

Reloc-Age

Contents

1	About	5
2	Introduction	7
2.1	Research questions	7
3	Overview of Registers	9
4	Description of data sets from SCB	13
4.1	Population	13
4.2	Lisa	13
4.3	Housing	13
5	Data cleaning and joining of raw data	15
5.1	SCB data	15
6	Descriptives by Age and Sex	17
7	Housing data	19
7.1	Specific data considerations	19

Chapter 1

About

This document outlines the aims, motivations, and research questions of the RELOC-AGE project. Descriptions on data retrieval, data processing, as well as some exploratory analysis and initial findings are also reported.

Using the sidebar found on the left, one may easily navigate the document to browse the contents.

Chapter 2

Introduction

RELOC-AGE: How Do Housing Choices and Relocation Matter for Active and Healthy Ageing?

To generate novel and significant knowledge on housing choices and relocation as related to active and healthy ageing, the objectives of this multistage mixed methods participatory project are to:

- Study housing choices, relocation and health patterns in the Swedish population aged 55+ (Register RELOC-AGE).
- Study housing choices and relocation and examine the effects on active and healthy ageing among people aged 55+ considering relocation (Prospective RELOC-AGE).
- Complete the development of a novel housing counselling intervention and a subsequent pilot study (Intervention RELOC-AGE).
- Contribute to theory development (Theory RELOC-AGE).

2.1 Research questions

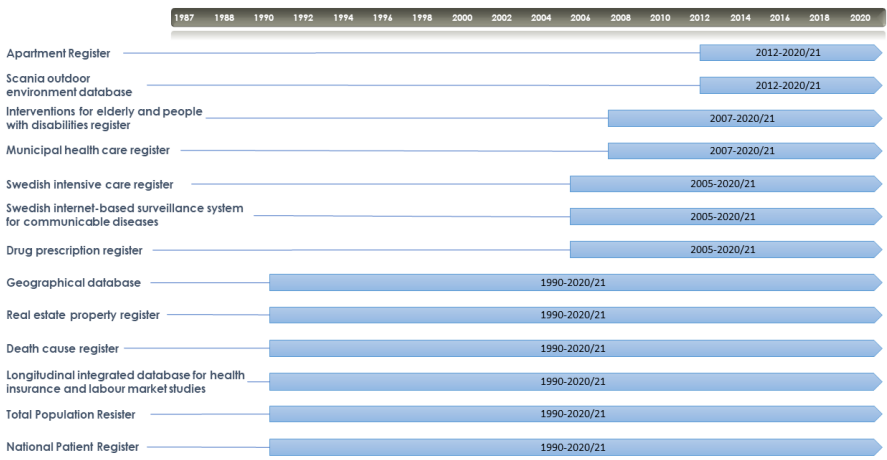
1. What are the trends over time and by age when it comes to housing types and tenures?
2. How do housing aspects and relocations affect future health outcomes?
 - How are these patterns affected by age, sex, civil status, country of origin, adverse health events, loss of a partner, socio-economic and neighborhood characteristics?
 - Given equal propensity of relocation based on baseline demographic, socio-economic and health characteristics, how do specific housing decisions affect future health outcomes?
3. What are the effects of adverse health events on housing choices and relocation patterns?

- What are the short- and long-term effects?
 - How do these effects differ between men and women, across different disease and/or disability profiles, civil status, country of origin and socioeconomic status?
4. What aspects of housing and health predict:
 - relocation to different housing options in the ordinary housing stock
 - relocation to residential care facilities
 - remaining in the present dwelling?
 5. How is the complex interaction between objective and perceived aspects of housing and social aspects associated with active and healthy ageing, and what are the characteristics and trajectories of such dynamics?
 6. What housing attributes do older adults considering relocation find important, and to what extent, when making their decisions on housing preferences?
 7. How do older adults considering relocation reason regarding:
 - different housing options and
 - motives for considering and effectuating relocation, and
 - to what extent are their motives fulfilled?
 8. Is the newly developed housing counselling intervention usable, feasible and acceptable for the Swedish municipality context, and what are the pros and cons of different delivery formats?
 9. Which outcomes should be used to investigate the effectiveness of housing counselling, and what are
 - the responsiveness and,
 - the intervention effects on the selected primary and secondary outcomes, as indicated by the results of the pilot study?
 10. What are the main concepts and pathways of a theory on housing choices, relocation and active ageing?

Chapter 3

Overview of Registers

The study period in Register RELOC-AGE ranges from 1990-2020 and is comprised of 13 registers with staggered coverage which is here illustrated. Below the figure are descriptions of each register contained in the study.



Register descriptions

Register

Description

Total Population Register (TPR) (1968)

Sex; birth date; civil status (duration, dates, changes); address (dates, changes); income; country of origin; citizenship; in-/emigration (dates); n of people in the

dwelling; housing tenure; socioeconomic indicators of the neighbourhoods (on postal code, municipal levels)

National Patient Register (NPR) (1987)

Hospitalization outcomes: total no. of hospitalizations/month; in-patient health outcomes based on ICD-10 chapters (e.g., for falls, fractures, stroke etc.)

Real Estate Property Register (REPR) (1908)

Objective housing characteristics for each dwelling: type of dwelling; price of dwelling; type of tenure; size; presence of stairs/elevator; floor; building and construction year; characteristics of the neighbourhood: communal facilities (e.g., roads), green areas; date of each relocation

Geographical database (GD) (1952)

DESO (demographical statistical unit); coordinates of the housing and address etc.

Death Cause Register (DR) (1952)

Death cause and date

Longitudinal integrated database for health insurance and labour market studies (LISA) (1990)

Education level, income, social insurance

Drug Prescription Register (DPR) (2005)

Drug prescriptions for chronic illnesses (ATC code, dose and date): endocrine; cardiovascular; hepatic; renal or neurological/ neuromuscular

Swedish Intensive Care Register (SIRI) (2001)

Intensive care for laboratory-confirmed influenza and (since 2020) COVID-19.

Swedish internet-based surveillance system for communicable diseases (SmiNet) (1997; 2004)

Laboratory-confirmed influenza and (since 2020) COVID-19

Municipal Health Care Register (MHCR) (2007)

Care received and date

Interventions for Elderly and People with Disabilities Register (IEPDR) (2007)

Home help and service type and no of hours/month/year: escorting, replace the relative, personal care, meal delivery, security alarm, daytime activities; short-term vs long-term

Apartment Register (AR) (2012)

Dwelling type; number of rooms; dwelling unit size; kitchen type

Scania Outdoor Environment Database (ScOut)

24 outdoor environemtn qualities 2008-2019

Chapter 4

Description of data sets from SCB

A significant amount of data from the registrars originates from Statistics Sweden (SCB). This data is delivered in text format (file extension .txt), and is partitioned, for the most part, into individual files separated by both year and grouped data set. While the data contained in this data originates from the specified data Registrars previously outlined, the data is received from SCB consolidated and grouped into various data sets which require further cleaning and processing. These grouped data sets are described below.

4.1 Population

Description of Population data set

4.2 Lisa

Description of Lisa data set

4.3 Housing

Description of Housing data set

Chapter 5

Data cleaning and joining of raw data

As illustrated above, the data from RELOC-AGE is comprised of from several registers and sources. In order to arrive at the final data set, a number of data cleaning actions, multiple joins, and many quality control steps have been taken to insure reliable analysis and data integrity. This section details steps taken.

5.1 SCB data

With data covering about 3 million individuals over decades and with a multitude of variables spread out across hundreds of very large files, the computational effort to complete these tasks are very time intensive, often taking hours for a large merge, with progress occasionally hindered by computational restrictions and small errors which arise in the data cleaning process. With this in mind, detailed documentation, contained both here and alongside code used in the data cleaning, is prioritized to reduce any need to repeat these time-intensive processes.

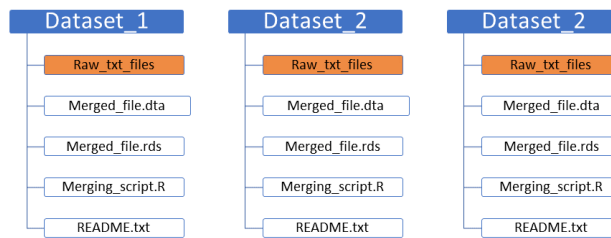
As a first step, the following initial data cleaning is performed:

- Raw data files are organized into folder structure where each folder contains all data from a particular data set.
- An individual script for each data set is written in R that reads the raw yearly .txt files and merges files into one data set.
- When required, a variable “year” is generated in the joined data set specifying which year the data originates from(taken from the name of the .txt file).
- Variables are renamed into lowercase with spaces and other delimiters transformed into underscores (_) for consistent naming conventions and

avoidance of future merge conflicts.

- The joined and cleaned data set is saved in the contained folder in both R's .rds and Stata's .dta formats.
- A README.txt file is created in each folder documenting the process.

The result consists of eleven folders each containing a data set's respective raw data, a documented R merging/cleaning script for full reproducibility, and a merged data set in both R and Stata format to be used in subsequent merging and further analysis.



Chapter 6

Descriptives by Age and Sex

Chapter 7

Housing data

Study period 2012-2020 when data is made available from SCB from the Real estate and Apartment registers. Data is taken from three SCB data sources: lev_lisa, lev_housing, and lev_population. The following steps are taken to arrive at a complete data set:

- The unique identifier **lopnr** in the lev_population data set is filtered with the following criteria: `index_p == 1 & ater_pnr == 0 & sen_pnr == 1`.
- Next this set of unique identifiers are matched to the full lev_lisa data set, resulting in lisa data that contains only the unique identifiers from the first step.
- 690 duplicate lopnr-year observations are identified and removed from the data set. Further inspection indicated this small number is a result of messy data, showing no trend or further information.
- As data from lev_lisa ends in 2019 and data from the lev_housing data set ends in 2020, time invariant variables(year of birth, sex, education, etc.) from unique individuals in 2019 are replicated for 2020 to facilitate the appropriate matching to the unique identifiers in the 2020 Housing data.
- Lastly, the data is joined by the unique lopnr-year combination with the lev_housing data.

7.1 Specific data considerations

7.1.1 Identifying partners

There are some discrepancies in the data when finding consistent partner matches to unique individuals across the multiple data sets.

We can find partner data in two of the datasets: **partners_rtb** and **samh**. In the **partners_rtb** dataset, we have three variables for every lopnr-year observation:

- Lopnrsamh – (no definition given in the excel sheet) 1987-1997.
- Lopnrsambo – “sambo’s personummer” from 1998.
- Lopnrmakpart – “make/maka/partners personummer” (I believe this is technically married) from 1998.

From the **samh** dataset, we have one variable: * LopNrSamh – (no definition given in the accompanying excel sheet).

Since no particular partner variable was consistent over time, a new variable is created, “partner”, that takes the value of whichever variable has valid data (of one of the above variables) for that lopnr/year. If there are two values, the priority is for the lopnrsamh from the **partners_rtb** dataset (it seems to have the best coverage).

A screenshot for an particular individual with multiple partners over time to illustrate.

	↑	lopnr	year	lopnrsamh	lopnrsambo	lopnrmakpart	LopNrSamh	partner
1	29	1992		46328	NA	NA	803685	46328
2	29	1993		803685	NA	NA	NA	803685
3	29	1994		803685	NA	NA	NA	803685
4	29	1995		NA	NA	NA	NA	NA
5	29	1996		NA	NA	NA	NA	NA
6	29	1997		NA	NA	NA	NA	NA
7	29	1998		NA	NA	NA	NA	NA
8	29	1999		NA	NA	NA	NA	NA
9	29	2000		NA	NA	2974318	NA	2974318
10	29	2001		NA	NA	2974318	NA	2974318
11	29	2002		NA	NA	2974318	NA	2974318
12	29	2003		NA	NA	NA	NA	NA
13	29	2004		NA	NA	NA	NA	NA
14	29	2005		NA	NA	NA	NA	NA
15	29	2006		NA	NA	NA	NA	NA
16	29	2007		NA	NA	NA	NA	NA
17	29	2008		NA	NA	NA	NA	NA
18	29	2009		NA	NA	NA	NA	NA
19	29	2010		NA	NA	NA	NA	NA
20	29	2011		NA	NA	NA	2789169	2789169
21	29	2012		NA	NA	NA	2789169	2789169
22	29	2013		NA	NA	NA	2789169	2789169
23	29	2014		NA	NA	NA	2789169	2789169
24	29	2015		NA	NA	NA	2789169	2789169
25	29	2016		NA	NA	NA	2789169	2789169

7.1.2 Recoding variables

Education, housing type, housing tenure

7.1.3 Identifying relocations

To identify when an individual has relocated in the data the following considerations are taken into account.

- The housing variables, **fast_lopnr** and **lghlopnr** appear to uniquely identify the housing location of a particular individual. Over time, a change in either variable should indicate that an individual has relocated (highlighted in orange below). This change appears to be the best indicator of whether an individual has relocated or not.

Lopnr	year	age	age_group	civil_status	education	housing	tenure	sex	fast_lopnr	lghlopnr	xkoord	ykoord	kommun	byggtyp	n_housing	change_housing	change_tenure
1	2012	58	55-74	single	Upper secondary	Småhus	Äganderätt	Man	574486	40905	186	30	3	no change	no change	no change	no change
	2013	59	55-74	single	Upper secondary	Småhus	Äganderätt	Man	574486	40905	186	30	3	no change	no change	no change	no change
	2014	60	55-74	single	Upper secondary	Småhus	Äganderätt	Man	574486	40905	186	30	3	no change	no change	no change	no change
	2015	61	55-74	single	Upper secondary	Småhus	Äganderätt	Man	1178270	254126	182	30	3	Småhus to Småhus	Äganderätt to Äganderätt	no change	no change
2	2016	62	55-74	single	Upper secondary	Småhus	Äganderätt	Man	574486	40905	186	30	3	Småhus to Småhus	Äganderätt to Äganderätt	no change	no change
	2017	63	55-74	single	Upper secondary	Småhus	Äganderätt	Man	574486	40905	186	30	3	no change	no change	no change	no change
	2018	64	55-74	single	Upper secondary	Övriga hus	Bostadsrätt	Man	995882	404307	180	30	3	Småhus to Övriga hus	Äganderätt to Bostadsrätt	no change	no change
	2019	65	55-74	single	Upper secondary	Övriga hus	Bostadsrätt	Man	995882	404307	180	30	3	no change	no change	no change	no change
3	2020	66	55-74	single	Upper secondary	Övriga hus	Bostadsrätt	Man	995882	404307	180	30	3	no change	no change	no change	no change
	2014	58	55-74	partner	Preschool/lower secondary	Småhus	Äganderätt	Man	698882	2331947	1380	30	2	no change	no change	no change	no change
	2015	56	55-74	partner	Preschool/lower secondary	Småhus	Äganderätt	Man	698882	2331947	1380	30	2	no change	no change	no change	no change
	2016	57	55-74	partner	Preschool/lower secondary	Småhus	Äganderätt	Man	272335	2342136	1315	30	2	Småhus to Småhus	Äganderätt to Äganderätt	no change	no change
4	2017	58	55-74	partner	Preschool/lower secondary	Småhus	Äganderätt	Man	272335	2342136	1315	30	2	no change	no change	no change	no change
	2018	59	55-74	partner	Preschool/lower secondary	Småhus	Äganderätt	Man	272335	2342136	1315	30	2	no change	no change	no change	no change
	2019	60	55-74	partner	Preschool/lower secondary	Småhus	Äganderätt	Man	272335	2342136	1315	30	2	no change	no change	no change	no change
	2020	61	55-74	partner	Preschool/lower secondary	Småhus	Äganderätt	Man	272335	2342136	1315	30	2	no change	no change	no change	no change
5	2012	63	55-74	partner	College/University/Research	Specialbostäder	Bostadsrätt	Kvinna	840863	1666168	180	33	1	no change	no change	no change	no change

- The coordinate variables, **xkoord** and **ykoord** appear to be occasionally inconsistent, giving different values for the same **fast_lopnr** and **lghlopnr** identifier in some of the data. This may be the result of data errors or possibly a GPS margin of error when measuring housing location (Highlighted yellow).
- The variable **n_housing** counts how many unique **fast_lopnr**'s are associated with each **lopnr**. For example, individual RED, has resided in three locations, Individual BLUE has resided in two locations, and Individual 40 has resided in one location during the sample period.
- The variables **change_housing** and **change_tenure** take the before and after values of **housing** and **tenure** (respectively) and return the before and after categories when a change (relocation) has occurred.
- Byggtyp** appears to follow the patterns of **fast_lopnr** and **lghlopnr**, but contains some missing data. The missing **Byggtyp** value seems to be associated with the values “Övriga hus” and “Specialbostäder” in **housing**.

7.1.4 Identifying change in kommun

In a similar fashion as identifying the change in a housing identifier over time for an individual, we can observe when an individual has relocated outside of a

geographical area when the geographical indicator changes.