New Zealand Biostatistics Conference 2023

Abstracts (by session, not including keynotes or invited talks)

Postgraduate Student Presentations I

Rory Miller

Avoiding double counting: the effect of bundling hospital events in administrative datasets for the interpretation of rural-urban differences in Aotearoa New Zealand

Rory Miller (Dept. GP and rural health), Gabrielle Davie (Dept. of Social and Preventative Medicine), Brandon De Graaf (Dept. of Social and Preventative Medicine), Sue Crengle (Ngāi Tahu Māori research unit), Garry Nixon (Dept. GP and rural health)

Background
All publicly funded hospital discharges in Aotearoa New Zealand (NZ) are recorded in the National Minimum Dataset (NMDS). Within the NDMS there are multiple discharges (rows) per health care encounter if there are intra- or inter-hospital transfers. The aim of this study was to determine whether different methods of bundling ambulatory sensitive hospitalisations (ASH) discharge events into health care encounters resulted in different rural-urban analyses.

Methods
Discharge events within the NMDS that had an admission date between June 1 2015 and December 31 2019 were bundled into encounters using either using a date based, admission source flag or event end type flag compared to no bundling. ASH incidence rates and rate ratios (IRR), the mean total length of stay and the percentage of interhospital transfers were estimated for each bundling method. These outcomes were compared across 4 categories of the Geographic Classification for Health (GCH).

Results
Compared with no bundling, using the dates method resulted in an 8% reduction (150 less events per 100 000 person years) in the estimated incidence rate. There was no difference in the interpretation of the rural-urban IRR. Length of stay was longer for all bundling methods used. Using dates identified twice as many inter-hospital transfers (5.7% vs. 12.4%) as using admission flags for patients that live in the most rural regions.

Conclusion
Consecutive events within administrative datasets should be bundled into encounters to reduce the effect of overcounting when estimating incidence rates and undercounting interhospital transfers and total length of stay.

Rosie Jackson

How does mild sleep deprivation affect diet, physical activity and wellbeing in children: a randomised cross-over trial

Rosie Jackson[1], Silke Morrison[1], Jillian Haszard[2], Barbara Galland[3], Kim Meredith-Jones[1], Rachael Taylor[1]

[1]Medicine, [2]Biostatistics, [3]Woman's and Children's Health, University of Otago, Dunedin, New Zealand

Background:
Children who sleep less are shown to be at increased risk of developing obesity, the reasons why are unknown.

Purpose:
The aim of this trial was to investigate the effects of mild sleep deprivation on diet, physical activity, and wellbeing.

Materials and methods:
105 children aged 8-12 years underwent a randomised cross-over trial where they were asked to go to bed 1 hour later each night for 1-week (sleep restriction), and 1 hour earlier each night for 1-week (sleep extension), separated by a 1-week washout. During the intervention weeks, assessment of diet (through dietary recalls), eating behaviour (by questionnaire), physical activity (accelerometers) and wellbeing (via a health-related quality of life questionnaire) were assessed. Differences (95% CI) in diet, physical activity and wellbeing were compared between sleep restriction and extension weeks.

Results:
In this cross-over trial, children achieved a sleep difference of 40 (33, 47) minutes per night between the two weeks. Intention to treat analyses (n=99) showed that in the week of sleep restriction, the participants consumed more total energy 233kJ (-42, 509) and caregivers reported increased emotional overeating (0.12; 0.01, 0.24) and emotional undereating (0.15; 0.03, 0.27). The participants also reported having a lower total health-related quality of life (SMD, −0.21; −0.34 to −0.08) in the week of sleep restriction. The extra awake time gained from sleeping less was re-allocated to physical activity and sedentary behaviour proportionally, rather than preferentially, indicating the physical activity behaviour did not appear to change in the week of sleep restriction compared to sleep extension.

Conclusion:
Our study demonstrated that mild sleep deprivation of 40 minutes per night resulted in increased energy intake and changes in eating behaviour rather than decreased physical activity. These results imply that the sleep – obesity relationship seems to be partially explained through excess energy intake, rather than lower energy expenditure.

Andre Mason

A test of the Three Step Theory for suicide in Aotearoa New Zealand university students

Andre Mason[1,2], Irie Schimanski[3], Gareth T Treharne[2], Charlene Rapsey[1] & Damian Scarf[2]

[1]Department of Psychological Medicine, Dunedin School of Medicine, University of Otago, Dunedin, Aotearoa New Zealand.

[2]Department of Psychology, University of Otago, Dunedin, Aotearoa New Zealand.
[3]Te Kura Tātai Hauora School of Health, Te Herenga Waka Victoria University of Wellington, Wellington, Aotearoa New Zealand.

Presenter's email: andre.mason@otago.ac.nz

Suicide is one of the leading causes of death among youth. Past global research demonstrates that theories of suicide can explain variance in suicidal ideation and suicide attempts, although there is limited evidence among university students and in specific contexts. The aim of this study was to comprehensively test the utility of the Three Step Theory (3ST) of suicide among university students in the context of Aotearoa New Zealand. A sample of 361 university students aged 17–51 (mean 19.83) completed an online survey covering all aspects of the 3ST. Pain, hopelessness, and the combination of the two, uniquely predicted variance in suicidal ideation. Thwarted belonging buffered against pain for the overall sample, but not among individuals experiencing high pain and high hopelessness. Practical suicide capability was observed to differentiate between those who had previously only thought about suicide and those who had previously attempted suicide. Overall, these findings provide support for Step One of the 3ST and partial support for Steps Two and Three. The findings have implications for future research and clinical practice seeking to reduce suicide attempts among university students.

Ellie Medcalf

Using the counterfactual framework to estimate non-ITT estimands in randomised controlled trials: a methodological scoping review

Authors: Ellie Medcalf (1), David Espinoza (2), Robin Turner (3), Katy Bell (1)

1. Sydney School of Public Health, The University of Sydney
2. National Health and Medical Research Council (NHMRC) Clinical Trials Centre, Sydney, NSW, Australia
3. Biostatistics Centre, University of Otago, Dunedin, Otago, New Zealand

Background: Traditionally, when treatment estimates are reported in randomised controlled trials (RCT), they are intention-to-treat (ITT) estimates. However, these estimates do not always provide the information that patients and clinicians want. For example, when non-adherence occurs in a trial, ITT estimates reflect the effect of being offered the intervention, rather than the benefits of adhering to the intervention. Moreover, it can often be of interest to understand what role mediators may play in facilitating the effects of an intervention. Estimating the effects of actually adhering to an intervention and of potential mediators requires alternatives to ITT estimates that can estimated using innovative causal inference and mediation methods.

Objective: Conduct a methodological scoping review to identify and summarise causal inference and mediation methods that can be used in RCTs.

Methods: We searched MEDLINE and EMBASE for articles in which authors discussed causal inference and mediation methods

in RCTs. One reviewer will undertake full-text screening and data extraction and a second reviewer will check extractions.

Results: 745 unique articles were retrieved from the database searches. After full-text screening, 62 studies were included. Preliminary findings show that using causal inference methods to account for nonadherence or to provide mediator effects often compares favourably to more traditional approaches. These methods are applied in both simple scenarios (e.g., where participants either adhere or don't adhere, or where there is a single mediator) and more complex scenarios (e.g., where participants partially adhere or where there are multiple mediators). There has also been increased application of novel approaches such as combining causal inference methods with machine learning techniques such as random forest and stacking. Data extraction of included studies is currently underway, and final results will be presented at the conference.

Conclusion: This scoping review will summarise evidence on causal inference and mediation methods and their use in RCTs. These methods allow estimation of potentially more informative estimates than ITT in RCTs, as they can assist in generating evidence that end-users are most interested in: what is the effect of an intervention if they actually adhere to it.

Postgraduate Student Presentations II

Andre Mason

Effects of vaping on uptake and cessation of smoking: longitudinal analysis in Aotearoa New Zealand adults

Andre Mason[1,2], Benjamin C Riordan[3], Taylor Winter[4], Tamlin Conner[2], Chris Sibley[5], Damian Scarf[2]

[1]Department of Psychological Medicine, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand,
[2]Department of Psychology, PO Box 56, University of Otago, Dunedin, New Zealand,
[3]Centre for Alcohol Policy Research, La Trobe University, Melbourne, Victoria, Australia
[4]School of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand
[5]School of Psychology, University of Auckland, New Zealand.

Presenter's email: andre.mason@otago.ac.nz

The prevalence of vaping, or e-cigarette use, is increasing. Initially posited as a mechanism to support smokers to quit, there is growing concern that vaping may also lead to smoking. This talk presents research exploring the prevalence of vaping and smoking in Aotearoa New Zealand and considers the longitudinal pathways between smoking status and vape use. We analysed data from three waves, between 2018-2020, of the New Zealand Attitudes and Values study, a large, representative, multi-wave study of adults living in New Zealand. Generally, the data indicated that the prevalence of vaping was increasing while the prevalence of smoking was decreasing. Despite these trends, however, analyses suggest that individuals are just as likely to transition from smoking to vaping as they are to shift from vaping to smoking; no differences were in the odds of transitioning between pathways. Overall, these findings suggest that vaping may be just as likely to

catalyse transition to smoking as it is to have a cessation effect. Our findings highlight the need for greater consideration regarding vaping-related policy and restriction.

Ellie Medcalf

Handling missing data in randomised controlled trials (RCTs): a methodological scoping review

Ellie Medcalf (1), Vicky He (2), David Espinoza (3), Robin Turner (4), Katy Bell (1)

1. Sydney School of Public Health, The University of Sydney
2. The University of Melbourne, Melbourne, Victoria
3. National Health and Medical Research Council (NHMRC) Clinical Trials Centre, Sydney, NSW, Australia
4. Biostatistics Centre, University of Otago, Dunedin, Otago, New Zealand

Background: Missing data can compromise the validity of inferences when conducting randomised controlled trials (RCTs), especially when complete case analysis (CCA) is used to handle the missing data. This approach can result in biased estimates and may exclude many in the target population for whom the intervention is intended, limiting use of the results in real-world settings.

Objective: Motivated by missing data issues encountered in a melanoma surveillance trial (MEL-SELF), we conducted a methodological scoping review to identify and summarise missing data methods in RCTs.

Methods: We searched MEDLINE, EMBASE, CENTRAL, and CINAHL for articles in which authors discussed missing data methods in RCTs. One reviewer undertook full-text screening and data extraction, while a second reviewer checked 20 data extractions. Data extractions were also discussed with three other reviewers.

Results: From 1872 records screened, 101 eligible papers were included in the review. 11 (11%) papers were frameworks that provided comprehensive guidance for handling missing data, including a machine learning framework and a control-based multiple imputation (MI) framework, which extends the use of MI to handle missing not at random (MNAR) data. 90 (89%) papers explored the use of one or more method(s), with 25 methods in total. 4 (16%) of these methods could be used under the missing at random (MAR) assumption, 15 (60%) under the MNAR assumption and 6 (24%) under a hybrid of MAR and MNAR. MI was the most common MAR method, while control-based methods were the most performed MNAR method. Novel methods under MNAR also featured including some using machine learning techniques. While these MAR and MNAR methods typically performed favourably compared to methods such as CCA, their performance was only directly compared in 39 (43%) papers.

Conclusion: This scoping review synthesises evidence on different statistical approaches for handling missing data in RCTs, and circumstances where one method may be preferred over another. Use of more robust methods, such as ones used under MNAR, may ensure better use of all trial data. However, further assessment of their performance, particularly that of novel

methods, is needed to better understand their applicability to RCT settings.

Sajeeka Nanayakkara

Development and internal validation of clinical risk prediction models for structural recurrence disease of thyroid cancer

Sajeeka Nanayakkara[1], Jiaxu Zeng[1], Robin Turner[1], Matthew Parry[1] and Mark Sywak[2]

[1]University of Otago, Dunedin, New Zealand.
[2]Endocrine Surgical Unit, University of Sydney, NSW, Australia.

Background: Risk prediction models are crucial for aiding clinical decision-making to improve healthcare. Shrinkage methods have proven beneficial for risk prediction due to preventing the problem of overfitting while improving the precision of predictions. This study focused on the development and internal validation of clinical prediction models for structural recurrence disease of thyroid cancer using shrinkage methods.

Methods: We used a tertiary thyroid cancer service database in Australia from 2000 to 2018. Thirteen demographic and clinical characteristics were considered as risk factors potentially predictable for thyroid cancer recurrence. Shrinkage methods, including ridge, lasso, and elastic net regression analysis were used to construct the prediction models. Predictive performance was compared with conventional logistic regression. Model validation was performed to assess the internal validity of the developed prediction models using optimism-corrected bootstrapping. The predictive performance of the models was assessed based on discrimination, calibration and clinical usefulness.

Results: This study included 2942 patients with thyroid cancer, of which 259 (8.8%) patients experienced the recurrence, which requires further surgeries. The most influential risk factors of structural disease recurrence included post-operative serum thyroglobulin level, pT classification, primary tumor size and lymph node ratio. The performance measures indicated that all prediction models performed well in predicting the risk of structural recurrence in thyroid cancer patients with a Brier score ranging from 0.059 to 0.061, c-statistic ranging from 0.847 to 0.851, and calibration slope ranging from 0.906 to 1.137. Furthermore, the decision curve analysis confirmed that these prediction models could obtain appropriate net benefits, demonstrating their potential clinical practicability.

Conclusions: The standard logistic regression and shrinkage models performed relatively similarly when the dataset is low-dimensional with no multicollinearity problem and with a good selection of relevant risk factors. Identifying the use of these methods for risk predictions in clinical practice remains challenging.

Tri Nisa Widyastuti

Trends in body mass index z-score of Indonesian children and adolescents between 1993 and 2014: a longitudinal population-based study

Nisa Widyastuti(1), Robin Turner(1), Helen Harcombe(2), Rachael McLean(2)

1. Biostatistics Centre, Division of Health Sciences, University of Otago, 18 Frederick Street, Dunedin, New Zealand, 9016
2. Department of Preventive and Social Medicine, Dunedin School of Medicine, University of Otago, 18 Frederick Street, Dunedin, New Zealand, 9016

Background:
Repeated cross-sectional surveys have shown the prevalence of overweight/obesity among Indonesian children and adolescents has increased in recent years. However, no longitudinal studies have examined this. To address this gap, this study was conducted to: (1) examine the trajectory of body mass index z-score (BMIZ) in Indonesian children and adolescents from 1993 to 2014, and (2) investigate different patterns by sex.

Methods:
Children and adolescents aged six to 18 years (n =27,394 individuals and 45,549 total observations) in the Indonesian Family Life Survey, a large-scale population-based longitudinal study from 1993 to 2014 were included in the study. The change in BMIZ across time was estimated using group-based trajectory models, then differences by sex was investigated using random effect (mixed) models.

Results:
The group-based trajectory models showed four distinct groups with each trajectory was quadratic in shape (except Group 1). The first group (11.7% of participants) was of moderately underweight on average throughout the time period; the second group (54.3%) was of normal weight on average with slow increase over time; and the third group (28.4%) was of normal weight throughout the period but the BMIZ was increasing faster than the second group; the fourth group (5.6%) started in the overweight range and moved to obese range by the end of the study period. The shape of these four trajectories differed by sex ($p<0.001$).

Conclusions:
The BMIZ trends confirm that Indonesia is facing the double burden malnutrition. Our findings support policy making in developing more targeted prevention and intervention strategies on handling the dual burden of malnutrition in Indonesia.

Methods and Applications I

Yoram Barak

The interRAI assessment for elder abuse: we can do better? lessons from the elder abuse prevention fund project in New Zealand

Prof Robin Turner[1], Prof Paul Glue[2] and A/Prof Yoram Barak[2]
.
1) Director of the Biostatistics Center, Division of Health Sciences, University of Otago, New Zealand.
2) Department of Psychological Medicine, Dunedin School of Medicine, University of Otago, New Zealand.

This study was supported by a grant from the Ministry of Social Development, Office of Seniors (EAPF), New Zealand.

Background:

Globally one in 6 older adults in the community will be a victim of abuse (elder abuse; EA). Despite these horrific statistics EA remains largely undetected and under-reported. Available screening methods and tools fail to accurately identify the phenomenon's true prevalence. The combined project by the Biostatistics Center and the Department of Psychological Medicine at the DSM aimed to collaborate in order to improve EA detection in NZ.

Methods:
We analysed a methodology to improve detection of EA using the interRAI-HC (International Resident Assessment Instrument - Home Care) which currently underestimates the extent of abuse. The interRAI is a suite of clinical assessment instruments. In Aotearoa New Zealand interRAI is mandatory in aged residential care and home and community services for older people living in the community. They are designed to show the assessor opportunities for improvement and any risks to the person's health.

Findings:
Analysis of 9 years of interRAI-HC data (July 2013 to June 2022) encompassing 186,713 individual assessments from an Aotearoa New Zealand cohort identified that through altering the criteria for suspicion of EA, capture rates of at-risk individuals could be more than doubled from 2.5% to 5.9%.

Interpretation:
We propose that via adapting the interRAI-HC criteria to include the "unable to determine" whether abuse occurred (UDA) category, identification of EA victims could be substantially improved, facilitating enhanced protection of this vulnerable population.

KEYWORDS: Elder abuse, InterRAI, Elderly, Older Adults, Abuse, Screening.

Gabrielle Davie

Estimating the ratio of standardised incidence rates: why, how and what standard population?

Assoc Prof Gabrielle Davie, Dept of Preventive and Social Medicine, University of Otago
Prof Sue Crengle, Ngāi Tahu Māori Health Research Unit, University of Otago
Prof Garry Nixon, Dept of General Practice and Rural Health, University of Otago

Background
Age standardisation enables comparison of incidence rates between populations while taking into account differences in age structure. Standardisation can be undertaken to compare rates between countries or between population groups within a country. In New Zealand, Māori have a younger age-distribution than non-Māori; in 2018, 57% of Māori were less than 30 years of age compared to 37% of non-Māori. Since rates of many diseases are higher in older age groups comparison of the crude rates between Māori and non-Māori may underestimate differences. Standardised incidence rates obtained from age standardisation often use a standard population to 'adjust' for the differences in the age structure between populations enabling an appropriate comparison of rates. In health research a range of standard populations are used, with differing results. Although statistical software packages

may contain commands for standardisation, obtaining the 95% Confidence Interval (CI) for the ratio of standardised rates, may require additional coding.

Purpose/objective
The purpose of this presentation is to use research undertaken recently comparing rates of amenable mortality during 2013-2017 for Māori and non-Māori in rural and urban areas,
1) to illustrate the impact of different standard populations on direct standardised incidence rates (SIRs) and Māori:non-Māori age-standardised incidence rate ratios (SIRRs) and
2) to present Stata commands that were used to obtain 95% CIs for SIRRs.

Conclusion
Using the Māori 2001 Census population as the standard population, the SIR for amenable mortality for Māori in urban areas was 134.9 per 100,000 person-years and 55.1 per 100,000 person-years for non-Māori in urban areas. The Māori:non-Māori SIRR for those living in urban areas was 2.45 (95%CI 2.36, 2.54). For those living in rural areas the SIRs were 152.9 and 65.5 per 100,000 person-years for Māori and non-Māori respectively giving an SIRR of 2.34 (95%CI 2.19, 2.49). In the presentation findings obtained using different standard populations, namely the Māori 2018 Census population, the New Zealand 2018 Census population and the WHO world population will be compared. A useful addition to Stata's direct standardisation command 'dstdize' would be the ability to compare age-standardised rates using ratios and 95% CIs.

Rebecca Harding

Intravenous iron interventions for anaemia during pregnancy in low- and middle-income countries: application and considerations of an adaptive trial design

Rebecca Harding[1], Cally Munro[2], Robert K Mahar[2], Julie A Simpson[2], Sabine Braat[1,2]

Affiliations:
[1] Population Health and Immunity, Walter and Eliza Hall Institute of Medical Research
[2] Biostatistics Unit, Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne

Background:
Anaemia in pregnancy is a significant global health issue, in particular in low- and middle-income countries (LMICs). Reducing maternal anaemia prevalence is a key 2025 World Health Organization (WHO) nutrition target. Approximately 50% of pregnant women with anaemia also have low iron stores (i.e., iron deficiency). Five parallel group individually randomised controlled trials (RCTs) investigating the effectiveness and safety of intravenous iron compared to oral iron are ongoing across Africa and Asia, involving approximately 7700 mothers. These trials use conventional designs without planned modifications to the sample size and with no or a single highly conservative look at the accumulating evidence while the study is ongoing. Adaptive trial designs that allow for modifications during the trial can potentially play a role in resource-poor and low-research capacity settings by reducing sample size and trial duration.

Objective:
To explore design adaptations to a fixed design three arm parallel group individually randomised placebo-controlled trial of two intravenous iron doses delivered in pregnant women with anaemia during pregnancy with the aim to assess their effect on anaemia prevalence just before delivery. The adaptation may increase trial efficiency (e.g., reduce sample size) while generating evidence for anaemia treatment in pregnant women across LMICs.

Principles:
Our adaptions to the fixed study design include adaptive population enrichment (APE), futility analysis, and sample size re-estimation (SSR) using frequentist approaches. APE explores whether the prevalence of anaemia just before delivery differs between iron-deficient and non-deficient participants during the trial, allowing to identify the patient population who are most likely to benefit from the iron interventions. Futility analyses identifies and stops recruitment to ineffective treatments, allowing to focus resources on potentially beneficial interventions. SSR ensures the trial maintains adequate power by accounting for changes from the sample size assumptions made at the study design stage during interim analyses.

Conclusion:
Re-designing the IV iron program into an adaptive program of RCTs offers a promising approach to optimize resources and generate evidence for anaemia treatment in pregnant women across LMICs. However, careful consideration of practical challenges and stakeholder concerns is essential for the successful execution of this novel trial design.

Karen Mumme

A run around the data from the Healthy Active Learning evaluation

Karen Mumme[1], Owen Mugridge[1], Jonathan Godfrey[2], Ajmol Ali[1]

[1]School of Sport, Exercise and Nutrition, Massey University, Auckland: k.mumme@massey.ac.nz, o.mugridge@massey.ac.nz, a.ali@massey.ac.nz
[2]School of Mathematical and Computational Sciences, Massey University, Palmerston North: a.j.godfrey@massey.ac.nz

Healthy Active Learning (HAL), part of the Child and Youth Wellbeing Strategy, is a joint government initiative between Sport NZ, Ministry of Health and the Ministry of Education. The initiative is supported by a government investment of $47 m to improve the wellbeing of children through healthy eating/drinking and quality physical activity. We are evaluating the HAL initiative using a longitudinal quasi-experimental mixed-methods design at three time points: 2020/21, 2022/23 and 2024/25. This abstract describes the data collected at timepoint 2.

Survey data was collected using Qualtrics from more than 3,000 students from 53 primary and intermediate schools over 8 months. The questionnaires measured student engagement, physical activity parameters and motivation for physical activity. Data were collected to create sociograms to assess social engagement within the classroom. A subset of students (n ~720) were fitted with an accelerometer on the non-dominant wrist to measure activity levels for 7 days. Another subset of students (n ~425) attended focus groups to discuss their physical education.

Survey data was collected from ~2500 parents (~250 schools). Questions were asked about community, school culture, physical activity opportunities, parents' values and their children's experiences, attitudes and level of engagement.

Survey data was collected from ~1200 teachers (~250 schools). Questions were asked about working with community, enjoyment, confidence, practice, professional learning and development, school culture and student engagement. A subset of teachers (n ~175) attended focus groups to discuss physical activity as part of the curriculum.

The Ministry of Education provided attendance records for each quarter for 4 years (2019-2022) for 838 schools. The data was aggregated at class level with student attendance at defined percentiles e.g., 90+% attendance and the number of justified and unjustified absences.

This evaluation of Healthy Active Learning will inform many interested parties. Is HAL improving equity efficiently? Has value been created and contributing to wellbeing? Is there a positive return on investment with regards to human and social capital? This evaluation can impact policy, improve the effectiveness of the initiative and ensure wellbeing and success of students which will benefit communities at a local, regional and national level.

Methods and Applications II

Alice Kim

Natural experiments in a health research setting

Alice Kim, University of Otago

Background: A natural experiment design can provide an alternative method for studying causal associations in situations where implementing the traditional experimental design is either unethical or unfeasible. However, the credibility and validity of natural experiment designs depend on the strength of ‘as if’ random exposure assignment and the balance of pre-exposure characteristics.

Purpose/objective: The key elements of natural experiment designs including their strengths and limitations are discussed using examples from studies that apply natural experiment designs in a health research context. We will examine how quantitative and qualitative evidence is used to establish the credibility of natural experiment design in analysing the prevalence of visible goitre in two regions geologically separated by subducting plates. We will also look at how the outcomes of an intervention study are analysed by examining the balance in pre-exposure characteristics. We will draw on lessons from an outcome evaluation study that analysed the one-month follow up outcomes of a mental health crisis response intervention.

James Stanley

Synthetic cohort studies using national level surveys linked to administrative health data

James Stanley, Ricci B Harris

Background: The current New Zealand “health data ecosystem” has many strengths when working with observational study designs based on administrative health data. These data sources,

as held by the Ministry of Health, can be used in cohort-design studies, and can be both time-efficient and cost-effective. However, these administrative data sources are limited in terms of personal level detailed information on exposures and confounders that would be collected within a traditional prospective cohort study, which can restrict their validity for some research questions.

Purpose/objective: This presentation outlines a study aiming to develop a synthetic cohort methodology by linking the New Zealand Health Survey to National Collections health outcome data. Such a design has potential for use as an efficient and cost-effective alternative to a cohort study that relies on primary data collection.

Methods: This presentation describes design and implementation decisions in developing a synthetic cohort study based on linking a large-scale national survey (New Zealand Health Survey, NZHS) to administrative health data sources (e.g. hospitalisation records, mortality records). This approach can provide detailed survey data to form the baseline of a prospective cohort study period, with the linked health data providing excellent follow-up outcomes for the entire sample, giving a cost-effective alternative to traditional cohort study designs.

Results: The presentation will focus on design and feasibility considerations, including data availability and coverage; implications of matching of records between NZHS and subsequent health data; accommodation in analysis of follow-up time, survey weights and other technical aspects; considerations of complexity of models for adjusting for confounding within the joint survey/administrative health data environment; limitations of the approach relative to other design options; and pragmatic challenges in conducting this research in the Stats NZ Datalab environment. Preliminary results may be presented from an exemplary analysis looking at the impact of unmet health need on subsequent hospitalisation risk.

Andrew Gray

Is ChatGPT any use at all for biostatistics?

Andrew Gray

Biostatistics Centre. University of Otago.

Background: ChatGPT and large language models (LLMs) more generally have gained considerable attention due to their ability to provide seemingly intelligent and informed, almost human-like, answers to a myriad of questions without users needing to provide significant amounts of background information. A substantial amount of research has emerged discussing the use of this technology for research and academic writing, with some articles examining this in the context of epidemiology and public health. However, little could be found specifically looking at statistics and biostatistics. As well as this and related technologies being used for code generation, it seems likely that some biostatisticians and other researchers are already using LLMs to find answers to biostatistical questions.

Purpose/objective: While LLMs may be appealing tools, their accuracy in answering biostatistics questions has not been well assessed, particularly for GPT-4 and related models. In this presentation, I assess ChatGPT-4 in terms of the quality of its

response to biostatistical queries concerning study design, data analysis, interpretation, and reporting. I also explore its use in literature searches, biostatistical education, and, for GPT-4, data processing.

Results: Using a series of queries and dialogues, some areas where ChatGPT-4 (with selected plugins enabled for drawing directed acyclic graphs and for performing some basic statistical analyses) performs surprisingly well are identified, alongside a larger number of areas where it fails more often than it succeeds. These experiments are used to formulate guidelines for its use by biostatisticians and non-biostatisticians.

Conclusions: Biostatisticians have no current need to worry that they might be replaced by LLMs. While the technology is not yet capable of consistently providing useful and correct guidance in biostatistics in general, it does show promise in certain scenarios and could be used by biostatisticians now in performing certain tasks. As well as deciding whether, and if so, how, to incorporate LLMs into their workflow, biostatisticians should also be prepared to offer advice to other (non-biostatistician) researchers about the use of LLMs in biostatistics.

Methods and Applications III

Dot Dumuid

Using compositional data analysis to explore the relationships between time use and health

Dorothea Dumuid, University of South Australia and Murdoch Children's Research Institute

Background: Time-use data are compositional because they are made up of mutually exclusive activities (such as sleeping, sitting, standing, walking) which always sum to 24 hours. The activities in a time-use composition are perfectly multi-collinear: if one increases the other(s) must decrease by the same amount to maintain a consistent total of 24 hours. This has made it difficult to include the complete 24-hour composition in standard linear regression models that explore the relationships between time use and health. Compositional data analysis removes the constant sum (to 24 hours) constraint by expressing the compositions as isometric log ratios before their inclusion in standard regression models.

Purpose/objective: This presentation will introduce compositional data analysis, specifically the log-ratio approach. It will demonstrate how compositional data analysis enables the composition to be included in linear regression models that explore the relationships between time use and health. Real population data from the Australian Child Health CheckPoint study (2015-2016, n=1874, 11-12 years, 51% males) will be used to describe time-use compositions within population subgroups, determine how time reallocations between activities are associated with health, and explore optimal 24-hour time distributions for a range of child health measures.

Funding: DD is funded by the Australian Research Council and the Australian Heart Foundation. The CheckPoint study was funded by the Australian National Health and Medical Research Council; the

National Heart Foundation of Australia; The Royal Children's Hospital Foundation; the Murdoch Children's Research Institute; The University of Melbourne; the Financial Markets Foundation for Children; and the Australian Department of Social Services.

Beatrix Jones

The geometry of diet: using projections to quantify the similarity between sets of dietary patterns

Beatrix Jones, Shiyu Fan (Department of Statistics, University of Auckland), Karen Mumme (College of Health, Massey University)

Background: Food consumption is complex and high dimensional. Nutrition researchers measure (or attempt to measure) consumption using food frequency questionnaires, food recalls, or food records. This generates high dimensional data which is frequently summarised using principal components, principal components with rotation, or factor analysis. The resulting axes in the high dimensional space are called "dietary patterns."

Purpose: We define a multivariate extension of Tucker's congruence coefficient (MTCC) to quantify how similar two different sets of dietary patterns are, producing a similarity measure that ranges from zero to one. While the MTCC can be applied to compare different populations assessed with the same dietary instrument, to contextualise the MTCC we compute the similarity for several datasets from the dietary pattern validation literature, where the same food questionnaire is given to the same people a few weeks apart.

Results: 'Good agreement' on conventional validation measures (correlation of dietary pattern scores, Bland-Altman analysis) frequently but not always corresponds to high values of the MTCC, and vice versa. The MTCC is affected by truncation or rounding of loadings values, but is fairly robust to truncation of values less than 0.15.

Robin Willink

Confidence limits for a proportion - a method for improving symmetry of coverage

Robin Willink, University of Otago Wellington, Wellington, New Zealand

robin.willink@otago.ac.nz

Background: The estimation of the binomial proportion is one of the most basic tasks in medical statistics. Existing confidence intervals for the proportion can show a lack of symmetry in the probability of non-coverage in the upper and lower tails. This means that the individual confidence limits are not necessarily suitable for one-sided analyses, for example, in one-tailed hypothesis tests.

Objective: This talk will describe a modification to the limits of the Wilson interval: a method of moments is applied with the first and third moments to improve the accuracy of the approximation to the relevant binomial tail probability. Optimality criteria are discussed, and the method is tuned accordingly. The method involves the solution of only a quadratic equation, and explicit expressions for the limits are only slightly more complicated than the equations for the limits of the Wilson interval. The individual limits can be used to form an interval, but if the only performance criterion relates to the two-sided question then the Wilson interval performs as well.

Andrew Gray

Five things that Stata does that I wish I'd known about earlier than I did (and their equivalents in R)

Andrew Gray

Biostatistics Centre. University of Otago.

Background: As the capabilities of statistical software increase, both in terms of core and user-written functionality, it becomes increasingly difficult to keep track of new features that we would benefit from knowing about. For a given feature, which could be a new command or a new option, the benefit is sometimes in terms of how we, as biostatisticians, perform our own work, but it can also be in terms of the advice we offer to non-biostatistician researchers, including research students. In the Biostatistics Centre at the University of Otago, we promote Stata for non-biostatisticians and focus on it for our teaching. R is also commonly used here, and used even more on other campuses, alongside SAS, SPSS, GraphPad Prism, and others. An earlier survey of statistical software use by biostatisticians at the University of Otago found up to six different programmes used for data wrangling and statistical analyses in research co-authored by individual biostatisticians over an 18 month period.

Purpose/objective: I identify and explain five Stata commands or options that I discovered later than I would have liked. Knowing about these sooner would have helped me with data wrangling, statistical analyses I performed myself, and advice I offered to research students and others.

Results: For each of the five commands/options, I explain why I wish I knew about it earlier and when I would use or recommend it (and some situations where I would not). I also consider the circumstances leading to my ignorance of these helpful commands/options. In order to avoid being too software specific, I identify analogous functions in R.

Conclusions: Staying up to date with the features of even one statistical analysis package is difficult, and many biostatisticians use two, three, or more such packages. I hope that by talking about some of these commands/options in Stata, others will find something new that they can incorporate into their research and/or teaching.

Virtual presentations

Farah Yasmine Ainun Nisa

Nurses' therapeutic communication and its effect on hospitalized patients' satisfaction

Farah Yasmine Ainun Nisa[1], Ratna Dwi Wulandari[2], Tatin Wahyanto[3]

[1,2]Faculty of Public Health, Airlangga University; Mulyorejo,Surabaya, Indonesia
[3]Faculty of Public Health, North Sumatera Health Institute ; Medan Tuntungan, Medan, Indonesia

Corresponding Email:farahyasmine39@yahoo.com

Patient satisfaction is one of the hospital indicators used to measure hospital quality. However, the result of patient satisfaction measurement in the inpatients' rooms showed that the level of patient satisfaction was below target. Therefore, it's crucial to carry out a series of therapeutic communication processes as an effort to improve patient satisfaction. This study sought to determine the effect of nurses' therapeutic communication on patient satisfaction in the inpatients' rooms. This was an analytical study with a cross-sectional design conducted from May to June 2022 at Muji Rahayu Hospital, with 70 respondents obtained using systematic random sampling technique. Primary data was gathered from questionnaires that had been distributed to respondents earlier. SPSS was used to analyze data with the Spearman test used. This study revealed that as many as 48,6% of nurses carried out therapeutic communication well. This study also showed that the orientation phase (p=0.000) and work phase (p=0.000) affected patient satisfaction, while the termination phase had no effect (p=0.081). Collectively, there was a significant effect between therapeutic communication on patient satisfaction (p=0.000).

Keywords: therapeutic communication, patient satisfaction, nurse, hospital

Methods and Applications IV

Robin Willink

A simple non-parametric test more powerful than Mann-Whitney

Robin Willink, University of Otago Wellington, Wellington, New Zealand

robin.willink@otago.ac.nz

The Wilcoxon rank-sum or Mann-Whitney (WMW) test is a standard procedure for detecting a difference between two populations. It seems to be used in clinical research whenever medical professionals and students encounter data that cannot be justifiably treated as having arisen from normal distributions. The procedure involves the transformation of the data into ranks. But it is not the most powerful available. Here we show by example that, in almost every realistic situation, power is greater with an alternative permutation-based procedure involving transformation of the data to 'normal scores'. This idea is not novel. The method described here is essentially the same as a test proposed by van der Waerden in 1952/53, which has been described as being "competitive" with the WMW test. But when we replace pre-computer-age results that apply to infinite sample sizes, symmetric distributions and simple shift alternatives by simulation results applying with realistic sample sizes, realistic distributional forms and realistic treatement effects, the superiority of van der Waerden's test becomes apparent. This test should be made available by default in statistical software.

John Pickering

Faking it to making it: The use of synthetic data to derive biomarker diagnostic thresholds

John Pickering

Background: Emergency Department (ED) doctors use the myocardial injury marker, troponin, to risk stratify patients being examined for possible myocardial infarction (MI; heart attack). The advent of high-sensitivity troponin assays means where troponin

concentrations are very low patients may be safely discharged home after just one blood draw. The determination of the troponin threshold for “very low” is hindered by small data sets and lack of understanding of the effect of measurement error.

Objective: To demonstrate how synthesising more troponin data from a research data set may improve estimation of diagnostic thresholds and understanding of the precision needed for an assay at these thresholds.

Methods: From base research data of patients being investigated for MI and with at least one troponin measurement multiple synthetic data sets were created using bootstrapping and sampling from probability distributions of the base data. The determination of a threshold with >99% sensitivity for MI is compared between raw data and synthetic data. How analytical precision affects safety (sensitivity) and effectiveness (proportion <threshold) of a threshold was simulated by creating several new data sets each of which had troponin where imprecision was simulated on a normal curve with standard deviations of 0 to 50% of the low-risk threshold.

Results: One million fake patients with characteristics similar to those in the baseline dataset were synthesised. Random draws of fake patients containt 50 to 5000 MI patients demonstrate that the sensitivity of the low-risk threshold to the size of the data set. The safety was unaffected at the low-risk threshold by even large imprecision. The effectiveness reduced marginally at high imprecision.

Conclusions: Synthetic data creation of biomarker data enabled insights into the numbers needed to determine optimal low-risk troponin thresholds. Additional insights gained on the effect of measurement impresion on the safety and effectiveness of troponin thresholds could be used to guide minimum standards for assay precision at these thresholds.

Andrew Gray

What is biostatisticalityness?

Andrew Gray and Robin Turner.

Biostatistics Centre. University of Otago.

Background: Despite advancements in reporting guidelines relating to statistics, and the adoption of statistical editorial boards and statistical review by some journals, the literature across the health sciences remains replete with conceptual and application errors in biostatistics. Given the current ratio of biostatisticians to other researchers, it is not possible for biostatisticians to be involved in all aspects of all quantitative studies. Alongside training new biostatisticians, building biostatistics capacity also involves educational interventions for non-biostatisticians, including research students. These interventions include workshops and short courses, such as those offered by the Biostatistics Centre at the University of Otago and elsewhere. To encourage researchers to attend such workshops/courses, it would be helpful if we were able to assess participants’ understanding of biostatistics, both before and following intervention, giving them feedback on their progress and identifying areas for further improvement. We also need to develop an evidence base for what works to encourage investment in biostatistical training by research institutions.

Purpose/objective: By ‘biostatisticalityness’, named with tongue firmly in cheek, we mean the understanding of biostatistics than enables researchers to, as safely as possible, interpret the literature and to design, collect data for, analyse, interpret, and report their own quantitative studies in the health sciences. In other words, the construct that we aim to promote through workshops/courses. We have performed a scoping review of existing instruments claiming to assess statistical knowledge. None of these captured quite what we mean by ‘biostatisticalityness’ and so we developed a novel instrument for this purpose.

Results: In this presentation, we talk about the requirements we have for such an instrument and describe our instrument, distinguishing it from related instruments in statistics. As well as discussing our experiences to date, we present a vision of how our instrument could be used going forward.

Conclusions: Whether it is ‘biostatisticalityness’ or some related construct, there is a pressing need for evidence-based research supporting biostatistics workshops and short courses. It is possible that this could also be used in introductory University biostatistics courses.