# Intergenerational Analyses Using the IDI

Irene (Run) Wu

Barry Milne

COMPASS Research Centre

August 2017

# Introduction

This project aimed to investigate the extent of which intergenerational links are preserved using the collection of de-identified administrative datasets from the Integrated Data Infrastructure (IDI). Intergenerational links are important to assess such things as intergenerational transfer of wealth, intergenerational socio-economic mobility, and familial influences on health and wellbeing (both genetic and environmental).

Less investigated are multi-generational effects (spanning more than two generations). Our investigations of intergenerational effects in the IDI have a particular interest in multi-generational links to open up possibilities for research on, e.g., (i) assessing the multi-generational effect of socio-economic status on health and other outcomes for those living today (e.g., does socio-economic influence span two generations and more?); (ii) documenting intergenerational residential mobility (e.g., by documenting geographical similarlty between births from succeeding generations); and (iii) to assess whether the biological effect of parental age (e.g., on psychiatric disorders)[1] extends across generations.

The datasets in the IDI have been linked at the person-level for the whole New Zealand population and covers different timeframes. Specifically, the Department of Internal Affairs (DIA) data include birth information dating back to the 1840s, with unique IDs for the child and for both their parents where they exist. We have focussed on the DIA dataset as this is the only IDI dataset with potential links prior to 1990 which covers most of the population (MSD records begin in 1990, Census records are only available for 2013).

Making use of these DIA intergenerational links, we sought to answer the following questions: (i) How many generations can be determined? (ii) What is the total number at each generation?

# Data

Department of International Affairs – Birth dataset. The following variables were of interest to the investigation:

| IDI Variable Name | Description |
| --- | --- |
| Snz_uid | IDI unique identifier allocated by SNZ |
| Snz_dia_uid | DIA identifier allocated by SNZ |
| Dia_bir_birth_month_nbr | Month child was born |

| Dia_bir_birth_year_nbr | Year child was born |
|---|---|
| Parent1_snz_uid | IDI unique identifier allocated by SNZ |
| Parent1_snz_dia_uid | DIA dataset identifier allocated by SNZ |
| Dia_bir_parent1_birth_month_nbr | Month parent1 was born |
| Dia_bir_parent1_birth_year_nbr | Year parent1 was born |
| Parent2_snz_uid | IDI unique identifier allocated SNZ |
| Parent2_snz_dia_uid | DIA dataset identifier allocated by SNZ |
| Dia_bir_parent2_birth_month_nbr | Month parent2 was born |
| Dia_bir_parent1_birth_year_nbr | Year parent2 was born |

The data dictionary for the Live event data reveals several key changes within the dataset:

1) The notification of births and deaths were made compulsory in 1848. Introduced through the Registration of Births and Deaths Act 1875.
2) The birth date for parents were only captured from 1972 onwards, previously the parents age was requested instead.
3) In 1998 the DIA moved to digital storage of paper records, referred to as the Record Back Capture Project.
4) Since 2005 the sex of the parents was recorded, prior to 2005 the first parent entry in the birth record was always the mother.

## Method

Within the IDI environment, intergenerational links were created through subsetting and merging the dataset using the procedure as follows:

1) Birth dates were categorised into decades starting from 1880 to 2020.
2) Two subsets were created: separating child links to parent1 and child links to parent2.
3) By decade, assess the proportion of parent1's and parent2's who themselves appear as a child in the DIA records

## Results

Figure 1 and Table 1 show that in the most recent decades (2001-2010; 2011-2017) intergenerational linkage rates were reasonable: 66-72%. This is perhaps what would be expected given the proportion of children born to parents who were also born in New Zealand. However, linkage rates drop away

dramatically prior to 2001: approximately 45% for 1991-2000, 33% for 1981-90; 20% for 1971-1980, and 0.1% or less for decades prior to 1971.

As these linkage rates were so low prior to 2001, multi-generational linkage are likely to be limited. As such, we did not direct compute the answers to our two questions – (i) the number of generations able to be linked, and (ii) the total numbers at each generation – as we can give approximate answers from the linkage rates we did compute. These approximations indicate that the number of generations able to be linked back ($n_g$) is $n_g = 1$ for about 69% of cases for parents having children since 2001, and close to no cases for parents having children prior to 1970. Even assuming very short generations (20 years) and a random probability of being linked back to a previous generation, it is likely that $n_g = 2$ for ~30% of cases for parents having children since 2011 (i.e., 0.675*0.456 = 0.308 for parent1, and 0.666*0.445=0.296 for parent2), and $n_g = 3$ for ~6% of cases (i.e., 0.675*0.456*0.189 = 0.058 for parent1, and 0.666*0.445*0.201=0.060 for parent2). Using similar methodology and assuming (more reasonably) generations=30 years, $n_g = 2$ for ~22% of cases for parents having children since 2011, and $n_g = 3$ for no cases. It is very unlikely there are any cases for $n_g = 4$.

*Table 1: Total number of intergenerational links in each decade for parent1 and for parent2*

| Decade | N of child IDs in parent1 dataset | N of individual IDs linked as parent1 | % of parent1 links | N of child IDs in parent2 dataset | N of individual IDs linked as parent2 | % of parent2 links |
|---|---|---|---|---|---|---|
| Before 1881 | 240516 | 0 | 0 | 241854 | 0 | 0 |
| 1881-1890 | 191235 | 0 | 0 | 186216 | 0 | 0 |
| 1891-1900 | 187845 | 0 | 0 | 181533 | 0 | 0 |
| 1901-1910 | 240516 | 0 | 0 | 233115 | 0 | 0 |
| 1911-1920 | 294018 | 0 | 0 | 286746 | 6 | 0 |
| 1921-1930 | 308799 | 0 | 0 | 301974 | 6 | 0 |
| 1931-1940 | 308526 | 6 | 0 | 303039 | 0 | 0 |
| 1941-1950 | 441441 | 21 | 0 | 434925 | 15 | 0 |
| 1951-1960 | 573921 | 63 | 0 | 566562 | 60 | 0 |
| 1961-1970 | 634017 | 348 | 0.1 | 620706 | 393 | 0.1 |
| 1971-1980 | 572856 | 108201 | 18.9 | 549156 | 110115 | 20.1 |
| 1981-1990 | 544269 | 180216 | 33.1 | 512439 | 172662 | 33.7 |
| 1991-2000 | 585924 | 267330 | 45.6 | 548598 | 244002 | 44.5 |
| 2001-2010 | 611007 | 439497 | 71.9 | 576036 | 406338 | 70.5 |
| 2011-2017 | 360450 | 243228 | 67.5 | 343623 | 228810 | 66.6 |
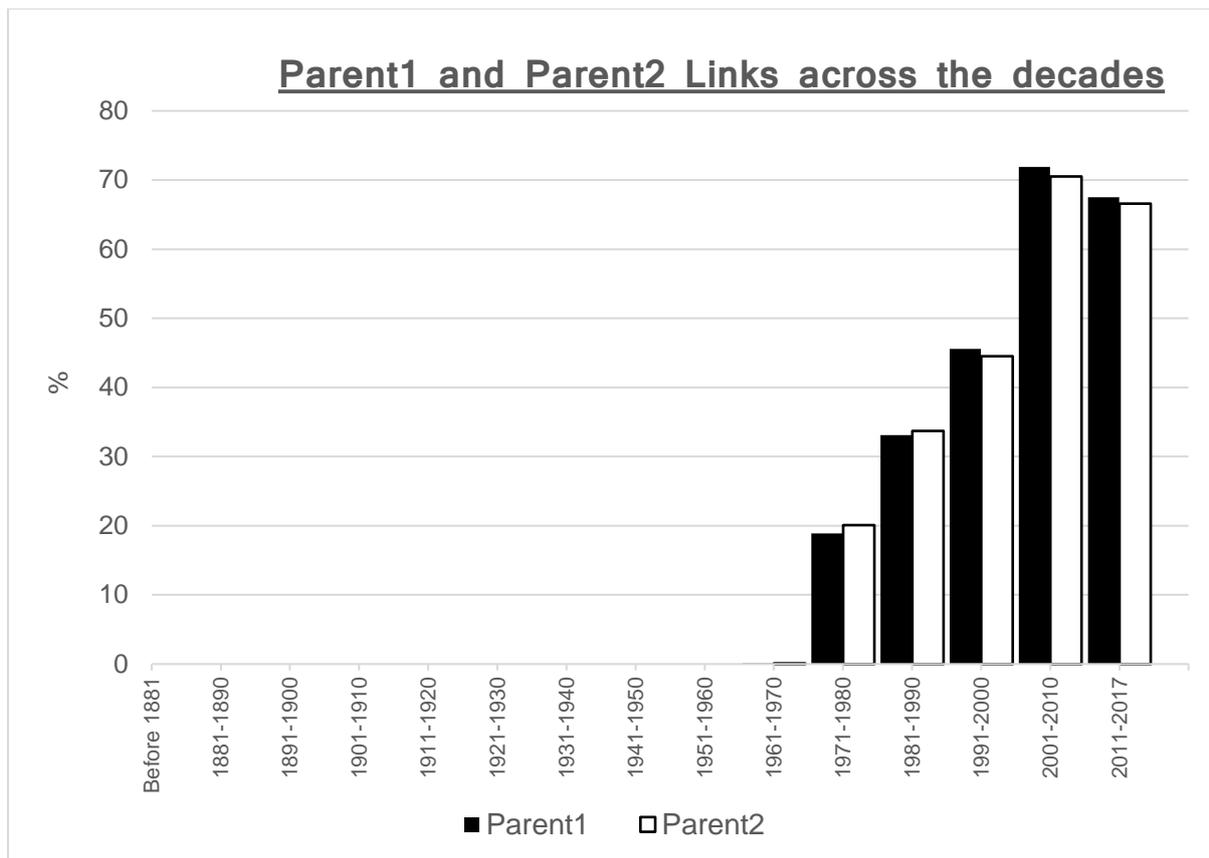
*Figure 1: Comparison of the proportion of individuals to child links in each decade from the IDI birth dataset.*

## Conclusion

Using DIA data in the IDI, one-generation intergenerational links are possible for large proportions of the New Zealand population for recent generations (since 2001) only. Numbers drop away prior to 2001. Multigenerational (2+ generation) links are likely to exist for a small proportion of recent generations (two generation links for 20-30% of those born since 2011; three generation links for <6% of those born since 2011; four generation links are unlikely to exist at all).

References

1. de Kluiver H, Buizer-Voskamp JE, Dolan CV, Boomsma DI. (2017) Paternal age and psychiatric disorders: A review. Am J Med Genet B Neuropsychiatr Genet;174(3):202-213.