Matching
Heather J. Baer, ScD
July 27, 2016

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 unexposed subjects so that they have same distribution of matching factor(s) as exposed subjects
  - Matching factor will no longer be a confounder
- Case-control study
  - Choose controls so that they have same distribution of matching factor(s) as cases
  - But matching alone does not 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.

Matching
Heather J. Baer, ScD
July 27, 2016

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

Example: Large Population

Males

	Disease +	Disease -	Total
E+	480	7520	8000
E-	40	1960	2000

RR = 3.0

Females

	Disease +	Disease -	Total
E+	60	1940	2000
E-	80	7920	8000

RR = 3.0

	Disease +	Disease -	Total
Exposed	540	9460	10,000
Unexposed	120	9880	10,000

$RR_{crude} = 4.5$

Sex is a confounder in the population ($RR_{crude} = 4.5 \neq RR_{adjusted} = 3.0$)

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

Study 1: Cohort Study, Matched on Sex

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

	Exposure +	Exposure -
Male	800 (80%)	800 (80%)
Female	200 (20%)	200 (20%)
Total	1000	1000

same number

Sex is not associated with exposure in the study: 80% of exposed subjects are male 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

Matching
Heather J. Baer, ScD
July 27, 2016

Study 1: Cohort Study, Matched on Sex

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

	Exposure	
	E+	E-
Male	48	16
Female	6	2
Total	54	18

Study 1: Cohort Study, Matched on Sex

Males

	Disease +	Disease -	Total
E+	48	752	800
E-	16	784	800

RR = 3.0

Females

	Disease +	Disease -	Total
E+	6	194	200
E-	2	198	200

RR = 3.0

	Disease +	Disease -	Total
Exposed	54	946	1000
Unexposed	18	982	1000

$RR_{crude} = 3.0$

Sex is a not a confounder in the study: $RR_{crude} = RR_{adjusted} = 3.0$

Study 1: Conclusion

- Sex is not a confounder in the matched cohort 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 case who is female, enroll control who is female
- 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	Control
Male	520	520
Female	140	140
Total	660	660

same number because of the match

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 cases by sex and exposure group (same as in the original population):

	Exposure		
	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 controls 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

Males

	Case	Control
E+	480	416
E-	40	104
Total	520	520

OR = 3.0

Females

	Case	Control
E+	60	28
E-	80	112
Total	140	140

OR = 3.0

	Case	Control
Exposed	540	444
Unexposed	120	216
Total	660	660

OR_{crude} = 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 not 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 not match on age, age distribution will differ between cases and controls (cases more likely to be older)
  - If stratify on age without 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 I and II 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 · Marianne N. Prout · Virginia P. 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 comorbidities 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 1,361 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 survivor cohort[a] (n = 1,361)		Comparison cohort[b] (n = 1,361)	
	No.	%	No.	%
Sociodemographic				
Age category at beginning of follow-up (years)				
70–74	502	37	502	37
75–79	417	31	417	31
80+	442	32	442	32
Race/ethnicity				
Caucasian, non-Hispanic	1,115	82	1,147	84
African-American, non-Hispanic	137	10	125	9.2
Hispanic	72	5.3	62	4.6
Asian/Pacific Islander	37	2.7	27	2.0
Native American	0	0	0	0

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

Example: Matched Cohort Study

Table 3 All-cause mortality after 10 years of follow-up in older 5-year early breast cancer survivors and matched comparison cohort

	Total deaths, n	Adjusted HR[c]	(95 % CI)
Age category at index date (years)			
70–74	263	1 (ref)	–
75–79	289	1.3	(1.1, 1.6)
80+	508	2.8	(2.4, 3.3)
Cohort			
Breast cancer survivors[a]	593	1.3	(1.1, 1.4)
Comparison[b]	467	1 (ref)	–
Any prevalent comorbidity (at 5 years after index date)	614	1.8	(1.6, 2.1)
Incident comorbidity (6–15 years after index date)	773	4.8	(4.1, 5.6)

[a] Older 5-year survivors (n = 1,361) diagnosed 1990–1994 with early Stage I and II breast cancer at age 65 or older followed for 10 years beginning 5 years after index date (date of breast cancer diagnosis)

[b] Comparison cohort (n = 1,361) matched for health system and age, who were free of breast cancer at matched index date and followed for 10 years beginning 5 years after index date

[c] Hazard ratio (HR) adjusted for study site location, age, and presence of prevalent comorbidity. *CI* confidence interval

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

Example: Matched Case-Control Study

ORIGINAL CONTRIBUTION

Selective Serotonin Reuptake Inhibitors and the Risk of Acute Pancreatitis

A Swedish Population-Based Case-Control Study

Richard Ljung, MD, PhD, Christian Rück, MD, PhD,† Fredrik Mattsson, BSc,* Tomas Sjöberg Bexelius, MD, PhD,* Jesper Lagergren, 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 84 years between 2006 and 2008 in Sweden. The Patient Register was used to identify 6161 cases of first-episode 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," "past," or

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

Example: Matched Case-Control Study

TABLE 1. Characteristics of Cases With Acute Pancreatitis and Frequency Matched Population Control Subjects in Sweden During the Period 2006 to 2008

	Cases		Controls		
	n	%	n	%	Total Number
Total	6161	100.0	61,637	100.0	67,798
Sex					
Women	2774	45.0	27,745	45.0	30,519
Men	3387	55.0	33,892	55.0	37,279
Age group, y					
40–44	522	8.5	5226	8.5	5748
45–49	533	8.7	5336	8.7	5869
50–54	648	10.5	6480	10.5	7128
55–59	749	12.2	7498	12.2	8247
60–64	883	14.3	8830	14.3	9713
65–69	776	12.6	7760	12.6	8536
70–74	702	11.4	7019	11.4	7721
75–79	689	11.2	6889	11.2	7578
80–84	659	10.7	6599	10.7	7258
Use of SSRIs*					
No use	5358	87.0	56,243	91.2	61,601
Current (1–114 d)	488	7.9	3374	5.5	3862
Recent (115–180 d)	68	1.1	441	0.7	509
Past (181–365 d)	92	1.5	675	1.1	767
Former (1–3.5 y)	155	2.5	904	1.5	1059

*Before index date.

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

Example: Matched Case-Control Study

TABLE 2. Current, Recent, Past, and Former Use of SSRIs and Other Antidepressants and Acute Pancreatitis, ORs, and 95% CI

	Age, Sex, and Calendar Year Adjusted OR (95% CI)	Fully Adjusted* OR (95% CI)
Use of SSRIs†		
Current (1–114 d)	1.5 (1.4–1.7)	1.1 (1.0–1.3)
Recent (115–180 d)	1.6 (1.3–2.1)	1.3 (1.0–1.7)
Past (181–365 d)	1.4 (1.2–1.8)	1.1 (0.9–1.4)
Former (1–3.5 y)	1.8 (1.5–2.2)	1.4 (1.2–1.7)
Use of other antidepressant drugs†		
Current (1–114 d)	1.7 (1.5–1.9)	1.2 (1.0–1.4)
Recent (115–180 d)	1.6 (1.2–2.2)	1.2 (0.9–1.6)
Past (181–365 d)	1.7 (1.3–2.2)	1.2 (0.9–1.6)
Former (1–3.5 y)	1.4 (1.2–1.7)	1.0 (0.8–1.3)

*Adjusted for age, sex, calendar year, alcohol, chronic obstructive pulmonary disorder, diabetes, education, ischemic heart disease, marital status, obesity, and use of opioids.
†Before index date.

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!