

Indstilling om godkendelse af projekt

Autoriseret institution

Institut for Folkesundhedsvidenskab, Københavns Universitet

Projekttitel

Danish Social Network Database: The influence of family and social relations in diabetes epidemiology

Projektbeskrivelse

Undersøgelsen er en analyse af hvorledes type 2 diabetes samt relaterede risikofaktorer og sygdomme akkumuleres i familiære og sociale netværk i Danmark. Forekomsten af type 2 diabetes i Danmark er kraftigt stigende til trods for at sygdommen til en vis grad kan forebygges. Derfor er nye forebyggelsestiltag påkrævede.

I denne undersøgelse vil vi derfor sammenkæde individer i sociale netværk via følgende relationer: familiære, ægtefæller, samboende, skolekammerater, arbejdskolleger samt geografisk nærområde. Når dette netværk er skabt vil vi analysere, hvordan type 2 diabetes og relaterede sygdomme samt risikofaktorer 'spredes' i sociale netværk. Analyserne vil blive gennemført ved brug af 'R' og tilhørende pakker til sociale netværks analyser ('sna', 'sand', 'igraph').

Resultaterne fra dette studie vil afdække i hvilken grad sundhed og type 2 diabetes-specifikke risikofaktorer deles af personer i samme husstand eller sociale netværk under kulturel, økonomisk, og geografisk variation, hvilket vil være af essentiel vigtighed for en informeret debat om, hvorvidt fremtidige sundhedsinterventioner skal fokusere på familier, husstande, arbejdspladser, skoler osv. frem for den hidtidige praksis med et mere snævert fokus på det enkelte individ. En detaljeret projektbeskrivelse på engelsk er vedlagt som bilag 1.

Population

Alle individer som var i live og boede i Danmark eller er født eller immigreret til DK efter d. 1. januar 1980. Dette drejer sig om potentielt 8.500.000 individer. At vi ønsker hele Danmarks population skyldes 1) De nævnte sociale relationer er ikke begrænset til et specifikt geografisk område i Danmark; og 2) Sociale relationer kan have forskellig betydning i forskellige geografiske områder såsom storbyer versus landsbyamfund.

Variabelindhold

- Demografi, bopæl (BOPA)
- Civilstatus, Familie og husstandsforhold (FAIN)
- Uddannelsesoplysninger (UDDA)
- Indkomster (INDH/INDK)
- Arbejdsplads(er)

Se i øvrigt specifikation af variable i vedlagt variabeliste (Bilag 2).

Særligt vedr. Lægemiddelsdata

Forskningsprojekter, der skal anvende data fra Lægemiddeldatabasen, skal godkendes hos Statens Serum Institut. Godkendelsesprocessen sker via den projektansvarlige medarbejder i Forskningsservice. Lægemiddelsdata må kun anvendes til formålet beskrevet i denne projektindstilling.

Såfremt lægemiddeldata skal anvendes til andre formål kræves en fornyet godkendelse fra Serum instituttet.

Eksterne data eller egne data

Eksterne data/egne data til dette projekt er:

Fra Sundhedsdatastyrelsen vil vi desuden søge om adgang til følgende register for data om diabetes samt relaterede sygdomme og risikofaktorer:

- *Kroniske sygdomsdiagnoser:*
 - Det Nationale Diabetesregister
 - Landspatientregistret
 - Cancerregisteret
- *Dødsårsag:*
 - Dødsårsagsregisteret
- *Medicinering kronisk sygdom:*
 - Lægemiddelstatistikdatabasen
- *Overvægt:*
 - Sessionsregisteret

Datatilsynet

Godkendelse fra Datatilsynet vedlagt i bilag 3.

Projektperiode

Tidspunkt for sletning af oplysninger: Der ønskes adgang til data frem til og med december 2021.

Autoriserede forskere

Navn: Jannie Nielsen, Postdoc

Institution: Institut for Folkesundhedsvidenskab, Københavns Universitet

E-mail: jannien@sund.ku.dk

Mobilnummer: +4522770909

Forsker-ident: Haves endnu ikke

Navn: Claus Thorn Ekstrøm, Professor

Institution: Institut for Folkesundhedsvidenskab, Biostatistisk Afdelinge

E-mail: ekstrom@sund.ku.dk

Mobilnummer: 35 32 75 97

Forsker-ident:

Navn: Daniel Witte, Professor

Institution: Institut for Folkesundhed, Aarhus Universitet

E-mail: Daniel.witte@ph.au.dk

Mobilnummer: +31 628423150

Forsker-ident: FEKS

Navn: Adam Hulman

Institution: Institut for Folkesundhed, Aarhus Universitet

E-mail: Adam.hulman@ph.au.dk

Mobilnummer: 93508208

Forsker-ident: FEKR

Navn: Bendix Carstensen

Institution: Steno Diabetes Center

E-mail: bxc@steno.dk

Mobilnummer: 30758738

Forsker-ident:

FSE, den

Godkendt til ekstern adgang

Ivan Thaulow

Bilag 1- Projektbeskrivelse

Danish Social Network Database: The influence of family and social relations in diabetes epidemiology

Introduction

It has been demonstrated in high-income (8,9) and middle-income countries (10,11), that the onset of Type 2 Diabetes (T2D) can be prevented or postponed in high-risk individuals through a healthy diet, increased physical activity, and weight loss. Although these trials were pivotal in providing unequivocal proof of principle that T2D can be prevented by a change in lifestyle, many outstanding issues have prevented the risk reduction demonstrated in these trials from being implemented in high-risk populations across the world. This failure may partly be due to the oversight of the direct social context as a main determinant of human behaviour. While psychology and sociology have extensively documented the close social circle as a determinant of human behaviour, diabetes studies, trials, and risk reduction programmes continue to focus either on large internally homogeneous groups or isolated individuals.

There are indications that obesity can cluster in and spread through social networks, consisting of both genetically related individuals (siblings, parents, children) and non-genetically related individuals (spouses, friends)(12). Spouses of people with diabetes have been shown to have an increased risk of diabetes of between 18% and 26% depending on BMI adjustment (13). However, the impact of wider social networks (beyond the spouse) on diabetes, other cardiovascular risk factors or on subclinical or clinical indicators of micro and macrovascular complications has not yet been studied. Yet, networks may be a key factor in the uptake of behaviours and beliefs that lead to obesity, and conversely of healthy lifestyle choices. If obesity and diabetes can 'spread' through social networks, these networks may also hold the key to successful prevention. In the same way vaccination of key individuals is an effective way to limit the spread of infectious diseases, targeting highly connected individuals in social networks for preventive efforts with regard to diabetes and CVD may yield a multiplicative preventive effect. Few population-based epidemiological studies to date include multiple members of a social network and information on the strength and nature of the network.

This project will generate a social network database for the entire population of Denmark, based on linkage and integration of data available through various national registers (for further details refer to the method section) and analyse how T2D and related diseases spread in social networks and thus under which conditions social networks could be a focus for T2D screening and prevention. This has never been attempted at the scale of an entire country including different geographical areas.

Purpose of the study

This study will investigate the hypothesis that clusters of prevalent diabetes can be identified in people who are closely related to each other in a social/family network (cross-sectional), and that the risk of diabetes incidence and related diseases can 'spread' via social network ties (prospective), depending on the number of degrees of separation from an affected individual. As a secondary more explorative objective, the study will investigate to which degree social network clustering and spread can be explained by clustering and spread of the risk factors for diabetes development obesity.

Problems

- How are social network connections (determinant) associated with the prevalence and incidence of diabetes and related cardiovascular disease events and are some social connections more dominant than others than others in different age groups, socio-economic groups or geographical areas?

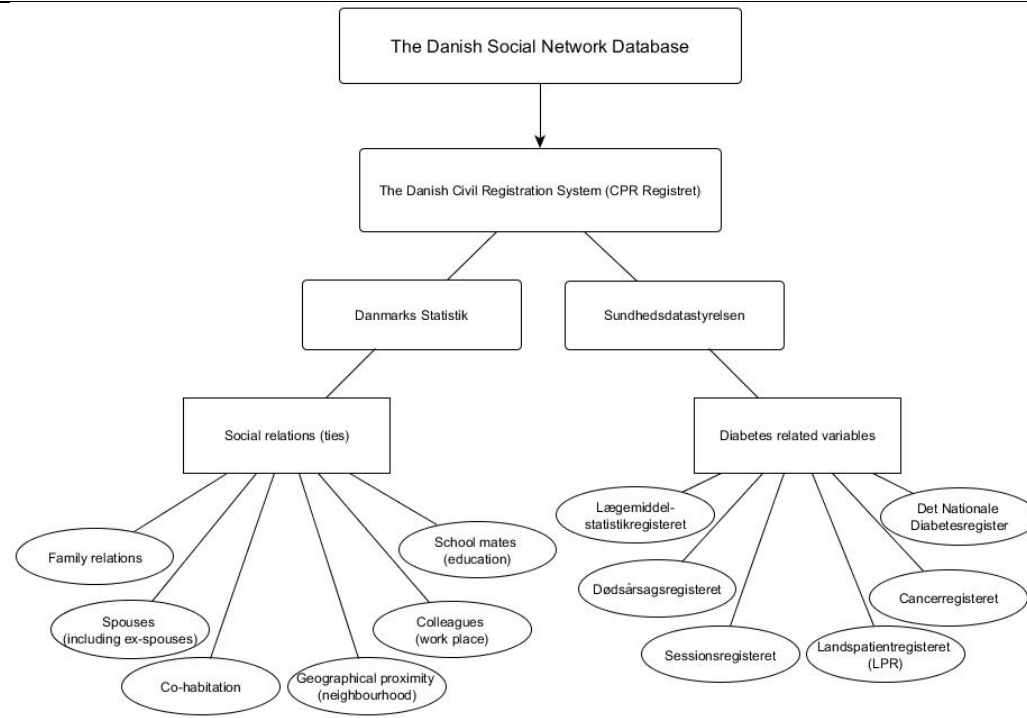
- How are social network connections (determinant) associated with risk factors and determinants of diabetes, such as obesity; and are some social connections more dominant than others in different age groups, socio-economic groups or geographical areas?

Data & Method

The project will include individuals who were alive or born after 1st of January 1980.

For an overview of included registries please refer to figure 1.

Figure 1. Overview of the registries used to generate the Danish Social Network Database



Data:

Generating the National Danish social network database:

By a register linkage approach, hosted at 'Statistics Denmark', we will create a data set of links between all inhabitants of Denmark (nodes), and social links between them (ties) based on:

Social Link	Source	Hosted by
(1) Marriage	CPR-registeret	Danmarks Statistik
(2) Co-habitation	CPR-registeret	Danmarks Statistik
(3) Family relation	CPR-registeret	Danmarks Statistik
a) Parents/children		
b) Siblings		
c) Third and fourth degree relatives		
(4) Social proximity during a particular time window		
a) Sharing an educational institution (folkeskole)	Elevregisteret	Danmarks Statistik
b) Sharing a place of work	IDA Databasen	Danmarks Statistik
c) Geographical proximity (sharing a local neighbourhood)	Bopælsinformation	Danmarks Statistik

Diabetes related variables:

This network database will subsequently be linked to two different sources of diabetes related outcomes:

<i>Variable</i>	<i>Register</i>	<i>Hosted by</i>
(1) Diabetes related outcomes for the entire Danish population		
(a) Diabetes diagnosis	Danish Diabetes Register	Sundhedsdatastyrelsen
(b) Related disease diagnoses and complications (hypertension, cardiovascular diseases, cancer etc. for calculating Charlson's co-morbidity index; foot ulcers/amputations)	Landspatientregisteret (LPR) Cancerregisteret Dødsårsagsregisteret Lægemiddelstatistik-registeret	Sundhedsdatastyrelsen/Danmarks Statistik
(c) Height and weight (body mass index)	Sessionsregistret	Sundhedsdatastyrelsen

The study will also include covariates like sex, age, socio-economic status (income, employment status and education), country of origin and migration status.

Data Analyses:

Data analyses will be performed primarily using 'R', equipped with relevant packages for social network analysis ('sna', 'sand', 'igraph') depending on the specific needs of the analysis at hand.

Analyses will first describe the topology (structure) of social networks, at the first possible baseline timepoint and as it develops over time. Characteristics of individuals (e.g. incidence of diabetes) will be modeled as a function of the characteristics of the individuals in the surrounding network, taking into account the network distance (first, second, third or higher degree of separation). For example, for each case of incident diabetes, the presence of pre-existing diabetes in members of the social network will be quantified and modeled according to the type of connection, giving risk estimates for being socially connected to a diabetic individual per connection type, duration and intensity. Statistical inferences will be made by comparison of observed networks to random networks, following the framework proposed by Erdős and Renyi.

For a more detailed description of statistical modeling please refer to Appendix A (Familial aggregation of Diabetes) and Appendix B (Assortative Mating).

Ethical considerations

The generated National Danish social network database will be generated and managed on servers hosted by Danmarks Statistik, and will be subject to their strict regulations regarding the handling of sensitive data. All CPR-numbers will be de-identified by Danmarks Statistik before the researchers (data handlers) will have access to the data. No individual data or data where individuals can be identified will be analyzed or transferred out of Danmarks Statistik.

All necessary permissions and registrations from Datatilsynet, Sundhedsdatastyrelsen, and Danmarks Statistik will be obtained before any data handling occurs.

Given that the link between people at a population level and regional/geographical variances are central feature of the project, data on the entire Danish population is necessary and therefore requested.

Perspectives

The project holds great potential of generating a novel understanding about the spread of disease in social network and the role that social network relationships play in the risk of T2D, which potentially can lead to a new and innovative approach for T2D prevention where several individuals at risk of T2D can be targeted at once instead of the high-cost individual T2D prevention approaches currently used. Such an understanding can also be utilised in the primary health sector to prevent high-risk individuals from developing diabetes and its complication.

Prospectively the Danish Social Network Database will include existing Danish cohort studies such as Addition and Inter99 to investigate to which degree social network clustering and spread can be explained by clustering and spread of the risk factors for diabetes development (e.g. obesity, physical inactivity, poor diet).

Reference List

1. IDF. IDF Diabetes Atlas Seventh Edition. 2015. International Diabetes Federation.
2. Gregg EW, Zhuo X, Cheng YJ, Albright AL, Narayan KM, Thompson TJ: Trends in lifetime risk and years of life lost due to diabetes in the USA, 1985-2011: a modelling study. *Lancet Diabetes Endocrinol* 2:867-874, 2014
3. Colagiuri S, Borch-Johnsen K, Glumer C, Vistisen D: There really is an epidemic of type 2 diabetes. *Diabetologia* 48:1459-1463, 2005
4. Wilson PW, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB, Sr.: Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. *Arch Intern Med* 167:1068-1074, 2007
5. Hu FB, Manson JE, Stampfer MJ, Colditz G, Liu S, Solomon CG, Willett WC: Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. *N Engl J Med* 345:790-797, 2001
6. Mozaffarian D, Kamineni A, Carnethon M, Djousse L, Mukamal KJ, Siscovick D: Lifestyle risk factors and new-onset diabetes mellitus in older adults: the cardiovascular health study. *Arch Intern Med* 169:798-807, 2009
7. Uusitupa MI, Stancakova A, Peltonen M, Eriksson JG, Lindstrom J, Aunola S, Ilanne-Parikka P, Keinanen-Kiukaanniemi S, Tuomilehto J, Laakso M: Impact of positive family history and genetic risk variants on the incidence of diabetes: the Finnish Diabetes Prevention Study. *Diabetes Care* 34:418-423, 2011
8. Knowler WC, Fowler SE, Hamman RF, Christophi CA, Hoffman HJ, Brenneman AT, Brown-Friday JO, Goldberg R, Venditti E, Nathan DM: 10-year follow-up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study. *Lancet* 374:1677-1686, 2009
9. Lindstrom J, Ilanne-Parikka P, Peltonen M, Aunola S, Eriksson JG, Hemio K, Hamalainen H, Harkonen P, Keinanen-Kiukaanniemi S, Laakso M, Louheranta A, Mannelin M, Paturi M, Sundvall J, Valle TT, Uusitupa M, Tuomilehto J: Sustained reduction in the incidence of type 2 diabetes by lifestyle intervention: follow-up of the Finnish Diabetes Prevention Study. *Lancet* 368:1673-1679, 2006
10. Pan XR, Li GW, Hu YH, Wang JX, Yang WY, An ZX, Hu ZX, Lin J, Xiao JZ, Cao HB, Liu PA, Jiang XG, Jiang YY, Wang JP, Zheng H, Zhang H, Bennett PH, Howard BV: Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance. The Da Qing IGT and Diabetes Study. *Diabetes Care* 20:537-544, 1997
11. Ramachandran A, Snehalatha C, Mary S, Mukesh B, Bhaskar AD, Vijay V: The Indian Diabetes Prevention Programme shows that lifestyle modification and metformin prevent type 2 diabetes in Asian Indian subjects with impaired glucose tolerance (IDPP-1). *Diabetologia* 49:289-297, 2006
12. Christakis NA, Fowler JH: The spread of obesity in a large social network over 32 years. *N Engl J Med* 357:370-379, 2007
13. Leong A, Rahme E, Dasgupta K: Spousal diabetes as a diabetes risk factor: A systematic review and meta-analysis. *BMC Med* 12:12, 2014

Appendix A – Familial Aggregation of Diabetes

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), $s_d, d = 0, 1, 2, \dots$ — 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_d s 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 2nd 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 2nd 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.

Appendix B – Assortative Mating

Assortative Mating

Claus Ekstrøm

14 Mar 2016

Assortative mating and its influence on family/network studies

When modeling related individuals it is customary assumed that an individual chooses his/her spouse completely randomly from the population. This is an assumption that is convenient from a mathematical perspective but is not very realistic. Humans, for example, have a significant correlation between spouses with respect to height, skin color, IQ, social status, religion, and some medical conditions.

Positive assortative mating cause identical alleles to come together more often than in the case of random mating, which will cause an increase in homozygosity just as inbreeding does. Similar arguments can be made for shared environmental risk factors if we disregard the assortative mating so modeling. In fact, unmeasured environmental factors may behave in the same way as genetic risk factors which consequently means that modeling individuals from the same family may result in biased estimates if random mating is assumed.

Modeling assortative mating

The genetic and environmental relationships for outbred families are typically modeled through correlation matrices representing the relationship between the individuals in a family/network. For example, the genetic and environmental relationship between a father, mother and child (first, second, and third rows, respectively) are typically viewed as

$$\begin{bmatrix} 1 & 0 & 0.5 \\ 0 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{bmatrix}$$

and

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix},$$

respectively.

The correlation due to assortative mating between related individuals can be modeled through path analysis and essentially introduces extra parameters, ρ and θ in the two matrices above so they become

$$\begin{bmatrix} 1 & \rho & (1+\rho)/2 \\ \rho & 1 & (1+\rho)/2 \\ (1+\rho)/2 & (1+\rho)/2 & 1 \end{bmatrix}$$

and

$$\begin{bmatrix} 1 & \rho' & (1+\rho')\theta \\ \rho' & 1 & (1+\rho')\theta \\ (1+\rho')\theta & (1+\rho')\theta & 1 \end{bmatrix},$$

respectively.

Thus, the assortative mating enables us to separate the genetic and environmental risk factors and provide estimates closer to the real situation with which individuals choose their partners.

Estimating assortative mating

While assortative mating *may* be a potential problems we do not know if it is a big problem in practice. Information on cohabitation may be used to infer and compare the correlation between phenotypes of married individuals and cohabitating but not married individuals and to the correlation between the same phenotypes of unrelated individuals. This will give us not only information about the scope of the potential problem but will also provide information that can be used as input to the modeling of general networks.

Importance for the study of familial aggregation of DM

Identification of assortative mating patterns in the Danish population will enable us to better partition the contributions of risk factors for DM into genetic and environmental components and provide a better statistical model for the shared environmental factors found in larger networks.

Bilag 2. Udtræksbeskrivelse Danmarks Statistik

For hele populationen ønskes følgende data på alle individer, der var i live eller født efter 1/1-1980.

Data sociale relationer		
Løbenummer (Entydigt ID)		
Familære relationer		Resource
<i>Periode:</i>		
Personnummer, barn	BARN_ID	DST
Personnummer, barn (henvisning til mor el. far)	PNRB	DST
Personnummer, far	PNRF	DST
Personnummer på mor	PNRM	DST
Faderens person ID	FAR_ID	DST
Moderens person ID	MOR_ID	DST
Familie id	C_FAMILIE_ID	DST
Familietype	FAMILIE_TYPE	DST
Personens plads i familien	PLADS	DST
Familiemedlemskab fra	MED_VFRA	DST
Familiemedlemskab fra	MED_VTIL	DST
Civil status relationer		
Civil status	CIVST	DST
Ægtefælle ID	AEGTE_ID	DST
Civilstandsdato fra	CIV_VFRA	DST
Civilstandsdato til	CIV_VTIL ??	DST
Partnerens ID	C_FAELE_ID	DST
Samboende		
Bopælsadresse_id	ADRESSE_ID	DST
Id for bopælsadresse	BOPAEL_ADRESSE_ID	
Dato for hvornår adressen gælder fra	ADR_VFRA	DST
Dato for hvornår adressen gælder til	ADR_VTIL	DST
Uddannelse		
Institutionsnummer	INSTNR	DST
Starttidspunkt	ELEV3_VFRA	DST
Sluttidspunkt	ELEV3_VTIL	DST
Uddannelseskode	UUD	DST
Uddannelsesdel	UDEL	DST
Uddannelsesform	UFORM	
Arbejdsplads		
Personlig arbejdssted adresse	PERSADR	
Fra?		
Til?		

Determinanter/Covariater		Source
Løbenummer (Entydigt ID)		
<i>Periode:</i>		
Fødselsdato	FOED_DAG	DST
Sex	KOEN	DST
Adoption	MADOPT	DST

Indvandrere, efterkommere, personer med dansk oprindelse	IE_TYPE	DST
Indvandringsland	INDV_LAND	DST
Oprindelsesland	OPR_LAND	DST
Indvandringsland	VAN_VTIL	DST
Uddannelse (højest færdiggjorte uddannelse)	HFAUDD	DST
Minimum uddannelsestid i måneder	PRIA	DST
Indkomst (individuel) Skattemæssig personlig indkomst beregnet af Skat?	PERINDKP	DST
Beskæftigelse (Employment)	BESKST (udgår i 2001 – ny?)	DST
Socioøkonomisk klassifikation	SOCIO (udgår i 2001 – ny?)	DST
Familietype	FAMILIE_TYPE	DST
Personens plads i familien	PLADS	DST
Antal personer i husstanden	ANTPERH	DST
Antal børn i husstanden	ANTBRNH	DST
Familieindkomst	FAMINDKOMSTALT (udgår i 2013; ny?)	DST
Familiens socioøkonomiske gruppe	FAMSOCIOGRUP	DST

Bilag 3 – Tilladelse fra Datatilsynet

Institut for Folkesundhedsvidenskab

Sendt til: Jannien@sund.ku.dk

Kopi til: Steffen Loft, Instituttleder, IFSV



20. JANUAR 2016

Tilladelse til behandling af personoplysninger i projektet ”The influence of family and social relations in diabetes epidemiology”

Institut for Folkesundhedsvidenskab har den **7. december 2015** ansøgt om tilladelse til behandling af personoplysninger i forbindelse med ovennævnte projekt.

Det Sundhedsvidenskabelige Fakultet (SUND) meddeler hermed tilladelse til behandling af personoplysninger i projektet. Tilladelsen er meddelt inden for rammerne af SUNDs fællesanmeldelse¹, og er givet efter reglerne i § 45, stk. 1, nr. 3, i persondataloven om tilladelse til behandling af personoplysninger i videnskabeligt eller statistisk øjemed. Læs mere om persondataloven på [Datatilsynets hjemmeside](#).

Tilladelsen gælder indtil den 31. december 2021. Al behandling (herunder også opbevaring) af personoplysninger efter tilladelsens udløb er en overtrædelse af persondataloven.

Tilladelsen forudsætter, at der ikke overføres personoplysninger til de nævnte samarbejdspartnere, før der er indgået databehandleraftaler for disse.

Projektet medtages på oversigten over godkendte forskningsprojekter, som administreres af fakultetet. Datatilsynet kan til enhver tid anmode om kopi af oversigten med henblik på kontrol, herunder kontrol med sikkerhedshændelser.

Tilladelsen er betinget af, at behandlingen af personoplysninger i projektet sker i overensstemmelse med de vilkår, som er fastsat af SUND. Den projektansvarlige er ansvarlig for, at persondatalovens regler og vilkårene overholdes. Vilklårene fremgår af bilaget til denne tilladelse og er gældende indtil videre. SUND forbeholder sig ret til senere at tage vilkårene op til revision.

Eventuelle vilkår, der fastsættes efter anden lovgivning, forudsættes overholdt.

Med venlig hilsen
Anders Berntsen
Fuldmægtig

FORSKNING OG INNOVATION

BLEGDAMSVEJ 3B
2200 KØBENHAVN N

DIR 35 33 49 77
MOB 23 64 94 20

anders.berntsen@sund.ku.dk

Sagsnr. SUND-2016-05

¹ Datatilsynets udtalelse (j.nr. 2015-57-0121) om Fakultetets fællesanmeldelse findes her: <https://anmeld.datatilsynet.dk/frontend/fortegnelse/vis.off.asp?myid=97056&pub=yes&myid=2015-57-0121>

Vilkår fastsat i forbindelse med tilladelse til behandling af personoplysninger i projekt under SUNDs fællesanmeldelse

1 Generelle vilkår

- Personoplysninger må kun anvendes til brug for projektets gennemførelse. Oplysningerne må ikke indgå i administrativ eller konkret sagsbehandling. Oplysningerne må heller ikke anvendes som grundlag for konkrete retlige eller faktiske foranstaltninger over for de omhandlede personer eller andre personer. Det er kun resultatet af den videnskabelige eller statistiske bearbejdning af personoplysninger, der kan bruges i administrativ sammenhæng, og kun under forudsætning af, at anvendelsen af resultaterne sker på en sådan måde, at det ikke er muligt at identificere enkeltpersoner. (Persondatalovens § 10, stk. 2).
- Behandlingen af personoplysningerne må kun foretages af den projektansvarlige eller af personer, som den projektansvarlige bemyndiger hertil. Medarbejderne skal have instruktion og oplæring i, hvad de må gøre med personoplysningerne, og hvordan de skal beskytte oplysningerne. Adgang til personoplysningerne skal begrænses til personer, der har et sagligt behov for adgang. Det skal være så få personer som muligt.
- Personoplysninger må ikke opbevares på en måde, der giver mulighed for at identificere de registrerede i et længere tidsrum end det, der er nødvendigt af hensyn til projektets gennemførelse.

2 Informationssikkerhed

- Behandlingen af personoplysninger er omfattet af regler og retningslinjer i [sikkerhedshåndbogen for Københavns Universitet](#).
- Behandling af oplysninger skal tilrettelægges således, at oplysningerne ikke hændeligt eller ulovligt tilintetgøres, fortabes eller forringes. Der skal endvidere foretages den fornødne kontrol for at sikre, at der ikke behandles urigtige eller vildledende oplysninger. Urigtige eller vildledende oplysninger eller oplysninger, som er behandlet i strid med persondataloven eller disse vilkår, skal berigtiges eller slettes.

2.1 Fysisk sikkerhed

- Lokaler, der benyttes til opbevaring og behandling af projektets oplysninger, skal være indrettet med henblik på at forhindre uvedkommende adgang. Manuelt projektmateriale, udskrifter, fejl- og

kontrollister, m.v., der direkte eller indirekte kan henføres til bestemte personer, skal opbevares forsvarligt aflåst og på en sådan måde, at uvedkommende ikke kan gøre sig bekendt med indholdet.

2.2 Adgangsstyring -passwords

- Adgangen til projektdata må kun finde sted ved benyttelse af et personligt fortroligt password. Kun personer, der er beskæftiget med eller har et andet sagligt formål til projektet må tildeles et password til projektets data. Passwordet skal afgrænses således, at den enkelte projektdeltager alene har rettigheder til de funktioner, der er relevante for denne og autorisationerne skal angive, i hvilket omfang brugeren må forespørge, inddatere eller slette personoplysninger. Læs mere på KUNet om "[det gode password](#)".
- Hvert halve år skal det kontrolleres, at projektdeltagerne har de korrekte rettigheder.

2.3 Logning

- Der skal foretages registrering af alle afviste adgangsforsøg. Hvis der inden for en fastsat periode er registreret et nærmere fastsat antal på hinanden følgende afviste adgangsforsøg fra samme arbejdsstation eller med samme brugeridentifikation skal der blokeres for yderligere forsøg. Der skal følges op på afviste adgangsforsøg.
- Der skal foretages logning af alle anvendelser af personoplysninger i forbindelse med projektet. Loggen skal mindst indeholde oplysning om tidspunkt, bruger, type af anvendelse og angivelse af den person, de anvendte oplysninger vedrørte eller det anvendte søgekriterium.
- Såfremt identifikationsoplysninger enten er krypterede eller erstattet af et id-nummer, skal loggen alene indeholde oplysninger om bruger og tidspunktet for behandlingen (se, gemme, søge, opdatere m.v.)
- Loggen skal opbevares i 6 måneder, hvorefter den skal slettes. Ved særligt behov kan loggen opbevares i op til 5 år.
- Personoplysningerne skal i videst muligt omfang behandles i en form, hvor de ikke er umiddelbart personhenførbare, fx i krypteret form eller under et løbenummer i stedet for under personnummer. Nøglefiler – krypteringsnøgle, kodenøgle m.v. – skal opbevares forsvarligt og adskilt fra personoplysningerne.

2.4 Overførsel af personoplysninger uden for KU netværk

- Ved behov for adgang til projektets data uden for Københavns Universitets netværk, skal dette ske via en VPN-forbindelse. Kontakt SUND-ITs servicedesk for at få en VPN-forbindelse.
- Ved overførsel af personhenførbare oplysninger via Internet eller andet eksternt netværk uden for Københavns Universitet skal der træffes de fornødne sikkerhedsforanstaltninger mod, at oplysningerne kommer til uvedkommendes kendskab. Oplysningerne skal som minimum være forsvarligt krypteret under hele transmissionen. Overføres data inden for Københavns Universitets netværk sker overførslen krypteret/sikkert. Læs mere i [Københavns Universitets krypteringsvejledning](#).
- Data må ikke opbevares lokalt på fx C-drevet på pc'er eller USB-nøgler. Dog kan der ske sikkerhedskopiering af data til udtagelige lagringsmedier. Disse skal opbevares forsvarligt aflåst og således, at uvedkommende ikke kan få adgang til oplysningerne.

2.5 Databehandleraftale

- De fastsatte vilkår skal iagttages ved behandling, der foretages af databehandler. Ved brug af databehandler indgås databehandleraftale mellem Det Sundhedsvidenskabelige Fakultet og databehandleren.

3 Den registreredes (forsøgspersonens) rettigheder

3.1 Oplysningspligt

- Hvis der skal indsamles oplysninger hos den registrerede (ved interview, spørgeskema, klinisk eller paraklinisk undersøgelse, behandling, observation m.v.) skal der uddeles/fremsendes nærmere information om projektet. Den registrerede skal heri oplyses om den dataansvarliges navn, formålet med projektet, at det er frivilligt at deltage, og at et samtykke til deltagelse til enhver tid kan trækkes tilbage.
- Den registrerede skal endvidere oplyses om, at projektet er anmeldt til Datatilsynet via Det Sundhedsvidenskabelige Fakultets fælles-anmeldelse og at der er fastsat nærmere vilkår til beskyttelse af den registreredes privatliv.

4 Videregivelse af personoplysninger

4.1 Videregivelse til databehandlere i Danmark

- Videregivelse af personoplysninger til andre databehandlere i Danmark må kun ske med henblik på modtagerens udførelse af en undersøgelse i statistisk eller videnskabeligt øjemed, jf. § 10, stk. 2.
- Oplysningerne skal i videst muligt omfang videregives i en form, hvor de ikke er umiddelbart personhenførbare for modtageren, dvs. uden f.eks. personnummer, navn eller adresse på de personer, der videregives oplysninger om.
- Det Sundhedsvidenskabelige Fakultet skal meddele forudgående tilladelse til videregivelsen. Ansøgning om tilladelse til videregivelse sendes ved anvendelse af blanketten ”ansøgning om tilladelse til videregivelse af personoplysninger”.

4.2 Videregivelse til dataansvarlige i udlandet

- Videregivelse til behandling for dataansvarlige i udlandet kan alene ske efter særskilt, forudgående tilladelse fra Datatilsynet. Anmodning om tilladelse til videregivelse til udlandet skal indgives på en særlig blanket, som findes på [Datatilsynets hjemmeside under ”Anmeldelse og tilladelse”](#).

5 Ændringer i projektet

- Væsentlige ændringer i projektet skal anmeldes/meddeles til Det Sundhedsvidenskabelige Fakultet (som ændring af eksisterende anmeldelse).
- Ændring af tidspunktet for projektets afslutning skal altid anmeldes/meddeles til Det Sundhedsvidenskabelige Fakultet.

6 Ved projektets afslutning

- Senest ved projektets afslutning skal oplysningerne slettes, anonymiseres eller tilintetgøres, således at det efterfølgende ikke er muligt at identificere enkeltpersoner, der indgår i projektet, medmindre fortsat opbevaring kræves efter anden lovgivning. Manuelt projektmateriale skal tilintetgøres.
- Sletning af oplysninger fra elektroniske medier skal ske på en sådan måde, at oplysningerne ikke kan genetableres. Der bør i denne forbindelse tages kontakt til SUND-IT.

- Sletning af oplysninger fra elektroniske medier skal ske på en sådan måde, at oplysningerne ikke kan genetableres. Der bør i denne forbindelse tages kontakt til SUND-IT.

6.1 Overførsel af data til opbevaring i Rigsarkivet

- Alternativt kan oplysningerne overføres til videre opbevaring i Rigsarkivet (herunder Dansk Dataarkiv) efter arkivlovens regler.
- Data klausuleret af DFF- Sundhed og Sygdom:
 - a. Forskningsrådet indsætter klausul i rammebevillingen vedr. arkivering af data hos Rigsarkivet.
 - b. Snarest efter Rigsarkivets modtagelse af kopi af rammebevillingen fra forskningsrådet fremsendes et brev til den ansvarlige forsker vedr. forventet dato for aflevering af data og dokumentation til Rigsarkivet.
 - c. Når datoen for aflevering nærmer sig, kontakter Rigsarkivet den ansvarlige forsker med information om, hvordan afleveringen skal finde sted.

Hvis undersøgelsen af den ene eller anden grund ikke indeholder unikke data, så anbefales det, at forskeren kontakter Rigsarkivet desangående, så vi kan få afsluttet sagsbehandlingen til gavn for alle parter.

- Ikke-klausulerede, unikke forskningsdata skal anmeldes til Rigsarkivet via følgende anmeldelsesblanket: <https://www.sa.dk/aflevering-arkivet/private-og-forskere/anmeldelse-af-offentlige-og-private-forskningsdata>, hvorefter Rigsarkivet tager stilling til om data skal arkiveres.

6.2 Præsentation af resultater

- Den dataansvarlige skal til enhver tid sikre sig, at dokumenter og andre præsentationer, som publiceres eller på anden måde gøres tilgængelig for andre på internettet, USB-nøgle eller på andet elektronisk medie, ikke indeholder personoplysninger. Der skal vises særlig agtpågivenhed i forbindelse med brug af grafiske præsentationer i Excel og PowerPoint, da de uforvarende kan indeholde indlejrede persondata i form af regneark, tabeller mv. Præsentationer, der gøres tilgængelig på internettet, skal derfor omformateres til PDF-format, da dette fjerner eventuelle indlejrede Excel-tabeller.
- En eventuel offentliggørelse af undersøgelsens resultater må ikke ske på en sådan måde, at det er muligt for udenforstående at identificere enkeltpersoner.