

Practicals in Nested and Matched Case-Control studies

Department of Biostatistics, University of Copenhagen

<http://BendixCarstensen.com/AdvEpi>

November 2015

Exercise version

Compiled Friday 30th October, 2015, 20:32
from: /home/bendix/teach/Epi/KU-cc/pracs/pracs.tex

Bendix Carstensen Steno Diabetes Center, Gentofte, Denmark
& Department of Biostatistics, University of Copenhagen
bxc@steno.dk
<http://BendixCarstensen.com>

Contents

1	Case-control study of anal-cancer.	1
2	<i>Salmonella</i> Typhimurium outbreak at Fyn, autumn 1996	1

1 Case-control study of anal-cancer.

The following is based on the paper:

Morten Frisch, Jørgen Olsen & Mads Melbye: Malignancies that Occur before and after Anal Cancer: Clues to their Etiology, *American Journal of Epidemiology*, **140**, 1, pp 12–19, 1994.

which is available at <http://BendixCarstensen.com/AdvEpi>

The article describes a case-control study and a follow-up study, both based on material from the Danish Cancer Registry. The following questions mainly relate to the case-control study.

1. How are cases and controls defined?
2. In which cohort is the case-control study nested?
3. What rate-ratios are estimated in the case-control study? That is what are the exposures in question, and what is the outcome.
4. Describe how the corresponding complete-follow-up study could be constructed.
5. Specifically, what are the timescales for the anal cancer incidence considered in the case-control study?
6. Discuss the pros and cons of using a nested case-control design, respectively a complete follow-up study.

2 *Salmonella* Typhimurium outbreak at Fyn, autumn 1996

In the autumn 1996 an unusually large number of cases of infection with *S. Typhimurium* was reported in Fyn county in Denmark.

The Danish Zoonosis Center conducted a case-control study to find the causes of the infection. It was conducted by telephone-interviews with verified cases and individually matched controls.

In the file `fyn.sas` is a SAS-program that read in data from the study. There are the following variables available:

`inddato`: date of interview

`intw`: interviewer-id.

`par`: stratum (matched set).

`lbnr`: serial number.

`pk`: case / control status: P — case (patient), K — control.

`foddat`: birth date.

sex: sex: M — males, K — females.

The following variables are 0-1 variables, coded 1 for exposed, and 0 for unexposed. Note that there are also missing values (“.”) in the dataset:

udland: been abroad within the last two weeks.

okskod: has eaten beef.

svinkod: has eaten pork.

kalvkod: has eaten veal.

fjerkod: has eaten poultry.

kodpaal: has eaten meat on bread.

paal8: has eaten liver-paste.

gronsag: has eaten vegetables.

frugt: has eaten fruit.

eggret: has eaten egg-dishes.

slagt7: has eaten meat bought in a store supplied by slaughter house no. 7.

1. Read in data and make a table of the number of cases and controls in each matched set.
2. Screen the variables to see if they have an association with the outcome *Salmonella* Typhimurium infection.

Consider in particular how the missing values should be treated — note that `proc phreg/proc logistic` just ignores observations from the dataset where one of the variables in the model is missing.

3. Construct a final model for the material. In particular check if there is an interaction between any of the variables, and choose a suitable parametrization of the interaction. Write a conclusion of the analysis, preferably with a tabular exposition of the results.