Check for updates

STUDY PROTOCOL

REVISED *Streptococcus pyogenes* carriage acquisition, persistence and transmission dynamics within households in The Gambia (SpyCATS): protocol for a longitudinal household cohort study

[version 2; peer review: 2 approved]

Edwin P. Armitage [1], Alex J. Keeley[1], Gabrielle de Crombrugghe [2],
Elina Senghore[1], Fatoumatta E. Camara[1], Musukoi Jammeh[1], Amat Bittaye[1],
Haddy Ceesay[1], Isatou Ceesay[1], Bunja Samateh[1], Muhammed Manneh[1],
Abdul Karim Sesay[3], Beate Kampmann [1], Adam Kucharski[4],
Thushan I. de Silva[1,5], Michael Marks [6-8], MRCG StrepA Study Group

[1]Vaccines and Immunity Theme, Medical Research Council Unit The Gambia at the London School of Hygiene & Tropical Medicine, Banjul, The Gambia
[2]Molecular Bacteriology Laboratory, Faculty of Medicine, Free University of Brussels, Brussels, Belgium
[3]Genomics Strategic Core Platform, Medical Research Council Unit The Gambia at the London School of Hygiene & Tropical Medicine, Banjul, The Gambia
[4]Centre for Mathematical Modelling of Infectious Diseases, London School of Hygiene and Tropical Medicine, London, WC1E 7HT, UK
[5]The Florey Institute and Department of Infection, Immunity and Cardiovascular Disease, Medical School, University of Sheffield, Sheffield, S10 2TN, UK
[6]Division of Infection and Immunity, University College London, London, WC1E 6BT, UK
[7]Department of Clinical Research, Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London, WC1E 7HT, UK
[8]Hospital for Tropical Diseases, University College London Hospital, London, NW1 2BU, UK

Abstract

Background

Streptococcus pyogenes (StrepA) causes a significant burden of disease globally from superficial infections to invasive disease. It is responsible for over 500,000 deaths each year, predominantly in low- and middle-income countries (LMIC). Superficial StrepA infections of the skin and pharynx can lead to rheumatic heart disease, the largest cause of StrepA-related deaths in LMIC. StrepA can also asymptomatically colonise normal skin and the pharynx (carriage), potentially increasing infection risk. *Streptococcus dysgalactiae subsp. equisimilis* (SDSE) carriage is also common in LMIC and may interact

Open Peer Review

Approval Status ✔✔

	1	2
version 2 (revision) 30 Oct 2023		✔ view
version 1 27 Jan 2023	✔ view	? view

1. **Shiranee Sriskandan** [ID], Imperial College London, Hammersmith Hospitals campus, London, UK

with StrepA. This study aims to investigate StrepA and SDSE carriage and infection epidemiology, transmission dynamics and naturally acquired immunity within households in The Gambia.

Methods

A longitudinal household observational cohort study will be conducted over one year. 45 households will be recruited from the urban area of Sukuta, The Gambia, resulting in approximately 450 participants. Households will be visited monthly, and available participants will undergo oropharyngeal and normal skin swabbing. Incident cases of pharyngitis and pyoderma will be captured via active case reporting, with swabs taken from disease sites. Swabs will be cultured for the presence of group A, C and G beta-haemolytic streptococci. Isolates will undergo whole genome sequencing. At each visit, clinical, socio-demographic and social mixing data will be collected. Blood serum will be collected at baseline and final visit. Oral fluid and dried blood spot samples will be collected at each visit. Mucosal and serum anti-StrepA antibody responses will be measured.

Outcome

This study will report StrepA and SDSE clinical epidemiology, risk factors, transmission dynamics, and serological responses to carriage and infection. Detailed social mixing behaviour will be combined with phylogenetic relatedness to model the extent of transmission occurring withing and between households. The study will provide data to help meet global strategic StrepA research goals.

Keywords
Streptococcus pyogenes, asymptomatic carriage, pharyngitis, pyoderma, longitudinal cohort study, transmission modelling, The Gambia, rheumatic heart disease

2. **Stephan Brouwer**, The University of Queensland, Saint Lucia, Australia

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Edwin P. Armitage (edwin.armitage@lshtm.ac.uk)

Author roles: Armitage EP: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Keeley AJ**: Conceptualization, Data Curation, Funding Acquisition, Investigation, Methodology, Project Administration, Visualization, Writing – Review & Editing; **de Crombrugghe G**: Conceptualization, Data Curation, Investigation, Methodology, Project Administration, Visualization, Writing – Review & Editing; **Senghore E**: Investigation, Methodology, Project Administration, Writing – Review & Editing; **Camara FE**: Investigation, Methodology, Project Administration, Writing – Review & Editing; **Jammeh M**: Investigation, Methodology, Project Administration, Writing – Review & Editing; **Bittaye A**: Investigation, Methodology, Project Administration, Writing – Review & Editing; **Ceesay H**: Investigation, Writing – Review & Editing; **Ceesay I**: Investigation, Writing – Review & Editing; **Samateh B**: Investigation, Writing – Review & Editing; **Manneh M**: Investigation, Project Administration, Writing – Review & Editing; **Sesay AK**: Methodology, Supervision, Writing – Review & Editing; **Kampmann B**: Conceptualization, Methodology, Supervision, Writing – Review & Editing; **Kucharski A**: Conceptualization, Methodology, Supervision, Writing – Review & Editing; **de Silva TI**: Conceptualization, Funding Acquisition, Methodology, Supervision, Writing – Review & Editing; **Marks M**: Conceptualization, Funding Acquisition, Methodology, Supervision, Writing – Review & Editing;

How to cite this article: Armitage EP, Keeley AJ, de Crombrugghe G *et al.* **Streptococcus pyogenes carriage acquisition, persistence and transmission dynamics within households in The Gambia (SpyCATS): protocol for a longitudinal household cohort study [version 2; peer review: 2 approved]** Wellcome Open Research 2023, **8**:41 https://doi.org/10.12688/wellcomeopenres.18716.2

First published: 27 Jan 2023, **8**:41 https://doi.org/10.12688/wellcomeopenres.18716.1

Introduction

Streptococcus pyogenes (Group A *Streptococcus*, StrepA) is a beta-haemolytic Gram positive bacterium that is a major cause of infectious disease burden globally, responsible for over 500,000 annual deaths[1–4]. It causes a wide spectrum of disease from superficial skin and pharynx infections to invasive disease, in addition to the immunological sequelae of acute rheumatic fever, rheumatic heart disease (RHD) and acute post-streptococcal glomerulonephritis[3,5]. Each year an estimated 1.8 million invasive StrepA, 111 million pyoderma and 616 million pharyngitis cases occur globally[2]. Most clinical StrepA infections are thought to occur in low- and middle-income countries (LMIC), though data from such countries is lacking[1,2,6,7]. Moreover, RHD, the most serious immunological consequence of StrepA infection, causes over 300,000 deaths each year, predominantly in LMIC[3], where diagnosis and surveillance is poor[3,8].

Despite the significant burden of StrepA disease and its immunological sequelae, there has been little focus on StrepA carriage and transmission in LMIC[8,9]. Furthermore, the understanding of the natural history of StrepA carriage, transmission and infection is limited[9,10]. A better understanding of carriage incidence, prevalence, persistence (duration), seasonal variation, transmission and the associated risk factors within high-disease burden settings in LMIC is crucial to design and implement interventions targeting StrepA in such countries.

The epidemiology of superficial StrepA infections in The Gambia is poorly understood. In 2018[11], a cross-sectional study in 1441 Gambian children under five years old found a high prevalence of bacterial pyoderma (17.4%), scabies infestation (15.9%), and of StrepA culture-positive pyoderma (8.8%). There was also a significant increase in pyoderma detected during the rainy season compared to before (8.9% vs. 23.1%, adjusted prevalence ratio 2.42, CI 1.39-4.23).

Whole genome sequencing (WGS) has transformed our ability to understand StrepA epidemiology, giving significantly better resolution than *emm* typing to determine linkage between strains. This has been used to gain valuable insights into transmission dynamics and to inform outbreak investigation in HICs[12–16]. In low-income settings where the molecular epidemiology of StrepA is notably different, combining WGS data, clinical and behavioural data with mathematical models can provide new insights into transmission dynamics and potential intervention strategies[17,18].

Rationale

In 2018, the World Health Assembly stated that RHD and StrepA research should be a global priority[19]. The WHO then published a Group A *Streptococcus* Vaccine Development Technology Roadmap highlighting key strategic areas for research including to improve global estimates of disease burden and epidemiology of StrepA infections, and to further describe the spectrum of natural disease history and immunity in longitudinal studies[9,10]. Our limited understanding of StrepA transmission dynamics and immunity is mostly derived from studies in high-income countries (HIC)[20–22]. However, in LMIC such as The Gambia, a higher prevalence and incidence of StrepA carriage and a wider diversity of the circulating *emm* types may underlie the higher burden of StrepA-related clinical infections and immune sequelae seen[23–25].

Very few longitudinal studies of StrepA exist[20,26,27], and high-quality longitudinal data from a high-prevalence country in sub-Saharan Africa combining classical epidemiology with detailed social mixing behaviour and next generation WGS techniques to model disease transmission will be highly informative in growing our understanding of StrepA epidemiology and meeting global strategic StrepA research goals on the road towards a StrepA vaccine.

Study objectives
Primary:

1. To determine the prevalence, incidence, duration and transmission dynamics of asymptomatic StrepA carriage and clinical StrepA infections within households.

2. To establish risk factors for pharyngeal and skin clinical StrepA infection, including detailed characterisation of the relationship with individual and household asymptomatic carriage, *emm* type and seasonality.

3. To develop a mathematical model of household StrepA transmission using clinical, behavioural and phylogenetic relatedness data to calibrate it, to allow for estimation of the relative contributions of between and within household transmission.

Secondary:

1. To determine risk factors for asymptomatic StrepA pharyngeal and skin carriage.

2. To describe the role of asymptomatic StrepA skin and pharyngeal carriage in StrepA transmission and infection.

3. To describe seasonal variation in StrepA carriage and clinical infection throughout the year.

4. To describe StrepA *emm* type diversity.

5. To investigate the extent of StrepA tissue tropism of *emm* types identified.

6. To determine the prevalence and incidence of *Streptococcus dysgalactiae subspecies equisimilis* (groups C and G streptococcus; SDSE) carriage and clinical infection.

7. To describe the prevalence, incidence and transmission dynamics of *Staphylococcus aureus* skin carriage and infection within households.

8. To describe variations in bacterial density by site, season and clinical characteristics using quantitative PCR.

9. To identify non-human sources of StrepA within households and the presence of airborne StrepA indoors using settle plates.

10. To describe the antimicrobial sensitivity of StrepA isolates identified.

11. To describe age-stratified anti-StrepA antibody titres.

12. To explore StrepA-specific serological and mucosal immune activity in response to colonisation and disease.

13. To investigate the relationship between anti-StrepA antibody titres and risk of incident colonisation and infection to explore serological correlates of protection.

Protocol
Study setting
The Gambia is a small country in West Africa with a population of approximately two million people. It was ranked 174th by the United Nations Human Development Index in 2021, making it one of the least developed countries in the world. It is a predominantly Muslim country, comprising several tribal groups, the largest being Mandinka, Wolof, Fula and Jola.

Sukuta is an urban area, part of the coastal region's sprawling conurbation, where most of the population live. It is a majority Mandinka area, with a population of 47,048, and an average household size of 8.1 in the census in 2013.

The climate is sub-tropical with a long dry season from November to May, and a short rainy season between June and October each year.

Study design
SpyCATS is a prospective, longitudinal (open) cohort study within households in Sukuta, The Gambia. Households will be recruited, and all household members present at the time of the visit will be asked to participate. Households will be followed for 12 months, with monthly visits, and more frequently for some subgroups of participants (described below).

A sample size target of 450 participants was determined (see below). With an average household size of approximately 10 people in Sukuta (based on our previous research and experience in the area given the census data is substantially out of date), 45 households will be recruited, and every available consenting household member included as an individual participant.

Selection of participants
The study will enrol participants as individuals within households. Households will be identified using a process of GPS random selection. No complete sampling frame of households exists for Sukuta, however geographic information system data exist from the 2013 census of The Gambia. These data will be utilised to obtain a random set of GPS sampling locations stratified by population density. A list of GPS coordinates for the locations will be identified and for each location and the nearest household will be approached for participation. Each location on the list will be approached in order until the desired sample size is reached. Households will only be enrolled if over 50% of household members consent to participate in the study.

For the purposes of enrolment in the study, a household will be defined according to The Gambia Demographic and Health Survey 2013 definition: "a household [is] defined as a person or a group of related or unrelated persons who live together in the same dwelling unit(s) or in connected premises, who acknowledge one adult member as the head of the household, and who have common arrangements for cooking and eating."

Inclusion criteria
Households must:

- Be within the boundary of Sukuta as determined by the 2013 census

- Have at least 3 members including at least one child under age 18

Individuals must:

- Provide signed (or thumb-printed) informed consent for study participation (obtained from a parent or guardian for children under the age of 18)

- Be willing and have capacity to participate and comply with the study protocol as judged by a member of the study team

- Be resident in the household, with no plans to move outside of the household during the period of study participation

Exclusion criteria
Households:

- Less than 50% of individuals living in the household, as defined by the The Gambia Demographic and Health Survey 2013 definition, provide consent to participate

Individuals:

- Consent not provided

- Has any condition or any other reason that may lead to difficulty or discomfort in obtaining all the necessary samples

- Is judged by the study team member to be unable or unlikely to participate and comply with the study protocol for the entire study period. Examples could include individuals with severe mental health conditions, communication barriers that cannot be overcome, or frequent absences from the household.

NB. There are no restrictions on enrolment of pregnant women or newly born babies.

Field activities
Overview. Households will be enrolled for 12 months covering both the dry and rainy seasons, with enrolment having commenced in July 2021. Every household will undergo an enrolment visit (MV0), then monthly visits (MV1, MV2, MV3 etc., up to MV12) unless practical constraints arise (see Figure 1). At each visit, the household size will be determined by the number of individuals who slept in the household the previous night, and those household members present will be asked to participate. Household members not available to be seen will be still allocated an ID number, in order to capture relevant information regarding their social mixing with other household members, and if they are present at later visits, they will be asked to consent and enrol. Participants whose baseline (enrolment) visit occurs after MV0 will be asked why they were not available previously. Reasons for missed visits and late enrolment will be captured.

At each visit an oropharyngeal swab (OPS), normal skin swab (NSS), oral fluid sample (OF) and dried blood spot (DBS) will be taken from all enrolled individuals present and data collected on socio-demographics, social mixing behaviour and clinical examination findings. In addition, a blood sample for serum (BS) will be taken at the beginning and end of the study for detailed functional immune responses. Wound swabs (WS) will also be taken from any pyoderma lesions and swabs taken from non-infected skin overlying scabies lesions (scabies skin swab, SSS). Additionally, environmental swabs (ES) will be collected from common touch points in the household and settle plates (SP) placed inside households. Throughout the study, enrolled participants reporting symptoms potentially consistent with StrepA infection will have an unscheduled visit including a physical examination, an appropriate swab (OPS or WS), OF and DBS. Swabs will be transported the
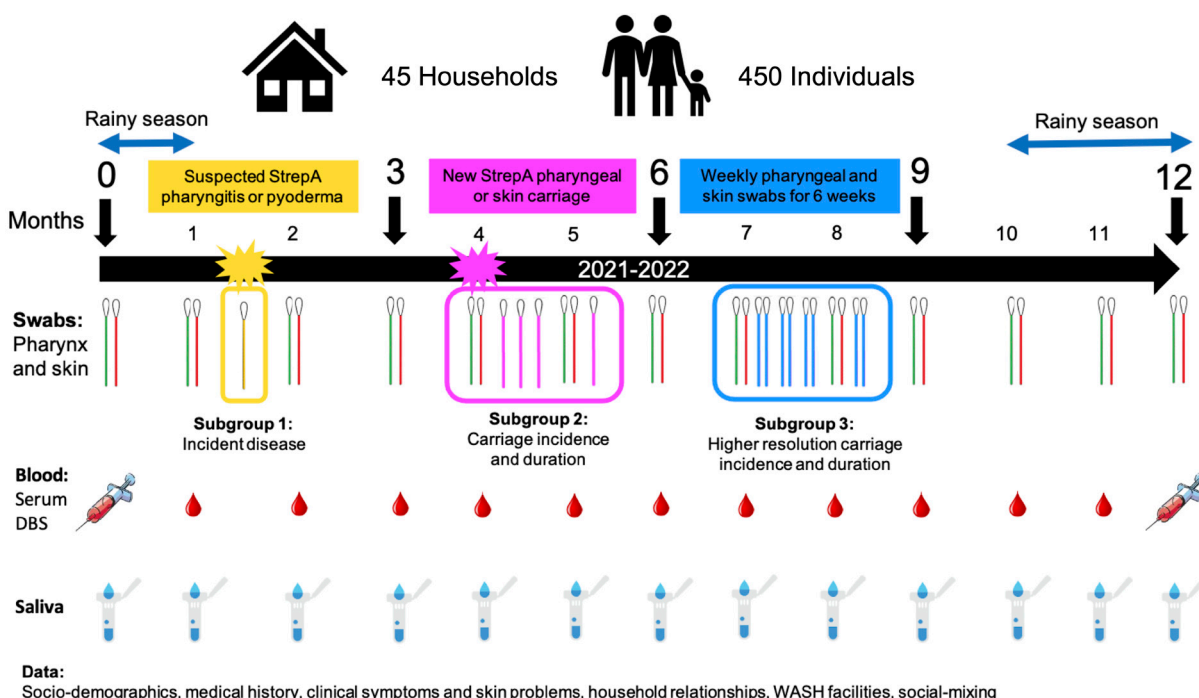

Figure 1. SpyCATS study diagram.

same day to the microbiology laboratory and plated for culture. The culture will identify the presence of catalase-negative, beta-haemolytic streptococci, and determine the group by latex agglutination testing. Antibiotic resistance will be determined, and any isolates will be stored in glycerol for later use. *Staphylococcus aureus* isolates cultured from NSS and WS will also be documented and stored for later characterisation. DBS and OF samples will be used for serological and mucosal immunology objectives. StrepA and SDSE isolates identified will undergo whole genome sequencing (WGS).

Enrolment visit. At the enrolment visit (MV0), each individual will undergo a baseline survey including participant's sociodemographic data, medical history, social mixing and behavioural factors. A physical examination will be carried out including a full body examination (taking care to maintain privacy) looking for any evidence of skin infections, and OPS and NSS will be collected. Additionally, an OF sample will be obtained, and venepuncture performed for BS and DBS. Any participant who is exhibiting signs and symptoms of pharyngitis (sore throat and pharyngeal inflammation) will be further examined and managed as they would be at an unscheduled visit for possible StrepA infection (see below). Any participant who is identified as having evidence of pyoderma (bacterial skin infection involving pus or crusts), will have a WS taken, and will be provided antibiotic treatment according to local guidelines as appropriate. Any additional abnormal finding requiring further investigation or treatment will be managed according to local practice or an appropriate referral made according to the nature of the finding.

Alongside the questionnaires on individuals, data will be collected on households such as household setup. Two ES will be taken from common touch points such as door handles and soft furnishings and a SP will be placed indoors to investigate airborne transmission.

Monthly visits (MV1, MV2, MV3 etc. up to MV12). Participants will be enrolled for 12 months undergoing an enrolment visit (MV0), then monthly visits (MV1, MV2, MV3 etc. up to MV12), though due to the open cohort design new participants can enrol at visits later than MV0 (this includes babies born during the study period if their parent/guardian consents).

At each monthly visit following enrolment, the study team will collect further survey data from each individual household member, collect an OPS, NSS, OF and DBS. Survey data collected will include socio-demographic information, social mixing, behavioural factors and clinical examination findings. Participants with any evidence of pyoderma will have a WS taken and will be offered antibiotic treatment; any other abnormal finding requiring further investigation or treatment will be managed according to WHO guidelines or best local practice or an appropriate referral made according to the nature of the finding. Additionally, any use of antibiotics or other medication, or any attendance at a healthcare setting since the previous visit will be recorded.

Unscheduled visits for possible StrepA infections. All enrolled households will be provided with an on-call number to call at any time when an enrolled household member is experiencing symptoms that could be compatible with a StrepA infection or any other acute illness. A study nurse will then complete a rapid assessment over the phone and arrange for an unscheduled visit as appropriate.

A study nurse will collect data on the history of the complaint, any medication taken and other relevant information. They will perform a physical examination, including vital signs under the supervision of a study clinician. If the symptoms are consistent with possible StrepA infection (pharyngitis or pyoderma), then a study swab will be taken (OPS or WS) for culture alongside an OF and DBS sample. For cases of pharyngitis and pyoderma, treatment will be provided empirically according to locally devised guidelines based on WHO and local guidelines and medication availability (see Table 1).

Participants presenting with symptoms of systemic infection or possible invasive disease will be urgently referred to the MRCG clinical services department for management. Participants presenting with other minor acute medical complaints will be provided appropriate treatment by the study team. All other medical presentations will be referred to the appropriate local healthcare facility.

Sub-studies

Intensified incident surveillance cohort. A subgroup of 16 randomly selected households will undergo intensified swabbing. This subgroup will be used to assess incidence and duration of StrepA and SDSE carriage and disease with greater resolution than the main cohort. During the second half of the study, the 16 households will undergo blocks of 6 weeks of intensified swabbing in batches of 4 households in turn, until all 16 have had a 6-week block of weekly swabbing. The

Table 1. SpyCATS empirical treatment guidelines for potential StrepA infections.

Diagnosis	Signs and symptoms	First-line treatment	Alternative treatment
Pyoderma	Purulent or crusted skin lesion	Cloxacillin (weight-based dose for children under 12 years; over 12 years 500mg qds 5–7 days)	Azithromycin (12 mg/kg up to max 500mg od 5 days)
Pharyngitis	Sore throat, pharyngeal/tonsillar erythema	Penicillin V (weight-based dose for children under 12 years; over 12 years 500mg qds 5–7 days)	Azithromycin (12 mg/kg up to max 500mg od 5 days)

households will be visited weekly for each of the 6 weeks of the intensive swabbing period. All household members present will undergo swabbing (OPS and NSS) in addition to more detailed social mixing behaviour data being collected. These data will be used in combination with WGS data from any StrepA isolated to inform the household transmission model.

Estimating duration of StrepA carriage. Following MV0, any participant who becomes an asymptomatic StrepA (or SDSE) carrier (i.e. was negative at baseline or the previous visit, and then becomes positive at a monthly visit without symptoms) will have weekly swabs taken from the same site that was positive (OPS or NSS), until two negative swabs have been received in a row.

Nested cross-sectional study of personal hygiene behaviour. At a MV11 or MV12, participants will be requested to undergo an additional survey on their personal hygiene behaviours during the last week including laundry, handwashing, bathing and soap and disinfectant use. Attitudes towards wound care and usual practice of participants in response to wounds will be captured.

At the same visit, additional environmental swabs (ES) will be collected from the household including four commonly touched locations within the household and a sample of water from the main household greywater source.

These data, combined with individuals' carriage and infection data from the wider study will be used to assess the relationship between individual- and household-level hygiene behaviours and StrepA, SDSE and *S. aureus* carriage, infection and non-human source presence within households in this setting.

Clinical and field evaluations

Socio-demographics and household set-up. At the enrolment and later visits where necessary, a questionnaire will be asked of each individual participant in relation to their socio-demographic information including their date of birth, sex, tribal group, educational level and occupation. Any relevant medical information that is identifiable from ante-natal cards (ANC), or infant welfare cards (IWC) (especially for younger children) will be recorded such as birthweight, previous medical diagnoses and allergies.

For each household, data will be collected relating to the household set-up including the number of buildings, family relationships, number of rooms, accessibility for non-household members, sleeping arrangements, mosquito net use, water access, sanitation and hygiene (WASH) facilities and proximity to community meeting points.

At subsequent monthly visits, individuals will be asked to update some of their sociodemographic details such as occupation, work place, school name and attendance and any other factors that may change throughout the year, and to complete any missing data. Similarly, alterations to household set-up will also be collected.

Social mixing behaviour. For all participants at their enrolment and monthly visits, and in more detail for the intensified incidence cohort at each weekly visit, data will be collected on individuals' social mixing behaviour. Participants will be asked information about which other household members they had close contact with the previous day, and who they shared a food bowl, room and/or a bed with. They will be asked about their non-household social contacts from the previous day, including each contacts age, relationship to them, the location of meeting, whether there was physical contact and the duration. For the more detailed extended social mixing behaviour collected in the intensified swabbing cohort, the above data will also be collected for household social contacts.

Medical and drug history. At enrolment, a focused past medical history will be taken from participants including any regular medication taken, previous diagnoses and previous history of skin or throat infections and acute rheumatic fever specifically. At enrolment and at each subsequent visit a brief history of recent medication (particularly antibiotics) and current clinical symptoms will be taken, including details of any recent healthcare setting attendance including traditional healers.

At unscheduled visits, a clinical history of the presenting complaint, medication usage and healthcare attendance will be taken to capture information related to any potential StrepA infections, but also to inform immediate and subsequent medical management of other complaints.

Clinical examination and vital signs. At the enrolment visit, all participants will undergo a physical examination including vital signs to provide a baseline. Vital signs collected will include axillary temperature, pulse rate and respiratory rate. Adults (over 18 years) will also have blood pressure recorded.

Participants will then undergo a physical examination which will include an examination of the pharynx and associated lymph nodes, and a full body examination of the participant's skin, to identify any pyoderma lesions, and other relevant skin conditions (particularly scabies and fungal infections). Care will be taken to perform the full body examination with appropriate privacy and verbal consent obtained at the time. Participants' genitals will only be examined if they specifically report (or the parent reports, in the case of children) the presence of a lesion and verbally consent for the study nurse examine them.

At each subsequent visit participants will undergo the physical examination as described including throat and skin but will not have vitals recorded unless they are reporting any symptoms. If they have any medical complaint, a clinical history will be taken and fuller clinical examination of the presenting complaint will be done, in addition to recording vital signs. At unscheduled visits participants will also have their vitals recorded in the same way, and a clinical history and focused clinical examination will be done.

Clinical samples

At each visit, participants will have clinical samples collected according to the sampling schedule outlined in Table 2.

Oropharyngeal swab. Oropharyngeal swabs (Copan Transystem™ 140C rayon swabs in liquid Amies medium) will

Table 2. Visit data and sampling schedule for the various cohorts. *only at MV0 and MV12.

Visit timing	Visit window	Data and samples	Main cohort	Intensive incident cohort (16 households)	Duration cohort (new carriers)
Month 0 enrolment visit (MV0)	-	Eligibility	X	-	-
		Socio-demographics	X	-	-
		Social mixing behaviour	X	-	-
		Household setup and WASH	X	-	-
		Previous and recent medical history	X	-	-
		Pharyngeal and skin examination including vitals	X	-	-
		Oropharyngeal swab	X	-	-
		Normal skin swab	X	-	-
		Blood serum*	X	-	-
		Dried blood spot	X	-	-
		Oral fluid	X	-	-
		Environmental swabs (x2)	X	-	-
		Settle plate	X	-	-
Weekly visits for duration and intensive incident surveillance	+/- 7 days	Recent medical history	-	X	X
		Extended social mixing behaviour	-	X	-
		Pharyngeal and skin examination	-	X	X
		Oropharyngeal swab	-	X	(X) *if previously positive*
		Normal skin swab	-	X	(X) *if previously positive*
Monthly visits (MV1, MV2, MV3 etc. up to MV12)	+/- 14 days	Update socio-demographics, household setup and WASH	X	X	-
		Recent medical history	X	X	-
		Social mixing behaviour	X	-	-
		Extended social mixing behaviour	-	X	-
		Pharyngeal and skin examination	X	X	-
		Oropharyngeal swab	X	X	-
		Normal skin swab	X	X	-
		Blood serum*	X	X	-
		Dried blood spot	X	X	-
		Oral fluid	X	X	-
		Environmental swabs (x2)	X	X	-
		Settle plate	X	X	-

Visit timing	Visit window	Data and samples	Main cohort	Intensive incident cohort (16 households)	Duration cohort (new carriers)
Unscheduled visits (may occur at scheduled visits if symptoms present)	-	Clinical history and examination	X	X	X
		Wound or oropharyngeal swab	(X) *if applicable*	(X) *if applicable*	(X) *if applicable*
		Oral fluid	(X) *if applicable*	(X) *if applicable*	(X) *if applicable*
		Dried blood spot	(X) *if applicable*	(X) *if applicable*	(X) *if applicable*
		Scabies skin swab	(X) *if applicable*	(X) *if applicable*	(X) *if applicable*
Personal hygiene visit (done at another monthly visit)	-	Personal and household hygiene behaviour	X		
		Extended environmental swabs (x5)	X		

be collected from each participant by swabbing the posterior pharynx (both tonsils, posterior wall, uvula and any area of inflammation or exudation), avoiding touching the tongue, cheeks and lips. After sample collection, the swab be aseptically placed in liquid Amies transport solution and placed in a cold box until processing in the laboratory.

Oropharyngeal swabs will be collected in the same way for participants complaining of symptoms that could be consistent with acute pharyngitis at unscheduled visits

Normal skin swab. Normal skin swabs (CITOSWAB® flocked nylon fibre mini-tip swabs in 1ml liquid Amies medium) will be collected with the intention of identifying any StrepA present on the skin, rather than differentiating skin site. Therefore, to maximise sensitivity, a single swab will be used on multiple skin sites.

Swabs will be obtained using modification of a standard skin microbiota swabbing technique[28–31] in which the swab head is moistened with sterile water prior to skin swabbing. The swab will be taken from 2 by 2cm squares of skin on the forehead, then a larger area (5 by 20cm) on both forearms and both lower legs, and then placed aseptically in liquid Amies transport medium and stored in a cold box until processing in the laboratory.

SSS will be collected in the same was as NSS but from a 2 by 2cm patch of skin overlying typical non-infected scabies lesions.

Pyoderma wound swab. Pyoderma WS (Copan Transystem) will be taken at from participants any visit with evidence of pyoderma (a skin infection with pus or crusts). Pus will be expressed if necessary. WS will be placed in liquid Amies transport medium in cold boxes until processing in the laboratory.

Dried blood spot. DBS samples will be collected using dried blood spot collection cards (Whatman 903 protein saver

snap-apart cards with four sample spots) from a finger prick on the participant. Four drops of blood will fill the four spots on the DBS card. The finger will be cleansed with alcohol and allowed to dry before the finger prick is made with a lancet. Following collection, the DBS card will be left to dry at room temperature before transportation. Transportation to the laboratory will be in a cold box.

Blood sampling. The study team will be trained to perform venepuncture in the field. In the case that the head of the household, all participants aged over 18, and all guardians of children under 18 verbally consent to venepuncture for blood to be taken on site, this will be performed within the household. Alternatively, an appointment will be made at a specified time to attend Sukuta Health Centre where venepuncture for blood serum will be performed by members of the study team.

Peripheral blood will be collected into serum separation tubes using aseptic technique, ensuring appropriate PPE is used. BS samples will not be obtained from participants under the age of 2 years or those who do not verbally consent.

Oral fluid samples. OF samples will be collected using an ORACOL® salivary collection device (Malvern Medical Developments, S10) from participants at the time points specified in Table 1. The oracol swab will be placed in the buccal cavity of the participant between the gums and the cheek for two minutes. Once obtained, the swab will be immediately placed in the collection tube according to the manufacturer's instructions. OF tubes will be transported to the laboratory in a cold box.

Environmental swabbing. At monthly household visits, two ES (Copan Transystem) will be taken from common touch points in the household. At the enrolment visit, the study team will identify two surfaces to swab within the household which are commonly touched by multiple people. Swabbing points might include door handles, table surfaces, curtains, benches,

chair handles etc. Once two surfaces have been decided for the household, those will be the two surfaces swabbed at each subsequent visit.

The ES tip will be soaked in sterile water prior to swabbing and will be rubbed slowly and thoroughly over the surface (up to 50cm) three times reversing direction between strokes. Once collected, the swab will be aseptically placed in liquid Amies transport medium and stored in a cold box until processing in the laboratory.

For the personal and household hygiene visit, additional environmental swabs will be collected from a wider range of common touch points in the house, and a swab will be soaked in water taken from the household greywater source.

Settle plates. A settle plate will be used at each monthly household visit to passively capture the presence of airborne StrepA within households. A culture petri dish pre-prepared with Colombia blood agar will be placed at a suitable point in the main social indoor room of the household. Ideal placement of the settle plate would be at least one metre off the floor, one metre away from the walls and other large obstacles. The plate will be left for one hour, then retrieved and stored in a cold box until processing in the laboratory.

Laboratory evaluations

Sample transport. All swabs and clinical samples taken in the field, except for DBS cards which will be dried first at room temperature, will be stored as soon as possible in a cold box maintained at approximately 2–8°C. All samples will be transported in the cold box to the MRCG Fajara laboratories for processing the same day.

Culture procedures. OPS, NSS, SSS, WS and ES will be processed in the same way. Swabs will arrive at the laboratory in 1ml of liquid Amies transport medium. On arrival at the lab, after ensuring that the swab is inside and the lid is properly closed, the swab will be briefly vortexed in the transport medium. Next the swab will be removed and streaked onto a Colombia blood agar culture plate, and then discarded. Colombia blood agar was chosen for its availability MRCG and utility in culturing both *Streptococci* and *Staphylococci*, though use of selective blood agars may have been preferable to prevent overgrowth. The remaining liquid Amies will be stored at -70°C without the addition of glycerol for subsequent use. It should be noted that the absence of glycerol in the stored medium precludes the possibility of reculturing from these samples, and the addition of glycerol will be considered in future studies.

The culture plates will be incubated for 18–24 hours at 37°C and assessed for the presence of beta-haemolytic colonies. Colonies with clear beta-haemolysis will be picked and replated for purity for a further 18–24 hours at 37°C. Pure growth colonies will then undergo catalase testing, and if negative then latex agglutination testing for Lancefield group A, C and G. Colonies identified from the primary plate as possible *Staphylococcus aureus* from NSS, SSS and WS will be identified based

on their morphology and will also be replated for purity, then if catalase positive, will be tested for *S. aureus* using staphylococcal latex testing. Latex agglutination tests will be used for initial identification of streptococcal groups due to their practicality in our setting. However, we acknowledge the limitations of this method in distinguishing between *S. pyogenes* and SDSE. Liquid Amies transport medium from the swabs will be stored for future PCR validation, specifically targeting the *speB* gene.

Single colonies identified from the purity plate of any group A, C or G streptococci isolates identified will then be stored in glycerol broth at -70°C for later revival, DNA extraction and whole genome sequencing (WGS), which will provide detailed information on species, antigen presence and carbohydrate expression. *S. aureus* colonies will also be stored for later analysis. In future studies, consideration will be given to additionally storing a sweep of colonies from the original culture plate.

Antimicrobial susceptibility testing by disc diffusion using standard CLSI procedures will also be performed on group A, C or G streptococcal isolates identified.

Whole genome sequencing (WGS). Isolates will be revived, and DNA extracted using established methods. Library preparations and WGS (Illumina short read and Oxford Nanopore technology long read platforms) will be undertaken. Quality control, *de novo* genome assembly, and core genome determination will be performed, followed by basic phylogenetic reconstruction using maximum likelihood. *Emm* and MLST typing and AMR will be performed. Genotypically-linked isolates will be determined by analysing genetic diversity and relationships between isolates.

Dried blood spot processing. Upon arrival at the laboratory, DBS cards will be dried at room temperature overnight, then stored at -20°C for later elution. To elute the blood, 6mm punches of dried blood filter paper will be obtained and eluted using a buffer solution. The resulting eluate will then undergo serological analysis for anti-StrepA antibodies, including antibodies to Streptolysin O, SpyCEP, SpyAD, GAC, DNAseB, Enn, Mrp and M protein. Responses to the hypervariable M protein will be explored by selecting three types of antigens: first, *emm*-cluster-representative M peptides; second, M peptides from emm-types identified in The Gambia; and third, M peptides from emm-cluster-representative M protein vaccine antigens. A similar framework will be applied to selecting representative Mrp and Enn proteins. The antigens were chosen to provide a mixture of well-established markers of infection and to cover antigens used in some leading vaccine candidates. Our preliminary (unpublished) evidence suggests that DBS samples are adequate for monitoring the development of immunity to StrepA infections and carriage events compared to serum, as has been shown elsewhere[32].

Serum blood. BS samples will be used to assess serological activity to StrepA antigens including Streptolysin O, SpyCEP, SpyAD, GAC, DNAseB, Enn, Mrp, and M protein at baseline

and at the end of the cohort. BS taken at the same time-points as DBS samples will additionally contribute to validation of DBS in this setting as a reliable and reproducible method for measuring anti-StrepA antibodies. Serum samples will be stored for further immunological work including streptococcal killing assays and opsonophagocytic assays to explore correlates of protection from StrepA asymptomatic carriage and clinical disease.

Oral fluid samples OF samples will be mixed with antibody stabilising buffer on the day of collection. OF samples will be used to assess for mucosal antibody activity to StrepA antigens including Streptolysin O, SpyCEP, SpyAD, GAC, DNAseB, Enn, Mrp, and M protein. Samples will be stored for further immunological work.

Modelling

Using data generated on swab positivity time, participant relationships, WGS data on phylogenetic relatedness of strains, geographic distance between households, and assortativity of social mixing in this setting, we will attempt to identify likely transmission events between individuals using R packages such as *outbreaker2* and *o2geosocial*. These models use Bayesian techniques to compute the likelihood that transmission occurred between individuals or not and hence allows for reconstruction of likely transmission chains.

Utilisation of the novel data in this project will allow estimation of relative contributions of between and within household transmission, and transmission between symptomatic and asymptomatic individuals. To our knowledge this has never been done for StrepA carriage and infection in Africa. The household model will also be valuable in evaluating potential intervention strategies for future implementation within LMICs. Once past events have been estimated, it will be possible to calibrate the model to simulate forward to predict likely onward transmission in the case of an individual with certain characteristics becoming positive.

Sample size considerations

The primary outcome measures used to determine sample size were:

1. Monthly StrepA carriage prevalence, and

2. StrepA carriage and infection incidence over 12 months.

In HIC, StrepA pharyngeal carriage prevalence in children is 2–17%[20,21], and in Uganda is 15.9%[33]. Our study also includes adults, in whom carriage is lower, but will use pooled skin and pharyngeal carriage as our outcome measure, which will likely increase prevalence in turn. We therefore estimate a pooled prevalence of 15%.

StrepA pharyngeal carriage yearly incidence in children in the US was shown to be 27–32%[20]. We found a skin infection incidence of 592/1000 child years in The Gambia during an influenza vaccine study follow-up (unpublished data) and of which ~50% are likely due to StrepA[11]. As we are including adults with a likely lower incidence, we estimate a yearly incidence of 20%.

The sample size was calculated for the primary objective, StrepA carriage prevalence, using the formula below to measure the estimated prevalence of 15% with a precision of ±5%.

$$n = \frac{Z_{\alpha/2}^2 \times p \times (1-p)}{e^2}$$

Where p is predicted proportion and e is desired precision.

Using Z=1.96 for α=0.05, p=0.15 and e=0.05 we require a sample of 196. Intracluster correlation is unknown, therefore we used a conservative design effect of 2, which allowing for 10% drop-out rate gives a required sample size of 431.

We therefore propose to recruit 45 households, which with an average household size of 10, will equal approximately 450 individuals for the main cohort.

This sample size would provide adequate power for precise estimates of prevalence and incidence of StrepA carriage (precision between ±4 and ±5%) and to detect risk factors for StrepA carriage with prevalence (or incidence) rate ratios of greater than 2 with 80% power.

Data analysis

The clinical epidemiology of StrepA and SDSE carriage and infection will be presented using descriptive statistics. Baseline and monthly prevalence of skin and pharyngeal StrepA and SDSE carriage will be reported, including seasonal (monthly) variation. For pharyngeal and skin infection, baseline prevalence will be reported, then monthly and annual incidence for the duration of the study. The typical patterns of transmission observed between individuals within households will be described.

Logistic regression models will be used to look for socio-demographic and medical risk factors for carriage and infection at baseline. Survival analysis (extensions to Cox proportional hazards models) will be used to explore risk factors for carriage and infection throughout the study period taking into account household clustering, repeated events and time-dependent co-variates. The impact of carriage and infection in a close contact or family member in the month prior to new acquisitions of carriage or new infection will be investigated. The relationship between SDSE carriage and StrepA carriage and infection, whether SDSE presence impacts StrepA *emm*-type diversity, whether SDSE carriage is itself a personal risk factor for StrepA will also be explored. Additional risk factors to be explored include the impact of scabies, social mixing patterns, and socio-demographic factors.

To explore the protective association of antibody titres further, regression analysis will be performed to establish the association of antibody titre for each conserved antigen with incident disease, carriage or no carriage/infection accounting for covariates including age, sex and household size.

Data collection and handling

Field data will be collected on electronic case report forms inputted on tablet computers by the field team. The

questionnaires will be designed using REDCap™ electronic data capture software hosted at MRCG. Data will be collected offline and synced with the secure database at the end of each day. Data generated in the laboratory will be inputted onto the same database. Written informed consent will be sought from all participants prior to any study activities and before any data is collected.

Questionnaires will be designed with up-front data quality checks including reference ranges and dropdown menus to minimise incorrect data entry. Additionally, after completion of the study, a data checking process will be performed running queries to check for incomplete or nonsense data.

All data will be kept confidentially, and electronic data encrypted. Each participant will be assigned a unique study ID, so that no person identifiable data will be kept on the database. Any person identifiable data will be held securely and will not be available to anyone other than those in the investigator team. Data will be handled in accordance with the data management SOPs of MRCG which is fully compliant with GDPR regulations. Anonymised data will be held in the study database for a minimum of 10 years following project completion, in compliance with LSHTM's Records Retention and Disposal Schedule. Anonymised raw study data and analysis code will be deposited in the LSHTM Data Compass repository on publication of study outputs and will be available upon request for scientific purposes.

Whole genome sequencing data will be generated using both short read and long read platforms. Raw sequence data will be in the form of fastq files and initially stored on high-performance clusters (HPCs). Data management and analysis will be performed on pipelines established on both the MRCG HPC and the University of Sheffield HPC, as well as cloud-based servers such as the MRC CLIMB platform. Raw sequence data will be archived at MRCG according to their data archiving procedures. Certain processed data fields from genomic analysis (e.g. *emm* type) will be included as variables in the study REDCap™ database. Sequence data along with links to relevant metadata will be submitted to a public sequence repository (e.g. genbank) as is standard practice, on publication of study outputs. Analysis pipeline and code will be made openly available via GitHub on publication of the study.

Ethics and informed consent
This study will be conducted in accordance with the principles set forth in the ICH Harmonised Tripartite Guideline for Good Clinical Practice and the Declaration of Helsinki in its current version, whichever affords the greater protection to the participants.

Ethical approval has been obtained for the study from the MRC Scientific Coordinating Committee and the joint MRC/Gambia Government Research Ethics Committee, as well as the LSHTM ethics committee. Ethical approval reference number LEO24005.

Sensitizing potential study participants will precede the formal recruitment period to ensure that they are aware of the study as far in advance as is practical and therefore are given as much chance as possible to consider their potential involvement prior to providing informed consent. Sensitization will be approached using community and household/individual level strategies.

At the informed consent visit (at least 24 hours after sensitization), the study team will discuss the study with the household head and other household members to confirm that they have understood the consequences of study participation and to answer any remaining questions. If all the inclusion are met and none of the exclusion criteria are, the study team member will proceed to obtain informed consent from all household members.

To obtain informed consent from the household members, in the presence of a literate witness, a member of the study team will translate the informed consent document (ICD), which is in English, line-by-line into the local language spoken by the consenting individual (e.g. Mandinka or Wolof). Once the entire ICD has been translated, the study team member will answer any questions that the individual may have. If the consenting individual remain willing to participate and to provide informed consent, they will be asked to sign or thumbprint signature page of the ICD. If the participant is not literate, the witness will write the date and time and the participant will be asked to thumb-print the signature portion of the ICD. For participants under the age of 18 the child's parent will be required to sign (or thumbprint) the ICD on their behalf. Children aged between 12 and 17 years inclusive will be asked to provide assent to participate in the study in addition to the informed consent provided by the child's parent or legal guardian.

Dissemination
This observational cohort study is registered on ClinicalTrials.gov (NCT05117528). The study results along with raw data and analysis code will be published promptly in peer-reviewed journals and promoted through the MRCG communications department and through social media where appropriate. Data will be submitted as abstracts to be presented at international conferences such as the European Congress of Clinical Microbiology and Infectious Diseases and the Lancefield International Meeting on Streptococci and Streptococcal Disease.

Study status
Field work for this study is now complete. Study enrolment commenced on 27th July 2021 and the final MV12 visit was completed on 28th September 2022. 442 individuals were recruited from 44 households, with 3 households and 160 individuals being lost to follow up.

Data availability
Underlying data
No underlying data are associated with this article.

Extended data

Zenodo. REDCap data dictionary. DOI: https://doi.org/10.5281/zenodo.7463052[34]

Zenodo. SpyCATS informed consent documents (adult and child). DOI: https://doi.org/10.5281/zenodo.7501168[35]

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

References

1. Bowen AC, Mahé A, Hay RJ, et al.: **The Global Epidemiology of Impetigo: A Systematic Review of the Population Prevalence of Impetigo and Pyoderma.** *PLoS One.* 2015; **10**(8): e0136789.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

2. Carapetis JR, Steer AC, Mulholland EK, et al.: **The global burden of group A streptococcal diseases.** *Lancet Infect Dis.* 2005; **5**(11): 685–94.
   PubMed Abstract | **Publisher Full Text**

3. Watkins DA, Johnson CO, Colquhoun SM, et al.: **Global, Regional, and National Burden of Rheumatic Heart Disease, 1990-2015.** *N Engl J Med.* 2017; **377**(8): 713–22.
   PubMed Abstract | **Publisher Full Text**

4. Zühlke LJ, Beaton A, Engel ME, et al.: **Group A Streptococcus, Acute Rheumatic Fever and Rheumatic Heart Disease: Epidemiology and Clinical Considerations.** *Curr Treat Options Cardiovasc Med.* 2017; **19**(2): 15.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

5. Walker MJ, Barnett TC, McArthur JD, et al.: **Disease manifestations and pathogenic mechanisms of Group A Streptococcus.** *Clin Microbiol Rev.* 2014; **27**(2): 264–301.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

6. Steer AC, Jenney AW, Kado J, et al.: **High burden of impetigo and scabies in a tropical country.** *PLoS Negl Trop Dis.* 2009; **3**(6): e467.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

7. Romani L, Steer AC, Whitfeld MJ, et al.: **Prevalence of scabies and impetigo worldwide: a systematic review.** *Lancet Infect Dis.* 2015; **15**(8): 960–7.
   PubMed Abstract | **Publisher Full Text**

8. Watkins DA, Beaton AZ, Carapetis JR, et al.: **Rheumatic Heart Disease Worldwide: JACC Scientific Expert Panel.** *J Am Coll Cardiol.* 2018; **72**(12): 1397–416.
   PubMed Abstract | **Publisher Full Text**

9. Vekemans J, Gouvea-Reis F, Kim JH, et al.: **The Path to Group A Streptococcus Vaccines: World Health Organization Research and Development Technology Roadmap and Preferred Product Characteristics.** *Clin Infect Dis.* 2019; **69**(5): 877–83.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

10. World Health Organization: **Group A streptococcus vaccine development technology roadmap: priority activities for development, testing, licensure and global availability of group A streptococcus vaccines.** Geneva: World Health Organization; 2018.
    Reference Source

11. Armitage EP, Senghore E, Darboe S, et al.: **High burden and seasonal variation of paediatric scabies and pyoderma prevalence in The Gambia: A cross-sectional study.** *PLoS Negl Trop Dis.* 2019; **13**(10): e0007801.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

12. Sharma H, Ong MR, Ready D, et al.: **Real-time whole genome sequencing to control a Streptococcus pyogenes outbreak at a national orthopaedic hospital.** *J Hosp Infect.* 2019; **103**(1): 21–6.
    PubMed Abstract | **Publisher Full Text**

13. Quainoo S, Coolen JPM, van Hijum S, et al.: **Whole-Genome Sequencing of Bacterial Pathogens: The Future of Nosocomial Outbreak Analysis.** *Clin Microbiol Rev.* 2017; **30**(4): 1015–63.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

14. Bergin SM, Periaswamy B, Barkham T, et al.: **An Outbreak of Streptococcus pyogenes in a Mental Health Facility: Advantage of Well-Timed Whole-Genome Sequencing Over emm Typing.** *Infect Control Hosp Epidemiol.* 2018; **39**(7): 852–60.
    PubMed Abstract | **Publisher Full Text**

15. Tagini F, Aubert B, Troillet N, et al.: **Importance of whole genome sequencing for the assessment of outbreaks in diagnostic laboratories: analysis of a case series of invasive Streptococcus pyogenes infections.** *Eur J Clin Microbiol Infect Dis.* 2017; **36**(7): 1173–80.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

16. Cordery R, Purba AK, Begum L, et al.: **Frequency of transmission, asymptomatic shedding, and airborne spread of Streptococcus pyogenes in schoolchildren exposed to scarlet fever: a prospective, longitudinal, multicohort, molecular epidemiological, contact-tracing study in England, UK.** *Lancet Microbe.* 2022; **3**(5): e366–e75.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

17. Bowen AC, Harris T, Holt DC, et al.: **Whole genome sequencing reveals extensive community-level transmission of group A Streptococcus in remote communities.** *Epidemiol Infect.* 2016; **144**(9): 1991–8.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

18. Lacey J, Marcato A, Chisholm R, et al.: **Widespread Streptococcus pyogenes Transmission between Impetigo Lesions and Asymptomatic Throat Carriage in a Longitudinal Cohort from Remote Communities.** [preprint], 2022.
    Publisher Full Text

19. World Health Assembly: **Rheumatic fever and rheumatic heart disease.** Geneva: World Health Organization; Contract No.: 71, 2018.
    Reference Source

20. Martin JM, Green M, Barbadora KA, et al.: **Group A streptococci among school-aged children: clinical characteristics and the carrier state.** *Pediatrics.* 2004; **114**(5): 1212–9.
    PubMed Abstract | **Publisher Full Text**

21. Oliver J, Malliya Wadu E, Pierse N, et al.: **Group A Streptococcus pharyngitis and pharyngeal carriage: A meta-analysis.** *PLoS Negl Trop Dis.* 2018; **12**(3): e0006335.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

22. DeMuri GP, Wald ER: **The Group A Streptococcal Carrier State Reviewed: Still an Enigma.** *J Pediatric Infect Dis Soc.* 2014; **3**(4): 336–42.
    PubMed Abstract | **Publisher Full Text**

23. Jabang S, Erhart A, Darboe S, et al.: **Molecular Epidemiology of Group A Streptococcus Infections in The Gambia.** *Vaccines (Basel).* 2021; **9**(2): 124.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

24. Lorenz N, Ho TKC, McGregor R, et al.: **Serological Profiling of Group A Streptococcus Infections in Acute Rheumatic Fever.** *Clin Infect Dis.* 2021; **73**(12): 2322–5.
    PubMed Abstract | **Publisher Full Text**

25. Whitcombe AL, McGregor R, Bennett J, et al.: **Increased breadth of Group A Streptococcus antibody responses in children with Acute Rheumatic Fever compared to precursor pharyngitis and skin infections.** *J Infect Dis.* 2022; **226**(1): 167–176.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

26. Tagg JR, Ragland NL, Dickson NP: **A longitudinal study of Lancefield group A streptococcus acquisitions by a group of young Dunedin schoolchildren.** *N Z Med J.* 1990; **103**(897): 429–31.
    PubMed Abstract

27. McDonald MI, Towers RJ, Andrews RM, et al.: **Low rates of streptococcal pharyngitis and high rates of pyoderma in Australian aboriginal communities where acute rheumatic fever is hyperendemic.** *Clin Infect Dis.* 2006; **43**(6): 683–9.
    PubMed Abstract | **Publisher Full Text**

28. Ogai K, Nagase S, Mukai K, et al.: **A Comparison of Techniques for Collecting Skin Microbiome Samples: Swabbing Versus Tape-Stripping.** *Front Microbiol.* 2018; **9**: 2362.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

29. Grice EA, Kong HH, Conlan S, et al.: **Topographical and temporal diversity of the human skin microbiome.** *Science.* 2009; **324**(5931): 1190–2.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

30. Capone KA, Dowd SE, Stamatas GN, et al.: **Diversity of the human skin microbiome early in life.** *J Invest Dermatol.* 2011; **131**(10): 2026–32.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

31. Paulino LC, Tseng CH, Strober BE, et al.: **Molecular analysis of fungal microbiota in samples from healthy human skin and psoriatic lesions.** *J Clin*

Microbiol. 2006; **44**(8): 2933–41.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

32. Whitcombe AL, Han F, McAlister SM, *et al*.: **An eight-plex immunoassay for Group A streptococcus serology and vaccine development.** *J Immunol Methods*. 2022; **500**: 113194.
PubMed Abstract | **Publisher Full Text**

33. DeWyer A, Scheel A, Webel AR, *et al*.: **Prevalence of group A β-hemolytic streptococcal throat carriage and prospective pilot surveillance of** **streptococcal sore throat in Ugandan school children.** *Int J Infect Dis.* 2020; **93**: 245–51.
PubMed Abstract | **Publisher Full Text**

34. Armitage EP: **SpyCATS REDCap Data Dictionary.** [Data set]. Zenodo. 2022.
http://www.doi.org/10.5281/zenodo.7463052

35. Armitage EP: **SpyCATS informed consent documents and consent forms (adult and child).** *Zenodo*. 2023.
http://www.doi.org/10.5281/zenodo.7501168

Open Peer Review

Current Peer Review Status: ✓ ✓

Version 2

Reviewer Report 10 November 2023

https://doi.org/10.21956/wellcomeopenres.22555.r69402

✓ **Stephan Brouwer**

The University of Queensland, Saint Lucia, Queensland, Australia

I would like to inform you that I am happy with the revisions, and I do not have any additional comments on the revised version.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 24 August 2023

https://doi.org/10.21956/wellcomeopenres.20754.r64027

? **Stephan Brouwer**
[1] The University of Queensland, Saint Lucia, Queensland, Australia
[2] The University of Queensland, Saint Lucia, Queensland, Australia

Summary:
This is a well-written and carefully designed study protocol by Armitage *et al*. that will provide valuable insights into the true burden of *Streptococcus pyogenes* disease in low-middle income settings. There is a limited understanding of evolution and transmission pathways in regions with

the highest incidence *of S. pyogenes* diseases, such as Africa and the Pacific region. This longitudinal household observational study will make an important contribution to our understanding on the prevalence, incidence, and transmission dynamics of *S. pyogenes* in The Gambia. In addition, this study aims to investigate carriage and infection epidemiology of closely related *Streptococcus dysgalactiae* subsp. *equisimilis* (SDSE), which may also cause significant infections in humans and is highly prevalent in low-middle income settings. However, since the study aims to discriminate the two species and their specific contributions to health burden, I felt that there are a few limitations regarding the selection of molecular markers that may not allow to differentiate between *S. pyogenes* and SDSE (see below) with 100% confidence, thus rather reporting on the burden of "beta-haemolytic streptococci".

Major points:
S. pyogenes and SDSE share extensive genetic homology, including genes encoding conserved virulence factors (e.g. *emm*, *slo*, *sls*, streptokinase and C5a peptidase). Moreover, several studies have reported the (rare) occurrence of Group A-positive SDSE, making it even more difficult to distinguish the two species. The authors suggest the use of latex agglutation tests for the identification of the streptococcal group. However, these have important limitations when trying to distinguish between *S. pyogenes* and SDSE (McMillan *et al.*, 2010[1]) and alternative or additional assays should be carried out to validate results, such as PCR amplification of the *speB* gene and the intergenic region upstream of the *scpG* gene (Davies *et al.*, 2007[2], Watanabe *et al.*, 2016[3], McMillan *et al.*, 2010[1]). Also, it may be very difficult to use antibodies to SLO, SpyCEP, SpyAD, GAC, DNAase B, Enn, Mrp and M protein as *S. pyogenes*-specific serological markers to explore *S. pyogenes* prevalence and infection rates given the genetic similarity to SDSE. How do the authors rule out the possibility that antibody titres are a correlate of SDSE infection?

Page 11: '*Dried blood spot processing*': Antibodies to Enn, Mrp and M protein is very generic. There are more than 200 different *emm* types. It is not clear how the authors address this variability and possible issues with serotype specificity and/or (cross-)reactivity.

Minor points:
Page 4: '*To identify non-human reservoirs of StrepA within households*'. The term "reservoir" is misleading and incorrect as it implies that *S. pyogenes* flourishes and multiplies or can be permanently maintained in these environments, which is not the case. Please change.

Page 5: '*Exclusion criteria*': Can the authors please provide an example how team members judge a participant unable or unlikely to participate and comply?

Page 5: In '*Field activities*' it is stated that enrolments have commenced in July 2021. Are enrolments currently still ongoing, or are they finalised? Can the authors please provide an update on the enrolment situation and also comment on the participation success rate? Also, I understand that enrolment commenced at a time when the country witnessed a severe wave of COVID-19 infections that was mainly associated with the delta variant from July to September 2021 (Gambia-COVID-19-Situational–Report_2021_7th_8th_November_No-399). Have participants with or without symptoms been tested for COVID-19? These would be important data to include in this study. The authors should also comment on the possible impact this wave may have had on carriage/infection rates. Could these data be used to investigate a potential correlation of co-infection?

References

1. McMillan DJ, Vu T, Bramhachari PV, Kaul SY, et al.: Molecular markers for discriminating Streptococcus pyogenes and S. dysgalactiae subspecies equisimilis.*Eur J Clin Microbiol Infect Dis*. 2010; **29** (5): 585-9 PubMed Abstract | Publisher Full Text

2. Davies MR, McMillan DJ, Beiko RG, Barroso V, et al.: Virulence profiling of Streptococcus dysgalactiae subspecies equisimilis isolated from infected humans reveals 2 distinct genetic lineages that do not segregate with their phenotypes or propensity to cause diseases.*Clin Infect Dis*. 2007; **44** (11): 1442-54 PubMed Abstract | Publisher Full Text

3. Watanabe S, Takemoto N, Ogura K, Miyoshi-Akiyama T: Severe invasive streptococcal infection by Streptococcus pyogenes and Streptococcus dysgalactiae subsp. equisimilis.*Microbiol Immunol*. 2016; **60** (1): 1-9 PubMed Abstract | Publisher Full Text

Is the rationale for, and objectives of, the study clearly described?

Yes

Is the study design appropriate for the research question?

Yes

Are sufficient details of the methods provided to allow replication by others?

Partly

Are the datasets clearly presented in a useable and accessible format?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Bacterial pathogenesis, Streptococcus pyogenes, host-pathogen interaction

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 26 Oct 2023

Edwin Armitage

Dear Dr Brouwer, Thank you for your review of this protocol paper. Please find below detailed responses to your specific comments.

Summary:
This is a well-written and carefully designed study protocol by Armitage et al. that will provide valuable insights into the true burden of Streptococcus pyogenes disease in low-middle income settings. There is a limited understanding of evolution and transmission pathways in regions with the highest incidence of S. pyogenes diseases, such as Africa and the Pacific region. This longitudinal household observational study will make an important contribution to our understanding on the prevalence, incidence, and transmission dynamics of S. pyogenes in The Gambia. In addition, this study aims to investigate carriage and infection epidemiology of closely

related Streptococcus dysgalactiae subsp. equisimilis (SDSE), which may also cause significant infections in humans and is highly prevalent in low-middle income settings. However, since the study aims to discriminate the two species and their specific contributions to health burden, I felt that there are a few limitations regarding the selection of molecular markers that may not allow to differentiate between S. pyogenes and SDSE (see below) with 100% confidence, thus rather reporting on the burden of "beta-haemolytic streptococci".

Thank you for your thoughtful and constructive review of our study protocol. Your recognition of the study's potential contribution to our understanding of ipyogenes and *Streptococcus dysgalactiae subsp. equisimilis* (SDSE) in low-middle income settings is greatly appreciated. We acknowledge your concerns regarding the limitations in differentiating between *S. pyogenes* and SDSE. We will address these specific points in detail below.

Major points:
S. pyogenes and SDSE share extensive genetic homology, including genes encoding conserved virulence factors (e.g. emm, slo, sls, streptokinase and C5a peptidase). Moreover, several studies have reported the (rare) occurrence of Group A-positive SDSE, making it even more difficult to distinguish the two species. The authors suggest the use of latex agglutation tests for the identification of the streptococcal group. However, these have important limitations when trying to distinguish between S. pyogenes and SDSE (McMillan et al., 2010[1]) and alternative or additional assays should be carried out to validate results, such as PCR amplification of the speB gene and the intergenic region upstream of the scpG gene (Davies et al., 2007[2], Watanabe et al., 2016[3], McMillan et al., 2010[1]). Also, it may be very difficult to use antibodies to SLO, SpyCEP, SpyAD, GAC, DNAase B, Enn, Mrp and M protein as S. pyogenes-specific serological markers to explore S. pyogenes prevalence and infection rates given the genetic similarity to SDSE. How do the authors rule out the possibility that antibody titres are a correlate of SDSE infection?

Thank you for your detailed and insightful comments. We acknowledge the limitations of latex agglutination tests in distinguishing between *S. pyogenes* and SDSE. Due to the nature of the study and resource constraints, latex agglutination was the most viable method for real-time identification. However, we stored the liquid Amies transport medium with the intention of performing PCR at a later stage to validate our results. This could have been made clearer in the protocol, and we have now updated it accordingly. Regarding the concern about antibody titres potentially correlating with SDSE infection, we aim to minimise this risk through careful event definition and selection. Moreover, the possibility that these antibodies may be stimulated by SDSE would itself be an interesting finding, suggesting that SDSE infection may induce immunity against StrepA. We are currently in the process of setting up a PCR assay for *SpeB* at our unit for further validation. Whole-genome sequencing of all isolates is also planned to provide additional insights into species, antigen presence, and carbohydrate expression.

Page 11: 'Dried blood spot processing': Antibodies to Enn, Mrp and M protein is very generic. There are more than 200 different emm types. It is not clear how the authors address this variability and possible issues with serotype specificity and/or (cross-)reactivity.

Thanks for the comment. We welcome the opportunity to provide more detail. In order to explore the serological responses to M protein within the cohort, we will select

hypervariable M peptide fragments as well as conserved M protein antigens. We will select M peptides firstly from emm-cluster-representative M proteins, secondly from emm types that we have identified in The Gambia and thirdly from other emm-cluster representative M protein vaccine antigens. The same framework will be applied to select representative Mrp, Enn proteins for exploration of serological responses. We have added this detail to the text.

Minor points:
Page 4: 'To identify non-human reservoirs of StrepA within households'. The term "reservoir" is misleading and incorrect as it implies that S. pyogenes flourishes and multiplies or can be permanently maintained in these environments, which is not the case. Please change.

Thank you for pointing out the misleading use of the term "reservoir." We appreciate the clarification and have amended the text to use the term "sources" instead.

Page 5: 'Exclusion criteria': Can the authors please provide an example how team members judge a participant unable or unlikely to participate and comply?

Thank you for raising the question about how team members judge a participant's ability to comply with the study. In the "Exclusion Criteria" section, we have added examples to illustrate how team members might judge a participant's ability to comply. For instance, an individual living in the household with a mental health condition or learning difficulties may be considered not to have capacity to judge the risks and benefits of study participation, even if they were willing to provide consent, or to answer questions related to medical history, demographics, or social mixing behaviour. Other situations could include individuals with communication barriers, such as language difficulties, where translation is not feasible, making it challenging to obtain accurate data, individuals who are frequently absent from the household due to work or other commitments, making it difficult to collect consistent data. These examples aim to provide a clearer understanding of the criteria used to assess a participant's suitability for the study.

Page 5: In 'Field activities' it is stated that enrolments have commenced in July 2021. Are enrolments currently still ongoing, or are they finalised? Can the authors please provide an update on the enrolment situation and also comment on the participation success rate? Also, I understand that enrolment commenced at a time when the country witnessed a severe wave of COVID-19 infections that was mainly associated with the delta variant from July to September 2021 (Gambia-COVID-19-Situational–Report_2021_7th_8th_November_No-399). Have participants with or without symptoms been tested for COVID-19? These would be important data to include in this study. The authors should also comment on the possible impact this wave may have had on carriage/infection rates. Could these data be used to investigate a potential correlation of co-infection?

Thank you for your relevant questions about enrolment and the potential impact of the COVID-19 pandemic on our study. We have updated the "Study status" section to include the final recruitment numbers. While we successfully reached our target enrolment, we did experience a higher-than-expected rate of loss to follow-up. Regarding COVID-19, we did not conduct specific testing for the virus, as a national testing service was available. Participants who met the symptom criteria were advised to utilise this service, although we

did not record this data. This was not one of our study aims, and incorporating it would have required a protocol amendment that was not possible in the timeframe. It's worth noting that several members of our field team did test positive for COVID-19 at various points and followed the prescribed self-isolation protocols. Fortunately, there were no severe cases requiring hospitalisation among our participants. We did not observe any noticeable impact of the pandemic on either enrolment or follow-up. We appreciate your suggestion to consider the impact of COVID-19 on StrepA carriage and infection, but unfortunately without viral nasopharyngeal swabs an individual-level analysis would not be possible. However, we could consider COVID-19 waves in our seasonal analysis. While this is an interesting point, it was not an original aim of our study, and we are hesitant to include it in the protocol at this stage. It is also worth noting that a larger COVID-19 household cohort study was also being conducted in The Gambia at the same time, samples from which could retrospectively be analysed for StrepA presence to address this question. Thanks again for taking the time to review this protocol and provide such detailed comments. I hope that the above responses are to your satisfaction.

Competing Interests: No competing interests were disclosed.

Reviewer Report 18 August 2023

✔️

Shiranee Sriskandan 🆔

[1] Gram-Positive Molecular Pathogenesis Group, Department of Infectious Diseases, Imperial College London, Hammersmith Hospitals campus, London, W12 0NN, UK
[2] Gram-Positive Molecular Pathogenesis Group, Department of Infectious Diseases, Imperial College London, Hammersmith Hospitals campus, London, W12 0NN, UK

The protocol for the household study of Streptococcus pyogenes acquisition and transmission dynamics in The Gambia is adequately detailed and easy to follow. Of note the fieldwork for this study is complete; comments are made with that knowledge. The primary aims of the study are likely to be met, provided that adequate numbers of children are recruited in the households, and the predicted asymptomatic carriage of 15% is broadly correct. A few clarifications in the protocol will enable others to follow or repeat what was done.

1. Bacteriology:
- 1.1: Environmental samples. Columbia Blood Agar has been used; as a suggestion, selective blood agar might be helpful to avoid overgrowth.

- 1.2: Human swabs. Copain Liquid Amies medium swabs are used - these are equivalent (if not better) than gel transport medium for recovery of bacteria. In this study, liquid medium

is used after vortexing (Van Horn *et al.*, 2008[1]). The protocol states the tubes are vortexed and the swabs are used for plating - please clarify.

○ 1.3: The liquid Amies medium is being stored frozen at -70; is this with or without glycerol?

○ 1.4: Groups A, C, or G streptococci are being stored in glycerol for WGS; are single colonies or a sweep of colonies from the original plate stored.

○ 1.5: How is probable Staphylococcus aureus identified? (prior to catalase testing).

2. Serological responses:
○ 2.1: Serum at time zero and at 12 months is collected with blood spots in between. Given that the study will capture incident active S. pyogenes infections and carriage events, is there preliminary evidence that the blood spots are adequate to monitor development of immunity to the infection?

○ 2.2: Eight antigens are proposed for measurement of immunity; is there any explanation of how/why these were selected? (the authors have recently published data using an overlapping slightly different panel of antigens).

○ 2.3: Is there any control measurement to be made in the samples, for example total IgG or a vaccine-related antibody response that is expected to be uniform in the population?

3. Sample size and the model:
○ 3.1: Recently published cross sectional studies from Yemen and Ethiopia suggest asymptomatic carriage rates of 12-15% in children; these data support a predicted rate of 15% and the planned sample size however these are based on children only. How many children are expected in study?

○ 3.2: The protocol states that household size is expected to be 10/household, but also says the 2013 data yielded a figure of 8 per household. Which is correct (the latter might yield only 360 participants)?

○ 3.3: In terms of transmission, is it an aim to determine the effect of strain type on transmission (for example comparing two types of strain or two types of carriage)?

4. Subgroup for frequent swabbing:

The protocol states 16 households will be selected for intense swabbing, but then states that, in the second half of the study, 4 households will undergo intense swabbing. Does this mean that in fact 4 households (out of a possible 16) will eventually undergo this process - or that 16 households will undergo intense swabbing in the first half and four of these will repeat this in the second half?

5. Information about social mixing:

The focus of the protocol is the household which - if understood correctly - may include multiple

smaller outbuildings. If so, please clarify if information about shared sleeping vs cooking/eating arrangements are collected. Although information about non household contacts is being collected, is place and name of school or work collected, since valuable information might be obtained therein?

6. Ethics:

The study has been approved locally and at the sponsor institution. I was surprised that assent was not required from children under the age of 12 but assume this is acceptable in the region being studied. There is no comment about pregnant women and babies but one assumes they will be part of the study.

References
1. Van Horn KG, Audette CD, Sebeck D, Tucker KA: Comparison of the Copan ESwab system with two Amies agar swab transport systems for maintenance of microorganism viability.*J Clin Microbiol* . 2008; **46** (5): 1655-8 PubMed Abstract | Publisher Full Text

Is the rationale for, and objectives of, the study clearly described?
Yes

Is the study design appropriate for the research question?
Yes

Are sufficient details of the methods provided to allow replication by others?
Partly

Are the datasets clearly presented in a useable and accessible format?
Not applicable

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Streptococcal clinical research: transmission and pathogenesis

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

> Author Response 26 Oct 2023
> **Edwin Armitage**
>
> Dear Prof. Sriskandan, Thank you for taking the time to review this protocol paper and providing detailed suggestions to improve the article. Please find below my responses to your specific comments and details of revisions made.
>
> *The protocol for the household study of Streptococcus pyogenes acquisition and transmission dynamics in The Gambia is adequately detailed and easy to follow. Of note the fieldwork for this study is complete; comments are made with that knowledge. The primary aims of the study are*

likely to be met, provided that adequate numbers of children are recruited in the households, and the predicted asymptomatic carriage of 15% is broadly correct. A few clarifications in the protocol will enable others to follow or repeat what was done.

1. Bacteriology:

○ *1.1: Environmental samples. Columbia Blood Agar has been used; as a suggestion, selective blood agar might be helpful to avoid overgrowth.*

Thank you for your suggestion regarding the use of selective blood agar to avoid bacterial overgrowth. We appreciate the insight and will consider using selective blood agar in future studies to minimise this issue. We had hoped to use a selective agar for only the streptococcal culturing, but it was not possible within the timeframe and resource constraints. We have updated the "Culture Procedures" section to clarify the choice of Columbia Blood Agar and its potential limitations.

○ *1.2: Human swabs. Copain Liquid Amies medium swabs are used - these are equivalent (if not better) than gel transport medium for recovery of bacteria. In this study, liquid medium is used after vortexing (Van Horn et al., 2008[1]). The protocol states the tubes are vortexed and the swabs are used for plating - please clarify.*

Thank you for your query. To clarify, after vortexing the tube containing the swab and liquid Amies medium, the swab was used for streaking onto the Columbia blood agar culture plate. The remaining liquid Amies medium is stored at -70°C for future use. The swab was used for streaking on the culture plate in order to preserve as much liquid Amies as possible for the future. We believe the original text accurately describes this procedure, but we appreciate the opportunity to clarify.

○ *1.3: The liquid Amies medium is being stored frozen at -70; is this with or without glycerol?*

Thank you for bringing this to our attention. The liquid Amies medium is stored at -70°C without the addition of glycerol. This was an oversight and, as a result, prevents the possibility of reculturing from the stored Amies medium. We have updated the 'Culture Procedures' section to include this detail and its implications.

○ *1.4: Groups A, C, or G streptococci are being stored in glycerol for WGS; are single colonies or a sweep of colonies from the original plate stored.*

In this study, single colonies were stored from the purity plate for emm-typing and whole-genome sequencing (WGS). We acknowledge the potential value of storing a sweep of colonies from the original plate and will consider this approach in future studies. The 'Culture Procedures' section has been updated to reflect this.

○ *1.5: How is probable Staphylococcus aureus identified? (prior to catalase testing).*

Thank you for your question regarding the identification of probable *Staphylococcus aureus*. In this study, these colonies were initially identified based solely on morphology, prior to catalase testing. We have updated the 'Culture Procedures' section to clarify this point.

2. Serological responses:

○ *2.1: Serum at time zero and at 12 months is collected with blood spots in between. Given that the study will capture incident active S. pyogenes infections and carriage events, is there preliminary evidence that the blood spots are adequate to monitor development of immunity to the infection?*

Thank you for your question. We have included a statement in the "Dried blood spot processing" section to indicate that we have preliminary (unpublished) evidence that suggests that DBS samples are adequate for this purpose compared to serum, but it has

also been demonstrated elsewhere (https://pubmed.ncbi.nlm.nih.gov/34801540/). In the "Serum blood" section it states that serum samples taken at the same time-points as DBS samples will serve to validate the use of DBS for measuring anti-StrepA antibodies in this context.

  ○ *2.2: Eight antigens are proposed for measurement of immunity; is there any explanation of how/why these were selected? (the authors have recently published data using an overlapping slightly different panel of antigens).*

Thank you for this comment. The selection of the eight antigens was based on their relevance to *S. pyogenes* pathogenesis. A combination of some highly conserved antigens, which align with the Combo4 StrepA vaccine, as well as some well-established markers of StrepA infection used in our previous research and the work of others (https://pubmed.ncbi.nlm.nih.gov/34801540/). We have updated the protocol to clarify this.

  ○ *2.3: Is there any control measurement to be made in the samples, for example total IgG or a vaccine-related antibody response that is expected to be uniform in the population?*

Thank you for your question regarding control measurements in the samples. We were not clear about the specific purpose or the nature of the control measurements suggested. If the query was about ensuring that DBS measurements from Luminex samples were valid and of sufficient quality, we are confident that the assay we have adapted for us shows excellent correlation between serum and DBS for GAS antibody measurement (see above reference). We also have unpublished data showing similar results.

3. Sample size and the model:

  ○ *3.1: Recently published cross sectional studies from Yemen and Ethiopia suggest asymptomatic carriage rates of 12-15% in children; these data support a predicted rate of 15% and the planned sample size however these are based on children only. How many children are expected in study?*

Thank you for your question regarding the expected number of children in the study and the predicted carriage rate. Given the demographic profile of The Gambia, where the median age is around 15, we anticipated that at least half of the sample will be children. Additionally, as stated, our 15% prevalence estimate is a pooled figure that accounts for both pharyngeal and potential skin carriage (which is unknown). Furthermore, we employed a conservative design effect of 2 in our calculations, therefore we felt that an estimated prevalence of 15% was reasonable.

  ○ *3.2: The protocol states that household size is expected to be 10/household, but also says the 2013 data yielded a figure of 8 per household. Which is correct (the latter might yield only 360 participants)?*

Thank you for your question regarding the expected household size. We based our estimate of 10 individuals per household on our previous studies and research experience in Sukuta. The 2013 data is considered outdated and less relevant to the current study. We have updated the protocol to make this clearer.

  ○ *3.3: In terms of transmission, is it an aim to determine the effect of strain type on transmission (for example comparing two types of strain or two types of carriage)?*

Thank you for suggesting this possible avenue of analysis. While this was not explicitly an aim of the study, particularly given the expected large diversity of emm types, it could be considered as an exploratory aim if some dominant emm types emerge. We believe the current wording of the protocol allows for this possible exploratory aim.

4. Subgroup for frequent swabbing:

The protocol states 16 households will be selected for intense swabbing, but then states that, in the second half of the study, 4 households will undergo intense swabbing. Does this mean that in fact 4 households (out of a possible 16) will eventually undergo this process - or that 16 households will undergo intense swabbing in the first half and four of these will repeat this in the second half?

Thank you for your question regarding the intense swabbing of households. We apologise for the lack of clarity in the original text. To clarify, 16 households will be randomly selected for intensified swabbing to be done in the second half of the study. These 16 households will be divided into four blocks of 4 households each. Each block will undergo 6 weeks of intensified swabbing in turn until all 16 households have had a block of 6 weeks of swabbing. This was planned this way so as not to overload the field team with weekly visits We have updated the protocol to make this clearer.

5. Information about social mixing:

The focus of the protocol is the household which - if understood correctly - may include multiple smaller outbuildings. If so, please clarify if information about shared sleeping vs cooking/eating arrangements are collected. Although information about non household contacts is being collected, is place and name of school or work collected, since valuable information might be obtained therein?

Thank you for your suggestion regarding the collection of information on social mixing. We would like to clarify that the protocol does already include collection of data on shared sleeping, cooking/eating arrangements, and other household set-up details in the "Socio-demographics and household set-up" and "Social mixing behaviour" sections. We also collected data on school and work places, so to make it more explicit, we have added a sentence to highlight that the name of the school and place of work will also be recorded, as this information was indeed collected.

6. Ethics:

The study has been approved locally and at the sponsor institution. I was surprised that assent was not required from children under the age of 12 but assume this is acceptable in the region being studied. There is no comment about pregnant women and babies but one assumes they will be part of the study.

Thank you for your comments on the ethical considerations of the study. We followed the MRCG informed consent guidelines, which require assent only from individuals aged 12 and above. However, we do always seek informal, ongoing verbal assent from all participants. We appreciate your point about the inclusion of pregnant women and babies; they are indeed part of the study. We have added a sentence to the protocol to clarify this.

Thank you again for taking the time to review this protocol paper in detail. I hope that these changes and clarifications are to your satisfaction. We believe that the clarifications made will make the methodology clearer and easier to replicate.