is a 55-yo male, find unexposed subject who is a 55-yo male
- Frequency matching: category matching that balances the proportion of people with a confounding factor in compared groups
 - Example: cohort study where 20% of exposed are 40-49 yo, 40% are 50-59 yo, 20% are 60-69 yo, and 20% are ≥ 70 yo
 - Frequency matching would ensure same age distribution in unexposed group (same % in each age category)

Motivation for Matching

- · To avoid confounding
 - Fulfilled for cohort studies
 - But still need to account for matching in analysis for case-control studies
- To improve efficiency of analysis
 - Better ability to control for a strong confounder in the analysis using stratification or regression

Example: Large Population

Suppose you are interested in exposure-disease relation, but sex is a potential confounder.

Association between sex and exposure in the population:

Exposure

	E+	E-	Total
Male	8,000 (80%)	2,000 (20%)	10,000
Female	2,000 (20%)	8,000 (80%)	10,000
Total	10,000	10,000	20,000

Sex is associated with exposure: 80% of exposed subjects are male, 20% of unexposed subjects are male

matching in case control study may affect selection bias.

Example: Large Population

Risk of disease in the population, by sex and exposure group:

		Exposure
	E+	E-
Male	0.06	0.02
Female	0.03	0.01

Sex is an independent risk factor for the outcome; among unexposed, RR for males vs. females = 0.02 / 0.01 = 2.0

Example: Large Population

Expected number of outcomes in the population, by sex and exposure group (# subjects x risk for each group):

		Exposure
	E+	E-
Male	480	40
Female	60	80
Total	540	120

the way of matching

Study 1: Cohort Study, Matched on Sex

- For each exposed subject who is male, enroll unexposed subject who is male
- For each exposed subject who is female, enroll unexposed subject who is female
- Result: same sex distribution in exposed and unexposed groups

Example: Large Population

	<u>M</u>	ales			<u> </u>	emales	
	Dis	ease				Disease	
	+	-	Total		+	-	Total
+	480	7520	8000	E+	60	1940	2000
-	40	1960	2000	E-	80	7920	8000
	1	RR = 3.0				RR = 3.0	
				Disease			
		+		-		Total	
Ex	posed	540)	9460		10,000	
Uı	nexposed	120)	9880		10,000	
			R	R _{crude} = 4.5			

Study 1: Cohort Study, Matched on Sex

Sex distribution among 1000 exposed and among 1000 matched unexposed subjects:

Exposure

	+	 same number
Male	800 (80%)	800 (80%)
Female	200 (20%)	200 (20%)
Total	1000	1000

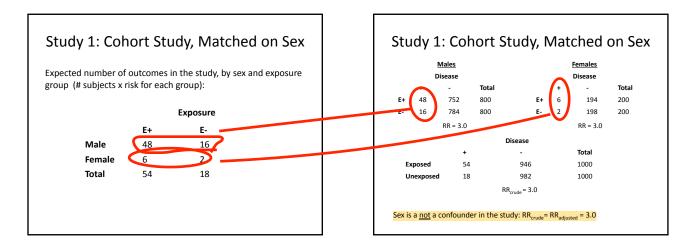
Sex is \underline{not} associated with exposure in the study: 80% of exposed subjects are male \underline{and} 80% of unexposed subjects are male

Study 1: Cohort Study, Matched on Sex

Risk of disease in the study, by sex and exposure group: (same as in the original population)

		Exposure
	E+	E-
Male	0.06	0.02
Female	0.03	0.01

Sex is an independent risk factor for the outcome; among unexposed, RR for males vs. females = 0.02 / 0.01 = 2.0



Study 1: Conclusion

- Sex is not a confounder in the <u>matched cohort</u> study
- RR_{crude} = RR_{adjusted} = 3.0
- Matching eliminated confounding by sex
 - No need to do stratified analysis
 - Get valid estimate from the crude analysis (assuming no other confounding or bias)

Study 2: Case-Control Study, Matched on Sex

- For each case who is male, enroll control who who is male
- For each <u>case</u> who is <u>female</u>, enroll <u>control</u> who is <u>female</u>
- Result: same sex distribution in case and control groups

the number of case = cont

Study 2: Case-Control Study, Matched on Sex

Sex distribution among 660 cases and 660 matched controls (if took all cases from the original population):

	Case	same number pecause of the mate	h
Male	520	520 Secause of the mate	,
Female	140	140	
Total	660	660	

Question: What would be the Odds Ratio for sex in the

matched case-control study?

once you match in case control odds =1 (ad/bc))

sex became the not risk

Study 2: Case-Control Study, Matched on Sex

Number of <u>cases</u> by sex and exposure group (same as in the original population):

	Ex	posure	
	E+	E-	Total
Male	480	40	520
Female	60	80	140
Total	540	120	660

Study 2: Case-Control Study, Matched on Sex

Number of <u>controls</u> by sex and exposure group (based on exposure distribution in the original population):

Exposure

	E+	E-	Total
Male	416 (80%)	104 (20%)	520
Female	28 (20%)	112 (80%)	140
Total	444	216	660

In the population, 80% of males and 20% of females were exposed; therefore, we would expect 80% of male controls and 20% of female controls to be exposed

Study 2: Case-Control Study, Matched on Sex

	ļ	Males			<u> </u>	emales
	Case	Control			Case	Control
E+	480	416		E+	60	28
E-	40	104		E-	80	112
Total	520	520		Total	140	140
	OR	t = 3.0			QR = 3	0_
			Case	Co	ntrol	
	Expo	osed	540	44	4	
	Une	xposed	120	21	6	
	Tota	l	660	66	0	
			OR _{cruc}	le= 2.2		

Study 2: Conclusion

- $OR_{crude} = 2.2 \neq OR_{adjusted} = 3.0$
- Sex still appears to be a confounder in the matched case-control study - why?
- Controls <u>not</u> sampled independently of exposure (selection bias)
 - Matched controls are more similar to cases in terms of exposure, because matching factor (e.g., sex) is associated with exposure
 - In matched case-control study, crude OR will be biased toward null (closer to 1.0 than adjusted OR)
- Still need to account for the matching in the analysis!

So Why Match in a Case-Control Study?

- To improve efficiency of stratified analyses
 - Matching helps ensure sufficient numbers of cases and controls in each category of matching factor
- Example: case-control study of risk factors for prostate cancer, matched on age (± 2 years)
 - If did <u>not</u> match on age, age distribution will differ between cases and controls (cases more likely to be older)
 - If stratify on age <u>without</u> matching, some age strata may have many cases but few controls and vice versa
 - Matching on age will tend to make constant case:control ratio across age strata; thus, better ability to control for age in the stratified analysis than if had not matched

How to Account for Matching in Analysis

- · Stratify by matching factor
- · Adjust for matching factor in regression model
- Preserve matched pairs ("matched" analysis)
 - McNemar's test
 - Paired t-test or Wilcoxon signed-rank test
 - Conditional logistic regression

When to Match in a Case-Control Study

- Matching factor is strong potential confounder
 Definitely planning to control for it in the analysis
- Not interested in examining the effect of the matching factor on outcome
- Information on subjects' value for matching factor is easy to obtain (e.g., age, sex)
- May be interested in examining matching factor as effect modifier (subgroup analyses)

Ratio of Controls to Cases

- May only be able to do 1:1 ratio of controls to cases, given limited resources
- If have sufficient resources, can increase precision by increasing number of controls per case (2:1, 3:1, etc.)
- Marginal increase in precision from increasing ratio of controls to cases beyond 4:1
- Best way to increase precision in a case-control study is to increase number of cases by widening base geographically or temporally

Final Comments about Matching

- · Rarely used in cohort studies
 - Challenging logistically
 - Can reduce sample size if can't find matches
 - Propensity score matching sometimes used (matching on summary confounder "score", rather than on individual confounders)
- · More often used in case-control studies
 - Doesn't "fix" confounding, but can make it easier to control for it in analysis (i.e., increased efficiency)
 - Choose matching factors carefully and be sure account for matching in the analysis

Example: Matched Cohort Study

Incident comorbidities and all-cause mortality among 5-year survivors of Stage 1 and 11 breast cancer diagnosed at age 65 or older: a prospective-matched cohort study

Jennifer H. Jordan ^{*} Soe Soe Thwin ^{*} Timothy L. Lash ^{*} Diana S. M. Buist Terry S. Field ^{*} Reina Haque ^{*} Pamala A. Pawloski ^{*} Hans V. Petersen ^{*} Mariame N. Prout ^{*} Virginda ^{*} Quinn ^{*} Marianne Ulcickas Yood ^{*} Rebecca A. Silliman ^{*} Ann M. Geiger

Abstract Five-year breast cancer survivors, diagnosed after 65 years of age, may develop more incident comes biddities than similar populations free of cancer. We investigated whether older breast cancer survivors have a similar comorbidity burden 6-15 years after cancer diagnosis to matched women free of breast cancer at start of follow-up and whether incident comorbidities are associated with all-cause mortality. In this prospective cohort study, 1,361 older 5-year early-stage breast cancer survivors diagnosed between 1990 and 1994 and 1,51 age- and health system-matched women were followed for 10 years.

Jordan et al. Breast Cancer Res Treat 2014; 146:401-409

Example: Matched Cohort Study

| Table 1 Characteristics of older 5-year survivors of early breast cancer and matched comparison cohort | Breast cancer | Comparison cohort | Ge = 1.561 | Comparison cohort | Ge = 1.561 | Cohort |

Jordan et al. Breast Cancer Res Treat 2014; 146:401-409.

Example: Matched Cohort Study

Table 3 dis-uses mortality after 10 years of follow-up in older 5-year and phrase cancer un-trease and marked corresporates deaths, at 1872.

Age category at index date cyeans.

70-74. 260 1 torfy.

70-75-79. 270 1 torfy

Jordan et al. Breast Cancer Res Treat 2014; 146:401-409.

Example: Matched Case-Control Study

Selective Serotonin Reuptake Inhibitors and the Risk of Acute Pancreatitis

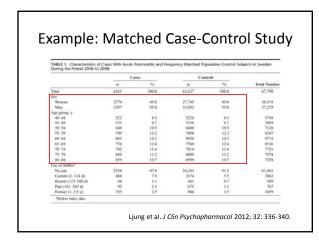
A Swedish Population-Based Case-Control Study

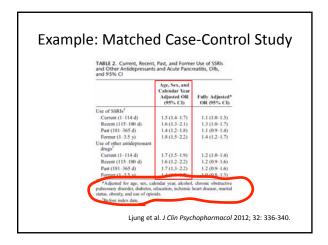
Richard Ljung, MD, PhD.* Christian Rick, MD, PhD.† Fredrik Mattson, RSc.* ionax Sjöberg Beselius, MD, PhD.* Jesper Lagergen, MD, PhD.*‡ and Mats Lindblad, MD, PhD*‡

Abstract: Case reports have indicated an increased risk of acute pancreatitis during use of selective serotonin reuptake inhibitors (SSRIs), an association not found in a few epidemiological studies. We studied the use of SSRI in relation to risk of acute pancreatitis in a population-based case-control study of people aged 40 to 34 years between 2006 and 2008. In Sweden. The Patient Recister was used to identify 6161 cases of first-

in Sweden. The Patient Register was used to identify 6161 cases of firstepisode acute pancreatitis. The Register of the Total Population was used to randomly select 61,637 control subjects from the general population using frequency-based density sampling, matched for age, sex, and calendar year. Use of SSRI was defined as "current," "recent," "pust," or

Ljung et al. J Clin Psychopharmacol 2012; 32: 336-340.





Summary

- Matching is one way to address confounding in design of a study
 - Matching in a cohort study eliminates confounding by the matching factors, but can be challenging logistically
 - Matching in a case-control study does not eliminate confounding; still need to account for matching factors in the analysis, but matching can improve efficiency
 - Be cautious about matching!