PHC 6010

EPIDEMIOLOGY METHODS I

Measurement of Effects: Matched Studies

Hamisu Salihu, MD, PhD

Switching Gears…

Matching in Epidemiologic Studies

Measuring Effects: matching

- In many medical studies a group of cases, individuals with a disease under investigation, are compared with a group of controls, people without the disease but who are comparable in other respects.
- This happens in epidemiological case-control studies, where a possible risk factor is compared between cases and controls to investigate the nature of the disease.
- In both types of study, cases and controls are sometimes matches. This means that for every case there is a control who has the same (or closely similar) values of the matching variables. Matching may be by sex, age to within five years, ethnic group, etc. Sometimes there are two or more such controls for each case.
- Matching is particularly useful in small studies, where we might not have sufficient subjects to adjust for several variables at once.

Measuring Effects: matching

Reasons for matching

- To ensure that controls and cases are similar in variables which may be related to the variable we are studying but are not of interest in themselves.
- For example, in many epidemiological case-control studies age is an important predictor of exposure to the risk factor under investigation. There are strong cohort effects in variables such as cigarette smoking and diet.
- If we do not take age into account we may get spurious differences between cases and controls because, for example, cases are older than controls.
- Matching ensures that any difference between cases and controls cannot be a result of differences in the matching variables. However, we cannot then examine the effects of the matching variables. [ǂ]

Measuring Effects: matching

- At times matching is ignored in the analysis of the data.
- If the matching variables are important, this is inefficient.
- Matching variables, such as age and sex, may be strongly related to the variable of interest. If we allow for the matching in the analysis the variation due to these variables is removed.
- If the matching is ignored then the variability, which is related to the variation, may mask important differences.

Measuring Effects: matching

- For example, if we compare the mean blood pressure of subjects with a disease to that of their age matched controls, the variability in blood pressure, which is associated with its increase with age, will be part of the residual variance and will increase the standard error of the difference between the means.
- Instead, we should use the differences between individual matched cases and their controls. Appropriate simple methods include the paired *t* test for means, McNemar's test for proportions, and the sign test for ordinal data.

Measuring Effects: matching

- The simplest instance of matching is pair-matching, where the units considered in the study are not individual items but pairs of components with something in common (e.g., twins or siblings).
- Matching of twins or siblings could even control for several genetic and environmental characteristics shared by the matched individuals.

Measuring Effects: matching

- Let's have a brief introductory note about twins before we go on to demonstrate concrete examples of matching
- Twins are a multiple birth in which a mother gives birth to two babies from same pregnancy.
- May be either of same gender or different gender
- There are about 125 million human twins worldwide (circa 1.9% of the world's population)
- And just 10 million identical twins (genetically similar; same DNA). This represents 8% of all twins, and 0.2% of the world's population

Measuring Effects: matching

- Twins can be either monozygotic (genetically identical) or dizygotic (fraternal or genetically non-identical; fraternal because similar to brothers and sisters and the only difference is that these are born at the same time)
- Monozygotic twins (genetically identical twins) arise from the same fertilized egg. The same embryo cleaves itself. In other words, the embryo clones itself.
- Hence monozygotic twins are the same individual, one being the natural clone of the other.
- The DNA of monozygotic twins are 100% the same.

Measuring Effects: matching

- In about 1% of identical twins, the splitting occurs late enough to result in conjoined twins. Hence, conjoined twins are necessarily identical and monozygotic.
- Conjoined twins are therefore, the result of failure of natural process of cloning
- Dizygotic twins (fraternal twins; non-identical twins): Usually occur when two fertilized eggs (hence, two different individuals; eggs = different; spermatozoa different) are implanted in the uterine wall at the same time.
- Both eggs are independently fertilized.
- Hence, the genetic material = different.

Measuring Effects: matching

- From the above discussion it is now clear that twin pairs are a good example of natural matching.
- But what are the likely patterns of matching in twins?
- Individuals could differ fundamentally based on genetic and environmental influences. Based on this consideration, we could examine matching of twins as given in the next slide.

Measuring Effects: matching

Type of twining	Environment	Genetic
Monozygotic	Same	Always Same
Monozygotic	Different (e.g, one adopted)	Always Same
Dizygotic	Same	Always different
Dizygotic	Different (e.g, one adopted)	Always different

Measuring Effects: matching

- From the table, we could glean that there are always two sources of variation that could occur for dizygotic twins (environmental and genetic) whereas, for monozygotic twins, there could only be one (environmental)
- This means then that studies that use dizygotic twins as study subjects could give us information on the effects of both environmental and genetic influences on outcomes of interest.
- On the other hand, studies that use monozygotic twins will give us information on the influences of environmental factors alone because the genetic component is same (controlled and non-variant or no difference).

Measuring Effects: matching

- Matched pairs may arise from a cross-sectional, cohort or case-control design
- In a typical cohort study matching is performed at baseline by equating certain characteristics of interest between the exposed and unexposed groups. In a one-to-one match, each pair will consist of an exposed and an unexposed member, and the occurrence of disease in one or both members of a pair is noted.
- In a case-control study, cases (diseased) are matched to controls (non-diseased). In a one-to-one match, each pair will consist of a case and a control and the presence of the exposure in one or both members of a pair is noted.

Measuring Effects: matching

Unexposed pair

Exposed pair	cases	non-cases
cases	a	b
non-cases	c	d

2*2 table for a matched cohort study design

Measuring effects: Matching

Control

Case	E	$\bar{E}$
E	a	b
$\bar{E}$	c	d

2*2 table for a matched case-control study design

Measuring effects: Matching

- In the above example of a matched case-control study design, there are four possible exposure combinations for each pair:
  - Both case and control are exposed
  - Case exposed and control non-exposed
  - Case unexposed and control exposed
  - Both case and control unexposed
- Denoting the numbers of outcomes in these four categories by n_{11},n_{10}, n_{01}, and n_{00}, respectively, the results of the study may thus be presented in a 2*2 table as shown in the next slide

Measuring effects: Matching

Control

Case	E	$\bar{E}$
E	n_{11}	n_{10}
$\bar{E}$	n_{01}	n_{00}

2*2 table for a matched case-control study design

Measuring effects: Matching

- In the table above: the n_{11} cases and controls are concordant for exposure level because members of any pair have the same level of exposure
- Similarly, the n_{00} cases and controls are concordant because members of any pair do not have the exposure of interest.

Measuring effects: Matching

- On the other hand, consider the other two cells:
- the n_{10} cases and controls are discordant for exposure level because members of any pair differ in exposure level. In this cell, all cases in a pair have the exposure of interest while all controls in a pair are unexposed
- Similarly, the n_{01} cases and controls are discordant for exposure level because members of any pair differ in exposure level. In this cell, all controls in a pair have the exposure of interest while all cases in a pair are unexposed

Measuring effects: Matching

- Now, for two individuals to be different their characteristics based on which they are being measured must be different, else they do not differ on those characteristics.Ŧ
- One example: Two populations A and B being compared on the "Richness index" based on the number of cars, homes and boats.
- Population A has 4 homes, 2 cars and 5 boats. Population B on the other hand has 10 homes, 6 cars and 5 boats.
- Do you think we gain any information by comparing the two populations based on the number of boats they have? Do you think it helps us to distinguish between the two populations based on richness. Absolutely not. In fact, that info is not useful because they cancel out on that basis.

Measuring effects: Matching

- In a similar token to the above analogy, an odds ratio cannot be estimated from a study in which all the subjects have the same exposure level. It is like an empty flat rectangle with a single chamber.
- Therefore, the matched pairs in which the case and control are either both exposed or both unexposed, contribute no information about the odds ratio. We could therefore, ignore them when making comparisons and trying to determine differences via a summary measure like the odds ratio.

Measuring effects: Matching

- It follows then that the relevant information is confined to the discordant pairs, namely (n_{01}, n_{10}). And we could ignore the concordant pairs (n_{11}, n_{00}).
- Let's examine further the members of the discordant pairs since they are the ones that will provide us with relevant information

Measuring effects: Matching

- The cell n_{10} contains information that describes the presence of an exposed case and an unexposed control
- In other words the discordant cell n_{10} expresses the frequency of occurrence of an unexposed control given that the case is exposed.
- In a similar fashion, the discordant cell n_{01} expresses in magnitude the occurrence of an unexposed case given that the control is exposed.
- How else could we describe the discordant cell n_{10} n_{01} in terms of probability? The next slide teaches you how.

Measuring effects: Matching

- In previous lectures, we have denoted p_1 as the probability of a case being exposed. Hence, the probability of the case being unexposed will be $1- p_1$
- Let also p_0 represent the probability of the control being exposed. Similarly then the probability of the control being unexposed will be $1- p_0$
- Using the law of conditional probability, what is the probability of a case being exposed and a control being unexposed? This is given by $p_1(1- p_0)$

Measuring effects: Matching

- And what will be the probability of a control being exposed given that a case is unexposed? This is given by $p_0(1- p_1)$
- Hence, in probability terms, $p_1(1- p_0)$ and $p_0(1- p_1)$ are actually equivalent to the frequency information in n_{10} n_{01} respectively.
- Let's make this clearer using the more familiar designations of a 2*2 table

Measuring effects: Matching

Control

Case		E	Ē
	E	a	b
	Ē	c	d

p1(1- p0) = b

p0(1- p1) = c

And we know that p1(1- p0) / p0(1- p1) = the odds ratio

Hence, **p1(1- p0) / p0(1- p1)** = b/c; which is the odds ratio

Measuring effects: Matching

- Thus, the odds ratio in a matched case-control study may be estimated by taking the ratio of the number of positive to negative discordant pairs, namely,

$$OR = b/c$$

And the formula for the standard error of the natural logarithm of the odds ratio is given by

$$SE [\ln (OR)] = \sqrt{(1/b+1/c)}$$

Measuring effects: Matching

The confidence interval for the odds ratio could therefore, be expressed thus:

$$e^{\ln OR \pm z * SE(\ln OR)}$$

Where *e* is the base on the natural logarithms (*e* =2.71828), z is a Standard Normal Deviate corresponding to the desired level of confidence (z =1.645 for 90% confidence level, z = 1.96 for 95% confidence level, and z = 2.576 for 99% confidence level)

Measuring effects: Matching

- The null hypothesis of no association between the exposure and outcome may be tested using a chi-squared statistic (with 1 degree of freedom) as suggested by McNemar (1947). The formula is

$$\chi^2 = (b-c)^2/(b+c)$$