NORTH DENMARK REGION

Self-Controlled Case Series

Theory and a Stata command -sccsdta-

Niels Henrik Bruun,
Research Data and Biostatistics,
Aalborg University Hospital

NORTH DENMARK REGION

Study designs

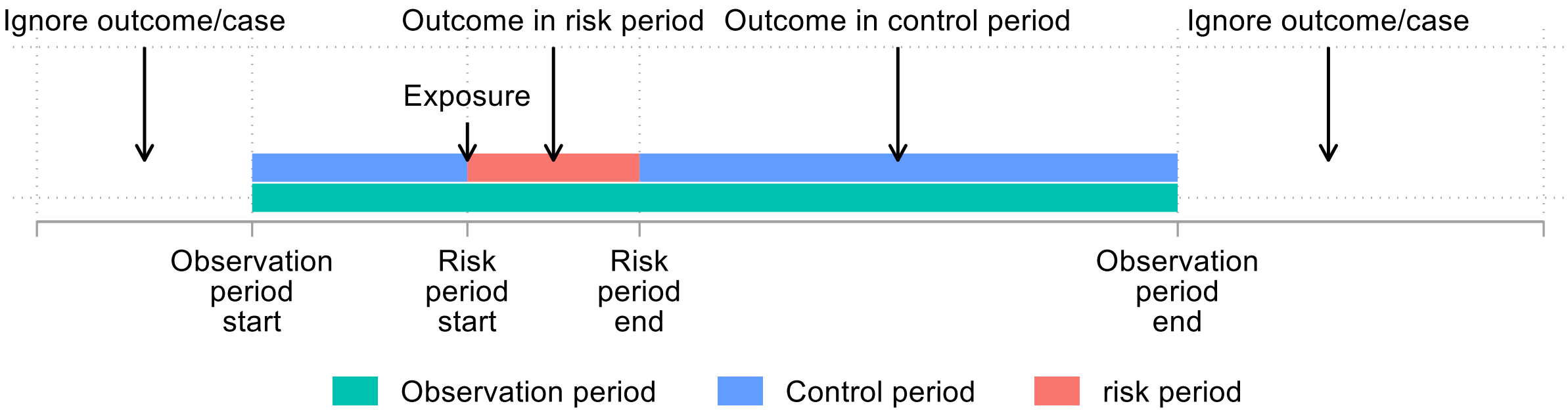
	With and without exposure	With and without outcome
Compare people	Cohort study	Case-control study
Compare periods	Self-Controlled Case study	Case-crossover study

Table 1 in Iwagami and Takeuchi (2021)

- Time-varying exposure -> an outcome event
  - E.g., meningitis after vaccination
- The eligibles have both exposure and event
- See Whitaker et al. (2006) and Petersen, Douglas, and Whitaker (2016) for an introduction
- Also see Farrington, Whitaker, and Weldeselassie (2018)

NORTH DENMARK REGION

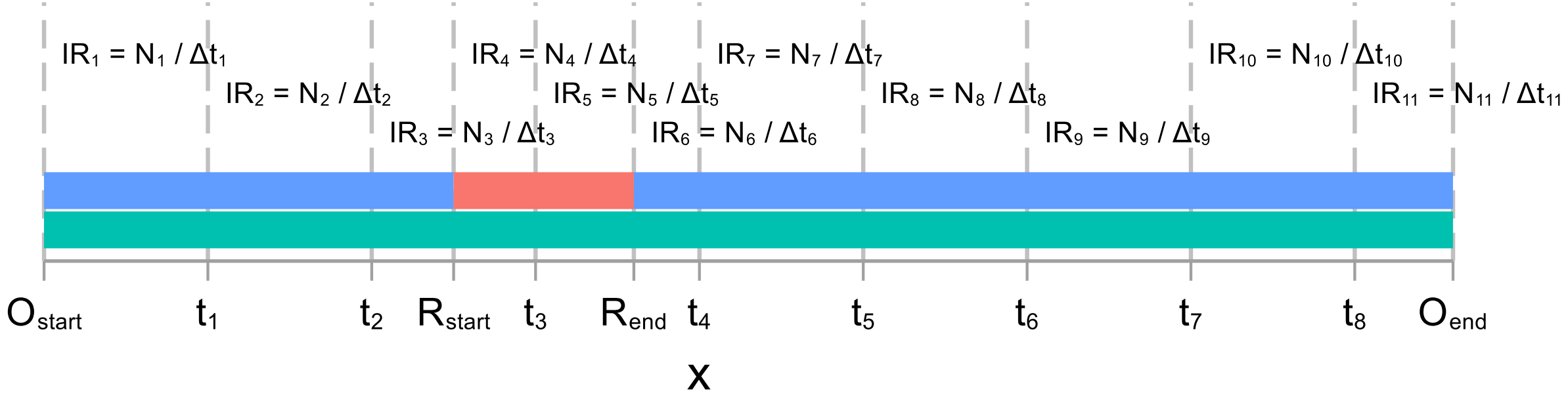
Observation, control, and risk periods



For each case

NORTH DENMARK REGION

Adding time effects during the observation period



Observation period Control period risk period

A set of sub intervals from a non-homogeneous Poisson process for each case

NORTH DENMARK REGION

The Model

- Three multiplicative elements (independence):

$$case \times time \times risk$$

- Self-controlled by fixed-effect regression (**eliminate the *case* element**)
  - using -xtpoisson-
  - or Poisson Pseudo-Likelihood regression with High-Dimensional Fixed Effects (PPMLHDFE)
- All cases has the same time effects during the observation period
- All cases the same risk effects during the observation period
- Estimates are:
  - relative incidence rates
  - adjusted for time

NORTH DENMARK REGION

Comparison to other designs

In 1998, three studies investigated whether the MMR vaccine may cause autism.

Study	Sample size	IRR	(95% CI)
Cohort	537303 children, 316 cases	0.92	(0.68, 1.24)
Case–control	1294 cases, 4469 controls	0.86	(0.68, 1.09)
Self-Controlled Case Series	357 cases	0.88	(0.40, 1.95)

Table 6 in Farrington and Whitaker (2006).

- **Few observations**
- estimates similar to Cohort and Case-control
- confidence interval wider (possibly due to a long post period)
- **No problem with unknown or known confounders**

NORTH DENMARK REGION

Violation of assumptions, impact, and solutions

- Event temporarily changes probability of exposure
  - Include a pre-exposure period (See sensitivity)
- No exposure after event
  - Begin observation period at exposure
  - For single or multiple exposures, see Farrington, Whitaker, and Hocine (2009)
- Events increases P(death)
  - Sensitivity excluding cases with deaths due to event
  - adjust for bias, Farrington et al. (2011)

Table 1 in Petersen, Douglas, and Whitaker (2016)

NORTH DENMARK REGION

Description, -sccsdta- (not complete)

Syntax: sccsdta varlist(event time and exposure time) [, options]

Required options

- enter({varname|number}) - Varying observation period starting point
- riskpoints(numlist) - risk period endpoints relative to exposure time

Optional (not complete)

- exit(varname) - Varying observation period endpoint
- timepoints(numlist)
  - The time interval end points - are relative to enter unless option absolutetimepoints is set
- absolutetimepoints - Whether timepoints are relative to the start of the observation period or fixed within
- eventtimes - Use the event times as absolute time period end points.
- knots(3-7) - Number of knots to use when generate restricted cubic splines on the event time period endpoints
  - Splines are used when there are many event times such that xtpoisson may not converge
  - knots(3-7) also sets eventtimes and absolutetimepoints
- Preserve - Preserve the original dataset

NORTH DENMARK REGION

Dataset description, example

- Study the association between oral polio vaccine (OPV) and intussusception in infants aged 28 to 365 days.
  - a dangerous condition where the bowel folds in on itself causing an obstruction of the intestine
- 218 episodes
  - with 11 infants having more than one episode

- Considered two-week risk periods after taking the OPV
  - 14-27 and 28-41 days from the vaccination day

NORTH DENMARK REGION

The dataset

```
. use eventday agep3 cutp? using ///
        "http://fmwww.bc.edu/RePEc/bocode/s/sccsdta%20intuss.dta", clear
. list in 1/5
     +-----------------------------------+
     | eventday   agep3   cutp1   cutp2 |
     |-----------------------------------|
  1. |      156     114      27     365 |
  2. |      221     147      92     365 |
  3. |      107     160      27     365 |
  4. |      197     160      27     365 |
  5. |      148     138      27     365 |
     +-----------------------------------+
```

```
. sumat *, statistics(n mean sd min max) ///
  decimals((0,2,2,0))

-----------------------------------------
             n    mean      sd  min   max
-----------------------------------------
eventday   218  191.14   73.70   36   357
agep3      218  161.19  108.52  109  1234
cutp1      218   31.35   24.21   27   212
cutp2      218  362.28   19.97  165   365
-----------------------------------------
```

NORTH DENMARK REGION

Using -sccsdta- and relative timepoints

```
. sccsdta eventday agep3, enter(cutp1) exit(cutp2) riskpoints(13 27 41) ///
        timepoints(0(90)270 366)
The SCCS regression summary table

                                    |        IRR        [95%        CI]    P(IRR=1)
------------------------------------+-----------------------------------------------
At risk                             |
                               ctrl |      1.000          .          .          .
                           ]13; 27] |      2.087      1.300      3.350      0.002
                           ]27; 41] |      1.271      0.716      2.256      0.413
------------------------------------+-----------------------------------------------
Time group                          |
                 ]enter+0; enter+90] |      1.000          .          .          .
              ]enter+90; enter+180] |      1.845      1.250      2.722      0.002
             ]enter+180; enter+270] |      1.570      1.064      2.318      0.023
             ]enter+270; enter+366] |      0.638      0.371      1.097      0.104
```

NORTH DENMARK REGION

The dataset after -sccsdta- for the first row

- _rowid - Row id in the original dataset
- _start - Interval start
- _stop - Interval end
- _nevents - Numbers of events in the interval
- _exgr - The risk group category of the interval
- _tmgr - The time group category of the interval
- _interval - The interval width

```
. list _* if _rowid == 1, noobs sepby(_tmgr) abbreviate(12)
  +-----------------------------------------------------------------------------+
  | _rowid   _start   _stop   _nevents      _exgr                    _tmgr   _interval |
  |-----------------------------------------------------------------------------|
  |      1       27     117          0       ctrl      ]enter+0; enter+90]          90 |
  |-----------------------------------------------------------------------------|
  |      1      117     127          0       ctrl    ]enter+90; enter+180]          10 |
  |      1      127     141          0   ]13; 27]    ]enter+90; enter+180]          14 |
  |      1      141     155          0   ]27; 41]    ]enter+90; enter+180]          14 |
  |      1      155     207          1       ctrl    ]enter+90; enter+180]          52 |
  |-----------------------------------------------------------------------------|
  |      1      207     297          0       ctrl   ]enter+180; enter+270]          90 |
  |-----------------------------------------------------------------------------|
  |      1      297     365          0       ctrl   ]enter+270; enter+366]          68 |
  +-----------------------------------------------------------------------------+
```

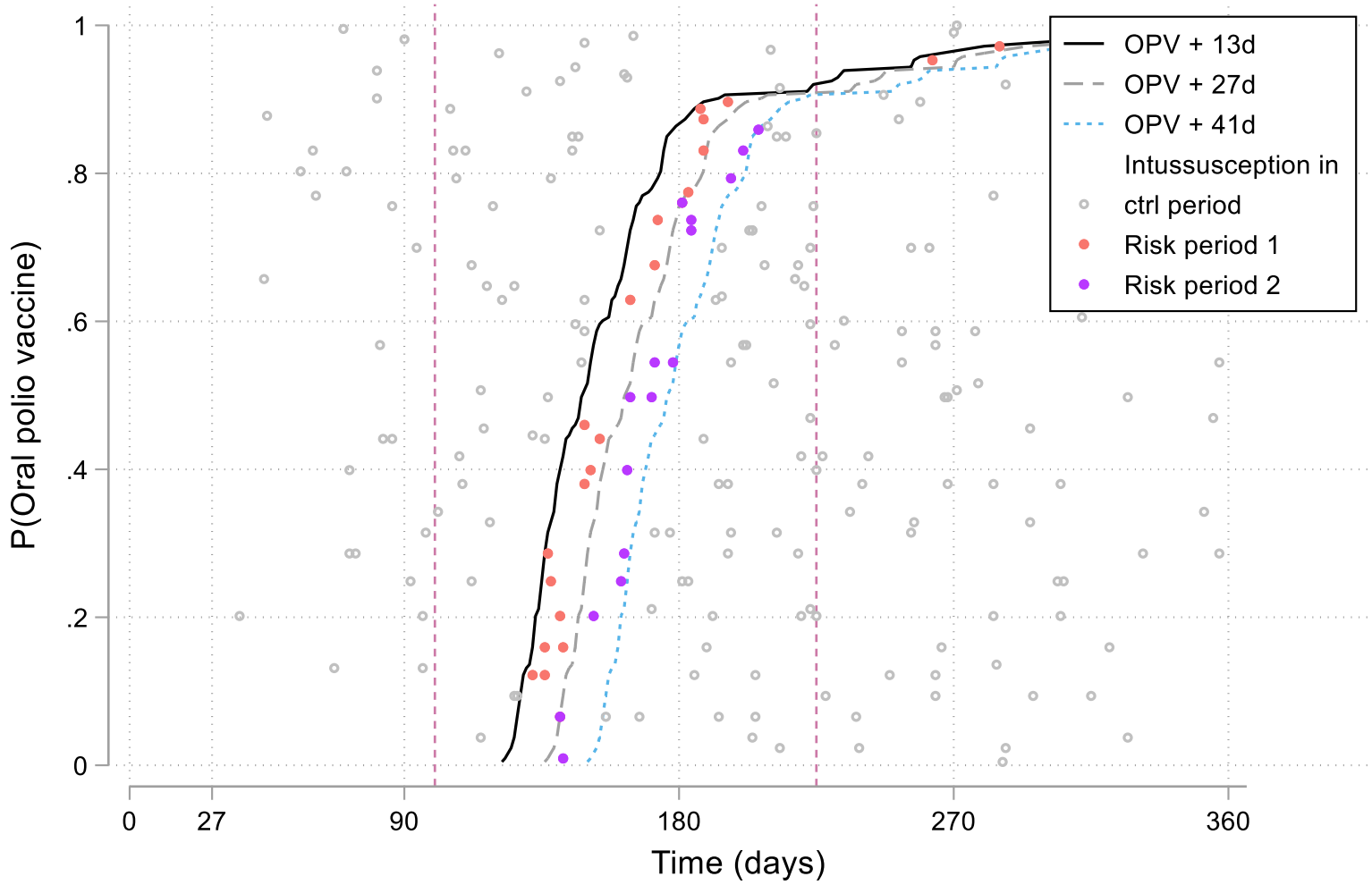
NORTH DENMARK REGION

Sufficient data for reproduction/meta analysis

```
. collapse (sum) _nevents _interval, by(_exgr _tmgr)
. list, noobs abbreviate(32) sepby(_exgr)
  +-----------------------------------------------------+
  |    _exgr                      _tmgr   _nevents   _interval |
  |-----------------------------------------------------|
  |     ctrl      ]enter+0; enter+90]         43       19508 |
  |     ctrl    ]enter+90; enter+180]         57       14384 |
  |     ctrl   ]enter+180; enter+270]         62       18565 |
  |     ctrl   ]enter+270; enter+366]         19       13866 |
  |-----------------------------------------------------|
  | ]13; 27]      ]enter+0; enter+90]          0          50 |
  | ]13; 27]    ]enter+90; enter+180]         21        2604 |
  | ]13; 27]   ]enter+180; enter+270]          2         183 |
  | ]13; 27]   ]enter+270; enter+366]          0          69 |
  |-----------------------------------------------------|
  | ]27; 41]      ]enter+0; enter+90]          0          62 |
  | ]27; 41]    ]enter+90; enter+180]         14        2560 |
  | ]27; 41]   ]enter+180; enter+270]          0         217 |
  | ]27; 41]   ]enter+270; enter+366]          0          73 |
  +-----------------------------------------------------+
```

NORTH DENMARK REGION

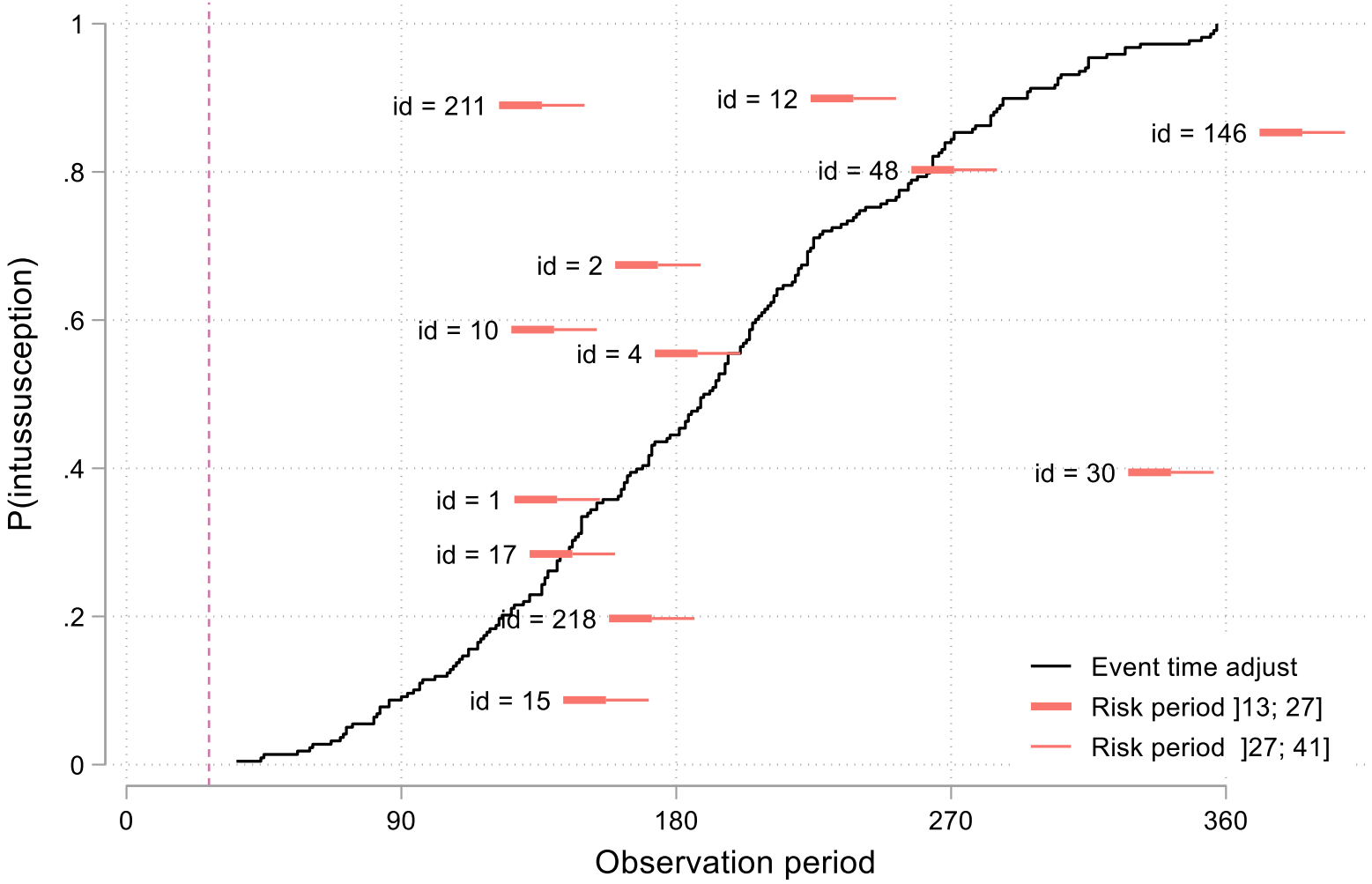
Every case has an oral polio vaccine (OPV) exposure



Most OPV are given within 100 and 225 days (See sensitivity)

NORTH DENMARK REGION

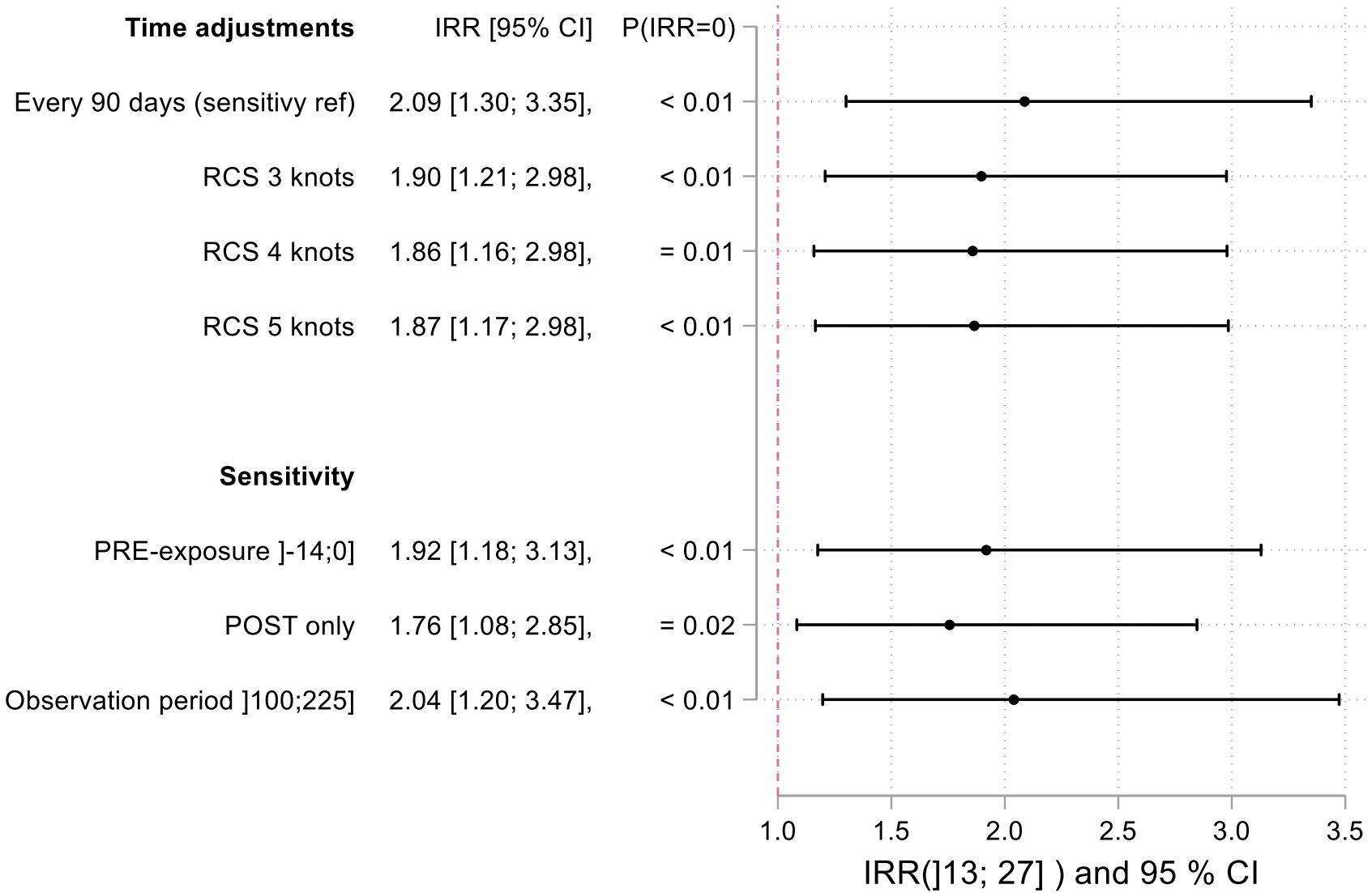
Every case has an event



Some risk periods and the probability of an event.

NORTH DENMARK REGION

Model validation summary graph



NORTH DENMARK REGION

-sccsdta- code for the model validation summary graph

Time adjustments

```
. sccsdta eventday agep3, enter(cutp1) exit(cutp2) ///
        riskpoints(13 27 41) timepoints(0(90)270 366) preserve
. sccsdta eventday agep3, enter(cutp1) exit(cutp2) ///
        riskpoints(13 27 41) nk(3) preserve
. sccsdta eventday agep3, enter(cutp1) exit(cutp2) ///
        riskpoints(13 27 41) nk(4) preserve
. sccsdta eventday agep3, enter(cutp1) exit(cutp2) ///
        riskpoints(13 27 41) nk(5) preserve
```

Sensitivity

```
. sccsdta eventday agep3, enter(cutp1) exit(cutp2) ///
        riskpoints(-14 0 13 27 41) timepoints(0(90)270 366) preserve
. sccsdta eventday agep3, enter(agep3) exit(cutp2) ///
        riskpoints(13 27 41) timepoints(0(90)270) preserve
. sccsdta eventday agep3, enter(100) ///
        riskpoints(13 27 41) timepoints(100(25)225) preserve absolutetimepoints
```

NORTH DENMARK REGION

References

Farrington, C. Paddy, Karim Anaya-Izquierdo, Heather J. Whitaker, Mounia N. Hocine, Ian Douglas, and Liam Smeeth. 2011. “Self-Controlled Case Series Analysis with Event-Dependent Observation Periods.” *Journal of the American Statistical Association* 106 (494): 417–26.

Farrington, C. Paddy, Heather J Whitaker, and Mounia N Hocine. 2009. “Case Series Analysis for Censored, Perturbed, or Curtailed Post-Event Exposures.” *Biostatistics (Oxford, England)* 10 (1): 3–16.

Farrington, C. P., and H. J. Whitaker. 2006. “Semiparametric Analysis of Case Series Data.” *Applied Statistics* 55 (5): 553–94.

Farrington, Paddy, Heather Whitaker, and Yonas Ghebremichael Weldeselassie. 2018. *Self-Controlled Case Series Studies: A Modelling Guide with R*. 2nd ed. Vol. 1. Chapman & Hall/Crc Biostatistics Series. Milton: CRC Press.

Iwagami, Masao, and Yoshinori Takeuchi. 2021. “Introduction to Self-Controlled Study Design.” *Annals of Clinical Epidemiology* 3 (3): 67–73.

Petersen, Irene, Ian Douglas, and Heather Whitaker. 2016. “Self Controlled Case Series Methods: An Alternative to Standard Epidemiological Study Designs.” *BMJ (Online)* 354: i4515–i4515.

Whitaker, Heather J., C. Paddy Farrington, Bart Spiessens, and Patrick Musonda. 2006. “Tutorial in Biostatistics: The Self-Controlled Case Series Method.” *Statistics in Medicine* 25 (10): 1768–97.