NEW ZEALAND AS A SOCIAL LABORATORY

MICROSIMULATION WITH THE NEW ZEALAND LONGITUDINAL CENSUS

MARTIN VON RANDOW

COMPASS ANNUAL COLLOQUIUM

13 AUGUST 2018

MICROSIMULATION

- Use real data to create an artificial world – starting population
- Describe how people change over time – statistical models and transition probabilities
- Carry out ‘virtual experiments’ and test ‘what-if’ scenarios by changing those rules – simulation
- Aggregate for estimated population distributions for different variables at different time points

COMPASS EXPERIENCE

- Modelling Social Change (MoSC, 2005–2008)
  - Cross-sectional census data 1981–2001
- Primary Care in an Ageing Society (PCASo, 2005–2008)
  - NZ & Australian Health Surveys; NZ National Primary Medical Care Survey
- Balance of Care in an Ageing Society (BCASo, 2009–2012)
  - NZ Health Surveys, NZ Disability Survey, NZ Census figures
- Modelling the Early Life-Course (MEL-C, 2009–2013)
  - Longitudinal data sets: CHDS, DMHDS, THNR, PIFS, and an NZ Census basefile
- Knowledge Laboratory of the Early Life-Course (K-LAB, 2013–2016)
  - MEL-C, NZ Health Survey, outcomes derived from international systematic reviews

NEW ZEALAND LONGITUDINAL CENSUS

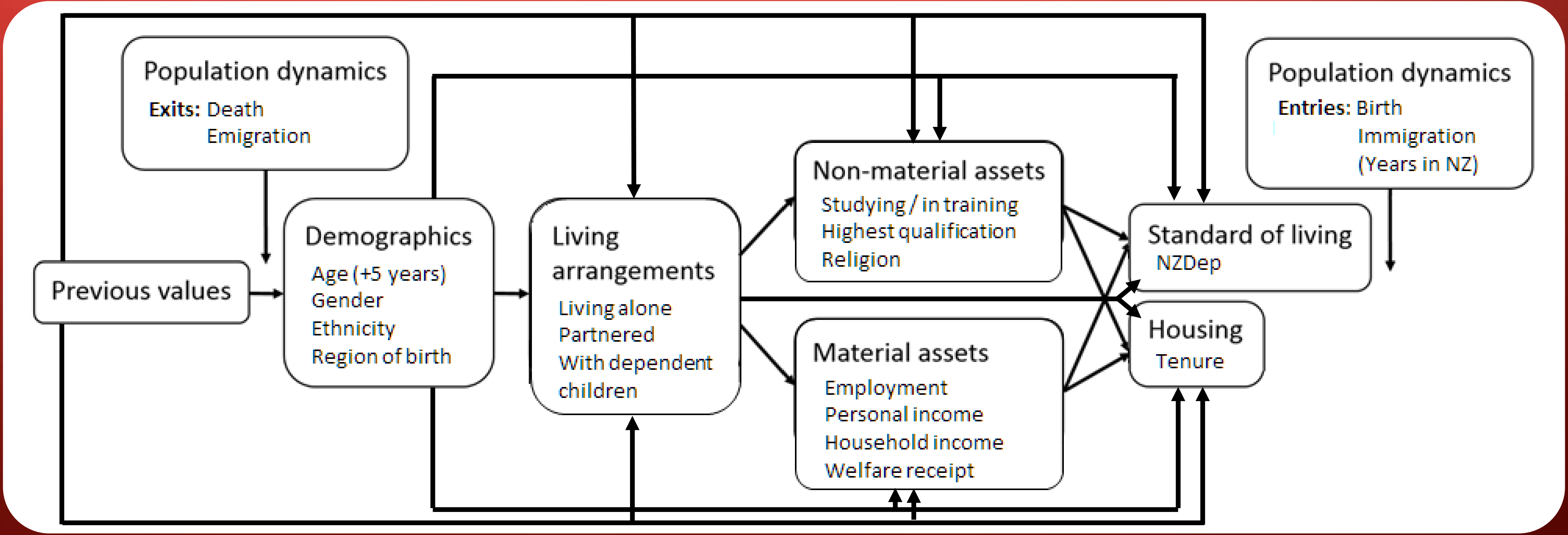
- Then COMPASS Director Peter Davis envisaged microsimulation based on the entire New Zealand population
- Linked censuses back 2006–1981: “where did you live 5 years ago?”

Cohort	Number of Censuses	1981	1986	1991	1996	2001	2006	% linked
2006–2001	2					2,311,000		70.3
2001–1996	2				2,171,000			69.5
1996–1991	2			2,174,000				72.0
1991–1986	2		2,220,000					75.9
1986–1981	2	2,078,000						72.1

- We used the linkage bias weights that had been constructed through earlier work at COMPASS

NEW ZEALAND AS A SOCIAL LABORATORY

- Advent of the NZ Longitudinal Census enabled us to analyse actual transitions between censuses & simulate based thereon
- Offered up the whole age range in one place, and a number of useful variables to look at changes in



THE BASEFILE

- Our basefile was people in private dwellings linked back 1986–1981, plus the 1981 “residual” file
- Problematic retaining household structure when working with links at the individual level
- Whole population too big for constructing models, so we took a 1% sample – for all of the data sets we constructed
- Led to a final basefile of just over 30,000 records. Across the years there were a total of around 7.5 million records linked

VARIABLES FOR EVERYONE

Sex	1 = Male; 2 = Female
Age	In years, continuous
Māori/Pacific/ Asian/Eurother	0 = Not x ethnicity; 1 = x ethnicity
Birthreg	0 = New Zealand; 1 = Oceania / Pacific Islands; 2 = Asia; 3 = Europe; 4 = The Americas; 5 = The Middle East & Africa
Religion	0 = No religion; 1 = Christian religion; 2 = Other religion
Years	0 = Born in New Zealand; 1 = ≤5 years since arrival in New Zealand; 2 = >5 years since arrival in New Zealand
Tenure	0 = Living in owned dwelling; 1 = Living in rented dwelling
Dep	1 = Lowest deprivation quintile; 2 = Second; 3 = Third; 4 = Fourth; 5 = Fifth
H_income	CPI-adjusted household income

EXTRAS FOR ADULTS

Livealone	0 = Not living alone; 1 = Living alone
Partner	0 = Not living with a partner; 1 = Living with a partner
Children	0 = Not living with children; 1 = Living with children
Education	0 = No educational qualification; 1 = Secondary school qualification 2 = Post-school non-university qualification; 3 = University qualification
Student	0 = Not currently in full-time study or training; 1 = Currently in full-time study or training
Emp	0 = In paid employment; 1 = Unemployed; 2 = Not in the labour force
P_income	CPI-adjusted personal income
P_benefits	0 = No personal benefit receipt; 1 = Personal benefit receipt

And for 15–49 year old females only

Newborn	0 = Not living with a 0-year-old child; 1 = Living with a 0-year-old child

PAIRED DATA SETS

- 1% samples were then taken of people linked between each pair of censuses, with those variables at both ends
- For modelling we decided that all missing values in our set of variables needed to be removed
- Enter Stata, which we used to perform Multiple Imputation by Chained Equations for each paired data set and the basefile
  - Paired data sets had to be split into four parts: for the two years and for the different variables held for adults and children

MULTIPLE IMPUTATION (1)

- After initial efforts we reduced the data to have no missing in: Birthreg, Years, and Religion, before the imputation
- Variables were then imputed in all data sets as:

Tenure	Binomial logistic regression
Dep	Ordinal regression
H_income	Linear regression, truncated to a minimum of zero
P_income	Linear regression, truncated to a minimum of zero
P_benefits	Binomial logistic regression
Education	Ordinal regression
Emp	Multinomial logistic regression
Partner	Binomial logistic regression

MULTIPLE IMPUTATION (2)

- Settings were for 9 runs of 20 iterations, with a random seed start, and then model results used as starting values for the next iteration, for the best prediction of imputed values

Variable	15+ Basefile	u15 Basefile	15+ 9691_91	u15 9691_91	15+ 0601_01	u15 0601_01
Tenure	78 (0.4%)	33 (0.4%)	120 (0.8%)	330 (6.9%)	372 (2.2%)	72 (1.5%)
Dep	153 (0.7%)	30 (0.4%)	1,245 (8.0%)	12 (0.2%)	1,566 (9.3%)	474 (9.7%)
H_income	3,537 (16.2%)	1,479 (17.6%)	1,932 (12.4%)	870 (18.1%)	2,427 (14.4%)	771 (15.7%)
P_income	1,764 (8.1%)		606 (3.9%)		771 (4.6%)	
P_benefits	237 (1.1%)		381 (2.4%)		-	
Education	126 (0.6%)		210 (1.3%)		1,332 (7.9%)	
Emp	21 (0.1%)		111 (0.7%)		-	
Partner	141 (0.6%)		54 (0.3%)		228 (1.3%)	
Total n	**21,786 (100%)**	**8,385 (100%)**	**15,591 (100%)**	**4,803 (100%)**	**16,890 (100%)**	**4,896 (100%)**

FINAL DATA SETS

- Distributions of variables with imputed values validated well against their original full distributions!
- We stacked all of the census pair data sets vertically, including "pair" as another variable that could be used as a covariate
- We took "year" data sets from the earlier year in each pair, drawing random samples of births and immigrants
  - Figures from Stats NZ lifetables and official statistics on migration and births added further value for demographic modelling
- These then construct the statistical models of changes over time, which end up in our R Shiny visualisation. Results are still to come!