New Zealand

simario: An R Package for Dynamic Microsimulation

2014 International Methodology Symposium

Statistics Canada, Gatineau, Québec

COMPASS RESEARCH CENTRE

FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

Jessica McLay

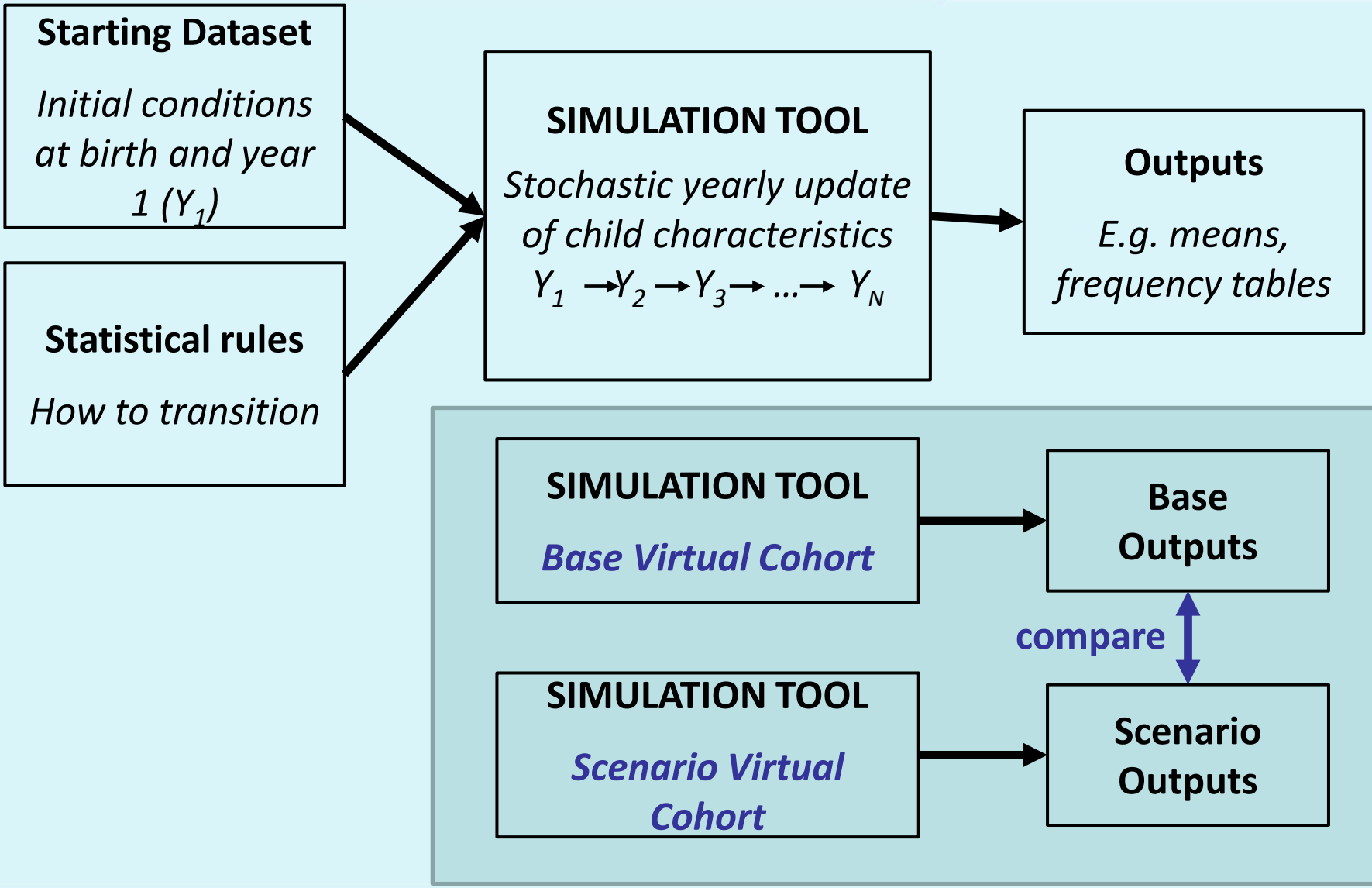
COMPASS Research Centre
Faculty of Arts
University of Auckland
New Zealand

MINISTRY OF BUSINESS, INNOVATION & EMPLOYMENT
HIKINA WHAKATUTUKI

The University of Auckland

What is Dynamic Microsimulation?

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau



Why R?

- Open source and free
  - Anyone can install, use and further develop
  - Availability of public critique and refinement
  - Existing user base
- Designed for data analysis and manipulation
- Flexible
  - Massive 3rd party contribution
  - Libraries for most anything statistical you may want to do

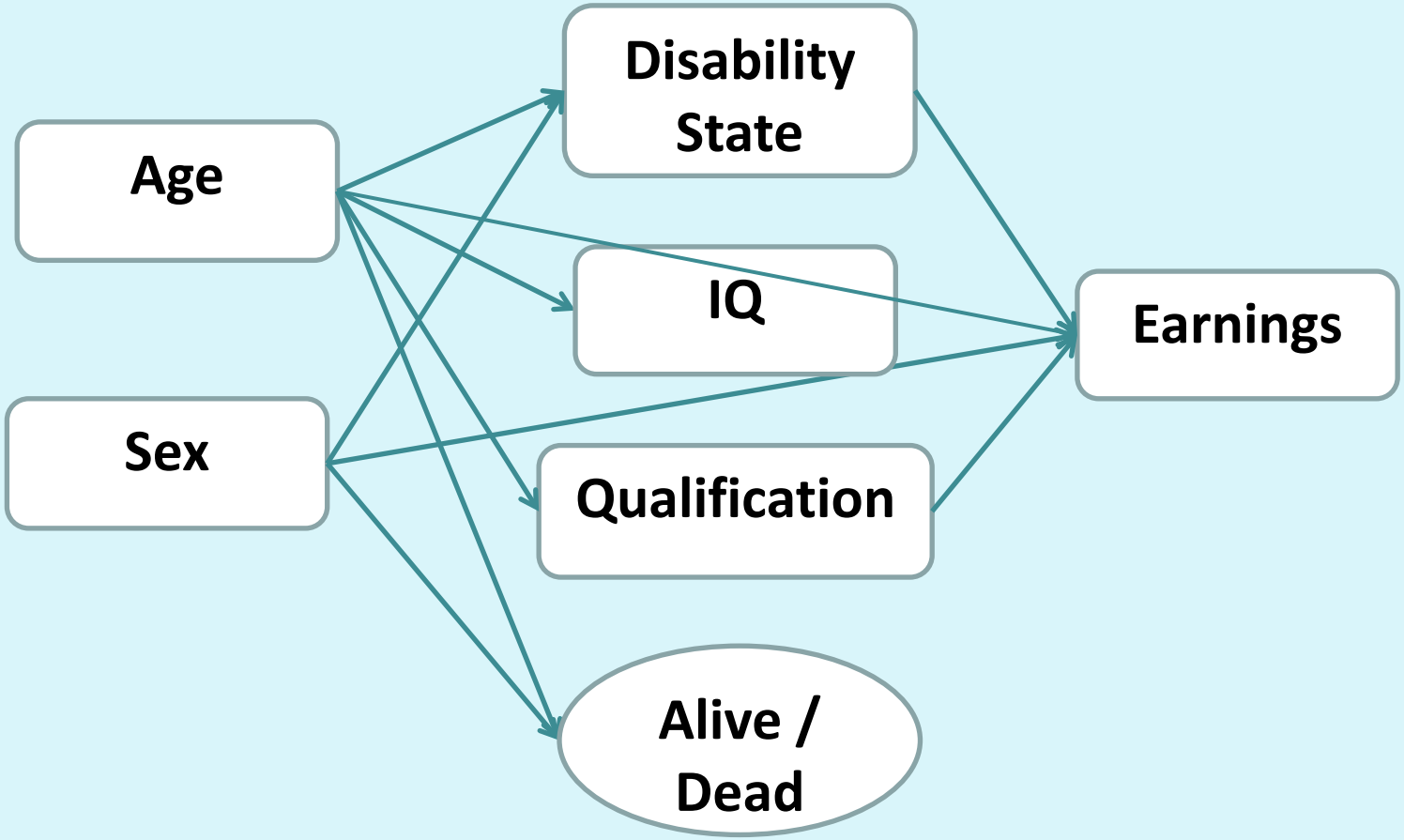
The simario R package

- Purpose: to provide a flexible framework of functions for creating a microsimulation in R
- R package: A collection of related R functions and other R objects (e.g. a dataset)
- *Given the required csv files, use simario functions to programme your microsimulation model from start to finish, then run scenarios*
- Illustration of simario:
  - Setting up (initiation files)
  - The simulation process
  - Outputs
  - Running scenarios
  - Viewing results

Two Types of Functions

- Simario functions:
  - Generic stand-alone functions, no reference to objects outside of the function
- Project-specific shell functions
  - Shell/outline for the programmer to complete with details of their specific microsimulation project

Demonstration Model



Simulate from birth, forward 100 years

Model Initiation

- Files needed prior to programming the simulation in R
  - Starting dataset (.csv)
  - Data dictionary file (.csv)
  - Statistical sub-models (.csv)

Model Initiation

- Complete project-specific initiation function
  - Shell function provided (`initDemo()`)
  - Point to the initiation csv files just discussed,
    - Imports statistical sub-models and starting dataset,
    - Creates objects in the R environment, e.g. empty lists and matrices that will be filled during the simulation
- Fill in other project-specific functions which are called by the initiation function

The Simulation Process

Simulate Run 1
- Year 1
  - Simulate disability state
  - Simulate qualification
  - Simulate IQ
  - Simulate earnings
  - Simulate alive / dead
  - Save simulated values
- Year 2
  - Simulate disability state
  - Simulate qualification
  - …
  - Save simulated values
- …
- Year 100
- Calculate summary statistics

Simulate Run 2

…

Simulate Run M

Calculate means of summary statistics

```
simulateRun <- function() {
  for (year in 2:NUM_YEARS) {
    simulate_disability_state()
    simulate_qualification()
    simulate_IQ()
    simulate_earnings()
    simulate_alive()
    store_current_values_in_outcomes()
  }
}

for (i in 1:total_runs) {
  simulateRun()
  map_outcomes_to_run_results()
}
collate_all_run_results()
```

Predict and Simulate Functions

	predSimNorm()	predSim Binom()	predSimPois()	predSim NBinom()
Variable type	Continuous	Dichotomous	Continuous	Continuous
Type of statistical sub-model	Linear regression	Logistic regression	Poisson regression	Negative binomial regression
Get predicted value for each individual				
Random draw from	Normal dist.	Binomial dist.	Poisson dist.	Negative binomial dist.
Other parameters	SD= residual standard error from model			Dispersion parameter

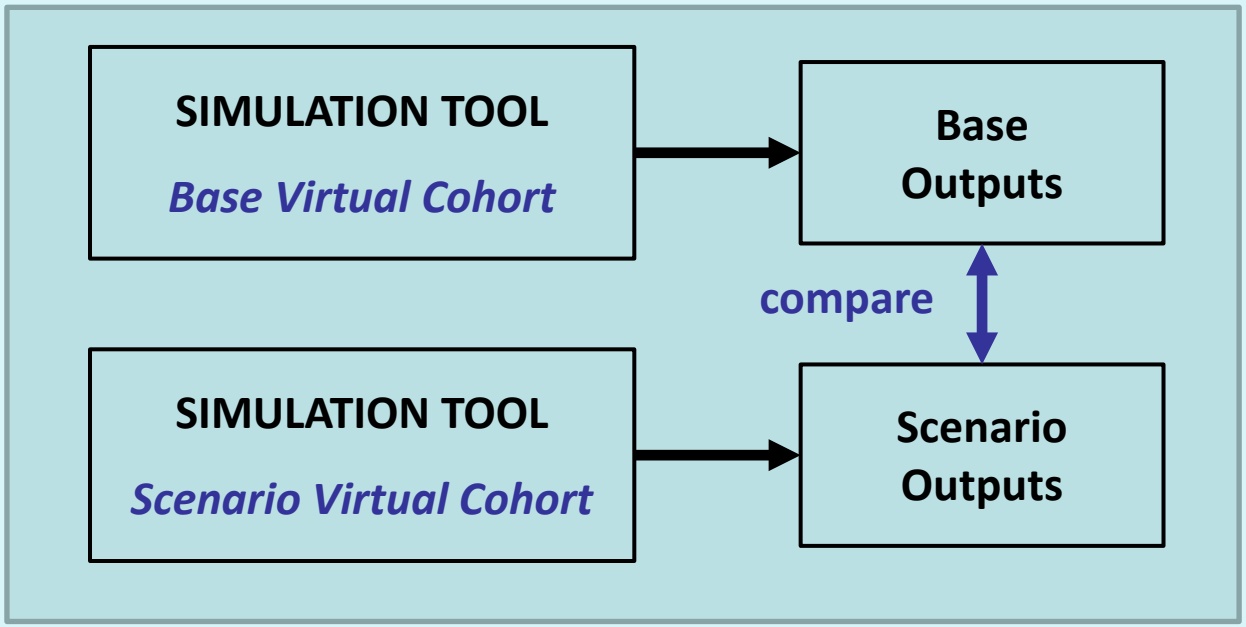
Outputs (Collated Results)

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

- Means
- Percentages
- Quantiles (min, 25th, 20th, 40th, median, 60th, 75th, 80th, max)
- Percentages for categorised continuous variables

The University of Auckland New Zealand

Running Scenarios



Running Scenarios: Scenario Specification

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

Fill in cat.adjustment matrices (created by the initiation function)

	None (%)	Secondary School (%)	Below Degree (%)	Degree (%)
Year 17	NA	NA	NA	NA
Year 18	0	90	10	0
Year 19	0	85	15	0
Year 20	0	80	20	0
Year 21	0	25	25	50
Year 22	0	15	25	60
Year 23	0	15	25	60
Year 24	0	15	25	60
Year 25	NA	NA	NA	NA

New Zealand
The University of Auckland

Running Scenarios: "Adjusting" Data

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

Requested Proportions:

	None (%)	Secondary School (%)	Below Degree (%)	Degree (%)
Year 21	0	25	25	50

Qual	Proportion in Base Simulation	Number in Base Simulation	Number Needed to Match Requested Proportions	Number to Change
None	2.10%	21	0	21
Secondary school	40.20%	402	250	152
Below degree	50.00%	500	250	250
Degree	7.70%	77	500	-423

New Zealand

The University of Auckland

Subgroup Scenarios

- Idea for a program in high schools: Mentors to encourage 16 and 17 years old boys to finish school
- Scenario: **For males only**, increase proportion with a secondary school qualification at age 17/18
- Can program very specific subgroups
  - e.g. `subgroupExpression` **`<- "disability_state==1 & earnings>70000"`**
- Additional "by subgroup" outputs generated for subgroup scenarios:
  - e.g. **`means_by_subgroup`**, **`means_by_subgroup_base_data`**

The tableBuilder() Function

- Results can be viewed by
  - Looking at outputs automatically created
  - Using R manually to investigate/summarise the simulated data (which is stored for each run)
  - Using the tableBuilder() function

The tableBuilder() Function


```
tableBuilder(envName="Base", statistic="means",
variableName="earnings", grpbyName="sex", CI=FALSE)
```

gender		
NA	Male	Female
40	51032	38178
41	49042	38176
42	49551	37326
43	50504	38023
44	49249	36675
45	47808	35814
46	45849	34173
47	43938	34891
48	44599	33541
49	42912	33769
50	41934	32245

Summary: Limitations and Disadvantages of simario

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

- Most suited to dynamic closed cohort models
  - Simulating a set group of individuals over time (no current capacity for individuals to enter or leave the simulation, births and deaths)
- Need to be confident using R
- Level of complexity to fitting all the functions together
- No current capacity for scenarios where the effect of one variable on another is changed

New Zealand
The University of Auckland

Summary: Advantages of simario

- Simario provides a framework for creating a microsimulation model in R
- Good for scenarios that examine the effect of changing peoples actions
- Very flexible
  - Simulating variables
  - Specifying outputs
  - For a given variable, can use different parameters (statistical sub-models) for different cases
  - Confident R programmers can expand and change functions to suit their own purposes

Acknowledgments

- Oliver Mannion
- Barry Milne
- Janet Pearson
- Mengdan Yu
- Roy Lay-Yee
- Martin von Randow
- Peter Davis

More information:

- simario to be published as an R package on CRAN for free download
- Article providing instructions on how to use simario to be published
- Code currently available on google code (search "simario")
- jessica.mclay@auckland.ac.nz

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau
New Zealand
The University of Auckland

New Zealand

Appendix

COMPASS RESEARCH CENTRE

FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND

Whare Wānanga o Tāmaki Makaurau

The University of Auckland

The Simulation Process

Simulating Reading score: Rule from statistical model:

E[reading score] = 13.00 + .91*reading.score.previous + .07*months.breast.fed +
1.04*father.tertairy.qualification + .87*father.secondary.qualification

	Child 1
Characteristics	
Reading score at age 8	40
Number of months breast fed	12
Father's Education	Secondary
Predicted reading score at age 9	`13.00 + .91*40 + .07*12 + .87` = 50.58
Random draw from a normal distribution	50.23
Reading score assigned at age 9	50

Apply Rule

Expected value

Stochastic component

Starting Dataset

- One row per individual
- Provides the starting values from which to simulate all other variables and years

sex	IQ	qualification	disability_state
1	81	1	1
1	72	1	4
1	88	1	1
1	103	1	1
1	101	1	1
1	91	1	1
1	110	1	2
2	111	1	1
2	112	1	1

E.g. Data Dictionary

Varname	Description	Codings_Expr
age	age	
sex	gender	c('Male'=1, 'Female'=2)
Alive	alive	c('Alive'=T, 'Dead'=F)
disability_state	disability state	c('No disability'=1, 'Mild disability'=2, 'Moderate disability'=3, 'Severe disability'=4)
IQ	IQ	
IQ_previous	IQ (prev year)	
qualification	highest qualification	c('None'=1, 'Secondary School'=2, 'Below Degree'=3, 'Degree'=4)
earnings	earnings to date	

Statistical Sub-Model

Sub-model for earnings:

Variable	ClassVal0	Estimate
Intercept		5.23
IQ		0.03
age		0.13
age*age		0.00
qualification	1	-0.30
qualification	2	-0.13
qualification	3	0.11
qualification	4	0.32
sex	2	-0.16
sex	1	0.16
disability_state	2	-0.11
disability_state	3	-0.35
disability_state	4	-0.68
_Alpha		0.61

The Simulation Process

Earnings outcomes from 1 run

Year / Age

Individual	15	16	17	18	19	20	21	22
1	0	1200	1024	9084	7964	12236	8407	10005
2	0	4099	11148	3331	1896	4450	11829	8802
3	0	877	7913	4568	8763	4954	18273	13343
4	0	9927	9376	19271	16069	17514	12998	23982
5	0	6212	2656	43013	18059	21338	15455	89382
6	0	3881	33907	1356	7104	35060	24946	38773
7	0	985	10450	19073	32143	6613	4297	17246
8	0	42974	44194	12982	69105	7547	20857	4481
9	0	13514	30753	36516	20983	30387	37576	12393

The Simulation Process

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

run_results

Year / Age	Earnings means run 1	Earnings means run 2
15	0	0
16	13178	12562
17	14461	13265
18	16133	16501
19	17556	18817
20	19475	21040
21	21846	23968
22	23976	23937

run_results_collated

	Earnings mean of means		
	Mean	Lower	Upper
15	0	0	0
16	13013	12808	13132
17	14771	14566	15073
18	16210	16076	16341
19	17968	16779	18728
20	19650	18745	20321
21	21817	21216	22387
22	23720	23309	24238

New Zealand
The University of Auckland

R Object Structure

COMPASS RESEARCH CENTRE

FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND

Whare Wānanga o Tāmaki Makaurau

New Zealand

The University of Auckland

The tableBuilder() Function

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

Argument	Specifies:	Options / Examples
envName	Which set of simulated data to use	Base or scenario
statistic	Which statistic to calculate	**`frequencies`**, **`means`**, **`quintiles`**
variableName	The variable on which to calculate the statistic	**`earnings`**
grpbyName	An optional variable to group the results by	**`disability_state`**
CI	Whether to calculate confidence intervals	**`TRUE`** or **`FALSE`**
logiset	An optional string expression that defines a group. Only data from this group will be using in calculating the specified statistics.	**`age>20 & age<65`**

New Zealand
The University of Auckland

Running Scenarios: "Adjusting" Data

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

Requested Proportions:

	None (%)	Secondary School (%)	Below Degree (%)	Degree (%)
Year 21	0	25	25	50

Qual	Proportion in Base Simulation	Number in Base Simulation	Number Needed to Match Requested Proportions	Number to Change
None	2.10%	21	0	21
Secondary school	40.20%	402	250	152
Below degree	50.00%	500	250	250
Degree	7.70%	77	500	-423

New Zealand
The University of Auckland

Running Scenarios: "Adjusting" Data

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

Requested Proportions:

	None (%)	Secondary School (%)	Below Degree (%)	Degree (%)
Year 21	0	25	25	50

Qual	Proportion in Base Simulation	Number in Base Simulation	Number Needed to Match Requested Proportions	Number to Change
None	2.10%	21	0	21
Secondary school	40.20%	402	250	152
Below degree	50.00%	500	250	250
Degree	7.70%	77	500	-423

Move 21

New Zealand
The University of Auckland

Requested Proportions:

	None (%)	Secondary School (%)	Below Degree (%)	Degree (%)
Year 21	0	25	25	50

After one step:

Qualification	Number	Proportion	Number to Change
None	0	0	0
Secondary school	423	42.3	173
Below degree	500	50	250
Degree	77	7.7	-423

Requested Proportions:

	None (%)	Secondary School (%)	Below Degree (%)	Degree (%)
Year 21	0	25	25	50

After one step:

Qualification	Number	Proportion	Number to Change
None	0	0	0
Secondary school	423	42.3	173
Below degree	500	50	250
Degree	77	7.7	-423

Move 173

Requested Proportions:

	None (%)	Secondary School (%)	Below Degree (%)	Degree (%)
Year 21	0	25	25	50

After two steps:

Qualification	Number	Proportion	Number to Change
None	0	0	0
Secondary school	250	25	0
Below degree	673	67.3	423
Degree	77	7.7	-423

Requested Proportions:

	None (%)	Secondary School (%)	Below Degree (%)	Degree (%)
Year 21	0	25	25	50

After two steps:

Qualification	Number	Proportion	Number to Change
None	0	0	0
Secondary school	250	25	0
Below degree	673	67.3	423
Degree	77	7.7	-423

Move 423