School of Optometry

Follow-up (Cohort) Studies

Dr. John Waterbor
UAB School of Public Health
October 25, 2010

Outline

- General Description
- Common Closing Date; Ending Follow-Up On A Subject
- Validity and Precision Considerations
- 3 Varieties of Follow-Up Studies
- Limitations of this design
- Choice of comparison group
- Basic Data Analysis (using Incidence Rate, I)
- Basic Data Analysis (using Cumulative Incidence, CI)
- Logistics of Evaluating "Dose-response"
- Interpretation of "Dose-response"
- Categorization of Data and the "ile" Approach
- Mantel Extension Test of ☞ For Linear Trend

Of the ***observational*** epidemiologic study designs, by far the most common are the ***Follow-Up Study*** (or, ***Cohort Study***) and the ***Case-Control Study***.

Note that ***neither design is superior to the other***. They each have their ***strengths*** and ***weaknesses***, but either can be done very well or very poorly.

General Description of a Follow-up Study

1) Individuals whose ***exposure status*** to ***E*** is known (***E+*** vs. ***E-***, or level of ***E***), ***and*** whose ***disease D status is known to be negative*** (i.e., they are ***free of disease D),*** are followed from "***time zero***" to "***time t***" for occurrence (***incidence***) of ***D*** or for death (***mortality***) from that ***D***.
2) For each ***exposure group*** (***E+*** vs. ***E-,*** or level of ***E***) the ***I***, ***CI*** or ***M*** is calculated.
3) The measures of disease frequency are then compared via the appropriate ***Relative Risk*** (***RR***) measure.
4) Note that the incidence of (or mortality from) ***more than one*** disease ($D_1, D_2, D_3, \ldots$) can be measured in a given study.

The Common Closing Date

The ***Common Closing Date (CCD)*** (or, time "***t***") is the date established by the investigator, on which the collection of outcome data (***incident cases*** or ***deaths***) ***stops.*** Outcomes occurring ***after*** the ***CCD*** "don't count" because the "books are closed" on the ***CCD.***

But . . . not all subjects are followed to the CCD. Follow-up ends ***earlier*** than the ***CCD*** for a subject as soon as he or she is ***no longer at risk*** for disease ***D*** or is ***lost to follow-up***.

End Follow-up on any Subject the Moment that He or She . . .

1) ***Gets the disease*** of interest (or, in a ***mortality*** study, ***dies*** from the disease) – a person is not "at risk" for a disease that he or she already has

2) ***Dies*** (from any cause) - dead people cannot be "followed" because they are no longer at risk of becoming an incident case!

3) ***Becomes*** "***lost to follow-up***" for any reason (refused, moved away, etc.) making his or her disease status ***D+ or D- unascertainable*** from that time forward so that he or she is effectively no longer "at risk" (in our data set)

4) ***Biologically*** leaves the at-risk pool (e.g., a woman who receives a hysterectomy cannot get uterine cancer)

Differences of a Follow-up Study from an Experimental Study

- The investigator ***does not*** control ***E status*** in a follow-up study as he or she does in an ***experiment.***
- There is ***no randomization*** of people to one study group or the other (***E+*** or ***E-***) in a follow-up study . . . so that ***bias*** and ***confounding*** are generally more of a problem than they are in an experiment.

Validity of an Epidemiologic Study

The ***validity*** of an epidemiologic study is the extent to which the ***associations*** found in the study of ***samples*** reflect the "***unknowable***" ***associations*** that are true in the ***populations***.

If a study is ***valid*** this means that the ***RR*** point estimate (***sample statistic***) reflects ***"the truth"*** about the ***E – D*** relationship in the ***population (population parameter)***.

We will discuss validity issues in detail later. For now . . . how do we ***maximize validity*** in a follow-up study?

How to Do a Valid Follow-up Study

1) Avoid or minimize ***selection bias***
2) Avoid or minimize ***information bias***
3) Avoid or minimize ***confounding*** by means of:
   a) Subject characteristic restriction in the ***design*** phase (a ***limitation*** being that the results may not be ***generalizable*** to people dissimilar from your subjects)
   b) Standardization (or, ***adjustment***) in the ***analysis*** phase (a ***limitation*** being that adjustment will not work if there are ***extreme differences*** in the distribution of the confounding variable in the two populations compared)
   c) Both ***a)*** and ***b)*** are necessary in most studies

How to do a Precise Follow-up Study

This one is a little easier . . . ***study "enough" people*** (***E+*** and ***E-***, or a range of ***E***'s) so that the ***95% confidence interval*** around the ***RR point estimate*** is narrow enough that it does ***not*** include the ***null value 1***.

Having a ***95% CI that excludes 1*** means that our study was "worth doing" in a statistical sense, because our study was big enough to effectively ***exclude*** the possibility (with 95% certainty!) of ***the null*** ("***no association***").

Varieties of Follow-up Studies

a) ***Prospective follow-up studies***: Follow-up begins "today" and continues into the future

b) ***Retrospective follow-up studies***: Follow-up begins in the past and ends in past

c) ***Ambidirectional follow-up studies***: Follow-up begins in the past and continues into the future (i.e., a combination of ***both*** varieties above)

Note that for all varieties, follow-up is ***forward*** in time, from ***E+*** or ***E-*** to ***D+*** or ***D-***.

Limitations of Follow-Up Studies

1) ***Not*** an ideal design for studying a ***rare*** disease, because in order to observe ***any*** incident cases you must observe a ***great many*** people and/or follow them for a ***long*** period of time for generation of ***lots of person-years*** . . . otherwise you will find ***very few or perhaps no cases***!

2) ***Prospective follow-up*** studies require the investigator to wait ***a long time*** (perhaps, an ***induction period***!) for the disease to occur – for chronic diseases this could be years or decades – so that ***results are not readily available.***

3) ***Retrospective follow-up*** studies require "historical" documentation (or circumstantial evidence, or inferences, or guesses?) about ***exposure level*** at ***time-zero, time-zero*** being sometime in the past, perhaps the distant past.

Uncertainty in Choosing the Most Appropriate "Unexposed" Group

When exposure is measured ***dichotomously*** (***E+*** vs. ***E-***) it is usually easy to identify the ***exposed*** group (e.g., heavy smokers) because this is the group of people that have come to our attention to study.

Often more ***problematic*** is choice of the so-called "***unexposed***" group, ideally people like the exposed people in all relevant ways ***other than for the exposure.***

In some situations the "unexposed" group we desire must be ***totally unexposed*** (***exposure-free***) whereas in other situations people with ***minimal*** but non-zero exposure make acceptable members of the "unexposed" group.

Using an External Comparison Group for the "Unexposed" Group

An ***external*** comparison group such as the U.S. population can be used if we can ***reasonably assume*** that the vast majority of the population is ***unexposed*** or ***minimally exposed***.

Having a very ***large*** comparison group, as one might find externally, is good for statistical reasons.

The external comparison group may be acceptable or even preferable for studies of ***uncommon*** exposures such as most industrial chemicals.

Limitations of an External Comparison Group

A ***limitation*** of choosing an ***external*** "unexposed" comparison group is that too many of them may actually be ***exposed***!

Would you choose the ***U.S. population*** to serve as your "unexposed" comparison group in a study of the health effects of cigarette smoking when you know that the ***prevalence*** of cigarette smoking in the U.S. population is ***high***, about 23%? ***NO!*** Bad choice.

When your so-called ***unexposed*** group is truly ***partially exposed***, does this drive your ***RR*** point estimate ***toward 1*** ("***bias toward the null***") or ***away from 1*** ("***bias away from the null***")?

Another ***limitation*** of the external comparison group is the possible lack of ***comparability*** of the ***unexposed*** to the ***exposed*** with respect to other important factors (***confounders***) that we need to control for.

Using an Internal Comparison Group for the Unexposed Group

An ***internal*** comparison group (people "internal" to the study) is usually preferred to an ***external*** comparison group.

Choose people who live in the same community or who work in the same factory as do the ***exposed*** people, but who we ***know*** are ***unexposed*** or who we ***believe*** are ***unexposed***, often based on circumstantial evidence.

A ***limitation*** of the internal comparison group is that the people chosen may not ***truly*** be unexposed if they live, work, or otherwise associate with people in the ***exposed*** group.

If the so-called ***unexposed*** group was truly ***partially exposed***, would this drive our ***RR*** point estimate ***toward*** or ***away*** from the null?

Example of Data Layout for an Incidence Rate–Type (or, Incidence Density–Type) Follow-up Study

The following data were gathered over a 3-year follow-up period:

	E^+	E^-
cases	**24**	**8**
p-yrs	**1152**	**811**

Calculation of the Incidence Rate Ratio, IR

- ***IR*** **= I_{E+} / I_{E-} = (24 / 1152) / (8 / 811) = 2.1**
- Interpretation: the incidence rate of this disease is ***2.1 times higher*** among the ***exposed*** than among the ***unexposed***

Example of Data Layout for a Cumulative Incidence – Type Follow-up Study

Example: These are the same subjects followed for 3 years, but here is a different analysis, using a ***2 x 2*** or ***fourfold*** table:

	E+	E-	
D+	24	8	32
D-	376	292	668
	400	300	700

Calculation of the Cumulative Incidence Ratio, CIR

- ***CIR*** = CI_{E+} / CI_{E-} = (24 / 400) / (8 / 300) = **2.25**

- Interpretation: the cumulative incidence of this disease is ***2.25 times higher*** among the ***exposed*** than among the ***unexposed***

Comparison of Two Analyses of the Same Data Set

Although the same data set was analyzed twice, we did ***not*** get exactly the same result:

IR = **2.1** and ***CIR*** = **2.25**

In both calculations the ***numerators*** "24" and "8" were the same but the ***denominators*** were different.

Person-years of follow-up time (the sum of the ***each individual's*** "exposure opportunity") are the ***denominators*** of each ***I*** calculation

Populations at risk at time-zero are the ***denominators*** of each ***CI*** calculation. Each ***CI*** does ***not*** take into account anyone leaving or entering the population as follow-up proceeds.

Evaluation of Dose-Response

It is better to measure ***E*** according to ***level (dose)*** rather than classifying ***E dichotomously*** as simply ***E+*** or ***E-*** (thereby ***degrading*** the data).

"***Dose-response***" can be evaluated if ***E*** is measured (or classified) on at least ***3 levels.*** As ***dose*** of ***E*** increases, does ***incidence*** of ***D*** increase, decrease, or stay the same? If ***D*** increases or decreases, we have evidence of ***"dose-response"*** . . . which may or not reach statistical significance.

The argument for ***causation*** is ***enhanced*** if ***dose-response*** (one of ***Hill's Criteria***) is present because the alternative explanation of ***confounding*** for the non-null ***RR*** that you found, is less likely.

But if ***dose-response*** is ***absent***, causation is ***still possible.*** There could be a ***threshold effect*** whereby a certain dose raises the risk and causes disease, but a higher dose does not raise the risk any further.

Calculation of Dose & Grouping of Subjects into Exposure Categories

"***Dose***" is based on ***duration*** of exposure (months, years, etc.) and also on ***intensity*** of exposure (concentration or amount of exposure on a typical day in a typical year).

For example, the "***dose***" of ***cigarette smoking*** would be based on ***number of years smoked***, and the ***average number of cigarettes smoked*** per day during that time.

Once each subject's ***dose*** has been measured or estimated, subjects are ***grouped*** into ***exposure categories*** so that ***I*** or ***CI*** can be measured in each category. How should the ***exposure categories*** be formed?

Use of "Natural" Exposure Level to Form Exposure Categories

"***Natural***" exposure categories can be used if the ***exposure data*** fall into a dozen or fewer categories and the data set is sufficiently ***large***.

For example, if in our data set the number of ***cups of coffee*** consumed per day ranges from ***0 to 10 cups***, these "natural" 11 categories could be used for ***dose-response*** analysis if the data set is ***large.***

Caution! for statistical reasons we should ***avoid using so many categories that there are zero, or very few, subjects in some of them,*** making some ***I*** or ***CI*** estimates ***0*** or very "***unstable***" (imprecise).

Inventing Categories When there are Too Many Natural Categories

In situations when the raw data on exposure are recorded on a virtually ***continuous*** scale (e.g., BP, cholesterol level, grams of food intake) the investigator must ***group*** the subjects into a small number of ***E levels*** (often, ***3, 4 or 5***).

The "ile" method: divide your subjects into categories of approximately ***uniform*** size to form ***tertiles (3 levels), quartiles (4 levels) or quintiles (5 levels)*** so that there are 3, 4 or 5 levels on which to judge ***dose-response.***

Choosing the Number of Categories of Exposure Level to Invent

If your data set is ***small***, use few categories (but you need ***at least 3 to evaluate dose-response***).

Larger data sets will allow a finer assessment of ***dose*** (or, "***exposure gradient***") so that 5 or more categories may be used.

Often, additional categories ***beyond 5*** do not help in the evaluation of ***dose-response*** because the biological effects of very similar doses are negligible.

How Convincing is the Evidence of Dose-Response in this Data Set?

E level (cigs/day)		People	Cases	CI	CIR=CI_{E+}/CI_{E-}
1.	0	50	1	.020	1.0 (Referent)
2.	1-9	80	2	.025	1.25
3.	10-19	70	3	.043	2.15
4.	20-29	90	5	.056	2.8
5.	30-39	40	4	.100	5.0
6.	40+	40	5	.125	6.25

In this table, the E levels, People, and Cases data are ***given*** to you.

The ***CI***s are calculated as Cases / People and the ***CIR***s are calculated by dividing each ***CI*** by the ***common referent CI*** corresponding to E level 1 (0 cigs/day).

Be sure that you understand how the ***CI***'s and ***CIR***'s are calculated!

Graphical and Statistical Evaluation of Dose-Response

Draw a ***graph*** of ***CIR*** versus exposure level 1-6.

Regression techniques can be used to establish the "***best***" line that runs through the 6 data points.

The ***slope*** of the best regression line is statistically compared to the ***slope*** of a horizontal line (zero). The horizontal line indicates ***no association*** between ***dose*** and incidence. If the regression line ***significantly differs*** from the horizontal line, there is evidence of a ***positive association*** (***I*** or ***CI*** rises with increasing ***E***) or an ***inverse association*** (***I*** or ***CI*** falls with increasing ***E***).

Mantel Extension Test, χ^2 for Linear Trend

For ***stratified*** data (incidence according to ***dose***), the "***Mantel Extension Test of χ^2 for a Linear Trend***" (covered in lab) is used to assess whether an observed trend (***dose-response*** relationship) as seen graphically is ***statistically significant***.

A ***limitation*** of this statistical test is that it comes out ***"significant"*** for ***any*** scatter of points that looks non-random (non-null). A significant result does not necessarily mean that the dose-response is ***linear***; it simply means there is some sort of relationship (linear, curvilinear, other) between ***RR*** and dose, ***not independence.***

--JW (follow08)