MRC typhus cohort

NCT04506944

STUDY PROTOCOL – Version 4

THE EPIDEMIOLOGY OF SCRUB TYPHUS AND RICKETTSIAL INFECTIONS IN A HIGHLY ENDEMIC RURAL SETTING IN SOUTH INDIA: POPULATION-BASED COHORT STUDY

12-3-2021

SUMMARY

Background

There is evidence to suggest that scrub typhus and spotted fever group rickettsioses are common causes of febrile illness in India. Serological evidence also exists for murine typhus, but is rarely tested for. Incidence, risk factors, clinical features and molecular epidemiology of these three infections are poorly understood. Delays in disease recognition and treatment may cause thousands of preventable deaths across India. We do not know how many cases develop fever after infection and how many cases with fever develop severe infection. We further do not know how scrub typhus infection is transmitted in the community, i.e. whether infestation with infected chiggers occurs largely in the fields, or within the village environment.

Objectives

The objectives of the research are to determine the incidence and risk factors of scrub typhus, spotted fever and murine typhus by severity, and to determine clinical features of these neglected and often unrecognized infections. Further to study the effect of previous infection on incidence and severity of subsequent infections. Finally to study the association between vector parameters and risk of infection.

Methods: The study will be conducted in South India (Tamil Nadu) and follows up about 30,000 people living in affected villages. Participants will be followed up at 4 to 6 weeks intervals to ask for the occurrence of undifferentiated fever since the last contact. We will take blood samples of all identified fever cases and test for the three infections (scrub typhus, murine typhus and spotted fever) using ELISA and/or IFA. Participants will be asked to notify the study team in the case of any fever or come to a study clinic. Participants will be asked questionnaires regarding living conditions, socio-economic data and occupational factors. 4000 participants will be followed up through annual serological testing to determine the incidence of serological infection. We will catch rodents (the main carriers of infected mite larvae) at different locations in the village environment (outside in field, village edge, village centre) and explore whether there is a relationship between the number of infected mite larvae on rodents and the occurrence of human cases in the neighbourhood.

Outcome: The data collected in this study will be used to estimate the incidence of scrub typhus, spotted fever and murine typhus by severity in the community. This will help establish the "severity pyramid" of infection, i.e. the proportions of serological infections that are clinically apparent, lead to health care use and to complications. We will calculate household and spatial risk factors for scrub typhus and the economic impact of the three infections in the community.

1. BACKGROUND:

Scrub typhus is a febrile illness caused by *Orientia tsutsugamushi*, bacteria related to the genus *Rickettsia* (both are part of the family *Rickettsiaceae*).[1] The infection is transmitted by the larvae (chiggers) of trombiculid mites which infect mammals as incidental hosts.[2] The immune response to scrub typhus involves cellular and humoral immunity (IgG and IgM), and is complicated by the great antigenic diversity of *O tsutsugamushi*.[2] Scrub typhus occurs over much of tropical and subtropical East Asia, South Asia and South-east Asia.[3] The disease has recently been identified in Chile[4] and possibly East Africa.[5] In many endemic areas, scrub typhus accounts for over 15% to 20% of febrile illness leading to health care utilisation.[6,7] At the Christian Medical College in Vellore (CMC, Tamil Nadu, South India, the study location) scrub typhus has been confirmed in 35.9% of admissions for undifferentiated fever (dengue: 30.6%).[8] Mortality in complicated cases remains substantial despite treatment.[2,9] Case fatality of all cases at CMC has been estimated at 9%.[10] Mortality for patients with Acute Respiratory Distress Syndrome (ARDS), the most common complication, was 25%.[10] Other complications include meningo-encephalitis, shock and renal failure.[8,10] Complications may be avoided by early administration of antibiotics. Adverse pregnancy outcomes resulting in stillbirth, prematurity and low birthweight may occur in 40% of pregnant women with scrub typhus.[11] The population-based incidence of severe scrub typhus is not known, but it has been suggested that globally, deaths may exceed those of dengue fever.[2] A recent editorial in the New England Journal of Medicine described scrub typhus as "probably the single most prevalent, under-recognized, neglected, and severe but easily treatable disease in the world".[12] There is a paucity of even basic population-based epidemiological data. The vast majority of studies on scrub typhus are hospital-based, single institution studies or cross sectional sero-surveys.[13] Small-scale repeated sero-surveys in selected villages involved fewer than 200 subjects.[14] Estimating incidence by disease severity requires large-scale community-based cohort studies. Determining the risk of severe infection will help estimate the burden of disease and deciding on the merit of presumptive treatment for scrub typhus, which is practised at some centres. Little knowledge on the immunogenicity of scrub typhus infection is available. In particular, it is not know whether previous infection has the potential to enhance subsequent infection as in dengue fever (antibody-dependent enhancement). The study will contribute to obtaining such data, which will be essential to immunologists and vaccine developers. By estimating the economic impact of scrub typhus, the study will help calculate the cost effectiveness of disease prevention and vaccination as compared to focussing on early case recognition. Although hospital-based risk factor studies have been published,[15-17] the study is community-based and will allow a much more refined spatial-temporal and risk factor analysis, which may add to the primary prevention of the disease. Methodologically, the work, by linking paired serological data to active surveillance data, will echo the landmark cohort studies available for influenza[18] and dengue.[19] The apparent significant increase in the incidence of scrub typhus in recent years and the emergence of new areas of endemicity across the globe underscores the timeliness of this study.[12,15] In addition, we will study two other under-researched infections caused by *Rickettsiaceae*: flea-borne murine typhus (*Rickettsia typhi*) and tick-borne spotted fever (spotted fever group *Rickettsiae*). These occur in the study area,[20] and share clinical features and management with scrub typhus, while very little knowledge exists regarding their epidemiology.

2. AIMS AND OBJECTIVES

The aim of the cohort study is to better understand the epidemiology, sero-epidemiology and transmission of scrub typhus, murine typhus and tick-borne spotted fever. The overall focus of the work will be on scrub typhus, being the most important of the three from the public health perspective. We aim to determine the "severity pyramid" of serological, symptomatic and severe infection. The results from this study will help clinicians to understand the natural course of scrub typhus infection, health policy makers to estimate the

global burden and the scope for community-based interventions, and immunologists to better understand the immunogenicity of single and repeat infections.

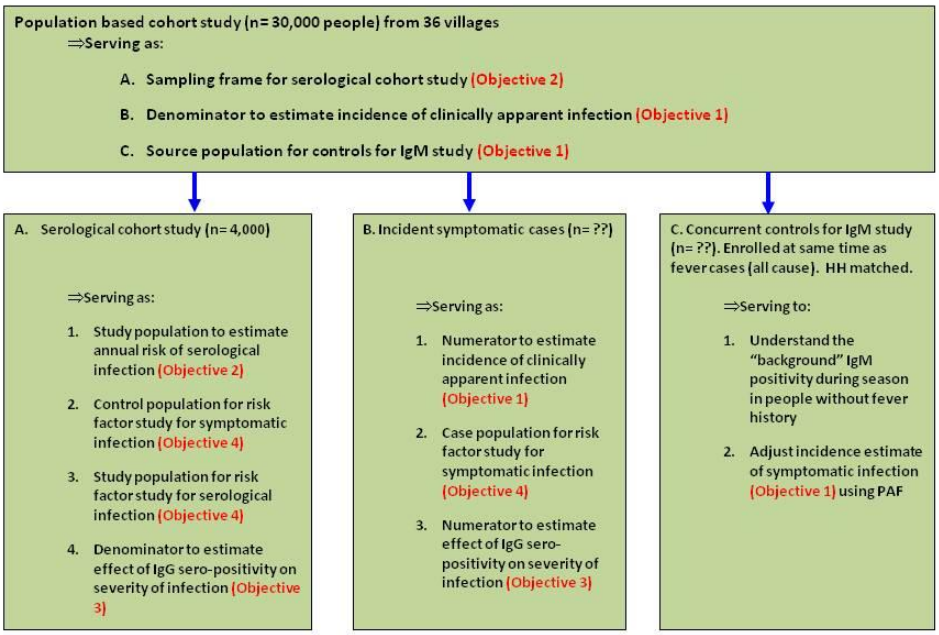
The specific objectives are to:

1) Estimate the incidence of symptomatic and severe scrub typhus, spotted fever and murine typhus infection in the community.
2) Estimate the incidence of serological (asymptomatic or symptomatic) scrub typhus, spotted fever and murine typhus infection.
3) Estimate the effect of previous scrub typhus infection (sero-positive at baseline) on the risk of subsequent infection and disease severity (subclinical vs clinically apparent).
4) Estimate the effect of potential risk factors such as age, gender, occupation, water/sanitation access and comorbidity on the risk of scrub typhus, spotted fever and murine typhus infection.
5) Estimate the effect of spatial-temporal risk factors for scrub typhus, spotted fever and murine typhus by identifying high risk areas, accounting for human population density and land use.
6) Compare the rate of rodent trappings and vector parameters of chiggers between high risk and low risk areas identified under 5) in three environments: outside of the village, village edge and village centre.
7) Estimate the disease burden associated with scrub typhus, spotted fever and murine typhus and the economic impact of scrub typhus, spotted fever and murine typhus in terms of loss of earnings and health expenditure.

3. STUDY OVERVIEW

The study will be a population-based cohort study over two years (i.e. two scrub typhus seasons) in about 30,000 people living in approximately 35 villages with a sero-prevalence of at least 15%. These individuals will be followed-up through active surveillance for recent febrile illnesses (clinical cohort). For passive surveillance, participants will be encouraged to notify the study team in case of fever or come to one of 4 local study clinics. The study will include a smaller cohort (sero-cohort) for serological surveillance of 4000 people drawn at random from the 30,000 people of the clinical cohort. Longitudinal serological surveys linked to active surveillance are essential to estimate true attack rates unbiased by disease severity.[21] The sero-cohort will consist of three annual blood testing in the same group of participants over two scrub typhus seasons (baseline, inter-season, endline). In addition, for each fever case identified in the main cohort study, we will aim to enrol one control without history of fever from the same household to understand the "background" IgM positivity ("IgM study"). As the diagnosis of symptomatic cases is mainly based on IgM ELISA in this study, understanding the "background" IgM positivity may be needed to be able to adjust estimates (see SAP).

The different study populations with respect to the study objectives are shown in the flow diagram below (excluding vector study):



4. STUDY SETTING

The study will be conducted in villages in Vellore district in the Indian States of Tamil Nadu. In the area, the scrub typhus season approximately lasts from June to February.[8,16,17] Urban areas will be excluded as a pilot study showed an unwillingness of urban populations to provide blood samples which we did not encounter in the villages.[30] Also, scrub typhus is likely to be higher in rural settings. The epidemiology of spotted fever and murine typhus is not well known in the area. The number of tests positive for spotted fever (including *Candidatus Rickettsia kellyi*) at CMC's microbiology department is about 10% of the corresponding figure for scrub typhus (JAJP, unpublished). A community-based sero-survey in the neighbouring district of Thiruvannamalai done as part of the RiQVest study revealed an IgG sero-prevalence of 9.4% for spotted fever and 13.6% for murine typhus (JAJP, unpublished).

Target population- A pilot study showed that scrub typhus infection is highly focal with 17 of 48 included villages having 0% IgG prevalence and 10 villages having a prevalence of >40%.[30] The intra-class correlation coefficient of positive scrub typhus IgG at village level was 0.16 (controls only). This makes it difficult to define a target population for the purpose of making inferences. As described above, we found that villages with ≥15% seroprevalence had a higher proportion of fever cases attributable to scrub typhus than those with <15% prevalence. We will define the target population in terms of the village level sero-prevalence. The aim is to correlate the incidence of scrub typhus with the sero-prevalence at village level. This will allow using the results of the study to predict the expected incidence and burden of scrub typhus in other settings with a known sero-prevalence, and eventually estimate the global burden of scrub typhus (not part of the study but a potential extension of the work). The target population for the study is therefore "villages in which the IgG sero-prevalence is at least 15%".

5. RECRUITMENT OF VILLAGES AND PARTICIPANTS

Data collection will begin in January 2020 with a population census in the 25 villages with ≥15% sero-prevalence identified in the pilot study (about 26,000 people). Additional villages will be identified based on results from the RiQVest study and based on clinic records at CMC to meet the sample size of over 30,000 people of all ages from about 7000 households, which will require a total of about 35-40 villages. For villages recruited based on CMC records, those with at least 2 scrub typhus cases testing positive at the Microbiology Department between 2016 and 2019 will be eligible. The census will include GPS mapping of households. After enrolment of villages we will obtain at least one blood sample per household. These will be used to estimate the effect of prior scrub typhus infection (IgG positive) on the risk of subsequent symptomatic infection (objective 3). Within a household, we will select a household member for blood sampling at random. A random sample of 4000 people aged ≥12 years will be enrolled for the sero-cohort. This is an open cohort. Participants newly moving or being born into a household already included in the study will be enrolled for the clinical cohort (but not for the serological cohort).

Prior to enrolment we will inform district and block administrators, and community leaders about the study. We will conduct an awareness campaign in the village to make the study known to potential participants. Providing venous blood samples for research was well accepted by villagers in the pilot study. Enrolment of cases will begin in mid-August 2020.

6. PARTICIPANTS' ELIGIBILITY

For the clinical cohort, all residents including young children permanently resident in a study village will be enrolled and will be asked for a blood sample in case of fever meeting the case definition. For the sero-cohort, we will use the following enrolment criteria: 1) Permanently living in the study area and likely to be resident for another two years. 2) Old enough to give informed consent or assent (minors) but not younger than 12 years.

7. ETHICAL CONSIDERATIONS AND CONSENT PROCEDURES

Vulnerable groups:
In endemic areas, everybody is at risk of scrub typhus and rickettsial infection. It is reasonable to potentially enroll anyone living in these villages for the cohort. At CMC children as young as one month are regularly treated for complicated scrub typhus. In children under the age of 12 and in adults not able to understand the purposes of the study (including those with a severe febrile illness infection too unwell to give consent), we will only perform blood tests in the case of an acute febrile illness if this is judged to be in the best interest of the patient (for example to make or rule out a critical diagnosis for the acute fever). We will not do blood tests in these person groups in the absence of a fever episode (i.e. they will be excluded from the serological cohort, and baseline blood testing), although we may take a sample after the fever episode has been resolved to confirm diagnosis ("convalescent blood sample"). Convalescent samples will not be taken from children if parents / carers do not give written consent for blood sampling, or if the child is unusually distressed by the procedure. In patients who have been unable to give consent during the febrile illness we will obtain assent/consent as soon as this is feasible, in addition to consent obtained from relatives.

Women will be enrolled independent of pregnancy status, as the study procedures pose no specific risks to pregnant women. Pregnant women may particularly benefit from rapid testing offered in this study as this may lower their exposure to antibiotics not indicated for an infection.

Procedures: The study uses procedures which are regularly performed in any medical facility for determining the cause of ill health. In our study, we will collect blood samples and eschar samples where available from individuals who are willing to join the study. The study uses standard techniques which are routinely used for identification of the disease causing agents and poses very little risk to humans.

Consent: The technicians/nurse recruited for the study will be collecting the blood samples from eligible individuals after obtaining informed consent (in writing) in the field. Assent will be obtained from children old enough to understand the purposes of the study. Information sheets will be provided in the local languages (Tamil and Telugu).

Benefits: We will not provide any financial reward for study participation. Acute cases and their doctors will benefit from the availability of tests results from rapid tests performed acutely. Scrub typhus and rickettsial infections are underdiagnosed and are neglected tropical disease (NTD). This study attempts to measure the incidence of these infections in the community. The new knowledge may help to raise awareness to the disease and help with public health planning and treatment decision making.

Animal Welfare: The study has been submitted for approval by the Animal Welfare and Ethics Review Boards at LSHTM and CMC.

8. DATA COLLECTION

ACTIVE SURVEILLANCE- All households of the clinical cohort will be visited about every 4-6 weeks. The sampling interval and recall period of 4-6 weeks has been chosen to balance logistics and the tiring out of study participants against their ability to remember recent febrile illnesses.[22] In particular, due to ongoing Covid-19 transmission in the study area, participants may be reluctant to report fever. The participants themselves or the female head of the household will be asked to report the occurrence of any fever in participants living in the household since the last visit. The pilot study suggested that only 60% of the cases will be available for blood testing at the time of the screening visit. As in the pilot we will schedule an appointment with absent participants for blood testing. Individuals or whole households temporarily absent or thought to have permanently moved out of the study area will be recorded to exclude that time from the calculation of total person time observed.

PASSIVE SURVEILLANCE – At enrolment and during follow up visits participants will be encouraged to call a 24h phone number in the case of any acute febrile illness for which they have visited a clinic or have been admitted. Acute (ongoing) cases making contact by phone or are identified during active surveillance but have not visited a clinic or have not been admitted will be referred to a clinic as per their usual practice if signs of severe infection are present or the fever is prolonged (>3 days). They will have the option to visit one of 4 study clinics part of the study where testing for rickettsial infection will be offered for free. Eligible cases will be those that reside in the study villages (clinical cohort).

Cases identified through active and passive surveillance will undergo a brief screening questionnaire developed for the pilot study to confirm the diagnosis of “unspecified febrile illness” (see below). Cases will be asked further questions regarding duration of illness and concomitant symptoms, including potential danger signs.

Test results will be made available to the treating doctor and communicated to the patient. In case of positive test, treating doctors will be informed that the antibiotics effective against scrub typhus are doxycycline and azithromycin. Independent of the test result, treating doctors will be given the

option to discuss the case over the phone with a Medicine consultant at CMC throughout the patient's hospital admission. A blood sample will be taken from all acute cases and brought to the CMC microbiology lab. These will be tested for scrub typhus, spotted fever and murine typhus using IgG/IgM ELISA and PCR. Eschars[23] will be photographed and confirmed by two independent observers. If the patient agrees we will remove the eschar scab and take a swab from the eschar ulcer for PCR analysis. Clinical information about the patient's progress until discharge will be collected.
Households of confirmed cases will be visited for outcome confirmation (alive or dead at discharge), geo-referencing and a risk factor questionnaire.

SERO-COHORT - We will conduct three rounds of serological testing (IgG / IgM ELISA) in the subsample of 4000 participants enrolled in the sero-cohort, the first before August 2020 ahead of the first season, the second after the end of the first season (March to June 2021) and a final round at the end of the second season (March to June 2022).

NESTED CASE-CONTROL STUDY – Risk factors for infection (symptomatic and asymptomatic) will be estimated using a nested case-control approach.

Enrolment of cases and controls:

1) Cases and controls for symptomatic infection:

Symptomatic cases are diagnosed in active or passive surveillance. These will be enrolled continuously throughout the study. The risk factor questionnaire will be administered to all cases diagnosed with a rickettsial infection. Controls will be enrolled concurrently throughout the season. The serological cohort will serve as the control population. We aim at enrolling all participants of the serological cohort into the risk factor study, excluding participants with a serological infection (see 2).

2) Cases and controls for asymptomatic infection:

The risk factor study for asymptomatic infection will be done in all participants of the serological cohort, i.e. the controls will be the same as described under 1). Cases will be identified after each season, based on paired pre-season/post-season data from the serological cohort. Participants reporting a febrile illness since the pre-season blood test will be excluded.

IgM STUDY

One limitation of using mainly IgM ELISA to diagnose symptomatic cases lies in the possibility of IgM being positive due to an earlier asymptomatic infection unrelated to the fever episode for which the blood test was done. Our pilot study suggested a prevalence of IgM positivity of about 0.5% in the study area after the season.[30] During the season this figure may be higher as IgM may wane weeks after an infection. To allow estimating the extent to which this "background" IgM positivity we will enrol study participants as "IgM controls" throughout the season. As blood sampling in asymptomatic people may be difficult to justify to the study population, we will enrol these IgM controls from households where a fever case (of any aetiology) was identified during active follow up. Enrolment will be done at random in the household among those present at the time and willing to give a blood sample. If no control can be enrolled in that household, a control may be enrolled from an immediately neighbouring household.

9. OUTCOMES

Outcome variables:

- Scrub typhus IgM seropositivity: we will use an optical density (OD) of 1.0.
- Spotted fever IgM sero-positivity: we will use the same cut off of 1.0 as for scrub typhus.
- Murine typhus IgM sero-positivity: We have no knowledge on OD cut off values to define murine typhus infection. During the study we will initially define cases based on manufacturer's specifications, and refine the cut-off by following up these cases over time following infection.
- Where possible we will obtain acute and convalescence samples to determine the increase in IgG using IFA (four fold titre increase) or sero-conversion in ELISA using 1.0 OD as a cut off.

Exposure variables:

- Hospitalisation: defined as any admission lasting for 1 night or longer (to exclude those who were only waiting for results)
- Outpatient treatment: defined as any health care utilisation not defined as hospitalisation. Those receiving diagnosis and treatment, solely from pharmacies will also be included into this category.
- Socio-demographic and socio-economic variables
- Geographic location (GPS)
- Land use (questionnaire items on farming practices, satellite images)

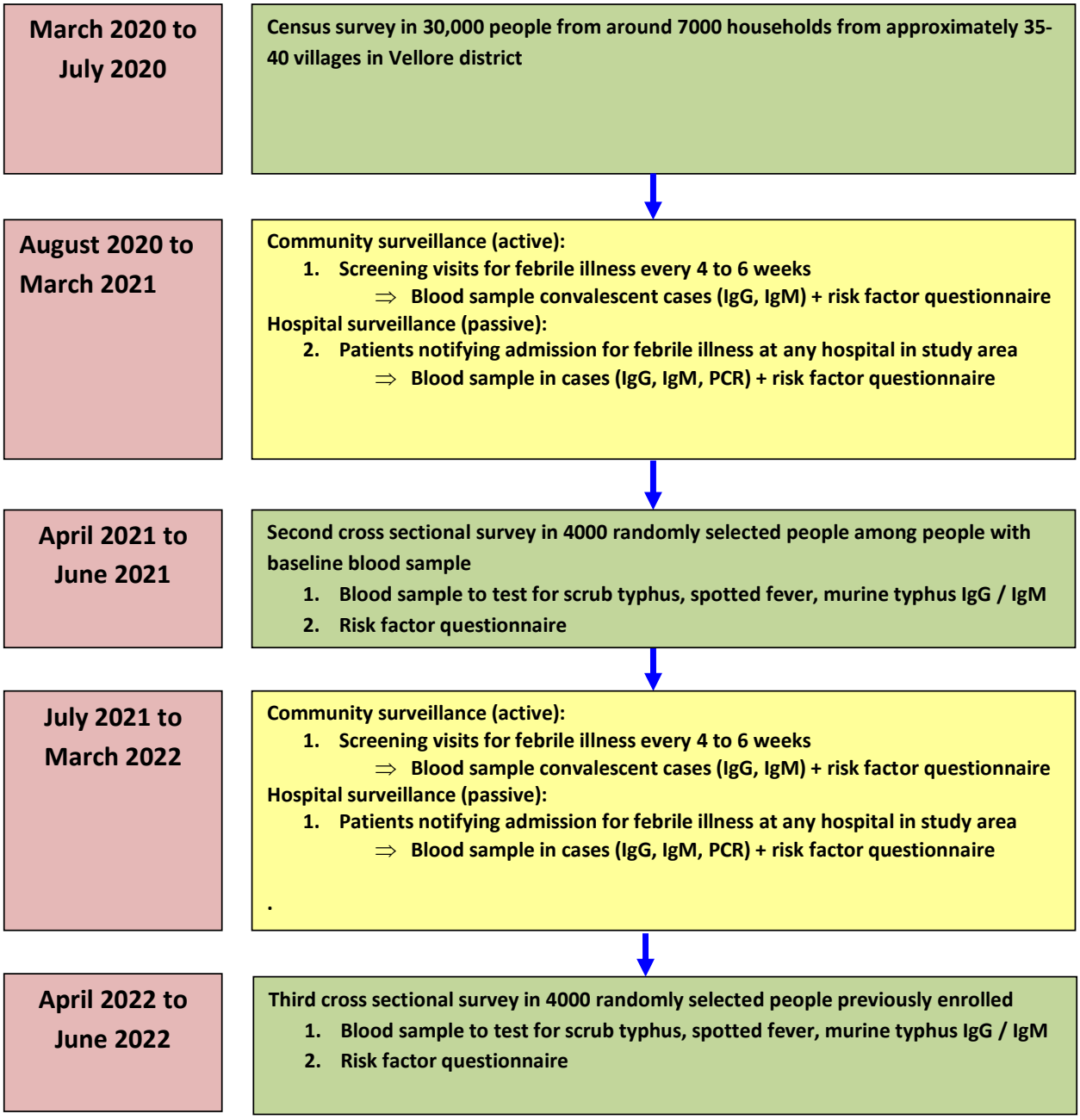
Case reports

A case report form will be used to document the clinical features of scrub typhus, spotted fever and murine typhus, including the following details: age, gender, date of fever onset, date of eschar onset, 24 hour recall of events and visits prior to onset of fever, antibiotics taken prior to this visit, number of clinics/hospitals visited prior this, breathing difficulty, chest pain, headache, number of hours fever subsided following antibiotic administration.

Risk factor questionnaire

Cases recruited will be followed up at their home by a field worker. A questionnaire will be administered with the following details: GIS location, occupations, number of hours spent in the field work alone, open defecation practice, collecting fodder/vegetation/firewood, owning any livestock/poultry/domestic animals, number of hours taken to graze cows/goats. The same questionnaire will be administered to the controls enrolled from the serological cohort study. However, only the cases will be asked questions on treatment costs and income loss due to the fever.

10. STUDY FLOW DIAGRAM



11. CASE DEFINITIONS

Unspecified febrile illness – 1) a self-reported febrile illness occurring since the last visit, 2) absence of leg infection, 3) no other surgical cause for fever identified from patients memory or available health records, 4) absence of culture-confirmed urinary tract infection based on available health records, 5) duration of fever at least 2 days or duration of fever not known, 5) no other obvious cause for fever identified.

Serological case – A study participant with a blood sample which is 1) newly positive for IgG (IFA or ELISA) or 2) shows an increase in IgG titre from the previous survey. The amount of titre increase to define infection will be determined post-hoc based on the data and measurement error.[21]

Clinical case – A study participant with a febrile illness and 1) a positive PCR for scrub typhus/spotted fever/murine typhus, 2) a positive IgM (ELISA) with other serologies being negative. If multiple rickettsia IgM ELISA serologies are positive in a convalescent sample, IFA IgG titration will be done with the highest titre of the three infections being treated as diagnostic. If multiple rickettsia IgM ELISA serologies are positive in an acute sample, we will attempt to obtain a convalescent sample (see 3). IFA IgG titration will be done with the highest titre of the three infections being treated as diagnostic. 3) An increase in IFA IgG titre from acute to convalescence if paired data are available if multiple infections show an IFA IgG titre increase. 4) A confirmed eschar in the presence of fever will be treated as diagnostic for scrub typhus, unless spotted fever IgM is positive and scrub typhus IgM is negative.

Severe case:

- Lung involvement –oxygen saturation below 92% and tachypnea at any time during admission. Tachypnea was defined as >20 breaths/minute in adults; children: >30/min aged 2-5 years, >25/min 5-12 years.
- Shock – adults: documented hypotension (systolic <90mmHg) at presentation or during treatment not responding to a single fluid bolus, or any documented use of inotropes. Children: documented hypotension <80mmHg (2-5 years), <90mmHg (5-12 years), or any documented use of inotropes; or capillary refill time >2s with tachycardia >150/min (2-5 years), >130/min (5-12 years).
- Kidney injury – any creatinine of 3.0 mg/dl or higher in the absence of a known, pre-existing chronic kidney disease.
- Central nervous system (CNS) – any focal neurological deficit, or any elevated white blood cell counts in a cerebrospinal fluid sample, or any focal or generalised seizure in an adult, or any focal or generalised seizure in a child not diagnosed as simple febrile seizure. Simple febrile seizure in children less than 6 years of age was assumed if there was no more than one generalised seizure lasting less than 15 minutes.
- Myocarditis: New onset heart failure confirmed by echocardiography with an elevated troponin T in a patient with no known heart condition.
- Large vessel occlusion, eg. peripheral gangrene or organ infarct.
- Severe bleeding manifestation: purpura fulminans, gastro-intestinal and urinary tract haemorrhage.
- Any natural death during the hospital stay.

12. LABORATORY METHODS

In ongoing cases, malaria RDT, Dengue RDT and Scrub typhus RDT (InBios) may be performed in the field or using finger prick method collected blood samples. Test results will be made available to the treating doctor and communicated to the patient.

All acute and convalescent samples will be tested for IgG/IgM ELISA (scrub typhus, spotted fever, murine typhus) and IFA where appropriate. Blood samples will be brought in cool condition to the laboratory, where they will be aliquoted and stored at -70°C until testing.

Base line census and sero-cohort samples: We will obtain clotted venous blood samples (4ml) from participants at baseline census and during the sero-cohort. At baseline and at end of one year and end of two years we will test for IgG/IgM antibodies to scrub typhus, spotted fever and murine typhus by ELISA

All assays will be performed as per the SOP followed in the Dept of Clinical Microbiology which is an NABL (ISO15189:2012) accredited laboratory.

Real-time PCR (qPCR)

- Samples**:** Whole blood (buffy coat) and eschar swab or eschar biopsy; skin rash biopsy
- Samples will be collected from subjects with ongoing fever of ≥ 3days
- In case rash is present, viral exanthem and drug rash will be ruled out using clinical criteria
- DNA extraction will be performed using the DNeasy Blood and Tissue kit (Qiagen) as per the manufacturer's protocol.
- The quality of the extracted DNA will be assessed by RNAse P qPCR
- Scrub typhus, spotted fever and murine typhus DNA will be detected by amplifying the 47 kDa, ompA and the gltA genes respectively by qPCR.

Nested PCR Samples positive by qPCR will be subjected to nested PCR followed by sequencing so the scrub typhus specific 56-kDa type-specific antigen (TSA) of *O. tsutsugamushi or the ompA gene for spotted fever rickettsia and the gltA will be analysed using MEGA7 .*

13. VECTOR AND RODENT STUDY

Our main aim is to link the vector to human cases. This is more challenging than e.g. for dengue or malaria where human bait catches or collection of vectors in the household environment can be done. In scrub typhus, it is not even known where infestation with infected chiggers takes place.

We are using two different sampling approaches for vector identification through spatial analysis. First, we will compare low and high prevalence villages in terms of rodent and vector parameters. Second, we will explore within a village whether infestation with infected chiggers takes place in the village environment or in the fields. Rodent and vector sampling inside and at the edge of a village, as well as at the more distantly located agricultural fields will help to identify vectors and determine place of infestation with infected chiggers.

We will use cage traps with bait in three locations of an index house (case or control house): 1) peridomestic in the centre of villages, 2) village edge, 3) agricultural land closest to the village edge. About 20 traps will be placed at 10-20m intervals at each location: 20 in the compound of the index house and the 19 nearest neighbours of the index house, 20 at the nearest edge of the settlement, between 20m to 50 away from houses, and 20 in the fields at least 50m away from the settlement. Traps will be put in place on Monday, and then visited daily in the morning (between 8.30 and 9.30am) and checked for trapped rodents. Trapped rodents will be taken inside the cage to a vehicle and brought to the field lab (up to 45 min drive from any of the study villages). Traps with a rodent will be replaced by a fresh trap at the same location. Since rodents are foraging between dusk and dawn, the maximum time spent by a rodent inside the cage including transport and before being killed is about 16 hours (7pm to 11am).

Trapped rodents once brought to the laboratory alive, will be anaesthetized with Isoflurane at a lethal dose and killed.

We will conduct the following procedures after killing the rodent:

1. Rodents will be screened for ectoparasites (ticks and fleas).
2. We will take 5 different measures (tail length, body length, etc) to identify rodent species.

3. We will take the following tissue samples from rodents for freezing and later analysis: lung, spleen, kidney, blood, liver
4. For larval trombiculid mites, ears will be clipped and screened for chiggers using stereo microscope to obtain an approximate number of chiggers on each rodent.
5. Individual trombiculid mites will be mounted on a slides using Hoyer's medium for permanent fixation to enable later speciation.

14. DATA ENTRY

Most data entry will occur in the field using the mWater app on android phones. Logic checks will be incorporated in the questionnaire design. The digitalized questionnaires, once submitted to the server will be checked for plausibility before finally being accepted. In some cases (e.g. clinic questionnaires) we may use paper formats for data collection, though we aim to minimize this. Completed paper questionnaires will be brought to the department along with samples and filed according to villages. Data will be entered twice on a daily basis on the following day after collection with daily updates sent to the research associate. Those sheets that are incomplete will be informed to the field supervisor on a daily basis to ensure completion of forms.

15. DATA ANALYSIS (SEE STATISTICAL ANALYSIS PLAN)

16. SAMPLE SIZE

The sample size has been informed by a pilot study conducted from March to June 2018 in Vellore district. [30] This pilot was conducted as a retrospective cohort. Through screening of 11964 households (42965 people aged ≥12 years) in 48 villages we identified 301 cases (60 hospitalisations and 241 outpatient visits) of febrile illness occurring over the last scrub typhus season (June 2017-February 2018). A random sample of 529 people served as control. Cases and controls underwent blood testing for scrub typhus IgG and IgM. We found that 25 of the 48 villages had an IgG sero-prevalence among the controls of ≥15%. In these villages, the sero-positivity for scrub typhus IgG was 64.3% in people reporting hospital admission due to fever, 37.5% in people reporting a fever related outpatient visit, and 33.6% in controls. We used these data to calculate population attributable fractions (PAF) of fever attributable to scrub typhus. While this approach has some limitations for recurrent infectious diseases, it allowed us estimating that 46% of hospital admissions, and 6% of outpatient visits for undifferentiated febrile illness were due to scrub typhus.

Serological infection – A small cohort in Malaysia published in 1978 found an annual risk of serological infection of 14.5%.[14] In our pilot study, using a relatively low OD cut-off of 0.5 to account for the IgM decline in the months after infection, IgM sero-positivity was 3.6% among controls, which may indicate infection in the last season. In an unpublished one year follow up of controls in the pilot study we found an annual risk of serological infection of 6%. We wish to estimate this proportion with a margin of error of 1% (95% CI of 5%, 7%). This will require 2166 person-years. Since we are observing over two years, this amounts to 1083 persons. We multiply this estimate by a design effect for repeat infections (1.3) and assume 10% participants lost to follow-up, resulting in 1565 individuals in the sero-cohort. Adding a design effect of 1.2 to account for household clustering results in 1877 individuals. In addition we wish to obtain more precise estimates by age group and by village which will enable us to calculate age specific infection rates and correlate village level sero-prevalence with incidence. We therefore aim to include 4000 individuals in the serological study.

Clinical cases – A Malaysian study from the 1970s which relied on passive case finding, but appears to include many mild cases found an annual incidence of 12 per 1000.[6] Our pilot study using the PAFs and the rate of fever related hospital admission and outpatient use found a crude incidence of

1.5/1000 per year in villages with a prevalence of 15% or greater. This is almost certainly an underestimate as a consequence of the long recall period and (due to funding constraints) the use of two relatively untrained field nurses. A preparatory study to test the questionnaire using the same recall period but done by medical interns identified twice as many fever cases per 1000 people. We assume that at a minimum, the incidence of clinically apparent scrub typhus to be 3/1000 per year, with a third being hospitalised cases. This is likely to be a conservative estimate also because the pilot ignored cases not seeking any health care. A study population of 30,000 (60,000 person years over two years of study) will then result in 180 cases over 2 years. This corresponds to an overall incidence rate of 0.003 people with a confidence interval of 0.0026 to 0.0035. Assuming a design effect of 3 due to the village-level sampling approach we expect a confidence interval of 0.0022 to 0.0038. Based on CMC records for admissions vs outpatient treatments we assume that 33% (n= 60) of the expected symptomatic 180 infections will lead to hospitalisation (incidence 0.001, 95%CI 0.0007, 0.0012). We have no data on complicated infection but estimate that 33% of hospitalisations (20 cases) will have complications.

Nested case control study (Risk factor study)

Symptomatic infection

We assume to observe 180 cases of symptomatic infection (see above). Enrolling 4 controls for each case (ratio 1:5) will result in a total sample size of 1054 (176 cases, 878 controls). This sample size will allow the detection of a 0.115 prevalence difference in exposure between cases and controls (assuming a control prevalence of exposure of 0.5). This is equivalent to an odds ratio of 1.6. The detectable effect of the study may be somewhat larger than that, as we enrol all participants of the serological cohort as controls (excluding those with serological infection).

Asymptomatic infection (sero-cohort)

We are expecting about 4000 individuals to be followed over 2 years (8000 person years). We expect an annual risk of 6% which results in 480 cases of sero-infection. The effective sample size may be lower as some individuals may have repeated serological infection. Assuming a design effect of 1.3 due to repeated infection results in an effective sample size of 370 cases. A total sample size of 740 cases and controls will allow detection of a 0.105 prevalence difference in exposure between cases and controls, assuming a prevalence of exposure in controls of 0.5 (OR 1.53). The detectable effect of the study may be somewhat larger than that, as we enrol all participants of the serological cohort as controls (excluding those with history of fever).

Vector study – Rodent trapping and vector parameters of chiggers show great geographic variation. Regarding the prevalence of rodents infested with chiggers, we estimate that 165 rodents from high and low risk areas each will yield 80% power to detect a 15% prevalence difference (75% vs. 60%). We estimate that the prevalence of chigger pools from a single rodent infected with Orientia tsutsugamushi will be 5% and 15% in low and high risk areas, which requires 207 rodents in high / low risk areas each, for 80% power (stata command sampsi for comparison of two proportions). These parameters were estimated based on studies in India[28] and Korea.[29]
We anticipate to observe similar differences between “outbreak” and “non outbreak” villages (see sampling scheme under 3.) The total number of rodents we need to trap will therefore be 2*2*207= 828. As a result, we plan to trap about 828 rodents over the total study period.

17. DATA MANAGEMENT PLAN

1. Description of the data

1.1 Type of study

This is a population-based cohort study. The study consists of a large cohort which will only be followed up through passive surveillance and a small cohort (nested in the large cohort) which will undergo active and passive surveillance. Embedded will be a cross sectional rodent / entomological vector study.

1.2 Types of data

Three types of data will be collected: field data, clinical and laboratory data. The field data will be collected at household level, and will include demographic information, clinical information, geographic positions and (for the vector study) rodent and entomological information. It will include both categorical and quantitative data. Clinical data from patients will be obtained questionnaires (patients / relatives) and from clinical records at participating hospitals (binary, categorical and quantitative data). Laboratory data will be generated from ELISA assays, PCR and biochemistry. Laboratory data will be quantitative and used as continuous (e.g. titres) and binary (positive/negative).

1.3 Format and scale of the data

Data will be collected using paper case report form for clinical cases (CRF) and personal digital assistants for household / participants demographics. The paper based data will be entered through an epidata mask and exported as Access. At database lock, all data will be extracted from the final databases to the statistical packages STATA and R (for spatial analysis). These datafiles could be shared and archived for long term.

2. Data collection / generation

2.1 Methodologies for data collection / generation

Data will be collected into the CRFs (paper/PDA) through oral interviews of the study participants or their relatives (if too ill to respond), transcription from the participants' health records and transcription from lab and entomological / rodent analysis.

Data from laboratory assays will be managed using well established standard operating procedures. Laboratory data will be recorded daily in dedicated notebooks and archived electronically. Data will be discussed critically at regular meetings within and between institutional groups. A data manager will be appointed to oversee data entry and management.

2.2 Data quality and standards

The completed CRFs will be checked to assure completion and consistency by study personnel immediately after filling. At a second level, there will be a quality control (QC) responsible person (data manager) who will prospectively review and verify 100% of all data that is collected and entered into the participant's CRFs. This QC check will be performed as close to the time of data capture as possible to ensure the collection and accurate documentation of all key data points.

A project-specific quality management plan will be developed which will include a checklist of items to be verified. All laboratory procedures will be strictly monitored according to good laboratory practice. Staff will be fully trained in the use of protocols and data will be regularly checked by the lab technician, data manager and senior microbiological staff.

3. Data management, documentation and curation

3.1 Managing, storing and curating data.

There will be a controlled booking system for all CRFs processed which includes real time tracking. When not being processed, CRFs will be locked in cabinets in secure locations. All electronic data will be kept in secure password protected databases housed on a secure network. Strict access controls will

apply both to the servers and the databases.

The database will include QC checks to flag any data entry errors or inconsistencies in the data being encoded. The database will have in-built audit trail to keep track of all the changes/modifications made in the database.

All data will be backed up on a daily basis on to portable hard drives. Data on the hard drives will be encrypted. In addition, anonymised collated data will be uploaded on to the LSHTM network each week using Filr (http://www.lshtm.ac.uk/its/staffservices/filr/). This system is maintained in accordance with LSHTM's Information Management & Security Policy, backed-up and virus scanned on a daily basis.

3.2 Metadata standards and data documentation

Documentation will be developed carefully to ensure it provides sufficient information to understand the content, without providing information that may be used to re-identify individuals. Each output will be assigned a unique identifier to enable data linking across multiple outputs. Names, labels and descriptions for all variables, fields, records and their values contained within tabular datasets will be recorded in a data dictionary. Processing activities performed in STATA will be captured as a set of *do* or R command files, which will be commented throughout in order to provide a clear record of its purpose. All laboratory data will be recorded as per protocol and cleaning strategies employed will be documented.

3.3 Data preservation strategy and standards

At the end of the study, all the electronic study data will be in STATA and R datasets and housed in the study data folder. The locked database will have highly controlled access. All paper records and laboratory books / forms will be stored in a temperature and pest controlled data archiving room for a period of 10 years, compliant with the Indian Government's regulations on records keeping. However, there are no barriers that prevent it being retained in perpetuity.

4. Data security and confidentiality of potentially disclosive information

4.1 Formal information/data security standards

Study team members will be trained on the principles of data security covered in Good Clinical Practice and will follow the LSHTM Information Management and Security Policy at all times.

4.2 Main risks to data security

Main risk to confidentiality is the personal information on subjects kept in the study file by the PI and authorised senior staff for study follow up verification purposes. These will be kept separate from the source data and not available to the study staff so that no linkages of personal data to study records would be easily made. The study data are stored in limited access, password protected databases so that only staffs who have the required permissions can view the study records. Personal information will not be available to study staff except the PI and senior staff authorized by the PI.

5. Data sharing and access

5.1 Suitability for sharing

Data will be of high quality and in a format that will be suitable to share with other interested researchers. We will also share these data with those interested in scrub typhus research including mathematical modelers who may use our data to develop transmission models.

5.2 Discovery by potential users of the research data

Data outputs will be deposited with LSHTM Data Compass (http://datacompass.lshtm.ac.uk/) for long-term curation and preservation, where it will be assigned a Digital Object Identifier (DOI). The DOI will be cited in project reports and journal publications through a Data Access Statement or citation list. Descriptive metadata on each data collection held in LSHTM Data Compass will be made available through OAI-PMH, RSS and ATOM in various formats (Dublin Core, MODS, METS) for use by third party services, such as research data catalogues. Information on each data collection will be provided to the

MRC for inclusion in their Research Data Gateway for Health Sciences catalogue (https://www.datagateway.mrc.ac.uk/).

5.3 Governance of access

An aggregated dataset will be made available as open access using a permissive licence, such as Creative Commons Attribution (CC-BY). Access requests will be reviewed by a Data Access Committee (consisting of project members and external experts).

5.4 The study team's exclusive use of the data

Any request for use of data of the completed study needs approval by the Principal Investigator. The scientists and leaders will be considered for preferential approval up to three years after completion of the project. All data must ultimately be managed in line with the LSHTM requirements for data sharing.

5.5 Restrictions or delays to sharing, with planned actions to limit such restrictions

Requestor will only have access to anonymized datasets. If the request for use of data is within the scope of the study and covered by the signed informed consent, the PI/or his designee will be the authorized person to grant the access to the dataset. If the request is out of the scope of the study, the request will be submitted to the ethics committee to obtain their favourable opinion before the data could be shared with the third party.

5.6 Regulation of responsibilities of users

Any requests for data sharing will be assessed by a Data Sharing Committee and a Data Transfer Agreement will need to be signed upon appropriate conditions being met. Interested parties will need to confirm that they do not attempt to re-identify any participants.

6. Responsibilities

A dedicated data manager will be employed at CMC to ensure the data is stored securely. The data manager will be in charge of checking all field data for consistency and ensuring timely encoding of field and lab data. The majority of analysis will be performed by WPS and NA, who will ensure all meta-data is documented throughout and report to the group.

7. Relevant institutional, departmental or study policies on data sharing and data security

Please complete, where such policies are (i) relevant to your study, and (ii) are in the public domain, e.g. accessible through the internet.

Add any others that are relevant

Policy	**URL or Reference**
Data Management Policy & Procedures	http://researchonline.lshtm.ac.uk/612422/ or http://www.lshtm.ac.uk/research/researchdataman/rdm_policy.html
Data Security Policy	http://www.lshtm.ac.uk/its/informationsecurity/policy/index.html
Data Sharing Policy	See RDM Policy, principle 7 http://www.lshtm.ac.uk/research/researchdataman/rdm_policy.html#principle07
Institutional Information Policy	Held on LSHTM Intranet – available on request
Other:	-
Other	-

18. STUDY MANAGEMENT GROUP

The composition of the study management group will be as follows:

Name	Position Held	Organisation
John Jude Prakash (Co-PI)	Professor	Department of Clinical Microbiology, Christian Medical College,
Wolf-Peter Schmidt (PI)	Assistant Professor	DCD, LSHTM, London UK
Neal Alexander (study statistician)	Professor	Tropical Epidemiology Group, LSHTM
Mary Cameron (lead acarologist)	Professor	DCD, LSHTM, London UK
Carol S Devamani (senior study manager)	Assistant Professor	RUHSA Department, Christian Medical College

The study management group will be responsible for protocol development, day-to-day management of the study, staff supervision, data management and data analysis.

19. STEERING COMMITTEE (SC)

The composition of the SC will be as follows:

Name	Position Held	Organisation
Daniel Chandramohan (Chair)	Professor	LSHTM, London UK
Paul Newton	Professor	Oxford / LOMWRU
Punam Mangtani	Professor	LSHTM, London UK

The role of the SC is to provide overall supervision of the study. In particular, the SC will concentrate on the progress of the study, adherence to the protocol, participants' safety and consideration of new information. The SC must be in agreement with the final Protocol and, throughout the study, will take responsibility for:

- major decisions such as a need to change the protocol for any reason;
- monitoring and supervising the progress of the trial;
- reviewing relevant information from other sources;
- informing and advising the Trial Management Group on all aspects of the trial.

20. REFERENCES

1. Tamura A, et al. Classification of Rickettsia tsutsugamushi in a new genus, Orientia gen. nov., as Orientia tsutsugamushi comb. nov. *Int J Syst Bacteriol* 1995; **45**(3): 589-91.
2. Paris DH, et al.. Unresolved problems related to scrub typhus: a seriously neglected life-threatening disease. *Am J Trop Med Hyg* 2013; **89**(2): 301-7.

3. Kelly DJ, et al.. Scrub typhus: the geographic distribution of phenotypic and genotypic variants of Orientia tsutsugamushi. *Clin Infect Dis* 2009; **48 Suppl 3**: S203-30.
4. Weitzel T, et al. Endemic Scrub Typhus in South America. *N Engl J Med* 2016; **375**(10): 954-61.
5. Maina AN, et al. Q Fever, Scrub Typhus, and Rickettsial Diseases in Children, Kenya, 2011-2012. *Emerg Infect Dis* 2016; **22**(5): 883-6.
6. Brown GW, et al.. Scrub typhus: a common cause of illness in indigenous populations. *Trans R Soc Trop Med Hyg* 1976; **70**(5-6): 444-8.
7. Phongmany S,et al. Rickettsial infections and fever, Vientiane, Laos. *Emerg Infect Dis* 2006; **12**(2): 256-62.
8. Abhilash KP, et al. Acute Undifferentiated Febrile Illness in Patients Presenting to a Tertiary Care Hospital in South India: Clinical Spectrum and Outcome. *J Glob Infect Dis* 2016; **8**(4): 147-54.
9. Taylor AJ, et al. A Systematic Review of Mortality from Untreated Scrub Typhus (Orientia tsutsugamushi). *PLoS Negl Trop Dis* 2015; **9**(8): e0003971.
10. Varghese GM, et al. Clinical profile and improving mortality trend of scrub typhus in South India. *Int J Infect Dis* 2014; **23**: 39-43.
11. McGready R, et al. Pregnancy outcome in relation to treatment of murine typhus and scrub typhus infection: a fever cohort and a case series analysis. *PLoS Negl Trop Dis* 2014; **8**(11): e3327.
12. Walker DH. Scrub Typhus - Scientific Neglect, Ever-Widening Impact. *N Engl J Med* 2016; **375**(10): 913-5.
13. Bonell A, et al. Estimating the burden of scrub typhus: A systematic review. *PLoS Negl Trop Dis* 2017; **11**(9): e0005838.
14. Brown GW, et al. Serological evidence for a high incidence of transmission of R. tsutsugamushi in two Orang Asli settlements in Peninsular Malaysia. *Am J Trop Med Hyg* 1978; **27**(1 Pt 1): 121-3.
15. Wei Y, et al. Rapid increase of scrub typhus: an epidemiology and spatial-temporal cluster analysis in Guangzhou City, Southern China, 2006-2012. *PLoS One* 2014; **9**(7): e101976.
16. Trowbridge P,et al. Prevalence and risk factors for scrub typhus in South India. *Trop Med Int Health* 2017; 22(5):576-582
17. Varghese G, et al. Epidemiology&risk factors of scrub typhus in S India. *Ind J Med Res* 2016; **144**: 76-81.
18. Fragaszy EB, et al. Cohort Profile: The Flu Watch Study. *Int J Epidemiol* 2016.
19. Sangkawibha N, et al. Risk factors in dengue shock syndrome: a prospective epidemiologic study in Rayong, Thailand. I. The 1980 outbreak. *Am J Epidemiol* 1984; **120**(5): 653-69.
20. Prakash JA, et al. Molecular detection and analysis of spotted fever group Rickettsia in patients with fever and rash at a tertiary care centre in Tamil Nadu, India. *Pathog Glob Health* 2012; **106**(1): 40-5.
21. Cauchemez S, et al. Influenza infection rates, measurement errors and the interpretation of paired serology. *PLoS Pathog* 2012; **8**(12): e1003061.
22. Schmidt WP, et al. Sampling strategies to measure the prevalence of common recurrent infections in longitudinal studies. *Emerg Themes Epidemiol* 2010; **7**(1): 5.
23. Kundavaram AP, et al. Eschar in scrub typhus. *J Postgrad Med* 2013; **59**(3): 177-8.
24. Liyanapathirana VC, et al. Seroepidemiology of rickettsioses in Sri Lanka. *BMC Infect Dis* 2011; **11**: 328.
25. Schmidt WP, et al. Population density, water supply, and the risk of dengue fever in Vietnam: cohort study and spatial analysis. *PLoS Med* 2011; **8**(8): e1001082.
26. Greenland K, Schmidt WP. Household cost of illness: summary of findings from four countries. *https://blogslshtmacuk/envhealthgroup/files/2015/05/Household-Cost-of-Illness-Summary-of-findings-from-four-countriespdf* 2015.
27. Vallee J, et al. Contrasting spatial distribution and risk factors for past infection with scrub typhus and murine typhus in Vientiane City, Lao PDR. *PLoS Negl Trop Dis* 2010; **4**(12): e909.
28. Tilak R, et al. Emergence of Schoengastiella ligula as the vector of scrub typhus outbreak in Darjeeling: has Leptotrombidium deliense been replaced? *Indian J Public Health* 2011; **55**(2): 92-9.

29. Song HJ, et al. [Population density of chigger mites, the vector of tsutsugamushi disease in Chollanam-do, Korea]. *Korean J Parasitol* 1996; **34**(1): 27-33.
30. Devamani CS, Prakash JAJ, Alexander N, Suzuki M, Schmidt WP. Hospitalisations and outpatient visits for undifferentiated fever attributable to scrub typhus in rural South India: Retrospective cohort and nested case-control study. PLoS Negl Trop Dis. 2019 Feb 25;13(2):e0007160. doi: 10.1371/journal.pntd.0007160.