PyRrHiC: Pharmaceutical Benefits Scheme Replication and Harmonisation Challenge using Australian dispensing data

Malcolm Gillies, Benjamin Daniels, Chrianna Bharat, Ximena Camacho, Kelly Hall, Lan Kelly, Erin Kelty, Jialing Lin, Melisa Litchfield, Derrick Lopez, Firouzeh Noghrehchi, Jacques Raubenheimer, Sallie-Anne Pearson, Claire Vajdic, Bianca Varney, Nicole Pratt

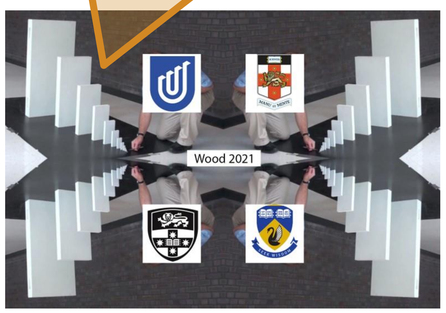
Presenting author: Kelly Hall

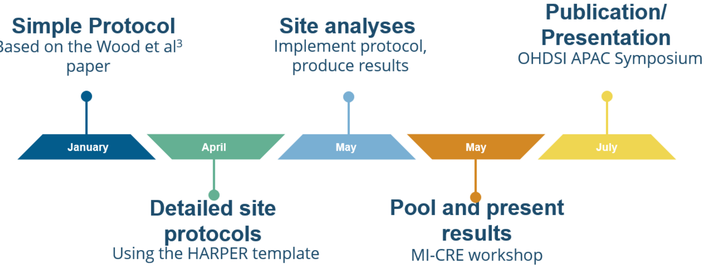
Background

- Replicability of study findings by different research teams is of fundamental importance to pharmacoepidemiology research but may be challenging in practice.
- Researchers make different, apparently minor, technical choices during data preparation and analysis that can lead to different results.
- In 2021, OHDSI embarked on a reproducibility challenge, where nine teams aimed to reproduce the cohort logic for the target, comparator and outcome cohorts[1]. That study found that only the simplest criteria were easy to reproduce and on average, the teams did not reproduce 60% of the criteria.
- Using this challenge as our inspiration, we aimed to conduct our own replicability study (PyRrHiC) to inform research practices using a medicine dispensing research dataset available in Australia (PBS10% sample). Four sites in the Medicines Intelligence Centre of Research Excellence (MI-CRE) participated. Each site completed the HARmonized Protocol Template to Enhance Reproducibility (HARPER) protocol[2].
- **Our study aimed to:**
  1. identify variation in data preparation and analysis for drug utilisation studies and measure its impact on replicability;
  2. develop guidance on data preparation and analysis for drug utilisation studies;
  3. develop documentation standards.

Methods

Each site accessed the same simple protocol based on an existing treatment dynamics study[3]. Each site independently produced a detailed protocol using the HARPER template[2]. Analysis was then conducted separately at each site in a variety of languages (SAS, Stata and R).





The study characterised the treatment dynamics of adults initiating metformin using an Australian pharmaceutical claims dataset. Two cohorts were studied to account for data collection changes over time. Outcomes of interest were the proportion that discontinue, switch or intensify treatment and the effect of age and sex on the time of each outcome.

After completion of analysis, results were compared in a standard format and pooled. Concordance measures were determined *a priori* to allow for comparison of results across the four sites. A traffic light rating was applied to each measure to indicate whether there was concordance across the four sites.

Rating	Definition
Good	1. Deviation of counts <5% (relative) from median value 2. Deviation of proportions <5% (absolute) from median value 3. Deviation of HR_i/HR_{median} <5% (relative) 4. Deviation of ages, median survival times <0.1 (standardised) In addition, for HR: 1. Point estimates of HR all on same side of unity and all significant; *OR* 2. Significance tests all non-significant
Moderate	1. Deviation of counts <10% (relative) from median value 2. Deviation of proportions <10% (absolute) from median value 3. Deviation of HR_i/HR_{median} <10% (relative) 4. Deviation of ages, median survival times <0.2 (standardised)
Poor	Does not meet definition of Good or Moderate

References

1. Ostropolets A, Albogami Y, Conover M, Banda JM, Baumgartner WA, Blacketer C, et al. Reproducible variability: assessing investigator discordance across 9 research teams attempting to reproduce the same observational study. J Am Med Inform Assoc. 2023 Apr 19;30(5):859-868.
2. Wang SV, Pottegard A, Crown W, Arlett P, Ashcroft DM, Benchimol EI, et al. HARmonized Protocol Template to Enhance Reproducibility of hypothesis evaluating real-world evidence studies on treatment effects: A good practices report of a joint ISPE/ISPOR task force. Pharmacoepidemiol Drug Saf. 2023;32(1):44-55.
3. Wood S, Magliano DJ, Bell JS, Shaw JE, Ilomäki J. Treatment Dynamics in People Who Initiate Metformin or Sulfonylureas for Type 2 Diabetes: A National Cohort Study. *Front Pharmacol*. 2021;12.

Results

	Site	2007 - 2010	2015 - 2019
Cohort size (N)	Site 1	23,847	54,405
	Site 2	23,245	51,848
	Site 3	24,334	55,273
	Site 4	23,182	51,850
Sex Male (%)	Site 1	48.9	43.2
	Site 2	48.9	43.1
	Site 3	48.8	42.9
	Site 4	48.9	43.1
Age (Median (IQR))	Site 1	63 (52-71)	54 (39-66)
	Site 2	63 (52-71)	53 (38-65)
	Site 3	63 (52-71)	54 (39-66)
	Site 4	63 (52-71)	53 (38-65)
Concession* (%)	Site 1	93.9	44.9
	Site 2	94.4	44.7
	Site 3	83.8	42.6
	Site 4	83.8	42.3
Died (%)	Site 1	0.9	0.5
	Site 2	4.2	2.5
	Site 3	5.1	3.1
	Site 4	1.7	0.9

A similar number of patients were included during each study period with comparable demographic profiles for sex and age, indicated by a rating of 'good'.

Exposure duration was calculated differently in each site and different approaches to defining discontinuation, switching and intensification were used. This resulted in deviations in the percentage of patients reaching each outcome at 12 months.

Variation in the coding of concession status and death introduced differences into the cohorts used at each site. The difference is indicated by ratings of 'moderate' and 'poor'.

*Concession or health care card holders eligible to get cheaper medicines.

Conclusions

What we learnt	• Data curation and preparation differs across sites • Different approaches used to calculate exposure duration and to define outcomes • Starting from the same dataset and basic protocol does not guarantee replicability
How OHDSI will help?	• Using a shared mapping of PBS item codes to Australian Medicines Terminology (AMT) and RxNorm will allow for a standardized approach to estimation of exposure duration across Australia • Consistent definitions of outcomes (discontinuation, switching, intensification) will be developed using validated phenotypes • Application of OHDSI tools will facilitate standardized analytics overcoming potential differences introduced when a variety of statistical software is used
Final thoughts	Drug utilisation studies are difficult to implement consistently but are critical in Australia where a universal medicines subsidy framework is used. These data are important for our medicines subsidy committee, PBAC, to ensure that robust cost-effectiveness estimates are available for their decision making. Collaborating to create harmonized approaches will ensure rigor and build trust in those estimates. Understanding how minor changes in data curation and preparation, as well as definitional and analytical differences, impact on results is critical to generalizability and interpretability. In addition to this, careful sensitivity analyses are needed to ensure results adequately reflect true analytical uncertainty.

Acknowledgements: This research is supported by the National Health and Medical Research Council (NHMRC) Centre of Research Excellence in Medicines Intelligence (ID: 1196900). We thank Services Australia for the use of the PBS 10% dataset for this project and acknowledge that the PBS data has limitations and is not suitable for all drug utilisation studies.