#### Forskningsservice Dato: 23-07-2015

Opgave nr. 704550

# Indstilling om godkendelse af projekt

### Opdatering d. 23. juli 2015

Projektet er udvidet med yderligere ca. 2300 lymfompatienter, samt oplysninger om deres behandling. Disse patienter er køns- og aldermatchet med ca. 23000 kontroller ud fra diagnosedatoen hhv. datoen for afsluttet behandling. Opdateret SAS program er vedlagt.

### Autoriseret institution

Gentofte Hospital

### Projekttitel

Overlevelse med lymfom

### Projektbeskrivelse

Formålet med projektet er at undersøge overlevelsen med lymfom i forhold til baggrundsbefolkningen.

### Population

Populationen er 291 lymfom cases og 2910 køns- og aldersmatchede kontroller.

### Variabelindhold

Alder og køn

Død og dødsårsager

Sygehuskode

Diagnose

Detaljer om cancer

Se i øvrigt specifikation af variable i vedlagt SAS-program

### Autoriserede forskere

Berit Jamie Nielsen (ident: fbpr) Martin Bøgsted, Anders Ellern Bilgrau

FSE, den

Godkendt til ekstern adgang

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Ivan Thaulow

SAS program der genererer udtræk:

libname lymfom 'X:\Data\Workdata\_Hurtig\703573\Lymfom 4550';

libname cancer 'X:\Data\Rawdata\_Hurtig\703573\cancerregister';

libname pop 'X:\Data\Rawdata\_Hurtig\703573\pop';

libname dst 'X:\Data\Rawdata\_Hurtig\703573\DST';

options mergenoby=error;

\*Rawdata;

**data** rawdata; set cancer.danmarks\_statistik\_23\_03\_2014;

datofordod = input(datofordd,ANYDTDTE10.);

format datofordod MMDDYY10.;

drop datofordd;

hvem=**1**; /\*1=Anders\*/

**run**;

**data** lymfom.cases\_2014; set rawdata;

group=**1**;

byear=Year(Fdselsdato);

rename Fdselsdato=fdato;

if KnMmandFkvinde='F' then sex=**0**;

if KnMmandFkvinde='M' then sex=**1**;

drop KnMmandFkvinde;

rename diagnosedato=event;

**run**;

\*Rawdata;

**data** lymfom.cases\_2015; set cancer.lymfom2;

doedsdato2 = input(doedsdato\_new,ANYDTDTE10.);

format doedsdato2 MMDDYY10.;

drop doedsdato\_new;

rename doedsdato2=doedsdato\_new;

hvem=**2**; /\*2=Lasse\*/

**run**;

**data** casesLasse; set lymfom.cases\_2015; /\*Matcher på diagnosetidspunkt\*/

group=**1**;

byear=Year(F7);

rename F7=fdato;

if F6='F' then sex=**0**;

if F6='M' then sex=**1**;

drop F6;

event=GC\_LDX\_DATO\_DIAGNOSE;

**run**;

**data** casesLasse2; set lymfom.cases\_2015; /\*matcher på sluttidspunkt for behandling\*/

group=**1**;

byear=Year(F7);

rename F7=fdato;

if F6='F' then sex=**0**;

if F6='M' then sex=**1**;

drop F6;

event=GC\_LB\_KEMOSLUT;

if GC\_LB\_STRAALBEH1DATO ne **.** then event=GC\_LB\_STRAALBEH1DATO; \*enkelte pt mangler straalbehfrakt;

if GC\_LB\_STRAALBEH1DATO ne **.** and GC\_LB\_STRAALBEHFRAKT ne **.** then event=GC\_LB\_STRAALBEH1DATO+GC\_LB\_STRAALBEHFRAKT;

format event date10.;

**run**;

\*all cases;

**data** cases; set lymfom.cases\_2014 casesLasse; **run**;

\*pnr - på cases;

**data** pnrcases; set rawdata (keep=pnr hvem) lymfom.cases\_2015 (keep= pnr hvem); **run**;

/\*proc sort data=pnrcases nodupkey; by pnr; run; -OK alle obs er unikke\*/

\*Population, hvor mathces skal findes;

**data** pop; set pop.pop; **run**;

\*Første indvandring og seneste udvandring fra DK;

**data** vandring\_ind; set dst.vnds2012;

where indud='1';

**run**;

**proc** **sort** data=vandring\_ind; by pnr haend\_dato;

**data** vandring\_ind; set vandring\_ind;

retain first\_ind last\_ind;

by pnr;

if first.pnr then

do;

first\_ind=haend\_dato; last\_ind=**.**;

end;

if last.pnr then

do;

last\_ind=haend\_dato;

output;

end;

drop indud;

format last\_ind first\_ind date7.;

**run**;

**data** vandring\_ud; set dst.vnds2012;

where indud='2';

**run**;

**proc** **sort** data=vandring\_ud; by pnr haend\_dato;

**data** vandring\_ud; set vandring\_ud;

retain first\_ud last\_ud;

by pnr;

if first.pnr then

do;

first\_ud=haend\_dato; last\_ud=**.**;

end;

if last.pnr then

do;

last\_ud=haend\_dato;

output;

end;

drop indud;

format last\_ud first\_ud date7.;

**run**;

\* IEtype og indvandringsland;

**%macro** ***iepe***;

data iepe;

set %do i= **1980** %to **2013**; dst.iepe&i %end;;

run;

**%mend**;

%***iepe***;

**proc** **sort** data=iepe; by pnr; **run**;

**data** iepe; set iepe;

by pnr;

if last.pnr; **run**;

**run**;

\* Første registrering i Familie FAIN og BEF;

**%macro** ***fain***;

%do i=**1980** %to **2007**;

data fain&i; set dst.fain&i (keep=pnr);

length bef\_year **3**;

bef\_year=&i;

run;

%end;

%do i=**1986** %to **2013**;

data bef&i; set dst.bef&i (keep=pnr);

length bef\_year **3**;

bef\_year=&i;

run;

%end;

data bef\_fain; set

%do i=**1980** %to **2007**; fain&i %end;

%do i=**1986** %to **2013**; bef&i %end;;

run;

**%mend**;

%***fain***;

**proc** **sort** data=bef\_fain; by pnr bef\_year; **run**;

**data** bef\_fain; set bef\_fain;

by pnr;

if first.pnr;

**run**;

\*Samle befolkning;

**proc** **sort** data=pop; by pnr; **run**;

**proc** **sort** data=vandring\_ind; by pnr; **run**;

**proc** **sort** data=vandring\_ud; by pnr; **run**;

**proc** **sort** data=iepe; by pnr; **run**;

**proc** **sort** data=bef\_fain; by pnr; **run**;

**data** pop; merge pop vandring\_ind vandring\_ud iepe bef\_fain; by pnr;

if not first.pnr then delete; /\*fjerner 2 dobbeltgængere om hvem man intet ved\*/

\*fjerner personer uden køn og definerer vanlig sex;

if kon='' then delete;

sex=kon=**1**; drop kon;

\* fjerner mærkelige cpr-numre;

if \_cprchk\_=**0**; if \_cprtype\_=**1**;

drop \_cprchk\_ \_cprtype\_ haend\_dato;

**run**;

\*døde;

**data** doede; set dst.dod2013;

if doddato>**'01JAN2002'd** then

do;

doedsaars1=''; doedsaars2=''; doedsaars3=''; doedsaars4='';

end;

**run**;

**data** doede2; set dst.dodsaasg2012 ;

drop C\_bopkom c\_region c\_bopkomf07 c\_bopamtF07 C\_liste14 C\_liste49 d\_dodsdato d\_findedato d\_statdato c\_dodsmaade

c\_dodssted c\_praecis\_dodssted C\_FINDESTED C\_PRAECIS\_FINDESTED C\_HAENDELSESSTED C\_OBDUKTION c\_operation c\_laegefunktion

v\_alder c\_sex /\*15.3.12- tilgrundlæggende død ikke fjernet længere\*/ ;

**run**;

**proc** **sort** data=doede; by pnr; **run**;

**proc** **sort** data=doede2; by pnr; **run**;

**data** doede; merge doede doede2; by pnr;

if first.pnr;\* Removes 4 doublicate death records;

if doddato<**'01JAN2002'd** then

do;

if (substr(doedsaars1,**1**,**3**)in('I21','I22') ) or substr(doedsaars1,**1**,**3**)='410' or

(substr(doedsaars2,**1**,**3**)in('I21','I22') ) or substr(doedsaars2,**1**,**3**)='410' or

(substr(doedsaars3,**1**,**3**)in('I21','I22') ) or substr(doedsaars3,**1**,**3**)='410'

then dod\_ami=**1**;

if (substr(doedsaars1,**1**,**3**)in('I61','I62','I63','I64')) or **431**<=substr(doedsaars1,**1**,**3**)<=**436** or /\*rettet fra 435 til 436 \*/

(substr(doedsaars2,**1**,**3**)in('I61','I62','I63','I64')) or **431**<=substr(doedsaars2,**1**,**3**)<=**436** or

(substr(doedsaars3,**1**,**3**)in('I61','I62','I63','I64')) or **431**<=substr(doedsaars3,**1**,**3**)<=**436**

then dod\_stroke=**1**;

if substr(doedsaars1,**1**,**1**)='I' or **400**<=substr(doedsaars1,**1**,**3**) <=**451** or

substr(doedsaars2,**1**,**1**)='I' or **400**<=substr(doedsaars2,**1**,**3**) <=**451** or

substr(doedsaars3,**1**,**1**)='I' or **400**<=substr(doedsaars3,**1**,**3**) <=**451**

then dod\_cv=**1**;

end;

array doedsaars[**13**] c\_dodtilgrundl\_acme c\_dod\_1A c\_dod\_1b c\_dod\_1c c\_dod\_1d c\_dod\_21 c\_dod\_22 c\_dod\_23 c\_dod\_24 c\_dod\_25 c\_dod\_26 c\_dod\_27 c\_dod\_28; /\* stavefejl rettet 15.3.12 \*/

if doddato>=**'01JAN2002'd** then

do i=**1** to **13**;

if substr(doedsaars[i],**1**,**3**) in ('I21','I22') then dod\_ami=**1**;

if substr(doedsaars[i],**1**,**3**) in ('I61','I62','I63','I64') then dod\_stroke=**1**;

if substr(doedsaars[i],**1**,**1**)='I' then dod\_cv=**1**;

end;

**run**;

**proc** **sort** data=pop; by pnr; **run**;

**proc** **sort** data=doede; by pnr; **run**;

**data** pop; merge pop doede; by pnr ;

**run**;

**data** popny; set pop;

dod=doddato ne **.**;

if doddato= **.** then doddato=**'01jan2014'd**; /\*default dato for personer uden dødsdato\*/

**run**;

**data** pnrpop; set popny (keep=pnr); **run**;

**data** controls;

set popny;

group=**0**;

byear=Year(fdato);

format doddato MMDDYY10.;

format fdato MMDDYY10.;

rename doddato=event;

**run**;

**proc** **sort** data=cases; by pnr; **run**;

**proc** **sort** data=controls; by pnr; **run**;

**proc** **sort** data=casesLasse2; by pnr; **run**;

\*laver datasæt med controller og cases til brug i match macro;

**data** samlet; set cases (keep=pnr sex event group byear) controls (keep=pnr sex event group byear); by pnr;

if not first.pnr then delete;

**run**;

**data** samlet2; set casesLasse2 (keep=pnr sex event group byear) controls (keep=pnr sex event group byear); by pnr;

if not first.pnr then delete;

**run**;

**proc** **freq** data=samlet2; tables group; **run**;

/\*\*\*\*\*\*\*\*\* Greedy match macrosyntax - Dr. Gunni 19-10-08 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\* statement (syntax) til at køre op greedy match macro \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\* nedenstående copy-pastes ind i dit program udfyldes korrekt og køres derfra \*\*\*\*\*/

/\*\*\*\*\*\*\*\*\* justeret libnames for den nye server 13-12-2010 / GG \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

%include 'G:\Data\Alle\Help programs\_703421\Match macro\greedy\_match\_macro.sas';

%***GMATCH***(DATA=samlet, /\* datasæt som indeholder case og kontrol samt de nødvendige variabler - oblikatorisk!\*/

GROUP=group, /\* variablen som definerer case og kontrol 1/0 - oblikatorisk!\*/

ID=pnr, /\* variable som identificeret hver individ i datasæt - personnummer - oblikatorisk!\*/

MVARS=sex byear, /\* variabler (numeriske) som matches på - kan være flere en én og adskilles med mellemrum - oblikatorisk! \*/

WTS=**1** **1**, /\* vægtningen på hver enkelt variable som matches på i forhold til hinanden (mellemrum i mellem hver tal - oblikatorisk) \*/

DMAXK=**0**, /\* Maximale afstand der tillades for matchning for hver variable der matches på - fx. 0 for køn hvis der skal matches på køn og 5 for alder hvis der maximalt tillades 5 års aldersforskel \*/

DMAX=, /\* ikke nødvendig og lades som regel være blank, men hvis man har meget strikte matchnings kriteria så skal dette tal være lavere en vægtningen for variablen som blev sat ovenfor \*/

DIST=**1**, /\*Indicates type of distance to calculate. 1=weighted sum(over matching vars) of absolute case-control differences(default) 2=weighted Euclidean distance \*/

NCONTLS=**10**, /\* antal kontrol der matches på hver case - minimum 1 \*/

TIME=event, /\* (JBO 29.7.2009) lades stå blank hvis alle inkluderes 1.1.1997, ved løbende inklution af case'ses sikre TIME: at kontrolerne er i live når case'n inkluderes. Denne variable skal indeholde index dato for alle case'ses og dodsdato for alle de mulige kontroller \*/

TRANSF=**0**, /\*Indicates whether all matching vars are to be transformed (using the combined case+control data) prior to computing distances. 0=no(default), 1=standardize to mean 0 and variance 1, 2=use ranks of matching variables.\*/

SEEDCA=**87877**,SEEDCO=**987973**, /\* Tal der sikrer random fordeling af cases og kontrol - lades stå uændret \*/

PRINT=y, /\* skal resultaterne printes i output y/n - default y \*/

OUT=matchedData, /\* navn på outputdatsæt hvor case og kontrol er definerede \*/

OUTNMCA=nomatchcases, /\* navn på outputdatasæt der indeholder cases der ikke kunne matches - default \_\_NMCA\*/

OUTNMCO=nomatchcontrols /\* navn på outputdatasæt der indeholder kontrol der ikke kunne matches - default \_\_NMCO\*/

);

**data** match; set matchedData (keep= \_\_catime \_\_cotime \_\_idca \_\_idco \_\_ca1);

format \_\_catime DDMMYY10.;

format \_\_cotime DDMMYY10.;

rename \_\_catime=diagdato;

rename \_\_cotime=dodsdatocontrol;

rename \_\_idca=pnr;

rename \_\_idco=pnrcontrol;

if \_\_ca1=**1** then kon='M';

if \_\_ca1=**0** then kon='F';

drop \_\_ca1;

**run**;

**data** lymfom.match; set match; **run**;

%***GMATCH***(DATA=samlet2, /\* datasæt som indeholder case og kontrol samt de nødvendige variabler - oblikatorisk!\*/

GROUP=group, /\* variablen som definerer case og kontrol 1/0 - oblikatorisk!\*/

ID=pnr, /\* variable som identificeret hver individ i datasæt - personnummer - oblikatorisk!\*/

MVARS=sex byear, /\* variabler (numeriske) som matches på - kan være flere en én og adskilles med mellemrum - oblikatorisk! \*/

WTS=**1** **1**, /\* vægtningen på hver enkelt variable som matches på i forhold til hinanden (mellemrum i mellem hver tal - oblikatorisk) \*/

DMAXK=**0**, /\* Maximale afstand der tillades for matchning for hver variable der matches på - fx. 0 for køn hvis der skal matches på køn og 5 for alder hvis der maximalt tillades 5 års aldersforskel \*/

DMAX=, /\* ikke nødvendig og lades som regel være blank, men hvis man har meget strikte matchnings kriteria så skal dette tal være lavere en vægtningen for variablen som blev sat ovenfor \*/

DIST=**1**, /\*Indicates type of distance to calculate. 1=weighted sum(over matching vars) of absolute case-control differences(default) 2=weighted Euclidean distance \*/

NCONTLS=**10**, /\* antal kontrol der matches på hver case - minimum 1 \*/

TIME=event, /\* (JBO 29.7.2009) lades stå blank hvis alle inkluderes 1.1.1997, ved løbende inklution af case'ses sikre TIME: at kontrolerne er i live når case'n inkluderes. Denne variable skal indeholde index dato for alle case'ses og dodsdato for alle de mulige kontroller \*/

TRANSF=**0**, /\*Indicates whether all matching vars are to be transformed (using the combined case+control data) prior to computing distances. 0=no(default), 1=standardize to mean 0 and variance 1, 2=use ranks of matching variables.\*/

SEEDCA=**87877**,SEEDCO=**987973**, /\* Tal der sikrer random fordeling af cases og kontrol - lades stå uændret \*/

PRINT=y, /\* skal resultaterne printes i output y/n - default y \*/

OUT=matched2, /\* navn på outputdatsæt hvor case og kontrol er definerede \*/

OUTNMCA=nomatchcases2, /\* navn på outputdatasæt der indeholder cases der ikke kunne matches - default \_\_NMCA\*/

OUTNMCO=nomatchcontrols2 /\* navn på outputdatasæt der indeholder kontrol der ikke kunne matches - default \_\_NMCO\*/

);

**data** match2; set matched2 (keep= \_\_catime \_\_cotime \_\_idca \_\_idco \_\_ca1);

format \_\_catime DDMMYY10.;

format \_\_cotime DDMMYY10.;

rename \_\_catime=diagdato;

rename \_\_cotime=dodsdatocontrol;

rename \_\_idca=pnr;

rename \_\_idco=pnrcontrol;

if \_\_ca1=**1** then kon='M';

if \_\_ca1=**0** then kon='F';

drop \_\_ca1;

**run**;

**data** lymfom.match\_2015\_2; set match2;

rename diagdato=slutbehdato;

**run**;

**data** cases\_2014; set lymfom.cases\_2014;

drop datoforsammenkringmedLPR group byear sex;

**run**;

**data** cases\_2015; set lymfom.cases\_2015;

**run**;

**proc** **sort** data=match; by pnr; **run**;

**proc** **sort** data=match2; by pnr; **run**;

**proc** **sort** data=cases\_2014; by pnr; **run**;

**proc** **sort** data=cases\_2015; by pnr; **run**;

**data** lymfom.matched\_2014; merge match cases\_2014 (in=data); by pnr;

if data;

**run**;

**data** lymfom.matched\_2015; merge match cases\_2015 (in=data); by pnr;

if data;

**run**;

**data** lymfom.matched\_2015\_2; merge match2 cases\_2015 (in=data); by pnr;

if data;

**run**;

/\* Gemmer pnr på kontroller\*/

**data** lymfom.pnrcontrols\_2014; set lymfom.matched\_2014 (keep= pnrcontrol);

**run**;

**data** lymfom.pnrcontrols\_2015; set lymfom.matched\_2015 (keep= pnrcontrol);

**run**;

**data** lymfom.pnrcontrols\_2015\_2; set lymfom.matched\_2015\_2 (keep= pnrcontrol);

**run**;

/\* Gemmer fødsesldato og vitalstatus på kontroller \*/

**data** control; set controls (keep= pnr fdato event dod);

rename fdato=foddtocont;

rename event=doddtocont;

rename pnr=pnrcontrol;

**run**;

**proc** **sort** data=lymfom.pnrcontrols\_2014; by pnrcontrol; **run**;

**proc** **sort** data=lymfom.pnrcontrols\_2015; by pnrcontrol; **run**;

**proc** **sort** data=lymfom.pnrcontrols\_2015\_2; by pnrcontrol; **run**;

**proc** **sort** data=control; by pnrcontrol; **run**;

**data** lymfom.controls\_2014; merge lymfom.pnrcontrols\_2014 (in=data) control (in=data2); by pnrcontrol;

if data; if data2;

**run**;

**data** lymfom.controls\_2015; merge lymfom.pnrcontrols\_2015 (in=data) control (in=data2); by pnrcontrol;

if data; if data2;

**run**;

**data** lymfom.controls\_2015\_2; merge lymfom.pnrcontrols\_2015\_2 (in=data) control (in=data2); by pnrcontrol;

if data; if data2;

**run**;

/\*Samler data om kontroller og cases i een fil\*/

**proc** **sort** data=lymfom.controls\_2014; by pnrcontrol; **run**;

**proc** **sort** data=lymfom.controls\_2015; by pnrcontrol; **run**;

**proc** **sort** data=lymfom.controls\_2015\_2; by pnrcontrol; **run**;

**proc** **sort** data=lymfom.matched\_2014; by pnrcontrol; **run**;

**proc** **sort** data=lymfom.matched\_2015; by pnrcontrol; **run**;

**proc** **sort** data=lymfom.matched\_2015\_2; by pnrcontrol; **run**;

**data** lymfom.samlet\_2014; merge lymfom.controls\_2014 lymfom.matched\_2014;

by pnrcontrol;

**run**;

**data** lymfom.samlet\_2015; merge lymfom.controls\_2015 lymfom.matched\_2015;

by pnrcontrol;

**run**;

**data** lymfom.samlet\_2015\_2; merge lymfom.controls\_2015\_2 lymfom.matched\_2015\_2;

by pnrcontrol;

**run**;

**proc** **sort** data=lymfom.samlet\_2014; by pnr pnrcontrol; **run**;

**proc** **sort** data=lymfom.samlet\_2015; by pnr pnrcontrol; **run**;

**proc** **sort** data=lymfom.samlet\_2015\_2; by pnr pnrcontrol; **run**;