BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (http://bmjopen.bmj.com).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

Searching for a Technology-Driven Acute Rheumatic Fever Test: the START study protocol

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-053720
Article Type:	Protocol
Date Submitted by the Author:	25-May-2021
Complete List of Authors:	Ralph, A; Charles Darwin University; Royal Darwin Hospital Webb, Rachel; Counties Manukau District Health Board; Starship Children's Hospital McGregor, Reuben; The University of Auckland Lassmann, Timo; Telethon Kids Institute Benothman, Rym; Telethon Kids Institute Yan, Jennifer; Menzies School of Health Research, Charles Darwin University; Royal Darwin Hospital Bennett, Julie; University of Otago Wilson, Nigel; Starship Children's Hospital Moreland, Nikki; The University of Auckland, School of Medical Sciences and Maurice Wilkins Centre Bosco, Anthony; Telethon Kids Institute, The University of Western Australia Broadhurst, David; Edith Cowan Barnett, Timothy; Telethon Kids Institute, Wesfarmers Centre for Vaccines and Infectious Diseases Remenyi, Bo; Menzies School of Health Research, Charles Darwin University, Global and Tropical Health Division; Royal Darwin Hospital, Paediatrics Mayo, Mark; 1. Menzies School of Health Research, Charles Darwin University Pearson, Glenn; Telethon Kids Institute, Aboriginal Health Institute Leadership Team Kollmann, Tobias; Telethon Kids Institute Carapetis, Jonathan; Telethon Kids Institute
Keywords:	Paediatric rheumatology < PAEDIATRICS, Valvular heart disease < CARDIOLOGY, Immunology < TROPICAL MEDICINE, Diagnostic microbiology < INFECTIOUS DISEASES

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

For peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Searching for a Technology-Driven Acute Rheumatic Fever Test: the START study protocol

Authors

Anna P. Ralph[1, 2*]
Rachel Webb[3,4,5]
Nicole J. Moreland[6]
Reuben McGregor[6]
Anthony Bosco[7]
David Broadhurst[8]
Timo Lassmann[7]
Timothy Barnett[9]
Rym Benothman[7]
Jennifer Yan[1, 2]
Bo Remenyi[1, 2]
Julie Bennett[10]
Nigel Wilson[4]
Mark Mayo[1]
Glenn Pearson[7]
Tobias Kollmann[7]
Jonathan R. Carapetis[7,11]

1. Menzies School of Health Research, Charles Darwin University, Darwin, Northern Territory, Australia
2. Royal Darwin Hospital, Darwin, Northern Territory, Australia
3. KidzFirst Hospital, Counties Manukau District Health Board, Auckland, New Zealand
4. Starship Children's Hospital, Auckland, New Zealand
5. Department of Paediatrics; Child and youth Health, University of Auckland, Auckland, New Zealand
6. School of Medical Sciences and Maurice Wilkins Centre, The University of Auckland, Auckland New Zealand
7. Telethon Kids Institute, Perth, Western Australia; Perth Children's Hospital, Perth, WA.
8. Centre for Integrative Metabolomics & Computational Biology, Edith Cowan University, Perth, WA.
9. Wesfarmers Centre for Vaccines and Infectious Diseases, Telethon Kids Institute, University of Western Australia, Nedlands, WA, Australia.
10. University of Otago, Wellington, New Zealand
11. University of Western Australia, Perth, Western Australia.

*corresponding author anna.ralph@menzies.edu.au

Word count: Abstract 295; Manuscript 4016

Keywords: Acute rheumatic fever; biomarker; transcriptomics; rheumatic heart disease

For peer review only

Abstract

Introduction

The absence of a diagnostic test for Acute Rheumatic Fever (ARF) is a major impediment om managing this serious childhood condition. ARF is an autoimmune condition triggered by infection with Group A *Streptococcus*. It is the precursor to rheumatic heart disease (RHD), a leading cause of health inequity and premature mortality for Indigenous peoples of Australia, New Zealand and internationally.

Methods and Analysis

'Searching for a Technology-Driven Acute Rheumatic Fever Test' (START) is a biomarker discovery study that aims to detect and test a biomarker signature that distinguishes ARF cases from non-ARF, and use systems biology and serology to better understand ARF pathogenesis. Eligible participants with ARF diagnosed by an expert clinical panel according to the 2015 Revised Jones Criteria, aged 5-30 years, will be recruited from three hospitals in Australia and New Zealand. Age, sex and ethnicity-matched individuals who are healthy or have non-ARF acute diagnoses or RHD, will be recruited as controls. In the discovery cohort, blood samples collected at baseline, and during convalescence in a subset, will be interrogated by comprehensive profiling to generate possible diagnostic biomarker signatures. A biomarker validation cohort will subsequently be used to test promising combinations of biomarkers. By defining the first biomarker signatures able to discriminate between ARF and other clinical conditions, the START study has the potential to transform the approach to ARF diagnosis and RHD prevention.

Ethics and dissemination The study has approval from the Northern Territory Department of Health and Menzies School of Health Research ethics committee and the New Zealand Health and Disability Ethics Committee. It will be conducted according to ethical standards for research involving Indigenous Australians and New Zealand Māori and Pacific Peoples. Indigenous investigators and governance groups will provide oversight of study processes and advise on cultural matters.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

33 **Article summary**

34

35 **Strengths and limitations of this study.**

36 • START addresses a critically important health problem, acute rheumatic fever, which
37 disproportionately affects global Indigenous populations and for which there is no
38 diagnostic test nor effective treatment able to limit progression to rheumatic heart
39 disease

40 • The large sample size recruited across international sites and the use of unbiased,
41 comprehensive immune profiling will maximise the likelihood of being able to define
42 a biomarker signature which can discriminate between ARF and other clinical
43 conditions

44 • The most robust 'gold standard' rheumatic fever diagnostic process currently available
45 will be used, comprising the revised 2015 Jones Criteria applied by an expert clinical
46 panel

47 • Culturally safe processes with Indigenous governance guide the development and
48 conduct of the study

49 • An inherent limitation is heterogeneity among people with ARF with regards to
50 demographic factors, ARF type (first or recurrent), diagnostic category (possible,
51 probable, definite), clinical phenotype (carditis, chorea, arthritis, skin and soft tissue)
52 and timing in relation to ARF onset (variable time taken for healthcare to be reached
53 after disease onset, consent gained and blood samples collected)

54

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Introduction

The absence of a diagnostic test for acute rheumatic fever (ARF) is a major impediment to management and epidemiological understanding of this serious childhood condition. ARF is one of the best examples of human autoimmunity caused by an infection, with Group A *Streptococcus* (GAS) the identified trigger. It is the precursor to rheumatic heart disease (RHD), a leading cause of health inequity and premature mortality for Indigenous peoples of Australia, New Zealand and internationally[1]. Prompt diagnosis of ARF allows timely commencement of secondary prophylaxis with long-acting penicillin to prevent repeated GAS infections that drive ARF recurrences. ARF recurrences are in turn the main mechanism responsible for development of RHD. Under-diagnosis of ARF appears to be a major contributor to the high rates of RHD in Aboriginal communities in Australia's Northern Territory (NT), where 76% of patients with RHD lack a prior ARF diagnosis[2]. Similarly a cohort study in New Zealand (NZ) found the majority of RHD cases (65%) had never been previously hospitalised with ARF[3]. Similar findings are reported in African countries.[4]

The Jones Criteria (Table 1) have been the diagnostic tool for ARF for nearly 80 years[5]. This is a regularly-reviewed clinical algorithm, now available in a user-friendly smart device application ('app')[6], most recently modified to optimise sensitivity for use in high-risk settings and optimise specificity for low-risk populations[7]. Despite these revisions, ARF diagnosis remains challenging and subjective. Diagnostic biomarkers for ARF that improve the performance of these criteria or provide a definitive diagnostic test would be a major advance in improving ARF detection, and thereby, RHD prevention.[8] Accurate case ascertainment is also needed to monitor progress towards disease control targets.

A further barrier to ARF management is the absence of an identified effective immunomodulatory treatment to alter long-term cardiac outcomes. Non-steroidal anti-inflammatory drugs alleviate joint symptoms but do not alter the long-term disease trajectory. Corticosteroids appear ineffective in limiting development or severity of RHD[9], but in some settings are prescribed as a treatment of last resort for severe carditis[10,11]. Hydroxychloroquine use has recently been reported in two cases of ARF with carditis, based on plausible *in vitro* data[12] but efficacy is as yet unknown[13]. Improved knowledge of ARF immune pathogenesis is needed to inform potential immune modulatory strategies targeted to the immune pathology.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Compounding complexity, different pathological processes may occur, accounting for the heterogeneous clinical phenotypes of ARF (Table 1, see major manifestations).

Most investigations into the immune basis of ARF have used biased (directed) approaches, with limited utility. For example, elevated TNFα, IL-6 and IL-8 have been reported in patients with active ARF and RHD, but are unreliable disease markers[14-16]. In difficult-to-diagnose diseases, the use of a multiplicity of biomarkers ('multivariate') to identify the presence or progression of disease has proven useful. Using micro-arrays, characteristic gene expression signatures have been reported in white blood cells from patients with Kawasaki disease (another inflammatory cardiac condition), lupus and tuberculosis demonstrating that multivariate expression profiling has potential to identify biomarker composite signatures as diagnostic tools[17-20]. Multivariate approaches to analysing biomarkers in sera or plasma have also shown promise. In a bead-based multiplex assay, a number of *M. tuberculosis*-reactive antibodies were successfully used to diagnose tuberculosis in non-human primates[21], and antibody glycosylation patterns can discriminate latent from active TB in humans[22]. Most importantly, our recent studies show proof-of-principle for immunopathogenesis studies to understand ARF, identifying a dysregulated cytokine axis using multivariate approaches including RNAseq and flow cytometry[12] as well as a linked IgG3-C4 response in early ARF with multiplex bead-based assays[23]. Serum- and plasma-borne biomarkers in addition to transcriptional profiling and circulating cellular components thus constitute a strong and accessible means of diagnosis once useful biomarkers have been identified.

Our hypothesis is that unbiased, multivariate analyses hold the greatest potential for identifying meaningful, translatable information on the immunopathogenesis of ARF. The aims of the 'Searching for a Technology-Driven Acute Rheumatic Fever Test' (START) study are to determine whether this approach can identify a biomarker signature that accurately classifies ARF diagnoses according to a defined 'gold standard' diagnosis (the 2015 revised Jones Criteria applied by a panel of clinical experts), and to better understand ARF immune pathogenesis (Table 2). Specifically, we aim to develop, in a discovery cohort and validate in a second cohort, a profile of metabolic and immunological biomarkers that distinguishes ARF cases from a range of non-ARF conditions and healthy controls.

Methods

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Design

START is a cross-national, prospective, observational study with a quasi case-control design. We will prospectively recruit children and young adults with ARF, and people matched on age, sex and ethnicity who have eligible control diagnoses, in Australia and NZ. The study will comprise a discovery cohort (up to 65 definite cases and 65 controls) and a validation cohort (up to 25 definite cases and 65 controls). Probable and possible ARF diagnoses will then be tested independently (Figure 1). Eligible control conditions are shown in Table 2. Participants will have blood collected for comprehensive immune profiling.

Ethics and dissemination

The study has approval from the NT Department of Health and Menzies School of Health Research ethics committee (18/3126) and the NZ Health and Disability Ethics Committee (18/CEN/197). It will be conducted according to Good Clinical Practice, Good Laboratory Practice and the National Health and Medical Research Council and ethical standards for research involving Indigenous Australians and NZ Māori and Pacific Peoples.

Knowledge diseemination will be through academic channels and community discussion forums. The study team is well placed to foster translation of findings into practical tools, namely, an improved diagnostic, and immunomodulatory therapy.

Study sites

Participating hospitals across the two international sites: Royal Darwin Hospital, Darwin, NT, Australia, and Starship and Middlemore Hospitals, Auckland, NZ. Clinical guidelines at all three study sites require that all people with suspected ARF are hospitalised for diagnostic workup and initiation of management[10,24]. Royal Darwin Hospital is the major tertiary referral hospital in Australia's NT. Thirty percent of the NT population is Indigenous (Aboriginal or Torres Strait Islander)[25]. ARF occurs at rates >300/100,000 total Aboriginal population in highest-burden communities in the NT[26]. Starship and Middlemore Hospitals serve a combined urban population of approximately 1.1 million people in the Auckland region. Māori and Pacific peoples make up approximately 27% of the population of greater Auckland region, where the incidence of ARF among Māori and Pacific children is approximately 33/100,000 and 93/100,000, respectively, for those 5–15 years old[24,27].

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Study Participants

Inclusion criteria for ARF cases are: individuals with suspected or confirmed ARF (the vast majority of whom are expected to identify as being of Aboriginal, Torres Strait Islander, Māori or Pacific ethnicity) aged 5-30 years (Darwin) or 2-20 years (NZ). Written, informed consent will be obtained from guardians or the participant themself if aged ≥16 years in NZ or ≥18 years in Australia. Differences between sites are consistent with local epiodemiology and local ethical recommendations.

Exclusion criteria comprise individuals with severe anaemia in whom collection of the blood volume required for the study collection would be relatively contraindicated; profound immunosuppression other than corticosteroids for two months prior to study entry, or as appropriate depending on the half-life of the immunosuppressive agent; and unstable social situation precluding discussion about consent for research.

Controls are recruited into one of four groups: A. Non-ARF Streptococcal infections; B. Other non-ARF acute inflammatory presentations; C. Established RHD without an intercurrent acute diagnosis; and D. Healthy individuals (Table 2). All must meet the above-stated age, ethnicity and consent inclusion criteria and none of the exclusion criteria. Controls will be matched at data analysis stage to cases by site on sex, age within 10 years, and ethnicity. Individuals in Groups A, B and D will have an echocardiogram to exclude underlying RHD.

Cultural safety

Cultural safety[28] is a core underpinning principle of the study. In NZ, the research is conducted in accordance with core Te Tiriti o Waitangi principles affording protection, participation and partnership for Māori participants. Research staff enrolling participants will engage approriate family members (in NZ: whānau) and respect differences in decision-making processes. Interpreters and translated participant materials will be used where study participants and their families primarily speak another language. In Darwin, patient information has been recorded in six Aboriginal languages. In NZ, participant materials have been translated into Te Reo Māori, Samoan and Tongan.

Patient and Public Involvement

The Indigenous investigators and collaborators at each site form governance groups to provide oversight of study design, conduct, reporting and dissemination. Patient and public

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

1
2
3 190 involvement, drawing on our established consumer networks, will support knowledge
4 191 dissemination.
5
6 192
7
8 193 **Enrolment**
9
10 194 It will be made clear that participation is voluntary and any benefits of participation are to
11
12 195 society through knowledge advancement, rather than to the individual. Participants will be
13 196 advised that wherever possible, blood collection for the study will coincide with collection
14
15 197 done for clinical purposes, to avoid extra venepunctures. If a participant or their
16
17 198 parent/guardian withdraws consent, no further samples will be collected unless a new consent
18 199 form is signed. Participants are asked whether their blood may be stored for future use.
19
20 200
21
22 201 People with suspected ARF will be enrolled as early as possible during admission, prior to a
23
24 202 final diagnosis being assigned, to ensure that an acute sample is collected. A final diagnosis
25 203 will be assigned once all required diagnostic information has become available. Participants in
26
27 204 Group C (RHD) may additionally be recruited through outpatient clinics. Group D (healthy)
28
29 205 will be sought from among family members and friends of other enrolled participants, or
30 206 healthy members of the community in Darwin.
31
32 207
33
34 208 **Clinical data collection**
35
36 209 After assigning a sequential study code, clinical details will be recorded on a paper Case Report
37 210 Form (CRF) then entered into an electronic database (Medrio™ Electronic Data Capture
38
39 211 System), including: date of illness onset, date of blood sample collection, demographics,
40
41 212 clinical presentation, Jones criteria, clinical laboratory results, medications received and
42 213 diagnosis assigned by the clinical treating team.
43
44 214
45
46 215 **Assignment of diagnosis**
47
48 216 Patients presenting with suspected ARF may have a final diagnosis of definite ARF, probable
49 217 ARF, possible ARF or a non-ARF condition. *Probable ARF* is defined in Australian guidelines
50
51 218 as an acute presentation not fulfilling criteria, missing a major or minor criterion or lacking
52
53 219 evidence of preceding streptococcal infection, but ARF is still considered the most likely
54 220 diagnosis. *Possible ARF* applies to the same presentation type, but where ARF is considered
55
56 221 uncertain but cannot be ruled out[10]. For this study, participants ultimately diagnosed with a non-
57
58 222 ARF condition will be allocated by the clinical panel as group A or B controls.
59
60 223

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

For peer review only

Management will be directed by treating clinicians. Clinical diagnoses will be assigned according to local Australian and NZ Diagnostic Criteria by a panel of ≥2 study clinicians expert in ARF. The decision on diagnostic category requires clinical judgement, after evaluating all aspects of the case. The Jones criteria app[6] which provides a result of definite, probable or possible ARF or not ARF depending on information provided to the algorithm, will be used by the panel at their discretion. The diagnosis of carditis will be made if rheumatic valvulitis is evident on echocardiogram as per Jones criteria. The diagnosis of rheumatic heart disease will be made in accordance with features described by the World Heart Federation[29]. The 2015 revised Jones Criteria[7,30] will be applied as the gold standard diagnostic for definite ARF. This represents a departure from normal practice in NZ where local diagnostic criteria differ: polyarthralgia is not considered a major criterion in NZ[24]. In addition to the Jones criteria, comprehensive clinical and laboratory information such as results of synovial fluid testing, serology for alternative infectious aetiologies, autoantibodies, radiological findings, diagnosis assigned by the treating team, family history and local epidemiology, will be considered in adjudicating the diagnosis. ARF type will be further specified as initial episode or recurrence. Where the adjudication panel's diagnosis differs from that of the treating team with implications for management, this will be communicated to the treating specialist. The clinical panel will also assign final diagnoses for controls.

Blood collection

Blood will be collected at baseline, and for patients with ARF, on follow-up occasions during the convalescent period as able e.g. during prolonged hospitalisation or at a later outpatient appointment (Table 3). A safe maximum volume of 28.5 mL depending on age will be obtained in Australia[31], and 14.5 mL in NZ. Baseline samples will be collected as soon as possible during the acute presentation, timed to coincide with routine blood testing wherever possible. Convalescent samples will be used to determine persistence and decay of immunopathological signatures post ARF. After collection, samples will be transported immediately to research laboratories at Menzies School of Health Research (Australia) or the University of Auckland (NZ) for centrifugation, serum/plasma separation and viable peripheral blood mononuclear cell (PBMC) preparation. Timely freezing and storage of serum/plasma aliquots, PAXgene tubes at -80°C and PBMCs in gas-phase liquid nitrogen. Samples will be shipped periodically to relevant laboratories.

Sample size

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

It is not possible to predict *a priori* the combined discriminatory ability of independently measured factors. This study has elected to enrol up to 120 ARF cases (comprising approximately 90 definite ARF, 15 probable, 15 possible) and 130 controls (30 each of control groups A, B and C, and 40 control group D; Figure 1 and Table 2). These numbers are feasible and should enable characterisation of different ARF clinical phenotypes (carditis, arthritis, chorea, skin/soft tissue manifestations or a combination of these) and phases (first or recurrent episodes), acknowledging heterogeneity among participants (ethnicity, age, sex; on corticosteroids or not; at different stages of illness).

In the multivariate approach it is expected that most, if not all, of the influential factors in the optimal model will be significant ($p<0.05$) in univariate testing. Assuming α of 5% and 95% power ($\beta = 0.05$) the minimum sample size is 26 cases. These calculations are only a guideline for removing clear false positives, as the standard procedure for producing robust screening models is to have a separate blinded analysis of a test subset of samples.We will utilise Discovery and Validation Cohorts, together with stratified bootstrap cross-validation to internally optimize the structural parameters in each model. We further used the standard inferential approach to sample size estimation in diagnostic test studies of biomedical informatics[32]. An effective multivariate predictor of ARF with positive predictive value of 0.8 (clinically useful), and a 95% confidence interval of +/- 0.1, would require 65 ARF cases in the discovery cohort.

The discovery cohort will therefore comprise a target of 65 definite ARF cases and 65 controls. The validation cohort will comprise a target of 25 definite ARF cases and 65 controls. Probable and possible diagnoses (~15 in each group) will be tested independently (Figure 1). Cases will be allocated to the discovery or validation cohorts using a computer-generated random selection of unique study identification numbers.

Laboratory methods

The overarching aim is to develop and validate a profile of host related biomarkers that distinguishes ARF cases from non-ARF conditions and healthy controls. This will be achieved through the following analyses on peripheral blood samples.

Immunophenotyping:

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

PBMC will be stained with labeled antibody panels to identify specific cell populations. Flow cytometry raw data will be analyzed manually using Flowjo software and via an automated gating platform as described[33,34]. Such high throughput, automated analysis of big flow cytometry data offers several advantages over manual gating, including increased throughput while increasing quality control (such as pre-processing removal of anomalous events via the flowCut algorithm), and identification of specific cell populations with up to 50-dimensional datasets[34-36]. Plasma samples will be analyzed utilizing multiplex cytokine assays to quantify cytokine and chemokine plasma concentrations, detailed elsewhere[33,37-39].

Metabolome analyses:

Untargeted metabolomic profiling (>1000 metabolites) will be performed on plasma samples using liquid chromatography coupled to high resolution mass-spectrometry (LC-HRMS). Data will be acquired using three modes of operation: reverse-phase/UPLC-MS/MS with positive ion mode electrospray ionization (ESI), reverse-phase/UPLC-MS/MS with negative ion mode ESI, and HILIC/UPLC-MS/MS with negative ion mode ESI. All identified metabolites will be annotated using appropriate orthogonal analytical techniques applied to the metabolite of interest against a chemical reference standard.

Blood Transcriptomics:

RNA will be extracted from stabilized whole blood samples (PAXgene tubes) as described[40,41], libraries prepared (TruSeq Stranded Total RNA with Ribo-Zero Globin reduction, Illumina) and NextGen sequencing undertaken (Illumina HiSeq2500, 50-bp single-end reads). Read alignment and gene-level quantification (counts) will be performed using Hisat[42]. Negative binomial models will be employed for differential expression analysis, with false discovery rate control for multiple testing[43]. The analyses will be adjusted for batch effects and variations in cellular composition, which will be estimated employing RUVSeq and CIBERSORT respectively[44,45]. Genes will be mapped to blood transcriptional modules to provide a systems-level view of the responses and reduce the dimensionality of the data[46].

CD4 T cell transcriptomic responses to GAS:

A two-phased approach will be used to examine CD4 T cell responses to GAS. Firstly, PBMCs from a subset of definite ARF cases and healthy controls will be cultured under a variety of conditions to identify optimal conditions for the second phase. Cells will be harvested at multiple time points post-stimulation (6h, 24h and 48h) with heat- and antibiotic-killed ARF-

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

associated GAS strains[12,47,48] and a selection of candidate ARF antigens. Total RNA from PBMCs cultured under each condition will be profiled by RNA-Seq (Illumina 100bp paired-end reads, 20M reads), and culture supernatants will be examined by Luminex (Bio-Plex 48-plex Pro Human Cytokine Screening Panel) to identify optimal conditions for maximal discrimination of responses between cases and controls. In the second phase, an expanded set of ARF cases and controls for which PBMC samples are available will be examined following culture under the identified, optimal conditions. Gene expression patterns and cytokine production will be profiled by RNA-Seq and Luminex, as described above.

Antibody analysis:

For unbiased investigation of autoantibodies, selected sera will be screened against planer protein microarrays comprised of over 42,000 protein fragments representing some 19,000 human proteins[49]. Protein fragments that are more significantly bound by autoantibodies in ARF compared to controls will be identified using a *P*-value of >0.05 and fold change of 2.0 as cut-off. A suspension bead array will then be designed[50] comprised of up to 380 potential autoantigens to assess autoantigen reactivity in all sera. The bead-antigens will be selected based on the planar array results, previously completed screens using HuProt high-content protein arrays (McGregor *et al.,* submitted) that contain over 15,000 autoantigens[51], and targets from the literature. Finally, candidate ARF autoantigens identified in the suspension bead-array will be orthogonally validated as individual antigens in ELISA or Luminex bead-based assays to determine sensitivity and specificity.

Bioinformatics and statistical analyses:

The data collected for each patient across the immune phenotyping, transcriptomic, proteomic and metabolomic technologies will be integrated computationally to identify patterns of covariance within and between designated clinical outcome groups. Convergence of signatures across diverse systems biology domains will provide independent functional validation. The use of "multi-omic" data integration techniques will also enable the possibility of deriving novel biological information that will not be revealed in a single dataset alone.

Data integration methods applying different computational strategies will be used to explore the complex covariance structure within and between the multi-omic data blocks. Initially, analysis will focus on each data block in isolation employing a combination of classical generalised linear modelling (GLM), multivariate projection regression models (PCA, PLS-

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

DA) and unsupervised cluster analysis. The multiple blocks will then be integrated into a single computational model to enable multi-omic functional mapping, which in turn will allow us to uncover the underlying biochemical mechanisms. Several methods for performing multi-block data integration will be investigated, from which a consensus model will be derived. These will include: protein−protein interaction networks using NetworkAnalyst[52], Multi-block Component Analysis[53], regularized Canonical Correlation Analysis[54], the Data Integration Analysis for Biomarker discovery using Latent cOmponent (DIABLO) framework[55] and Similarity Network Fusion[56].

This systems biology analysis should result in both a domain-specific and domain-integrative summary of biological phenomena that are associated with clinical outcome. Likely outcomes are a signature biological pathway associated with ARF as well as individual candidate markers that may provide the basis for the development of novel diagnostic tests. Furthermore, the data generated through this effort will allow crisply defined power calculations for future narrowly targeted assessment of potential biomarkers in clinical trials.

Discussion

We anticipate that study findings will provide the most comprehensive knowledge of the immunopathogenesis of ARF to date, be used as the basis for development of a diagnostic test and provide a pathway towards development of targeted immunomodulatory treatment(s). START commenced recruitment in Australia in November 2018 and in NZ in May 2019. Anticipated completion of recruitment is end-2021. Laboratory analyses will be batched and run together at completion of recruitment.

Both under-and over-diagnosis of ARF pose major challenges to individuals and health systems.[10] Whether the use of the 'probable' and 'possible' ARF diagnostic categories is reducing underdiagnosis or contributing to overdiagnosis is currently unknown. Should findings from this study successfully identify a discriminatory biomarker profile differentiating definite ARF from non-ARF, this will provide a mechanism for accurate diagnosis to guide appropriate management. This will be a critically important advance in the diagnosis and management of ARF globally.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

1
2
3 393 A minority of any population are at risk of ARF after exposure to group A streptococci. The
4 394 estimated lifetime cumulative incidence of ARF was previously calculated at 5.7% in
5
6 395 Australia's Northern Territory[57]. This was considered likely to be an under-estimate, and
7
8 396 indeed more recent data highlight the burden of unrecognised ARF in Aboriginal communities
9
10 397 in the NT where repeated infection with GAS is ubiquitous from early childhood[2,58], suggesting
11
12 398 that the lifetime risk of ARF in these populations is higher than 6%. In US military camps, 2-
13 399 3% of recruits developed ARF after a single bout of streptococcal pharyngitis[59].
14
15 400
16
17 401 Previous efforts to develop ARF diagnostic tests and elucidate the immunopathogenesis have
18 402 laid foundations for the START study design and analysis. In 2018, Kim *et al.* analysed
19
20 403 responses to GAS challenge of PBMC from ARF patients at Royal Darwin Hospital using
21
22 404 multiplex cytokine array, flow cytometric analysis, and global gene expression analysis by
23
24 405 RNA sequencing[12]. They identified a dysregulated interleukin-1β-granulocyte- macrophage
25
26 406 colony-stimulating factor (GM-CSF) cytokine axis in PBMCs from ARF patients, and the
27
28 407 potential to suppress this response by hydroxychloroquine. The authors proposed that
29 408 hydroxychloroquine, an immunomodulatory agent, could be repurposed to reduce the risk or
30
31 409 severity of RHD after ARF[12]. Clinical use of hydroxychloroquine, now reported as safe in two
32
33 410 cases[13], requires further investigation. A study using sera from patients with ARF in NZ
34
35 411 combined multiplex bead-based assays and systems immunology data analysis to identify a
36
37 412 linked IgG3-C4 response that may have utility as a clinical biomarker in early ARF[23]. However,
38
39 413 small sample sizes have been a limiting factor. Key strengths of the current study include the
40 414 larger sample size, multicentre enrolment across different countries to maximise relevance, and
41
42 415 use of the latest laboratory approaches.
43
44 416
45
46 417 Limitations include that capture of eligible participants early in disease is challenging, due to
47
48 418 the variable time taken for participants to be hospitalised after disease onset, to identify and
49
50 419 gain consent from participants, and obtain blood samples. At the Australian site, the majority
51 420 of eligible participants will be Aboriginal children living in remote communities, often
52
53 421 hundreds of kilometres distant from the enrolment site. Although Auckland is a large urban
54
55 422 city inequitable healthcare access is experienced by many Māori and Pacific people living in
56
57 423 socio-economic deprivation and this can lead to delayed hospitalisation for many children
58
59 424 presenting with ARF. Given these inherent delays, a proportion of baseline blood samples will
60 425 be collected after the peak inflammatory phase has passed. Another limitation is heterogeneity
426 among people with ARF with regards to ARF type (first or recurrent), diagnostic category

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

For peer review only

(possible, probable, definite), clinical phenotype (carditis, chorea, arthritis etc) and demographic variations; however, we limited enrolment to people below 30 years and only of Aboriginal, Torres Strait Islander, Māori or Pacific ethnicity to minimize heterogeneity. Although a distinguishing biomarker signature may be identified, the ability to readily translate study findings into a feasible diagnostic able to be used in health service laboratories outside a research environment, is uncertain.

Successful completion of this study will considerably improve knowledge of ARF immunology. Innovative strategies to improve the clinical management of ARF are core components of the overall suite of activities required to achieve the goal RHD elimination[60].

For peer review only

Acknowledgements: We are very grateful to Caroline Motteram for project management, Yolanda Hernandez, Laura Francis and Vincent Mithen for recruiting participants in Darwin and Shirley Lawrence, Rhonda Holloway, Renee Clark, Alicia Stanley, and Roxanne Buchanan for recruiting participants in New Zealand. We thank Alex Kaethner and NT Cardiac for facilitating echocardiography. Vanessa Rigas, Erin Gargan, Gabriella Minigo and Michael Serralha are gratefully acknowledged for laboratory support; Kyran Kuzio, Tasha Cole, Alana Whitcombe, Lauren Carlton and Timothy Ho for laboratory assistance; Amy Baker and Rebecca Trowman for program and financial management.

Authors' contributions: JRC, APR, RW, NJM, AB, DB, TB, MM, GP, TK conceived and designed the study. NJM, RMG, AB, DB,TL,TB, RB, JB, TK will perform the assays and analyse the data. GP and MM provide project governance. APR, RW, JY, BR, NW and JRC provided clinical care, patient recruitment, assessment and clinical review panel participation. APR create the first draft and managed the edits to the manuscript. All authors reviewed and approved the study protocol.

Funding statement: The study is funded by the Australian National Health and Medical Research Council (NHMRC) grant #1147531. APR is supported by NHMRC fellowship #1142011. TCB is supported by the Western Australia Department of Health (#not applicable) and a Career Development Fellowship from Improving Health Outcomes in the Tropical North (NHMRC #1131932)

Competing interests statement. The authors have no conflicts of interest to declare.

Data statement: Not applicable. No data have been presented in this protocol paper.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Tables

Table 1: Revised Jones Criteria

Legend: Reproduced with permission from The 2020 Australian guideline for prevention, diagnosis and management of acute rheumatic fever and rheumatic heart disease (3rd edition)[2]

	High-risk groups†	Low-risk groups
Definite initial episode of ARF	2 major manifestations + evidence of preceding GAS infection, OR 1 major + 2 minor manifestations + evidence of preceding GAS infection‡	
Definite recurrent § episode of ARF in a patient with a documented history of ARF or RHD	2 major manifestations + evidence of preceding GAS infection, OR 1 major + 2 minor manifestations + evidence of preceding GAS infection‡, OR 3 minor manifestations + evidence of a preceding GAS infection‡	
Probable or possible ARF (first episode or recurrence§)	A clinical presentation in which ARF is considered a likely diagnosis but falls short in meeting the criteria by either: • one major or one minor manifestation, OR • no evidence of preceding GAS infection (streptococcal titres within normal limits or titres not measured) Such cases should be further categorised according to the level of confidence with which the diagnosis is made: • Probable ARF (previously termed 'probable: highly suspected') • Possible ARF (previously termed 'probable: uncertain')	
Major manifestations	Carditis (including subclinical evidence of rheumatic valvulitis on echocardiogram) Polyarthritis¶ or aseptic monoarthritis or polyarthralgia Sydenham chorea†† Erythema marginatum ‡‡ Subcutaneous nodules	Carditis (including subclinical evidence of rheumatic valvulitis on echocardiogram) Polyarthritis¶ Sydenham chorea†† Erythema marginatum‡‡ Subcutaneous nodules
Minor Manifestations	Fever §§ ≥38°C Monoarthralgia ¶¶ ESR ≥30 mm/h or CRP ≥30 mg/L Prolonged P-R interval on ECG†††	Fever ≥38.5°C Polyarthralgia or aseptic monoarthritis¶¶ ESR ≥60 mm/h or CRP ≥30 mg/L Prolonged P-R interval on ECG†††

† High-risk groups are those living in communities with high rates of ARF (incidence >30/100,000 per year in 5–14-year-olds) or RHD (all-age prevalence >2/1000). Aboriginal and Torres Strait Islander peoples living in rural or remote settings are known to be at high risk. Data are not available for other populations but Aboriginal and Torres Strait Islander peoples living in urban settings, Māori and Pacific Islanders, and potentially immigrants from developing countries, may also be at high risk.
‡ Elevated or rising antistreptolysin O or other streptococcal antibody, or a positive throat culture or rapid antigen or nucleic acid test for GAS infection.
§ Recurrent definite, probable or possible ARF requires a time period of more than 90 days after the onset of symptoms from the previous episode of definite, probable or possible ARF
¶ A definite history of arthritis is sufficient to satisfy this manifestation. Note that if polyarthritis is present as a major manifestation, polyarthralgia or aseptic monoarthritis cannot be considered an additional minor manifestation in the same person.
†† Chorea does not require other manifestations or evidence of preceding GAS infection, provided other causes of chorea are excluded.
‡‡ Care should be taken not to label other rashes, particularly non-specific viral exanthems, as erythema marginatum.
§§ In high-risk groups, fever can be considered a minor manifestation based on a reliable history (in the absence of documented temperature) if anti-inflammatory medication has already been administered.
¶¶ If polyarthritis is present as a major criterion, monoarthritis or arthralgia cannot be considered an additional minor manifestation.
††† If carditis is present as a major manifestation, a prolonged P-R interval cannot be considered an additional minor manifestation.
CRP, C-reactive protein; ECG, electrocardiogram; ESR, erythrocyte sedimentation rate.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 2: Control groups

Group A (n=50):	Non-ARF Streptococcal infections or toxin mediated condition or post-streptococcal condition Examples: - active infection with a proven or likely causative streptococcal organism (e.g. β- haemolytic streptococcus, α-haemolytic streptococcus, *S. pneumoniae* etc) from sterile or non-sterile site - scarlet fever - acute post-streptococcal glomerulonephritis AND - No evidence of definite RHD on echocardiogram
Group B (n=50):	Other acute condition Examples: - inflammatory conditions, including autoimmune inflammatory conditions (e.g. lupus with arthritis), pancreatitis etc - infective condition: confirmed arboviral infection (eg Ross River virus), septic arthritis, bacterial endocarditis, etc - trauma requiring hospitalisation or operative procedure AND - No evidence of definite RHD on echocardiogram
Group C (n=50):	RHD - Echocardiographically confirmed RHD with no active inflammation OR - past history of definite ARF but no current disease activity.
Group D (n=50):	Healthy - No intercurrent medical condition as determined by clinical questionnaire - Minor trauma allowed AND - No evidence of definite RHD on echocardiogram

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 3: Blood collection

	Australian recruitment site		
	Participant age		**Collection tube**
Testing procedures	**5-9 years**	**≥ 10 years**	
Transcriptional profiling	2.5mL	2.5mL	PaxGene tube
Peripheral blood mononuclear cells for flow cytometry, Metabolomics	4.0mL	22.0mL	Sodium Heparin
Antibody studies	4.0mL	4.0mL	Serum
TOTAL	10.5mL	28.5mL	
	New Zealand recruitment sites		
	Participant age		**Collection tube**
Testing procedures	**2-10 years**	**≥ 11 years**	
Transcriptional profiling	2.5mL	2.5mL	PaxGene tube
Peripheral blood mononuclear cells for flow cytometry, Metabolomics	4.0mL	8.0mL	Sodium Heparin
Antibody studies	4.0mL	4.0mL	Serum
TOTAL	10.5mL	14.5mL	

For peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

References

1. Ralph AP, Read C, Johnston V, et al. Improving delivery of secondary prophylaxis for rheumatic heart disease in remote Indigenous communities: study protocol for a stepped-wedge randomised trial. *Trials* 2016;17(1):51. doi: 10.1186/s13063-016-1166-y
2. Hardie K, Ralph AP, de Dassel JL. RHD elimination: action needed beyond secondary prophylaxis. *Aust N Z J Public Health* 2020;44(5):427. doi: 10.1111/1753-6405.13002 [published Online First: 2020/06/09]
3. Oliver J, Robertson O, Zhang J, et al. Progression of Acute Rheumatic Fever to Recurrence, Rheumatic Heart Disease, and Death in New Zealand Children and Youth: A Cohort Study. *Heart, Lung and Circulation* 2019;28
4. Zuhlke L, Engel ME, Karthikeyan G, et al. Characteristics, complications, and gaps in evidence-based interventions in rheumatic heart disease: the Global Rheumatic Heart Disease Registry (the REMEDY study). *European heart journal* 2015;36(18):1115-22a. doi: 10.1093/eurheartj/ehu449
5. Jones TD. The diagnosis of rheumatic fever. *J Am Med Assoc* 1944;126:481-84.
6. RHDAustralia. iPhone and Android Apps. http://www.rhdaustralia.org.au/apps 2020 [14 August 2014].
7. Gewitz MH, Baltimore RS, Tani LY, et al. Revision of the Jones Criteria for the diagnosis of acute rheumatic fever in the era of Doppler echocardiography: a scientific statement from the American Heart Association. *Circulation* 2015;131(20):1806-18. doi: 10.1161/CIR.0000000000000205
8. McMillan DJ, Rafeek RAM, Norton RE, et al. In Search of the Holy Grail: A Specific Diagnostic Test for Rheumatic Fever. *Front Cardiovasc Med* 2021;8:674805. doi: 10.3389/fcvm.2021.674805
9. Cilliers A, Adler AJ, Saloojee H. Anti-inflammatory treatment for carditis in acute rheumatic fever. *Cochrane database of systematic reviews* 2015(5):CD003176. doi: 10.1002/14651858.CD003176.pub3
10. RHDAustralia (ARF/RHD writing group). The 2020 Australian guideline for prevention, diagnosis and management of acute rheumatic fever and rheumatic heart disease (3rd edition) https://www.rhdaustralia.org.au/arf-rhd-guideline. Menzies School of Health Research, 2020.
11. Dean GL, Edwards SG, Ives NJ, et al. Treatment of tuberculosis in HIV-infected persons in the era of highly active antiretroviral therapy. *Aids* 2002;16(1):75-83.
12. Kim ML, Martin WJ, Minigo G, et al. Dysregulated IL-1beta-GM-CSF Axis in Acute Rheumatic Fever That Is Limited by Hydroxychloroquine. *Circulation* 2018;138(23):2648-61. doi: 10.1161/CIRCULATIONAHA.118.033891 [published Online First: 2018/12/21]
13. Wilson NJ, Concannon A, Malcolm J, et al. The Treatment of Acute Rheumatic Fever: Novel Use of Hydroxychloroquine. *Pediatr Infect Dis J* 2020 doi: 10.1097/INF.0000000000002647 [published Online First: 2020/03/30]
14. M. Hafez M, EL-Morsy Z, EL-Shennawy F, et al. Susceptibility to Over Production of Cytokines in Acute Rheumatic Carditis and Their Role in the Pathogenesis. *Journal of Medical Sciences* 2002 doi: 10.3923/jms.2002.65.73
15. Yegin O, Coskun M, Ertug H. Cytokines in acute rheumatic fever. *Eur J Pediatr* 1997;156(1):25-9. doi: 10.1007/s004310050545 [published Online First: 1997/01/01]
16. Narin N, Kutukculer N, Narin F, et al. Anticardiolipin antibodies in acute rheumatic fever and chronic rheumatic heart disease: is there a significant association? *Clin Exp Rheumatol* 1996;14(5):567-9. [published Online First: 1996/09/01]

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

17. Wright VJ, Herberg JA, Kaforou M, et al. Diagnosis of Kawasaki Disease Using a Minimal Whole-Blood Gene Expression Signature. *JAMA Pediatr* 2018;172(10):e182293. doi: 10.1001/jamapediatrics.2018.2293 [published Online First: 2018/08/08]
18. Berry MP, Graham CM, McNab FW, et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* 2010;466(7309):973-7. doi: 10.1038/nature09247 [published Online First: 2010/08/21]
19. Bennett L, Palucka AK, Arce E, et al. Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. *J Exp Med* 2003;197(6):711-23. doi: 10.1084/jem.20021553 [published Online First: 2003/03/19]
20. Anderson ST, Kaforou M, Brent AJ, et al. Diagnosis of childhood tuberculosis and host RNA expression in Africa. *N Engl J Med* 2014;370(18):1712-23. doi: 10.1056/NEJMoa1303657
21. Khan IH, Ravindran R, Yee J, et al. Profiling antibodies to Mycobacterium tuberculosis by multiplex microbead suspension arrays for serodiagnosis of tuberculosis. *Clinical and vaccine immunology : CVI* 2008;15(3):433-8. doi: 10.1128/CVI.00354-07 [published Online First: 2007/12/14]
22. Lu LL, Chung AW, Rosebrock TR, et al. A Functional Role for Antibodies in Tuberculosis. *Cell* 2016;167(2):433-43 e14. doi: 10.1016/j.cell.2016.08.072 [published Online First: 2016/09/27]
23. Chung AW, Ho TK, Hanson-Manful P, et al. Systems immunology reveals a linked IgG3-C4 response in patients with acute rheumatic fever. *Immunol Cell Biol* 2020;98(1):12-21. doi: 10.1111/imcb.12298 [published Online First: 2019/11/20]
24. Heart Foundation of New Zealand. New Zealand Guidelines for Rheumatic Fever: Diagnosis, Management and Secondary Prevention of Acute Rheumatic Fever and Rheumatic Heart Disease: 2014 Update. https://www.heartfoundation.org.nz/resources/acute-rheumatic-fever-and-rheumatic-heart-disease-guideline, 2014.
25. Australian Bureau of Statistics. Estimates of Aboriginal and Torres Strait Islander Australians 2016 [Available from: https://www.abs.gov.au/ausstats/abs@.nsf/mf/3238.0.55.001 accessed 27 April 2020.
26. Francis J, R, , Gargan C, Remenyi B, et al. A cluster of acute rheumatic fever cases among Aboriginal Australians in a remote community with high baseline incidence. *Aust N Z J Public Health* 2019;4(3):288-93.
27. Institute of Environmental Science and Research Limited. Rheumatic fever report, January-December 2018. EpiSurv Science for Communities https://surv.esr.cri.nz/PDF_surveillance/RheumaticFever/Rheumaticfeverbi-annualreportJan-Dec2018.pdf. 2018
28. Curtis E, Jones R, Tipene-Leach D, et al. Why cultural safety rather than cultural competency is required to achieve health equity: a literature review and recommended definition. *International journal for equity in health* 2019;18(1):174.
29. Remenyi B, Wilson N, Steer A, et al. World Heart Federation criteria for echocardiographic diagnosis of rheumatic heart disease--an evidence-based guideline. *Nature reviews Cardiology* 2012;9(5):297-309. doi: 10.1038/nrcardio.2012.7
30. Beaton A, Carapetis J. The 2015 revision of the Jones criteria for the diagnosis of acute rheumatic fever: implications for practice in low-income and middle-income countries. *Heart Asia* 2015;7(2):7-11. doi: 10.1136/heartasia-2015-010648 [published Online First: 2015/01/01]
31. Seattle Children's Hospital. Maximum Allowable Blood Draw From Infants. Mayo Foundation for Medical Education and Research 2020 [28 April 2020]. Available from: https://seattlechildrenslab.testcatalog.org/show/1000721-1.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

32. Hajian-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. *J Biomed Inform* 2014;48:193-204. doi: 10.1016/j.jbi.2014.02.013 [published Online First: 2014/03/04]
33. Ben-Othman R, Cai B, Liu AC, et al. Systems Biology Methods Applied to Blood and Tissue for a Comprehensive Analysis of Immune Response to Hepatitis B Vaccine in Adults. *Frontiers in immunology* 2020;11:580373. doi: 10.3389/fimmu.2020.580373 [published Online First: 2020/12/01]
34. Rahim A, Meskas J, Drissler S, et al. High throughput automated analysis of big flow cytometry data. *Methods* 2018;134-135:164-76. doi: 10.1016/j.ymeth.2017.12.015 [published Online First: 2017/12/31]
35. Malek M, Taghiyar MJ, Chong L, et al. flowDensity: reproducing manual gating of flow cytometry data by automated density-based cell population identification. *Bioinformatics* 2015;31(4):606-7. doi: 10.1093/bioinformatics/btu677 [published Online First: 2014/11/08]
36. Ge Y, Sealfon SC. flowPeaks: a fast unsupervised clustering for flow cytometry data via K-means and density peak finding. *Bioinformatics* 2012;28(15):2052-8. doi: 10.1093/bioinformatics/bts300 [published Online First: 2012/05/19]
37. Amenyogbe N, Dimitriu P, Cho P, et al. Innate Immune Responses and Gut Microbiomes Distinguish HIV-Exposed from HIV-Unexposed Children in a Population-Specific Manner. *J Immunol* 2020;205(10):2618-28. doi: 10.4049/jimmunol.2000040 [published Online First: 2020/10/18]
38. Amenyogbe N, Dimitriu P, Smolen KK, et al. Biogeography of the Relationship between the Child Gut Microbiome and Innate Immune System. *mBio* 2021;12(1) doi: 10.1128/mBio.03079-20 [published Online First: 2021/01/14]
39. Moncunill G, Dobano C, Gonzalez R, et al. Association of Maternal Factors and HIV Infection With Innate Cytokine Responses of Delivering Mothers and Newborns in Mozambique. *Front Microbiol* 2020;11:1452. doi: 10.3389/fmicb.2020.01452 [published Online First: 2020/08/09]
40. Lee AH, Shannon CP, Amenyogbe N, et al. Dynamic molecular changes during the first week of human life follow a robust developmental trajectory. *Nat Commun* 2019;10(1):1092. doi: 10.1038/s41467-019-08794-x [published Online First: 2019/03/14]
41. Idoko OT, Smolen KK, Wariri O, et al. Clinical Protocol for a Longitudinal Cohort Study Employing Systems Biology to Identify Markers of Vaccine Immunogenicity in Newborn Infants in The Gambia and Papua New Guinea. *Front Pediatr* 2020;8:197. doi: 10.3389/fped.2020.00197 [published Online First: 2020/05/20]
42. Pertea M, Kim D, Pertea GM, et al. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc* 2016;11(9):1650-67. doi: 10.1038/nprot.2016.095 [published Online First: 2016/08/26]
43. Anders S, McCarthy DJ, Chen Y, et al. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nat Protoc* 2013;8(9):1765-86. doi: 10.1038/nprot.2013.099 [published Online First: 2013/08/27]
44. Risso D, Ngai J, Speed TP, et al. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat Biotechnol* 2014;32(9):896-902. doi: 10.1038/nbt.2931 [published Online First: 2014/08/26]
45. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* 2015;12(5):453-7. doi: 10.1038/nmeth.3337 [published Online First: 2015/03/31]
46. Li S, Rouphael N, Duraisingham S, et al. Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. *Nat Immunol* 2014;15(2):195-204. doi: 10.1038/ni.2789 [published Online First: 2013/12/18]

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

47. Smoot JC, Barbian KD, Van Gompel JJ, et al. Genome sequence and comparative microarray analysis of serotype M18 group A Streptococcus strains associated with acute rheumatic fever outbreaks. *Proceedings of the National Academy of Sciences of the United States of America* 2002;99(7):4668-73. doi: 10.1073/pnas.062526099 [published Online First: 2002/03/28]
48. Bryant PA, Smyth GK, Gooding T, et al. Susceptibility to acute rheumatic fever based on differential expression of genes involved in cytotoxicity, chemotaxis, and apoptosis. *Infect Immun* 2014;82(2):753-61. doi: 10.1128/IAI.01152-13 [published Online First: 2014/01/31]
49. Sjoberg R, Mattsson C, Andersson E, et al. Exploration of high-density protein microarrays for antibody validation and autoimmunity profiling. *N Biotechnol* 2016;33(5 Pt A):582-92. doi: 10.1016/j.nbt.2015.09.002 [published Online First: 2015/09/30]
50. Just D, Manberg A, Mitsios N, et al. Exploring autoantibody signatures in brain tissue from patients with severe mental illness. *Transl Psychiatry* 2020;10(1):401. doi: 10.1038/s41398-020-01079-8 [published Online First: 2020/11/20]
51. Pan J, Song G, Chen D, et al. Identification of Serological Biomarkers for Early Diagnosis of Lung Cancer Using a Protein Array-Based Approach. *Mol Cell Proteomics* 2017;16(12):2069-78. doi: 10.1074/mcp.RA117.000212 [published Online First: 2017/10/13]
52. Xia J, Gill EE, Hancock RE. NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. *Nat Protoc* 2015;10(6):823-44. doi: 10.1038/nprot.2015.052 [published Online First: 2015/05/08]
53. De Roover K, Ceulemans E, Timmerman ME. How to perform multiblock component analysis in practice. *Behav Res Methods* 2012;44(1):41-56. doi: 10.3758/s13428-011-0129-1 [published Online First: 2011/07/13]
54. Wilms I, Croux C. Robust sparse canonical correlation analysis. *BMC Syst Biol* 2016;10(1):72. doi: 10.1186/s12918-016-0317-9 [published Online First: 2016/08/16]
55. Singh A, Shannon CP, Gautier B, et al. DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays. *Bioinformatics* 2019;35(17):3055-62. doi: 10.1093/bioinformatics/bty1054 [published Online First: 2019/01/19]
56. Wang B, Mezlini AM, Demir F, et al. Similarity network fusion for aggregating data types on a genomic scale. *Nat Methods* 2014;11(3):333-7. doi: 10.1038/nmeth.2810 [published Online First: 2014/01/28]
57. Carapetis JR, Currie BJ, Mathews JD. Cumulative incidence of rheumatic fever in an endemic region: a guide to the susceptibility of the population? *Epidemiol Infect* 2000;124(2):239-44. [published Online First: 2000/05/17]
58. Francis JR, Fairhurst H, Hardefeldt H, et al. Hyperendemic rheumatic heart disease in a remote Australian town identified by echocardiographic screening. *Med J Aust* 2020;213(3):118-23. doi: 10.5694/mja2.50682 [published Online First: 2020/07/08]
59. Rammelkamp CH, Wannamaker LW, Denny FW. The Epidemiology and Prevention of Rheumatic Fever. *Bull N Y Acad Med* 1952;28(5):321-34. [published Online First: 1952/05/01]
60. Wyber R, Noonan K, Halkon C, et al. Ending rheumatic heart disease in Australia: the evidence for a new approach. *Med J Aust* 2020;213 Suppl 10:S3-S31. doi: 10.5694/mja2.50853 [published Online First: 2020/11/16]

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

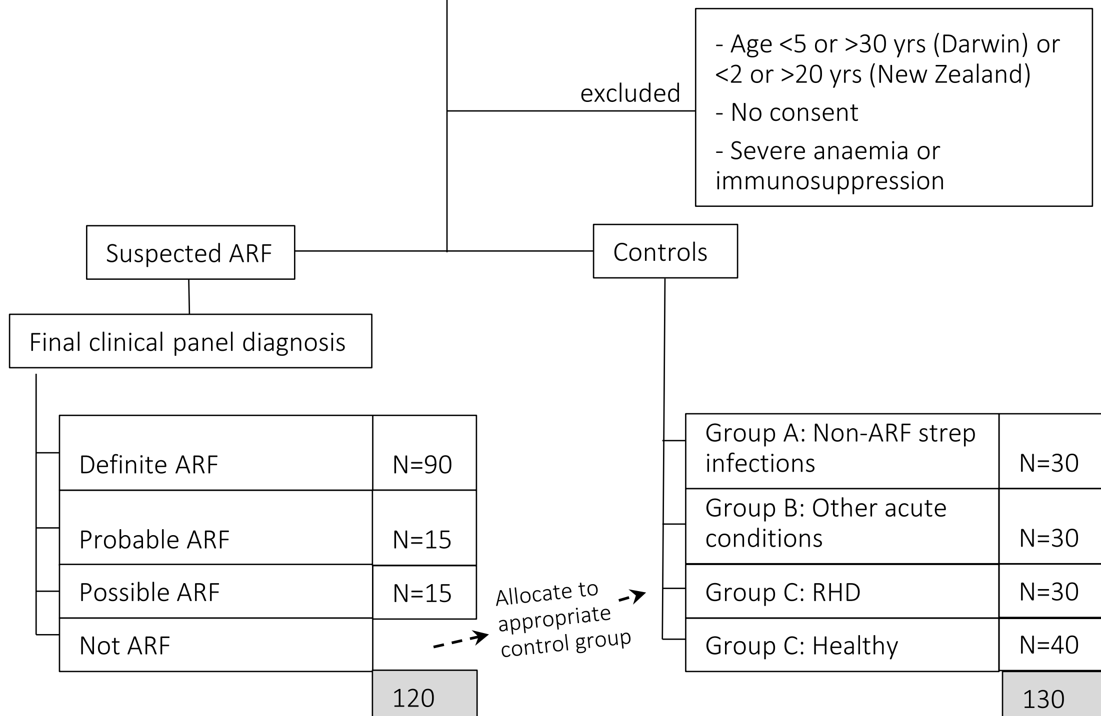
Figures

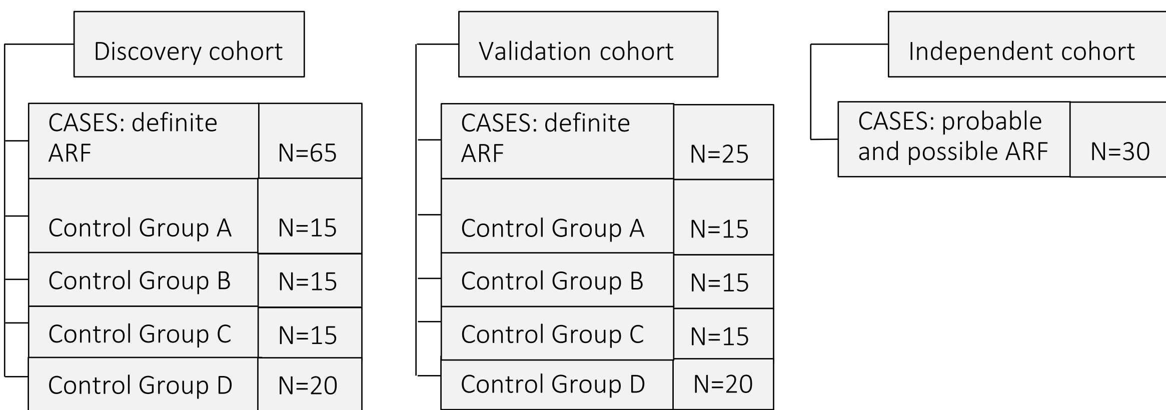
Figure 1: Study diagram showing target sample sizes
ARF = acute rheumatic fever; RHD = rheumatic heart disease

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Patients presenting to enrolment site or actively recruited through heathy volunteer drive





For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

BMJ Open

Searching for a Technology-Driven Acute Rheumatic Fever Test: the START study protocol

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-053720.R1
Article Type:	Protocol
Date Submitted by the Author:	04-Aug-2021

Complete List of Authors:	Ralph, A; Charles Darwin University; Royal Darwin Hospital Webb, Rachel; Counties Manukau District Health Board; Starship Children's Hospital Moreland, Nikki; The University of Auckland, School of Medical Sciences and Maurice Wilkins Centre McGregor, Reuben; The University of Auckland Bosco, Anthony; Telethon Kids Institute, The University of Western Australia Broadhurst, David; Edith Cowan Lassmann, Timo; Telethon Kids Institute Barnett, Timothy; Telethon Kids Institute, Wesfarmers Centre for Vaccines and Infectious Diseases Benothman, Rym; Telethon Kids Institute Yan, Jennifer; Menzies School of Health Research, Charles Darwin University; Royal Darwin Hospital Remenyi, Bo; Menzies School of Health Research, Charles Darwin University, Global and Tropical Health Division; Royal Darwin Hospital, Paediatrics Bennett, Julie; University of Otago Wilson, Nigel; Starship Children's Hospital Mayo, Mark; 1. Menzies School of Health Research, Charles Darwin University Pearson, Glenn; Telethon Kids Institute, Aboriginal Health Institute Leadership Team Kollmann, Tobias; Telethon Kids Institute Carapetis, Jonathan; Telethon Kids Institute

Primary Subject Heading:	Global health
Secondary Subject Heading:	Immunology (including allergy), Infectious diseases, Diagnostics
Keywords:	Paediatric rheumatology < PAEDIATRICS, Valvular heart disease < CARDIOLOGY, Immunology < TROPICAL MEDICINE, Diagnostic microbiology < INFECTIOUS DISEASES

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

For peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Note from the Editors: Instructions for reviewers of study protocols

Since launching in 2011, BMJ Open has published study protocols for planned or ongoing research studies. If data collection is complete, we will not consider the manuscript.

Publishing study protocols enables researchers and funding bodies to stay up to date in their fields by providing exposure to research activity that may not otherwise be widely publicised. This can help prevent unnecessary duplication of work and will hopefully enable collaboration. Publishing protocols in full also makes available more information than is currently required by trial registries and increases transparency, making it easier for others (editors, reviewers and readers) to see and understand any deviations from the protocol that occur during the conduct of the study.

The scientific integrity and the credibility of the study data depend substantially on the study design and methodology, which is why the study protocol requires a thorough peer-review.

BMJ Open will consider for publication protocols for any study design, including observational studies and systematic reviews.

Some things to keep in mind when reviewing the study protocol:

- Protocol papers should report planned or ongoing studies. The dates of the study should be included in the manuscript.
- Unfortunately we are unable to customize the reviewer report form for study protocols. As such, some of the items (i.e., those pertaining to results) on the form should be scores as Not Applicable (N/A).
- While some baseline data can be presented, there should be no results or conclusions present in the study protocol.
- For studies that are ongoing, it is generally the case that very few changes can be made to the methodology. As such, requests for revisions are generally clarifications for the rationale or details relating to the methods. If there is a major flaw in the study that would prevent a sound interpretation of the data, we would expect the study protocol to be rejected.

For peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Searching for a Technology-Driven Acute Rheumatic Fever Test: the START study protocol

For peer review only

Authors

Anna P. Ralph[1, 2*]
Rachel Webb[3,4,5]
Nicole J. Moreland[6]
Reuben McGregor[6]
Anthony Bosco[7]
David Broadhurst[8]
Timo Lassmann[7]
Timothy Barnett[9]
Rym Benothman[7]
Jennifer Yan[1, 2]
Bo Remenyi[1, 2]
Julie Bennett[10]
Nigel Wilson[4]
Mark Mayo[1]
Glenn Pearson[7]
Tobias Kollmann[7]
Jonathan R. Carapetis[7,11]

1. Menzies School of Health Research, Charles Darwin University, Darwin, Northern Territory, Australia
2. Royal Darwin Hospital, Darwin, Northern Territory, Australia
3. KidzFirst Hospital, Counties Manukau District Health Board, Auckland, New Zealand
4. Starship Children's Hospital, Auckland, New Zealand
5. Department of Paediatrics; Child and youth Health, University of Auckland, Auckland, New Zealand
6. School of Medical Sciences and Maurice Wilkins Centre, The University of Auckland, Auckland New Zealand
7. Telethon Kids Institute, Perth, Western Australia; Perth Children's Hospital, Perth, WA.
8. Centre for Integrative Metabolomics & Computational Biology, Edith Cowan University, Perth, WA.
9. Wesfarmers Centre for Vaccines and Infectious Diseases, Telethon Kids Institute, University of Western Australia, Nedlands, WA, Australia.
10. University of Otago, Wellington, New Zealand
11. University of Western Australia, Perth, Western Australia.

*corresponding author anna.ralph@menzies.edu.au

Word count: Abstract 295; Manuscript 4016

Keywords: Acute rheumatic fever; biomarker; transcriptomics; rheumatic heart disease

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Abstract

Introduction

The absence of a diagnostic test for Acute Rheumatic Fever (ARF) is a major impediment om managing this serious childhood condition. ARF is an autoimmune condition triggered by infection with Group A *Streptococcus*. It is the precursor to rheumatic heart disease (RHD), a leading cause of health inequity and premature mortality for Indigenous peoples of Australia, New Zealand and internationally.

Methods and Analysis

'Searching for a Technology-Driven Acute Rheumatic Fever Test' (START) is a biomarker discovery study that aims to detect and test a biomarker signature that distinguishes ARF cases from non-ARF, and use systems biology and serology to better understand ARF pathogenesis. Eligible participants with ARF diagnosed by an expert clinical panel according to the 2015 Revised Jones Criteria, aged 5-30 years, will be recruited from three hospitals in Australia and New Zealand. Age, sex and ethnicity-matched individuals who are healthy or have non-ARF acute diagnoses or RHD, will be recruited as controls. In the discovery cohort, blood samples collected at baseline, and during convalescence in a subset, will be interrogated by comprehensive profiling to generate possible diagnostic biomarker signatures. A biomarker validation cohort will subsequently be used to test promising combinations of biomarkers. By defining the first biomarker signatures able to discriminate between ARF and other clinical conditions, the START study has the potential to transform the approach to ARF diagnosis and RHD prevention.

Ethics and dissemination The study has approval from the Northern Territory Department of Health and Menzies School of Health Research ethics committee and the New Zealand Health and Disability Ethics Committee. It will be conducted according to ethical standards for research involving Indigenous Australians and New Zealand Māori and Pacific Peoples. Indigenous investigators and governance groups will provide oversight of study processes and advise on cultural matters.

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Article summary

Strengths and limitations of this study.

- START addresses a critically important health problem, acute rheumatic fever, which disproportionately affects global Indigenous populations and for which there is no diagnostic test nor effective treatment able to limit progression to rheumatic heart disease
- The large sample size recruited across international sites and the use of unbiased, comprehensive immune profiling will maximise the likelihood of being able to define a biomarker signature which can discriminate between ARF and other clinical conditions
- The most robust 'gold standard' rheumatic fever diagnostic process currently available will be used, comprising the revised 2015 Jones Criteria applied by an expert clinical panel
- Culturally safe processes with Indigenous governance guide the development and conduct of the study
- An inherent limitation is heterogeneity among people with ARF with regards to demographic factors, ARF type (first or recurrent), diagnostic category (possible, probable, definite), clinical phenotype (carditis, chorea, arthritis, skin and soft tissue) and timing in relation to ARF onset (variable time taken for healthcare to be reached after disease onset, consent gained and blood samples collected)

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Introduction

The absence of a diagnostic test for acute rheumatic fever (ARF) is a major impediment to management and epidemiological understanding of this serious childhood condition. ARF is one of the best examples of human autoimmunity caused by an infection, with Group A *Streptococcus* (GAS) the identified trigger. It is the precursor to rheumatic heart disease (RHD), a leading cause of health inequity and premature mortality for Indigenous peoples of Australia, New Zealand and internationally[1]. Prompt diagnosis of ARF allows timely commencement of secondary prophylaxis with long-acting penicillin to prevent repeated GAS infections that drive ARF recurrences. ARF recurrences are in turn the main mechanism responsible for development of RHD. Under-diagnosis of ARF appears to be a major contributor to the high rates of RHD in Aboriginal communities in Australia's Northern Territory (NT), where 76% of patients with RHD lack a prior ARF diagnosis[2]. Similarly a cohort study in New Zealand (NZ) found the majority of RHD cases (65%) had never been previously hospitalised with ARF[3]. Similar findings are reported in African countries.[4]

The Jones Criteria (Table 1) have been the diagnostic tool for ARF for nearly 80 years[5]. This is a regularly-reviewed clinical algorithm, now available in a user-friendly smart device application ('app')[6], most recently modified to optimise sensitivity for use in high-risk settings and optimise specificity for low-risk populations[7]. Despite these revisions, ARF diagnosis remains challenging and subjective. Diagnostic biomarkers for ARF that improve the performance of these criteria or provide a definitive diagnostic test would be a major advance in improving ARF detection, and thereby, RHD prevention.[8] Accurate case ascertainment is also needed to monitor progress towards disease control targets.

A further barrier to ARF management is the absence of an identified effective immunomodulatory treatment to alter long-term cardiac outcomes. Non-steroidal anti-inflammatory drugs alleviate joint symptoms but do not alter the long-term disease trajectory. Corticosteroids appear ineffective in limiting development or severity of RHD[9], but in some settings are prescribed as a treatment of last resort for severe carditis[10,11]. Hydroxychloroquine use has recently been reported in two cases of ARF with carditis, based on plausible *in vitro* data[12] but efficacy is as yet unknown[13]. Improved knowledge of ARF immune pathogenesis is needed to inform potential immune modulatory strategies targeted to the immune pathology.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Compounding complexity, different pathological processes may occur, accounting for the heterogeneous clinical phenotypes of ARF (Table 1, see major manifestations).

Most investigations into the immune basis of ARF have used biased (directed) approaches, with limited utility. For example, elevated TNFα, IL-6 and IL-8 have been reported in patients with active ARF and RHD, but are unreliable disease markers[14-16]. In difficult-to-diagnose diseases, the use of a multiplicity of biomarkers ('multivariate') to identify the presence or progression of disease has proven useful. Using micro-arrays, characteristic gene expression signatures have been reported in white blood cells from patients with Kawasaki disease (another inflammatory cardiac condition), lupus and tuberculosis demonstrating that multivariate expression profiling has potential to identify biomarker composite signatures as diagnostic tools[17-20]. Multivariate approaches to analysing biomarkers in sera or plasma have also shown promise. In a bead-based multiplex assay, a number of *M. tuberculosis*-reactive antibodies were successfully used to diagnose tuberculosis in non-human primates[21], and antibody glycosylation patterns can discriminate latent from active TB in humans[22]. Most importantly, our recent studies show proof-of-principle for immunopathogenesis studies to understand ARF, identifying a dysregulated cytokine axis using multivariate approaches including RNAseq and flow cytometry[12] as well as a linked IgG3-C4 response in early ARF with multiplex bead-based assays[23]. Serum- and plasma-borne biomarkers in addition to transcriptional profiling and circulating cellular components thus constitute a strong and accessible means of diagnosis once useful biomarkers have been identified.

Our hypothesis is that unbiased, multivariate analyses hold the greatest potential for identifying meaningful, translatable information on the immunopathogenesis of ARF. The aims of the 'Searching for a Technology-Driven Acute Rheumatic Fever Test' (START) study are to determine whether this approach can identify a biomarker signature that accurately classifies ARF diagnoses according to a defined 'gold standard' diagnosis (the 2015 revised Jones Criteria applied by a panel of clinical experts), and to better understand ARF immune pathogenesis (Table 2). Specifically, we aim to develop, in a discovery cohort and validate in a second cohort, a profile of metabolic and immunological biomarkers that distinguishes ARF cases from a range of non-ARF conditions and healthy controls.

Methods

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Design

START is a cross-national, prospective, observational study with a quasi case-control design. We will prospectively recruit children and young adults with ARF, and people matched on age, sex and ethnicity who have eligible control diagnoses, in Australia and NZ. The study will comprise a discovery cohort (up to 65 definite cases and 65 controls) and a validation cohort (up to 25 definite cases and 65 controls). Probable and possible ARF diagnoses will then be tested independently (Figure 1). Eligible control conditions are shown in Table 2. Participants will have blood collected for comprehensive immune profiling.

Ethics and dissemination

The study has approval from the NT Department of Health and Menzies School of Health Research ethics committee (18/3126) and the NZ Health and Disability Ethics Committee (18/CEN/197). It will be conducted according to Good Clinical Practice, Good Laboratory Practice and the National Health and Medical Research Council and ethical standards for research involving Indigenous Australians and NZ Māori and Pacific Peoples.

Knowledge diseemination will be through academic channels and community discussion forums. The study team is well placed to foster translation of findings into practical tools, namely, an improved diagnostic, and immunomodulatory therapy.

Study sites

Participating hospitals across the two international sites: Royal Darwin Hospital, Darwin, NT, Australia, and Starship and Middlemore Hospitals, Auckland, NZ. Clinical guidelines at all three study sites require that all people with suspected ARF are hospitalised for diagnostic workup and initiation of management[10,24]. Royal Darwin Hospital is the major tertiary referral hospital in Australia's NT. Thirty percent of the NT population is Indigenous (Aboriginal or Torres Strait Islander)[25]. ARF occurs at rates >300/100,000 total Aboriginal population in highest-burden communities in the NT[26]. Starship and Middlemore Hospitals serve a combined urban population of approximately 1.1 million people in the Auckland region. Māori and Pacific peoples make up approximately 27% of the population of greater Auckland region, where the incidence of ARF among Māori and Pacific children is approximately 33/100,000 and 93/100,000, respectively, for those 5–15 years old[24,27].

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

156 **Study Participants**

157 Inclusion criteria for ARF cases are: individuals with suspected or confirmed ARF (the vast
158 majority of whom are expected to identify as being of Aboriginal, Torres Strait Islander, Māori
159 or Pacific ethnicity) aged 5-30 years (Darwin) or 2-20 years (NZ). Written, informed consent
160 will be obtained from guardians or the participant themself if aged ≥16 years in NZ or ≥18
161 years in Australia. Differences between sites are consistent with local epiodemiology and local
162 ethical recommendations.
163
164 Exclusion criteria comprise individuals with severe anaemia in whom collection of the blood
165 volume required for the study collection would be relatively contraindicated; profound
166 immunosuppression other than corticosteroids for two months prior to study entry, or as
167 appropriate depending on the half-life of the immunosuppressive agent; and unstable social
168 situation precluding discussion about consent for research.
169
170 Controls are recruited into one of four groups: A. Non-ARF Streptococcal infections; B. Other
171 non-ARF acute inflammatory presentations; C. Established RHD without an intercurrent acute
172 diagnosis; and D. Healthy individuals (Table 2). All must meet the above-stated age, ethnicity
173 and consent inclusion criteria and none of the exclusion criteria. Controls will be matched at
174 data analysis stage to cases by site on sex, age within 10 years, and ethnicity. Individuals in
175 Groups A, B and D will have an echocardiogram to exclude underlying RHD.
176
177 **Cultural safety**

178 Cultural safety[28] is a core underpinning principle of the study. In NZ, the research is conducted
179 in accordance with core Te Tiriti o Waitangi principles affording protection, participation and
180 partnership for Māori participants. Research staff enrolling participants will engage approriate
181 family members (in NZ: whānau) and respect differences in decision-making processes.
182 Interpreters and translated participant materials will be used where study participants and their
183 families primarily speak another language. In Darwin, patient information has been recorded
184 in six Aboriginal languages. In NZ, participant materials have been translated into Te Reo
185 Māori, Samoan and Tongan.
186
187 **Patient and Public Involvement**

188 The Indigenous investigators and collaborators at each site form governance groups to provide
189 oversight of study design, conduct, reporting and dissemination. Patient and public

For peer review only

involvement, drawing on our established consumer networks, will support knowledge dissemination.

Enrolment

It will be made clear that participation is voluntary and any benefits of participation are to society through knowledge advancement, rather than to the individual. Participants will be advised that wherever possible, blood collection for the study will coincide with collection done for clinical purposes, to avoid extra venepunctures. If a participant or their parent/guardian withdraws consent, no further samples will be collected unless a new consent form is signed. Participants are asked whether their blood may be stored for future use.

People with suspected ARF will be enrolled as early as possible during admission, prior to a final diagnosis being assigned, to ensure that an acute sample is collected. A final diagnosis will be assigned once all required diagnostic information has become available. Participants in Group C (RHD) may additionally be recruited through outpatient clinics. Group D (healthy) will be sought from among family members and friends of other enrolled participants, or healthy members of the community in Darwin.

Clinical data collection

After assigning a sequential study code, clinical details will be recorded on a paper Case Report Form (CRF) then entered into an electronic database (Medrio™ Electronic Data Capture System), including: date of illness onset, date of blood sample collection, demographics, clinical presentation, Jones criteria, clinical laboratory results, medications received and diagnosis assigned by the clinical treating team.

Assignment of diagnosis

Patients presenting with suspected ARF may have a final diagnosis of definite ARF, probable ARF, possible ARF or a non-ARF condition. *Probable ARF* is defined in Australian guidelines as an acute presentation not fulfilling criteria, missing a major or minor criterion or lacking evidence of preceding streptococcal infection, but ARF is still considered the most likely diagnosis. *Possible ARF* applies to the same presentation type, but where ARF is considered uncertain but cannot be ruled out[10]. For this study, participants ultimately diagnosed with a non-ARF condition will be allocated by the clinical panel as group A or B controls.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Management will be directed by treating clinicians. Clinical diagnoses will be assigned according to local Australian and NZ Diagnostic Criteria by a panel of ≥2 study clinicians expert in ARF. The decision on diagnostic category requires clinical judgement, after evaluating all aspects of the case. The Jones criteria app[6] which provides a result of definite, probable or possible ARF or not ARF depending on information provided to the algorithm, will be used by the panel at their discretion. The diagnosis of carditis will be made if rheumatic valvulitis is evident on echocardiogram as per Jones criteria. The diagnosis of rheumatic heart disease will be made in accordance with features described by the World Heart Federation[29]. The 2015 revised Jones Criteria[7,30] will be applied as the gold standard diagnostic for definite ARF. This represents a departure from normal practice in NZ where local diagnostic criteria differ: polyarthralgia is not considered a major criterion in NZ[24]. In addition to the Jones criteria, comprehensive clinical and laboratory information such as results of synovial fluid testing, serology for alternative infectious aetiologies, autoantibodies, radiological findings, diagnosis assigned by the treating team, family history and local epidemiology, will be considered in adjudicating the diagnosis. ARF type will be further specified as initial episode or recurrence. Where the adjudication panel's diagnosis differs from that of the treating team with implications for management, this will be communicated to the treating specialist. The clinical panel will also assign final diagnoses for controls.

Blood collection

Blood will be collected at baseline, and for patients with ARF, on follow-up occasions during the convalescent period as able e.g. during prolonged hospitalisation or at a later outpatient appointment (Table 3). A safe maximum volume of 28.5 mL depending on age will be obtained in Australia[31], and 14.5 mL in NZ. Baseline samples will be collected as soon as possible during the acute presentation, timed to coincide with routine blood testing wherever possible. Convalescent samples will be used to determine persistence and decay of immunopathological signatures post ARF. After collection, samples will be transported immediately to research laboratories at Menzies School of Health Research (Australia) or the University of Auckland (NZ) for centrifugation, serum/plasma separation and viable peripheral blood mononuclear cell (PBMC) preparation. Timely freezing and storage of serum/plasma aliquots, PAXgene tubes at -80°C and PBMCs in gas-phase liquid nitrogen. Samples will be shipped periodically to relevant laboratories.

Sample size

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

It is not possible to predict *a priori* the combined discriminatory ability of independently measured factors. This study has elected to enrol up to 120 ARF cases (comprising approximately 90 definite ARF, 15 probable, 15 possible) and 130 controls (30 each of control groups A, B and C, and 40 control group D; Figure 1 and Table 2). These numbers are feasible and should enable characterisation of different ARF clinical phenotypes (carditis, arthritis, chorea, skin/soft tissue manifestations or a combination of these) and phases (first or recurrent episodes), acknowledging heterogeneity among participants (ethnicity, age, sex; on corticosteroids or not; at different stages of illness).

In the multivariate approach it is expected that most, if not all, of the influential factors in the optimal model will be significant ($p<0.05$) in univariate testing. Assuming α of 5% and 95% power ($\beta = 0.05$) the minimum sample size is 26 cases. These calculations are only a guideline for removing clear false positives, as the standard procedure for producing robust screening models is to have a separate blinded analysis of a test subset of samples.We will utilise Discovery and Validation Cohorts, together with stratified bootstrap cross-validation to internally optimize the structural parameters in each model. We further used the standard inferential approach to sample size estimation in diagnostic test studies of biomedical informatics[32]. An effective multivariate predictor of ARF with positive predictive value of 0.8 (clinically useful), and a 95% confidence interval of +/- 0.1, would require 65 ARF cases in the discovery cohort.

The discovery cohort will therefore comprise a target of 65 definite ARF cases and 65 controls. The validation cohort will comprise a target of 25 definite ARF cases and 65 controls. Probable and possible diagnoses (~15 in each group) will be tested independently (Figure 1). Cases will be allocated to the discovery or validation cohorts using a computer-generated random selection of unique study identification numbers.

Laboratory methods

The overarching aim is to develop and validate a profile of host related biomarkers that distinguishes ARF cases from non-ARF conditions and healthy controls. This will be achieved through the following analyses on peripheral blood samples.

Immunophenotyping:

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

PBMC will be stained with labeled antibody panels to identify specific cell populations. Flow cytometry raw data will be analyzed manually using Flowjo software and via an automated gating platform as described[33,34]. Such high throughput, automated analysis of big flow cytometry data offers several advantages over manual gating, including increased throughput while increasing quality control (such as pre-processing removal of anomalous events via the flowCut algorithm), and identification of specific cell populations with up to 50-dimensional datasets[34-36]. Plasma samples will be analyzed utilizing multiplex cytokine assays to quantify cytokine and chemokine plasma concentrations, detailed elsewhere[33,37-39].

Metabolome analyses:

Untargeted metabolomic profiling (>1000 metabolites) will be performed on plasma samples using liquid chromatography coupled to high resolution mass-spectrometry (LC-HRMS). Data will be acquired using three modes of operation: reverse-phase/UPLC-MS/MS with positive ion mode electrospray ionization (ESI), reverse-phase/UPLC-MS/MS with negative ion mode ESI, and HILIC/UPLC-MS/MS with negative ion mode ESI. All identified metabolites will be annotated using appropriate orthogonal analytical techniques applied to the metabolite of interest against a chemical reference standard.

Blood Transcriptomics:

RNA will be extracted from stabilized whole blood samples (PAXgene tubes) as described[40,41], libraries prepared (TruSeq Stranded Total RNA with Ribo-Zero Globin reduction, Illumina) and NextGen sequencing undertaken (Illumina HiSeq2500, 50-bp single-end reads). Read alignment and gene-level quantification (counts) will be performed using Hisat[42]. Negative binomial models will be employed for differential expression analysis, with false discovery rate control for multiple testing[43]. The analyses will be adjusted for batch effects and variations in cellular composition, which will be estimated employing RUVSeq and CIBERSORT respectively[44,45]. Genes will be mapped to blood transcriptional modules to provide a systems-level view of the responses and reduce the dimensionality of the data[46].

CD4 T cell transcriptomic responses to GAS:

A two-phased approach will be used to examine CD4 T cell responses to GAS. Firstly, PBMCs from a subset of definite ARF cases and healthy controls will be cultured under a variety of conditions to identify optimal conditions for the second phase. Cells will be harvested at multiple time points post-stimulation (6h, 24h and 48h) with heat- and antibiotic-killed ARF-

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

associated GAS strains[12,47,48] and a selection of candidate ARF antigens. Total RNA from PBMCs cultured under each condition will be profiled by RNA-Seq (Illumina 100bp paired-end reads, 20M reads), and culture supernatants will be examined by Luminex (Bio-Plex 48-plex Pro Human Cytokine Screening Panel) to identify optimal conditions for maximal discrimination of responses between cases and controls. In the second phase, an expanded set of ARF cases and controls for which PBMC samples are available will be examined following culture under the identified, optimal conditions. Gene expression patterns and cytokine production will be profiled by RNA-Seq and Luminex, as described above.

Antibody analysis:

For unbiased investigation of autoantibodies, selected sera will be screened against planer protein microarrays comprised of over 42,000 protein fragments representing some 19,000 human proteins[49]. Protein fragments that are more significantly bound by autoantibodies in ARF compared to controls will be identified using a P-value of >0.05 and fold change of 2.0 as cut-off. A suspension bead array will then be designed[50] comprised of up to 380 potential autoantigens to assess autoantigen reactivity in all sera. The bead-antigens will be selected based on the planar array results, previously completed screens using high-content protein arrays[51], and targets from the literature. Finally, candidate ARF autoantigens identified in the suspension bead-array will be orthogonally validated as individual antigens in ELISA or Luminex bead-based assays to determine sensitivity and specificity.

Bioinformatics and statistical analyses:

The data collected for each patient across the immune phenotyping, transcriptomic, proteomic and metabolomic technologies will be integrated computationally to identify patterns of covariance within and between designated clinical outcome groups. Convergence of signatures across diverse systems biology domains will provide independent functional validation. The use of "multi-omic" data integration techniques will also enable the possibility of deriving novel biological information that will not be revealed in a single dataset alone.

Data integration methods applying different computational strategies will be used to explore the complex covariance structure within and between the multi-omic data blocks. Initially, analysis will focus on each data block in isolation employing a combination of classical generalised linear modelling (GLM), multivariate projection regression models (PCA, PLS-DA) and unsupervised cluster analysis. The multiple blocks will then be integrated into a single

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

computational model to enable multi-omic functional mapping, which in turn will allow us to uncover the underlying biochemical mechanisms. Several methods for performing multi-block data integration will be investigated, from which a consensus model will be derived. These will include: protein−protein interaction networks using NetworkAnalyst[52], Multi-block Component Analysis[53], regularized Canonical Correlation Analysis[54], the Data Integration Analysis for Biomarker discovery using Latent cOmponent (DIABLO) framework[55] and Similarity Network Fusion[56].

This systems biology analysis should result in both a domain-specific and domain-integrative summary of biological phenomena that are associated with clinical outcome. Likely outcomes are a signature biological pathway associated with ARF as well as individual candidate markers that may provide the basis for the development of novel diagnostic tests. Furthermore, the data generated through this effort will allow crisply defined power calculations for future narrowly targeted assessment of potential biomarkers in clinical trials.

Discussion

We anticipate that study findings will provide the most comprehensive knowledge of the immunopathogenesis of ARF to date, be used as the basis for development of a diagnostic test and provide a pathway towards development of targeted immunomodulatory treatment(s). START commenced recruitment in Australia in November 2018 and in NZ in May 2019. Anticipated completion of recruitment is end-2021. Laboratory analyses will be batched and run together at completion of recruitment.

Both under-and over-diagnosis of ARF pose major challenges to individuals and health systems.[10] Whether the use of the 'probable' and 'possible' ARF diagnostic categories is reducing underdiagnosis or contributing to overdiagnosis is currently unknown. Should findings from this study successfully identify a discriminatory biomarker profile differentiating definite ARF from non-ARF, this will provide a mechanism for accurate diagnosis to guide appropriate management. This will be a critically important advance in the diagnosis and management of ARF globally.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

A minority of any population are at risk of ARF after exposure to group A streptococci. The estimated lifetime cumulative incidence of ARF was previously calculated at 5.7% in Australia's Northern Territory[57]. This was considered likely to be an under-estimate, and indeed more recent data highlight the burden of unrecognised ARF in Aboriginal communities in the NT where repeated infection with GAS is ubiquitous from early childhood[2,58], suggesting that the lifetime risk of ARF in these populations is higher than 6%. In US military camps, 2-3% of recruits developed ARF after a single bout of streptococcal pharyngitis[59].

Previous efforts to develop ARF diagnostic tests and elucidate the immunopathogenesis have laid foundations for the START study design and analysis. In 2018, Kim *et al.* analysed responses to GAS challenge of PBMC from ARF patients at Royal Darwin Hospital using multiplex cytokine array, flow cytometric analysis, and global gene expression analysis by RNA sequencing[12]. They identified a dysregulated interleukin-1β-granulocyte- macrophage colony-stimulating factor (GM-CSF) cytokine axis in PBMCs from ARF patients, and the potential to suppress this response by hydroxychloroquine. The authors proposed that hydroxychloroquine, an immunomodulatory agent, could be repurposed to reduce the risk or severity of RHD after ARF[12]. Clinical use of hydroxychloroquine, now reported as safe in two cases[13], requires further investigation. A study using sera from patients with ARF in NZ combined multiplex bead-based assays and systems immunology data analysis to identify a linked IgG3-C4 response that may have utility as a clinical biomarker in early ARF[23]. However, small sample sizes have been a limiting factor. Key strengths of the current study include the larger sample size, multicentre enrolment across different countries to maximise relevance, and use of the latest laboratory approaches.

Limitations include that capture of eligible participants early in disease is challenging, due to the variable time taken for participants to be hospitalised after disease onset, to identify and gain consent from participants, and obtain blood samples. At the Australian site, the majority of eligible participants will be Aboriginal children living in remote communities, often hundreds of kilometres distant from the enrolment site. Although Auckland is a large urban city inequitable healthcare access is experienced by many Māori and Pacific people living in socio-economic deprivation and this can lead to delayed hospitalisation for many children presenting with ARF. Given these inherent delays, a proportion of baseline blood samples will be collected after the peak inflammatory phase has passed. Another limitation is heterogeneity among people with ARF with regards to ARF type (first or recurrent), diagnostic category

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

(possible, probable, definite), clinical phenotype (carditis, chorea, arthritis etc) and demographic variations; however, we limited enrolment to people below 30 years and only of Aboriginal, Torres Strait Islander, Māori or Pacific ethnicity to minimize heterogeneity. Although a distinguishing biomarker signature may be identified, the ability to readily translate study findings into a feasible diagnostic able to be used in health service laboratories outside a research environment, is uncertain.

Successful completion of this study will considerably improve knowledge of ARF immunology. Innovative strategies to improve the clinical management of ARF are core components of the overall suite of activities required to achieve the goal RHD elimination[60].

For peer review only

Acknowledgements: We are very grateful to Caroline Motteram for project management, Yolanda Hernandez, Laura Francis and Vincent Mithen for recruiting participants in Darwin and Shirley Lawrence, Rhonda Holloway, Renee Clark, Alicia Stanley, and Roxanne Buchanan for recruiting participants in New Zealand. We thank Alex Kaethner and NT Cardiac for facilitating echocardiography. Vanessa Rigas, Erin Gargan, Gabriella Minigo and Michael Serralha are gratefully acknowledged for laboratory support; Kyran Kuzio, Tasha Cole, Alana Whitcombe, Lauren Carlton and Timothy Ho for laboratory assistance; Amy Baker and Rebecca Trowman for program and financial management.

Authors' contributions: JRC, APR, RW, NJM, RM, AB, DB, TB, MM, GP, TK conceived and designed the study. NJM, RM, RMG, AB, DB,TL,TB, RB, JB, TK will perform the assays and analyse the data. GP and MM provide project governance. APR, RW, JY, BR, NW and JRC provided clinical care, patient recruitment, assessment and clinical review panel participation. APR create the first draft and managed the edits to the manuscript. All authors reviewed and approved the study protocol.

Funding statement: The study is funded by the Australian National Health and Medical Research Council (NHMRC) grant #1147531. APR is supported by NHMRC fellowship #1142011. TCB is supported by the Western Australia Department of Health (#not applicable) and a Career Development Fellowship from Improving Health Outcomes in the Tropical North (NHMRC #1131932)

Competing interests statement. The authors have no conflicts of interest to declare.

Data statement: Not applicable. No data have been presented in this protocol paper.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Tables

Table 1: Revised Jones Criteria

Legend: Reproduced with permission from The 2020 Australian guideline for prevention, diagnosis and management of acute rheumatic fever and rheumatic heart disease (3rd edition)[2]

<table>
<tr><th></th><th>High-risk groups†</th><th>Low-risk groups</th></tr>
<tr><td>Definite initial episode of ARF</td><td colspan="2">2 major manifestations + evidence of preceding GAS infection, OR
1 major + 2 minor manifestations + evidence of preceding GAS infection‡</td></tr>
<tr><td>Definite recurrent § episode of ARF in a patient with a documented history of ARF or RHD</td><td colspan="2">2 major manifestations + evidence of preceding GAS infection, OR
1 major + 2 minor manifestations + evidence of preceding GAS infection‡, OR
3 minor manifestations + evidence of a preceding GAS infection‡</td></tr>
<tr><td>Probable or possible ARF (first episode or recurrence§)</td><td colspan="2">A clinical presentation in which ARF is considered a likely diagnosis but falls short in meeting the criteria by either:
• one major or one minor manifestation, OR
• no evidence of preceding GAS infection (streptococcal titres within normal limits or titres not measured)
Such cases should be further categorised according to the level of confidence with which the diagnosis is made:
• Probable ARF (previously termed 'probable: highly suspected')
• Possible ARF (previously termed 'probable: uncertain')</td></tr>
<tr><td>Major manifestations</td><td>Carditis (including subclinical evidence of rheumatic valvulitis on echocardiogram)
Polyarthritis¶ or aseptic monoarthritis or polyarthralgia
Sydenham chorea††
Erythema marginatum ‡‡
Subcutaneous nodules</td><td>Carditis (including subclinical evidence of rheumatic valvulitis on echocardiogram)
Polyarthritis¶
Sydenham chorea††
Erythema marginatum‡‡
Subcutaneous nodules</td></tr>
<tr><td>Minor Manifestations</td><td>Fever §§ ≥38°C
Monoarthralgia ¶¶
ESR ≥30 mm/h or CRP ≥30 mg/L
Prolonged P-R interval on ECG†††</td><td>Fever ≥38.5°C
Polyarthralgia or aseptic monoarthritis¶¶
ESR ≥60 mm/h or CRP ≥30 mg/L
Prolonged P-R interval on ECG†††</td></tr>
</table>

† High-risk groups are those living in communities with high rates of ARF (incidence >30/100,000 per year in 5–14-year-olds) or RHD (all-age prevalence >2/1000). Aboriginal and Torres Strait Islander peoples living in rural or remote settings are known to be at high risk. Data are not available for other populations but Aboriginal and Torres Strait Islander peoples living in urban settings, Māori and Pacific Islanders, and potentially immigrants from developing countries, may also be at high risk.
‡ Elevated or rising antistreptolysin O or other streptococcal antibody, or a positive throat culture or rapid antigen or nucleic acid test for GAS infection.
§ Recurrent definite, probable or possible ARF requires a time period of more than 90 days after the onset of symptoms from the previous episode of definite, probable or possible ARF
¶ A definite history of arthritis is sufficient to satisfy this manifestation. Note that if polyarthritis is present as a major manifestation, polyarthralgia or aseptic monoarthritis cannot be considered an additional minor manifestation in the same person.
†† Chorea does not require other manifestations or evidence of preceding GAS infection, provided other causes of chorea are excluded.
‡‡ Care should be taken not to label other rashes, particularly non-specific viral exanthems, as erythema marginatum.
§§ In high-risk groups, fever can be considered a minor manifestation based on a reliable history (in the absence of documented temperature) if anti-inflammatory medication has already been administered.
¶¶ If polyarthritis is present as a major criterion, monoarthritis or arthralgia cannot be considered an additional minor manifestation.
††† If carditis is present as a major manifestation, a prolonged P-R interval cannot be considered an additional minor manifestation.
CRP, C-reactive protein; ECG, electrocardiogram; ESR, erythrocyte sedimentation rate.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 2: Control groups

Group A (n=50):	Non-ARF Streptococcal infections or toxin mediated condition or post-streptococcal condition Examples: - active infection with a proven or likely causative streptococcal organism (e.g. β- haemolytic streptococcus, α-haemolytic streptococcus, *S. pneumoniae* etc) from sterile or non-sterile site - scarlet fever - acute post-streptococcal glomerulonephritis AND - No evidence of definite RHD on echocardiogram
Group B (n=50):	Other acute condition Examples: - inflammatory conditions, including autoimmune inflammatory conditions (e.g. lupus with arthritis), pancreatitis etc - infective condition: confirmed arboviral infection (eg Ross River virus), septic arthritis, bacterial endocarditis, etc - trauma requiring hospitalisation or operative procedure AND - No evidence of definite RHD on echocardiogram
Group C (n=50):	RHD - Echocardiographically confirmed RHD with no active inflammation OR - past history of definite ARF but no current disease activity.
Group D (n=50):	Healthy - No intercurrent medical condition as determined by clinical questionnaire - Minor trauma allowed AND - No evidence of definite RHD on echocardiogram

For peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 3: Blood collection

	Australian recruitment site		
	Participant age		**Collection tube**
Testing procedures	**5-9 years**	**≥ 10 years**	
Transcriptional profiling	2.5mL	2.5mL	PaxGene tube
Peripheral blood mononuclear cells for flow cytometry, Metabolomics	4.0mL	22.0mL	Sodium Heparin
Antibody studies	4.0mL	4.0mL	Serum
TOTAL	10.5mL	28.5mL	
	New Zealand recruitment sites		
	Participant age		**Collection tube**
Testing procedures	**2-10 years**	**≥ 11 years**	
Transcriptional profiling	2.5mL	2.5mL	PaxGene tube
Peripheral blood mononuclear cells for flow cytometry, Metabolomics	4.0mL	8.0mL	Sodium Heparin
Antibody studies	4.0mL	4.0mL	Serum
TOTAL	10.5mL	14.5mL	

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

References

1. Ralph AP, Read C, Johnston V, et al. Improving delivery of secondary prophylaxis for rheumatic heart disease in remote Indigenous communities: study protocol for a stepped-wedge randomised trial. *Trials* 2016;17(1):51. doi: 10.1186/s13063-016-1166-y
2. Hardie K, Ralph AP, de Dassel JL. RHD elimination: action needed beyond secondary prophylaxis. *Aust N Z J Public Health* 2020;44(5):427. doi: 10.1111/1753-6405.13002 [published Online First: 2020/06/09]
3. Oliver J, Robertson O, Zhang J, et al. Progression of Acute Rheumatic Fever to Recurrence, Rheumatic Heart Disease, and Death in New Zealand Children and Youth: A Cohort Study. *Heart, Lung and Circulation* 2019;28
4. Zuhlke L, Engel ME, Karthikeyan G, et al. Characteristics, complications, and gaps in evidence-based interventions in rheumatic heart disease: the Global Rheumatic Heart Disease Registry (the REMEDY study). *European heart journal* 2015;36(18):1115-22a. doi: 10.1093/eurheartj/ehu449
5. Jones TD. The diagnosis of rheumatic fever. *J Am Med Assoc* 1944;126:481-84.
6. RHDAustralia. iPhone and Android Apps. http://www.rhdaustralia.org.au/apps 2020 [14 August 2014].
7. Gewitz MH, Baltimore RS, Tani LY, et al. Revision of the Jones Criteria for the diagnosis of acute rheumatic fever in the era of Doppler echocardiography: a scientific statement from the American Heart Association. *Circulation* 2015;131(20):1806-18. doi: 10.1161/CIR.0000000000000205
8. McMillan DJ, Rafeek RAM, Norton RE, et al. In Search of the Holy Grail: A Specific Diagnostic Test for Rheumatic Fever. *Front Cardiovasc Med* 2021;8:674805. doi: 10.3389/fcvm.2021.674805
9. Cilliers A, Adler AJ, Saloojee H. Anti-inflammatory treatment for carditis in acute rheumatic fever. *Cochrane database of systematic reviews* 2015(5):CD003176. doi: 10.1002/14651858.CD003176.pub3
10. RHDAustralia (ARF/RHD writing group). The 2020 Australian guideline for prevention, diagnosis and management of acute rheumatic fever and rheumatic heart disease (3rd edition) https://www.rhdaustralia.org.au/arf-rhd-guideline. Menzies School of Health Research, 2020.
11. Dean GL, Edwards SG, Ives NJ, et al. Treatment of tuberculosis in HIV-infected persons in the era of highly active antiretroviral therapy. *Aids* 2002;16(1):75-83.
12. Kim ML, Martin WJ, Minigo G, et al. Dysregulated IL-1beta-GM-CSF Axis in Acute Rheumatic Fever That Is Limited by Hydroxychloroquine. *Circulation* 2018;138(23):2648-61. doi: 10.1161/CIRCULATIONAHA.118.033891 [published Online First: 2018/12/21]
13. Wilson NJ, Concannon A, Malcolm J, et al. The Treatment of Acute Rheumatic Fever: Novel Use of Hydroxychloroquine. *Pediatr Infect Dis J* 2020 doi: 10.1097/INF.0000000000002647 [published Online First: 2020/03/30]
14. M. Hafez M, EL-Morsy Z, EL-Shennawy F, et al. Susceptibility to Over Production of Cytokines in Acute Rheumatic Carditis and Their Role in the Pathogenesis. *Journal of Medical Sciences* 2002 doi: 10.3923/jms.2002.65.73
15. Yegin O, Coskun M, Ertug H. Cytokines in acute rheumatic fever. *Eur J Pediatr* 1997;156(1):25-9. doi: 10.1007/s004310050545 [published Online First: 1997/01/01]
16. Narin N, Kutukculer N, Narin F, et al. Anticardiolipin antibodies in acute rheumatic fever and chronic rheumatic heart disease: is there a significant association? *Clin Exp Rheumatol* 1996;14(5):567-9. [published Online First: 1996/09/01]

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

17. Wright VJ, Herberg JA, Kaforou M, et al. Diagnosis of Kawasaki Disease Using a Minimal Whole-Blood Gene Expression Signature. *JAMA Pediatr* 2018;172(10):e182293. doi: 10.1001/jamapediatrics.2018.2293 [published Online First: 2018/08/08]
18. Berry MP, Graham CM, McNab FW, et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* 2010;466(7309):973-7. doi: 10.1038/nature09247 [published Online First: 2010/08/21]
19. Bennett L, Palucka AK, Arce E, et al. Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. *J Exp Med* 2003;197(6):711-23. doi: 10.1084/jem.20021553 [published Online First: 2003/03/19]
20. Anderson ST, Kaforou M, Brent AJ, et al. Diagnosis of childhood tuberculosis and host RNA expression in Africa. *N Engl J Med* 2014;370(18):1712-23. doi: 10.1056/NEJMoa1303657
21. Khan IH, Ravindran R, Yee J, et al. Profiling antibodies to Mycobacterium tuberculosis by multiplex microbead suspension arrays for serodiagnosis of tuberculosis. *Clinical and vaccine immunology : CVI* 2008;15(3):433-8. doi: 10.1128/CVI.00354-07 [published Online First: 2007/12/14]
22. Lu LL, Chung AW, Rosebrock TR, et al. A Functional Role for Antibodies in Tuberculosis. *Cell* 2016;167(2):433-43 e14. doi: 10.1016/j.cell.2016.08.072 [published Online First: 2016/09/27]
23. Chung AW, Ho TK, Hanson-Manful P, et al. Systems immunology reveals a linked IgG3-C4 response in patients with acute rheumatic fever. *Immunol Cell Biol* 2020;98(1):12-21. doi: 10.1111/imcb.12298 [published Online First: 2019/11/20]
24. Heart Foundation of New Zealand. New Zealand Guidelines for Rheumatic Fever: Diagnosis, Management and Secondary Prevention of Acute Rheumatic Fever and Rheumatic Heart Disease: 2014 Update. https://www.heartfoundation.org.nz/resources/acute-rheumatic-fever-and-rheumatic-heart-disease-guideline, 2014.
25. Australian Bureau of Statistics. Estimates of Aboriginal and Torres Strait Islander Australians 2016 [Available from: https://www.abs.gov.au/ausstats/abs@.nsf/mf/3238.0.55.001 accessed 27 April 2020.
26. Francis J, R, , Gargan C, Remenyi B, et al. A cluster of acute rheumatic fever cases among Aboriginal Australians in a remote community with high baseline incidence. *Aust N Z J Public Health* 2019;4(3):288-93.
27. Institute of Environmental Science and Research Limited. Rheumatic fever report, January-December 2018. EpiSurv Science for Communities https://surv.esr.cri.nz/PDF_surveillance/RheumaticFever/Rheumaticfeverbi-annualreportJan-Dec2018.pdf. 2018
28. Curtis E, Jones R, Tipene-Leach D, et al. Why cultural safety rather than cultural competency is required to achieve health equity: a literature review and recommended definition. *International journal for equity in health* 2019;18(1):174.
29. Remenyi B, Wilson N, Steer A, et al. World Heart Federation criteria for echocardiographic diagnosis of rheumatic heart disease--an evidence-based guideline. *Nature reviews Cardiology* 2012;9(5):297-309. doi: 10.1038/nrcardio.2012.7
30. Beaton A, Carapetis J. The 2015 revision of the Jones criteria for the diagnosis of acute rheumatic fever: implications for practice in low-income and middle-income countries. *Heart Asia* 2015;7(2):7-11. doi: 10.1136/heartasia-2015-010648 [published Online First: 2015/01/01]
31. Seattle Children's Hospital. Maximum Allowable Blood Draw From Infants. Mayo Foundation for Medical Education and Research 2020 [28 April 2020]. Available from: https://seattlechildrenslab.testcatalog.org/show/1000721-1.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

32. Hajian-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. *J Biomed Inform* 2014;48:193-204. doi: 10.1016/j.jbi.2014.02.013 [published Online First: 2014/03/04]
33. Ben-Othman R, Cai B, Liu AC, et al. Systems Biology Methods Applied to Blood and Tissue for a Comprehensive Analysis of Immune Response to Hepatitis B Vaccine in Adults. *Frontiers in immunology* 2020;11:580373. doi: 10.3389/fimmu.2020.580373 [published Online First: 2020/12/01]
34. Rahim A, Meskas J, Drissler S, et al. High throughput automated analysis of big flow cytometry data. *Methods* 2018;134-135:164-76. doi: 10.1016/j.ymeth.2017.12.015 [published Online First: 2017/12/31]
35. Malek M, Taghiyar MJ, Chong L, et al. flowDensity: reproducing manual gating of flow cytometry data by automated density-based cell population identification. *Bioinformatics* 2015;31(4):606-7. doi: 10.1093/bioinformatics/btu677 [published Online First: 2014/11/08]
36. Ge Y, Sealfon SC. flowPeaks: a fast unsupervised clustering for flow cytometry data via K-means and density peak finding. *Bioinformatics* 2012;28(15):2052-8. doi: 10.1093/bioinformatics/bts300 [published Online First: 2012/05/19]
37. Amenyogbe N, Dimitriu P, Cho P, et al. Innate Immune Responses and Gut Microbiomes Distinguish HIV-Exposed from HIV-Unexposed Children in a Population-Specific Manner. *J Immunol* 2020;205(10):2618-28. doi: 10.4049/jimmunol.2000040 [published Online First: 2020/10/18]
38. Amenyogbe N, Dimitriu P, Smolen KK, et al. Biogeography of the Relationship between the Child Gut Microbiome and Innate Immune System. *mBio* 2021;12(1) doi: 10.1128/mBio.03079-20 [published Online First: 2021/01/14]
39. Moncunill G, Dobano C, Gonzalez R, et al. Association of Maternal Factors and HIV Infection With Innate Cytokine Responses of Delivering Mothers and Newborns in Mozambique. *Front Microbiol* 2020;11:1452. doi: 10.3389/fmicb.2020.01452 [published Online First: 2020/08/09]
40. Lee AH, Shannon CP, Amenyogbe N, et al. Dynamic molecular changes during the first week of human life follow a robust developmental trajectory. *Nat Commun* 2019;10(1):1092. doi: 10.1038/s41467-019-08794-x [published Online First: 2019/03/14]
41. Idoko OT, Smolen KK, Wariri O, et al. Clinical Protocol for a Longitudinal Cohort Study Employing Systems Biology to Identify Markers of Vaccine Immunogenicity in Newborn Infants in The Gambia and Papua New Guinea. *Front Pediatr* 2020;8:197. doi: 10.3389/fped.2020.00197 [published Online First: 2020/05/20]
42. Pertea M, Kim D, Pertea GM, et al. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc* 2016;11(9):1650-67. doi: 10.1038/nprot.2016.095 [published Online First: 2016/08/26]
43. Anders S, McCarthy DJ, Chen Y, et al. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nat Protoc* 2013;8(9):1765-86. doi: 10.1038/nprot.2013.099 [published Online First: 2013/08/27]
44. Risso D, Ngai J, Speed TP, et al. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat Biotechnol* 2014;32(9):896-902. doi: 10.1038/nbt.2931 [published Online First: 2014/08/26]
45. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* 2015;12(5):453-7. doi: 10.1038/nmeth.3337 [published Online First: 2015/03/31]
46. Li S, Rouphael N, Duraisingham S, et al. Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. *Nat Immunol* 2014;15(2):195-204. doi: 10.1038/ni.2789 [published Online First: 2013/12/18]

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

47. Smoot JC, Barbian KD, Van Gompel JJ, et al. Genome sequence and comparative microarray analysis of serotype M18 group A Streptococcus strains associated with acute rheumatic fever outbreaks. *Proceedings of the National Academy of Sciences of the United States of America* 2002;99(7):4668-73. doi: 10.1073/pnas.062526099 [published Online First: 2002/03/28]
48. Bryant PA, Smyth GK, Gooding T, et al. Susceptibility to acute rheumatic fever based on differential expression of genes involved in cytotoxicity, chemotaxis, and apoptosis. *Infect Immun* 2014;82(2):753-61. doi: 10.1128/IAI.01152-13 [published Online First: 2014/01/31]
49. Sjoberg R, Mattsson C, Andersson E, et al. Exploration of high-density protein microarrays for antibody validation and autoimmunity profiling. *N Biotechnol* 2016;33(5 Pt A):582-92. doi: 10.1016/j.nbt.2015.09.002 [published Online First: 2015/09/30]
50. Just D, Manberg A, Mitsios N, et al. Exploring autoantibody signatures in brain tissue from patients with severe mental illness. *Transl Psychiatry* 2020;10(1):401. doi: 10.1038/s41398-020-01079-8 [published Online First: 2020/11/20]
51. McGregor R, Tay ML, Carlton LH, et al. Mapping Autoantibodies in Children With Acute Rheumatic Fever. *Frontiers in immunology* 2021;12:702877. doi: 10.3389/fimmu.2021.702877 [published Online First: 2021/08/03]
52. Xia J, Gill EE, Hancock RE. NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. *Nat Protoc* 2015;10(6):823-44. doi: 10.1038/nprot.2015.052 [published Online First: 2015/05/08]
53. De Roover K, Ceulemans E, Timmerman ME. How to perform multiblock component analysis in practice. *Behav Res Methods* 2012;44(1):41-56. doi: 10.3758/s13428-011-0129-1 [published Online First: 2011/07/13]
54. Wilms I, Croux C. Robust sparse canonical correlation analysis. *BMC Syst Biol* 2016;10(1):72. doi: 10.1186/s12918-016-0317-9 [published Online First: 2016/08/16]
55. Singh A, Shannon CP, Gautier B, et al. DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays. *Bioinformatics* 2019;35(17):3055-62. doi: 10.1093/bioinformatics/bty1054 [published Online First: 2019/01/19]
56. Wang B, Mezlini AM, Demir F, et al. Similarity network fusion for aggregating data types on a genomic scale. *Nat Methods* 2014;11(3):333-7. doi: 10.1038/nmeth.2810 [published Online First: 2014/01/28]
57. Carapetis JR, Currie BJ, Mathews JD. Cumulative incidence of rheumatic fever in an endemic region: a guide to the susceptibility of the population? *Epidemiol Infect* 2000;124(2):239-44. [published Online First: 2000/05/17]
58. Francis JR, Fairhurst H, Hardefeldt H, et al. Hyperendemic rheumatic heart disease in a remote Australian town identified by echocardiographic screening. *Med J Aust* 2020;213(3):118-23. doi: 10.5694/mja2.50682 [published Online First: 2020/07/08]
59. Rammelkamp CH, Wannamaker LW, Denny FW. The Epidemiology and Prevention of Rheumatic Fever. *Bull N Y Acad Med* 1952;28(5):321-34. [published Online First: 1952/05/01]
60. Wyber R, Noonan K, Halkon C, et al. Ending rheumatic heart disease in Australia: the evidence for a new approach. *Med J Aust* 2020;213 Suppl 10:S3-S31. doi: 10.5694/mja2.50853 [published Online First: 2020/11/16]

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

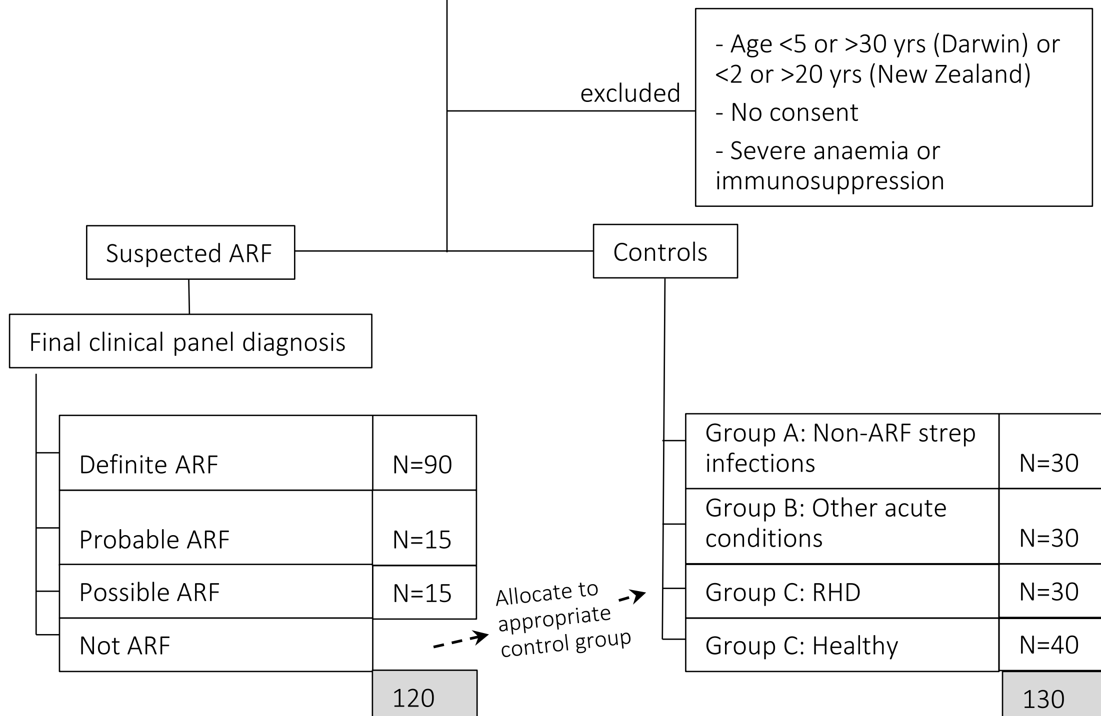
Figures

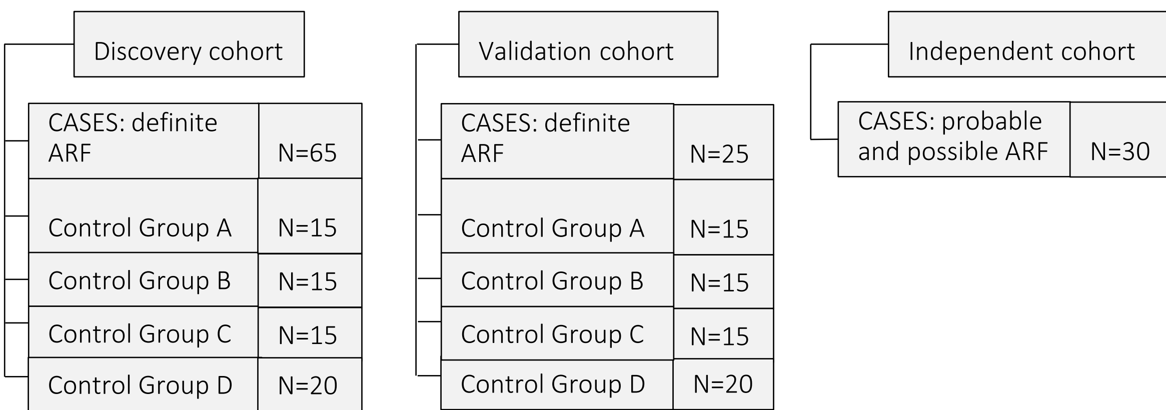
Figure 1: Study diagram showing target sample sizes

ARF = acute rheumatic fever; RHD = rheumatic heart disease

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml


Patients presenting to enrolment site or actively recruited through heathy volunteer drive
ENROLMENT AND DIAGNOSIS
excluded
- Age <5 or >30 yrs (Darwin) or <2 or >20 yrs (New Zealand)
- No consent
- Severe anaemia or immunosuppression
Suspected ARF
Controls
Final clinical panel diagnosis
Definite ARF N=90
Probable ARF N=15
Possible ARF N=15
Not ARF
120
Allocate to appropriate control group
Group A: Non-ARF strep infections N=30
Group B: Other acute conditions N=30
Group C: RHD N=30
Group C: Healthy N=40
130


ANALYSES
Discovery cohort
CASES: definite ARF N=65
Control Group A N=15
Control Group B N=15
Control Group C N=15
Control Group D N=20
Validation cohort
CASES: definite ARF N=25
Control Group A N=15
Control Group B N=15
Control Group C N=15
Control Group D N=20
Independent cohort
CASES: probable and possible ARF N=30