

Technical documentation of the BEFINN model, g2020

TALL

SOM FORTELLER

NOTATER / DOCUMENTS

2020/23

Michael Thomas, Rebecca Gleditsch and Astri Syse

In the series Documents, documentation, method descriptions, model descriptions and standards are published.

© Statistics Norway

When using material from this publication, Statistics Norway shall be quoted as the source.

Published 13 May 2020

ISBN 978-82-587-1137-4 (electronic)

ISSN 2535-7271 (elektronisk)

Symbols in tables	Symbol
Category not applicable	.
Data not available	..
Data not yet available	...
Not for publication	:
Nil	-
Less than 0.5 of unit employed	0
Less than 0.05 of unit employed	0.0
Provisional or preliminary figure	*
Break in the homogeneity of a vertical series	—
Break in the homogeneity of a horizontal series	
Decimal punctuation mark	.

Preface

The biennial production and publication of national population projections is a central task of Statistics Norway. This documentation describes the maintenance and practical applications of the underlying model for the population projections, BEFINN.

The documentation includes an overview of the model, including its components and different roles, and describes all necessary steps to run the model. As such, it works as a valuable resource to secure the production and delivery of the national population projections.

Statistics Norway, 20 May 2020

Linda Nøstbakken

Abstract

This document details the workflow and technical solutions in the BEFINN model, which constitutes Statistics Norway's main model for national population projections. Statistics Norway produces national population projections biennially. Work on model development, as well as substantive research relating to model assumptions, is undertaken in the intervening years.

The purpose of this documentation is to describe the maintenance and practical applications of the BEFINN model. In addition to the work implemented in this model, extensive external work is required in order to produce assumptions for fertility, mortality and international migration – with resulting files used as input in the BEFINN model. The external work is documented in the main report accompanying the g2020 national population projections and is therefore not discussed in this document.

A schematic overview of the model, including the programs and the input and output files, is included in Appendix A. Appendix B comprises a “cook book” detailing the order in which the various programs needs to be run. These brief descriptive summaries should be regularly consulted during the production of a new national population projection.

Contents

Preface	3
Abstract	4
1. Introduction	6
1.1. Modules and roles	6
1.2. Technical solutions.....	6
1.3. Operation of the model.....	6
1.4. Data sources	6
1.5. Base year.....	7
1.6. Start year and end year.....	7
1.7. Catalogue structure	7
1.8. Results	8
2. Preparation for a new projection	9
2.1. Archiving the last projection	9
2.2. Creation of new projection	9
2.3. Intervening years.....	9
2.4. Data sources.....	10
3. Operation of the model	11
3.1. Preparations.....	11
3.2. Overall workflow.....	11
3.3. The individual roles in BEFINN	12
4. Results, publication and post-production	20
4.1. Publishing in the StatBank	20
4.2. Control of results	20
4.3. Tables to 'Dagens statistikk' (DS) (the press release).....	23
4.4. The population projections website	23
4.5. Establishing the new production version	23
References	24
Appendix A: A flow chart of files and programs used in BEFINN	25
Appendix B: Cook book for producing new national population projections g2020 ...	26

1. Introduction

1.1. Modules and roles

The BEFINN model is built around different roles with independent areas of responsibilities. The model has four roles:

- Aggregation
- Mortality
- Migration
- Fertility

Each role is responsible for one module which together comprise the BEFINN model. The modules have been designed to be as similar as possible technically so that the lead on one role is able to take on the responsibilities and work for another role, should this become necessary. In this documentation report, the above colour coding is used throughout to facilitate readability, including figures documenting the workflow (see Figure 3.1) and a flow-chart of files and programs used in BEFINN (see Appendix A).

1.2. Technical solutions

The model is developed in SAS. Having in-depth knowledge of SAS is therefore beneficial when producing the national population projections. The model is executed on Linux, but only basic knowledge of Linux is required.

1.3. Operation of the model

In a regular update of the model, changing the input and projection years is necessary. These necessary changes are identified and easily executed using the module-specific update programs. The model draws on external files from the Population Statistics (BESTAT and BEREG), as well as other external files (`external_files` and `fixed_registers`). The format and content of these external files may change from year to year. It is thus important to ensure that all files comply with the model requirements. If not, adjustments must be made. Major external work is also necessary to produce new assumptions for fertility, mortality and migration. Description of this work is not included in this document but can be found in the main report for the national population projections 2020.

1.4. Data sources

The model uses three main sources for external files:

- Population Statistics (BESTAT and BEREG)
- Externally produced assumptions
- Fixed registers and standard classifications

For data obtained from the Population Statistics, access must be sought and granted in accordance with Statistics Norway's procedures. This data is updated annually and so the granting of access to these data must be repeatedly obtained from the producers.

Updated versions of externally calculated assumptions are used for fertility, mortality and migration, and must be provided by members of the project team before the actual projection models can be run.

Fixed registers are partly based on Statistics Norway's standard classification database, which is available from <https://www.ssb.no/en/klass/>. This includes information regarding countries and country groups. These are updated regularly, but not necessarily annually. Given the likelihood of changes it is strongly advised to obtain new versions of these data each time the model is planned to be run. The role of [Aggregation](#) is responsible for this.

1.5. Base year

Like many of Statistics Norway's other models, we apply the concept of a base year. This reflects the last whole calendar year for which actual, observed population figures have been collected. The first possible projection year will then be the year after the base year. Therefore, in all the modules in BEFINN, a variable `baseyr` (i.e. base year) will be defined that corresponds to the base year of the model.

1.6. Start year and end year

The first year after the base year is the start year (`startyr`), whereas the last projection year is termed end year (`endyr`).

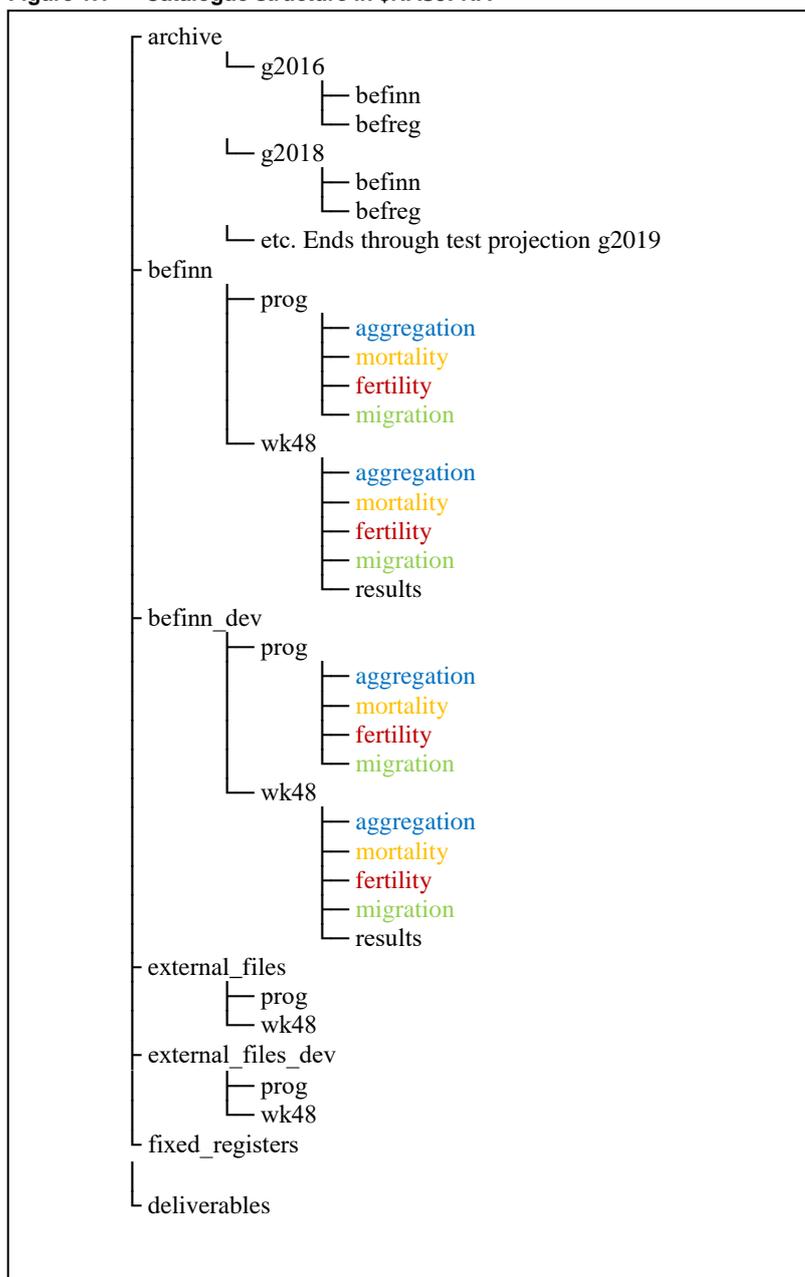
1.7. Catalogue structure

All program code and associated results from BEFINN are stored in the `$NASJFRA` library on SSB's Linux servers. Here we also store the externally produced assumptions as well as the fixed registers described in section 1.4. There are seven main directories in the library, and each of these directories has subdirectories that follow SSB's name standard. This means that all program code is stored in a directory named `prog`, while all data files are stored in a directory named `wk48`.

Figure 1.1 shows the directory structure. There are two `befinn` directories, where the one with the postscript `_dev` contains the version that is currently under development. The second, without a postscript, represents the production version, that is, the one that was last produced. The idea is that no files or programs in this production version (`befinn`) should be changed. They are only to be used as a reference or as source for the development version (`befinn_dev`) when expanding new vintages onto the datasets. For that reason, the production version should be "locked" for editing after each production, but it may be copied from.

The directories `external_files`, `external_files_dev` and `fixed_registers` contain the files that are not extracted from the population statistics, as described in Chapter 1.4. External files are used by all roles. The directory `archive` contains previous projections, both published and unpublished.

Figure 1.1 Catalogue structure in \$NASJFRA



If new catalogues and/or directories are made, it is important that all members of the core team are given access to read and (if warranted) write to the respective catalogues and directories.

1.8. Results

It is the [Aggregation](#) role that is responsible for the initial production of files required for later internal and external work – though all other roles also produce files for their own external work. Much of this output is therefore stored in each role's wk48 directory.

From the BEFINN model, two types of out-files are produced, `out1_XXX` and `out2_XXX`. The x can be H, M, L, 0, C or E, depending on the type of alternative being run. The letters represent the following alternatives: H=high; M=medium; L=low; 0=No in- or out-migration (the 0 is only used for migration); C=constant; and E=Equal in- and out-migration, i.e. zero net migration (the E is only used for

migration). The first position of *x* indicates fertility, the second position life expectancy, and the third position migration. For example, the out-file named `out1_MM0` represents a projection with medium fertility: medium life-expectancy and no in- or out-migration.

The `out1` files contain population figures, while the `out2` files contain figures that relate to the components: Fertility, mortality, and migration (immigration and emigration). Based on the output files, flat, comma-separated files are created and uploaded to Statistics Norway's StatBank. Programs for generating these data and uploading them are also part of the [Aggregation](#) role.

2. Preparation for a new projection

When the model is to be updated some preliminary preparations are required. This chapter details these preliminary preparations. Once completed, a projection can be made with a new base year, as described in the Chapter 3. It is necessary to run the model with all base years since the previous projection. This means that although we normally only publish projections biennially, the model must be run annually (see 2.3).

2.1. Archiving the last projection

First, it is important that the previous projection (which up to this point has been referred to as the production version) is filed. It should be transferred to the archives and marked with directory names stating the year when the projection was published, for example `g2018` (for the previous projection in 2018). The directory containing the files is named 'files' (not `wk48`). Otherwise the files will be deleted after 48 months. The role of [Aggregation](#) is responsible for this.

2.2. Creation of new projection

Next, the directory `befinn_dev` should be renamed `befinn` (i.e. the development version becomes the latest production version). Next, a new `befinn_dev` directory must be created, including all sub-directories. All sub-directories should be empty. The role of [Aggregation](#) is responsible for this. The new projection will be placed in the development catalogues (`wk48`). Next, each role is responsible for copying all the programs required from the role's sub-directory in the `befinn` directory, to the appropriate place in the newly created `befinn_dev` directory. The sub-directories for programs (`prog`) should then contain the files that exist in the production version. Once [Aggregation](#) has run the two update programs, the other roles can start work on making new projections.

2.3. Intervening years

The population projections are normally run every two years. As such, if there has been no intervening run between the two years, data for one year will be missing from the files. As stated in the introduction of this chapter, it will therefore be necessary to update with not just one, but two years of data, by running the updates twice with different base year datasets. However, in most cases developments are made also in the intervening years to improve the model. As such, it is usually not necessary to update with two years of data. These developments are logged and documented in the project-specific area on the S-drive in the administrative zone (`S:\Prosjekt\2005201-Nasjfram\`).

2.4. Data sources

As mentioned in Section 1.4, the models use data from three different sources. Updated versions of all data are needed to establish new versions of the models.

Population Statistics

The data from Population Statistics (BESTAT) are updated with new information annually. Accordingly, access to these data must be obtained from the producers.

For the BEFINN model, the datasets required are listed in Table 2.1, together with the specific role they relate to. Table 2.1 uses the projection in 2020 as an example. Whether datasets are drawn from the `archive` or `wk` directories will depend on BESTAT. The data format may vary from dataset to dataset, and across years. Datadok must be used to assess whether changes have been made, prior to the usage of any data from BESTAT. Each role is responsible for assessing any changes relevant to the role. Furthermore, depending on the timing of the projection, the datasets may also vary in terms of how they are stored, as either SAS-files or .dat files.

Table 2.1 Population Statistics files used in BEFINN

Used by	Dataset (mix of .dat- and SAS-files)
Aggregation	\$BEREG/snr_kat/ \$BESTAT/folkem/.../aggbos/g20XXm01d01
	\$BESTAT/flytt/.../g20XX
	\$BESTAT/innvandr/.../bosatt/g20XXm01d01
	\$BESTAT/innvandr/.../alle_fnr/g20XXm01d01
Fertility	\$BESTAT/fodte/.../g20XX
	\$BESTAT/innvandr/.../g20XXm01d01
Mortality	\$NASJFRA/fixed_registers/country_cat \$BEREG/snr_kat/
	\$BESTAT/dode/.../g20XX
Migration	\$NASJFRA/fixed_registers/country_cat
	\$BESTAT/innvandr/.../g20XXm01d01
	\$NASJFRA/fixed_registers/country_cat

None of the files should be copied anywhere, and they should only be accessed from where they are located (as specified in the table). This is to avoid double storage and any subsequent inconsistencies. The role of **Aggregation** is responsible for ensuring that `country_cat` is updated according to class (in Norwegian, `klass`), see <https://www.ssb.no/en/klass/>.

External files

These assumptions are calculated outside of BEFINN and delivered as external files to BEFINN.

Fertility:

- `fertility_x`
- `two_imm_par`

The files relating to the first bullet point contain the projected age-specific fertility rates in different alternatives. Here the alternatives (x) are low (l), medium (m) and high (h). The file relating to the second bullet point contains assumptions about the share of immigrant women who have children with an immigrant man, by the woman's country group of origin. Here only one alternative is made. The file is used to ascertain the projected numbers of Norwegian-born children of two immigrant parents in the overall projection output.

Mortality:

- `q_women_x`
- `q_men_x`

In both cases, the x alternatives are low (l), medium (m), high (h) or constant (c) and denote life expectancy. The constant alternative (c) is the first-year projected probabilities kept constant for men and women at every age level for all subsequent years.

Migration:

- `imm_x`

Here the x alternatives are low (l), medium (m), high (h), constant (c), zero net migration (e) and no in- or out-migration (0), and refer to the total number of immigrations.

In preparation for the new model, all the above must be updated with new versions. It is crucial to manually check the contents of these datasets to see if they are in the same format as previous deliveries.

Fixed registers

As part of the preparations for a new model version, checks are required to determine whether it is necessary to update the fixed registers. Since these are not changed every year, this may not be necessary. Since all roles use the fixed registers, it is natural that the role of [Aggregation](#) is responsible for updating them.

So long as the fixed registers are flat files, it is easiest to change them outside the SAS EG-system. Open Linux and use one of the SAS-servers. Find the following directory:

```
$NASJFRA/fixed_registers/
```

Enter «*nedit filename*» so that the file can be edited. Save afterwards and the register is updated.

Each role must then itself update the source code belonging to the role's area of responsibility if new versions of the fixed registers are available and these have been renamed.

3. Operation of the model

This chapter goes through the practical part of operating the model. The chapter's focus is on how parameters and other assumptions are specified and results are generated. The workflow during a projection is also described.

3.1. Preparations

Before work on a new model version can begin, all preparations described in Chapter 2 must be completed.

3.2. Overall workflow

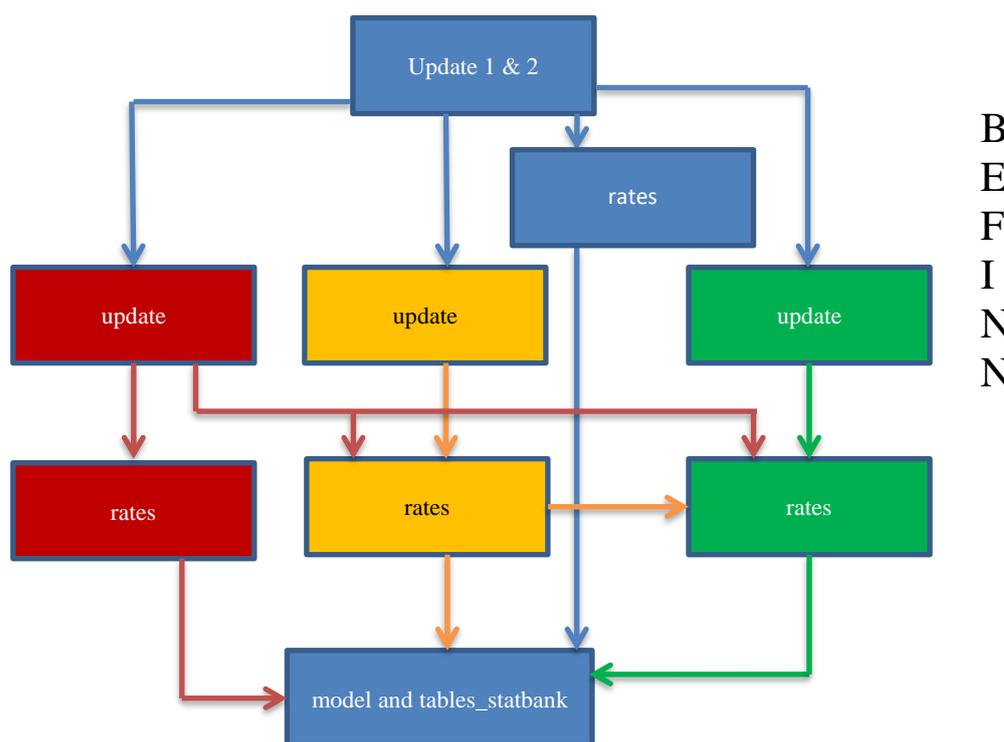
A new projection starts with the role of [Aggregation](#) updating the database with new annual information covering the period since the previous projection. Thereafter, the other roles may update the baseline data of their respective responsibilities as well as generate new rates, probabilities and figures.

With a few minor exceptions, there will be two different programs to run for each role, `update.sas` and `rates.sas`. They should be run in this order. Figure 3.1 shows the workflow in the projection work as well as the dependence between the

different modules. As the figure shows, the different roles are to some extent dependent on the other roles running their programs first. The programs must therefore be run in a relatively fixed order. A summary of this order can be found in Appendix B (*Cook book for the national population projections*). In Appendix A we show a detailed flow chart of how the different roles' programs depend on files produced by other roles' programs.

When all the rates and probabilities are available, and the external assumptions are in place, **Aggregation** can run the `model.sas` program, the `outfilecheck.sas` program and the `tables_statbank.sas` program and produce national projections.

Figure 3.1 Overall workflow in the model work



3.3. The individual roles in BEFINN

There are four roles in BEFINN: **Aggregation**, **Fertility**, **Mortality** and **Migration**. The work assignments of each role are explained below.

Aggregation

This work assignment is twofold, as the role is operated both before and after the other roles carry out their work.

Update. In the first phase, new population records that have been added since the last projection will be collected. This information should be formatted in such a way that it can be read by the programs used by the other roles.

Input data is taken from the Population Statistics (BESTAT) and the previous `befinn`. First, `update_1.sas` is run. The program uses two files:

- `kommuner` from the previous `befinn`
- `$BESTAT/folkem/.../aggbos/g20XXm01d01`

The program produces one permanent dataset:

- \$NASJFRA/befinn_dev/wk48/aggregation/nation

The dataset `nation` contains the total population for all cohorts for all years from 2006 to the base year.

Next, the program `update_2.sas` is run. This program reads information from the previous `befinn` and from the file `nation` created by `update_1` above:

- \$NASJFRA/befinn_dev/wk48/aggregation/nation

In addition, information is read from the following files:

- \$BESTAT/innvandr/.../bosatt/g20xxm01d01

This dataset contains information on all residents (with immigrant characteristics) at the end of the base year.

- \$BESTAT/innvandr/.../alle_fnr/g20xxm12d31

This dataset contains more detailed information about all residents.

- \$BESTAT/flytt/.../g20xx

This dataset contains more detailed information about migration, such as the country that migrants move to or from, and the year of migration.

Two datasets are read from the fixed registers, discussed in Section 2.4:

- \$NASJFRA/fixed_registers/country_cat
- \$BEREG/snr_kat/

The first dataset is a country catalogue that determines which country group a country belongs to. This country catalogue is our own file, wherein we classify all countries in the world into three broad country groups. It is necessary to check this file every year in case new countries are added to the original file, see <https://www.ssb.no/en/klasse/>. The second dataset shows, for each person in the population, current and past identification numbers as well as associated statistics numbers (s-nummer).

Table 3.1 details the data files that are produced, and which are subsequently used as input for [Aggregation](#), as well as other roles.

Table 3.1 Input files for BEFINN in the initial stage

Name	Description	Role
nation	Produced by <code>update_1</code> , contains the population by sex and one-year-age-group from 2006-base year.	Aggregation
stock	Number of persons in the base year, by age, sex, country of origin, immigration category and duration of stay	Aggregation
emigrations_red	Emigrations per year from 1990, by sex and age	Aggregation
immigrations_red	Immigrations per year from 1990, by sex and age	Aggregation
pop_ctrgr_immcat	Population by age, sex, immigration category and country group, back to 1989	Aggregation
immigrants_dur_stay	All immigrants (and their children) who have immigrated since 1990	Fertility Migration
immigrations_dur_stay_red	All immigrations (including those of children) that have taken place since 1990, reduced for repeated migrations by same individual	Migration
emigrations_dur_stay_red	All emigrations from Norway since 1990, reduced for repeated migrations by same individual	Migration
immigrants	All immigrants living in Norway for each year since 1990	Migration

Model. When all external work is completed, the model should be run. Input is taken from all roles as shown in Table 3.2. The results files that are generated are shown in Table 3.3.

Tabell 3.2 Input files for BEFINN after processing

Name	Description	Role
birth_l		
birth_m		Fertility
birth_h	Birth rates in three alternatives	
two_imm_par	Percentage born to two immigrant parents, assigning mother's country group of origin	Fertility
death_l		
death_m		
death_h		
death_c	Death probabilities in four alternatives	Mortality
emig	Emigration probabilities	Migration
immig	Distribution of immigrants by age, sex and duration of residence	Migration
imm_l		
imm_m		
imm_h		
imm_c		
imm_e		
imm_0	Immigrations figures in six alternatives	Migration
stock	Number of persons in the base year, by age, sex, country group, immigration category and duration of stay	Aggregation

The model is run 15 times with different combinations of L, M, H, C, E and 0:

- MMM Medium national growth (main alternative)
- LLL Low national growth
- HHH High national growth
- LMM Low fertility
- HMM High fertility
- MLM Low life expectancy
- MHM High life expectancy
- MML Low immigration
- MMH High immigration
- LHL Strong population ageing
- HLH Weak population ageing
- MME Zero net immigration
- MM0 No migration (closed borders)
- MKC No growth in life expectancy
- MMC Constant immigration

Using a macro in the BEFINN model program, all options can be run simultaneously.

For each of the fifteen combinations, the outputs in Table 3.3 are produced.

Table 3.3 Output from BEFINN

Name	Description
out1_xxx	Population by age, sex, year, country group, immigration category and duration of stay
out1_nat_xxx	Population by age, sex and year
out1_nat_integer_xxx	Population by age, sex and year – rounded
out1_ctrgr_gen_xxx	Population by age, sex, year, country group and immigration category
out1_ctrgr_gen_integer_xxx	Population by age, sex, year, country group and immigration category - rounded
out2_xxx	Components by age, sex, year, country group, immigration category and duration of stay
out2_nat_xxx	Components by age, sex and year
out2_nat_integer_xxx	Components by age, sex and year – rounded
out2_ctrgr_gen_xxx	Components by age, sex, year, country group and immigration category
out2_ctrgr_gen_integer_xxx	Components by age, sex, year, country group and immigration category – rounded

Tables to Statistics Norway's Statbank. After the model has been run in fifteen alternatives, tables to be uploaded into the Statbank can be created using the program `tables_statbank`. Before this program is run, new tables must be created in Statbank. This is done in collaboration with the Communications Department at Statistics Norway. Contrary to other producers of Statbank figures,

the national population projections creates new tables each year, instead of updating the old versions. The main reason for this is that we keep the old projection results available in Statbank. Furthermore, this ensures that users who refer to a specific table will find the specific figures in the same table even though new projections have been made. The tables that are created also need some registered (historical) population figures. Input is obtained from different roles to produce the tables in Table 3.4. After the tables have been successfully uploaded to the Statbank, all tables must be checked carefully in the internal Statbank. A project specific username and password is required, and can be obtained from the Communications Department. See section 4 for further details.

Table 3.4 Tables for StatBank from BEFINN

Table	Input data	Description of input data	Role
Table 1. Projected population, by sex, age, immigration category and country background, in 15 alternatives	pop_ctrgr_immcat	Registered (1989-base year) population by age, sex, year, country group and immigration category	Aggregation
	out1_ctrgr_gen_xxx	Projected population by age, sex, year, country group and immigration category	Aggregation
Table 2. Projected number of immigrants, by country background and time of residence, in 5 alternatives	stock	Registered (base year) population by age, sex, country group, immigration category and duration of stay	Aggregation
	out1_xxx	Projected population by age, sex, year, country group, immigration category and duration of stay	Aggregation
Table 3. Projected population changes, by immigration category and country background, in 9 alternatives	birth_agg	Registered number of births by country group for the ten previous years (up to and including base year)	Fertility
	death_ctrgr	Registered (1990-base year) number of deaths by country group	Mortality
	immigrations_ctrgr	Registered (1990-base year) number of immigrations by country group from base year	Migration
	emigrations_ctrgr	Registered number (1990-base year) of emigrations by country group from base year	Migration
	out2_ctrgr_gen_xxx	The components projected by country group, immigration category and year	Aggregation
Table 4. Projected fertility, by immigration category and country background, in 3 alternatives	pop_ctrgr_immcat	Registered (1989-base year) by age, sex, year, country group and immigration category	Aggregation
	mothers	Individual dataset with mothers who give birth in a given year and their characteristics (2007-base year)	Fertility
	out1_nat_xmm	Projected population by age, sex and year	Aggregation
	out2_nat_xmm	The components projected by year	Aggregation
	out1_ctrgr_gen_xmm	Projected population by age, sex, year, country group and immigration category	Aggregation
	out2_ctrgr_gen_xmm	The components projected by country group, immigration category and year	Aggregation
Table 5. Projected life expectancy, for men, women and both sexes combined, by age in 3 alternatives	women_base_LC	Estimated life expectancy (period), for men, women and both sexes combined, by age in 3 alternatives. The files are generated in R, and are one of four tables in StatBank that are not created using the files from BEFINN	Mortality
	women_l		
	women_m		
	women_h		
	men_base_LC		
	men_l		
	men_m		
	men_h		
	both_base_LC		
	both_l		
	both_m		
	both_h		
	Table 6. Projected death probabilities per 1,000, by sex and age, in 3 alternatives.		
q_women_m			
q_women_h			
q_men_l			
q_men_m			
Table 7. Projected cohort life expectancy, by age and sex, medium alternative	cohort_ex_m_men	Estimated life expectancy (cohort), by age and sex, medium alternative. The files are generated in R.	Mortality
	cohort_ex_m_women		
Table 8. Projected population with formal uncertainty ascertainment, in 5 percentiles	stoch_pop	Projected population by age, sex and year, in 5 percentiles. The table is made outside of BEFINN as a special delivery for the g2020 projections.	Other

The result is eight flat files that are automatically uploaded to StatBank, using the program `tables_statbank` in the [Aggregation](#) role. For g2020, Table 8 is an additional product resulting from a stochastic population projection. All files are stored at `$NASJFRA/befinn_dev/wk48/results/statbank/`, and their respective names are:

- Table 1. Population stock: `framskr2020t011.dat`
- Table 2. Immigrants: `framskr2020t031.dat`
- Table 3. Population changes: `framskr2020t041.dat`
- Table 4. Total fertility rate (TFR): `framskr2020t051.dat`
- Table 5. Period life expectancy: `framskr2020t061.dat`
- Table 6. Probabilities of death: `framskr2020t081.dat`
- Table 7. Cohort life expectancy: `framskr2020t091.dat`
- Table 8. Stochastically projected population: `framskr2020t101.dat`

In addition, six datasets of time series are created with registered and projected population figures that can be used in articles to check the results. These files are stored at `wk48/results/time_series` and include:

- `time_series_table_1`
- `time_series_table_2`
- `time_series_table_3`
- `time_series_table_4`
- `time_series_table_5`
- `time_series_table_6`

Fertility

Update and rates. The role of the fertility programs is to produce age-specific fertility rates by country group, immigration category and duration of residence in three alternatives. This is done in the programs `update` and `rates`. The program `rates` also draws on the program `smoothing`, which smooths the age-specific fertility rates.

Input data are obtained from various sources. From the Population Statistics (BESTAT), two files are utilised:

- `$BESTAT/fodte/arkiv/statfil/g20xx.dat`
- `$BESTAT/innvandr/arkiv/alle_fnr/g20xxm12d31.dat`

The first file contains information about those born last year. The second file contains general information about the residents last year.

A dataset produced by the role [Aggregation](#) (`update_1`) is also used:

- `Aggregation/nation`

It provides the total population in each cohort from 2006 through to the base year.

The **Fertility** `update` program creates four files:

- `gender_nation`
- `birth_agg`
- `mothers`
- `births`

The first and second files are used as input for the **Migration** rates program. The second and third files are used by `tables_statbank`, while the third and fourth files are used by the **Fertility** rates program.

A further dataset is obtained from the role the **Aggregation**, using the program `update_2`:

- `Aggregation/immigrants_dur_stay`

In addition, four datasets produced by the **Fertility** role are used in collaboration with the reference group:

- `External_files_dev/wk48/fertility_assump_1`
- `External_files_dev/wk48/fertility_assump_m`
- `External_files_dev/wk48/fertility_assump_h`
- `External_files_dev/wk48/two_immpar_assump`

The first three files contain national fertility assumptions in three alternatives. For each alternative, there is an annual factor that adjusts the age-specific fertility rates up or down. The files are produced by the role of **Fertility** in collaboration with the reference group. The fourth file contains assumptions about the share of immigrant women who have children with an immigrant man, by the woman's country group of origin, during the projection period.

The end result is three files that contain age-specific fertility rates by country group, duration of residence and immigrant category in three alternatives by year from the base year to the end year. In addition, a file is created with assumptions about the proportion of births to two immigrant parents, with origin group based on the mother's country group of origin. The four files are:

- `birth_l`
- `birth_m`
- `birth_h`
- `two_imm_par`

These four datasets are used by the role of **Aggregation** in the cohort component model itself. In addition, Table 3.5 details additional datasets that are used in the other roles.

Table 3.5 Output files from the fertility role used by other roles

Name	Description	Role
<code>birth_agg</code>	Number of births in country group 1-4 from 2007 to base year	Aggregation Migration
<code>Mothers</code>	Individual dataset with mothers who give birth the specific year and their characteristics from 2007 to base year	Aggregation

Mortality

Update and Rates. The **Mortality** `update` program first and foremost creates files (`death_women`, `death_men`, `pop_women` and `pop_men`) that are used as input in the external mortality work (i.e. outside BEFINN). However, it also creates a file of deaths by country group that is also used by the **Aggregation** role:

`death_ctrgr`

The input data for this file are taken from the Population Statistics (BESTAT) on deaths from the previous year.

The responsibility of the **Mortality** role is to produce age-specific death probabilities in four alternatives as well as the `death_ctrgr` used by [Aggregation](#).

Most of the work in death probabilities is done externally (i.e. outside BEFINN) by the role **Mortality**.¹ Six datasets are created that will be added to `external_files_dev/wk48/mortality/`. These files are:

- `q_women_l`
- `q_women_m`
- `q_women_h`
- `q_men_l`
- `q_men_m`
- `q_men_h`

Using the **Mortality** `rates` program, these death probabilities are converted to four files containing annual age-sex-specific death rates in four alternatives from the year after base year to end year:

- `death_l`
- `death_m`
- `death_h`
- `death_c`

The letters l, m, h and c refer to life expectancy, not mortality. This applies throughout the mortality component. These four datasets are used by the [Aggregation](#) role.

Life expectancy files are also generated externally. There are 12 SAS-files loaded to `external_files_dev/wk48/mortality/`.

The 12 SAS-files are:

- `women_l`
- `women_m`
- `women_h`
- `women_base_LC`
- `men_l`
- `men_m`
- `men_h`
- `men_base_LC`
- `both_l`
- `both_m`
- `both_h`
- `both_base_LC`

The external probability of death and life expectancy files are used by the [Aggregation](#) role in the program `tables_statbank`.

Migration

The role of **Migration** is to produce immigration figures (done externally, but basic figures are produced in this model), emigration probabilities and proportions for the distribution of immigrants by sex, age and duration of residence. This is done using the programs `update` and `rates`.

¹ Currently in collaboration with Dinh Q. Pham (Methods Section).

Update. External work requires updated figures on immigration to Norway. These are created in the **Migration** update program. Input data are obtained from **Aggregation** in BEFINN.

- immigrants
- immigrations_dur_stay_red
- emigrations_dur_stay_red
- pop_ctrgr_immcat

The program creates tables to be used for external work on migration projections – time series of immigration, emigration and resident immigrants, by country group:

- immig_ctrgr
- emig_ctrgr
- immig_ctrgr_31dec

In addition, tables are made for StatBank with immigrations and emigrations by country group:

- emigrations_ctrgr
- immigrations_ctrgr

Rates. The **Migration** rates program creates two datasets: one with rates for the distribution of immigrants arriving in the country (by age, sex, duration of residence, etc.), and a second with emigration probabilities (by age sex, duration of residence, etc.).

The program uses four files from **Aggregation**'s update program `update_2`:

- immigrants_dur_stay
- immigrations_dur_stay_red
- emigrations_dur_stay_red
- nation

Two further datasets are retrieved from the **Fertility** update program:

- gender_nation
- birth_agg

The latter file contains the time series for the number of births.

The end result of the `rates` program is two files containing the immigration distribution and emigration probabilities, respectively:

- immig
- emig

The immigration assumptions are generated externally, and iterations are run using the `model` program to account for network effects of migration. In the final stage, altogether six immigrations alternatives are loaded as `SAS_files`, and used in the `model` program:

- imm_x

Here the `x` alternatives are `l`, `m`, `h`, `c`, `e` and `0`, and refer to the total number of immigrations.

4. Results, publication and post-production

4.1. Publishing in the StatBank

When a new model version is available, the results will be published in Statistics Norway's StatBank. We have one program for transferring tables into the StatBank, `tables_statbank`, located in the role of [Aggregation](#).

Before running the StatBank program, it is important to contact the communications department to get new table numbers / table ids (otherwise the old numbers may be overwritten). Unlike all other publishers of Statbank tables at Statistics Norway, we require new table identifiers with each publication in order to prevent the first few years from the previous projection being left in the system when publishing new projections. Unique table identifiers also enable us to have results from previous projection rounds available in the Statbank.

When the new table identities are ready, three places in the program must be changed for each of the tables:

- At the beginning of the program, the last part of FILENAME is changed for all the tables, so it points to the new table identities
- In the 'load' macro, two changes for each table are specified:
 - 'load file' must be renamed (often this is the new name on the main table, but in lower case and with '1.dat' in the end)
 - 'main table' must be renamed

We can edit table titles, footnotes etc. in StatMetaEdit. This is performed in the production zone, by looking up 'StatBank administration' and logging in with the same username and password that we use to load and check tables in the internal StatBank.

4.2. Control of results

It is crucial to check the results of new projections as well as any changes that have been made to the model. This can be done in several different ways, depending on the type of files to be checked and in which programs the check is to be performed. Below we discuss some methods for checking.

Check routines for new projections

When producing a new projection, one should check both the files produced during the process as well as the output files for publication.

The logs. When the programs are run in SAS, a log is generated. This should always be checked for error messages and notifications.

Checking for updated time series. The results of the update programs in BEFINN can in many cases be checked against the published population figures in the StatBank. This is especially true for the population, but also the total numbers for the components (e.g. deaths, births and migrations can be found in StatBank).

From both SAS and StatBank you can export files to Excel and compare the figures there (remember that the figures used in BEFINN refer to 31.12, while those published in the StatBank refer to 1.1).

Checking rates and probabilities. Rates and probabilities to be included in the model programs can be compared to the corresponding rates / probabilities used in

the previous projection. There may be large files here and it is advised, for example, to select individual years or individual groups. For fertility, we check the total fertility rate (TFR) and for mortality we check life expectancy at birth (e_0) as well as remaining life expectancy across a range of ages (e.g. e_{60} , e_{70} and e_{80}). As an example, the program `ratescheck`, which is located in `befinn_dev/prog/migration/checks`, provides an example of how checks on migration rates can be done. The numbers generated can again be taken over into other programs, such as Excel, and studied further there.

Checking the effect of new in-files. It is relatively easy to check how new input files influence for the results, especially if we are interested in the overall effect on the population. One can simply run the model program twice, once with new files and once with the files from the previous projection. This can be done by changing the LIBNAME at the start of the model program such that it refers to the directory that is not `_dev`.

Checking the StatBank files. After all model input and output files are carefully checked, the files for the StatBank should be checked. This can be done in two ways.

The first way of checking uses Stata software prior to any files being loaded into the StatBank. The program files (.do files) plot the output of all tables that are to be published. It enables a visual check of fractures and peculiarities in the time series. The files for the check are named after the tables in StatBank and are located at `$NASJFRA/checks/WK48/prog/tabX.do`. The results (figures) are added to `$NASJFRA/checks/wk48/figures/tabX/`. These figures can be exported to the administrative zone using file transfer options. To export, go into Linux and into the relevant directory you want to move files from. From here, you must enter: `sftp filsluse`. You will then be prompted to enter a Linux password. To put all the image files in the lock, type `put *.png`. Then transfer the files to: `S:\Prosjekt\2005201-Nasjfram\checks\g20XX\tabX`. This approach requires manual checking of figures and can therefore be quite time consuming.

The second approach to checking StatBank files requires them to first be loaded into the StatBank (using the `tables_statbank` program). After the files have been uploaded, they can be checked in the internal StatBank (in the administrative zone). After log-in with username Last590 and the associated password, the unpublished tables will appear.

You can access both new and old projection results in the internal StatBank, but you cannot compile tables from different projections in this program. However, this should be done, and so respective tables must be moved to other programs (for example, Excel). The basis year of the projection will usually represent observed figures for the last registered year. These figures should be cross checked. This can be done for population, births, total fertility figures, deaths and net migration. Due to multiple migrants, the figures for gross immigration and emigration will not be comparable.

To change footnotes / contacts etc in the StatBank tables, we must use the *StatMetaRediger* tool. It is in the production zone. In the Start menu, look up StatBank Administration (a program). Log in with Last590 and the usual password and select *StatMetaRediger*. Then find the appropriate table in a (very long) list in the box at the top left. Text in the green boxes is subject to change. Similarly, we select *StatDataSlett (tid)* to remove years from the tables, such as historical years we will no longer include.

Checking the out-files. The output files can be compared to the previous projection files in SAS. Example programs to do this (`outfilecheck`) are located at `befinn_dev/prog/aggregation`.

Another method of checking out-files is to run the check program `befinn_dev/prog/aggregation/sjekking/befinnsjekk.do` in Stata. These programs read each of the nine model alternatives at the population level (`out1`) and component level (`out2`) and then plot developments over time and warn of negative and missing numbers. The results from `befinnsjekk.do` are stored in the same directory as the program under `befinn_komponent.pdf`, `befinn_nuller.pdf` and `befinn_folkemengde.pdf`.

Checking if the components output match the components input. For some of the components, the files produced in the model programs can be used to check if the components match with the respective assumptions that were added to the model.

For immigration this is particularly easy, as the assumptions that are entered are absolute numbers for the number of immigrants. This figure should be in line with the figures for the number of immigrants each year that are published in the StatBank table of projected population changes.

For fertility, it is important to check whether the TFR predicted by the model program for all women combined is in reasonable agreement with the TFR development that was assumed for native women.

Effect of model changes

When considering a change in the model, it is often desirable to examine whether the results are affected by the change.

In SAS, two datasets can be easily compared using the `PROC COMPARE` procedure to uncover any differences. The example below shows how this is done for new and old versions of the same file.

```
LIBNAME old
' $NASJFRA/befinn_dev/wk48/fertility/old';
LIBNAME new ' $NASJFRA/befinn_dev/wk48/fertility';

PROC COMPARE BASE = birth_m COMPARE =
new.birth_m;

RUN;
```

In the example above, SAS will respond with an overview that tells you how many observations are equal (`Equal`), possibly different (`Unequal`). Note that only variables that are common to both datasets are compared with each other, the others are excluded from the comparison.

The use of old projections

In some cases, it may be interesting to use files from old projections, to check differences with new files from the current projection. As described earlier in this documentation, files and programs from the previous projection will, as a general rule, be found in the directories without the `_dev` extension. Other projections back to the 2012 projection (the first restructured version) can be found in

\$NASJFRA/archive. The final results from older projections are also available in the Statbank and can be accessed directly.

Evaluation of projection accuracy

A system has been created for evaluating the population projections in terms of how accurate their estimates are. The procedure for this is described in more detail in the document *Manual for evaluating*, which can be found at `S:\Prosjekt\2005201-Nasjfram\evaluations`. So far, one national level publication, unfortunately only in Norwegian, has resulted (see Rogne, 2016).

4.3. Tables to 'Dagens statistikk' (DS) (the press release)

The tables that are posted on the main projection webpage (<https://www.ssb.no/folkfram>) are created in Excel. The reason we cannot use the table builder program, which is typically used to produce DS tables at Statistics Norway, is that the program is unable to update annual figures in the future. An Excel file must therefore be created for each table. The files are located at:

```
S:\Prosjekt\2005201-Nasjfram\publications
```

When a new projection is ready, the tables must be updated with new figures, which are extracted from the internal StatBank. In general, to enforce that projections are inherently uncertain, we attempt to produce rounded figures in these tables. When the Excel files are ready, they are sent to the platform that generates the finished tables. These can be previewed at the abovementioned webpage.

4.4. The population projections website

The population projection website is <https://www.ssb.no/folkfram>. All information on this website must relate to the current production, and information regarding previous productions must be clearly labelled as such. Documentation reports and/or other reports must be available at time of production.

We can edit parts of this page, such as 'About the statistics', contact information, publication date etc. To do this, first search for 'Dashboard fag' (dashboard subjects) on Byrånett (in the administrative zone) and click on the link suggested. Log in with your own username and password. In the list that follows, find 'folkfram'. Next to 'folkfram' you can enter and change the 'About statistics/Om statistikken' (English and Norwegian).

4.5. Establishing the new production version

When a new projection is completed, and the results published, the programs and files will no longer be in the development version, but in the production version. This must be done to ensure that published data, as well as programs used to generate these, are always available and accurate.

First, it is crucial to ensure that the existing production version is filed, as described in section 2.1.

The `befinn_dev` and `external_files_dev` directories are then changed to `befinn` and `external_files`, and new `befinn_dev` and `external_files_dev` directories can be created. Programs to be used further in the next projection are copied from the (new) production version to the (new) development version. Publications pertaining to previous projections are also available on the website (e.g. Syse et al. 2018).

References

- Rogne, A. F. (2016). Hvor godt treffer befolkningsframskrivingene? *Economic Surveys* 3/2016. Statistics Norway. [In Norwegian]
- Syse, A., Leknes, S., Løkken, S. & Tønnessen, M. (2018). Norway's 2018 population projections: Main results, methods and assumptions. *Reports 2018/22*. Statistics Norway.

Appendix A: A flow chart of files and programs used in BEFINN



Appendix B: Cook book for producing new national population projections g2020

- 1) All the programs needed are found at \$NASJFRA. If there has been changes in the country groups, take this into account by updating the country catalogue dat-file ('country_cat', in \$NASJFRA/fixed_registers/wk48).
- 2) Run the programs needed to a) update base-year population, b) do the external work on fertility, mortality and migration assumptions, and c) calculate rates for emigration and distribution of immigrations by age and sex. Before the programs are run, make sure the files from BESTAT are ready and ordered in the same way as earlier, see section 3.3 in this document and Datadok.
- 3) Before the program `tables_statbank` is run, new tables must be created in StatBank. This is done in collaboration with the Communications Department at Statistics Norway. Contrary to other producers of StatBank figures, the national population projections creates new tables each year, instead of updating the old versions. The main reason for this is that we keep the old projection results available in StatBank. Furthermore, this ensures that users who refer to a specific table will find the specific figures in the same table even though new projections have been made.

First, calculate base-year population:

- `befinn_dev/prog/aggregation/update_1*`

Calculate numbers of mothers and births:

- `befinn_dev/prog/fertility/update**`
- `befinn_dev/prog/fertility/rates` (without assumptions, must be run before the external work. This program calls upon a program called `smoothing.sas` which must also be in the appropriate folder)

After * is run:

Calculate numbers of population and deaths:

- `befinn_dev/prog/mortality/update`

Create datasets on immigrations/emigrations and immigrants/emigrants:

- `befinn_dev/prog/aggregation/update_2***`

After *** is run, calculate numbers of migrations and migrants (needed for external work on migration assumptions):

- `befinn_dev/prog/migration/update`

After ** and *** is run, calculate rates for emigrations and distributions of immigrants (without assumptions, must be run before the external work):

- `befinn_dev/prog/migration/rates`

- 4) EXTERNAL WORK (on mortality, fertility, immigration assumptions)
- 5) After the external work, run these programs:
 - `befinn_dev/prog/fertility/rates` (with assumptions)
 - `befinn_dev/prog/mortality/rates`

- [befinn_dev/prog/migration/rates](#) (with assumptions)

Next, the cohort-component model can be run along with the last programs, in this order:

- [befinn_dev/prog/aggregation/model](#) (all alternatives)
- [befinn_dev/prog/aggregation/outfilecheck](#) (checking)

Before the [tables_statbank](#) is run, mortality must check that Tables 5-7 are in a format that can be used by the [tables_statbank](#) program. If not, the program will not run. It is important to first run the program without the load macro, otherwise the tables will be uploaded before any checks have been made.

- [befinn_dev/prog/aggregation/tables_statbank](#)

After the tables have been successfully uploaded to the StatBank, all tables must be checked carefully in the internal StatBank. A project-specific username and password is required and can be obtained from the Communications Department.