Appendix to: Francia C, Johnston LM, Stacey I, et al. The epidemiology of acute rheumatic fever and rheumatic heart disease in Queensland, 2017–2021: a population-level cohort study using linked administrative data. *Med J Aust* 2025; doi: 10.5694/mja2.70052.

Supplementary materials and methods

This Queensland End RHD (QERHD) project supplementary methods and materials section commences with a description of the Queensland health system and geographical distribution of the 16 public health districts or Hospital and Health Services (HHSs). The rationale for the use of linked data is provided followed by a detailed explanation of the study design, population, data sources and variables, data linkage processes, and ascertainment of cases.

S1.1 Queensland Hospital and Health Services

The Queensland health system is made up of a network of public and private service providers. Public health services are administered through 16 separate Hospital and Health Services, overseen by the Queensland State Government's Department of Health. People with acute rheumatic fever (ARF) and rheumatic heart disease (RHD) receive hospital care within their local HHSs, but elective cardiac procedures and emergent care (e.g. decompensated heart failure) will require access to advanced cardiac services and rescue therapies (valve surgery, heart transplant, extracorporeal membrane oxygenation) located in facilities concentrated in metropolitan areas along the eastern coast of Queensland. This represents a geographical barrier for health care access for people with RHD and their families living in the endemic regions in the remote far north of Queensland who require long term specialist care and monitoring for disease progression. The urban south-east corner of Queensland is supported by six HHSs.

S1.2 Rationale for linked data methods

Retrospective identification of people with ARF/RHD in hospital administrative data is an important method for epidemiological disease surveillance and policy evaluation. The QERHD project dataset includes linked laboratory, hospital admissions, emergency presentation, death and RHD Register data to provide a comprehensive description of the Queensland ARF/RHD population and estimate disease burden. Linking of datasets in this way permits the records of people to be followed across different data collections, compensating somewhat for the data incompleteness of single data collections. This more complete picture also allows longitudinal analysis including estimates of initial disease onset and progression, and associated risk factors to be investigated in a series of QERHD project studies. Coding of disease diagnoses within these datasets uses the International statistical classification of diseases and related health problems, 10th revision, Australian modification

(ICD-10-AM).[1] However, there are concerns regarding the validity of ICD codes for identifying RHD cases due to the likelihood of misclassification of non-rheumatic valvular disease and ARF as RHD.[2,3] An algorithm developed by Bond-Smith et al[3] and validated using Queensland hospital data has demonstrated good performance in reducing the false-positive rate from misclassification of both ARF (0.59 to 0.27) and non-rheumatic valvular heart disease (0.7 to 0.22).[3] Therefore, these methods were adapted to optimise ascertainment of RHD cases from within our study data.

S1.3 Materials and methods

The QERHD project adapted elements of the multijurisdiction ERASE project methods for the purpose of single jurisdiction (Queensland) linked data analysis.[4]

S1.3.1 Data sources, data linkage and data-generating cohort

Figure 1 illustrates the sequence of steps involved in establishing the QERHD data collection, including ethics approvals, data sources and custodians, and the linkage process. The data-generating cohort comprised all Queensland residents with ARF or RHD, specifically: a record in the Queensland RHD Register or at least one Queensland hospital admission with ICD-10-AM codes for ARF (I00 to I02) and/or RHD (I05 to I09) in any diagnostic field. Queensland RHD Register data and Queensland Hospital Admitted Patient Data Collection (QHAPDC; inclusive of all private and public hospital patient records) data were deterministically linked by the Queensland Health Statistical Analysis Linkage Unit via the Queensland Health Master Linkage File (MLF). The MLF is an index which maps a person's records between multiple Queensland data collections using a unique person-key and personal identifiers, enabling person-level analysis of records between separate data collections. Of the 3749 unique persons from the Queensland RHD Registry, 3747 (99.9%) were matched to a MLF person-key. All existing QHAPDC data relating to all Queensland private and public hospital admissions (from 2007 and 2001, respectively) with a principle or other diagnosis for ARF (ICD-10-AM I00-I02) or RHD (ICD-10-AM I05-I09) were obtained and linked to the MLF. Combining the RHD Registry and QHAPDC data together resulted in 40,291 unique persons, after the exclusion of non-Queensland residents. At this stage, an 'index date' based on the earliest record of an ARF episode or RHD diagnosis from either data source was calculated for each cohort member to mark the earliest date of data extraction from the QHAPDC of hospital admissions and the Queensland Emergency Data Collection (EDC) of emergency presentations. Using the MLF, all QHAPDC and EDC records from 5 years prior to the index date to December 31st, 2021, with a minimum possible date of 1st of January 2001 for hospital admissions and 1st of July 2008 for emergency presentations, along with all death data after 2001 from the Queensland Death Register and Cause of Death Unit Record File, were extracted. Finally, Pathology Queensland (Queensland Health public laboratory) records of all positive Strep A cultures and bloods provided were linked separately to the MLF using deterministic and probabilistic methods based on patient demographic and address details. Of the total 233,588 lab numbers provided for linkage, 223,506 (95.7%) were linked to the MLF, and assigned a matching person-key. All records matching to the defined cohort were then extracted from the source Pathology Queensland records, comprising the data-generating cohort. The data-generating cohort was prepared for epidemiological analysis using methods adapted from the End RHD in Australia: Study of Epidemiology (ERASE) project.[4] In particular, it was important to address the known systemic biases in hospital administrative data, including misclassification of RHD ICD codes resulting in false-positive cases[3] and underestimation of Indigenous people due to misclassification or non-recording.[5,6]

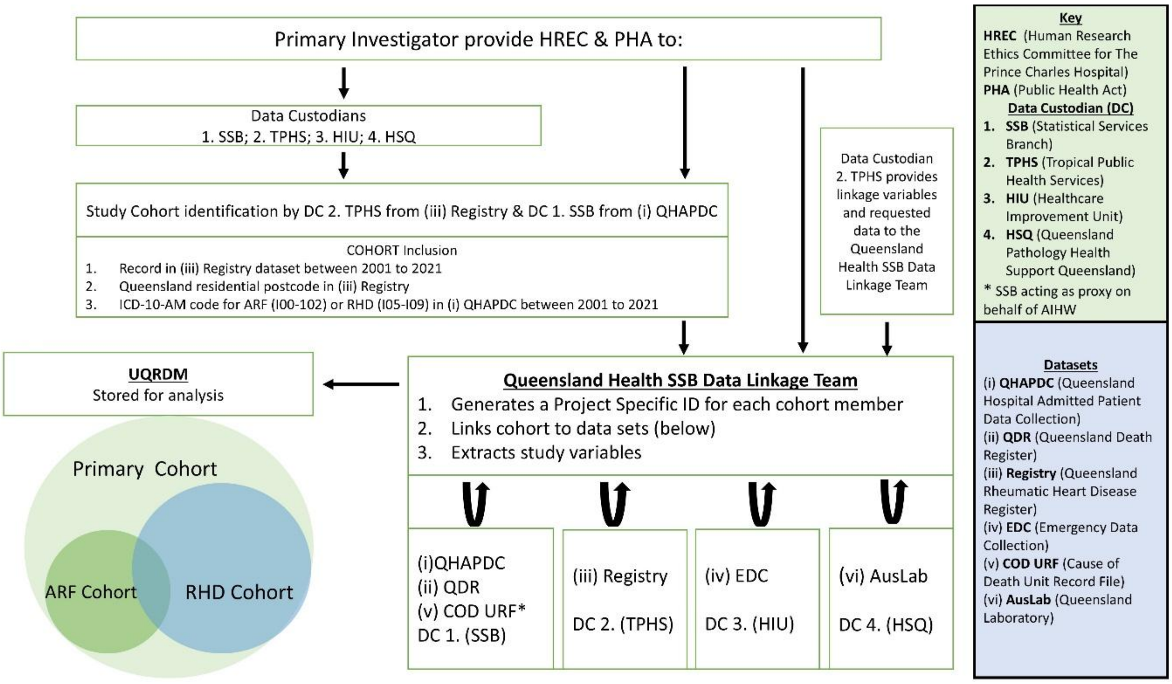

Figure 1. Diagram of the data linkage process for the Queensland End Rheumatic Heart Disease (QERHD) data collection, including datasets and associated data custodians at time of linkage (right set). ICD-10-AM, International statistical classification of diseases and related health problems, 10th revision, Australian modification; UQRDM, the University of Queensland Research Data Management.

CONSIDER statement

CONSIDER Governance

The study was endorsed by the Queensland Aboriginal and Islander Health Council (peak body) and Queensland Health Aboriginal and Torres Strait Islander Health Division. Our research team is led by a Torres Strait Islander clinician/researcher who led study planning, design, data management/analysis/interpretation, and writing up of study findings.

Prioritisation

Ending rheumatic heart disease (RHD) is a national priority for Aboriginal and Torres Strait Islander communities. The National Aboriginal Community-Controlled Health Association (NACCHO) have called for the elimination of RHD by 2031 through empowering the Community-Controlled Health sector to prevent/control acute rheumatic fever (ARF) and Strep A infections. This agenda is supported by a national guideline (developed in partnership with NACCHO), informed by the landmark national ERASE study through its provision of national disease burden estimates.

Relationships (Indigenous stakeholders/participants and research team)

Indigenous research governance was embedded throughout the life cycle of this study through my own ways of knowing and doing as Saibai Koedal augadgil, in consultation with fellow Indigenous researchers (Poche Institute for Indigenous Health, Menzies), clinical leaders (Qld Health [QH] Aboriginal and Torres Strait Islander Health Division; QH Aboriginal and Torres Strait Islander Clinical Network), policy leaders (Queensland Aboriginal and Islander Health Council) and affected community members. In the absence of a dedicated Aboriginal and Torres Strait Islander Human Research Ethics Committee in Queensland, we ensured the values of spirit and integrity, reciprocity, respect, equity, cultural continuity, and responsibility were embedded in all study activities (as described in our study protocol), through existing ‘mainstream’ ethical governance process.

Methodologies

De-identified prospectively collated data including indigenous status, ethnicity, demographic details, primary disease, comorbidities, and treatment modality were extracted from the Qld routinely collected data, including RHD register, death register, hospitalisations and emergency presentations. Data is presented to mitigate potential for re-identification of individuals (e.g. Presenting counts below 5 as ‘<5’).

In line with the national strategy, comprehensive community-centred ARF/RHD activities have been implemented in Qld. This has been supported by 2021 legislative reform requiring Qld’s 16 public Hospital and Health Services (HHSs) to partner with local Aboriginal and Torres Strait Islander communities to address local health priorities. This ‘*Health Equity’* legislative reform (unique to Qld) is an important ‘lever’ for Qld communities to advocate for, design and implement strategies to prevent Strep A/ARF/RHD with their local HHSs. It is our intention for our work to reclaim routinely collected health data for affected Aboriginal and Torres Strait Islander communities to support this local action, an example of our practical application of Indigenous Data Sovereignty principles.

As such, it was important that our findings (Aboriginal and Torres Strait Islander ARF and RHD disease burden estimates) were presented by geographical HHS region.

Participation

Public Health Act Approval was granted to justify access to routinely collected data without consent of individuals by balance of public benefit.

Analysis and interpretation

Our analysis was led by the primary investigator CF (Saibai Koedal Torres Strait Islander) in consultation co-authors comprising methodological/technical experts with a long track record of analysis and interpretation of linked Indigenous health data in collaboration with Indigenous health services and communities.

Dissemination

The manuscript ‘The epidemiology of acute rheumatic fever and rheumatic heart disease in Queensland, 2017-2021: a linked data study to inform place-based prevention strategies’ will be published in a peer-reviewed journal. First Author CF has presented preliminary findings to Qld Health and community stakeholders, both directly and through and scientific conferences.

S1.3.2 Ascertainment of cases

First ever 'record of ARF' or 'RHD' were identified by searching for the earliest record of ARF/RHD across all data collections within the available lookback period for each person. An ARF episode was defined as a record of ARF in ARF/RHD Register data or a diagnosis of ARF within a primary diagnostic field (ICD-10-AM I00-I02) in hospital, emergency department or death register data collections. Similarly, an RHD case was defined as a person having a record of RHD in ARF/RHD Register data or a diagnosis of RHD (ICD-10-AM I05-I09) within a primary diagnostic field in other data collections. As ARF episodes are potentially recurrent, the clinically informed cut-off of 90 days between ARF episodes was used to distinguish distinct episodes.[4] Therefore, a recurrent ARF episode was defined as occurring > 90 days after a previous ARF episode in any data collection. Conversely, RHD is a chronic condition, so the objective was to identify the earliest reliable record of RHD for each person to establish time of disease onset. A prediction algorithm adapted from the ERASE project was used to strengthen the robustness of our RHD case ascertainment. This algorithm was validated using a large dataset (n=7555) including known Queensland public hospital inpatient cases and non-cases, demonstrating excellent performance in reducing false-positive RHD cases misclassified from ARF (0.59 to 0.27) and non-rheumatic valvular disease (0.77 to 0.22).[3] In addition, an attempt to separate RHD severity into symptomatic severe RHD and mild or moderate RHD was made using the Australian RHD guideline system of classification.[7] The Queensland ARF/RHD Register data collection maintains records of disease severity, so these data were directly extracted for each person where available. In addition, efforts were made to identify asymptomatic latent RHD cases recorded in the ARF/RHD Register data collection detected through community- and school-based echocardiographic screening. Hospital and emergency data were also used to identify onset of 'severe' RHD although it was not possible to distinguish latent RHD cases. A 'severe' RHD case was determined as the first hospital admission or emergency presentation with heart failure (ICD-10-AM code I50), or hospital admission with an RHD-related invasive cardiac procedural or surgical intervention.

S1.3.3 Definitions and characteristics of study cohorts

Derivation of key demographics

Information from across all data sources was used to derive key demographics for each person, including Indigenous status, region of residence, sex, and country of birth using methods adapted from the landmark End RHD in Australia: Study of Epidemiology (ERASE) project.[4]

Indigenous status

The Indigenous status of each person was assigned initially within each data source. For hospital and emergency data, a person was coded as Indigenous depending on their number of total records. People with more than two records were coded as Indigenous if at least two records were recorded as 'Indigenous'. Whereas people with two or less records were coded as Indigenous if one record was recorded as 'Indigenous'. For ARF/RHD Register and death data sources, a person's recorded Indigenous status was assumed to be correct. A low threshold of 'ever' recorded as Indigenous across any data source for each person was then applied to derive the final Indigenous status of each person.

Country of birth, sex, and region of residence

To derive each person's country of birth, sex and region of residence, the mode of all entries within and across all data collections was chosen. This was performed as follows:

1. For hospital and emergency data, if a person had multiple records, the mode of all records was determined. If a person had an equal number of records with different values, then the value of the most recent record was assigned.
2. The mode was then calculated across ARF/RHD Register data, death data, and the modes of hospital and emergency data in the previous step. If no mode could be calculated due to equal number of differing records, then preference was given in the following order (descending): ARF/RHD Register, death data, hospital data, emergency data.

The region of residence of people was captured by two geographical variables across data collections: 1) 2016 Australian Statistical Geography Standard (ASGS) Statistical Area Level 2 (SA2) available in hospital, emergency, and death data; or 2) patient HHS (one of 16 Queensland public hospital areas, a non-ASGS data structure) available in hospital, emergency and ARF/RHD Register data. Note, while the Queensland Children's Hospital HHS serves as a statewide paediatric specialist referral hub located at south metropolitan Brisbane, residents of its small physical catchment area were assigned to the Metro South HHS region. Both were used to derive a third geographical data type called Indigenous Regions (IREG, a 2016 ASGS data structure). Where available, SA2s (the smallest geographical aggregation) were mapped to patient HHSs (the next highest aggregation) to address any missing entries, then both were used to map to IREGs (the highest aggregation).

Population category

People were grouped into one of three population categories: Indigenous; internationally born in a low or middle income country (ILIC); and other. Data recording Indigenous status, ethnicity, and minor country of birth from registry and hospital data (where available) were used to allocate people into population categories. The ILIC category included people who do not identify as Aboriginal and/or Torres Strait Islander but belong to another population known to be at high risk of ARF/RHD, including people born in low or middle income countries or of Māori/Pacific Islander descent.[8,9] If available, data directly reporting Māori/Pacific Islander descent were used to code population category. For ILIC, the population was assigned by evaluating the country of birth recorded for each person against World Bank country income data.[10] A person was coded as ILIC if their derived country of birth was a low to middle income country or New Zealand, the latter given that the person is most likely of Māori/Pacific Islander descent. People not categorised as Indigenous or ILIC were classified as ‘other’.

References

1. Australian Consortium for Classification Development. The international statistical classification of diseases and related health problems, tenth revision, Australian modification (ICD-10-AM). 10th ed. Adelaide: Independent Hospital Pricing Authority, 2017.

2. Katzenellenbogen JM, Nedkoff L, Cannon J, et al. Low positive predictive value of International Classification of Diseases, 10th Revision codes in relation to rheumatic heart disease: a challenge for global surveillance. *Intern Med J* 2019; 49: 400-403.

3. Bond-Smith D, Seth R, De Klerk N, et al. Development and evaluation of a prediction model for ascertaining rheumatic heart disease status in administrative data. *Clin Epidemiol* 2020; 9: 717-730.

4. Katzenellenbogen JM, Bond-Smith D, Seth RJ, et al. The end rheumatic heart disease in Australia study of epidemiology (ERASE) project: data sources, case ascertainment and cohort profile. *Clin Epidemiol* 2019; 15: 997-1010.

5. Australian Institute of Health and Welfare. Indigenous identification in hospital separations data: quality report. Canberra: AIHW, 2009. https://www.aihw.gov.au/getmedia/adcaf32e-d2d1-4df0-b306-c8db7c63022e/13630.pdf (viewed Nov 2024).

6. Australian Institute of Health and Welfare. Report on the use of linked data relating to Aboriginal and Torres Strait Islander people. Canberra: AIHW, 2013.

https://www.aihw.gov.au/reports/indigenous-australians/report-on-the-use-of-linked-data-relating-to-abori/contents/summary (viewed Sept 2024).

7. Bowen AC, Currie BJ, Wyber R, et al. The 2020 Australian guideline for prevention, diagnosis and management of acute rheumatic fever and rheumatic heart disease. 3rd ed. Darwin: Menzies School of Health Research, 2020 (3.2 Edition, March 2022). https://www.rhdaustralia.org.au/system/files/fileuploads/arf_rhd_guidelines_3.2_edition_march_2022.pdf (viewed Apr 2024).

8. Zühlke LJ, Steer AC. Estimates of the global burden of rheumatic heart disease. *Glob Heart* 2013; 8: 189-195.

9. Bennett J, Zhang J, Leung W, et al. Rising ethnic inequalities in acute rheumatic fever and rheumatic heart disease, New Zealand, 2000–2018. *Emerg Infect Dis* 2021; 27: 36-46.

10. World Bank Group. World bank country and lending groups. Washington: World Bank Group, 2020. https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups (viewed Dec 2024).

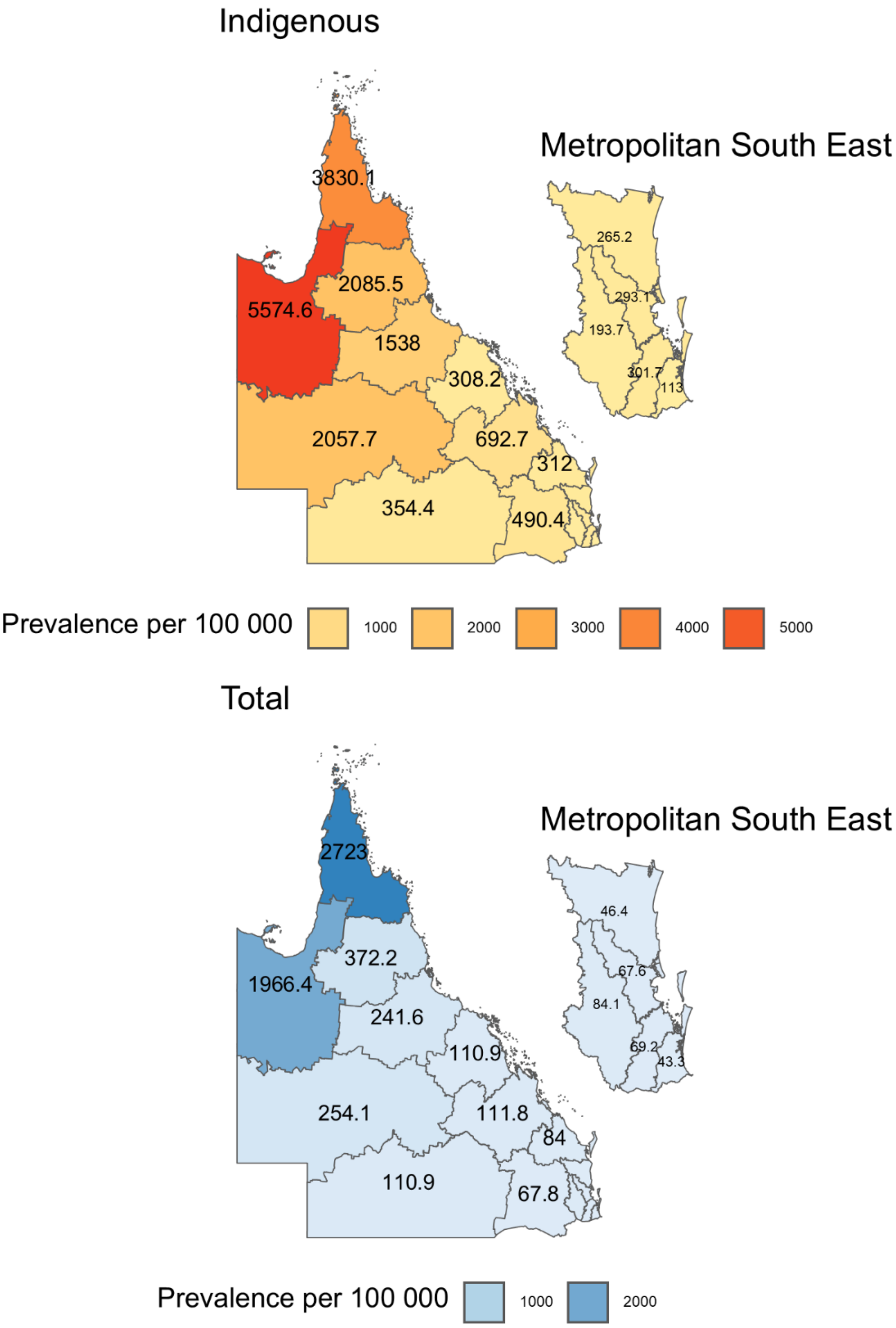

Figure 2. Mean age-standardised prevalence (< 55 years) of rheumatic heart disease in Indigenous (top) and total (bottom) populations of Queensland, by public Hospital and Health Service region, 2017–2021

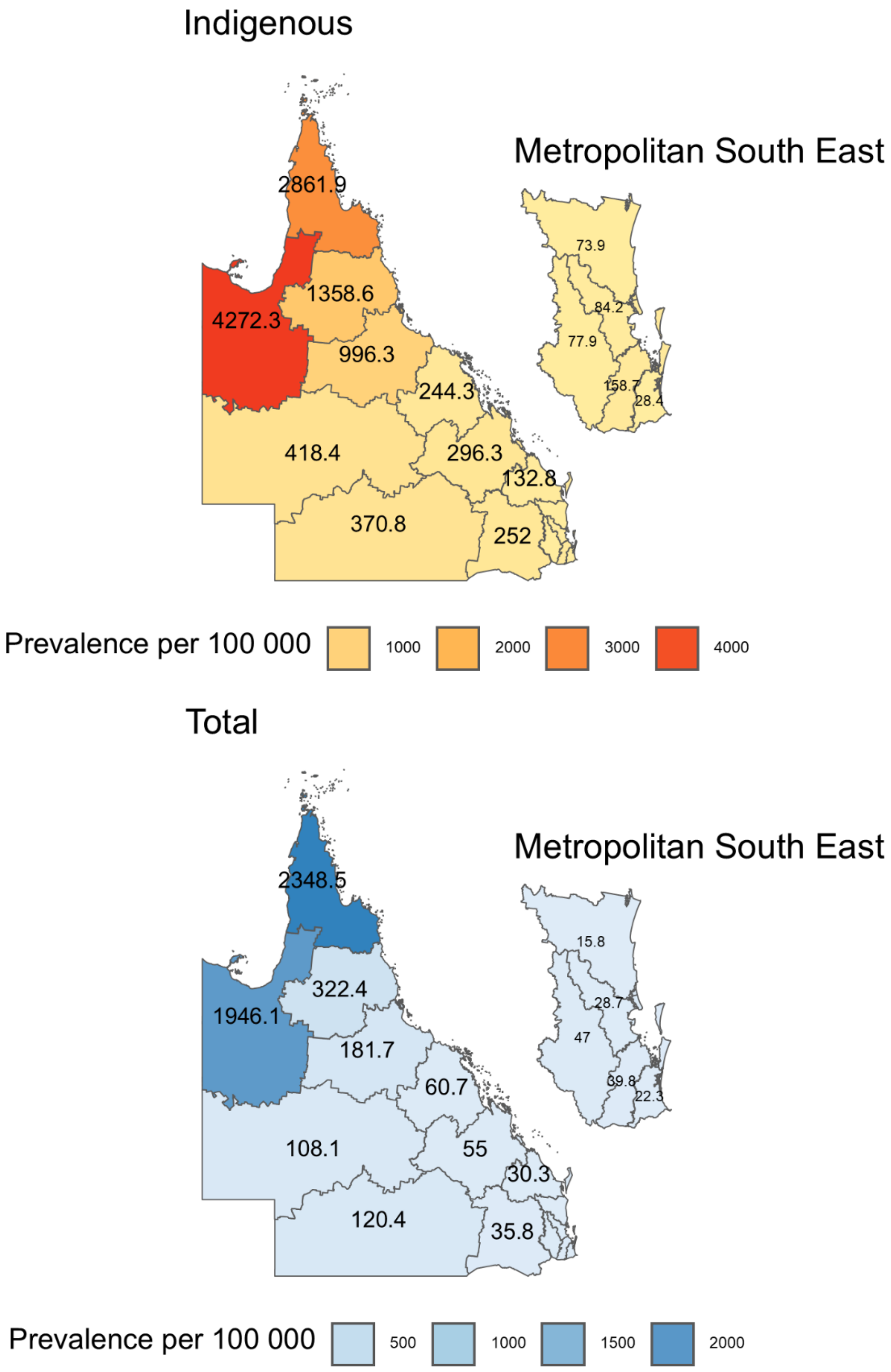

Figure 3. Mean age-standardised prevalence (< 25 years) of rheumatic heart disease in Indigenous (top) and total (bottom) populations of Queensland, by public Hospital and Health Service region, 2017–2021

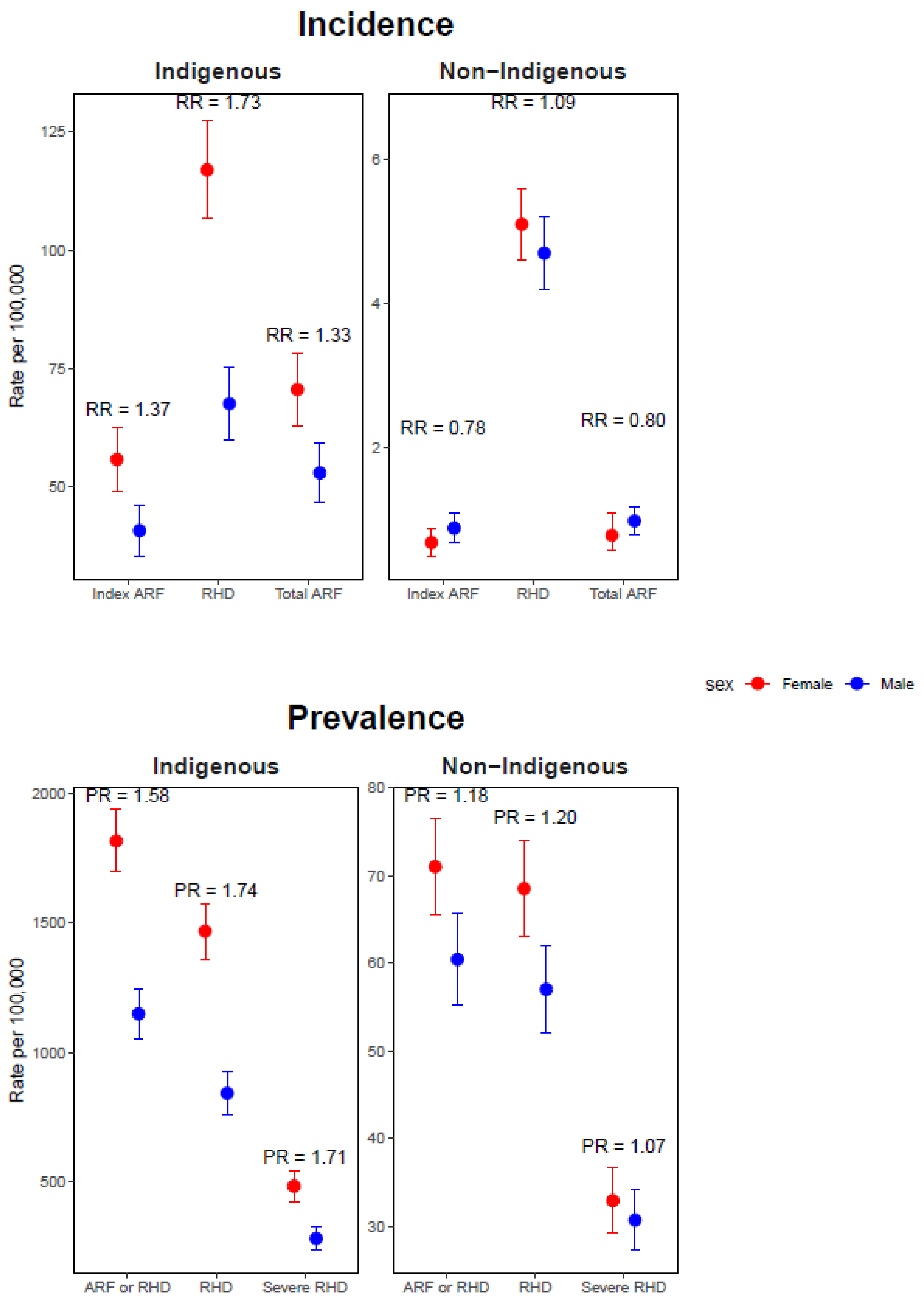

Figure 4. Sex differences in the age-standardised incidence and prevalence of acute rheumatic fever and rheumatic heart disease in Queensland (2017–2021), by Indigenous status.
RR is rate ratio; PR is prevalence ratio; RHD is rheumatic heart disease; ARF is acute rheumatic fever; ARF or RHD includes any person with a history of ARF or RHD; severe RHD includes any person with a record RHD and heart failure, a cardiac valvular intervention, or recorded as having severe RHD on the RHD Register; index ARF includes all first ever ARF episodes; total ARF includes all index ARF episodes plus all recurrent ARF episodes