Familiarity of DM

SDC / Jannie Nielsen, IFSV, KU
http://bendixcarstensen.com/SDC/FamDM
June 2015
Version 0

Compiled Friday 3rd July, 2015, 11:11
from: /home/bendix/sdc/extn/KU/JannieNielsen/fam

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

Contents

1 Familial aggregation of DM

With the construction of the NDR and of the family register it is possible to assess who is related to whom genetically (with some minor uncertainty from adoptions etc.) as parents/children and siblings.

The classical way of looking at genetic associations' influence on clinical/disease outcomes is with models that includes random effects, that is random variables that are supposed to represent shared frailty, susceptibility, clinical features etc.

These models easily extend to intensity models (incidence / mortality), normally called frailty models, referring to the interpretation of the random effects.

2 Marital aggregation of DM

As opposed to familial relations, marital (co-habitation included here) relations come and go, and so in terms of shared frailty models we would expect to have frailties that switched on and off during life as marriages come and go.

However it seems more appropriate to use the registers of marriage/cohabitation and diabetes occurrence to classify the life course of each person in (at least) the following states.

- single, never cohab
- cohab
- post-cohab
- widowed

In this setup, "cohab" can mean either cohabitation without marriage or married, optionally having separate categories for the two.

These states must be operationalized in such a way that any given person can be allocated to precisely one state at each point in time.

In order to capture the diabetes status and the effect of this on cohabitants, cohab (and subsequent) states must be subdivided by the state of the spouse into:

- no DM
- DM *at* cohabitation start
- DM *during* cohabitation

2.1 Generalization to networks

Cohabitation is a very simple and special type of network, with a very limited number of (DM relevant) states, but other and more general types of network may be defined in terms of work relationships or geographical proximity. If these contain a lot of people at any given time, explicit enumeration of all possible states (such as number of colleagues) is not feasible, and some sort of continuous score of your network (at any given time) is required, either as the *number* or *fraction* of your network affected by, say, diabetes.

On top of this, a person's network can be subdivided by proximity (distance), and the scores (number,fraction) recorded as a continuous function of proximity.

In the cohabitation setting you essentially only have (at most) one person in your network, so not only do you not have to bother about proximity, but you also only have a binary state of your network, namely either affected or not.

So general network analysis seems to involve some sort of time-varying exposure which essentially is an exposure (diabetes prevalence) at each point in your network. The simplest possible just a prevalence as a function of proximity, but the *number* in your network may be of particular interest too.

2.1.1 Modeling and reporting

The question of relevance is thus how exposures at different proximities influence your risk. The simplest cohabitation setting only has 0 or 1 in your proximity and that person is either DM or not, so it reduces to a 3-level categorical variable (single/cohab no DM/cohab DM), changing by (calendar) time.

In the general network setting, each person's network status (covariates, state membership, ...) is two functions of proximity; the number in the network and some diabetes score (prevalence for example).

Thus, follow-up of a person has for each little interval, two *functions* as covariates, so the question is how to model (quantify) and in particular report the effect of such functions. In the first place assuming that the effects are independent of age and other "classical" covariates we would have a score for each proximity, d, say, (distance), $s_d, d = 0, 1, 2, \ldots$ — think of this as for example the prevalence of diabetes in a certain distance from a person (at a give time).

The log-rate of diabetes occurrence could then be modelled as:

$$\log \lambda(t) = f(a, p, c) + \sum_d \beta_d s_d$$

If this type of analysis should be meaningful, we would presumably impose some sort of monotonicity constraint on the β_ds say $\beta_d = \alpha_0 + \alpha_1 d + \alpha_2 d^2$, so the model would be:

$$\begin{aligned}\log \lambda(t) &= f(a, p, c) + \sum_d (\alpha_0 + \alpha_1 d + \alpha_2 d^2) s_d \\ &= f(a, p, c) + \alpha_0 \sum_d s_d + \alpha_1 \sum_d d s_d + \alpha_2 \sum_d d^2 s_d\end{aligned}$$

This shows one way of reducing the recorded scores at a large number of distances to form a parametric model for the risk.

The reporting of the estimates from this model would be as a 2^{nd} order polynomial (parabola) as a function of proximity (d), under an assumption of say uniform distribution of scores over the network. Note that the setup with score as a predictor, requires a particular assumption of the scores in order to report model predictions.

Moreover, this is a particularly simple model assuming:

- there is no time-lag between the score (s_d) and the outcome
- the effect of proximity (the 2^{nd} order polynomial) is the same for all levels of the score

The former could be alleviated by lagging the network information (but there is very little information in actual data to support a particular lag). The latter is basically just an interaction, so that we instead of reporting a set of parallel polynomials showing the incidence rates as a function of proximity for different levels of scores, the polynomial would no longer be parallel.

3 Modeling

This is thus a classification of *all* lifetime in *all* persons in Denmark, and as such represents a categorical (time-varying) exposure.

The (relative) diabetes occurrence rates between the states reflect combined effects of common environment and preferential mating. If a detailed model for these effects is defined it may be possible to tease these effects apart.

3.1 Timescales

Furthermore, in the analysis of (DM incidence) rates we must include age (preferably separately for each sex) as well as time since cohabitation, time since partner's DM (in the relationship?) and presumably also time since start and end of cohabitation as such, and of course time since widowhood. Moreover some differences of these might be of interest, such as age at cohabitation, length of cohab at cessation etc.

In full generality, any multistate model can be expanded to let the transition intensities depend on time since entry into any previously visited state, as well as the point of entry evaluated on any timescale available. This of course becomes excessively weedy in practice, so the point is to make sufficiently simple and precise definitions.