Innovations in Applied Data Symposium: Abstracts

Te Rourou Tātaritanga and Social Wellbeing Agency

3 June 2021

There are 57 abstracts.

## Constructing a spatial index of public transport supply across Auckland, Canterbury and Greater Wellington.

### Adam Ward

*Head of Research at DOT Loves Data*

[adam@dotlovesdata.com](mailto:adam@dotlovesdata.com)

In this study we constructed a low-level spatial index of public transport (PT) supply across the 25,000 meshblocks in the Auckland, Canterbury and Greater Wellington regions. Using Open Route planner and Google Transit Feed data, for each meshblock in the study area we calculated the walking distance to the nearest PT location, the number of accessible locations (per capita), the number of accessible routes and (peak-time) services and the transit time from place of residence to place of work via PT networks. These metrics were then combined into an overall PT Supply Score. The results of this study are visualised in an interactive Power BI dashboard.

(105 words)

***Keywords (4):*** *Public Transport, Geo-spatial, Google Transit Feed Data, Power BI*

## Modelling the transfer of tacit knowledge on temporal bipartite networks

### Adrian Ortiz-Cervantes

*University of Auckland and Te Pūnaha Matatini*

[A.Ortiz-Cervantes@auckland.ac.nz](mailto:A.Ortiz-Cervantes@auckland.ac.nz)

In this work we analyze the employee mobility in New Zealand and its repercussions on knowledge and skills transfer amongst different industries, in the context of in temporal complex networks. Using tax records from all the employees and firms in New Zealand from the year 2000 to 2017, we created a bipartite temporal network of employers and employees that allow us to implement a model for the transfer of tacit knowledge between employees, and estimate the stock of knowledge inside firms and sectors across different skills.

(84 words)

***Keywords (3):*** *Complex Networks, Knowledge Transfer, IDI*

## Automating national land cover mapping using artificial intelligence

### Alexander Amies, Jan Schindler

*Manaaki Whenua – Landcare Research*

[amiesA@landcareresearch.co.nz](mailto:amiesA@landcareresearch.co.nz)

Land cover mapping is important for national- and regional-scale policy decision making, and environment and biodiversity monitoring. Previous time steps of Manaaki Whenua’s Land Cover Database (LCDB) have leveraged expert analysis of large volumes of multi-temporal satellite imagery to identify areas of change. This approach traditionally relies on rule-based decision systems in conjunction with time consuming manual data cleaning. We explore the potential of artificial intelligence for developing more flexible, automated processes for land cover classification and change detection. We compare machine learning and object-based methods with image-based convolutional encoder-decoders enabling more fine-grained spatial and temporal predictions.

(97 words)

***Keywords (5):*** *Remote sensing, land cover, machine learning, deep learning, artificial intelligence*

## Improved synthetic data method for Stats NZ datalab microdata output

### Alistair Ramsden

*Statistical Methods Census Methodology Team, Stats NZ*

[alistair.ramsden@stats.govt.nz](mailto:alistair.ramsden@stats.govt.nz)

Stats NZ has approved a variation to Microdata Output Guide confidentiality rules, for the ‘He Ara Poutama mō te reo Māori SURF’ datalab project.

The synthetic data method is Classification And Regression Tree (CART) modelling, implemented using the Synthpop R package. The difference compared to existing rules is to generate, test, and release synthetic data counts tables with noised counts values {0,3,6,9,12,…}. Previously such values were {‘Suppressed’,6,9,12,…}.

The new method releases data with better utility (inferential validity), yet retaining adequate safety (disclosure control).

Next steps are to calculate Differential Privacy (DP) parameters {epsilon, delta} for this data and method.

(95 words)

***Keywords (4):*** *confidentiality, methodology, datalab, synthetic data*

## Application of data science methodologies to explore, predict, and model wellbeing outcomes using the New Zealand Integrated Data Infrastructure (IDI).

### Anantha Narayanan, Tom Stewart, Scott Duncan

*Auckland University of Technology*

[anantha.narayanan.tl@aut.ac.nz](mailto:anantha.narayanan.tl@aut.ac.nz), [tom.stewart@aut.ac.nz](mailto:tom.stewart@aut.ac.nz), [scott.duncan@aut.ac.nz](mailto:scott.duncan@aut.ac.nz)

Wellbeing measures currently available in New Zealand are not sensitive enough to capture the effects of policy change – this is partly due to the lack of detailed population-level wellbeing data. Microdata within NZ’s IDI promises to provide a better understanding of wellbeing; however, detailed measures of wellbeing (from the General Social Survey) are only available for a smaller subset of IDI records (i.e., ~8,500 individuals). Extrapolating these wellbeing data to the full IDI population may be possible via advanced data modelling. In this research, we aim to apply machine learning techniques that can predict wellbeing outcomes from IDI administrative data.

(99 words)

***Keywords (3):*** *population-wellbeing, machine learning, big data*

## Epidemiology for non-epidemiologists

### Andrew Sporle, Daniel Barnett

*iNZight Analytics and Healthier Lives National Science Challenge*

[a.sporle@auckland.ac.nz](mailto:a.sporle@auckland.ac.nz)

(0 words)

## Can behavioural science help stop callers from hanging up?

### Caitlin Spence

*Behavioural Science Aotearoa, Ministry of Justice*

[Caitlin.Spence@justice.govt.nz](mailto:Caitlin.Spence@justice.govt.nz)

Thousands of people ring the Ministry of Justice every week, but some abandon the call (hang up) before they reach an agent. With the aim of keeping callers on the line, Behavioural Science Aotearoa worked with the contact centre to develop three sets of hold messages designed to reduce caller abandonment. The messages are currently being trialled using a quasi-experimental ‘reversal’ design. The team take advantage of administrative call-level data to measure impact by applying survival analysis techniques.

*Behavioural Science Aotearoa apply behavioural science across the Justice Sector to build more effective, people-centred justice services and policies that improve outcomes for all.*

(101 words)

***Keywords (3):*** *behavioural science, contact centre, survival analysis*

## How do victims of crime interact with the government?

### Callum Sleigh

*Taylor Fry*

[callum.sleigh@taylorfry.co.nz](mailto:callum.sleigh@taylorfry.co.nz)

A large proportion of victimisation in New Zealand is unreported. Because of this, the Ministry of Justice has commissioned a series of surveys to understand the nature and prevalence of unreported victimisation. By linking this survey data to administrative datasets in the IDI it is possible to understand government service use for victims who do not necessarily report to the Police.

(61 words)

***Keywords (3):*** *surveys, victimisation, administrative data*

## Adopting an improvement science approach to inform ongoing implementation of the Mana Ake initiative in Canterbury

### Caralyn Purvis

*Canterbury District Health Board*

[Caralyn.Purvis@cdhb.health.nz](mailto:Caralyn.Purvis@cdhb.health.nz)

Recognising the value of data and analytics saw the integration of an improvement science approach to inform service-delivery, within a cross-sector mental health initiative — Mana Ake. An increased prevalence of wellbeing concerns resulted in a holistic, locally-informed, collaborative initiative for tamariki. Internal review of mixed-methods data has allowed for a truly data-informed iterative approach to service improvement. Consequently, evaluation has influenced the responsiveness of the initiative to optimally meet the needs of the communities it supports. A collaborative initiative between health, education, and wider communities. Evaluation domains include: tamariki, whānau, schools/kura, communities, system.

More information: www.Manaake.health.nz

(95 words)

***Keywords (5):*** *Tamariki, wellbeing, health, education, evaluation*

## Using administrative data to better model receipt of transfers

### Cory Davis, Luke Symes

*Te Tai Ōhanga – The Treasury*

[Cory.Davis@treasury.govt.nz](mailto:Cory.Davis@treasury.govt.nz), [Luke.Symes@treasury.govt.nz](mailto:Luke.Symes@treasury.govt.nz)

Microsimulation models such as the New Zealand Treasury’s Tax and Welfare Analysis (TAWA) model play an important role in estimating the impacts of potential policy changes. A challenge facing these models is in estimating the take up of programmes like the Accommodation Supplement (AS). This can lead to significant uncertainty in estimated outcomes, such as fiscal costs and reductions in child poverty. In this presentation the Treasury’s TAWA team outlines work undertaken to better model AS by using administrative data on this programme contained in Stats NZ’s Integrated Data Infrastructure (IDI).

(91 words)

***Keywords (3):*** *Microsimulation, StatsNZ IDI, Welfare*

## Population outcome visualisation

### Daniel Barnett, Andrew Sporle

*The University of Auckland*

[daniel.barnett@auckland.ac.nz](mailto:daniel.barnett@auckland.ac.nz)

(0 words)

## Workload of university students in New Zealand

### Daniel Wrench

*Universities New Zealand*

[daniel.wrench@universitiesnz.ac.nz](mailto:daniel.wrench@universitiesnz.ac.nz)

This descriptive study uses education and taxation data to analyse patterns in employment for tertiary students in Aotearoa New Zealand. Using the Integrated Data Infrastructure to link tax data and tertiary enrolment data, and the Household Labour Force Survey to estimate hourly rates, we observe a strong seasonal pattern in employment rate and hours worked for full-time students, and to a lesser extent, part-time students. We also find key differences between international and domestic students. This work can give universities better insight into a key part of students’ lives, which has until now only been understood from surveys and anecdotes. To find out more, email Daniel Wrench at [daniel.wrench@universitiesnz.ac.nz](mailto:daniel.wrench@universitiesnz.ac.nz)

(109 words)

***Keywords (4):*** *students, employment, IDI, longitudinal*

## Smart Search for Electronic Health Records: An NLP-powered approach

### Edmond Zhang

*Orion Health*

[edmond.zhang@orionhealth.com](mailto:edmond.zhang@orionhealth.com)

This work looks at how advanced information search software can improve how doctors and nurses find what they need to know from patient electronic records. It is hoped that clinicians can find patient specific information faster, and more comprehensively and accurately. While electronic patient records can keep information in one physical area, this does not necessarily make it easier to find specific information. This is especially so for patients with complicated illnesses, or illnesses that affect many different organs at once. With Smart search clinicians should be more confident that they have found all the information needed for patient care.

(100 words)

***Keywords (4):*** *Search, EHR (Electronic health records), machine learning, NLP*

## Linked data is vital: evaluating the impact of prehospital care on mortality following major trauma

### Gabrielle Davie

*University of Otago*

[gabrielle.davie@otago.ac.nz](mailto:gabrielle.davie@otago.ac.nz)

The potentially fatal or severe consequences of many injuries can be reduced through an optimally structured prehospital trauma system that can provide timely and appropriate care. This retrospective cohort study linked data from St John and Wellington Free Ambulances, the NZ Trauma Registry, Ministry of Health (Mortality, Hospital discharge and NHI) and Coronial post-mortem reports. The project’s aim is to help identify opportunities to optimise the delivery of Emergency Medical Services (EMS) care in NZ. The collaborative multidisciplinary team includes researchers from St John, Universities of Auckland and Otago. More details about this project are in our protocol paper <https://pubmed.ncbi.nlm.nih.gov/33514568/>.

(100 words)

***Keywords (5):*** *trauma, prehospital, ambulance, mortality, data linkage*

## HEEP2 - Energy Insights from NZ Homes

### Greg Overton

*BRANZ*

[Greg.Overton@branz.co.nz](mailto:Greg.Overton@branz.co.nz)

The Household Energy End-Use Project #2 (HEEP2) is collecting data on how, and why, energy is used in households across New Zealand. From July 2021, HEEP2 will be recruiting 280 households from the Stats NZ Household Economic Survey. These households will each be monitored over 12 months, using a range of instrumentation and surveys, with the final up-to-date picture of energy use being available in the Stats NZ Data Lab. The HEEP2 data will allow researchers to answer a whole raft of questions, with the aim of understanding how to enable and motivate people to affordably create healthy home environments, in ways that contribute to a low-emissions economy.

For more information visit www.branz.co.nz/heep2

(108 words)

***Keywords (5):*** *Housing, Energy, Temperatures, Zero Carbon, Energy Hardship*

## Family Violence in the News: An analysis of media reporting of extreme family violence in New Zealand

### Harini Dissanayake

*Victoria University of Wellington & Data Scientist at DOT Loves Data*

[harini@dotlovesdata.com](mailto:harini@dotlovesdata.com)

This study investigated whether coverage of extreme family violence in New Zealand media is biased across a range of key factors gender, ethnicity, and age of victims, as well as the victim’s relationship to the killer. Our results were derived from a cohort of 946 articles published online by New Zealand media outlets. Analyzing the number of media articles relating to victims from each group (exposure) and their online presence (prominence), we found that although media coverage is generally quite equitable, certain groups of victims are severely under-represented - particularly true for victims from Pacifica communities and elderly victims.

(97 words)

***Keywords (8):*** *Media coverage, sentiment data, family violence, bias, underreporting, child abuse, domestic violence, racial bias*

## Using cell phone data to monitor population mobility and tourism recovery

### Hattie Plant

*Data Ventures*

[harriet@dataventures.nz](mailto:harriet@dataventures.nz)

Analytics on population movement has become increasingly important during the COVID-19 pandemic. In response, Data Ventures partnered with Vodafone and Spark to produce aggregated device counts at an hourly level for small geographies. We then developed a novel weighting methodology to create population estimates, broken down based on whether they are local to an area, domestic visitors, or international visitors. Our daily data provided insight into high mobility areas during lockdown. We now continue to use the data to inform the post COVID-19 recovery of international and domestic tourism.

For more information, see dataventures.nz and reports.dataventures.nz.

(95 words)

***Keywords (4):*** *Population mobility, Telecommunications, Data Ventures, COVID-19*

## New Zealand Activity Index

### Jamas Enright

*Statistics New Zealand*

[jamas.enright@stats.govt.nz](mailto:jamas.enright@stats.govt.nz)

Economic conditions evolve very quickly during crisis periods, and we (the New Zealand government) needed a timely and high-frequency measure of economic conditions to inform policy-making. Gross Domestic Product (GDP) is released quarterly and 10 weeks after the reference quarter. The New Zealand Activity Index (NZAC) was developed by combining eight source indicators and uses principal component analysis to capture as much co-movement of these indicators as possible. This measure is monthly and comes out 2-3 weeks after the end of the month. While not as comprehensive as GDP the NZAC provides a good leading indicator for monthly economic activity.

(98 words)

***Keywords (2):*** *Economic Activity, Principal Component Analysis*

## Exploratory Drivers of Transition as Applied on Workforce Modelling

### Joel E. Bancolita

*Social Wellbeing Agency*

[Joel.Bancolita@swa.govt.nz](mailto:Joel.Bancolita@swa.govt.nz)

Describing factors affecting individuals’ transitions addresses several policy questions. For instance, differentiating retention and exit rates of groups of the workforce based on geography, demography or intervention may inform decision-makers in formulating more targeted policies that can influence funding and the lives of people. We tried developing a framework analysing the drivers of transition in the IDI and employed transition modelling to answer these questions aimed at replicating this to several aspects of people’s lives, such as labour force, education and health. We also employed classification models to address some of limitations, like imputing occupation.

(95 words)

***Keywords (3):*** *Markov Models, Machine Learning, Journey Analysis*

## An analysis of household changes using Housing Register data

### Joel Gibbs

*Massey University*

[joel.a.gibbs@gmail.com](mailto:joel.a.gibbs@gmail.com)

There are a number of data sources and research methods that can be used to understand how households change over time. This project used Social Housing Register data collated via the Integrated Data Infrastructure to analyse the housing pathways of individuals and households who require public housing, and sought out to understand the likelihood that an individual or household that leaves the Housing Register will re-enter the Housing Register in the future. For individuals that re-enter the Housing Register, this research also looked at how households tend to change between applications.

(91 words)

***Keywords (4):*** *Housing Register, public housing, households, IDI*

## Bayesian Demography

### John Bryant

*Bayesian Demography Limited*

[john.r.bryant@gmail.com](mailto:john.r.bryant@gmail.com)

If your data consists of tables rather than individual records, and if one of your variables is age, then you are doing applied demography. Bayesian statistical methods help you do applied demography better. Bayesian methods are particularly helpful if your data is disaggregated, or if you are forecasting. The technical barriers to doing Bayesian demography can, however, be daunting. The talk describes a long-term project to develop methods and tools that reduce these barriers.

(74 words)

***Keywords (5):*** *demography, small area estimation, Bayesian, forecasting, R*

## Project Monty: an agent based model of the New Zealand transport system

### Julie Mugford

*Ministry of Transport*

[J.Mugford@transport.govt.nz](mailto:J.Mugford@transport.govt.nz)

The Ministry of Transport is developing an agent based model of the NZ transport system. The model builds a representation of the entire transport system as the sum of millions of different interactions as everyone seeks to achieve their own sets of goals and ambitions. We aim to offer better insights into complex systems by aggregating the interactions of individual agents (people) as they attempt to complete activities. Through their interactions with a transport network and each other, we get a more realistic representation of how individual travel choices impact the system.

(92 words)

***Keywords (5):*** *Agent based model, transport modelling, population synthesis, network generation, complex systems*

## NZ motor vehicle VKT (vehicle kilometres travelled) and fleet size estimation

### Kain Glensor

*NZ Ministry of Transport*

[K.Glensor@transport.govt.nz](mailto:K.Glensor@transport.govt.nz)

Information on the number of motor vehicles in use in NZ and how much they are driven is of great interest to the public and stakeholders. The process for using Waka Kotahi NZTA’s MVR (motor vehicle register) data to estimate these. How, in R, the data is cleaned, processed and projected forward to allow for a real-time estimate of the current figures for both. More information available from Kain Glensor ([k.glensor@transport.govt.nz](mailto:k.glensor@transport.govt.nz))

(71 words)

***Keywords (5):*** *transport, VKT, fleet, motor vehicles, R*

## Precision Driven health: Partnership model for health data science

### Kevin Ross

*Precision Driven Health*

[kevin.ross@orionhealth.com](mailto:kevin.ross@orionhealth.com)

Precision Driven Health is New Zealand’s health data science partnership. For the past five years, we have been forming collaborations between New Zealand’s health IT sector, health providers and universities, aimed at improving health outcomes through data science. Our work seeks to personalise health by integrating new data sources, developing predictive models, optimise decision making and empower people with new tools. We will share some lessons learned from over 100 projects, including the challenges of working with personal information and ensuring equity is improved by design.

(85 words)

***Keywords (4):*** *Health, collaboration, partnership, governance*

## Using risk-adjusted Days Alive and Out of Hospital (DAOH) to compare health outcomes across NZ after surgery

### Luke Boyle

*Department of Statistics, University of Auckland*

[lboy505@aucklanduni.ac.nz](mailto:lboy505@aucklanduni.ac.nz)

DAOH scores can be an effective way to measure health outcomes by collapsing many negative outcomes after surgery, such as death or readmission, into one number. This study used routine data from the Ministry of Health and applied novel risk adjustment methods to illustrate how DAOH scores can detect differences between patients, different types of operations and DHBs in NZ.

Using this data, we can identify important areas of difference, for example between hospitals, that can be further audited to improve outcomes after surgery or to investigate optimal patient pathways for recovery.

(91 words)

***Keywords (5):*** *Government data, Health, risk-adjustment, Surgery, comparisons*

## Ngā Tapuae: Stepping stones for Māori Student Transitions

### Marianna Pekar, Joel E. Bancolita

*Social Wellbeing agency*

[Marianna.Pekar@swa.govt.nz](mailto:Marianna.Pekar@swa.govt.nz), [Joel.Bancolita@swa.govt.nz](mailto:Joel.Bancolita@swa.govt.nz)

The Social Wellbeing Agency worked in partnership with Tokona Te Raki, a Ngāi Tahu-led collaborative established to increase Māori participation, success and progression in education and employment outcomes. Our research focuses on data discovery to learn actionable insights that identify the most important barriers, levers, and boosters that help young Māori to succeed. The research builds on the existing evidence base investigating the need for better transition pathways for Māori aged 15-29 years. Mixed methods were applied to answer the research question. We conducted the quantitative analysis in New Zealand’s Integrated Data Infrastructure (IDI).

(93 words)

***Keywords (5):*** *Administrative data, Sequence analysis, Mixed methods, Education, Māori*

## How people learn Te Reo Māori: Evidence from the IDI

### Matt Jones

*Tertiary Education Commission*

[Matt.Jones@tec.govt.nz](mailto:Matt.Jones@tec.govt.nz)

This presentation uses the IDI to combine Census and education data to estimate the number of young people who learn te reo Māori through education versus other methods. It also estimates the amount of formal education required to become proficient in the language. Finally it suggests how a model might work for understanding the growth of te reo speakers over time.

(61 words)

***Keywords (3):*** *Te Reo Māori, IDI, Education*

## TAWA - Treasury’s microsimulation model of the New Zealand personal tax and transfer system

### Michael Eglinton

*Te Tai Ōhanga The Treasury*

[Michael.Eglinton@treasury.govt.nz](mailto:Michael.Eglinton@treasury.govt.nz)

Tax and Welfare Analysis (TAWA) is the Treasury’s microsimulation model of the New Zealand personal tax and transfer system. The TAWA model uses a combination of survey and administrative data within Stats NZ Integrated Data Infrastructure to model potential policy changes. It is used extensively within Treasury and in external work related to policy analysis of tax and welfare settings.

(60 words)

***Keywords (4):*** *Treasury, microsimulation, tax, policy changes*

## Building for wellbeing

### Michael Nuth

*BRANZ*

[michael.nuth@branz.co.nz](mailto:michael.nuth@branz.co.nz)

The aim of this project is to develop a digital post occupancy evaluation to efficiently capture the self-reported qualitative perspectives of building occupants about the wellbeing performance of residential buildings. By doing so, my intention is to develop an app-based wellbeing assessment and data collection tool that is informed by the Government’s Living Standards Framework. Data collected via the app will ultimately help inform the planning, design and construction of residential buildings that meet the wellbeing needs of New Zealanders.

The app will be developed and trialled in collaboration with Auckland University of Technology, Tether Ltd and Kāinga Ora.

Contact Michael Nuth @michael.nuth@branz.co.nz for more information.

(104 words)

***Keywords (3):*** *wellbeing, eveluation, buildling and construction*

## Occupational Exposures and Ischaemic Heart Disease (IHD): results from the entire New Zealand population using the Integrated Data Infrastructure (IDI)

### Amanda Eng

*Research Centre for Hauora and Heath, Massey University*

[A.J.Eng@massey.ac.nz](mailto:A.J.Eng@massey.ac.nz)

Common occupational exposures have been associated with IHD, but evidence is conflicting. For the whole NZ adult working population at the time of the 2013 census, data were extracted from the Statistics NZ IDI, on occupation and incident IHD from 2013 to 2018. The number of working hours was extracted from the census, and exposure to sedentary work, loud noise, and night shift was assessed through NZ job exposure matrices. This study suggests occupational exposure to high levels of noise and night shifts increase IHD risk while there was no evidence of association with sedentary work and long working hours.

(97 words)

***Keywords (6):*** *Cardiovascular disease, Ischaemic heart disease, Occupational exposure, Occupational health, Survival analysis, integrated data infrastructure*

**NOTE:** Amanda and Marine giving related talks

## Association between Occupation and Ischaemic Heart Disease (IHD) by sex and ethnicity in the whole New Zealand population using the Integrated Data Infrastructure (IDI)

### Marine Corbin

*Research Centre for Hauora and Heath, Massey University*

[M.Corbin@massey.ac.nz](mailto:M.Corbin@massey.ac.nz)

Associations between IHD and occupations are poorly understood. This whole-population study aimed to identify occupations associated with increased IHD risk in NZ by sex and ethnicity. A cohort of workers was constructed for the whole NZ adult working population using the Statistics NZ IDI. Occupation was obtained from census data and incident IHD was determined using hospitalisation, prescription, and death records from 2013 to 2018. This study confirmed an increased IHD risk for several occupations previously identified as being high risk for IHD and has also identified some potentially new occupational groups. Important sex and ethnic differences were also observed.

(98 words)

***Keywords (6):*** *Cardiovascular disease, Ischaemic heart disease, Occupation, Occupational health, Survival analysis, integrated data infrastructure*

**NOTE:** Amanda and Marine giving related talks

## AIS Explorer — Integrating technology to prevent aquatic invasive species in Minnesota

### Nick Snellgrove

*Epi-interactive*

[nick@epi-interactive.com](mailto:nick@epi-interactive.com)

Minnesota in the United States is well known for its beautiful “10,000 lakes” - however the complex system of waterbodies is threatened by the spread of aquatic invasive species. We developed the AIS Explorer in collaboration with the Aquatic Invasive Species Research Center of the University of Minnesota to bridge research and decision making, for example to optimise watercraft inspections.

AIS Explorer: <https://www.aisexplorer.umn.edu/>

In this talk we will showcase how a diverse set of technologies, like R, R Shiny, Python and cloud computing was integrated to a create a tool that is easy to use and matches stakeholder needs.

(95 words)

***Keywords (5):*** *R Shiny, Python, cloud computing, analytics, dashboard*

## Increasing fine payments: A behavioural science approach

### Olivia Wills

*Ministry of Justice*

[Olivia.Wills@justice.govt.nz](mailto:Olivia.Wills@justice.govt.nz)

Over half a million New Zealanders owe fines to the Ministry of Justice, and total collections debt is around US$395m, or US$78 for every person in New Zealand. There are consequences for people who do not pay, as they can be hit with enforcement actions or may be summoned to Court. We used evidence from behavioural science to test new approaches to improve fines payment behaviour, using letters and text message reminders. Our results demonstrate that small tweaks to existing processes can be effective at changing behaviour, and pave the way for the future application of behavioural science in the fines collection space.

(103 words)

***Keywords (2):*** *behavioural science, fines*

## Urban Trees and Data Science: Examining Wellbeing

### Peter Edwards

*Manaaki Whenua Landcare Research*

[EdwardsP@landcareresearch.co.nz](mailto:EdwardsP@landcareresearch.co.nz)

Trees and forests have major impacts on planetary and human wellbeing.  In an increasingly urbanised world, urban trees and forests become important for human wellbeing. Many studies show the benefits of urban trees and forests – health, climate change, urban planning and ecological perspectives. Using a wellbeing framework, quantitative data from remote sensing, modelling and social and cultural administrative data, we aim to understand the impact of urban trees across a range of wellbeing domains. Using epidemiological methods, we aim to discover correlations and patterns between urban trees and human wellbeing in Singapore and Wellington. More information? Dr Peter Edwards — [EdwardsP@landcareresearch.co.nz](mailto:EdwardsP@landcareresearch.co.nz).

(92 words)

***Keywords (5):*** *Urban trees, data science, wellbeing, Wellington, Singapore*

## Augmented Decisions - even when you don’t have all the data!

### Peter Hong

*Houston We Have*

[peter.hong@houstonwehave.ai](mailto:peter.hong@houstonwehave.ai)

*Houston We Have* uses proprietary software, data, technology and mathematics to help humans make better decisions, especially when most at risk. Sectors where we already develop predictions, forecasts and risk models for enhanced strategic and operational decision making include defence, financial services and fintech, natural resources, and health. Our software, technology and data insight capability can be deployed in any area where better quality, visibility and traceability around decision making is important.

(72 words)

***Keywords (3):*** *Augmented Intelligence, analytics, decisions sorted*

## Epidemix- visualising analytical complexity to improve decision-making for disease control

### Petra Muellner

*Epi-interactive Ltd. & Massey University*

[petra@epi-interactive.com](mailto:petra@epi-interactive.com)

Epidemix[[1]](#footnote-134) allows users to develop an understanding of the impact of disease modelling assumptions on the trajectory of an epidemic and the impact of control interventions, without having to directly deal with the complexity of equations and programming languages. The app provides a visual interface for nine generic models, plus two disease-specific models of international relevance (COVID-19 and ASF). Epidemix supports the teaching of mathematical modelling to non-specialists – including policy makers by demonstrating key concepts of disease dynamics and control in a hands-on way. Funded by the Royal Veterinary College and the City University of Hong Kong (www.epidemix.app).

(121 words)

***Keywords (5):*** *SIR models, risk management, R, R Shiny, infectious disease*

## Sharing our experiences from COVID-19 and Te Pokapū Hātepe o Aotearoa, the New Zealand Algorithm Hub

### Pieta Brown

*Orion Health*

[pieta.brown@orionhealth.com](mailto:pieta.brown@orionhealth.com)

In support of the response to COVID-19, New Zealand initiated and rapidly delivered a solution for national algorithm management. During this project, we stood up a national instance of a machine learning platform, developed a governance process tailored to our local context in Aotearoa, deployed 30 models, and went live with a website (www.algorithmhub.co.nz) to support user engagement and interaction. The underlying technology solution has evolved over the last five years through our research partnership focused on the development and delivery of data science in healthcare; it was built in response to the need for tooling to deploy, manage and monitor algorithms safely and effectively in a healthcare context. This talk will share the key lessons learned from our COVID-19 experience and discuss the technology, user engagement and governance processes that made this successful.

(133 words)

***Keywords (5):*** *machine learning, data science, software, governance, accessibility*

## He Ara Poutama mō te Reo Māori — Forecasting te Reo Māori Acquisition

### Pip Bennett

*Nicholson Consulting*

[pip.b@nicholsonconsulting.co.nz](mailto:pip.b@nicholsonconsulting.co.nz)

[Presented in te reo Māori]

He Ara Poutama mō te Reo Māori is a project that focuses on forecasting levels of te Reo Māori acquisition into the future including consideration of the impacts of changing language revitalisation initiatives. We are working with subject matter experts and organisations to help ensure this taonga is accessible for present and future generations. Nicholson Consulting is a data science consultancy in Te Whanagnui-a-Tara; we are partnering with Kōtātā Insights to use quantitative methods to assist our clients on their haerenga to see Māori as a living language and as an ordinary means of communication. For more information contact [hello@nicholsonconsulting.co.nz](mailto:hello@nicholsonconsulting.co.nz)

(104 words)

***Keywords (5):*** *te Reo Māori, language acquisition, microsimulation, synthetic data, integrated data*

**NOTE:** Presenting twice, in each of English and te Reo

## He Ara Poutama mō te Reo Māori — Forecasting te Reo Māori Acquisition

### Pip Bennett

*Nicholson Consulting*

[pip.b@nicholsonconsulting.co.nz](mailto:pip.b@nicholsonconsulting.co.nz)

He Ara Poutama mō te Reo Māori is a project that focuses on forecasting levels of te Reo Māori acquisition into the future including consideration of the impacts of changing language revitalisation initiatives. We are working with subject matter experts and organisations to help ensure this taonga is accessible for present and future generations. Nicholson Consulting is a data science consultancy in Te Whanagnui-a-Tara; we are partnering with Kōtātā Insights to use quantitative methods to assist our clients on their haerenga to see Māori as a living language and as an ordinary means of communication. For more information contact [hello@nicholsonconsulting.co.nz](mailto:hello@nicholsonconsulting.co.nz)

(100 words)

***Keywords (5):*** *te Reo Māori, language acquisition, microsimulation, synthetic data, integrated data*

**NOTE:** Presenting twice, in each of English and te Reo

## Understanding gaps in income for fathers with new babies

### Raj Kulkarni, Tze Ming Mok

*Social Wellbeing Agency*

[Raj.Kulkarni@swa.govt.nz](mailto:Raj.Kulkarni@swa.govt.nz)

This piece of work followed the Having a Baby in South Auckland project, which applied the Social Wellbeing Agency’s representative timeline modelling method to produce insights into South Auckland families’ experiences around the birth of a child. As a result of one of these insights, we investigated fluctuations in new fathers’ income around time of the birth and found that a substantial proportion of fathers who usually earn around the minimum wage have income dips that suggest taking unpaid time off work. Yet, only half of them were eligible for two weeks of parental leave. This presentation discusses certain aspects of fathers’ life around the birth.

(106 words)

***Keywords (5):*** *Parental Leave, Fathers, Employer Attachment, Income Fluctuation, Collaboration*

## RAPping in the public sector: binding your legacy code into a pipeline with Python

### Shrividya Ravi

*Ministry of Transport*

[S.Ravi@transport.govt.nz](mailto:S.Ravi@transport.govt.nz)

Straggling legacy SAS scripts requiring manual steps are a common “feature” of data processing tasks in the public sector. However, refactoring such scripts to modern, reproducible analytical pipelines (RAP) can be challenging due to a lack of IT infrastructure or high complexity. In such situations, interim solutions can at least reduce manual effort and mental overhead. One such solution is using Python as an effective glue to create one click execution pipelines. Manual tasks like downloading data from email, updating new data file names in scripts, running scripts in sequence and more, can be managed with Python and its rich ecosystem of packages. In this talk, I will showcase how three Python packages, exchangelib, jupyter and saspy, can create quick and easy automated versions of legacy SAS scripts that contain many types of manual steps.

(135 words)

***Keywords (5):*** *Python, data engineering, legacy code, analytics, Reproducible analytical pipelines (RAP)*

## Accelerating Dataset Assembly

### Simon Anastasiadis

*Social Wellbeing Agency*

[Simon.Anastasiadis@swa.govt.nz](mailto:Simon.Anastasiadis@swa.govt.nz)

Data preparation is a key stage for analytic and research projects. However, as the number of data sources increases so does the complexity of preparation. Without a consistent method for assembling analysis-ready datasets, this process can become time-consuming, expensive, and error prone.

In response to this, we have developed the Dataset Assembly Tool. By standardising and automating the data preparation and dataset assembly stages of analytic projects, the tool helps staff deliver higher quality work faster. The assembly tool is now available for other researchers to use. This presentation with describe the tool and its advantages.

(95 words)

***Keywords (4):*** *data preparation, data wrangling, tools, process*

## Representative timeline modelling

### Simon Anastasiadis

*Social Wellbeing Agency*

[Simon.Anastasiadis@swa.govt.nz](mailto:Simon.Anastasiadis@swa.govt.nz)

Timeline are sequences of events and periods which reflect significant parts of a person’s experience. They can be an effective tool for researchers seeking to understand life events, and interactions between events, over time.

To share timeline information while respecting privacy and confidentiality, we have developed a representative timeline methodology. This method produces a timeline that captures experiences that are common across a group of people — similar to a group average.

This presentation will demonstrate the technique, drawing examples from our first application of it – a study of South Auckland families’ experiences around the birth of a child.

(96 words)

***Keywords (3):*** *methodologies, tools, timeline*

## Pathway modelling to optimise long-term policy impact in New Zealand

### Suzanne Woodward

*University of Auckland, Massey*

[s.woodward@auckland.ac.nz](mailto:s.woodward@auckland.ac.nz)

Evaluating policy initiatives is inherently difficult. We can address this by considering how policy decisions inform social pathways across multiple sectors, and how pathway trajectories affect individual wellbeing. Trajectory data and narratives of lived experience that complement large, linked datasets, provide the opportunity to construct models that can accurately predict wellbeing outcomes at scale. Policymakers can use these models to assess the efficacy of policy initiatives, quantifying their contribution to wellbeing. This approach combines ideas and methods from the policy and social sciences, co-design, data science and mathematical modelling to address the wicked problem of policy design. Visit <https://orua.blogs.auckland.ac.nz/project/policy-planning/>.

(99 words)

***Keywords (4):*** *Wicked problems, modelling, policy, wellbeing*

## Who is a visitor and who you think is a visitor, are they the same thing?

### Taylor Winter

*Data Ventures*

[Taylor.Winter@stats.govt.nz](mailto:Taylor.Winter@stats.govt.nz)

Population definitions are well established into our mahi at Stats NZ. However, when Data Ventures reached out to our tourism customers, we identified a key disconnect between expected definitions of visitors and actual definitions. Specifically, our customers thought of visitors as tourists and needed these numbers to inform their COVID-19 response. We developed a new measure of visitors, named ‘short-term visitors’, that predominantly represents tourists and was validated against visa and migration data. We present a comparison between visitor definitions and discuss how these definitions may differ in value based on the context and customer expectation.

(97 words)

***Keywords (5):*** *Population, Telecommunications, Data Ventures, Tourism, Visitors*

## Growth monitoring Aotearoa: Scoping a national system for tamariki and rangatahi

### Teresa Gontijo de Castro

*University of Auckland*

[t.castro@auckland.ac.nz](mailto:t.castro@auckland.ac.nz)

In New Zealand (NZ) three in ten 2-14 year- olds are affected by overweight/obesity with ethnic disparities. The existing data used to monitor growth of 0-19-year-olds are fragmented and mostly aggregated at the national level. Thus, NZ lacks group-specific information on prevalence, trends, and determinants of healthy growth for this group, crucial information if equitable improvements are to be achieved. We will map and assess diverse national sources of anthropometric data of 0-19-year-olds (last 2 decades) to establish what would be required and which sources would be suitable to be included in a national monitoring platform. For more information: [t.castro@auckland.ac.nz](mailto:t.castro@auckland.ac.nz)

(98 words)

***Keywords (5):*** *Healthy Weight, Growth Monitoring, Anthropometric Data, Data Collation, Fine-grained Data*

**NOTE:** Attending online

## Analysing Surveys with iNZight

### Tom Elliott

*Te Rourou Tātaritanga, Victoria University and The University of Auckland*

[tom.elliott@vuw.ac.nz](mailto:tom.elliott@vuw.ac.nz)

Survey data has a huge importance for many research groups, but many software tools require users to understand the survey design and know how to specify it to the relevent program. iNZight is a graphical user interface for R that lets researchers quickly and efficiently visualise and explore their data. Recent changes to iNZight allow users to forget the design and focus on what matters: exploring their data.

For more information, visit <https://inzight.nz>.

(72 words)

***Keywords (3):*** *surveys, tools, GUI*

## Te Matatini o te Horapa - a contagion network model for Aotearoa NZ

### Dion O’Neale

*University of Auckland and Te Pūnaha Matatini*

[d.oneale@auckland.ac.nz](mailto:d.oneale@auckland.ac.nz)

Contagion models for infectious disease can only be as good as the assumptions that they are built on. One common simplifying assumption is that the disease spreads through a well-mixed, homogeneous population (or populations). This has some obvious consequences for modeling the differential impact of disease on different groups. I will introduce an individual level, network based, contagion model, built from a range of data sources, including census data, and will show how it has been used for modelling the spread of CVOID-19 in Aotearoa NZ.

(86 words)

***Keywords (5):*** *Contagion, COVID-19, interaction network, individual based modelling, Equity*

**NOTE:** TPH Talk 1

## Constructing an individual level interaction network for modelling COVID-19 in Aotearoa NZ

### Emily Harvey

*ME Research and Te Pūnaha Matatini*

[emily@me.co.nz](mailto:emily@me.co.nz)

Illnesses and deaths from the COVID-19 pandemic have not been evenly distributed across communities. Risk factors for serious disease (including age, ethnicity, & health conditions), and risk factors for infection (including dwelling size, inter-generational living, & type of work) are not independent. In order to capture the interactions between these factors we build on data from linked StatisticsNZ Integrated Data Infrastructure (IDI) to construct a network of the ~5million individuals and their interaction contexts, in particular homes, workplaces, and schools. I will highlight our progress so far, and priorities for future refinements.

(90 words)

***Keywords (5):*** *Contagion, COVID-19, micro-simulation, individual based modelling, Equity*

**NOTE:** TPH Talk 2

## Integrating Contact Patterns into Simple Models of Disease Spread

### Nic Steyn

*University of Auckland and Te Pūnaha Matatini*

[nicholas.steyn@auckland.ac.nz](mailto:nicholas.steyn@auckland.ac.nz)

Simple models of disease spread often assume homogeneous mixing. While these are easy to construct and quick to solve, they ignore important detail. Contact matrices are an easy way of including heterogeneities in these simple models without introducing much complexity but require robust data. I will highlight why these contact matrices are important for our COVID-19 response, outline the best data we have, discuss what we are missing, and give examples of what other countries are doing in this space.

(80 words)

***Keywords (5):*** *COVID-19, disease modelling, contact matrix, contact survey, equity*

**NOTE:** TPH Talk 3

## Participatory surveillance of influenza and COVID-19 symptoms in New Zealand

### Frank Mackenzie

*University of Auckland and Te Pūnaha Matatini*

[frank.mackenzie@auckland.ac.nz](mailto:frank.mackenzie@auckland.ac.nz)

FluTracking is an online participatory surveillance system that has been used for collecting data on influenza seasons in Australia and New Zealand, wherein participants fill out a weekly survey on whether they have experienced symptoms. This has several advantages over sentinel-based data collection, allowing for near real-time reporting of symptoms and mitigating some bias. This study analyses New Zealand FluTracking data from April 2020 to April 2021, examining the impact of various factors on the incidence of flu- and COVID-like symptoms. I will discuss the various ways bias is introduced into these data, the methods employed to mitigate bias, and the potential of systems like FluTracking for public health surveillance.

(108 words)

***Keywords (4):*** *influenza, disease incidence, real-time reporting, participation bias*

**NOTE:** TPH Talk 4

## Uncertainty Quantification for Complex Network Contagion Simulation Models

### Frankie Patten-Elliott

*University of Auckland and Te Pūnaha Matatini*

[fpat084@aucklanduni.ac.nz](mailto:fpat084@aucklanduni.ac.nz)

As the complexity of a model increases, so does the uncertainty inherent in the model’s output. I consider Uncertainty Quantification (UQ) techniques to assess the reliability of a complex network contagion model we have developed for informing Aotearoa New Zealand’s COVID-19 response. By fitting Gaussian process surrogate models to output from the network contagion model, we can quickly and efficiently apply UQ methods and perform inference on model parameters. Using New Zealand COVID-19 case number data, we can also condition these surrogates to make predictions without having to rerun the computationally expensive network contagion model.

(95 words)

***Keywords (5):*** *Uncertainty Quantification, Surrogate Models, Network Contagion Model, Parameter Estimation, COVID-19*

**NOTE:** TPH Talk 5

## Connecting research data with stakeholders – the Gambling Data Explorer

### Uli Muellner

*Epi-interactive Ltd.*

[uli@epi-interactive.com](mailto:uli@epi-interactive.com)

The National Gambling Study (NGS) is the first New Zealand population representative longitudinal study into gambling behaviours and attitudes, health, and lifestyles. To support the dissemination and use of the findings from the study the National Gambling Study Explorer provides easy and user-friendly access for stakeholders to information on gambling participation and epidemiological risk factors. The Gambling Data Explorer is freely available at <https://www.health.govt.nz/publication/national-gambling-study-data-explorer>. This project was delivered for the Ministry of Health in collaboration with AUT. It demonstrates an innovative way to share research findings and data with stakeholders through online dashboards.

(93 words)

***Keywords (5):*** *gambling, research, visualisation, outreach, stakeholder*

## Providing customised health intelligence in real-time – COVID-19 dashboards

### Uli Muellner

*Epi-interactive Ltd.*

[uli@epi-interactive.com](mailto:uli@epi-interactive.com)

Real-time health intelligence is essential to an effective public health response. The NZ COVID-19 dashboards were built with this need in mind drawing on ESR’s EpiSurv database for notifiable diseases. Once published, the dashboards and underlying data streams had to evolve as the response progressed e.g. to differentiate between cases occurring in the community and MIQ or to provide insight into identified outbreaks. Using a R Shiny framework allowed us to dynamically adjust the dashboards to response requirements, which was critical. The COVID-19 dashboards were built for the Ministry of Health in collaboration with ESR. Public dashboard available at: <https://nzcoviddashboard.esr.cri.nz/>.

(100 words)

***Keywords (6):*** *COVID-19, response, health intelligence, operational epidemiology, real-time, R Shiny*

## Seeking feedback on the Administrative Population Census

### Vinayak Anand Kumar

*Stats NZ*

[vinayak.anand-kumar@stats.govt.nz](mailto:vinayak.anand-kumar@stats.govt.nz)

Stats NZ’s Census Transformation programme will release the first iteration of the experimental Administrative Population Census (APC) in August 2021. The APC uses administrative data to construct an annual census file, and will be an opportunity for groups to provide feedback on an early iteration of admin first Census outputs. At this presentation, we will outline: \* the features of the 2021 APC to demonstrate its potential value to the research community; \* how future iterations of the APC will build on the 2021 release .

Please contact Christine Bycroft ([Christine.bycroft@stats.govt.nz](mailto:Christine.bycroft@stats.govt.nz)) or Vinayak Anand-Kumar ([Vinayak.anand-kumar@stats.govt.nz](mailto:Vinayak.anand-kumar@stats.govt.nz)) to get more information about the APC.

(96 words)

***Keywords (4):*** *Census, Administrative Data, Experimental, Stats NZ*

## Data Wrangling At Warp Speed with Segna

### Will Haringa

*Segna*

[will@segna.io](mailto:will@segna.io)

As a Data scientist you want your data in an analysis-ready form faster, delivered through pipelines that don’t break as the shape of your data changes. Harnessing the power of machine learning, Segna’s smart wrangling tool aggregates and cleans data from multiple sources up to 1,600 times faster!

More information: <https://segna.io>

(49 words)

***Keywords (2):*** *data wrangling, automated data science*

## Calibrating Economic Agent-Based Models with Microdata

### Will Scarrold

*Te Pūnaha Matatini (University of Auckland) and the Institute for New Economic Thinking (University of Oxford)*

[will.scarrold@gmail.com](mailto:will.scarrold@gmail.com)

Despite the complexity of modern socio-economic systems, current benchmark models assume the economy is much simpler than it really is: households are often assumed to be identical, and firms are assumed to use the same equipment to produce the same “representative” product.

Recent developments in Agent-Based Modelling have provided an alternative to conventional models. Such models show comparable forecasting performance despite being a relatively early stage of development.

This presentation will discuss ABM modelling and the use of microdata from the New Zealand IDI to calibrate these models.

More information on recent macroeconomic agent-based models can be found at tinyurl.com/econABM

(97 words)

***Keywords (4):*** *Agent-Based Modelling, Integrated Data Infrastructure, Economics, Information Geometry*

1. Muellner U, Fournie G, Muellner P, Ahlstrom C, Pfeiffer D. epidemix - an Interactive Multi-Model Application for Teaching and Visualizing Infectious Disease Transmission. Epidemics, doi: 10.1016/j.epidem.2017.12.003, 2017. [↑](#footnote-ref-134)