KIWI. A knowledge-based inquiry tool for policy development using micro-simulation

COMPASS RESEARCH CENTRE

FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND

Whare Wānanga o Tāmaki Makaurau

Statistics Directorate Seminar
OECD, Delta, Room 6104
01 June 2015

Peter Davis and Colleagues

COMPASS Research Centre

University of Auckland

New Zealand

www.compass.auckland.ac.nz

MINISTRY OF BUSINESS, INNOVATION & EMPLOYMENT

HIKINA WHAKATUTUKI

Outline

- **Introduction**
  - The team – from COMPASS research centre
  - The product – micro-simulation (early life-course)
- **Construction**
  - The "end-users" – policy advisers
  - The inquiry system – ingredients/construction
- **Application**
  - Assessing the "social determinants of health" model
- **Extensions**
  - Knowledge "laboratory"; 'Open source' micro-simulation
- **Conclusion**

COMPASS
RESEARCH CENTRE

FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND

Whare Wānanga o Tāmaki Makaurau

New Zealand

The University of Auckland

- **Research-Policy Collaboration – Published 2014**

EVIDENCE
& POLICY

A journal of research, debate and practice

A collaborative approach to bridging the research-policy gap through the development of policy advice software

Barry John Milne, b.milne@auckland.ac.nz
Roy Lay-Yee, r.layyee@auckland.ac.nz
Jessica McLay, jessica.mclay@auckland.ac.nz
University of Auckland, New Zealand
Martin Tobias, martin_tobias@moh.govt.nz
Pat Tuohy, pat_tuohy@moh.govt.nz
Ministry of Health, New Zealand
Ann Armstrong, ann.armstrong@minedu.govt.nz
Ministry of Education, New Zealand
Robert Lynn, robert.lynn@justice.govt.nz
Ministry of Justice, New Zealand
Janet Pearson, je.pearson@auckland.ac.nz
Oliver Mannion, o.mannion@auckland.ac.nz
Peter Davis, pb.davis@auckland.ac.nz
University of Auckland, New Zealand

We have developed a software-based tool to support a dynamic micro-simulation model of life-course development (to age 13) as an aid to policy makers assessing the impact of policies affecting children. We demonstrate how this approach bridges the research-policy gap by creating: (1) an easy transfer of evidence in a form that policymakers can use (for example, 'What is the policy influence of X on Y?'); and (2) a 'pull' system of knowledge transfer by which policy makers control the knowledge they access. The advantage of close collaboration with policy makers in the development and implementation phases is also discussed.

3

Senior RF – Roy Lay-Yee

COMPASS RESEARCH CENTRE

FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND

Whare Wānanga o Tāmaki Makaurau

Determinants and Disparities – Published 2015

Determinants and disparities: A simulation approach to the case of child health care

CrossMark

Roy Lay-Yee[*], Barry Milne, Peter Davis, Janet Pearson, Jessica McLay

Centre of Methods and Policy Application in the Social Sciences, Faculty of Arts, University of Auckland, New Zealand

ABSTRACT

Though there is much agreement on the importance of the social determinants of health, debate continues on suitable empirically-based models to underpin efforts to tackle health and health care disparities. We demonstrate an approach that uses a dynamic micro-simulation model of the early life course, based on longitudinal data from a New Zealand cohort of children born in 1977, and counterfactual reasoning applied to a range of outcomes. The focus is on health service use with a comparison to outcomes in non-health domains, namely educational attainment and antisocial behaviour. We show an application of the model to test scenarios based on modifying key determinants and assessing the impact on putative outcomes. We found that appreciable improvement was only effected by modifying multiple determinants; structural determinants were relatively more important than intermediary ones as potential policy levers; there was a social gradient of effect; and interventions bestowed the greatest benefit to the most disadvantaged groups with a corresponding reduction in disparities between the worst-off and the best-off. Our findings provide evidence on how public policy initiatives might be more effective acting broadly across sectors and across social groups, and thus make a real difference to the most disadvantaged.

New Zealand

The University of Auckland

4

COMPASS RESEARCH CENTRE

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

New Zealand

The University of Auckland

Regression Estimation for Dynamic Microsimulation (McLay et al.)

- ACCEPTED WITH REVISIONS (International Journal of Microsimulation)

Abstract: Microsimulation models seek to represent real-world processes and can generate extensive amounts of synthetic data. Most often, the parameters that drive the data generation process are estimated by statistical modelling techniques, such as regression. But which techniques are best suited to this purpose? We assess the performance of five regression-style estimation techniques: ordinary least squares regression with a lagged dependent variable, random effects with and without an autoregressive order 1within-unit error structure, a hybrid model combining features from both econometric fixed effects and random effects models, and a dynamic panel model estimated with system generalised method of moments. The criterion for good performance was the proximity of fit of simulated data to empirical data on various characteristics. It was found that ordinary least squares regression with a lagged dependent variable out-performed the other techniques. Random effects with autoregressive errors of the first order was the next best, followed by standard random effects. The dynamic panel model came fourth followed by the hybrid model. This empirical assessment provides practical guidance to those contemplating dynamic microsimulation and other applications using regression-style techniques of synthetic data generation.

5

The Product: Micro-simulation.

- **We start with a sample of individuals (children)**
  - Real (studies) / synthetic (derived from Census)

- **We derive statistical rules to create a 'virtual cohort' through to age 13**
  - Derive rules best able to reproduce study data
  - Apply these rules to the base file to create a synthetic sample of children with typical biographies

- **We then simulate what might happen if policy were to change, by altering parameters**
  - Using software application to test counterfactuals

New Zealand

The University of Auckland

6

Outline

- Introduction

    - **ANY BRIEF QUESTIONS AT THIS POINT?**

- Construction
    - The "end-users" – policy advisers
    - The inquiry system – ingredients/construction
- Application
    - Assessing the "social determinants of health" model
- Extensions
    - Knowledge "laboratory"; 'Open source' micro-simulation
- Conclusion

The "End Users": Policy advisers

- **End Users Group:**

  Ministry of Social Development (MSD)

  Ministry of Health (MoH)

  Ministry of Education (MinEdu)

  Ministry of Justice (MoJ)

- Drive development
- Collaborative approach
- Suggest scenarios

New Zealand

The University of Auckland

Scenarios to test

1. Are children in households where both parents are working better off?

2. How does smoking in pregnancy affect later outcomes?

3. How can we improve early literacy, school achievement and reduce failure in the job market?

4. How does single parenting affect later conduct problems?

5. What interventions have impact on later (health, wealth, social, education, justice) outcomes for Māori, Pacific or low-socio-economic status groups?

New Zealand

The University of Auckland

9

The Inquiry System: six key ingredients

- **Knowledge-based inquiry system (KIWI)**

  1. A synthetic base file representative of the population

  2. A number of real-world longitudinal studies

  3. A technique for combining the data from 3 studies

  4. A statistical model mimicking life-course biographies

  5. A tool that helps interrogation of these biographies

  6. [Parameter estimates drawn from the literature]

1. Synthetic Base File
## (this work due to Barry Milne)

- Subset NZ 2006 Census to include just new-borns (0-year olds) and their parents
  - Randomly select 10,000 (from 50,000)

- Calculate distance (Euclidean) between each of the 10,000, based on 52 Census characteristics.
  - Done separately by family-type

- Choose the closest 2 ranks to form 10,000 clusters of 3 individuals

Methods

- **Randomly choose which "real" child's characteristics are used for each synthetic one**
  - Characteristic by characteristic

Characteristic	Cluster of 3 Children			Random Draw {1,2,3}	Synthetic child
	Child 1	Child 2	Child 3		
Child sex	Male	Female	Female	2	Female
Mother age	29	41	31	1	29
Father age	32	40	38	1	32
Home ownership	Owned	Owned	Rented	3	Rented
Deprivation score (1-10)	9	7	8	3	8

1. Synthetic Base File

- ▣ Voilà! A synthetic base-file of 10,000 composite individuals
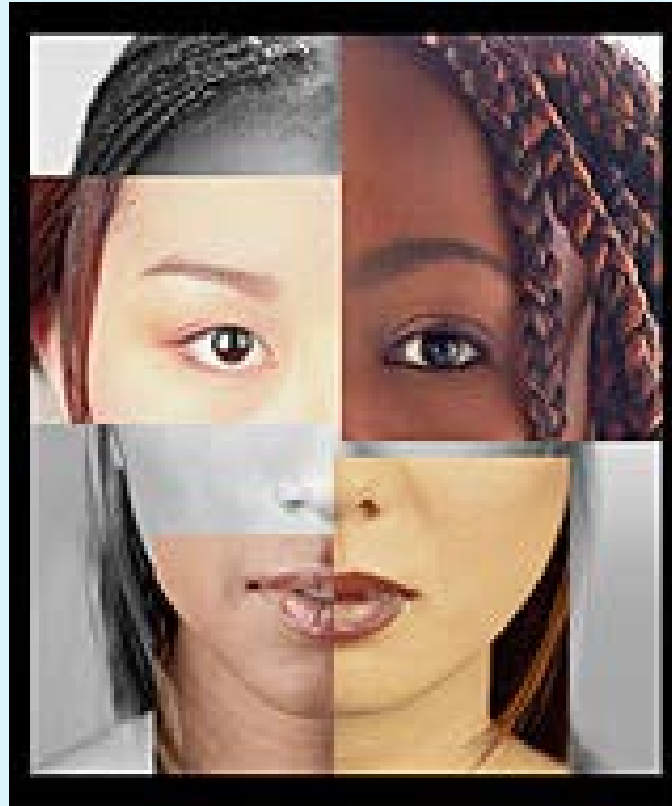
2. Four Studies

- Christchurch Health & Development Study (CHDS)
  - 1265 children born in Christchurch 1977. Followed since

- Dunedin Multidisciplinary Health & Development Study (DMHDS)
  - 1037 children born in Dunedin 1972/3. Followed since

- Pacific Islands Families Study (PIFS)
  - 1398 children born at Middlemore Hospital, 2000, with at least one parent of Pacific Islands ethnicity. Followed since

- Te Hoe Nuku Roa Study (THNR) **[calibration only]**
  - Longitudinal study of Māori households (beginning 1995)
    - Auckland, Wellington, Manawatu, Gisborne, Northland, Southland, Nelson
  - 568 children (0-12) assessed at least twice in four waves

New Zealand

The University of Auckland

3. Data Integration
(due to Barry Milne and Jessica McLay)

- Associations between X & Y assessed using longitudinal regression analyses
  - Utilises data from all the ages available from the three studies (THNR not used)

Age	Y_{CHDS}	Y_{DMHDS}	Y_{PIFS}	X_{CHDS}	X_{DMHDS}	X_{PIFS}
Birth	✓	✓	✓	✓	✓	✓
1	✓		✓	✓		✓
2	✓		✓	✓		✓
3	✓	✓		✓	✓	
4	✓		✓	✓		✓
5	✓	✓		✓	✓	
6	✓		✓	✓		✓
7	✓	✓		✓	✓	

Stack All Three Datasets

CHDS
(n=1265)

Dunedin
(n=1037)

PIFS
(n=1398)
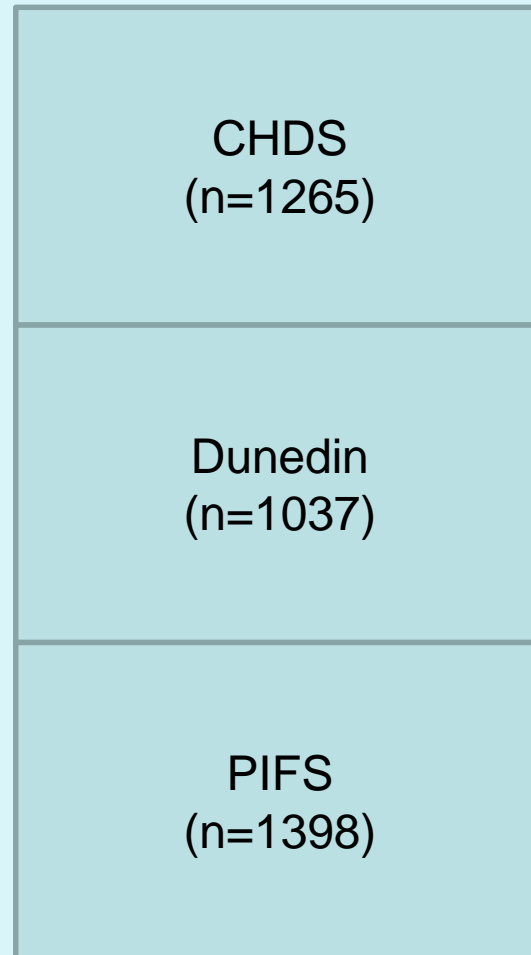
N=3700

4. A Statistical Model

(this work due to Jessica McLay)

Regression Techniques for Dynamic Micro-simulation:

An Empirical Performance Assessment

- Background

- Aims

- Statistical Modelling Techniques

- Empirical Assessment Methods

- Results

- Conclusion

The Simulation Process

Simulating Reading score: Simplified 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 A
Characteristics	
Reading score at age 8	40
Number of months breast fed	12
Father's Education	Secondary

Apply Rule

Predicted reading score at age 9	13.00 + .91*40 + .07*12 + .87 = 50.58

Expected value

Random draw from a normal distribution	50.23

Stochastic component

Reading score assigned at age 9	50

New Zealand

The University of Auckland
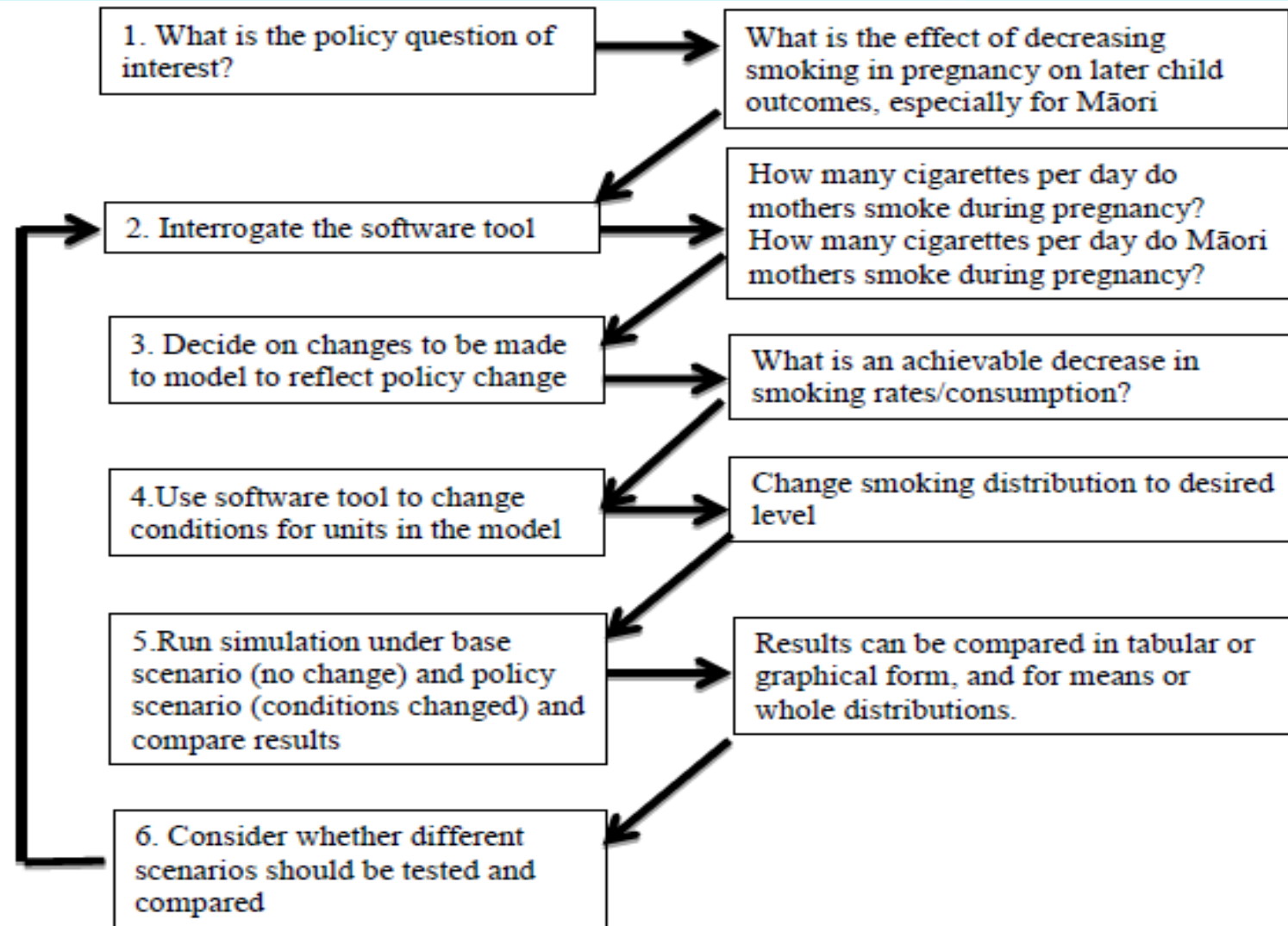
Virtual versus real cohort: family doctor visits, reading ability, and conduct problems, by year of age

Year	Real cohort (CHDS) n=1017	Virtual cohort (simulated) n=1017	Absolute error	Absolute error / CHDS mean
Family doctor visits (mean (95% CI))				
1	5.82	5.82	-	-
2	5.34	5.28	0.06	-
3	3.31	3.18	0.13	-
4	3.13	3.15	0.02	-
5	3.22	3.12	0.10	-
6	3.35	3.32	0.03	-
7	2.43	2.41	0.02	-
8	2.14	2.15	0.01	-
9	1.96	1.90	0.06	-
10	1.65	1.68	0.03	-
All years	3.24	3.20 *(3.15-3.25)*	0.04	1.2%
Reading ability: BURT score (mean (95% CI))				
8	45.3	45.3	-	-
9	54.4	54.7	0.3	-
10	64.1	63.7	0.4	-
11	72.8	71.9	0.9	-
12	79.5	78.9	0.6	-
13	85.2	84.6	0.6	-
All years	66.9	66.5 *(65.7-67.4)*	0.4	0.6%
Conduct problems (mean (95% CI))				
6	10.6	10.6	-	-
7	24.6	24.8	0.2	-
8	24.4	25.0	0.6	-
9	24.7	25.3	0.6	-
10	24.9	25.6	0.7	-
All years	21.8	22.3 *(22.1-22.4)*	0.5	2.3%

5. Inquiry Tool
(due to Barry Milne)

1. What is the policy question of interest? → What is the effect of decreasing smoking in pregnancy on later child outcomes, especially for Māori

2. Interrogate the software tool → How many cigarettes per day do mothers smoke during pregnancy? How many cigarettes per day do Māori mothers smoke during pregnancy?

3. Decide on changes to be made to model to reflect policy change → What is an achievable decrease in smoking rates/consumption?

4. Use software tool to change conditions for units in the model → Change smoking distribution to desired level

5. Run simulation under base scenario (no change) and policy scenario (conditions changed) and compare results → Results can be compared in tabular or graphical form, and for means or whole distributions.

6. Consider whether different scenarios should be tested and compared

New Zealand

The University of Auckland

20

Outline

- **Introduction**

- **Construction**

  - **ANY BRIEF QUESTIONS AT THIS POINT?**

- **Application**

  - Assessing the "social determinants of health" model

- **Extensions**

- **Conclusion**

Enhancing social policy outcomes: How important are structural factors?

Sociology Seminar Series

21 May 2014

Roy Lay-Yee and COMPASS colleagues

COMPASS Research Centre

University of Auckland, New Zealand

www.compass.auckland.ac.nz

Social determinants (SDs) of health

- Health disparities are rooted in SDs that confer differential vulnerability to poor health or exposure to conditions that produce poor health

- *Structural* factors comprise SDs of health disparities (that are also SDs of health) while *intermediary* factors comprise other SDs of health (only)

- Debate as to relative importance, as effective policy levers, of structural or intermediary factors

Assessing counterfactuals

▣ **Counterfactual paradigm of causal reasoning**

- ⊞ If the putative causal factor had not been present, we would not have observed the recorded outcome.

  - Randomised Controlled Trials (RCTs)

  - Experimental and quasi-experimental methods

  - Observational designs and statistical analysis

- ⊞ **Simulation techniques**

Research questions

- What is the effect of improving various factors on levels of family doctor (GP) visits among children?

  - Single or multiple factors?

  - Are structural or intermediary factors more influential?

  - Is there greater impact on socially disadvantaged groups?

- (Do the same mechanisms operate for outcomes in other domains, e.g. reading ability or conduct problems?)

Model of structural and intermediary influences on child outcomes (Christchurch study data only)

Structural factors (fixed)

Child
- Gender
- Ethnicity

Parental
- Age (at birth of child)
- Ethnicity
- Education

Familial
- Socio-economic position (at birth of child)

Proxy indicators (modifiable)

Family composition
- Single- or two-parent
- Number of children

Income source
- Parent employment
- Welfare dependence

Intermediary factors (modifiable)

- Owned/ rented home
- Overcrowding
- Accommodation type
- Change of parent
- Change of residence
- Parental smoking

Outcomes

GP visits

Determinants of health inequities (and health)

Determinants of health

26

Scenario testing procedure

1. We 'improved' *single* factors and assessed the degree of impact on outcome

2. We 'improved' *multiple* factors simultaneously

3. We compared the relative effects of 'improving' structural and intermediary factors

4. We posed 'best case scenarios' by 'improving' structural and intermediary factors *simultaneously*

New Zealand

The University of Auckland

GP Visits. Disparities: absolute change
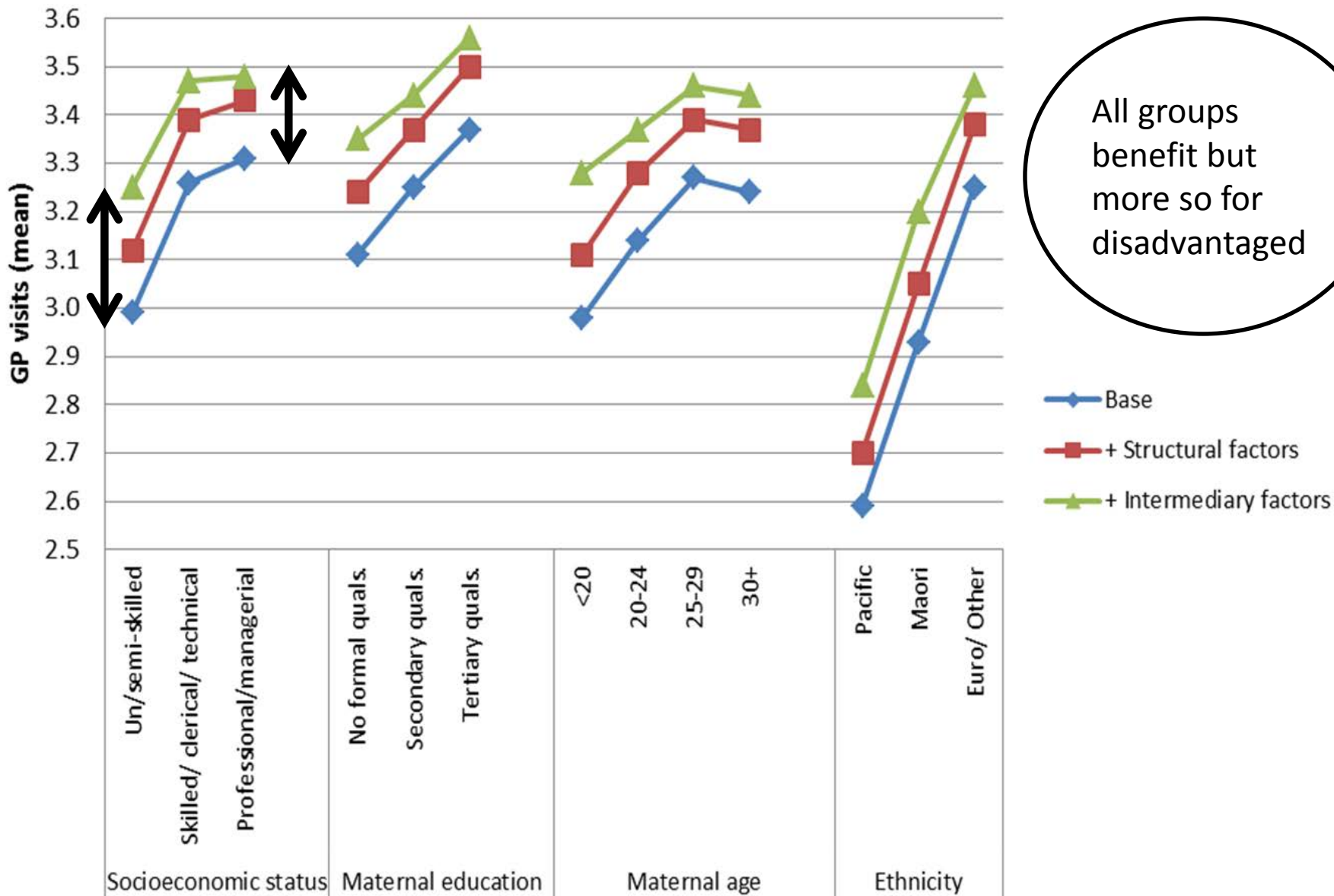
All groups benefit but more so for disadvantaged

GP visits (mean)

y-axis: 2.5, 2.6, 2.7, 2.8, 2.9, 3.0, 3.1, 3.2, 3.3, 3.4, 3.5, 3.6

Legend:
- Base
- + Structural factors
- + Intermediary factors

Socioeconomic status: Un/semi-skilled, Skilled/ clerical/ technical, Professional/managerial

Maternal education: No formal quals., Secondary quals., Tertiary quals.

Maternal age: <20, 20-24, 25-29, 30+

Ethnicity: Pacific, Maori, Euro/ Other

GP visits: Determinants

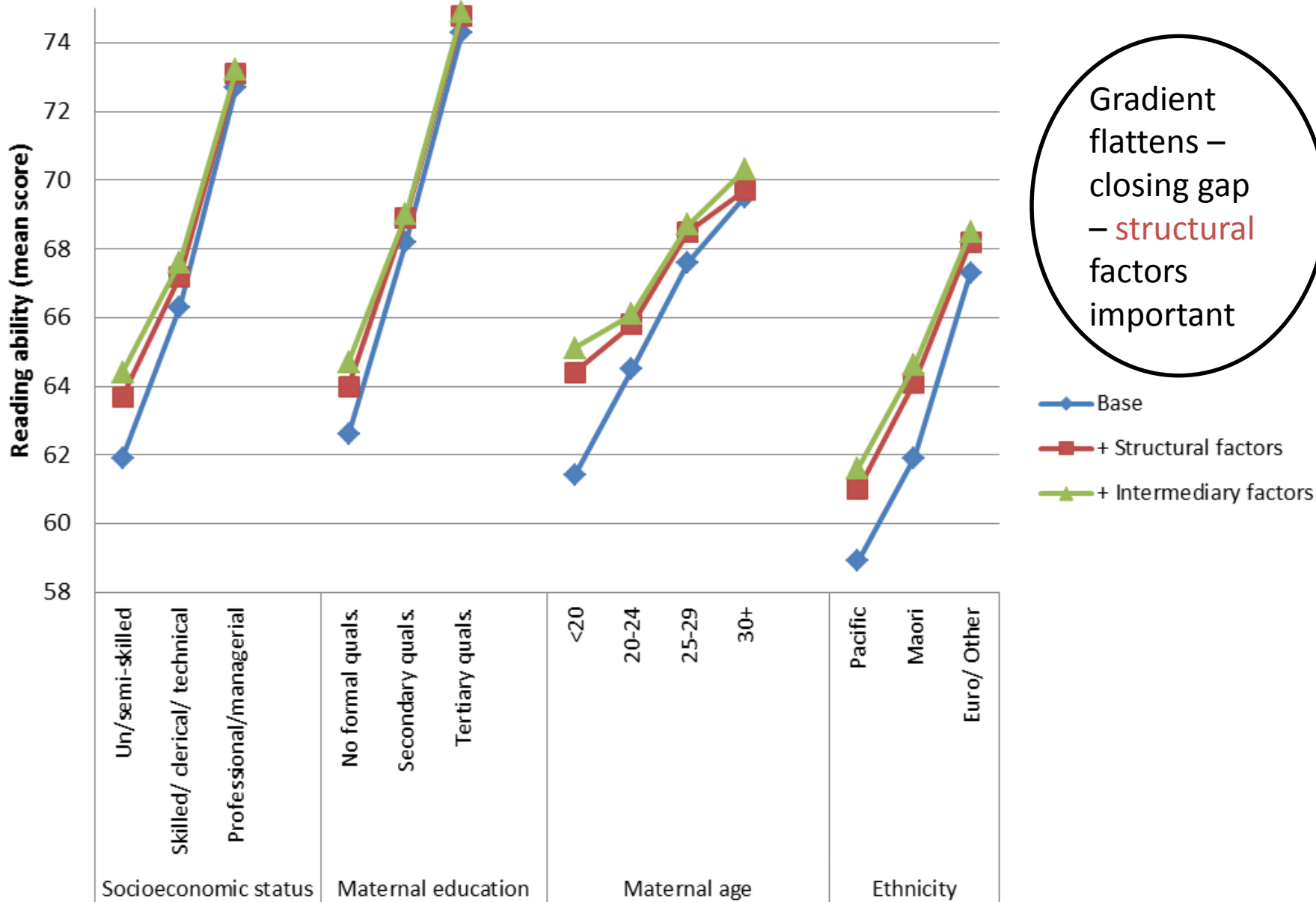| Scenarios	GP Visits (years 1-10) n=1017	
	Mean p.a.	% change
1. Base	3.20	
2. Improve structural factors only		
Fewer children	*3.31*	*+3.4%*
ALL	3.33	+4.1% *
3. Improve intermediary factors only		
Own home	*3.26*	*+1.9%*
ALL	3. 28	+2.5%
4. Best scenario: Improve both structural and intermediary factors	3.41	+6.6% *
		* p<0.05

Outcome: Reading ability

Increasing the reading score is interpreted as an improvement in outcome

Reading Ability. Disparities: absolute change

Gradient flattens – closing gap – structural factors important

Legend:
- Base
- + Structural factors
- + Intermediary factors

Y-axis: Reading ability (mean score)

X-axis categories:
- Socioeconomic status: Un/semi-skilled, Skilled/ clerical/ technical, Professional/managerial
- Maternal education: No formal quals., Secondary quals., Tertiary quals.
- Maternal age: <20, 20-24, 25-29, 30+
- Ethnicity: Pacific, Maori, Euro/ Other

Outcome: Conduct problems

Reducing the number of conduct problems per year is interpreted as an improvement in outcome

Conduct Problems. Disparities: absolute change
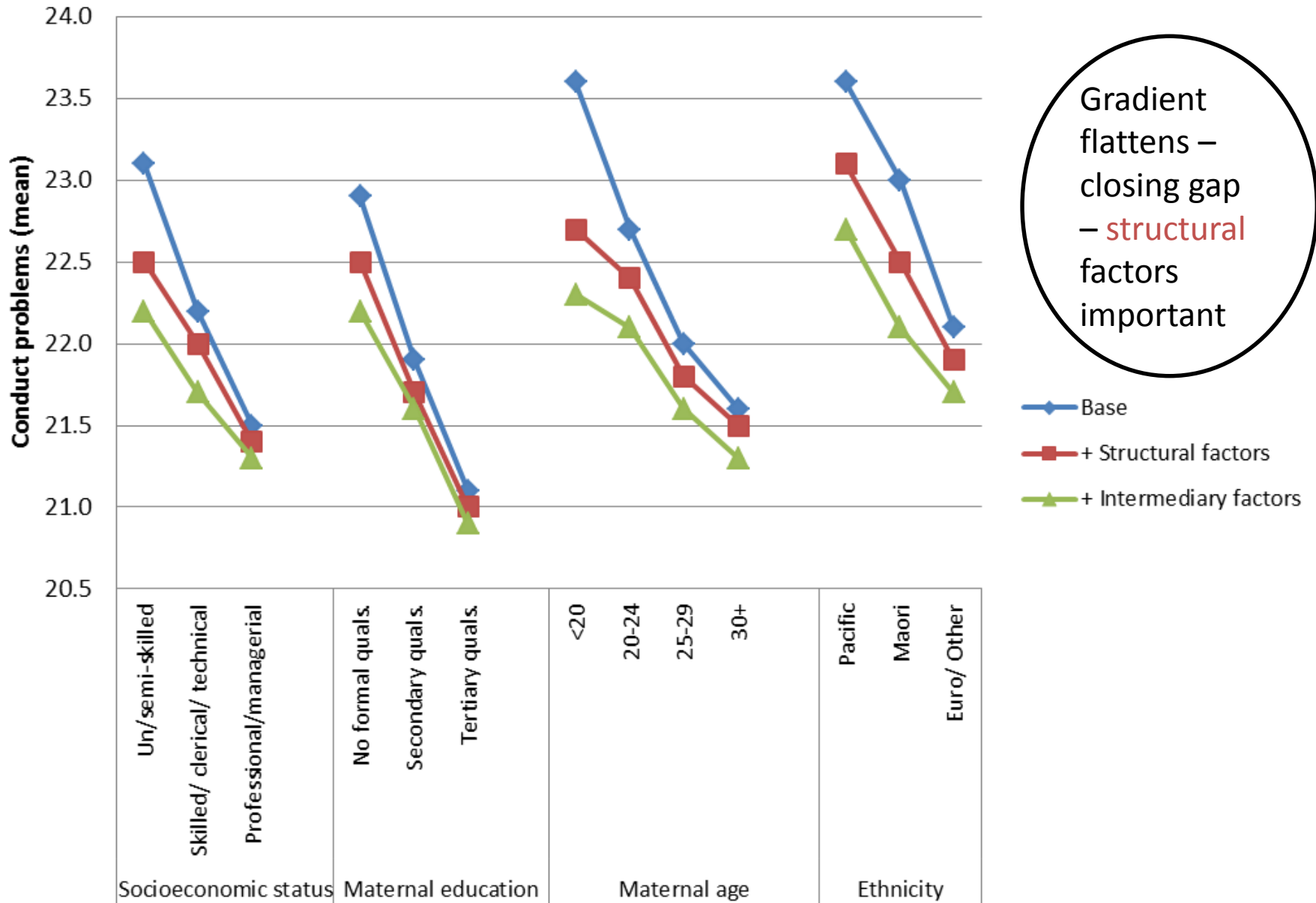
Gradient flattens – closing gap – structural factors important

Legend:
- Base
- + Structural factors
- + Intermediary factors

Y-axis: Conduct problems (mean), 20.5 to 24.0

Categories:
- Socioeconomic status: Un/semi-skilled, Skilled/ clerical/ technical, Professional/managerial
- Maternal education: No formal quals., Secondary quals., Tertiary quals.
- Maternal age: <20, 20-24, 25-29, 30+
- Ethnicity: Pacific, Maori, Euro/ Other

Summary of results

- Q1: Changing a single factor has slight effect on outcome - appreciable effect only by changing multiple factors

- Q2: Effect of modifiable structural factors is greater than of intermediary factors

- Q3: "Inverse" effect gradient: i.e. progressively more positive impact on outcome with greater social disadvantage

- [Q4: Similar findings for range of outcomes in different domains]

New Zealand

The University of Auckland

Outline

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

- Introduction

- Construction

- Application
  - **ANY BRIEF QUESTIONS AT THIS POINT?**
- Extensions
  - Knowledge "laboratory"; 'Open source' micro-simulation
- Conclusion

New Zealand

The University of Auckland

Inquiry System
- Strategic Observations

- **Astute observation 1**
  - There are many well-established estimates for factors that impact the lives of children, but these exist in isolation; micro-simulation offers a way to bring these together
    - John Lynch, Professor of Public Health, University of Adelaide

- **Astute observation 2**
  - 'Best' estimates are thought to be derived from systematic reviews/meta analyses, but it is difficult to test their validity.
    - David Gough, Professor of Evidence Informed Policy and Practice, Institute of Education/UCL

Knowledge Laboratory - Plan

- **Identify key determinants of child and adolescent outcomes**

- **Integrate estimates from systematic reviews/meta analyses into working model of early life course**
  - Developed from stage one; extended in breadth (more determinants and outcomes), and length (to age 18)

- **Use as knowledge laboratory**
  - Test the validity of 'best' estimates
  - Test policy scenarios using validated model

New Zealand

The University of Auckland

Knowledge Lab - Current work

■ Focussing on modelling six constructs

Alcohol and drug use	Ethnicity	Justice contacts	Physical activity
Ambulatory Sensitive Hospitalisations	Family transitions – formation/disintegration	Lead Maternity Carer enrolment	School type (single sex/co-ed)
Asthma/respiratory health	Food in schools	Maltreatment	Smoking
Birth weight/gestational age	Health visits	Mental Health	Socioeconomic measures (income, deprivation, living standards)
Books in home	Home visiting	Nutrition	Suicide
Breastfeeding	Housing quality	Obesity	Teaching quality
Conduct disorder	Immunisation	Otitis Media	Transfer payments
Early Childcare education (amount, quality, type)	Injuries	Parental and intergenerational welfare dependence	Transition to employment
Early parenting	Involvement in Child Health groups (e.g., plunket)	Parental involvement in schools	Violence in families
Education		Parental mental health	

New Zealand

The University of Auckland

Extension 2 – open source model generator

- ➡ **Modgen (Model generator)**
  - ➕ Modgen (Model generator) is a generic micro-simulation programming language supporting the creation, maintenance and documentation of dynamic micro-simulation models.

- ➡ **OpenM++:open source platform**
  - ➕ OpenM++ is an open source micro-simulation platform inspired by and compatible with Modgen. OpenM++, compared to its closed source predecessor, has many distinct features like portability, scalability and open source.

Extension 2 – Engaging with key policy actors

NGOs	Topic	Life Stage	Topic	Government
Child Poverty Action Group	Poverty	**Childhood**	Performance	Treasury
Ngai Tahu tribe	Work benefits	**Adulthood**	Tax credits	Inland Revenue Department
Age Concern	Future of Super	**Retirement**	Future of Super	Ministry of Social Development

New Zealand

The University of Auckland

Conclusion

- **Computational social science has much to offer:**

  - Can mimic social processes (e.g. life course trajectory)
  - Can address data shortcomings (e.g. sample size, data sources)
  - May also provide an approximation to causal analysis

- **Micro-simulation and decision support/inquiry system**

  - With the right empirical and conceptual "anchoring", and working closely with colleagues in the policy process, our tool (KIWI) could be the basis of a more "evidence-informed" policy approach

- **Future plans:**

  - Insert effect estimates from the literature (knowledge "laboratory")
  - Assess more complex interventions and outcomes
  - Improve causal power of underlying statistical analysis