Intergenerational Analyses Using the IDI: An update

Barry Milne

Eileen Li

Andrew Sporle

COMPASS Research Centre

March 2020

Statistics New Zealand Disclaimer

The results in this paper are not official statistics. They have been created for research purposes from the Integrated Data Infrastructure (IDI), managed by Statistics NZ.

The opinions, findings, recommendations, and conclusions expressed in this paper are those of the author(s), not Statistics NZ, or The University of Auckland.

Access to the anonymised data used in this study was provided by Statistics NZ under the security and confidentiality provisions of the Statistics Act 1975. Only people authorised by the Statistics Act 1975 are allowed to see data about a particular person, household, business, or organisation, and the results in this paper have been confidentialised to protect these groups from identification and to keep their data safe.

Careful consideration has been given to the privacy, security, and confidentiality issues associated with using administrative and survey data in the IDI. Further detail can be found in the Privacy impact assessment for the Integrated Data Infrastructure available from `www.stats.govt.nz`.

Citation: Milne, B. J., Li, E., & Sporle, A. (2020). Intergenerational Analyses Using the IDI: An update. COMPASS Research Centre, `https://www.auckland.ac.nz/content/dam/uoa/auckland/arts/our-research/research-institutes-centres-groups/compass/whole-population-data-analysis/Intergenerational-Links-IDI-Update.pdf`

Acknowledgement: This research was funded by an MBIE Endeavour Grant, ref 62506 ENDRP. We would like to thank Gabrielle Davie, Sheree Gibb, Andrea Teng and Martin von Randow for useful comments.

Executive summary

Intergenerational links are important to assess intergenerational transfer of wealth, intergenerational socio-economic mobility, and familial influences on health and wellbeing (both genetic and environmental). This report documents the potential for intergenerational links in the Integrated Data Infrastructure (IDI), a large database of de-identified administrative and survey data about people and households in New Zealand linked at the individual level.

The Department of Internal Affairs (DIA) dataset in the IDI includes birth information dating back to 1848, with unique IDs for the child and for both their parents where they were recorded. It is the only IDI dataset that can be used reliably to link to previous generations for most of the population. Making use of DIA records for the most current IDI refresh (at the time of writing this was the 20/10/2019 refresh), this report assesses the extent to which the parents, grandparents and great-grandparents of individuals identified in the DIA birth records can themselves be identified in the DIA birth records. Three questions are answered: (i) How many people can be linked back one, two, and three generations?; (ii) How does this vary by decade of birth?; and (iii) How does this vary by ethnicity and deprivation (measured using the NZDep13 scale of neighbourhood deprivation)?

Results indicate that over 1.9 million individuals can be linked back to at least one parent. Links to both parents are available for $> 50\%$ of individuals born since 1991. Links are possible to at least one parent for 26% of individuals born in the 1971–1980 decade, and for 69%, 85%, 81%, and 76% of individuals born in the four subsequent decades, respectively (note, the most recent 2011–2019 decade contains only 8.5 years). At least one grandparental link was available for more than half of individuals born in 2011–2019 and nearly a third of individuals born in 2001–2010. Great-grandparental links were rare — only 4% of individuals born in the most recent decade, though far more likely for Māori (11%).

Māori individuals are slightly more likely to have intergenerational links than European individuals, but greatly more likely than Pacific and Asian individuals. Multi-generational links (e.g., to grandparents and great-grandparents) are far more common for Māori than any other ethnic group. One-generation links do not differ by deprivation, but two- and three-generation links do, in that they are more common among those born in deprived neighbourhoods.

Five factors are offered to explain the high number of links for Māori individuals. Firstly, the Māori population is relatively young compared with the total NZ population, having a median age of approximately 24 years, resulting in a higher proportion of the total population having some form of digital administrative record. Secondly, Māori may be most likely to have at least one parent born in New Zealand (generating a record in the DIA birth records). Thirdly, because Māori are typically younger at the time of childbirth than non-Māori, Māori parents are more likely to be born after routine digitisation of DIA-birth records. Fourth, Māori may

be more likely to have a digitised birth record prior to routine digitisation, since digitisation often occurred when a birth certificate was required for the exercise of state custodial powers, and Māori are highly over-represented in the justice system in New Zealand (Department of Corrections, 2007). Fifth, linkage varies by deprivation, and half of the Māori population lives in the lowest 3 NZDep decile areas.

Implications of the ethnic differential in intergenerational links include that (i) those able to be linked do not form a representative cohort of New Zealanders, as Pacific and Asian New Zealanders are under-represented; (ii) if the interest is in Māori and/or European New Zealanders specifically, then intergenerational links will be available for the vast majority; (iii) multigenerational links are common for Māori, which may allow for whakapapa-based research, though issues of data governance and data sovereignty will need to be addressed (Te Mana Raraunga, 2018); and (iv) the relatively high representation of Māori in intergenerational records may be viewed as a result of Māori being the subjects of data surveillance activities by the state, without having the power or resources to have input into how researchers and decision-makers use the data (Kukutai & Cormack, 2019).

The report concludes that there is enormous scope for intergenerational research using IDI data, albeit with challenges in generalisability to the total population, and challenges regarding data governance and data sovereignty. Linkage is possible not only for those born in recent decades, who in many cases can be linked back to grandparents, but also for those born in the 1970s and 1980s. However, it should be noted that there may be data availability issues that impact research, including the completeness of birth records on (e.g.) birth weight, gestational age and parental occupation, and also the availability of historic data from other IDI tables, many of which do not include data before the year 2000 or even later. Nonetheless, possibilities for and issues regarding intergenerational research should be explored to assess the extent to which the breadth of intergenerational data available in the IDI can be used to answer questions of interest to researchers, policy makers, iwi and other groups.

Introduction

This is an update on the report by (Wu & Milne, 2017) Intergenerational Analyses Using the IDI, which summarised the ability to establish intergenerational links in the Integrated Data Infrastructure (IDI), a large database of de-identified administrative and survey data about people and households in New Zealand linked at the individual level (for a description, see (Milne et al., 2019)). The (Wu & Milne, 2017) report used the most up-to-date data in the IDI at that time (using the IDI refresh as at 20/04/2017). The current report updates numbers able to be linked across generations with the addition of 2.5 years of IDI data (using the IDI refresh as at 20/10/2019) and with Department of Internal Affairs (DIA) records now

fully digitised back to 1990 (compared to 1998 for earlier refreshes). In addition, this update more fully investigates links across multiple generations. In the initial report, low return of one-generational links meant that multi (2+) generational links were likely to be low, and so were estimated; in the current report, these have been directly calculated. Further, this report describes who is able to be linked (in terms of ethnicity and deprivation) to determine the extent to which the full diversity of the New Zealand population is represented in those able to be linked intergenerationally.

Background

Intergenerational links are important to assess 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). Multi-generational links 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?); and (ii) assessing whether the biological effect of parental age (e.g., on psychiatric disorders, (de Kluiver, Buizer-Voskamp, Dolan, & Boomsma, 2017)) extends across generations.

The Department of Internal Affairs (DIA) dataset in the IDI includes birth information dating back to 1848, with unique IDs for the child and for both their parents where they were recorded. DIA is the only IDI dataset that can be used reliably to link to previous generations for most of the population. Because the link is through birth records, links are only possible for individuals who have one or more parents born in New Zealand — those whose parents were not born in New Zealand will not be represented and therefore linkage can never be possible for the whole population using this method. Nonetheless, it is better than alternatives. For example, while some links are possible using Ministry of Social Development (MSD) data from 1990 onwards, these will only cover MSD clients for whom family pedigree data have been collected (links are not possible for MSD clients without family pedigree data, nor for those who are not MSD clients). Links are also possible using Census records but only for 2013; we note that Census data from 2018 is soon to be added.

Using DIA data, (Wu & Milne, 2017) found that 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).

Making use of these more recent (using the IDI refresh as at 20/10/2019) and more complete

(with all birth records digitised back to 1990) DIA intergenerational links, this report sought to answer the following questions: (i) How many people can be linked back one, two, and three generations?; (ii) How does this vary by decade of birth?; and (iii) How does this vary by ethnicity and deprivation?

Data and Methods

Data from the IDI refresh 20191020 were used. The DIA Birth dataset (DIA-Births) was used to identify people, their year of birth and the linking IDs (snz_uid) of the person's listed on their birth record as their 'parent1' (typically but not always the mother) and 'parent2' (typically but not always the father).

The Department of International Affairs Life events (i.e., births, deaths and marriages) data dictionary reveals the following key aspects about the dataset:

1. The birth dates for parents were only captured from 1972 onwards; previously the parents' age was requested instead.
2. In 1990 the DIA moved to digital storage of paper records.
3. Since 2005 the sex of the parents was recorded, prior to 2005 the first parent entry in the birth record was always the mother.

A one-generational link was determined if the snz_uid of the parent1 or parent2 of an individual could be found in the DIA-Births dataset (indicating a successful link to the birth record of a parent). Similarly, a two-generational link was determined if the snz_uid of the parent1 or parent2 of a parent1 or parent2 of an individual could be found in the DIA-Births dataset (indicating a successful link to the birth record of a grandparent). Three-generational links were similarly determined.

Year of birth was used to group individuals in decades from 1961–1970 to 2011–2019 (note, the final decade is incomplete — around 8.5 years instead of 10). Links were sought prior to 1961 but were rare ($n < 100$) so are not reported here.

Data on ethnicity were obtained from the source-ranked ethnicity variable in the personal_detail table in the IDI, which prioritises ethnicity records according to the quality of their source. Individuals were classified into one or more of the following Level 1 ethnic groups, using the Statistics New Zealand statistical standard for ethnicity (Statistics New Zealand, 2005): European; Māori; Pacific; Asian; Middle Eastern, Latin American and African (MELAA); Other. Note that individuals can (and commonly do) identify with more than one ethnic group (Shackleton et al., 2018). Because of low numbers we do not present data for MELAA and Other.

Data on deprivation were obtained using the New Zealand Deprivation Index 2013 (NZDep13, (Atkinson, Salmond, & Crampton, 2014) based on the deprivation characteristics of 'meshblocks' (small areas with a typical population of 60 – 110 people). The NZDep13 combines 2013 census data relating to income, employment, qualifications, communication, support, living space, transport and home ownership into a single measure of relative socio-economic deprivation. Meshblocks of each individual's address at birth were determined using the 'full address notification table' in the IDI. The most recent addresses within 60 days of birth were used. If there were no address updates within 60 days of birth, the mother's most recent address change prior to giving birth or within 60 days of giving birth was used. If the address information was still missing, the father's most recent address prior to their child's birth or within 60 days after their child's birth was used. If none of the above successfully identified the address, an enlarged time window for address changes was used (between 60 days and 365 days after birth). If no address information could be found in these time windows, then address (and therefore deprivation) was set to missing. These address meshblocks were then mapped to NZDep13 deciles using the meshblock-NZDep concordance table. Adjacent NZDep13 deciles were combined to form quintiles, from quintile 1 (Q1, least deprived, combining deciles 1 and 2) to quintile 5 (Q5, most deprived, combining deciles 9 and 10). Deprivation was only able to be assessed for the two most recent decades (2001-2010 and 2011–2019) as there was excessive missingness in earlier decades.

Results are presented as:

1. the number (percent) of individuals with one-, two-, and three-generational links to each of two, four and eight parents, grandparents, and great-grandparents, respectively;
2. the total number (percent) of individuals with
    - 0, 1, and 2 one-generational (parental) links;
    - 0, 1, 2, 3, and 4 two-generational (grandparental) links;
    - 0, 1, 2, 3, and 4+ three-generational (great-grandparental) links.

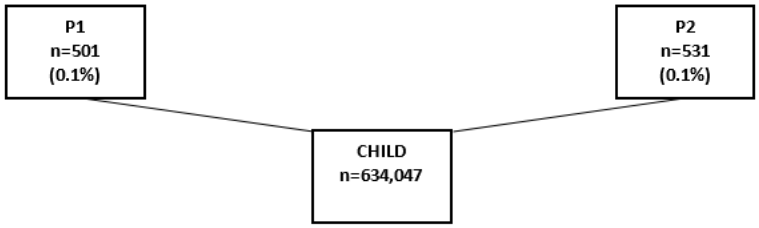
Results are presented for individuals by decade (for decades for which sufficient data exist) and stratified by ethnicity and deprivation.

Results

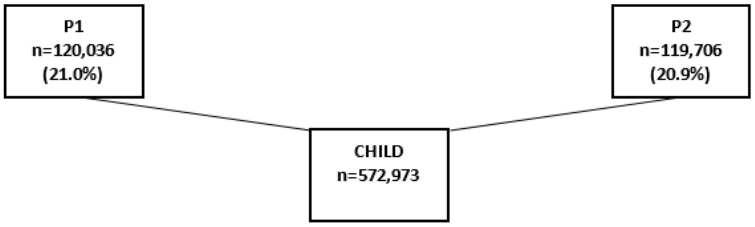
Intergenerational family pedigrees are shown in Figure 1. These show the number and percent linked back one generation, two generations, and three generations by decade of birth of the child. Because of diminishing numbers, results for two-generation links are shown only for

decades 1981–1990 to 2011–2019, and results for three generation links are shown only for the 2011–2019 decade.

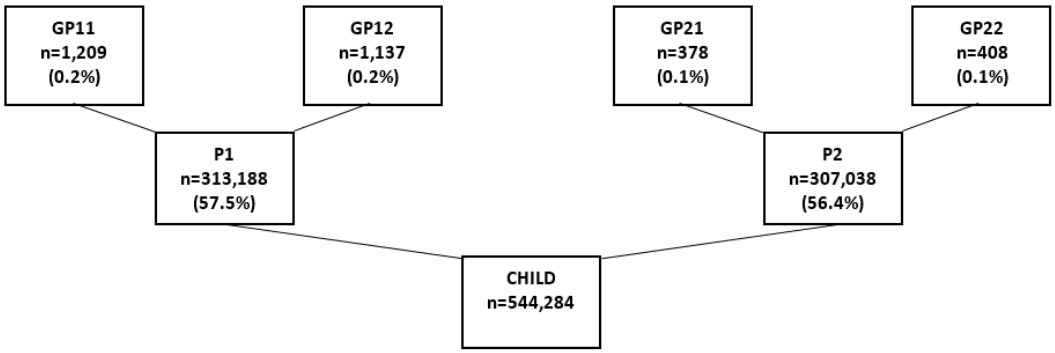
1961-1970



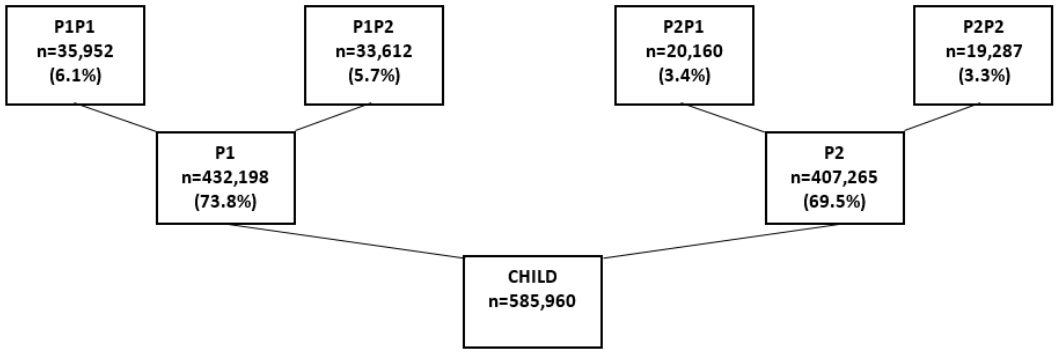
1971-1980



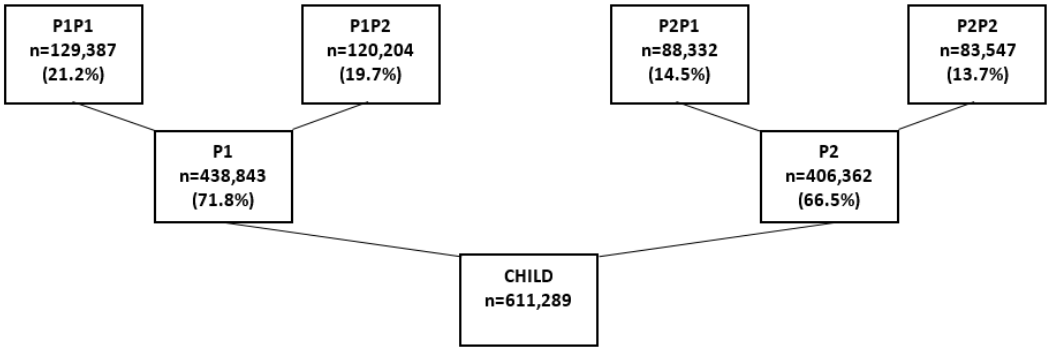
1981-1990



1991-2000



2001-2010



2011-2019

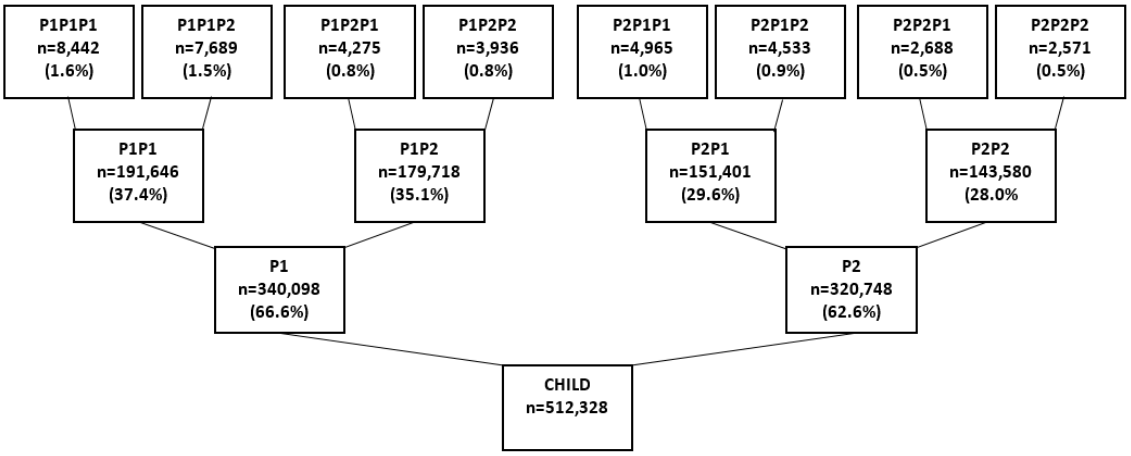


Figure 1: Intergenerational pedigrees by decade of birth. P1=Parent1; P2=Parent2; GP11=Grandparent – Parent1 of Parent1; GP12=Grandparent – Parent2 of Parent1, etc; GGP122=Great-grandparent – Parent2 of Parent2 of Parent1, etc. All numbers have been random-rounded to base 3 as per the confidentiality requirements of Stats NZ

One-generation links

For the 1961–1970 decade, very few individuals could be linked back to a parent (0.1%). For the 1971–1980 decade, around 21% of individuals could be linked to parent1 and 21% could also be linked to parent2. Slightly more individuals could be linked to parent1 than parent2 in the 1981–1990 decade (57.5% and 56.4%, respectively), 1991–2000 decade (73.8% and 69.5%, respectively), 2001–2010 decade (71.8% and 66.5%, respectively), and the 2011–2019 decade (66.6% and 62.6%, respectively). Overall, 1.9 million New Zealanders born since 1971 could be linked back to at least one parent (n=1,910,037).

Figure 2 shows the percentage of individuals able to be linked to 0, 1 and 2 parents by decade. This shows that around a quarter of individuals born in the 1971–1980 decade (26%) could be linked to at least one parent, which increased to 69%, 84%, 81% and 76% for the 1981–1990, 1991–2000, 2001–2010 and 2011–2019 decades respectively. Between 53% and 58% of individuals could be linked back to both parents in the most recent three decades.

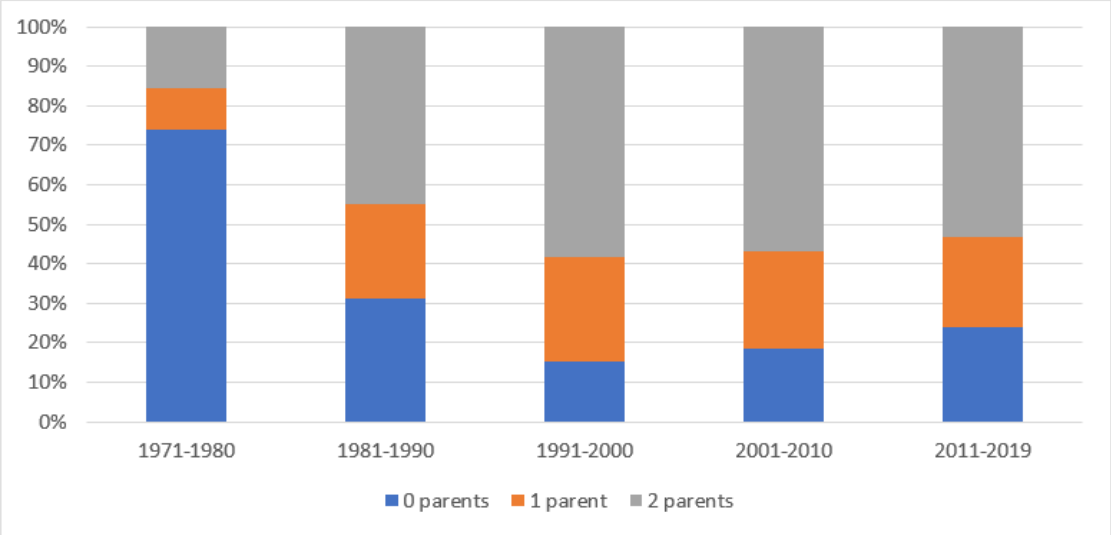


Figure 2: Percentage of individuals able to be linked to 0, 1 and 2 parents by decade.

Two-generation links

In general, two-generation links were more common on the parent1 (typically maternal) side than the parent2 (typically paternal) side. A link to a grandparent on the parent1 side was possible for slightly more than 1000 individuals (0.2%) born in the 1981–1990 decade, around 35,000 individuals (5.7% – 6.1%) born in the 1991–2000 decade, 120,000–130,000 individuals (19.7%–21.2%) born in the 2001–2010 decade, and 180,000-190,000 individuals (35.1%–37.4%) born in the 2001–2010 decade. The numbers for links on the parent2 side were around 400 individuals (0.1%) born in the 1981–1990 decade, around 20,000 individuals (3.3% – 3.4%)

born in the 1991–2000 decade, 83,000–88,000 individuals (13.7%−14.5%) born in the 2001–2010 decade, and 140,000–150,000 individuals (28.0%−29.6%) born in the 2001-2010 decade. Overall, more than 500,000 individuals born since 1981 could be linked back to at least one grandparent (n=531,900).

Figure 3 shows the percentage of individuals able to be linked to 0, 1, 2, 3 and 4 grandparents for the three most-recent decades. This shows that just over 10% of individuals (10.3%) born in the 1991–2000 decade, around 1/3 of individuals (32.5%) born in the 2001–2010 decade, and slightly more than half of individuals (53.1%) born in the 2001–2010 decade could be linked to at least one grandparent. The percent of individuals able to be linked to two or more grandparents was 6.3%, 23.7% and 43.0% for the 1991–2000, 2001–2010, and 2011–2019 decades, respectively.

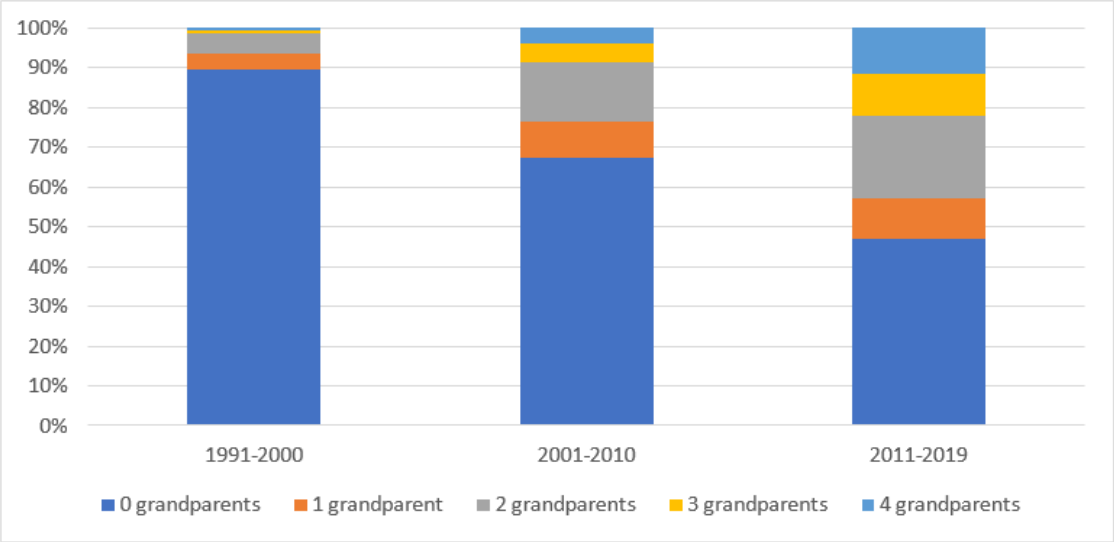


Figure 3: Percentage of individuals able to be linked to 0, 1, 2, 3 and 4 grandparents by decade.

Three-generation links

Three-generation links existed, but almost entirely only for the most recent decade (2011–2019). Only 0.3% had any links for the 2001–2010 decade (results not shown). For individuals born in 2011–2019, between 2,500 and 8,500 linked to specific great-grandparents (see 2011–2019 panel, Figure 1). n=20,514 (4.0%) individuals could be linked to at least one great-grandparent: n=8,061 (1.6%) could be linked to one only, n=8,601 (1.7%) could be linked to two, n=2,187 (0.4%) could be linked to three, and n=1,665 (0.3%) could be linked to four or more.

Inter-generational linkage by ethnicity and deprivation

Ethnicity

One-generation links. Figure 4 shows the number of individuals able to be linked to parents for each ethnic group for decades since 1981. For those born 1981–1990, around half of European and Māori individuals were able to be linked back to both parents, compared to only 13% of Pacific individuals and 11% of Asian individuals. More than 70% of European individuals and more than 80% of Māori individuals could be linked back to at least one parent. For those born 1991-2000, around 2/3 of European and Māori individuals were able to be linked back to both parents, compared to 22.5% of Pacific individuals and 9% of Asian individuals. Nearly all European individuals (94%) and Māori individuals (97%) were able to be linked back to at least one parent. For those born 2001–2010, around 2/3 of European individuals and nearly 3/4 of Māori individuals were able to be linked back to both parents, compared to 1/3 of Pacific individuals and 8% of Asian individuals. 92.5% of European individuals and 98% of Māori individuals were able to be linked back to at least one parent. For those born in the most recent decade (2011–2019), around 2/3 of European and over 3/4 of Māori individuals were able to be linked back to both parents, compared to 41% of Pacific individuals and just 6% of Asian individuals. 92% of European individuals and $> 99\%$ of Māori individuals were able to be linked back to at least one parent.

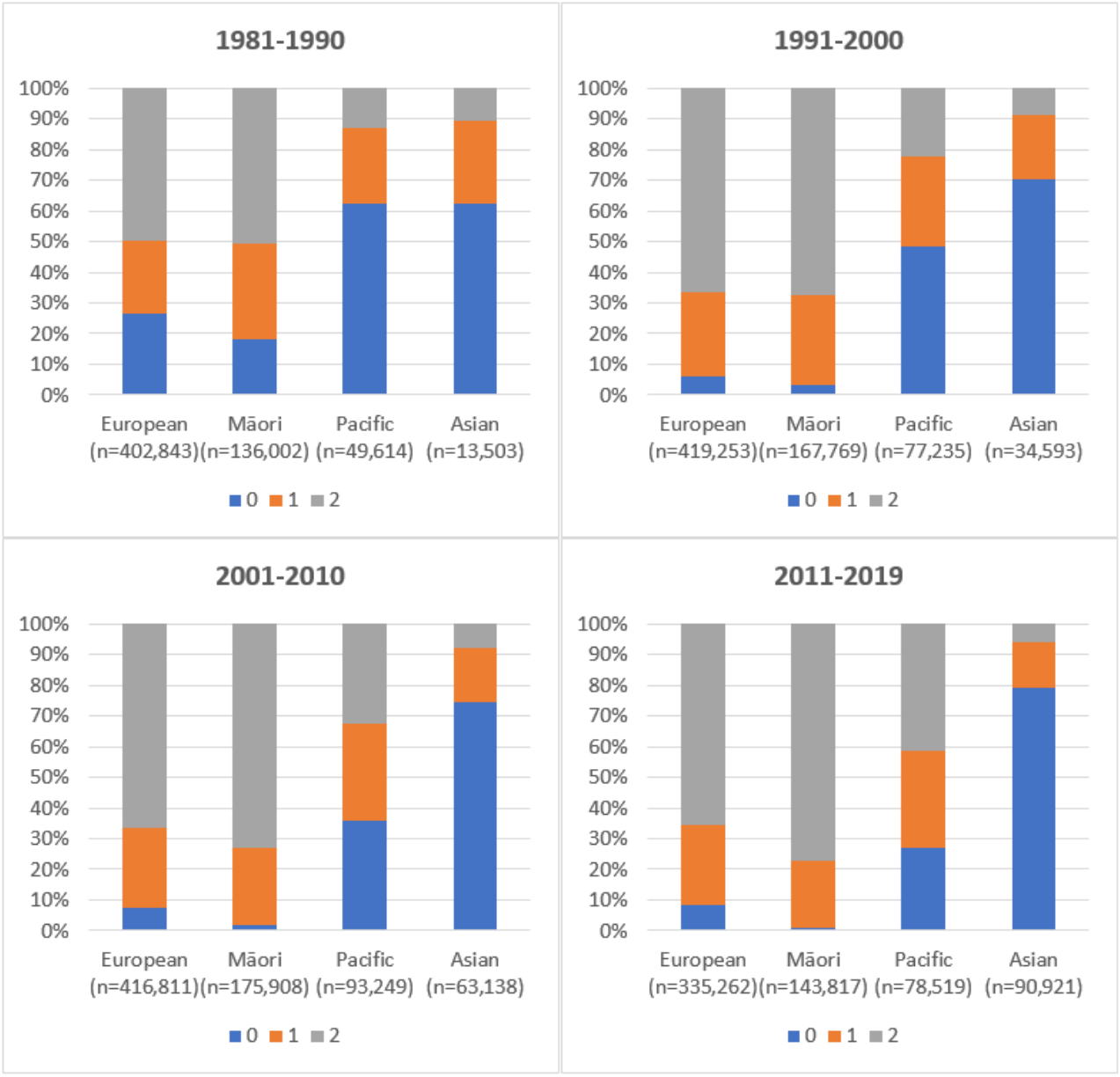


Figure 4: Percentage of individuals able to be linked to 0, 1 and 2 parents by ethnicity by decade.

Two-generation links. Figure 5 shows the number of individuals able to be linked to grandparents for each ethnic group for decades since 1991. For those born 1991–2000, very few European individuals (9%), Pacific individuals (9%) and Asian individuals (2%) could be linked to at least one grandparent. However, nearly a quarter of Māori individuals (23%) could be linked to at least one grandparent, and 13% could be linked to two or more grandparents. For those born 2001–2010, around 1/3 of European individuals, 59.5% of Māori individuals, 27% of Pacific individuals and just 7% of Asian individuals could be linked to at least one grandparent. Over 40% of Māori individuals (42.6%) could be linked to two or more grandparents. For those born in the most recent decade (2011–2019), 86% of Māori individuals, 64% of European individuals, 44% of Pacific individuals and 10% of Asian individuals could be linked back to at least one

grandparent; while 70.5% of Māori individuals, 53% of European individuals, 29% of Pacific individuals and 6.5% of Asian individuals could be linked back to at least two grandparents; 41% of Māori individuals, 27% of European individuals, and 11.6% of Pacific individuals could be linked back to three or four grandparents; and 20% of Māori individuals and 14.5% of European individuals could be linked back to all four grandparents.

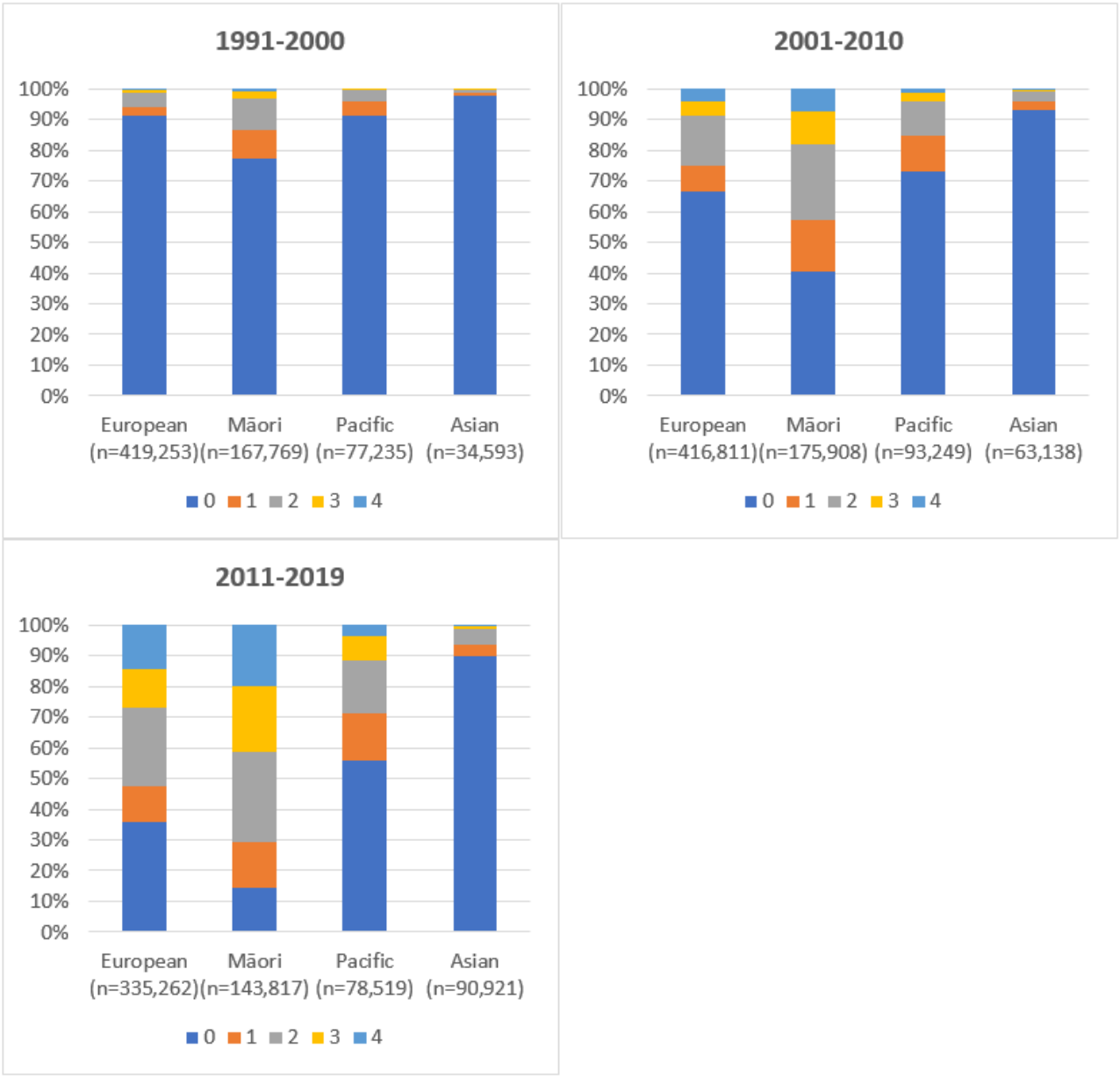


Figure 5: Percentage of individuals able to be linked to 0, 1, 2, 3 and 4 grandparents by ethnicity by decade.

Three generation links. Links to great-grandparents were rare for all ethnicities, but more than twice as common for Māori individuals born in 2011–2019 (11% were able to be linked to at least one great-grandparent) than Pacific (5.1%) and European (3.5%) individuals born in 2011–2019. Fewer than 0.5% of Asian individuals born in 2011–2019 were able to be linked to

at least one great-grandparent.

Deprivation

One-generation links. The proportion of individuals able to be linked back one generation (i.e., to parents) did not substantially vary by deprivation in either the 2001–2010 or 2011–2019 decades (Figure 6). Across deprivation quintiles, 85% – 90% of individuals were able to be linked back to one parent and 57% – 63% were able to be linked back to both parents in 2001–2010, while around 85% of individuals were able to be linked back to one parent and around 60% were able to be linked back to both parents in 2011–2019. Linkage was less common, however, among those missing deprivation information — only 29% (2001–2009) and 12% (2011–2019) of those with missing deprivation information were able to be linked back to at least one parent. Note that a substantial fraction — 10.4% and 12.5% — of individuals had missing deprivation information in 2001–2010 and 2011–2019, respectively.

Figure 6: Percentage of individuals able to be linked to 0, 1 and 2 parents by deprivation by decade.

Two-generation links. Figure 7 shows the percentage of individuals able to be linked to grandparents for deprivation quintiles in the previous two decades. Here a clear gradient appears, with linkage increasing with increasing deprivation. For example, for those born in 2001–2010, only 19% of individuals in quintile 1 (born in the least deprived neighbourhoods) could be linked back to at least one grandparent compared to 48% of individuals in quintile 5 (born in the most deprived neighbourhoods); while 14% of individuals in quintile 1 could be linked back to two or more grandparents compared to 34% of individuals in quintile 5. Similarly, for those born in 2011–2019, 48% of individuals in quintile 1 could be linked back to at

least one grandparent (39% could be linked to two or more) compared to 67% of individuals in quintile 5 (53% could be linked to two or more). In both decades, < 9% of those with missing deprivation information could be linked back to one or more grandparents.

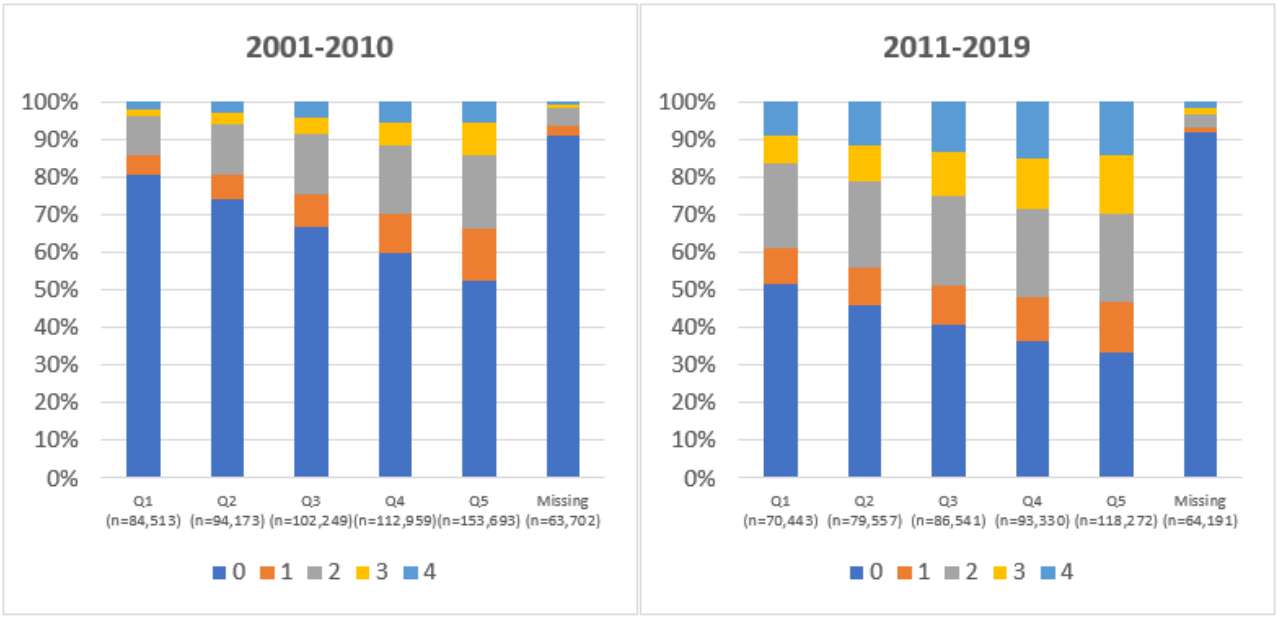


Figure 7: Percentage of individuals able to be linked to 0, 1, 2, 3 and 4 grandparents by deprivation by decade.

Three-generation links. There was also a deprivation gradient in linkage to great-grandparents: 1.2%, 1.9%, 3.2%, 5.1% and 8.7% of quintile 1, 2, 3, 4 and 5 individuals born in 2011–2019, respectively, were able to be linked back to at least one great-grandparent.

Discussion

This report documents the potential for intergenerational links in the IDI. It does so by analysing the extent to which the parents, grandparents and great-grandparents of individuals identified in the DIA birth records can themselves be identified in the DIA birth records, using IDI data for the 20/10/2019 refresh. Identification of individuals from previous generations in the DIA birth records means that (i) those individuals are necessarily in the IDI spine (excepting cases born prior to 1920, which will be rare) so can be linked to other IDI datasets; (ii) DIA birth information such as birthweight, gestational age, and number of elder siblings is potentially available for those individuals; and (iii) DIA birth information recorded about the parents of those individuals, such as parental age and occupation, is also potentially available. This last point means that where parents were linked, some data will be available on grandparents (i.e., their age and occupation at the time of the birth of their child); and

where grandparents were linked, some data will be available on great-grandparents; and where great-grandparents were linked, some data will be available on great-great-grandparents.

Using the 20/10/2019 IDI refresh, results indicate that over 1.9 million individuals can be linked back to at least one parent in the DIA births dataset, suggesting that data from subsequent IDI refreshes are likely to contain links for > 2 million New Zealanders. Links to both parents are available for $> 50\%$ of individuals born since 1991. Links are possible to at least one parent for 26% of individuals born as early as the 1971–1980 decade, and for 69%, 85%, 81%, and 76% of individuals born in the four subsequent decades, respectively. Note that 100% linkage is not possible since our methodology requires at least one parent to have been born in New Zealand. Decreasing intergenerational linkage rates since the 1991-2000 decade may be because a decreasing number of children born in New Zealand also have at least one parent born in New Zealand, as a result of New Zealand's increasing immigration. In line with this, the proportion of the usual resident population born in New Zealand has dropped from 82.5% at the 1996 Census to 80.5%, 77.1%, 74.8%, and 72.6% at the 2001, 2006, 2013, and 2018 Censuses, respectively (`https://www.stats.govt.nz/tools/nz-dot-stat`).

At least one grandparental link was available for more than half of individuals born in 2011–2019 and nearly a third of individuals born in 2001–2010. At least two grandparental links were available for 43% of individuals born in 2011–2019. Great-grandparental links were rare — only 4% of individuals born in the most recent decade, though far more likely for Māori (11%).

Links are more common on the 'parent1' side, which makes sense given a 'parent1' is always listed whereas a 'parent2' sometimes is not. Also, because the 'parent1' listed is most commonly the mother, and mothers are typically younger at the time of childbirth than fathers, the 'parent1' is more likely to be born after routine digitisation of DIA-birth records in 1990, and so more likely to be identified in the DIA birth records.

Intergenerational linkage differs by ethnicity. Links are most common for Māori individuals, for whom links are slightly more likely than European individuals, but greatly more likely than links for Pacific and Asian individuals. Multi-generational links (e.g., to grandparents and great-grandparents) are far more common for Māori than any other ethnic group. Five factors may explain this. Firstly, the Māori population is relatively young compared with the total NZ population, having a median age of approximately 24 years, resulting in a higher proportion of the total population having some form of digital administrative record. Secondly, Māori may be most likely to have at least one parent born in New Zealand (generating a record in the DIA birth records). Thirdly, because Māori are typically younger at the time of childbirth than non-Māori, Māori parents are more likely to be born after routine digitisation of DIA-birth records (as for 'parent1's above). Fourth, Māori may be more likely to have a digitised birth record prior to routine digitisation, since digitisation often occurred when a birth certificate

was required for the exercise of state custodial powers, and Māori are highly over-represented in the justice system in New Zealand (Department of Corrections, 2007). Fifth, linkage varies by socio-economic position as measured by NZDep (see below), and half of the Māori population lives in the lowest 3 NZDep decile areas.

Several implications of the ethnic differential in intergenerational links should be noted. First, those able to be linked do not form a representative cohort of New Zealanders, as Pacific and Asian New Zealanders are under-represented. Second, if the interest is in Māori and/or European New Zealanders specifically, then intergenerational links will be available for the vast majority. Third, multigenerational links are common for Māori, which may allow for whakapapa-based research, though issues of data governance and data sovereignty will need to be addressed (Te Mana Raraunga, 2018). Fourth, the relatively high representation of Māori in intergenerational records may be viewed as a result of Māori being the subjects of data surveillance activities by the state, without having the power or resources to have input into how researchers and decision-makers use the data (Kukutai & Cormack, 2019).

One-generation links do not differ by socio-economic circumstances (as measured by deprivation), but two- and three-generation links do, in that they are more common among those born in deprived neighbourhoods. It may be that the driver for this gradient is the same as that for the higher linkage rate for Māori — namely that individuals born in deprived neighbourhoods may be more likely to have parents who (i) were born in New Zealand, (ii) gave birth at a younger age, and (iii) were more likely to have a digitised birth record. It also perhaps highlights greater surveillance and engagement of state authority for those in deprived neighbourhoods.

Deprivation results should be interpreted with caution, however, given (i) the relatively high levels of missing (10.4% and 12.5% for the 2001–2010 and 2011–2019 decades, respectively), and (ii) the NZDep13 will characterise deprivation circumstances less well for years distant from 2013 (e.g., the early 2000s). Further investigation is required to work out who is unable to have address determined and this may shed light on why this group is under-represented in intergenerational links.

Research implications from this research include that there is enormous scope for intergenerational research using IDI data, albeit with challenges in generalisability to the total population, and challenges regarding data governance and data sovereignty. Linkage is possible not only for those born in recent decades, who in many cases can be linked back to grandparents, but also for those born in the 1970s and 1980s. For example, more than two thirds of those born in the 1981–1990 decade can be linked to at least one parent. Research on intergenerational mobility, intergenerational transmission of disease, birth order effects, and parental age effects all benefit from the availability of data capturing intergenerational links.

However it should be noted that there may be data availability issues that impact research,

including the completeness of birth records on (e.g.) birth weight, gestational age and parental occupation, and also the availability of historic data from other IDI tables, many of which do not include data before the year 2000 or even later (see `https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure#data-in-idi`). Nonetheless, possibilities for and issues regarding intergenerational research should be explored to assess the extent to which the breadth of intergenerational data available in the IDI can be used to answer questions of interest to researchers, policy makers, iwi and other groups.

References

Atkinson, J., Salmond, C., & Crampton, P. (2014). NZDep2013 Index of Deprivation. *Wellington: Department of Public Health, University of Otago*.

de Kluiver, H., Buizer-Voskamp, J. E., Dolan, C. V., & Boomsma, D. I. (2017). Paternal age and psychiatric disorders: A review. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *174*(3), 202-213. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/ajmg.b.32508 doi: 10.1002/ajmg.b.32508

Department of Corrections. (2007). *Over-representation of Māori in the criminal justice system: An exploratory report.* Wellington, New Zealand.

Kukutai, T., & Cormack, D. (2019). Mana motuhake a-raraunga: datafication and social science research in aotearoa. *Kotuitui: New Zealand Journal of Social Sciences Online*, *14*(2), 201-208. Retrieved from https://doi.org/10.1080/1177083X.2019.1648304 doi: 10.1080/1177083X.2019.1648304

Milne, B. J., Atkinson, J., Blakely, T., Day, H., Douwes, J., Gibb, S., ... Teng, A. (2019, 02). Data Resource Profile: The New Zealand Integrated Data Infrastructure (IDI). *International Journal of Epidemiology*, *48*(3), 677-677e. Retrieved from https://doi.org/10.1093/ije/dyz014 doi: 10.1093/ije/dyz014

Shackleton, N., Milne, B. J., Audas, R., Derraik, J. G. B., Zhu, T., Taylor, R. W., ... Taylor, B. (2018). Improving rates of overweight, obesity and extreme obesity in new zealand 4-year-old children in 2010–2016. *Pediatric Obesity*, *13*(12), 766-777. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/ijpo.12260 doi: 10.1111/ijpo.12260

Statistics New Zealand. (2005). *Understanding and working with ethnicity data: A technical paper.* Wellington: Statistics New Zealand.

Te Mana Raraunga. (2018). *Principles of Māori data sovereignty.* https://www.temanararaunga.maori.nz/nga-rauemi. Hamilton. Te Mana Raraunga.

Wu, R., & Milne, B. J. (2017). *Intergenerational analyses using the IDI.* https://cdn.auckland.ac.nz/assets/arts/research-centres/compass/documents/Intergenerational-Links-IDI.pdf. COMPASS Research Centre.