Target Trial Emulation for Real-world vaccine effectiveness using Big Data and Machine Learning

Vivek Jason Jayaraj (1)
Masliyana Husin (2), Peter Seah Keng Tok (2), Mohd Azahadi Omar (3), Sanjay Rampal (4), Sheamini Sivasampu (5)

1 : Digital Health Division, Ministry of Health Malaysia
2 : Institute for Clinical Research, National Institutes of Health, Ministry of Health Malaysia
3 : Biostatistics and Data Repository Sector, National Institutes of Health, Ministry of Health Malaysia
4 : Centre for Epidemiology and Evidence-based Practice, Department of Social and Preventive Medicine, Faculty of Medicine, Universiti Malaya, Malaysia
5 : National Public Health Laboratory, Ministry of Health Malaysia

Real-world evaluations of vaccine effectiveness (VE) against severe COVID-19 often cannot rely on randomised controlled trials (RCT) due to ethical and practical constraints. We aimed to replicate randomisation for counterfactual development through target trial emulation and propensity score matching, enhancing the estimation of the relative VE (rVE) of the fourth COVID-19 dose compared to the third in preventing hospitalisations and deaths. Our target trial framework capitalised on linked data from multiple administrative databases covering 25 million individuals. The population-level cohort included Malaysians aged 18 and above vaccinated between 1 May and 1 October 2022, with no prior SARS-CoV-2 infection, and had received at least three vaccine doses. Propensity score matching (PSM) was executed via the comparison of several machine learning algorithms—Random Forest, XGBoost, CatBoost, and an ensemble meta-learner—to estimate scores from covariates such as age, comorbidities, ethnicity, region, and vaccination date. Algorithms were assessed for computational efficiency and between-group balance using descriptive statistics, visualisations, and standardised mean differences (SMDs). The rVE was estimated using a Cox Proportional Hazards model with time-varying coefficients. Analysis was conducted in R and Python, leveraging Amazon Web Services for computing. Our matched cohort included 120,717 adults per group. The XGBoost nearest-neighbour algorithm achieved the most balanced covariates. The rVE for the fourth dose showed non-significant differences in COVID-19 hospitalisations and deaths compared to the three doses. Despite finding no significant additional VE from the fourth dose, this study showcases the utility of big data and machine learning in emulating pragmatic trials for VE studies, circumventing some of the limitations of RCTs. This study demonstrates a scalable approach to VE estimation using large datasets and advanced analytical techniques.

Selection bias in the test-negative case-control design

Yuto Kono (1)
Chieko Ishiguro (2), Yasuhiro Hagiwara (3), Haruhisa Fukuda (4), Yoshinori Takeuchi (5)

1 : Department of Data Science, Graduate School of Data Science, Yokohama City University, Kanagawa, Japan
2 : Section of Clinical Epidemiology, Department of Data Science, Center for Clinical Sciences, National Center for Global Health and Medicine, Tokyo, Japan
3 : Department of Biostatistics, School of Public Health, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan
4 : Department of Health Care Administration and Management, Kyushu University Graduate School of Medical Sciences, Fukuoka, Japan
5 : Department of Data Science, School of Data Science, Association of International Arts and Sciences, Yokohama City University, Kanagawa, Japan

Background : The test-negative case-control design (TND) has been widely applied to estimate vaccine effectiveness (VE), to mitigate bias from healthcare-seeking behavior (HSB) by including only symptomatic patients. Yet, it may induce selection bias when vaccination affects testing behavior, particularly with residual HSB. However, the impact of this selection bias has not been fully evaluated.

Objectives : To evaluate the impact of selection bias in the TND when vaccination influences testing behavior.

Method : We conducted simulation experiments by generating artificial cohorts consisting of 10,000, iterated 1,000 times, to compare the estimates obtained from TND, ordinary case-control (CC) design, and ideal cohort design. The TND enrolled test-positive cases and test-negative controls, the CC design enrolled controls irrespective of testing, and the cohort design assumed full ascertainment of outcomes irrespective of testing. We simulated scenarios for VE estimation, varying the influence of HSB and vaccination on testing behavior. Performance was assessed by bias and coverage of confidence interval. Moreover, we compared the results of these three designs for the existing study evaluating COVID-19 VE against infection from the Vaccine Effectiveness, Networking, and Universal Safety (VENUS) study data.

Results : When testing was influenced by HSB alone, TND and CC design estimates were slightly biased with reduced coverage. When vaccination influenced testing, TND estimates showed substantial bias in both directions and markedly lower coverage, CC design estimates showed slight bias with reduced coverage, and cohort design estimates were unbiased. We will show the findings from the VENUS study at the conference.

Conclusion : The TND insufficiently addresses bias from HSB and may introduce further selection bias. As the ideal cohort design is often infeasible and the CC design also yielded biased estimates, alternative approaches are needed to improve VE estimation.

Healthcare cost savings from delaying preterm birth in Northern Territory, Australia: A linked study

Tsegaye G. Haile (1,2)
Richard Norman (1), Winnie Chen (3,4), Ravisha Srinivasjois (1,5), Gavin Pereira (1,6,7), Gizachew A. Tessema(1,6,8)

1 : Curtin School of Population Health, Curtin University, Western Australia, Australia
2 : Department of Health Systems and Policy, Institute of Public Health, University of Gondar, Ethiopia
3 : Menzies School of Health Research, Charles Darwin University, Casuarina, Northern Territory, Australia.
4 : Leeder Centre for Health Policy and Economics, Faculty of Medicine and Health, University of Sydney, New South Wales, Australia
5 : Department of Paediatrics and Neonatology, Joondalup Health Campus, Western Australia, Australia.
6 : enAble Institute, Curtin University, Western Australia, Australia
7 : WHO Collaborating Centre for Climate Change and Health Impact Assessment, Faculty of Health Science, Curtin University, Western Australia, Australia.
8 : School of Public Health, University of Adelaide, South Australia, Australia

Background : Although survival after preterm birth (PTB) has improved, associated complications continue to impose a substantial burden on healthcare systems. Even modest prolongation of gestation, by as little as one week, may significantly reduce morbidity and healthcare costs. Thus, this study estimated the incremental healthcare cost savings of delaying PTB during early childhood in the Northern Territory (NT), Australia.

Methods : We conducted a retrospective cohort study of all births from July 1, 2000, to June 30, 2016, with records linked to hospitalisation, emergency department (ED) presentations, and cost-weight data through June 30, 2021. A total of 31,169 and 42,139 births were linked with hospital and ED records, respectively. Five-year healthcare cost savings from delaying preterm birth was calculated from a health system perspective and adjusted to June 2024 Australian dollars (AUD). A Generalised Additive Model (GAM) with gamma distribution and log link was used, with covariates selected through the least absolute shrinkage and selection operator (LASSO).

Results : Our findings indicated that delaying PTB by one week, from 23 to 24 weeks' gestation, yielded substantial per-child cost savings, up to AUD 58,412 for hospitalisations and AUD 11,343 for ED presentations. Shifting births from extremely preterm (23–27 weeks) to very preterm (28–32 weeks) resulted in average savings of AUD 43,946 (95% CI: 33,869–54,023) for hospitalisations and AUD 1,522 (95% CI: 62–2,983) for ED visits. Healthcare costs were non-linearly associated with gestational age, maternal age, and duration of hospital stay.

Conclusions : While per-child savings were greatest at the lowest gestational ages, the largest system-wide savings were observed among late preterm births due to their higher prevalence, highlighting the broad economic value of delaying PTB across all gestational ages.

LLM automation of Abstract Screening: Assessing Feasibility via Cross-Validation with Human Expert

Chai Hong Rim (1,2,3,4)
Kyung Hwa Lee (1,5)

1 : Korea University, College of Medicine, Dept. Radiation oncology, Seoul, Korea
2 : Korea University, College of Medicine, Dept. Medical Informatics, Seoul, Korea
3 : CURE analytics, Korea
4 : Korea University Ansan Hospital, Dept. Radiation Oncology, Gyeonggido, Korea
5 : Korea University Guro Hospital, Dept. Radiation Oncology, Seoul, Korea

Introduction : Systematic review (SR) derives high-level evidence from vast literature but requires substantial human labor. Abstract screening is the most labor-intensive step, and automation using large language models (LLMs) is increasingly needed. Prior studies compared LLM models on pre-designed datasets, yet practical perspectives from SR experts remain limited.

Method : With consultation from the SR company *CURE Analytics*, we analyzed two abstract databases: one on adjuvant radiotherapy after surgery for cholangiocarcinoma collected in 2020, and another on combined chemo-radiotherapy for metastatic lung cancer collected in 2025. Abstract screening was automated using the LLM (OpenAI GPT-4 via the Chat Completions API). Dataset-specific inclusion and exclusion criteria were embedded in system prompts, and the model produced both classification results and corresponding reasons. Results were compared with human expert screening, and discrepancies were qualitatively assessed by an oncologist of *CURE analytics* who created the databases.

Result : The cholangiocarcinoma database, after removing minor duplicate studies from previously published data, contained 696 abstracts, of which 89 (12.8%) were included. Agreement rate, sensitivity, specificity, and NPV were 90.4%, 60.3%, 94.5%, and 94.6%, respectively. The lung cancer database contained 431 abstracts, of which 17 (3.9%) were included. Corresponding rates were 97.0%, 82.4%, 97.6%, and 99.3%, respectively. The main reasons for discrepancies between human and LLM decisions included: deliberate leniency by humans to avoid missing data (43%), AI's lack of understanding of field-specific academic conventions (27%), and insufficient human accuracy (8%).

Conclusion : LLMs can serve as reliable assistants for SR abstract screening. By relying on LLM support to reduce excessive leniency and by designing prompts that better interpret field-specific expressions, LLM-assisted screening could become more efficient.

Reduction of Selection Bias Using a Transportability Formula: The Yokohama Health Study

Atsushi Goto (1)
Izumi Nakayama (1), Kosuke Inoue (2)

1 : Department of Public Health, Graduate School of Medicine, Yokohama City University, Yokohama, Japan.
2 : Graduate School of Medicine and School of Public Health, Department of Health Promotion and Human Behavior, Kyoto University, Kyoto, Japan.

Background : Insufficient response rates can introduce selection bias, undermining the validity of results. A transportability formula with national survey data can be used to reduce bias. However, the effectiveness of this approach in mitigating bias has not been thoroughly explored in population-based studies.

Methods : The Yokohama Health Study (YHS) is a population-based cohort study conducted from January to March 2023. A random sample of 80,000 residents aged 30 to 69 was drawn from the basic resident register of Yokohama City, and invitations to participate were mailed. The study achieved a response rate of 12%, with 9,582individuals participating. We also used data from the 2022 Comprehensive Survey of Living Conditions (CSLC), covering the same age group in Yokohama City (n = 1,963). Inverse odds of sampling weights were applied to correct YHS estimates for selection bias, then modified Poisson regression models calculated risk ratios (RRs) and 95% confidence intervals (CIs) for the association between education level and gastric cancer screening uptake.

Findings : Individuals with lower educational attainment were less likely to seek gastric cancer screenings than those with higher education. RRs for high school graduates vs. university were 0.56 (0.47-0.67) in CSLC, 0.76 (0.72-0.80) in unweighted YHS, and 0.56 (0.50-0.62) in weighted YHS. The weighted RR in YHS, excluding exposure and outcome variables for weight calculation, was 0.69 (0.63-0.76).

Conclusions : Estimates from the unweighted YHS were closer to null than CSLC, indicating selection bias from volunteering. The transportability formula applied to the CSLC data effectively reduced this bias, as the weighted YHS estimates became more consistent with those from the CSLC, even when exposure and outcome variables were not used for weight calculation. This highlights the importance of the transportability methodology in minimizing selection bias in population-based studies.