DEPARTMENT OF STATISTICS SINGAPORE

Data Integration in the Compilation of Cohort-Based Marriage and Divorce Indicators in Singapore

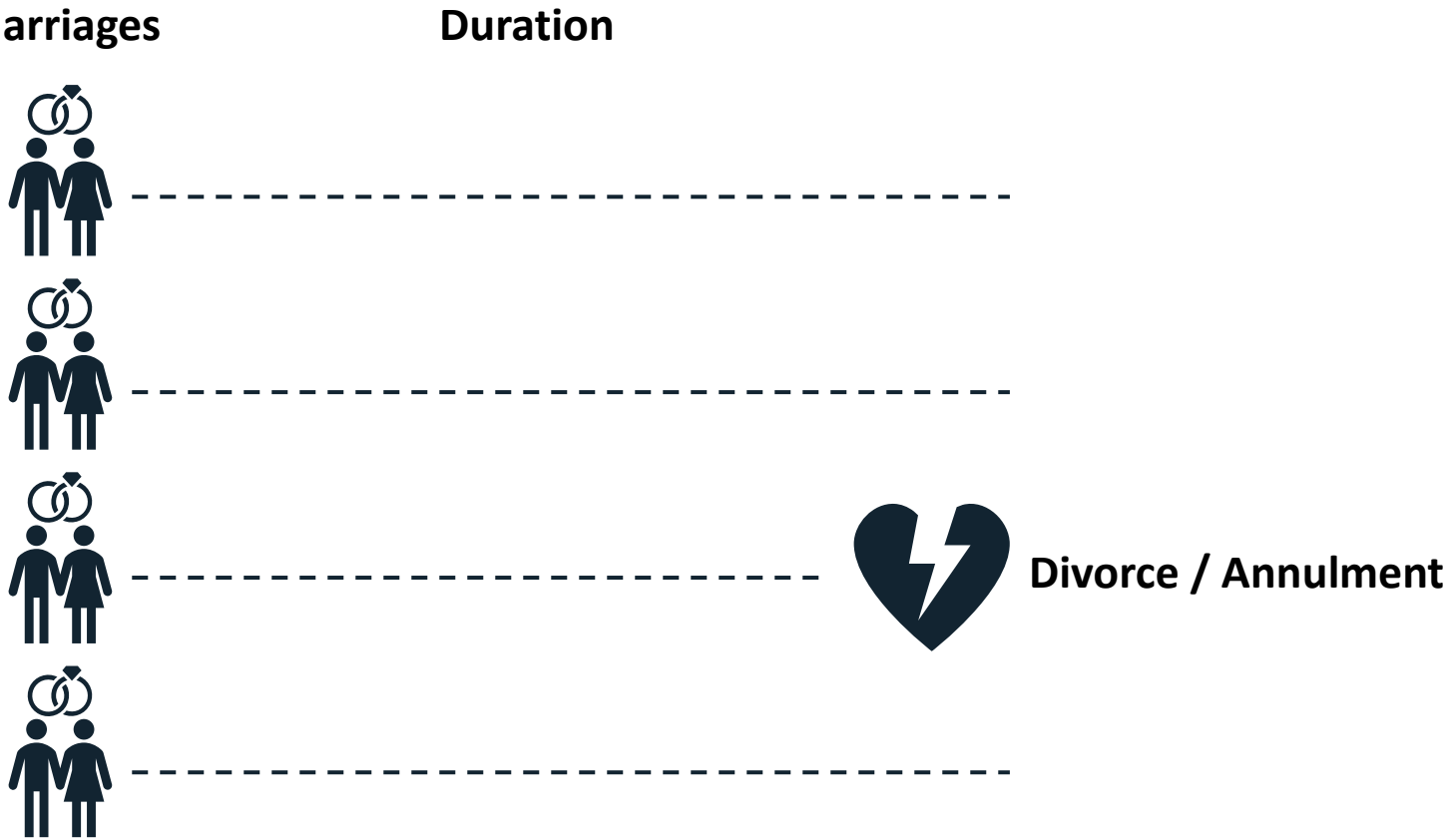


Cohort / Longitudinal Analysis

- Focuses on behaviour of individuals over time
- Building up a picture of collective history
- More time and effort required for compilation
- Gain valuable insights not available from period indicators

Cohort-based Marriage and Divorce Indicators

- Statistical dataset of marriages and divorces



Note: Singapore has a decentralised statistical system, and administrative records of marriages and divorces are collected by the respective administrative authorities.

Overview of Data Integration Process

1

Data Preparation



2

Multi-Step Matching Process

3

Clerical Review

Data Preparation

Basic Checks

Within each dataset

- Uniqueness of identifiers
- Completeness
- Validity of field values
- Consistency across related fields

Additional Checks

For integrating data across different sources

Important to ensure consistency in concepts, definitions, classifications, data formats and reference periods across different sources

Data Preparation

Data Standardization for Matching Variables

- Marriage Certificate Number

- Dates – Marriage date, Divorce date, Date of birth e.g. YYYYMMDD format

- National Registration Identity Card (NRIC) number i.e. Unique Identification Number (UIN) in Singapore

Data Standardization for Matching Variables

- String or character variables with a fixed format such as Marriage Certificate Number, UIN
  - E.g, “xxxM123456yyy”



Use “Regular Expressions” to find and extract a search pattern within a string of text

Data Preparation

Mapping Table

Source Agency 1		Source Agency 2	
Code	Code Description	Code	Code Description
NO	No qualification	0	No qualification
PR	Primary	1	Primary
UP	Secondary	3	Secondary
PS	Post Secondary (Non-Tertiary)	4	Post Secondary: General
		5	Post Secondary: Vocational
PO	Polytechnic	6	Polytechnic
PQ	Professional Qualification and other Diploma	7	Professional Qualification and other Diploma
TE	University	8	University First Degree
		9	University Postgraduate

Note: For illustration purposes only



Establish national statistical standards for adoption by administrative source agencies

Data Preparation

Code Conversion Table

Code Type	Org Code	Code Value	Code Description	Validity Indicator	Effected Date	Ceased Date
MS	ROMM	S	Single	N	20050701	20131231
MS	ROMM	F	Widower	N	20050701	20131231
MS	ROMM	W	Widow	N	20050701	20131231
MS	ROMM	D	Divorcee	N	20050701	20131231
MS	ROMM	M	Married	N	20050701	20131231
MS	ROMM	N	Unknown	N	20050701	20131231
MS	ROMM	1	Single	Y	20140101	
MS	ROMM	2	Married	Y	20140101	
MS	ROMM	3	Widowed	Y	20140101	
MS	ROMM	4	Divorced	Y	20140101	

Note: For illustration purposes only

Multi-Step Matching Process



Marriage



Marriage Certificate



Divorce / Annulment



Marriage Certificate



Marriage Certificate

+

Year of Marriage

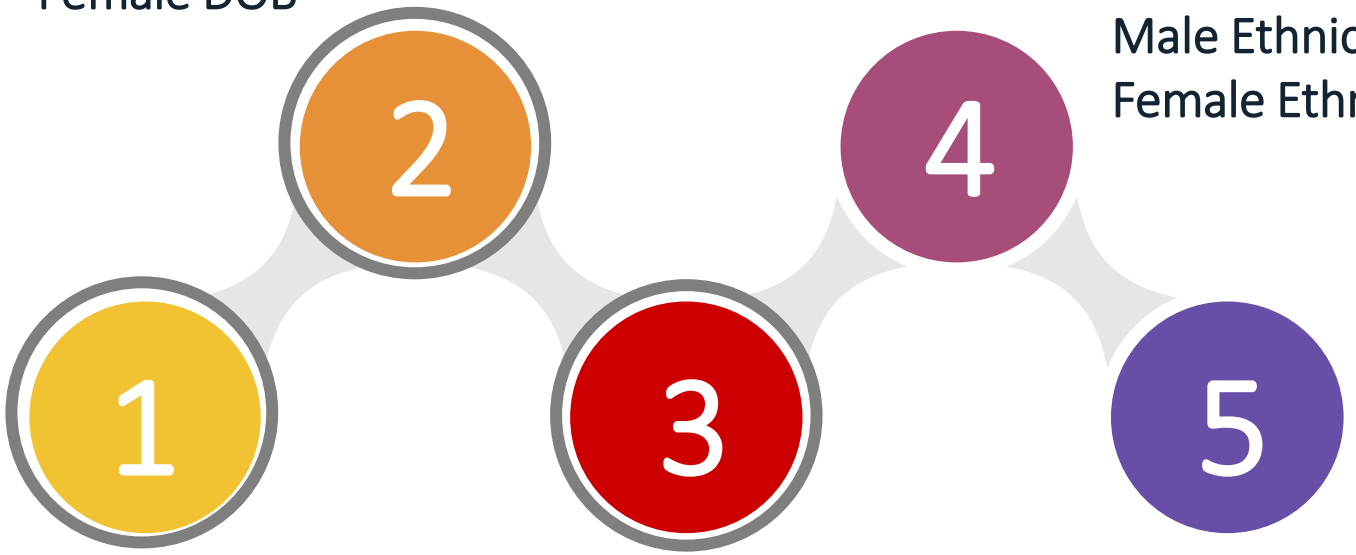
=

Matching Key between Marriage and Dissolution records

Multi-Step Matching Process

- High proportion of records with blank / invalid marriage certificate number in earlier years
- Adopted supplementary multi-step deterministic matching approach

Multi-Step Matching Process



Marriage Date
Male UIN
Female DOB

Marriage Date
Male UIN
Female UIN

Marriage Date
Female UIN
Male DOB

Marriage Date
Male DOB
Female DOB
Male Ethnicity
Female Ethnicity

Male UIN
Female UIN

Match with immediately preceding marriage

Multi-Step Matching Process

Consideration

It is unlikely for a person to be married to two different individuals on the same date.

Couple taken to be same if

Marriage date is the same and a combination of other matching variables such as UIN and date of birth tally reasonably well.

Multi-Step Matching Process

Consideration

It is possible for the marriage date to be reported or collected incorrectly during the filing of divorce/annulment.

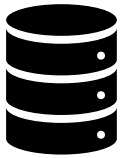
Couple taken to be same if

If a divorce/annulment record matches a preceding marriage record exactly on male and female UIN.

Multi-Step Matching Process

- Match keys chosen depends on topic of study
- Distinguishing power of a given matching variable may also differ across different data sources and countries
- Reasons for choosing match keys should be well thought-through and documented

Clerical Review



Integrated dataset



Automated checks



Inconsistencies flagged out

- More than one intact marriage at a given point
- Conflicting information for common variables between matched marriage and dissolution record



Clerical reviews are time-consuming, but necessary in maintaining data integrity and consistency

References

1. Statistical Singapore Newsletter “The Development of Cohort-Based Marriages and Divorces Statistics”

   www.singstat.gov.sg/-/media/files/publications/population/ssnsep15-pg5-9.pdf

2. Singapore’s National Statistical Standards and Classifications

   www.singstat.gov.sg/standards/standards-and-classifications

Thank You

Our Vision

A National Statistical Service of Quality, Integrity and Expertise

Our Mission

We deliver Insightful Statistics and Trusted Statistical Services that Empower Decision Making

DEPARTMENT OF STATISTICS SINGAPORE