

A short story about Danish register-based research – and about Statalist

Svend Juul

Stata User Group meeting
Aarhus, September 2014

1

A PhD student is working on a study of:
Health problems among children born by mothers with type 1 diabetes. A clinical database includes 1,300 such children (index children).

Statistics Denmark matched by gender and birth date
100 control children to each index child.

Follow-up to the age of 13-20 years.

2

By law, Statistics Denmark does not disclose identifiable information about individuals,

but it offers to perform analyses including register-based information about individuals.

The researcher submits a do-file in batch mode.
Within hours, she receives a mail with the output log – after it has been checked that it does not allow identifying individuals.

There are also other safeguards to prevent abuse of information.

3

Statistics Denmark drew the matched controls from the Civil Register

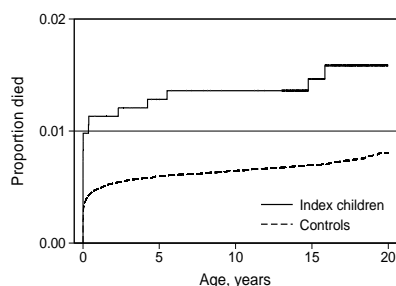
and made quite a bit of register-based information available to the study:

- Deaths – with diagnosis
- Hospitalizations and outpatient contacts – with diagnosis
- Medications – with ATC code
- Parental education

Among the health outcomes to be studied are mortality, frequency of hospital admissions, and frequency of medications.

4

Mortality



Number at risk					
Index children	1,326	1,287	1,279	928	88
Controls	131,884	127,709	125,559	91,036	8,353

5

The data on hospital admissions are clustered at two levels:

Admission data for each child constitute a cluster:
→ use a random effects model.

Each index child and its controls constitute a cluster:
→ use robust variance estimates.

Over-dispersion:
→ use negative binomial regression

The Stata command for this mixed-effects model is **menbreg**.

6

menbreg is computationally intensive, and we want some 600 analyses (six age groups, 20 diagnostic groups, etc.).

We reduced the number of controls to ten per index child, but some analyses would still take several hours, so we could not see the light at the end...

Is there a way to automatically stop an analysis that takes "too long time" and proceed with the next analysis?

(Some of the SUG participants will know how to do that, but I didn't know at the time.)

7

I sent the question to Statalist, and within five minutes I had two good suggestions from Clyde Schechter and Richard Williams:

Use the **iterate()** option;

Study **help maximize**.

It works, and the analyses are proceeding.

8

```
foreach DX of varlist dx1-dx20 {
  menbreg `DX' group, exposure(yrs) || id:, irr iter(100)
  display "Iterations: " e(ic) " Converged: " e(converged)
  if e(converged)==0 {
    nbreg `DX' group, exposure(yrs) vce(cluster id) irr
  }
}
```

9

Conclusion:

RTFM!

- but sometimes it helps to ask a friend.

Statalist is a friend.

10