

# DMreg på DST 707655

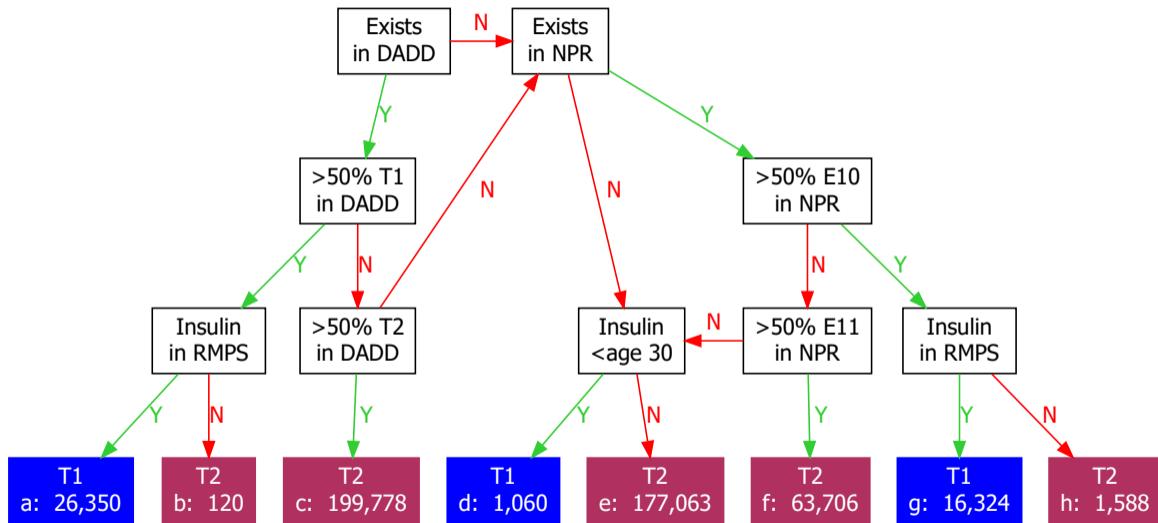
**Bendix Carstensen** Steno Diabetes Center  
Gentofte, Denmark  
<http://BendixCarstensen.com>

SDCC,  
26 august 2020

## Inclusion kriterier og -datoer for DMreg

- ▶ For hvert inklusions kriterium defineres den tidligste dato
  - ▶ Fodterapi for diabetes ptt. (Sygesikrings reg.)
  - ▶ Øjenscreening for diabetes ptt. (diaBase)
  - ▶ E10–E14 i LPR
  - ▶ OAD / Insulin, ATC A10Axx, A10Bxx (LMS)
  - ▶ DVDD indberetning — rapporteret diagnose dato
- ▶ RUKS-forbehold som er implementeret:
  - ▶ Inkluder ikke alene på baggrund af 1 LPR eller 1 recept
  - ▶ 2 LPR eller 2 recepter eller 1 af hver kræves
- ▶ Dette er den aktuelle algoritme i brug i DMreg — og en fodterapi, en øjen-screen eller en DVDD indberetning

# Definition of DM type i DMreg



## How is DMreg created

- ▶ A number of SAS-programs creating the SAS-dataset:  
`e:\workdata\707655\DMreg\data\DMreg.sas7bdat`
- ▶ It has 485,989 records, each representing 1 person
- ▶ A complete documentation is available in  
`v:\sdc\469drive\DMreg\tex\DMreg2018.pdf`
- ▶ Explains inclusion criteria etc. and a listing of all programs.

# SAS-version I

## The CONTENTS Procedure

Data Set Name	DMDAT.DMREG	Observations	486243
Member Type	DATA	Variables	21
Engine	V9	Indexes	0
Created	18/04/2020 14:39:42	Observation Length	136
Last Modified	18/04/2020 14:39:42	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label	Reconstructed DM register for Denmark		
Data Representation	WINDOWS_64		
Encoding	wlatin1 Western (Windows)		

## Engine/Host Dependent Information

Data Set Page Size	65536
Number of Data Set Pages	1011
First Data Page	*
Max Obs per Page	481
Obs in First Data Page	455
Number of Data Set Repairs	0
ExtendObsCounter	YES

# SAS-version II

Filename	E:\workdata\707655\DMreg\data\dmreg.sas7bdat
Release Created	9.0401M5
Host Created	X64_SR12R2
Owner Name	DSTFSE\FDIY7655
File Size	63MB
File Size (bytes)	66322432

## Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat	Label
21	DMtp	Char	*			Type of DM
*	PNR	Char	12	\$12.	\$10.	Personnummer
17	do2nd	Num	8	DDMMYY10.		Date of 2nd of Ins/OAD/NPR
12	doBth	Num	8	DDMMYY10.		Date of birth
15	doDM	Num	8	DDMMYY10.		Date of inclusion
16	doDVD	Num	8	DDMMYY10.		Date of DVDD
11	doDiaB	Num	8	DDMMYY10.	IS8601DA10.	Date of diaBase
14	doDth	Num	8	DDMMYY10.		Date of death
8	doIns	Num	4	DDMMYY10.		Date of 1st Ins
9	doIns2	Num	4	DDMMYY10.		Date of 2nd Ins
*	doNPR	Num	8	DDMMYY10.	DATE9.	Date of 1st NPR
4	doNPR2	Num	8	DDMMYY10.	DATE9.	Date of 2nd NPR
6	doOAD	Num	4	DDMMYY10.		Date of 1st OAD
7	doOAD2	Num	4	DDMMYY10.		Date of 2nd OAD

# SAS-version III

10	doPod	Num	8	DDMMYY10.	Date of Podiatry
5	dvdtyp	Char	*		Type from DVDD
20	hasdvd	Num	8		has DVDD record
18	inCr	Char	*		Incl. criterion
*	nprtyp	Char	*		Type from NPR
19	only1	Num	8		Only one criterion
13	sex	Num	8		sex

# The R-version

- ▶ An **R**-version where
  - ▶ `pnr` is **character**
  - ▶ other categorical variables are **factors**
  - ▶ date variables are `cal.yr` (fractional years)

is available in

`e:\workdata\707655\DMreg\data\DMreg.Rda`

- ▶ this **R**-file also contains a (named) vector of variable labels.



## SAS-version and R-version

- ▶ `pnr` is character in both versions
- ▶ dates in the SAS-version are numerical variables with date formats (`DDMMYY10.`)
- ▶ dates in the R-version are of class `cal.yr`, fractional years — a bit like `decimal_date` but without error
- ▶ `sex`, `DMtp` etc. are character variables in the SAS version
- ▶ these are `factors` in the R-version.

Otherwise identical contents, for practical use see:

`v:\sdc\469drive\DMreg\tex\DMreg2018-R.pdf`

## Complications files

e:\workdata\707655\DMreg\data\XXX.sas7bdat and  
e:\workdata\707655\DMreg\data\XXX.rda

- ▶ complications are derived from NPR for **total** DK population, only **first** occurrence included — except for **HypG** (hospitalized hypoglycaemia) and **Keto** (ketoacidosis)
- ▶ complications, 21 different types + 5 super categories
  - ▶ **fcomp** one record per (**pnr,comp1**), **compGr** and date, **doC**
  - ▶ **wcomp** one record per (**pnr**),  
26 date variables — one per complication
  - ▶ **rcomp** one record per (**pnr,comp1,doC**)  
dates and types of recurrent complications

## LABKA files

e:\workdata\707655\DMreg\data\labka\XXX.sas7bdat and  
e:\workdata\707655\DMreg\data\labka\XXX.rda

- ▶ LABKA data file contain 131 mil. records
- ▶ Takes forever to read both with SAS and **R**
- ▶ Read the total LABKA file with SAS, and split in 26 files
- ▶ Read each of these with **R**, store in 26 **R**-files

Details in

v:\sdc\469drive\DMreg\tex\DMreg2018-R.pdf

## LABKA files — what is XXX

e:\workdata\707655\DMreg\data\labka\XXX.sas7bdat and  
e:\workdata\707655\DMreg\data\labka\XXX.rda

hba1	"Hba1c"	sodi	"Natrium"
gluc	"Glukose"	tsh	"TSH"
glu0	"Glukose 0"	cpep	"C-peptid/Proinsulin"
gl30	"Glukose 30"	crp	"CRP"
gl120	"Glukose 120"	gad	"GAD65"
tchl	"Total kolesterol"	egfr	"eGFR"
ldl	"LDL kolesterol"	gfr	"GFR"
hdl	"HDL kolesterol"	alat	"ALAT"
vldl	"VLDL kolesterol"	alcp	"Basisk fosfatase"
trig	"Triglycerid"	cobl	"Cobalamin"
plcr	"Plasma Kreatinin"	trmb	"Trombocytter"
uacr	"Ualbcrea"	leuc	"Leucocytter"
pota	"Kalium"	hmgb	"Hæmoglobin"

# The status file

Some things change over time:

- ▶ place of residence
- ▶ family income
- ▶ attained education

This is collected for all persons in DK at 1 Jan 1996–2019

# The status file in SAS I

## The CONTENTS Procedure

Data Set Name	DMDAT.POPSTAT	Observations	131784868
Member Type	DATA	Variables	8
Engine	V9	Indexes	0
Created	11/08/2020 15:04:49	Observation Length	64
Last Modified	11/08/2020 15:04:49	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label	The population status at start of each year		
Data Representation	WINDOWS_64		
Encoding	wlatin1 Western (Windows)		

## Engine/Host Dependent Information

Data Set Page Size	65536
Number of Data Set Pages	129075
First Data Page	1
Max Obs per Page	1021
Obs in First Data Page	990
Number of Data Set Repairs	0

# The status file in SAS II

```
ExtendObsCounter      YES
Filename              E:\workdata\707655\DMreg\data\popstat.sas7bdat
Release Created       9.0401M5
Host Created          X64_SR12R2
Owner Name            DSTFSE\FDIY7655
File Size             8GB
File Size (bytes)    8459124736
```

## Variables in Creation Order

#	Variable	Type	Len	Format	Informat	Label
1	PNR	Char	12	\$12.	\$10.	personnummer
2	KOM	Char	3	\$KOM_V4_KT.		kommune
3	reg	Char	2	\$REG_V4_KT.		region
4	yr	Num	8			dato (år)
5	find	Num	8			disponibel fam. indkomst
6	udd	Num	8			uddannelseskode
7	udddk	Num	8	AUDD_HOVED_L5L5_KT.		grupperet uddannelse
8	eduen	Num	8	AUDD_LEVEL_L4L4_KT.		grouped education

`findec` — family income decile is added in the **R**-version.

**Note:** Deciles are computed per year.

```
> str( popstat, v=0 )
```

```
Classes 'tbl_df', 'tbl' and 'data.frame': 131784868 obs. of 9 variables:
```

```
$ pnr      : chr      ...  
  ..- attr(*, "label")= chr      ...  
  ..- attr(*, "format.sas")= chr      ...  
$ yr       : num      NULL ...  
  ..- attr(*, "label")= chr      ...  
$ kom      : Factor w/ 99 levels "København","Frederiksberg",...: NULL ...  
$ reg      : Factor w/ 5 levels "Nordjylland",...: NULL ...  
$ find     : num      NULL ...  
  ..- attr(*, "label")= chr      ...  
$ findec   : Factor w/ 10 levels "1","2","3","4",...: NULL ...  
  ..- attr(*, "label")= chr      ...  
$ udd      : num      NULL ...  
  ..- attr(*, "label")= chr      ...  
$ udddk    : Factor w/ 15 levels "Førskoleuddannelser",...: NULL ...  
$ eduen    : Factor w/ 9 levels "Early childhood education",...: NULL ...
```



# Reading the status file in R

```
> system.time(popstat <- read_sas("../data/popstat.sas7bdat"))
```

```
   user  system elapsed  
810.06    6.65  1903.40
```

```
> system.time( save(popstat, file="../data/popstat.Rda") )
```

```
   user  system elapsed  
493.72    5.12   571.10
```

```
> rm( popstat )
```

```
> system.time( load(                file="../data/popstat.Rda") )
```

```
   user  system elapsed  
112.73    1.56   168.81
```

## Summary

- ▶ A complete documentation of the `DMreg` is available in `v:\sdc\469drive\DMreg\tex\DMreg2018.pdf`
- ▶ A complete documentation of the `R`-version of `DMreg` is available in `v:\sdc\469drive\DMreg\tex\DMreg2018-R.pdf`  
— this document also contains accounts of the complications files, the LABKA files and the status file.

## Documentation of projects

- ▶ Large projects produce large amounts of documentation
- ▶ Construction of `DMreg` is by multiple SAS-programs
- ▶ ... the documentation is in  $\text{\LaTeX}$  format on the `v:` drive
- ▶ Documents pulled together are `.log` and `.lst` files produced by SAS, and `.tex` files produced by `R` (via Sweave / Rmd)
- ▶ It is **forbidden** to lift figures off the screen and publish them — the reason is that DST need a trail of data that you fished out from DST
- ▶ Therefore you must learn to produce files with your results, be that `.log`, `.lst`, `.tex`, `.txt`
- ▶ Once they are on your own computer you can use them

## Running SAS from command window

- ▶ open command window
- ▶ `cd` to your SAS-folder where `xxx.sas` is
- ▶ write:  
`copy \workdata\707655\DMreg\util\optslibs.sas .`  
— the space before the last dot is important
- ▶ this is a file the predefines your libnames etc.  
Look at it with notepad, and change the version you use.
- ▶ write:  
`\workdata\707655\util\start`
- ▶ write:  
`sj xxx`
- ▶ This runs the SAS-code

## Running SAS from command window

- ▶ writing `sj xxx` runs the SAS-code and gives you two new files:  
`xxx.log` and  
`xxx.lst`
- ▶ —and an absolute guarantee that what is in the `xxx.log` and `xxx.lst` is fully documented.
- ▶ there is no need to open SAS to see your results, you just open the files with `notepad` to see the results:  
`notepad xxx.log`
- ▶ be careful if you send these files home (micro data, counts < 4)
- ▶ you cannot use results in articles etc. **unless** you have sent home the files.