


# Administrative data – common errors

**Danilo Dolenc**



GCC-STAT Regional Workshop  
Muscat, Oman, 22-24 September 2019

# My country



# Your countries

Country	Population			% of non-nationals (GLMM)
	GCC-STAT	UNSD	GLMM	
Bahrain	1,50	1,57	1,50	55
Kuwait	4,23	4,20	4,64	70
Oman	4,56	4,83	4,66	44
Qatar	2,64	2,63	2,74	87
Saudi Arabia	32,61	33,55	33,41	38
UAE	9,12	9,54	9,12	87

GCC-STAT (Population Statistics 2017): <https://gccstat.org/en/statistic/statistics/population>

UNSD (Statistical Yearbook, 2018 edition): <https://unstats.un.org/unsd/publications/statistical-yearbook/>

GLMM (national statistics, 2017-2018): <https://gulfmigration.org/gcc-total-population-and-percentage-of-nationals-and-non-nationals-in-gcc-countries-national-statistics-2017-2018-with-numbers/>

## My way

- ▶ Census experience since 1981
- ▶ Head of demography statistics
  - ▶ Completely based on administrative sources
- ▶ Register-based Census Project Manager
  - ▶ 2011, 2015, 2018
- ▶ Responsible for managing administrative sources
  - ▶ New system developed in 2018

# General prerequisites for using administrative data (1)

- ▶ National legislation
  - ▶ Access to out of NSI data (administrative data, secondary data)
    - ▶ State / public owned
    - ▶ Private owned?
  - ▶ Right to linkage micro data at the individual level

# General prerequisites for using administrative data (2)

- Appropriate content of data in sources to cover national laws (e.g. census) and needs
- Availability of unique identifiers
  - Level of observation unit
    - Person
    - Building and dwelling
    - Enterprise

More about that in next sessions

# Administrative vs survey data (1)

- Administrative data are of .... quality than survey data
  - Better
  - Same
  - Worse
- How we measure quality in everyday life
  - Objective
  - Subjective

## Administrative vs survey data (2)

- Administrative data (de iure) reflect the reality..... than survey data (de facto)
  - Better
  - Same
  - Worse
- Administrative source
  - Population Register
- Methodological implication
  - Usual residence



# Administrative vs survey data (3)

- Respondents provide.... accurate data to the administrative authorities than to the statistical surveys
  - More
  - Same
  - Less
- Administrative source
  - Tax Register
- Methodological implication
  - Distribution of household income

# Administrative vs survey data (4)

- ▶ Use of administrative data in general .... the timeliness of results
  - ▶ Improve
  - ▶ Have no effect
  - ▶ Impair
- ▶ Administrative source
  - ▶ Real Estate Register
- ▶ Methodological implication
  - ▶ Re-definition of statistical process

# Errors in administrative data (1)

- ▶ The error can occur in every phase of the administrative process
  - ▶ Very similar to the data collection phase in statistical survey
  - ▶ Errors considered by statistical methodology could not be considered as errors by the administrative authority
    - ▶ Objective reporting delays - long-lasting administrative procedures
      - ▶ Ownership of real estate

## Errors in administrative data (2)

- ▶ Interaction between respondent and official
  - ▶ Intentional errors due to personal interest
    - ▶ Registration of place of residence
  - ▶ Accidental (random) errors
    - ▶ Wrong data entry (formally correct)
- ▶ Most data formally correct but not useful
  - ▶ Massive use of unknown category of classification
- ▶ Coverage could be questionable
  - ▶ Difficult to check
    - ▶ Possible use of aggregates

## Errors in administrative data (3)

- The most problematic systematic orientation of data keeper
  - Not to pay attention to some population sub-groups
    - Lower quality of foreigner's data
  - Not to collect data for all variables foreseen by the law
  - To consider statistics as the least important user
  - To cancel quality data sources
    - Household Register

# Common problems with administrative data from NSI point of view

- Missing identifiers
- Missing variables
- Item non-response (statistically speaking)
- Over-coverage
- Under-coverage
- Non-harmonized classifications
- Timeliness of data sources
- Coherency and consistency of data from different sources / periodicity

# Population Register QA (1)

- Stock data received monthly / quarterly
  - Agreement with data provider
    - Ministry of the Interior
- No missing personal identifiers (PIN)
  - Public identifiers could be transformed into statistical ones
- No missing addresses
  - Permanent and / or temporary

# Population Register QA (2)

- Full completeness of target population
  - Data extracted 3 months later
    - Including data on demographic events
- No editing needed for
  - Sex, age, citizenship, place of birth
- Data on relations available between
  - Spouses
  - Parents and children



# Population Register QA (3)

- Main problems with multiplicative effect
  - **Over-registration** (>1%)
    - Non-documented emigration
    - Administrative survivors
  - Some missing dwelling identifiers (part of address)
    - 13% in 2011 – only 1% in 2018
  - Item non-response in some variables of foreigners
    - Marital status, educational attainment
  - **Reliability of data on residence**
    - (6% - 8% - not living at registered address)

# Real Estate Register QA (1)

- ▶ Stock data received quarterly
  - ▶ Agreement with data provider
    - ▶ Surveying and Mapping Authority
- ▶ Publicly available data on web
- ▶ No missing identifiers
  - ▶ Building, part of building, owners (PIN)
- ▶ Data on actual use available
  - ▶ Classification of Types of Constructions (CC)

# Real Estate Register QA (2)

- Main problems with multiplicative effect
  - **Almost no updates for some variables of existing units**
    - Field data collection carried out in 2007
  - Long lasting administrative procedure for
    - New entries
    - Change of ownership
  - **Not satisfactory quality of basic housing variables**
    - High imputation rates

# Household Register QA (1)

- Stock data received monthly
  - Agreement with data provider
    - Ministry of the Interior
- All identifiers to link individual data available
  - Address, household ID, PIN of member
  - Relation to the reference person available
    - Key for automated family data derivation
- Excellent formal quality of data
  - No systematic errors

# Household Register QA (2)

- Main problems with multiplicative effect
  - **Under-coverage**
    - Data available for permanent residence only
      - 10% of household identifiers is missing for usual residence
    - Data refer to private households only
  - **Reliability of data on residence**
    - Household size and structure
    - Family types

# Quality assessment of register-based household data (1)

- Influence of de facto and de iure registration on relevancy and reliability of household and family data

HH 1	
Mother	58
Father	60
Son	30



HH 2	
Mother	28
Daughter	1
Partner	30

- The total number of households remains the same
- Average household size is not changed

## Quality assessment of register-based household data (2)

- The most important – transformation of family types
  - The main issue “statistical” increase of lone parent families as direct consequence of registration

HH 1	
Mother	58
Father	60
Son	30



HH 2	
Mother	28
Daughter	1
Partner	30

*Married couple with children*

***Married couple without children***

*Lone mother family*

***Cohabiting couple with children***

# Business Register QA (1)

- ▶ Data received daily
  - ▶ Agreement with data provider
    - ▶ Agency for Public Legal Records and Related Services
  - ▶ Publicly available data
- ▶ Excellent formal quality of data based on national laws
  - ▶ No systematic errors



# Business Register QA (2)

- Main problems with multiplicative effect
  - **The main activity of enterprise**
    - Could be intentionally wrong due the interest
      - For example, taxation reasons
  - **Registration of local units of enterprise**
    - It is not obligatory by the law
    - Three main negative quality consequences
      - Regional national accounts
      - Distribution of workplaces at the municipality level
      - Register-based labour migration

# Statistical vs. administrative

- ▶ Convert administrative data into statistics – the essence of register-based system
  - ▶ Administrative data could be used directly without any transformation
    - ▶ Some Population Register data
  - ▶ Administrative data had to be transformed or new statistical variables created due to
    - ▶ Linkage data from several sources
    - ▶ Methodology reasons = statistical concepts
- ▶ Statistical registers should be created

# Example - census concepts

Source	Administrative concepts	Statistical concepts
Central Population Register	Registered residence <ul style="list-style-type: none"><li>♦ permanent and/or temporary</li><li>♦ temporarily absence abroad</li></ul>	<b>Usual residence</b> <ul style="list-style-type: none"><li>♦ length and/or intention of stay</li><li>♦ priority rules in case of two registered residence</li></ul>
Household Register	Permanent residence only Private households only Relations to reference person only	Private households (housekeeping concept) Institutional households Generating families
Real Estate Register	Building Construction type Actual use of building / part of building	Conventional dwellings Collective living quarters Other housing units

# Example - Statistical Business Register

- ▶ Updated monthly
  - ▶ More administrative data sources included
    - ▶ Financial Administration (taxation, customs)
  - ▶ Statistical methods used for missing data
  - ▶ The main differences
    - ▶ Main activity of the enterprise
      - ▶ According to the turnover by product
    - ▶ Local units determined
      - ▶ Just for major enterprises with high turnover

# Administrative data quality drivers

- ▶ Permanent use in the statistical process
- ▶ Long-continued and long-lasting use in the statistical process
- ▶ Very close cooperation with the keepers
- ▶ Feedback from statistical evaluation implemented in the administrative source
  - ▶ No individual data are transferred