SECTION 1: Administrative information

1. **Title:**
    a. **Title: Statistical analysis plan MRC Typhus cohort**
    b. **Study registration number: https://clinicaltrials.gov/ct2/show/NCT04506944**
2. **Version 1, date 19/7/21**
3. **Refers to protocol "MRC Typhus cohort" Version 4, 12/3/2021**
4. **Revision history: NA**
5. **Contributors:**
    i. Neal Alexander, ISEG, LSHTM
    ii. Wolf-Peter Schmidt, DCD, LSHTM
6. **Signatures**
    a. **Person writing the Protocol:** Wolf-Peter Schmidt _____
    b. **Senior statistician:** Neal Alexander _____
    c. **Chief investigator:** Wolf-Peter Schmidt_____

SECTION 2: Introduction

7. **Rationale**

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 this 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 complications. We further do not know how scrub typhus infection is transmitted in the community and what role rodents play in attracting mite larvae.

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 3 to 6 weeks intervals to ask for the occurrence of undifferentiated fever since the last contact. We will seek blood samples of all identified fever cases and test for the three infections (scrub typhus, murine typhus and spotted fever). They will be asked to notify the study team in the case of any fever or come to a study clinic. At study clinics we will also enrol fever cases who are not part of the clinical cohort, i.e. have not been enrolled prior to their fever. In addition, we will enrol mother-child pairs where the child (under 5 years) has been diagnosed with scrub typhus, and test mothers for scrub typhus antibodies. In all ongoing fever cases we will do free rapid tests for common infections, and additional tests including PCR. Participants will be asked 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 study area 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.

8. 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). This includes the effect of antibodies in mothers on the risk of severe infection in young children via residual maternal antibodies.
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).

SECTION 3: Study Methods

9. Study design

The study will be a population-based open 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. At study clinics we will also enrol fever cases who are not part of the clinical cohort, i.e. have not been enrolled prior to their fever. Mothers of young scrub typhus cases (under 5 years) will undergo blood testing for scrub typhus IgG antibodies to explore antibody dependent enhancement due to maternal antibodies. The study will include a smaller cohort (sero-cohort) for serological surveillance of 4000 people drawn from the 30,000 people of the clinical cohort. 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).

10. Sample size calculations

The sample size has been informed by a pilot study conducted from March to June 2018 in Vellore district (Devamani et al. 2019). 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% (Brwon et al., 1978). In our pilot study (Devamani et al., 2019), 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 (i.e. in the months preceding the test). 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 1.3 to account for household clustering results in 2035 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 (Brown et al., 1976). 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 (Devamani et al., 2019). 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.

Because of the expected low numbers of severe cases, we further aim to enrol clinical cases at the 4 study clinics who are not part of the clinical cohort. This will help us estimate the proportion of cases seeking healthcare that are hospitalised and that have complications. We aim to enrol 1000 cases unrelated to the clinical cohort who come to the study clinics with fever of at least 5 days duration.

Assuming that about 30% of these cases will be due to scrub typhus infection, we expect about 300 scrub typhus cases to be enrolled, with a third (100 cases) being hospitalised and half of those having complications (n= 50). We hope to enrol at least 30 child mother pairs – based on clinical data at CMC which suggested children under the age of 5 accounting for 10% of scrub typhus cases.

11. Framework:

The primary outcomes constitute estimation of incidence of disease and infection.

12. Statistical interim analyses and stopping guidance

a. Due to the SARS-Cov-2 pandemic, the originally planned sample size of 10,000 HHs for the main cohort could not be met. The Steering Committee agreed to a reduction of the sample size to 7000 HHs (see updated calculation under 10.) subject to an interim analysis to be done in April 2021 after the first scrub typhus season. The interim analysis aims at verifying the assumptions made in the updated sample size calculations. Based on the results of the interim analysis, the Steering Committee may decide to extend the study by one year.
b. No adjustments for statistical multiplicity will be made.
c. No formal stopping rules have been adopted for the study.

13. Timing of final analysis

The final analysis will be conducted from November 2022 after the end of all data collection.

14. Timing of outcome assessment

- For the main clinical cohort, outcomes will be assessed during follow up rounds every 4-6 weeks over two years (Figure 1).
- For the serological cohort, outcomes will be assessed at annual serological testing between scrub typhus seasons (March to June).
- For the hospital cases not part of the main cohort, outcomes will be assessed continuously as cases are diagnosed at participating centres.

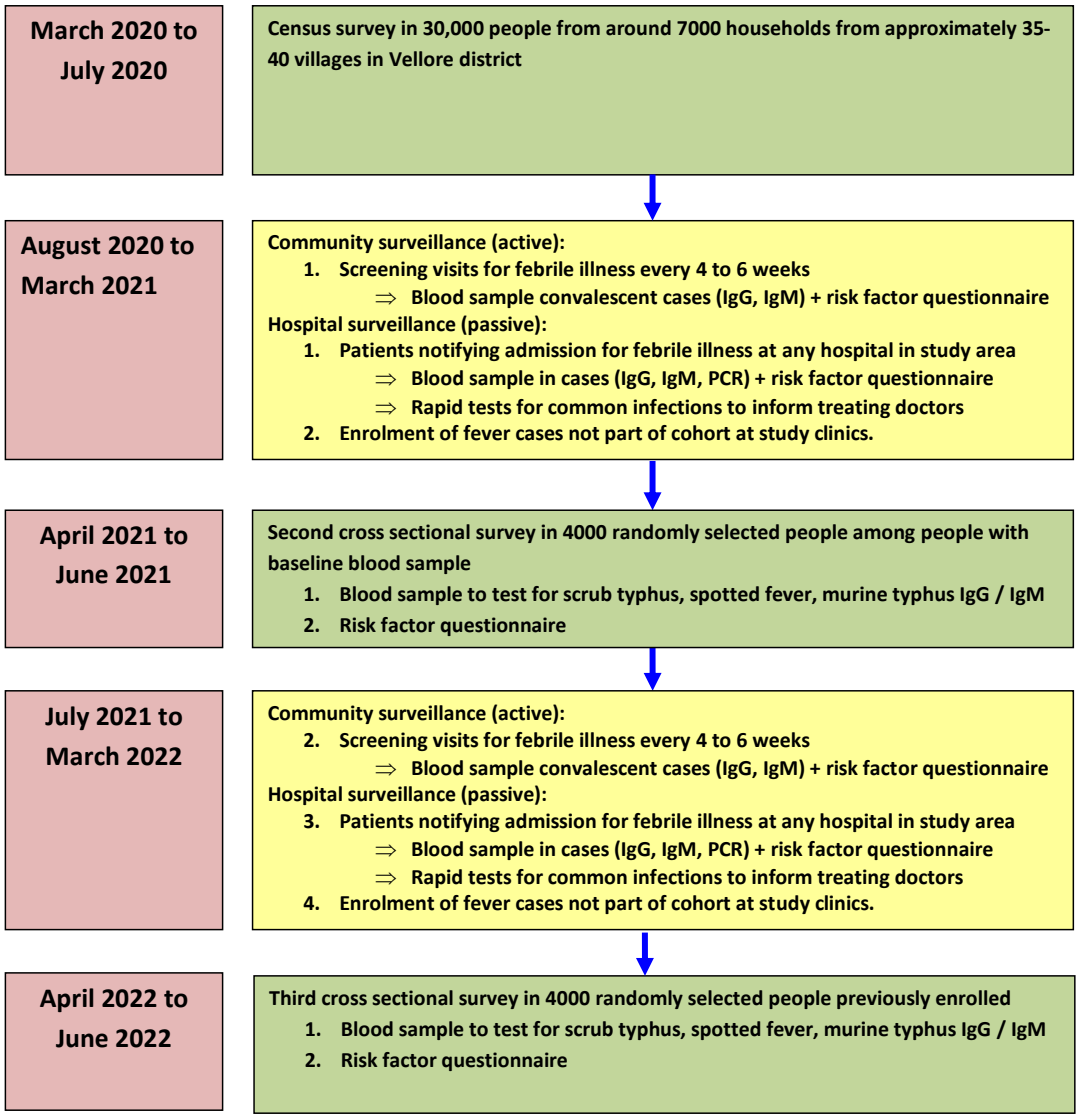

Figure 1. Timing of outcome assessment

SECTION 4: Statistical principles

15. Summary of eligibility criteria:

- **Main clinical cohort:** All residents including young children permanently residing (>6months) in a study village
- **Serological cohort (subset of the main clinical cohort):** 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.

- **Hospital cases not part of the main cohort:** Cases coming to study clinics with undifferentiated fever of at least 5 days duration.

16. Recruitment

- Main clinical cohort and serological cohort: Baseline survey February 28th to July 31st 2020.
- Ongoing recruitment of newly entered households and individuals throughout the follow up period (main clinical cohort).
- Hospital cases: September 2020 to August 2022

17. Baseline characteristics

- **Main clinical cohort/serological cohort**

Table 1. Baseline characteristics

Individual level	
Age	Years
Sex	Binary
Occupation	Binary: engaged in farming work yes/no
Household level	
Household size	Mean (SD)
Dog ownership	Binary

- **Fever cases (from main cohort or hospital cases not part of main cohort)**

Table 2. Characteristics of fever cases

Sex	Binary
Age	Years
number of fever days	Days
Level of treatment	Inpatient, outpatient, local doctor, none
Diagnosis at health care centre	COVID, scrub typhus, spotted fever, other
Treatment	Doxycycline, azithromycine, other
Dermatological manifestations	Eschar, maculopapular rash, petechial rash, purpura
Complications	ARDS, renal, shock, CNS, other

18. Outcome definitions:

Serological case

A study participant with a blood sample which is 1) newly positive for IgG by IFA or ELISA (i.e. the second paired sample was positive when the first was negative) or 2) shows an increase in IgG titre from the previous survey. As a default, we will be using a four-fold increase in titre (two step dilution). However, since a valid cut-off to define a titre increase indicating serological infection is not known (Cauchemez et al.), the definition may be changed post-hoc based on the data collected accounting for repeatability of IFA titration. Similarly, cut –offs for IFA positive samples will be defined post-hoc. Note: The case definition for a serological case requires availability of paired samples from the same person taken before and after the scrub typhus season.

Clinical case

Clinical cases will be identified in active and passive surveillance based on a history of past history of fever prior to a surveillance visit or acute ongoing fever. The presence of fever will be determined based on self-or proxy reported symptoms. The diagnostic criteria will depend on whether an acute sample, a convalescent sample or both are available. The flow diagrams below (Figures 2-4) show case definitions for these three scenarios. Of note, the IgG titre increase after scrub typhus infection is to date poorly defined, especially in cases who already have a high IgG level before an infection. Therefore, cases with paired samples not meeting any case definition for paired samples (Figure 3) can still be categorised as a probable case using the flow diagrams for single samples if they meet those case definitions (Figures 1 and 2).

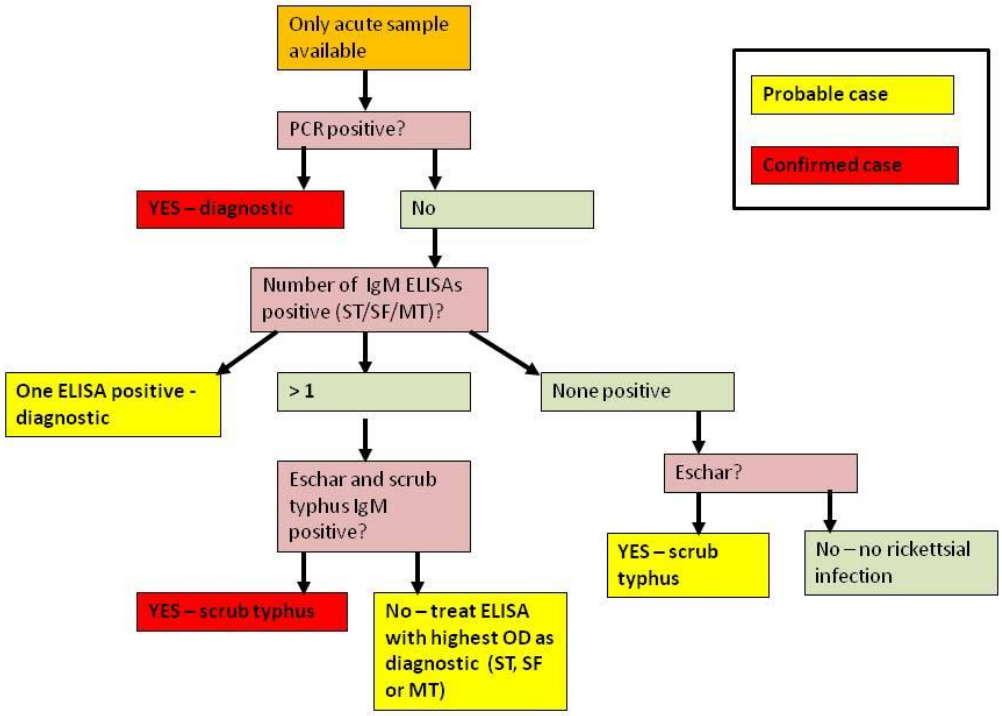

Figure 2. Flow diagram for diagnosis of clinical cases if only the acute sample is available. This applies to all three infections: scrub typhus (ST), spotted fever (SF), and murine typhus (MT). For eschar in spotted fever see Table 3.

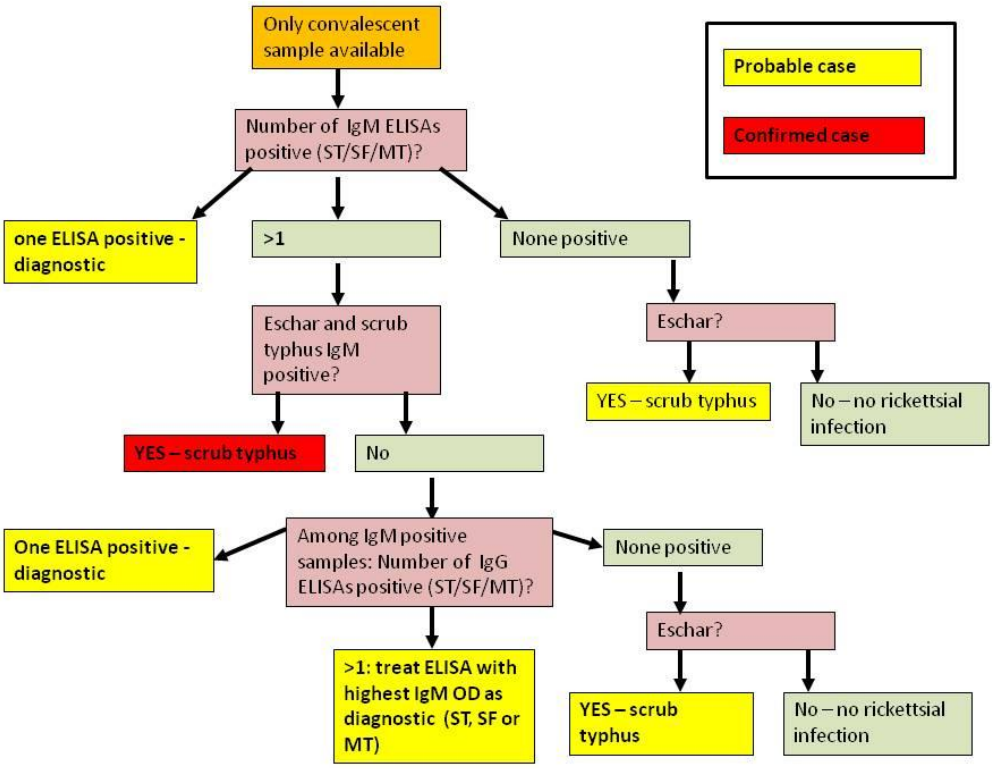

Figure 3. Flow diagram for diagnosis of cases if only the convalescent sample is available. For eschar in spotted fever see Table 3.

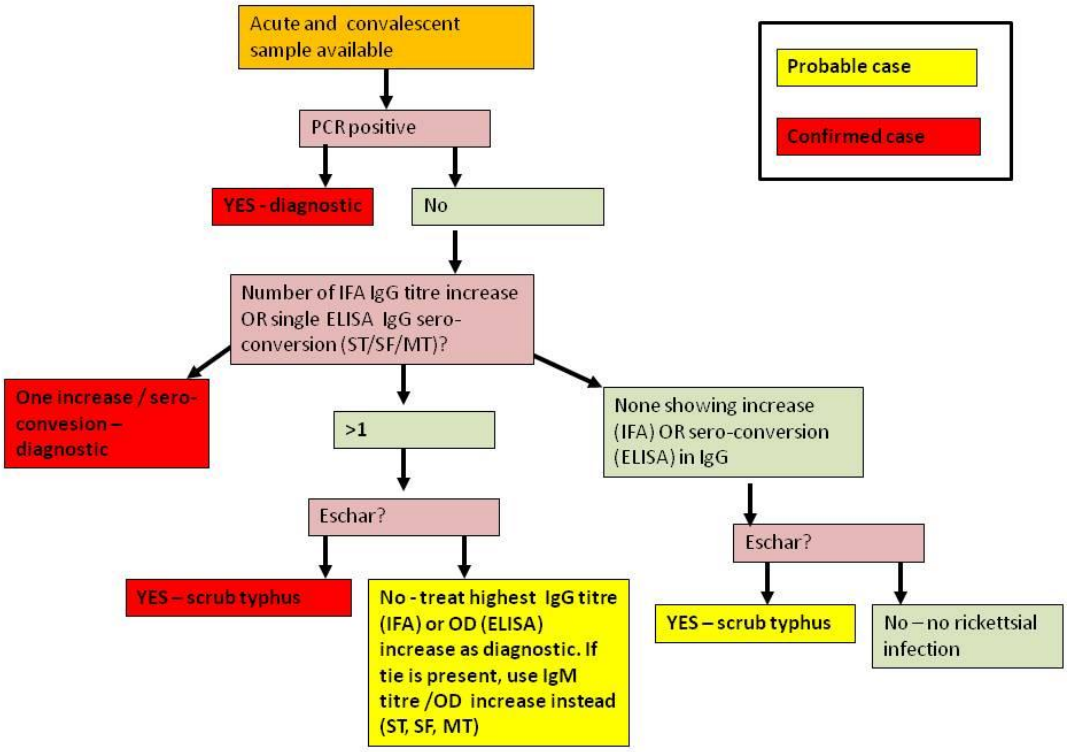

Figure 4. Flow diagram for diagnosis of cases if both the acute and convalescent samples are available ("paired sample"). For eschar in spotted fever see Table 3.

Complicated case (Schmidt et al. 2019)

- Acute Respiratory Distress Syndrome (ARDS) – any patient with oxygen saturation below 92% and tachypnea at any time during admission.
- Shock – any patient with documented hypotension at presentation or during treatment not responding to a single fluid bolus, or any documented use of inotropes;
- 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 will be assumed if there is no more than one generalised seizure lasting less than 15 minutes.
- Myocarditis: New onset heart failure in patient with no known heart condition.

- Large vessel occlusion, eg. peripheral gangrene and organ infarct
- Severe bleeding manifestation: purpura fulminans, gastro-intestinal and urinary tract
- Any natural death during the hospital stay.

Probable case vs confirmed clinical (symptomatic) case

Following the identification of cases using the flow diagrams, we will categorise clinical cases into probable cases and confirmed cases based on Table 1. The study outcomes will be calculated based on probable and confirmed cases combined.

Table 3. Probable vs confirmed cases

Probable case	**Confirmed case**
IgM positive case (ELISA OD cut-off 1.0). If IgM ELISA is positive for multiple organisms (ST/SF/MT), then highest OD is treated as indicating probable organism.	PCR positive case
Scrub typhus: Eschar with negative IgM for scrub typhus and spotted fever, or no test done	IgG titre increase according to definition (paired samples)
	ELISA sero-conversion (paired sample)
	Scrub typhus: Eschar with positive scrub typhus IgM regardless of spotted fever IgM
	Spotted fever: Eschar with positive spotted fever IgM and negative scrub typhus IgM

The above definitions apply to scrub typhus, spotted fever and murine typhus.

Definition of person time of observation to calculate incidence rate

Person-time of observation for an individual will be calculated as the observation time a participant was in the study. Time periods where the participant was not in the study (late entry, early exit, temporary absence), or during which it is not known whether the participant has become a case despite not being absent will be subtracted from that participant's person-time. For six weeks after start of a fever episode meeting the case definitions above, a participant will be treated as not at risk for the infection which was diagnosed for this episode.

Temporary absence is defined as a participant who is already enrolled and who is completely absent from a house (did not spend a single night) between two visits whether or not it is known if the participant has become a case in between.

19. Analysis methods for the different objectives:

All calculations will be done separately for scrub typhus, spotted fever and murine typhus. The analyses will be done in STATA unless otherwise stated.

1) Estimate the incidence of symptomatic (clinical) and complicated infection in the community.

Primary analysis: We will calculate the incidence rate for clinical and severe (complicated) clinical cases identified through active and passive surveillance by dividing the number of cases for each by the total person time of observation in the main clinical cohort. Confidence intervals will be calculated according to Clayton & Hills (Clayton & Hills, 1993)).

Planned additional analyses:

The estimate of the incidence of clinical and complicated infections will be adjusted to account for three potentially important limitations of the crude incidence estimate:

I. Missing blood sample in fever cases

We expect that in a certain proportion of fever cases identified through passive and active surveillance, no blood sample can be obtained to confirm the diagnosis of rickettsial infection, for example because the participant refuses the blood test. This is likely to be more common in cases who only had mild fever, and who did not seek treatment at one of the study clinics. The lack of blood testing in all observed fever cases is likely to lead to an under-estimate of the incidence of symptomatic infection, in particular uncomplicated cases.

We will estimate the probability of each case being positive by developing a logistic regression model for sample positivity in those cases who do have a sample result. The model will then be applied to those with the covariate information but not the sample result. These covariates may include seasonal (e.g. calendar month), level of treatment sought (e.g. local clinic vs hospital) and duration of fever.

II. Cases attributed to rickettsial infections due to coincidence with another condition

Among the controls in the IgM study (box “C” in flowchart in protocol), the proportion positive for each rickettsial infection will be calculated. The proportion of probable cases attributable to each infection will be calculated via classical attributable fraction methods (Smith T et al 1994).

III. Cross-reactivity

Some study participants with a febrile illness caused by an unrelated pathogen and who did not experience an asymptomatic or symptomatic rickettsial infection before or after the observed

episode, may still test positive for a rickettsial infection because of cross-reactivity. Cross-reactivity may lead to an over-estimate of the incidence.

A 4-fold titre increase in IgG can considered a true positive because it is a change within the person, and because cross-reactivity is less of a concern than with IgM. The concern will be those probable cases defined on a single high IgM. For these, a logistic regression model will be developed using those full data (IgM and paired IgG) to make a model for true status (defined by the paired IgG). Hospital cases will be included in this analysis. Duration of illness should be a covariate in the model. The model will then be applied to those with IgM and covariate information, but not paired IgG, to estimate the probability that they would have been positive on the latter.

As a further method to address coincident infections (II) and cross-reactivity (III) jointly, we will calculate the prevalence of IgG positivity for all acute cases and as in (II) use an attributable fraction approach (Smith T 1994) to calculate the proportion of cases attributable to each infection. The control group IgG sero-prevalence will be calculated from those study participants enrolled in the serological cohort.

Sensitivity analysis:

I. Restricted to confirmed cases (based on Table 1)
II. Restricted to confirmed cases and adjusted number of probable cases: In cases with available paired data (acute and convalescent) we will calculate the proportion of probable cases that also meet the definition for a confirmed case, and apply this proportion to all probable cases to obtain an estimated overall number of confirmed cases. This will allow us to account for cases having only one sample and where PCR was not done. This analysis assumes that the probability of getting two samples from a case is unrelated to the probability of a test being positive.
III. Using ELISA OD IgM cut-offs of 0.8 and 1.2 instead of 1.0.

2) Estimate the incidence of serological infection

Primary analysis: The incidence of serological infection (as defined above) will be calculated as the incidence rate per person-year of observation between two rounds of serological testing in participants of the sero-cohort, taking into account differing observation times between individuals. Point estimates and confidence intervals for the incidence rates will be calculated

using the *strate* command in STATA with calendar time split using the *stsplit* option to allow for seasonality.

Sensitivity analysis: We may conduct sensitivity analyses using a cut-off for the fold increase in titre other than 4 (e.g. 2-fold and/or 8-fold), depending on the observed repeatability of assays (Cauchemez et al.). A better repeatability may allow for a lower cut-off while maintaining an acceptable proportion of false positives.

3) Estimate the effect of previous infection (sero-positive before the season) on the risk of subsequent infection and disease severity (subclinical vs clinically apparent)

In clinical cases in the sero-cohort and with an available baseline sample, we will calculate the rate ratios for clinical and complicated infection comparing pre-season sero-negative with pre-season sero-positives using Poisson regression in STATA. In the primary analysis, estimates will be adjusted for age, since age may be associated with pre-season sero-positivity and the incidence of clinical infection. Pre-season sero-status will be treated as a time varying exposure variable (measured each year) as participants can become sero-positive after the first season. In further analyses estimates may be adjusted for sex.

Rate ratio for serological infection

We will similarly calculate the rate ratio for serological infection by pre-season sero-status, using Poisson regression in STATA. In study participants who are IgG sero-negative at baseline, serological infection is easily defined as IgG sero-conversion. Serological infection may be harder to define in participants already sero-positive at baseline in whom serological infection needs to be defined based on an IFA titre increase. Because of the difference in our ability to diagnose serological infection depending on initial sero-status, we will conduct a sensitivity analysis using different titre increase cut-offs as above.

Risk ratio for clinical scrub typhus in children

In a separate analysis of young children under the age of 5 years diagnosed with symptomatic scrub typhus as part of this study (community cohort and clinic cohort) with a blood sample available from the mother, we will calculate the effect of maternal sero-status for scrub typhus on the risk of complicated scrub typhus infection (defined above, section 18) vs uncomplicated symptomatic infection. The analysis will be done using modified Poisson regression (Zou et al) to estimate the risk ratio. The analysis will be stratified by age as the effect of maternal antibodies on the risk of complications in a child is likely to wane in the first 5 years of life. We expect maternal antibodies to be associated with complicated scrub typhus in very young children, but not in older children. Pre-defined age strata will be below 18 months of age and

18 months or older, with children 18 months or older serving as “control” in whom we do not expect to see an effect of maternal antibodies. If the sample size permits, we will calculate a test for interaction by age group.

4) Estimate the effect of potential risk factors such as age, gender, occupation, water/sanitation access and comorbidity on the risk of asymptomatic, symptomatic and complicated infection (described in section 18)

This analysis will be done as a nested case control study as described in the protocol. Controls will be enrolled approximately at the same time as cases are being diagnosed. We will conduct risk factor analysis on presumed determinants of infection using logistic regression models. Since some exposure variables, in particular agricultural activities, vary over time, the analysis will be stratified by calendar month to allow better comparison between cases and controls. Risk factors will include:

1) age
2) sex
3) occupation
4) level of education
5) comorbidities
6) village level sero-prevalence
7) presence of household animals
8) water access and sanitation

5) Estimate the effect of the spatial risk factors of infection

In these exploratory analyses we will use geo-referenced data from four sources: 1) Household GPS positions collected during the baseline survey, 2) geo-referenced locations of forest land and water bodies collected on the ground, 3) use of google earth in QGIS to determine boundaries of continuous settlements, and 4) remote sensing satellite images.

The data will be used to explore spatial determinants of the rate of clinical infection, the risk of serological infection and the prevalence of sero-positivity.

Analyses will include:

1) Population density

2) Distance to village edge (this is visible on remote sensing including Google Earth)
3) Distance to forest
4) Distance to water bodies (open drains, borewells, water tanks, ponds)
5) Vegetation index
6) Spatial correlation of infection

Rate ratios (clinical cases) and risk ratios (serological infection and sero-prevalence) will be calculated as in 4). The spatial correlogram for infection will be estimated using the *an appropriate* package for binary data in R (e.g. *ncf* package or *lorelogram* package, R project).

References

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.

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.

Cauchemez S, Horby P, Fox A, Mai le Q, Thanh le T, Thai PQ, *et al*. Influenza infection rates, measurement errors and the interpretation of paired serology. PLoS Pathog. 2012; **8**(12): e1003061.

Clayton D, Hills M. Statistical Models in Epidemiology. Oxford: Oxford University Press; 1993.

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. PMID: 30802243; PMCID: PMC6405239.

Smith, T., Armstrong Schellenberg, J., Hayes, R., 1994. Attributable fraction estimates and case definitions for malaria in endemic areas. Statistics in Medicine 13, 2345-2358.

Schmidt WP, Devamani CS, Rose W, Alexander N, Prakash JAJ. Antibody response following scrub typhus infection: clinical cohort study. *Trop Med Int Health* 2019; **24**(12): 1455-64.

Stata Manual 13 (https://www.stata.com/manuals13/ststrate.pdf)

Zou, G., 2004. A modified Poisson regression approach to prospective studies with binary data. Am J Epidemiol 159, 702-706.