Risk factors for acute rheumatic fever: Protocol for a case-control study in New Zealand Supplementary material

Michael G Baker*, Jason Gurney, Jane Oliver, Nikki Moreland, Deborah Williamson, Nevil Pierse, Nigel Wilson, Tony R Merriman, Teuila Percival, Colleen Murray, Catherine Jackson, Richard Edwards, Lyndie Foster Page, Florina Chan Mow, Angela Chong, Barry Gribben, Diana Lennon

*Corresponding author: Department of Public Health, University of Otago, Wellington
Box 7343 Wellington, New Zealand 6242
Email: michael.baker@otago.ac.nz

Contents of supplement:

Supplementary Table S1: Incidence of ARF hospitalisations (numbers and annual rates per 100,000) by District Health Board, 2010–2014

District Health Board	Population 2013	Population <20 years	ARF hospitalisations, total no. 2010-14	First ARF hospitalisations, total no. (deduct readmissions)	First ARF hospitalisations, <20 years old, total no.	First ARF hospitalisations <20 years, annual average no.	First ARF hospitalisations <20 years % of NZ total	First ARF hospitalisations <20 years, annual average rate (/100,000 children)
Northland*	151692	42378	90	63	47	9.4	8.9	**22.2**
Waitemata*	525555	144942	80	53	42	8.4	8.0	**5.8**
Auckland*	436341	109014	80	65	47	9.4	8.9	**8.6**
Counties Manukau*	469293	150099	377	282	207	41.4	39.2	**27.6**
Waikato*	359310	103344	78	63	50	10	9.5	**9.7**
Lakes*	98187	28773	37	27	16	3.2	3.0	**11.1**
Bay of Plenty*	205995	56508	49	35	25	5	4.7	**8.8**
Tairawhiti*	43653	13782	15	14	10	2	1.9	**14.5**
Hawkes Bay*	151692	43116	24	20	13	2.6	2.5	**6.0**
Taranaki	109752	30015	6	5	2	0.4	0.4	**1.3**
Whanganui	60120	16407	8	6	4	0.8	0.8	**4.9**
Midcentral	162564	44799	12	9	8	1.6	1.5	**3.6**
Wairarapa	41112	10752	2	2	1	0.2	0.2	**1.9**
Hutt Valley*	138378	38517	38	33	23	4.6	4.4	**11.9**
Capital and Coast*	283704	73626	32	29	19	3.8	3.6	**5.2**
North Island Total	**3237348**	**906072**	**928**	**706**	**514**	**102.8**	**97.3**	**11.3**
Nelson Marlborough	136995	33891	5	5	3	0.6	0.6	**1.8**
West Coast	32148	7974	2	2	1	0.2	0.2	**2.5**
Canterbury	482178	122784	21	17	8	1.6	1.5	**1.3**
South Canterbury	55626	13506	0	0	0	0	0	**0.0**
Southern	297423	77133	6	6	0	0	0	**0.0**
South Island Total	**1004694**	**255324**	**34**	**30**	**12**	**2.4**	**2.3**	**0.9**
Unknown	**-**	**-**	**2**	**2**	**2**	0.4	0.4	**-**
NZ total	**4242048**	**1161387**	**964**	**738**	**528**	**105.6**	**100**	**9.1**

* RFRF case-control study = 11 DHBs in North Island

Supplementary Table S2: Summary of previous ARF and RHD risk factor studies

This structured review used the following search strategy:

- Databases searched: Medline/Pubmed, Embase plus manual search of references
- Outcomes: Acute rheumatic fever, Rheumatic heart disease
- Study design: cohort, case-control, cross-sectional
- Exclusion: Ecological studies, Non-English language, No comparison group, No statistical analysis of risk factors

Author, year, ref	Design	Outcome	Place	Population	Age group	N	Risk factors or interventions	Effect : OR (CI), RR (CI) HR (CI), aHR (CI)
ARF								
Entine 1949 [1]	Case-control	ARF incident case	USA	Cases: Children diagnosed with ARF attending outpatient and ARF clinics, Controls: healthy children attending a periodontics appointment at Temple University School of Dentistry.	4-17yo	Cases: 100 Controls: 100	**Univariate:** Non-suppurative gingivitis Soft white tooth deposit/decalcification Discoloration (orange-brown) Average number of Carries/Restorations/ Extractions (CRE) were higher in ARF cases (13.2) compared with controls (8.1)	OR 11.82 (4.0-34.93) OR 2.75 (1.42-5.33) OR 9.60 (3.83-24.07) Statistical testing not reported
Grave 1957 [2]	Case-control	ARF incident case	Australia	Sydney children: Cases: diagnosed with ARF, Controls: out-patients with non-rheumatic conditions.	2-12 years	120 cases, 100 controls	**Univariate:** Breadwinner is unskilled labourer/ receiving social service benefit Substandard housing Sleeping space <300 cubic feet per person Damp house Poor sewage/drainage Low family income	"Stat. significant" "Stat. significant" "Stat. significant" "Stat. significant" "Stat. significant" "Stat. significant"

Author, year, ref	Design	Outcome	Place	Population	Age group	N	Risk factors or interventions	Effect : OR (CI), RR (CI) HR (CI), aHR (CI)
Coburn 1960 [3]	Cross-section	ARF incident case	USA	Families surveyed in the Bridgeport district to identify rheumatic children, cases verified by clinicians. Controls had no history of ARF/RHD.	7-15 years	Cross section 1039, cases 153, controls 886	**Univariate:** Low egg consumption Other dietary factors (low consumption of milk, protein vitamins A&C)	 p<0.01 NS
Adanja et al 1988 [4]*	Case-control	ARF incident case	Yugoslavia	Cases: ARF patients identified through reports to health administration within 1 year of their first attack. Controls: Healthy participants matched for age, sex, place of residence	Adults and children (95% participants <20 years old)	148 cases, 444 controls	**Univariate:** Living space <5m2 ≥ 2 people per room ≥ 2 people per bed Deteriorated house Damp house Other poor housing features Low education of mother Change in place of residence in last 5 years History of frequent sore throat Family history of ARF Parental unemployment	 RR 2.83 (1.19 - 6.71) RR 1.72 (1.08 - 2.72) RR 1.65 (1.02 - 2.66) RR 1.83 (1.12 - 2.98) RR 2.48 (1.34 - 4.61) RR 2.58 (1.38-4.83) RR 2.58 (1.38 - 4.83) RR 5.00 (1.52 - 7.93) RR 2.01 (1.41 - 2.89) RR 2.81 (1.68 - 4.69) RR 10.37 (5.31 - 20.24)
Bahr 1989 [5]	Case-control	ARF incident cases	Kuwait	Cases: ARF diagnosis based on the revised (1965) Jones Criteria. Controls: University and hospital staff	Cases: children. Control: Adults	Cases: 39 Controls: 90	**Univariate:** Vitamin D-binding protein Gc2 allele (2x more common in cases)	 p = 0.0024 (risk measure not given)
Vlajinac 1989 [6]*	Case-control	ARF incident case	Yugoslavia (same study populatio	Cases: ARF patients identified through reports to health administration	Adults and children (95%	148 cases, 444 controls	**Univariate:** ARF in family Under-nourishment	 RR 2.78 (1.67-4.63) RR 2.18 (1.33-3.58)

Author, year, ref	Design	Outcome	Place	Population	Age group	N	Risk factors or interventions	Effect : OR (CI), RR (CI) HR (CI), aHR (CI)
			n as above: Adanja, 1988)	within 1 year of their first attack. Controls: healthy participants matched for age, sex, place of residence	participants <20 years old)		>2 people per room Sharing bed Home dampness Low education of mother	RR 1.60 (1.05-2.44) RR 1.43 (1.04-2.13) RR 2.70 (1.31-4.28) RR 2.01 (1.18 - 3.41)
Vlajinac 1991 [7]*	Case-control	ARF incident case	Yugoslavia (same study population as above: Adanja, 1988	Cases: ARF patients identified through reports to health administration within 1 year of their first attack. Controls: healthy participants matched for age, sex, place of residence	Adults and children (95% participants <20 years old)	Cases 148, Controls 444	**Multivariate:** Home dampness Change in place of residence in last 5 years Body weight below normal Frequent sore throat Low education of mother Family history for ARF Space per person<5msq >2 people per room Sharing bed Unemployed parents	 RR 2.40 (1.26- 4.58) RR 3.62 (1.15-11.35) \ RR 1.42 (1.08- 1.86) RR 2.26 (1.49- 3.39) RR 2.52 (1.29- 4.92) RR 2.98 (1.68- 5.29) RR 1.72 (0.69- 4.25) RR 1.35 (0.61- 3.00) RR 1.04 (0.99- 1.09) RR 2.08 (0.85- 5.09)
Adanja 1991 [8] *	Case-control	ARF incident case	Yugoslavia (same study population as above: Adanja, 1988	Cases: ARF patients identified through reports to health administration within 1 year of their first attack. Controls: healthy participants matched for age, sex, place of residence	Adults and children (95% participants <20 years old)	Cases 148, Controls 444	**Univariate** Food consumption Meat <3 times/wk No milk No cheese No eggs Vegetables <3 times/wk No fruit Body weight: Below normal Above normal **Multivariate:**	 1.10 (0.73-1.66) 1.53 (0.91-2.57) 1.46 (0.83-2.57) 1.83 (0.65-5.13) 1.69 (0.56-5.13) 1.53 (0.64-3.65) 1.94 (1.23-3.07) 1.01 (0.67-1.53) 2.23 (1.09-4.53)

Author, year, ref	Design	Outcome	Place	Population	Age group	N	Risk factors or interventions	Effect : OR (CI), RR (CI) HR (CI), aHR (CI)
							Body weight below normal	1.28 (0.97-1.63)
							No milk in diet	
Thomas, 1994 [9]	Cross-section	ARF or scarlet fever history	United Kingdom	Farmworkers: Cases with history of ARF or scarlet fever (8); Controls with no history of these illnesses	16-75 years	Cases: 8 (ARF: 2, Scarlet Fever: 6) Controls: 396	**Univariate:**	
							Exposure to dairy cattle	RR 0.12 (0.02-0.99)
							Drinking raw milk	RR 0.10 (0.01-0.85)
Zaman 1997 [10] **	Case-control	ARF incident case	Banglades h	Cases with ARF, Controls with recent GAS infection	5-23 years	Cases: 44 Controls: 86	**Univariate:**	
							Low income	OR 2.37 (1.04-5.46)
							Substandard house	OR 2.93 (1.23-7.15)
							Low height for age	OR 2.68 (1.14-6.34)
							Low body weight for age	OR 1.36 (0.64-2.99)
							Small dwelling space	OR 2.14 [0.93-4.96
							Large family size	OR 2.03 (0.94-4.39)
							Age:	
							<9 years	OR 1.19 (0.48-2.89)
							15-19 years	OR 1.52 (0.56-4.01)
							>19 years	OR 1.52 (0.35-5.98)
							Female gender	OR 0.59 (0.28-1.23)
							Multivariate:	
							Substandard house	OR 3.18 (1.24-8.44)
							Family size >6 persons	OR 2.08 [0.88-5.041
							Low height for age	OR 2.68 (1.06-6.86)
Zaman 1998 [11] **	Case-control	ARF incident case	Banglades h	Cases diagnosed with ARF using Jones Criteria, Controls were hospitalised patients who did not meet the Jones criteria. Both cases and controls showed	5-20 years	Cases: 60 Controls: 104	**Univariate:**	
							Higher number of siblings	(p=0.001, t-test)
							Larger family size	(p=0.89, t-test)
							Sharing bedroom	(p=0.26, t-test)
							Higher number of people per room	(p=0.58, t-test)
							Less parental schooling	(p<0.0001, t-test) (p=0.002, t-test)

Author, year, ref	Design	Outcome	Place	Population	Age group	N	Risk factors or interventions	Effect : OR (CI), RR (CI) HR (CI), aHR (CI)
				evidence of recent β-haemolytic streptococci infection.			Low income	OR 3.83 (1.74-8.40)
							Low height for age	OR 2.41 (1.12-5.57)
							Low weight for age	OR 3.76 (l.87-7.89)
							Small upper arm circumference	OR 3.81 (1.95-7.63)
							Food intake low in:	OR 2.60(1.36-5.08)
							Egg	OR 1.67 (0.87-3.20)
							Milk	OR 1.33 (0.68-2.65)
							Beef	OR 2.62 (1.35-5.21)
							Mutton	OR 1.07 (0.56-2.06)
							Chicken	OR 1.98 (1.03-3.84)
							Fish	OR 2.29 (1.20-4.45)
							Pulses	OR 1.37 (0.71-2.66)
							Fruits	OR 3.15 (1.61-6.34)
							Parana	OR 1.02 (0.51-2.01)
							Ruli (bread)	OR 1.71 (0.99-3.27)
							Leafy vegetables	OR 0.28 (0.12-0.62)
							Other vegetables	
							Soybean	
							Multivariate:	OR 2.23 (0.97-5.53)
							Low height for age	OR 1.47(0.58-3.92)
							Low weight for age	OR 2.40 (1.04-5.77)
							Small upper arm circumference	
							Food intake low in:	OR 2.29 (1.01-5.27)
							Egg	OR 1.55 (0.66-3.61)
							Milk	OR 1.11 (0.49-2.48)
							Beef	OR 1.05 (0.46-2.38)
							Mutton	OR 1.60 (0.69-3.74)
							Chicken	OR 1.08 (0.48-2.47)
							Fish	OR 1.48 (0.67-3.31)
							Pulses	OR 1.66 (0.74-3.74)
								OR 0.81 (0.34-1.88)

Author, year, ref	Design	Outcome	Place	Population	Age group	N	Risk factors or interventions	Effect : OR (CI), RR (CI) HR (CI), aHR (CI)
							Fruits	OR 2.15 (0.96-4.85)
							Parana	OR 0.96 (0.42-2.15)
							Ruli (bread)	OR 1.38 (0.63-3.03)
							Leafy vegetables	OR 0.42(0.15-1.13)
							Other vegetables	
							Soybean	
Zaman 1998 [12] **	Case-control	ARF incident case	Bangladesh	Cases diagnosed with ARF using Jones Criteria, Controls were hospitalised patients who did not meet the Jones criteria. Both cases and controls showed evidence of recent β-haemolytic streptococci infection.	5-20 years	Cases: 44 Controls: 44	**Univariate:**	
							Higher total protein	OR 1.06 (0.98 - 1.15)
							Higher albumin	OR 0.72 (0.59 - 0.88)
							Higher cholesterol (total)	OR 0.80 (0.41 - 1.57)
							Higher HDL	OR 0.05 (0.01 - 0.49)
							Higher LDL	OR 1.44 (0.65 - 3.22)
							Higher Triglycerides	OR 0.50 (0.15 - 1.73)
							Higher Haemaglobin	OR 0.94 (0.90 - 0.99)
							Higher Packed cell Volume	OR 0.73 (0.59 - 0.91)
							Higher Iron	OR 0.81 (0.71 - 0.93)
							Higher Total iron binding capacity	OR 0.95 (0.91 to 1.00)
							Higher Transferrin saturation	OR 0.89 (0.83 to 0.96)
							Multivariate	
							Higher Total protein	OR 1.02 (0.92 - 1.14)
							Higher Albumin	OR 0.75 (0.60 - 0.95)
							Higher Cholesterol (total)	OR1.20 (0.46 - 3.16)
							Higher HDL	OR 0.38 (0.03 - 5.48)
							Higher LDL	OR 1.56 (0.51 - 4.74)
							Higher Triglycerides	OR 0.86 (0.15 - 4.82)
							Higher Haemaglobin	OR 0.96 (0.91 - 1.02)
							Higher Packed cell volume	OR 0.69 (0.50 - 0.97)
							Higher Iron	OR 0.82 (0.68 - 0.97)
								OR 0.96 (0.91 - 1.04)
								OR 0.90 (0.82 - 0.98)

Author, year, ref	Design	Outcome	Place	Population	Age group	N	Risk factors or interventions	Effect : OR (CI), RR (CI) HR (CI), aHR (CI)
							Higher total iron binding capacity	
							Higher transferrin saturation	
Berdeli 2004 [13] #	Case-control	ARF incident case	Turkey	Cases: Caucasian Turkish ARF patients meeting Jones Criteria at a University paediatric clinic. Controls: healthy child volunteers	Cases: 10.98 ±2.9 years, Controls : 8.71±1.3 (mean age ± SD)	Cases: 66, controls: 117	**Univariate:** FcgammaRIIA-R/H-131 polymorphisms:	
							RR genotype (Present in 66.7% cases, 39.3% controls)	OR 4.98 (1.81-13.70)
							HR heterozygosity (Present in 21.2% cases, 5.1% controls)	OR 3.09 (1.64-5.81)
Berdeli 2005 [14] #	Case-control	ARF incident case	Turkey	Cases: Caucasian Turkish ARF patients meeting Jones Criteria at a University paediatric clinic. Controls: healthy child volunteers.	Cases: 11.16±2.88 years, Controls : 8.71±1.3 (mean age ± SD)	Cases: 61, controls: 91	**Univariate:**	
							Arg753Arg genotype (TLR-2 polymorphism, present in 8.2% cases, 90.2% controls)	OR 0.01 (0.00-0.03)
							Arg753Gln genotype (TLR-2 polymorphism, present in 91.8% cases, 9.9% controls)	OR 100.00 (32.00–320.00)
							Presence of TLR-2 Gln allele	OR 16.0 (7.6–35.00)
Berdeli 2006 [15] #	Case-control	ARF incident case	Turkey	Cases: Caucasian Turkish ARF patients meeting Jones Criteria at a University paediatric clinic. Controls: healthy adult volunteers	Cases: 10.98 ±2.9 years, Controls : 27.86±6.8 years (mean age ± SD)	Cases: 66, controls: 107	**Univariate:** TNFalpha polymorphisms at -308	No association

Author, year, ref	Design	Outcome	Place	Population	Age group	N	Risk factors or interventions	Effect : OR (CI), RR (CI) HR (CI), aHR (CI)
Kurahara 2006 [16]	Case-control	ARF incident case	Hawaii	Cases: ARF cases Controls: Other heart condition at Paediatric cardiology clinic. All qualitied for Medicare	NS	Cases 26, Controls 41	**Univariate:** Polynesian ethnicity Polynesian and higher number of children per household More children in case bedroom Low parental education	 OR 4.80 (1.65-15.63) OR 6.53 (1.90-24.10) NS NS
Hounie 2007 [17]	Case-control	ARF incident case	Brazil	Cases: ARF outpatients attending a clinic and first degree relatives, Controls: orthopaedic outpatients attending a clinic and first degree relatives. University of Sao Paulo Medical School.	Case probands: Mean age 14.36 years (±4.60), Control probands: Mean age 11.51 ±3.29	Cases: 310, controls: 177	**Univariate:** Obsessive-compulsive spectrum disorders (ARF cases and first degree relatives)	 OR 2.21 (1.09-4.49)
Seixas 2008 [18]	Case-control	ARF incidence in family	Brazil	Cases: ARF cases or first degree family members presenting to RF Outpatient Clinic, Controls: patients or first degree family members presenting to Orthopedic Outpatient Clinic.	≥16 years old	Cases: 188 Controls:96	**Univariate:** Generalised anxiety disorder	 1.71 ($p<0.05$) 95%CI not supplied
Messias-Reason 2009 [19]	Case-control	ARF and RHD prevalent case	Brazil	Cases: History of ARF or RHD attending out-patients clinic Controls: healthy subjects from the	8-76 years	Cases: 244 (ARF=82, RHD=162) Controls: 420	**Univariate:** FCN2 haplotype AGA	 OR: 0.32 (0.13–0.77)

Author, year, ref	Design	Outcome	Place	Population	Age group	N	Risk factors or interventions	Effect : OR (CI), RR (CI) HR (CI), aHR (CI)
				same geographic region and socioeconomic background (not age matched).				
Walker 2011 [20]	Case-control	ARF incident case	South Africa	Cases: Index cases of ARF Controls: recruited from outpatient clinics at the time of venesection for other indications	Not stated in paper	Cases: 40 Controls: 47	**Univariate:** B-cells expressing D8/17	Mean difference between cases and controls cell % (not OR): 0.23 (95% CI: 0.02–0.43, p<0.05). Risk measure not given.
Col-Araz 2012 [21]	Case-control	ARF incident case	Turkey	Cases: ARF patients followed in the Pediatric Cardiology Clinic, Controls: healthy children referring to the Well-Child Outpatient Clinic for routine health checkups	5-15 years	Cases: 38, controls:40	**Univariate:** IFN- γ (+874) TT genotype IFN- γ (+874) gene T allele IFN- γ (+874) gene A allele	 OR 8.10 (2.41–27.27) OR 3.02 (1.57–5.79) OR 0.33 (0.17–0.64)
Riaz 2013 [22]	Case-control	ARF incident cases and RHD prevalent cases (echocardiography)	Bangladesh	Cases: Patients referred to National Centre for Control of Rheumatic Fever and Heart Disease, Controls: from same center or patients at Shaheed Suhrawardy Medical College Hospital with no ARF or RHD.	Not stated, no age limits implied	RF 103, RHD 102, Controls 207	**ARF risk factors:** **Multivariate:** Age >19 yo Sex (female) Urban residence Wall material (brick) Family size (>5 people) >2 siblings Higher family income Maternal illiteracy Maternal employment >3 people sharing living room Water supply (tubewell)	 OR 0.1 (0.03-0.1) OR 2.2 (1.1-4.3) OR 3.1 (1.2-8.4) OR 3.6 (1.6-8.1) OR 0.3 (0.2-0.7) OR 3.1 (1.5-6.3) OR 0.9 (0.4-1.8) OR 2.6 (1.2-5.8) OR 7.0 (2.0-24.2) OR 1.2 (0.6-2.4)

Author, year, ref	Design	Outcome	Place	Population	Age group	N	Risk factors or interventions	Effect : OR (CI), RR (CI) HR (CI), aHR (CI)
							Bed (floor)	OR 1.1 (0.4-2.6)
							Using toothpaste	OR 0.9 (0.3-3.4)
							Toothbrushing ≤1/day	OR 0.6 (0.2-1.7)
							Not tooth brushing after a meal	OR 0.9 (0.4-1.7)
							RHD risk factors (non-rheumatic controls):	OR 2.5 (1.0-6.3)
							Multivariate:	
							Age >19 yo	OR 0.1 (0.1-0.3)
							Sex (female)	OR 2.2 (1.2-4.2)
							Urban residence	OR 2.0 (1.2-7.0)
							Wall material (brick)	OR 2.8 (1.3-5.3)
							Family size (>5 people)	OR 0.5 (0.2-0.9)
							>2 siblings	OR 4.4 (2.2-8.7)
							Maternal illiteracy	OR 2.5 (1.2, 4.9)
							Maternal employment	OR 6.2 (2.1,18.4)
							>3 persons sharing living room	OR 1.9 (1.0-3.4)
							Water supply (tubewell)	OR 1.6 (0.7-3.6)
							Bed (floor)	OR 1.5 (0.5-4.1)
							Toothpaste	OR 0.7 (0.2-1.8)
							Toothbrushing ≤1/day	OR 1.3 (0.7-2.4)
							Not tooth brushing after a meal	OR 1.5 (0.6-3.9)
Thornley 2016 [23] ^	Cohort	ARF or RHD incident case	Auckland, NZ	20,333 children in Auckland who were free of RHD at enrolment were followed for a mean of 5 years, cases hospitalised with ARF or RHD, controls remained free of ARF and RHD.	Participants aged 5-6 years at first dental visit.	Cases 96, Controls 20,237	**Multivariate:** 5+ primary teeth affected by caries	aHR 1.57 (1.20 -2.06)

Author, year, ref	Design	Outcome	Place	Population	Age group	N	Risk factors or interventions	Effect : OR (CI), RR (CI) HR (CI), aHR (CI)
Thornley 2018 [24] ^	Cohort	ARF or RHD incident case	Auckland, NZ	213,957 children free of RHD at baseline, mean follow-up time 5.1 years. Cases ARF or RHD, controls remained free of ARF and RHD.	Participants aged 3-12 years at first dental Visit.	Cases 440, controls 213,517	**Multivariate** Children diagnosed with scabies during hospital admissions	aHR 8.98 (6.33 -20.2)
RHD								
Caughey 1975 [25]	Case-control	ARF or RHD prevalent case	NZ	Cases: 50 Maori and 50 Europeans with ARF or RHD Controls: 75 Maori and 514 European disease-free blood donors		50 Maori and 50 Europeans with ARF/RHD compared with each control group	**Univariate:** European case: HL-A28 reduced HL-A17 increased Maori cases: minor differences in frequency of HL-A3 increased HL-A8 increased HL-A10 decreased	RR 0.05 (0.00-0.90) RR 4.55 (2.12-9.77) RR 7.78 (0.37-165.63) RR 6.43 (0.70-59.37) RR 0.09 (0.01-0.70)
McLaren 1975 [26]	Cross sectional	RHD prevalent case (clinical)	South Africa	Pre-school and school children (Soweto). Cases: RHD diagnosed using clinical auscultation screening, Controls had no RHD on clinical auscultation.	2-18 years	Cross section 12,050 total children, RHD cases: 80, controls: 11,970	**Univariate:** Siblings >3 Language group School area Local area prevalence of pharyngeal GAS carriage Socioeconomic status No. people sharing bedroom	p<0.05 p>0.05 p>0.05 p>0.05 p>0.05 p=0.05
Anabwani 1989 [27]	Cross-sectional	RHD prevalent case	Kenya	School children, semi-rural area (Emuhaya). Cases: Identified as having RHD using	5-15 yo	Cross section 3631, RHD 6, Controls 3625	**Univariate** None identified (including socioeconomic status, family size, no. children per bedroom).	Nil

Author, year, ref	Design	Outcome	Place	Population	Age group	N	Risk factors or interventions	Effect : OR (CI), RR (CI) HR (CI), aHR (CI)
				echocardiography screening, Control: No RHD on echocardiography screening.				
Coggon 1993 [28]	Cohort (retro-spective)	RHD mortality during 1951 to 1989	UK	Chesterfield township address in 1936 housing survey and 1939 census. Cases identified in mortality records using ICD-9 coding to identify RHD as cause of death, Controls: RHD not listed as cause of death	No age limits implied	Cases: 76 Controls: 8,062	**Univariate** 1.5-2.49 persons per bedroom (compared with <1.5) >2.49 persons per bedroom (compared with <1.5)	RR 0.9 (0.5-1.6) RR 0.7 (0.3-1.6)
Longo-Mbenza 1998 [29]	Cross-sectional	RHD prevalent case (echocardiography)	Democratic Republic of Congo	Children living in Kinshasa town and adjoining slums. Cases: Identified as having RHD using echocardiography screening, Control: No RHD on echocardiography screening.	5-16 years	Cross section: 4848 children; RHD cases 68, controls 4780.	**Multivariate:** Birth in rainy season (vs. dry season) Low birthweight Low SES vs higher SES Low BMI >8 in household Migrant (vs Kongo native)	RR 2.2 (0.97–4.90) RR 1.81 (1.04–3.15) RR 2.68 (1.43–5.01) RR 2.64 (1.48–4.70) RR 4.10 (1.70–9.85) RR 4.79 (2.14–10.68)
Oli 1999 [30]	Cross-sectional	RHD prevalent case	Ethiopia	School children in an urban area. Cases: Identified as having RHD using echocardiography screening, Control: No RHD on echocardiography screening.	10-15yo	Cross section 9378, RHD cases 60, Controls: 9,318	**Univariate:** Sex (female) Low SES and Crowding conditions at home, in schools and in the bedrooms were not associated with risk of RHD after adjusting for confounders	OR 1.76 (1.01 – 3.06)

Author, year, ref	Design	Outcome	Place	Population	Age group	N	Risk factors or interventions	Effect : OR (CI), RR (CI) HR (CI), aHR (CI)
Rizvi 2004 [31]	Cross-sectional	RHD prevalent case	Pakistan	Rural population (Rahim Yaar Khan district) Cases: Identified as having RHD using echocardiography screening, Control: No RHD on echocardiography screening.	Not stated, no age limits implied, 50% pop. <15 yo.	Cross section 9430 screened, RHD cases 54, Controls: 9,376	**Univariate:** Older mean age (29.5 years cases/20.9 years controls) Female gender Home crowding (people per room) House construction House standard Lower SES Case education level	p<0.05 RR 1.86 (1.07-3.24) NS NS NS NS NS
Chou 2004 [32]	Case-control	RHD prevalent case	Taiwan	Cases: RHD confirmed by echocardiography Controls: age- and sex-matched unrelated healthy volunteers	26-80 years	Cases: 115, controls: 100	**Univariate:** ACE I/D II genotype ACE I allele	OR 2.12 (1.21-3.71) OR 1.50 (1.02-2.21).
Messias-Reason 2009 [19]	Case-control	RHD prevalent case	Brazil	Cases: Valvular RHD lesions confirmed on echocardiography in out-patients from Children's Cardiologic Unit and Cardiology Out-patient Clinic Controls: healthy subjects from the same geographic region and socioeconomic background.	8-76 years	Cases: 106 Controls: 210	**Univariate:** FCN2 haplotype GGA FCN2 haplotype AGA FCN2 haplotype AGG (compared with haplotype AAA)	OR 1.56 (1.10–2.30) OR 0.32 (0.13-0.77) OR 0.44 (0.23-0.82)

Author, year, ref	Design	Outcome	Place	Population	Age group	N	Risk factors or interventions	Effect : OR (CI), RR (CI) HR (CI), aHR (CI)
Steer 2009 [33]	Cross-sectional	RHD prevalent case	Fiji	Cases: RHD confirmed by echocardiography, Controls: No RHD on echocardiography.	5-15 years	Cross section 3462, RHD cases 359, Controls: 3103	**Univariate** Indigenous Fijian ethnicity Rural school Lower weight for age Lower height for age Lower BMI for age Older age **Multivariate** Female gender Indigenous Fijian ethnicity Rural school Impetigo Scabies	 RR 2.3 (0.97-5.6) RR 1.8 (1.0-3.1) p>0.05 p>0.05 p>0.05 p>0.05 RR 1.6 (0.8-3.2) RR 2.0 (0.8-5.1) RR 1.3 (0.7-2.4) RR 1.7 (0.7-4.2) RR 1.0 (0.4-2.4)
Azevedo 2010 [34]	Case-control	RHD prevalent case	Brazil	Cases: ARF patients who fulfilled the Jones criteria with features of RHD Controls: race-matched, healthy blood donors	Cases: 7-41 years, controls > 35 years old.	Cases 84, controls 84	**Univariate:** Polymorphism of the IL-1ra gene is a relevant factor for RHD severity	Allele 1 and genotype A1/A1 less frequent among patients with severe carditis compared with patients without this manifestation (OR=0.11, p=0.031; OR=0.092, p=0.017)
Saxena 2011 [35]	Cross-sectional	RHD prevalent case	India	School children living in rural Northern India. Cases: Identified as having RHD using echocardiography screening, Control: No RHD on echocardiography screening.	5-15 years	Cross section: 6270, RHD cases 128, Controls 6142	**Univariate:** Older age Female gender Studying in government funded school Substandard house Home crowding Height Weight <25th percentile, Low BMI	 p<0.001 p<0.001 p<0.005 p<0.005 p<0.05 p<0.05 p<0.05

Author, year, ref	Design	Outcome	Place	Population	Age group	N	Risk factors or interventions	Effect : OR (CI), RR (CI) HR (CI), aHR (CI)
							Waist circumference History of joint pain Auscultatory abnormalities **Multivariate:** Older age Female gender Government school Substandard house Home crowding	p<0.05 p<0.05 p<0.05 OR 1.93 (1.29-2.88) OR 1.84 (1.25 -2.72) OR 1.55 (1.02 - 2.34) OR 1.34 (0.84 – 2.17) OR 1.16 (0.75 - 1.78)
Ba-Saddik 2011 [36]	Cross-sectional	RHD prevalent case	Yemen	School children. Cases: Identified as having RHD using echocardiography screening, Control: No RHD on echocardiography screening	5-16 yo	Cross section 6000, RHD cases 219, controls 5781	**Univariate:** Older age Home crowding (>4 per bedroom) Low household income Poor housing conditions (ie. home not constructed with stone and lacked water supply)	 RRs not calc. but association tested p=0.001 p<0.001 p<0.001 p<0.001
Dobson 2012 [37]	Case-control (nested in cross-sectional study)	RHD prevalent case (echocardiogram)	Fiji	School children in rural and urban settings who participated in a previous screening study. Cases with documented RHD, Controls age & sex matched without RHD	5-15yo	Cases 80, Controls 80	**Univariate:** Trend toward increased risk of RHD and poor-quality housing, lower SES (but not stat. sig.) People in home People per bedroom Children per classroom Maternal unemployment Maternal education	 NS NS NS OR 2.6 (1.2–5.8) NS
Okello 2012 [38]	Case-control	RHD prevalent case	Uganda	Patients seen at the Mulago National Referral Hospital: cases were diagnosed with	5-60 years	Cases 243, Controls 243	**Univariate:** People per house >8 Space per person < 90 feet^2	 OR 1.98 (1.4–2.5) OR 8.3 (6.1–10.4)

Author, year, ref	Design	Outcome	Place	Population	Age group	N	Risk factors or interventions	Effect : OR (CI), RR (CI) HR (CI), aHR (CI)
				RHD, controls had normal echocardiograms.			Longer distance to nearest health unit	OR 1.48 (0.2–3.2)
							Education level (compared to <primary):	
							Primary	OR 0.38 (0.17–0.83)
							Secondary	OR 1.28 (0.59–2.74)
							Vocational	OR 1.47 (0.45–4.79)
							University	OR 2.19 (0.95–5.09)
							Unemployed	OR 3.09 (2.04–4.72)
							Income (compared with <25USD)	
							25-49.5USD	OR 1.95 (0.9–4.2)
							50-99.5USD	OR 2.19 (0.96–4.98)
							100USD+	OR 14.7 (5.96–36.1)
							Male gender	OR 1.26 (0.82–1.95)
							Multivariate:	
							Space per person < 90 square feet	OR 1.35 (1.10–1.56)
							Longer distance to nearest health unit	OR 0.70 (0.61–0.87)
							Higher number of people living in the house	NS
							Education level (compared to <primary):	OR 0.57 (0.22–1.43)
							Primary	OR 1.91 (0.75–4.88)
							Secondary	OR 1.8 (0.45–7.21)
							Vocational	OR 2.93 (1.04–8.19)
							University	OR 1.71 (1.05–8.19)
							Unemployed	
Rehman 2013 [39]	Case-control	RHD prevalent case	Pakistan	Cases: patients with RHD, Controls: healthy individuals from similar ethnic groups and	Cases: mean age 30 ± 14.5 years,	150 RHD cases, 204 controls	TNF-alpha(-308):	
							G/G genotype	OR 0.39 (0.20-0.76)
							G/A genotype	OR 1.97 (0.98–3.97)
							A/A genotype	

Author, year, ref	Design	Outcome	Place	Population	Age group	N	Risk factors or interventions	Effect : OR (CI), RR (CI) HR (CI), aHR (CI)
				geographic area as cases [aim to investigate the role of cytokine gene polymorphisms and their potential usefulness as biomarkers in RHD].	controls: 45 ± 12.7 years.		G allele A allele IL-6-174: G/G genotype G/C genotype C/C genotype G allele C allele IL-10–1082: G/G genotype G/A genotype A/A genotype G allele A allele IL-Ra VNTR: A1/A2 genotype A1/A3 genotype A2/A3 genotype A1 allele A2 allele A3 allele	OR 9.94 (1.21–217.31) OR 0.35 (0.20–0.64) OR 2.81 (1.55–5.14) OR 2.6 (1.17–5.85) OR 0.92 (0.56–1.52) OR 0.76 (0.48–1.20) OR 1.5 (1.04–2.16) OR 0.67 (0.46–0.96) OR 1.06 (0.66–1.71) OR 1.30 (0.81–2.07) OR 0.73 (0.46–1.17) OR 1.18 (0.87–1.61) OR 0.85 (0.62–1.15) OR 1.72 (0.92–3.24) OR 1.37 (0.34–5.58) Not calculable OR 0.52 (0.31–0.86) OR 1.50 (0.61–3.83) OR 0.0 (0.0–5.54)
Eriksson 2013 [40]	Cohort (retro-spective)	RHD hospitalisations and death	Helsinki, Finland,	Births in Helsinki maternity hospitals 1924-1944, cases identified using hospital and mortality records. Cases: Received a diagnosis of RHD, Controls: Not diagnosed with RHD.	25-80 yo	Cohort 20,431, RHD cases 101, Controls 20330	**Univariate:** Long umbilical cord (risk found for RHD mitral disease only) No. people living in home (compared with ≤3 people): 4 people 5 people ≥6 people People per room	HR 1.23 (1.04-1.45) HR 2.1 (1.1-4.0) HR 2.1 (1.1-4.2) HR 2.1 (1.1-4.3) NS

Author, year, ref	Design	Outcome	Place	Population	Age group	N	Risk factors or interventions	Effect : OR (CI), RR (CI) HR (CI), aHR (CI)
Mirabel 2015 [41]	Cohort (prospective)	RHD persistence in an RHD cohort compared with controls	New Caledonia	4th grade school children, RHD cases diagnosed using echocardiography, controls selected randomly from classmates without RHD, matched for ethnicity and classroom	9-10 years	RHD cases 114, Controls 227	**Univariate** Number of siblings House construction Mother's education Maternal employment Male gender Oceanic ethnicity Lived >1year out of household Usual mode of transport: private car Number of people per bedroom >2 **Multivariate :** ≥3 per bedroom Mother's education:	 p=0.2 p=0.9 p=0.048 p=0.4 p=0.3 p=0.9 p=0.2 p=0.4 p=0.003 OR 8.27 (1.67–41.08) OR 1.97 (0.69–5.64)
Azevedo 2016 [42]	Case-control	RHD prevalent case	NZ	Cases: confirmed RHD and of Māori and Pacific ancestry, controls: Polynesian/Maori ancestry and no autoimmune condition	Cases aged 3-32 years old at first ARF presentation, control age range not stated	204 RHD cases, 116 controls	**Univariate:** Variant of IL6 promoter (rs1800797 (-597G/A)) with RHD IL1RN variant (rs447713) with the severity of carditis	GG genotype 6.09 (CI 1.23; 30.23) times more likely to develop RHD than carriers of AA genotype (P=0.027). G allele (GG plus AG genotype) for IL1RN rs447713 SNP had 2.36 times (CI 1.00; 5.56) more severe carditis than those without this allele (the AA genotype) (P=0.049)
Parks 2017 [43]	GWAS meta-analyses	RHD prevalent case	Eight Oceanian countries	Melanesians (607 cases and 1,229 controls);	Age limits not	2,852 individuals.	**GWAS meta-analysis:** Immunoglobulin heavy chain locus - IGHV4-61	Each copy of IGHV4-61*02 associated with a

Author, year, ref	Design	Outcome	Place	Population	Age group	N	Risk factors or interventions	Effect : OR (CI), RR (CI) HR (CI), aHR (CI)
				Polynesians, South Asians and Mixed or other populations (399 cases and 617 controls)	stated (none implied)		gene segment (IGHV4-61*02 allele)	1.4-fold increase in the risk of RHD (OR 1.43, 1.27-1.61, P=4.1 x 10-9)
Gray 2017 [44]	GWAS	RHD prevalent case	Australia	Aboriginal population	≥18 years old	1263 total, 398 RHD cases; 865 controls	**GWAS analysis:** HLA-DQ locus was strongest genetic marker associated with RHD, eg HLA-DQA1*0101_ DQB1*0503	OR = 1.44, CI = 1.09–1.90, P = 9.56×10^{-3}

The following groups of papers all drew on the same or overlapping study populations:

* Adanja 1988, Vlajinac 1989, Vlajinac 1991, Adanja 1991

** Zaman 1997, Zaman 1998, Zaman 1998

Berdeli 2004, Berdeli 2005, Berdeli 2006

^ Thornley 2016, Thornley 2018

Abbreviations: NS=Not Significant/No association

OR=Odds Ratio, RR=Relative Risk, HR=Hazard Ratio

GWAS=Genome-Wide Association Study

Table S3: Factors investigated in the ARF Risk Factors Study

Category of risk and protective factors	Examples of variables including in the study	Variables included in the primary analysis
A. Preceding infection		
Potential GAS infections	• Preceding sore throat • Preceding skin infection • Preceding scabies • Skin cuts, grazes, wounds, insect bites	• Not included as these infections are considered a direct cause of ARF
B. Environmental risk factors		
High levels of social contact	• Attendance at social gatherings outside own home • School size (number of pupils) • Overseas travel and contact with overseas visitors	• Social gatherings outside home (composite of 9 Qu.)
Household crowding	• Occupancy (people per house) • Density (people per room, people per 100m^2) • Bedroom deficit (Canadian National Occupancy Standard (CNOS)) • Self-assessed 'overcrowding' and bedroom deficit	• Structural household crowding (people per room)
Functional crowding and bedroom crowding	• Functional crowding (eg sleeping in same room as others just to keep warm) • Sleeping in communal area (eg living room) • Sleeping with excess (≥ 2people) in bedroom	• Functional crowding (eg sleeping in same room as others just to keep warm)
Bed sharing and lack of standard bed	• Bed sharing with one or more others • 'Hot bedding' (ie using bed others have slept in) • Sleeping on floor or couch	• Bed sharing
Exposure to others with potential GAS	• Others in household with sore throat, cough, skin infection, scabies	• Not included

Household washing and laundry resources	• Lack of hot water for washing (delaying or having cold/lukewarm bath/shower) • Availability and use of bath or shower • Soap, own towel • Laundry facilities, frequency bedding changed • Regular swimming	• Lack of hot water for washing (composite of 2 Qu.)
House conditions	• Self-assessed house condition (eg poor or very poor) • Age of house (eg build prior to 1980)	• Not included
House indoor environment, including cold, damp, mould and contributing factors	• Self-assessed cold • Self-assessed damp, mould • Insulation, Heating • Power disconnections • Ventilation, unflued gas heaters	• Cold (composite of 4 Qu.) • Damp and mould (composite of 3 Qu.)
Tobacco smoke exposure	• Living with smokers (mother, other household members) • Exposure in house/car • Active smoker • Hair nicotine levels	• Living with a smoker
Exposure to animals and fleas	• Domestic animals (cats, dogs) • Fleas	• Not included

C. Healthcare assess		
Healthcare access	• Access barriers (timeliness, cost, transport, childcare) • Has a usual GP or medical centre	• Barriers to accessing primary healthcare (composite of 5 Qu)
Dental healthcare access	• Regular dental checks, timely visits, access problems • Has dental care provider • Frequency of teeth brushing • Personal toothbrush	• Not included
Sore throat and skin infection treatment	• Access to school-based RF Prevention Programme (RFPP)	• Attending school that provides RFPP

	• Consultation and treatment of recent sore throat / skin infection	
Healthcare knowledge (health literacy)	• Knowledge of ARF causes and consequences • Knowledge and reported management of key conditions (sore throat, skin infection)	• Not included
D. Health status and nutrition		
Personal health history and status	• Self-assessed general health • Asthma, eczema • Tonsils and adenoids removed • Frequency of previous hospitalisations, including infectious diseases (linked data)	• Not included
Oral health history and status	• Decayed, missing, and filled teeth (dmft/DMFT)(linked data) • Self-assessed oral health status • History of teeth filling and removal • Gum disease (bleeding after brushing)	• Decayed, missing, and filled teeth (dmft of primary teeth/DMFT of permanent teeth)
Dietary intake	• Sugar-sweetened drink intake • Fruit and vegetable intake • Takeaway food intake	• Sugar-sweetened drink intake
Nutrient status	• Vitamin D levels • Irons stores (Ferritin)	• Not included (only available on sample)
Body mass index (BMI)	• Height, weight, BMI	• Not included
E. Host socio-economic determinants		
Socioeconomic factors	• Area-based deprivation (NZDep)* • Caregiver Individual deprivation (NZiDep) • Caregiver education level • Household income	• NZDep* • NZiDep (caregiver)

Housing tenure and residential mobility	• Tenure (rented or owned) • Short housing duration (<1year) • Number of schools attended	• Tenure (rented or owned)
Location	• District Health Board (DHB)	• DHB*

F Predisposing host factors		
Demographic factors	• Age*, Sex*	• Age*, Sex*
Ancestry and genetic factors	• Ethnicity (prioritised)* • Ancestry (ethnicity of grandparents) • Born in a Pacific Island country • Specific genetic markers (eg HLA-DRB1)	• Ethnicity*
Family history of ARF	• Family history of ARF	• Family history of ARF
Immunological factors	• Family size, Birth order • Specific immunological markers	• Not included
Pregnancy and neonatal factors	• Pre-term delivery, Low birth weight • Low APGAR score	• Not included
Breast feeding	• History of breast feeding,	• Not included

F. Organism factors		
Characteristics of GAS organisms	• Characterisation (*emm* typing, whole genome sequence) of GAS organisms isolated from throat	• Not included (only available on sample)
Exposure to infectious cofactors	• Nasal detection of *S. aureus*	• Not included (only available on sample)

*Used for matched controls

Supplementary Table S4: Potential risks to study effectiveness and their management

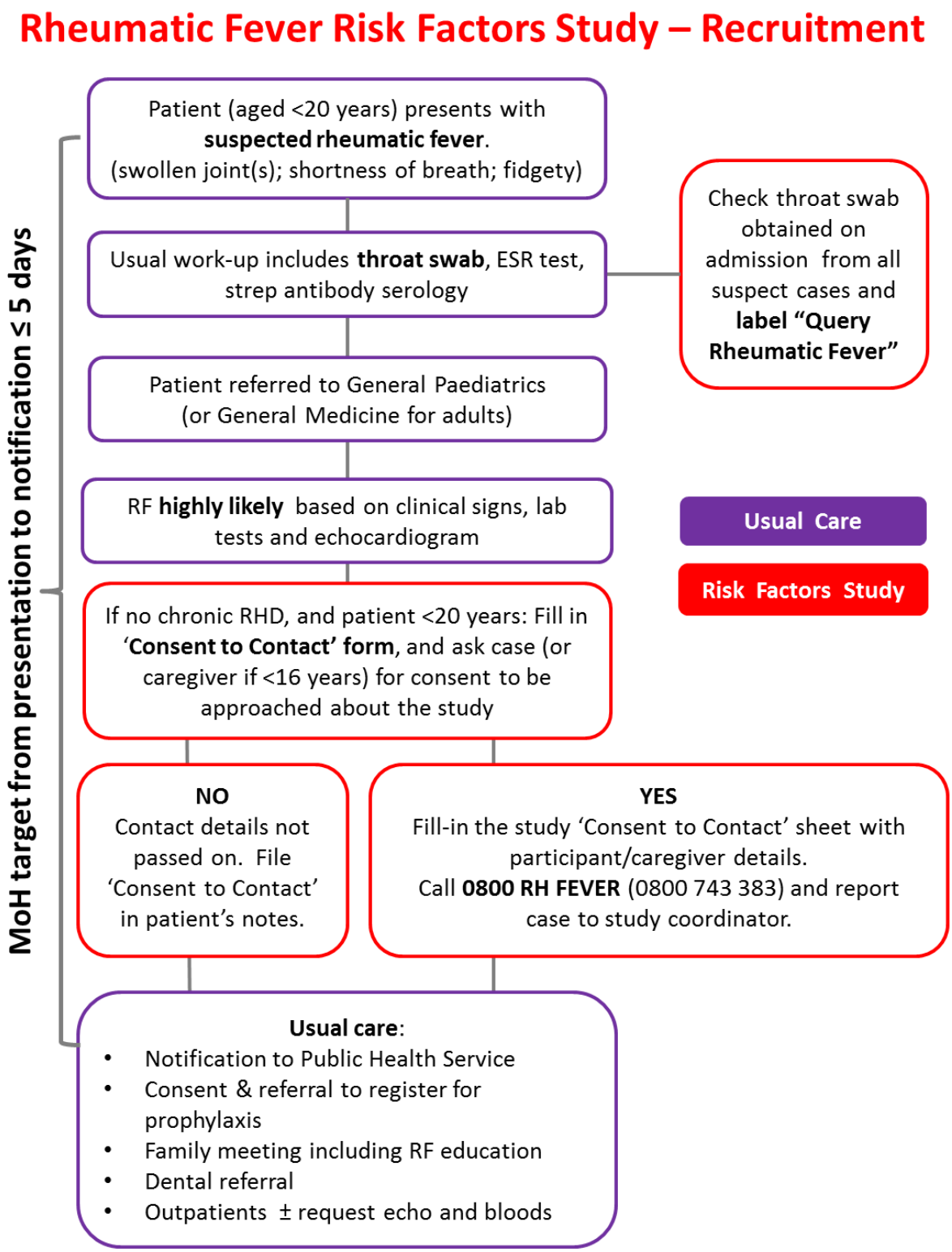
Potential Risk	Approach to Risk Management
1. **Missing key exposures from study design** - partly because of the complex nature of ARF as 2-stage autoimmune disease and lack of knowledge about its pathophysiology	• Extensive review of existing literature on likely risk factors for ARF • Knowledge of the biology of Streptococcal bacteria and their modes of transmission • Active review of study questionnaire and design by the multi-disciplinary research team
2. **Missing data** – at conclusion of study	• Direct data entry during interviews. • Built in systems for monitoring data completeness to identify and correct data gaps and improve processes • Optimise data collection and review completeness regularly, identify any items with high degree of missing data and take steps to improve completeness of collection • Consider analytic approaches to missing data.
3. **Problems accessing dental records** - .	• Early communication and engagement with appropriate dental service contacts. • Identification of data sources and logistics required to access records. • Careful statement of project objectives, confidentiality measures and ethical consent when approaching dental health providers.
4. **Difficulties with data analysis** -	• Establish a detailed data analysis plan. • Liaise with experienced epidemiologists. • Keep record of methods and steps in analysis including log of analytical code used

5.	**Insufficient cases** -	• Maximise study recruitment through engagement with clinicians and utilising public health services as back-up to avoid missing cases • Ensure study includes all DHBs with a moderate numbers of cases (eg 5+ expected during 2-year study period) • Extend study period and/or increase number of matched controls per case if required
6.	**Selection bias of cases** ie ARF cases included in study are not representative of ARF cases generally in terms of exposures	• Should be low as ARF is uncommon and all diagnosed cases are hospitalised, notified, and intensely scrutinised, and we anticipate high participation rates among identified cases. • Review case recruitment and if high proportion of eligible cases not participating investigate reasons and take steps to increase participation. • Collect data on non-participating eligible cases and compare characteristics of consenting cases with non-consenting case to see if there are any systematic differences.
7.	**Selection bias of controls** ie controls included in study are not representative of source population in terms of important exposures eg population controls recruited for NZHS may be a relatively more 'stable' population than is typical; matched controls may be somewhat atypical as they are recruited from NZHS participants who agree to being re-contacted, who are still	• NZHS control population is recruited by CBG using a "probability-proportional-to-size (PPS) sampling design" that aims to give a sample population that is representative of the NZ population. Characteristics of the NZHS control population will be compared with the NZ Census population to see if there are any systematic differences (in terms of characteristics such as ethnicity and deprivation). • Matched control population will be recruited from NZHS re-contact population in such a way as to have similar characteristics to cases for matched variables (age group, ethnicity, DHB, month of interview). Characteristics of these

	contactable, and who agree to being re-interviewed	matched controls will be compared with similar subjects in the original NZHS population to see if they differ in terms of important exposures (eg household crowding levels). • Consider use of quantitative bias analysis techniques if there is an indication of systematic selection bias in either control group.
8.	**Information bias– case/control misclassification** eg cases incorrectly diagnosed as ARF, controls incorrectly diagnosed as non-cases	• Because ARF is a serious illness with long-term consequences, considerable effort goes into establishing the diagnosis when suspected cases are hospitalised. As noted in the Protocol, all cases will undergo expert clinical review to assign them to a case category. Only 'definite' and 'probable' cases will be retained in the analysis, and a sensitivity analysis will assess the effect of excluding 'probable' cases. • Control misclassification as cases is unlikely as ARF is a rare disease, even in its sub-clinical forms. Controls will be specifically asked whether they have had ARF and will be excluded if they have.
9.	**Information bias– Poor recall** of some exposures by cases and controls (non-differential) especially early childhood exposures) or social desirability (eg smoking near children)	• Generally we will avoid asking about exposures that took places more than a few weeks previously. • Generally focus on asking about 'usual' exposures as recall is likely to be more accurate than asking about exposures at a specific time • Rely on linked health records for maternal and perinatal exposures
10.	**Information bias– Non reporting of some exposures**	• As above.

that are considered undesirable (eg smoking near children)	• Use laboratory testing to help validate recall (eg hair nicotine testing) in subset of subjects.
11. **Information bias – language difficulties and cultural differences** creating non-differential recall of exposures by cases and controls	• Using ethnically appropriate interviewers (i.e. matched ethnicity for Māori and Pacific participants) • Selection and training of interviewers to improve their interview technique
12. **Information bias - Differential recall** of exposures by cases and controls (eg differences in impact of case views on disease cause)	• Use identical questions and data collection methods for cases and controls • Minimise number of questions that ask about exposures prior to illness (for cases), which is a different period to controls (but inevitable for a small set of specific exposures) • Specifically ask subjects about their views on disease causality and use responses in sensitivity analysis (e.g. perform analyses that exclude cases who think a specific exposure is causal)
13. Information bias – Differential interview processes for cases and NZHS controls, eg cases/parents will usually be interviewed in hospital, but controls will be interviewed in their homes	• For children, specify a consistent approach to identifying interview subject (proxy) so that the type of person interviewed is similar across both groups (for children this will usually be the main caregiver in all instances). • Standard data collection methods and interviewer training about consistency of approach to cases and control regardless of setting.
14. **Information bias – Differential effort by interviewer** to identify exposures for cases and controls	• Interviewers cannot be blinded to case/control status • Use carefully constructed questions that focus on usual exposures

Supplementary Figure S1: Case recruitment algorithm used by clinicians



References

1. Entine, M., A survey of dental diseases as a diagnostic aid in rheumatic fever. *JAMA* **1949,** 38, (3), 303-8.
2. Grave, P. E., Social and environmental factors in the aetiology of rheumatic fever. *Med. J. Aust.* **1957,** 44, (18), 602-8.
3. Coburn, A. F., The concept of egg yolk as a dietary inhibitor to rheumatic susceptibility. *Lancet* **1960,** 1, (7129), 867-70.
4. Adanja, B.; Vlajinac, H.; Jarebinski, M., Socioeconomic factors in the etiology of rheumatic fever. *J. Hyg. Epidemiol. Microbiol. Immunol.* **1988,** 32, (3), 329-35.
5. Bahr, G. M.; Eales, L. J.; Nye, K. E.; Majeed, H. A.; Yousof, A. M.; Behbehani, K.; Rook, G. A., An association between Gc (vitamin D-binding protein) alleles and susceptibility to rheumatic fever. *Immunology* **1989,** 67, (1), 126-8.
6. Vlajinac, H.; Adanja, B.; Jarebinski, M., Socio-economic factors and rheumatic fever occurrence. Differences between patients with and without frequent sore throat. *J. Hyg. Epidemiol. Microbiol. Immunol.* **1989,** 33, (4), 471-6.
7. Vlajinac, H.; Adanja, B.; Marinkovic, J.; Jarebinski, M., Influence of socio-economic and other factors on rheumatic fever occurrence. *Eur. J. Epidemiol.* **1991,** 7, (6), 702-704.
8. Adanja, B. J.; Vlajinac, H. D.; Marinkovic, J. P.; Jarebinski, M. S., Rheumatic fever and diet. *Isr. J. Med. Sci.* **1991,** 27, (3), 161-3.
9. Thomas, D. R.; Salmon, R. L.; Kench, S. M.; Meadows, D.; Coleman, T. J.; Morgan-Capner, P.; Morgan, K. L., Zoonotic illness--determining risks and measuring effects: association between current animal exposure and a history of illness in a well characterised rural population in the UK. *J. Epidemiol. Community Health* **1994,** 48, (2), 151-5.
10. Zaman, M. M.; Yoshiike, N.; Chowdhury, A. H.; Jalil, M. Q.; Mahmud, R. S.; Faruque, G. M.; Rouf, M. A.; Haque, K. M.; Tanaka, H., Socio-economic deprivation associated with acute rheumatic fever. A hospital-based case-control study in Bangladesh. *Paediatr. Perinat. Epidemiol.* **1997,** 11, (3), 322-32.
11. Zaman, M. M.; Yoshiike, N.; Chowdhury, A. H.; Nakayama, T.; Yokoyama, T.; Faruque, G. M.; Rouf, M. A.; Haque, S.; Tanaka, H., Nutritional factors associated with rheumatic fever. *J. Trop. Pediatr.* **1998,** 44, (3), 142-7.
12. Zaman, M. M.; Yoshiike, N.; Rouf, M. A.; Haque, S.; Chowdhury, A. H.; Nakayama, T.; Tanaka, H., Association of rheumatic fever with serum albumin concentration and body iron stores in Bangladeshi children: case-control study. *BMJ* **1998,** 317, (7168), 1287-8.
13. Berdeli, A.; Celik, H. A.; Ozyurek, R.; Aydin, H. H., Involvement of immunoglobulin FcgammaRIIA and FcgammaRIIIB gene polymorphisms in susceptibility to rheumatic fever. *Clin. Biochem.* **2004,** 37, (10), 925-9.
14. Berdeli, A.; Celik, H. A.; Ozyurek, R.; Dogrusoz, B.; Aydin, H. H., TLR-2 gene Arg753Gln polymorphism is strongly associated with acute rheumatic fever in children. *J. Mol. Med. (Berl.)* **2005,** 83, (7), 535-41.
15. Berdeli, A.; Tabel, Y.; Celik, H. A.; Ozyurek, R.; Dogrusoz, B.; Aydin, H. H., Lack of association between TNFalpha gene polymorphism at position -308 and risk of acute rheumatic fever in Turkish patients. *Scand. J. Rheumatol.* **2006,** 35, (1), 44-7.
16. Kurahara, D. K.; Grandinetti, A.; Galario, J.; Reddy, D. V.; Tokuda, A.; Langan, S.; Tanabe, B.; Yamamoto, K. S.; Yamaga, K. M.; Kurahara, D. K.; Grandinetti, A.; Galario, J.; Reddy, D. V.; Tokuda, A.; Langan, S.; Tanabe, B.; Yamamoto, K. S.; Yamaga, K. M., Ethnic differences for developing rheumatic fever in a low-income group living in Hawaii. *Ethn. Dis.* **2006,** 16, (2), 357-61.
17. Hounie, A. G.; Pauls, D. L.; do Rosario-Campos, M. C.; Mercadante, M. T.; Diniz, J. B.; De Mathis, M. A.; De Mathis, M. E.; Chacon, P.; Shavitt, R. G.; Curi, M.; Guilherme, L.; Miguel, E.

C., Obsessive-compulsive spectrum disorders and rheumatic fever: a family study. *Biol. Psychiatry* **2007,** 61, (3), 266-72.
18. Seixas, A. A.; Hounie, A. G.; Fossaluza, V.; Curi, M.; Alvarenga, P. G.; De Mathis, M. A.; De Mathis, M. E.; Vallada, H.; Pauls, D.; de Braganca Pereira, C. A.; Miguel, E. C., Anxiety disorders and rheumatic Fever: is there an association? *CNS Spectr* **2008,** 13, (12), 1039-46.
19. Messias-Reason, I. J.; Schafranski, M. D.; Kremsner, P. G.; Kun, J. F., Ficolin 2 (FCN2) functional polymorphisms and the risk of rheumatic fever and rheumatic heart disease. *Clin. Exp. Immunol.* **2009,** 157, (3), 395-9.
20. Walker, K. G.; Cooper, M.; McCabe, K.; Hughes, J.; Mathiassen, W.; Lawrenson, J.; Wilmshurst, J. M., Markers of susceptibility to acute rheumatic fever: the B-cell antigen D8/17 is not robust as a marker in South Africa. *Cardiol. Young* **2011,** 21, (3), 328-33.
21. Col-Araz, N.; Pehlivan, S.; Baspinar, O.; Oguzkan-Balci, S.; Sever, T.; Balat, A., Role of cytokine gene (IFN-gamma, TNF-alpha, TGF-beta1, IL-6, and IL-10) polymorphisms in pathogenesis of acute rheumatic fever in Turkish children. *Eur. J. Pediatr.* **2012,** 171, (7), 1103-8.
22. Riaz, B. K.; Selim, S.; Karim, M. N.; Chowdhury, K. N.; Chowdhury, S. H.; Rahman, M. R., Risk factors of rheumatic heart disease in Bangladesh: a case-control study. *J. Health Popul. Nutr.* **2013,** 31, (1), 70-7.
23. Thornley, S.; Marshall, R. J.; Bach, K.; Koopu, P.; Reynolds, G.; Sundborn, G.; Ei, W. L., Sugar, dental caries and the incidence of acute rheumatic fever: a cohort study of Maori and Pacific children. *J. Epidemiol. Community Health* **2016**.
24. Thornley, S.; Marshall, R.; Jarrett, P.; Sundborn, G.; Reynolds, E.; Schofield, G., Scabies is strongly associated with acute rheumatic fever in a cohort study of Auckland children. *J. Paediatr. Child Health* **2018,** 54, (6), 625-632.
25. Caughey, D. E.; Douglas, R.; Wilson, W.; Hassall, I. B., HL-A antigens in Europeans and Maoris with rheumatic fever and rheumatic heart disease. *J. Rheumatol.* **1975,** 2, (3), 319-22.
26. McLaren, M. J.; Hawkins, D. M.; Koornhof, H. J.; Bloom, K. R.; Bramwell-Jones, D. M.; Cohen, E.; Gale, G. E.; Kanarek, K.; Lachman, A. S.; Lakier, J. B.; Pocock, W. A.; Barlow, J. B., Epidemiology of rheumatic heart disease in black shcoolchildren of Soweto, Johannesburg. *BMJ* **1975,** 3, (5981), 474-8.
27. Anabwani, G. M.; Amoa, A. B.; Muita, A. K., Epidemiology of rheumatic heart disease among primary school children in western Kenya. *Int. J. Cardiol.* **1989,** 23, (2), 249-52.
28. Coggon, D.; Barker, D. J.; Inskip, H.; Wield, G., Housing in early life and later mortality. *J. Epidemiol. Community Health* **1993,** 47, (5), 345-8.
29. Longo-Mbenza, B.; Bayekula, M.; Ngiyulu, R.; Kintoki, V. E.; Bikangi, N. F.; Seghers, K. V.; Lukoki, L. E.; Mandundu, M. F.; Manzanza, M.; Nlandu, Y., Survey of rheumatic heart disease in school children of Kinshasa town. *Int. J. Cardiol.* **1998,** 63, (3), 287-94.
30. Oli, K.; Porteous, J., Prevalence of rheumatic heart disease among school children in Addis Ababa. *East Afr. Med. J.* **1999,** 76, (11), 601-5.
31. Rizvi, S. F.; Khan, M. A.; Kundi, A.; Marsh, D. R.; Samad, A.; Pasha, O., Status of rheumatic heart disease in rural Pakistan. *Heart* **2004,** 90, (4), 394-9.
32. Chou, H. T.; Tsai, C. H.; Tsai, F. J., Association between angiotensin I-converting enzyme gene insertion/deletion polymorphism and risk of rheumatic heart disease. *Jpn. Heart J.* **2004,** 45, (6), 949-57.
33. Steer, A. C.; Kado, J.; Wilson, N.; Tuiketei, T.; Batzloff, M.; Waqatakirewa, L.; Mulholland, E. K.; Carapetis, J. R., High prevalence of rheumatic heart disease by clinical and echocardiographic screening among children in Fiji. *J. Heart Valve Dis.* **2009,** 18, (3), 327-336.
34. Azevedo, P. M.; Bauer, R.; Caparbo Vde, F.; Silva, C. A.; Bonfa, E.; Pereira, R. M., Interleukin-1 receptor antagonist gene (IL1RN) polymorphism possibly associated to severity of rheumatic carditis in a Brazilian cohort. *Cytokine* **2010,** 49, (1), 109-13.

35. Saxena, A.; Ramakrishnan, S.; Roy, A.; Seth, S.; Krishnan, A.; Misra, P.; Kalaivani, M.; Bhargava, B.; Flather, M. D.; Poole-Wilson, P. P., Prevalence and outcome of subclinical rheumatic heart disease in India: the RHEUMATIC (Rheumatic Heart Echo Utilisation and Monitoring Actuarial Trends in Indian Children) study. *Heart* **2011,** 97, (24), 2018-22.
36. Ba-Saddik, I. A.; Munibari, A. A.; Al-Naqeeb, M. S.; Parry, C. M.; Hart, C. A.; Cuevas, L. E.; Coulter, J. B. S., Prevalence of rheumatic heart disease among school-children in Aden, Yemen. *Ann Trop Paediatrics* **2011,** 31, (1), 37-46.
37. Dobson, J.; Steer, A. C.; Colquhoun, S.; Kado, J., Environmental factors and rheumatic heart disease in Fiji. *Pediatr. Cardiol.* **2012,** 33, (2), 332-6.
38. Okello, E.; Kakande, B.; Sebatta, E.; Kayima, J.; Kuteesa, M.; Mutatina, B.; Nyakoojo, W.; Lwabi, P.; Mondo, C. K.; Odoi-Adome, R.; Juergen, F., Socioeconomic and environmental risk factors among rheumatic heart disease patients in Uganda. *PLoS One* **2012,** 7, (8), e43917.
39. Rehman, S.; Akhtar, N.; Saba, N.; Munir, S.; Ahmed, W.; Mohyuddin, A.; Khanum, A., A study on the association of TNF-alpha(-308), IL-6(-174), IL-10(-1082) and IL-1Ra(VNTR) gene polymorphisms with rheumatic heart disease in Pakistani patients. *Cytokine* **2013,** 61, (2), 527-31.
40. Eriksson, J. G.; Kajantie, E.; Phillips, D. I.; Osmond, C.; Thornburg, K. L.; Barker, D. J., The developmental origins of chronic rheumatic heart disease. *Am. J. Hum. Biol.* **2013,** 25, (5), 655-8.
41. Mirabel, M.; Fauchier, T.; Bacquelin, R.; Tafflet, M.; Germain, A.; Robillard, C.; Rouchon, B.; Marijon, E.; Jouven, X., Echocardiography screening to detect rheumatic heart disease: A cohort study of schoolchildren in French Pacific Islands. *Int. J. Cardiol.* **2015,** 188, 89-95.
42. Azevedo, P. M.; Merriman, T. R.; Topless, R. K.; Wilson, N. J.; Crengle, S.; Lennon, D. R., Association study involving polymorphisms in IL-6, IL-1RA, and CTLA4 genes and rheumatic heart disease in New Zealand population of Maori and Pacific ancestry. *Cytokine* **2016,** 85, 201-6.
43. Parks, T.; Mirabel, M. M.; Kado, J.; Auckland, K.; Nowak, J.; Rautanen, A.; Mentzer, A. J.; Marijon, E.; Jouven, X.; Perman, M. L.; Cua, T.; Kauwe, J. K.; Allen, J. B.; Taylor, H.; Robson, K. J.; Deane, C. M.; Steer, A. C.; Hill, A. V. S.; Pacific Islands Rheumatic Heart Disease Genetics, N., Association between a common immunoglobulin heavy chain allele and rheumatic heart disease risk in Oceania. *Nat Commun* **2017,** 8, 14946.
44. Gray, L. A.; D'Antoine, H. A.; Tong, S. Y. C.; McKinnon, M.; Bessarab, D.; Brown, N.; Remenyi, B.; Steer, A.; Syn, G.; Blackwell, J. M.; Inouye, M.; Carapetis, J. R., Genome-wide analysis of genetic risk factors for rheumatic heart disease in Aboriginal Australians provides support for pathogenic molecular mimicry. *J. Infect. Dis.* **2017**.