STUDY PROTOCOL

REVISED Study design and protocol for investigating social network patterns in rural and urban schools and households in a coastal setting in Kenya using wearable proximity sensors [version 2; peer review: 2 approved]

Moses Chapa Kiti [1], Alessia Melegaro [2], Ciro Cattuto [3], David James Nokes [1,4,5]

[1]Epidemiology and Demography Department, KEMRI-Wellcome Trust Research Programme, Kilifi, 80108, Kenya
[2]Department of Social and Political Sciences, Bocconi University, Milan, Italy
[3]Data Science Laboratory, Institute for Scientific Interchange Foundation, Turin, Italy
[4]Zeeman Institute of Systems Biology and Infectious Disease Research, University of Warwick, Coventry, UK
[5]School of Life Sciences, University of Warwick, Coventry, UK

v2 **First published:** 22 May 2019, **4**:84 (https://doi.org/10.12688/wellcomeopenres.15268.1)
Latest published: 22 Aug 2019, **4**:84 (https://doi.org/10.12688/wellcomeopenres.15268.2)

Abstract

Background: Social contact patterns shape the transmission of respiratory infections spread via close interactions. There is a paucity of observational data from schools and households, particularly in developing countries. Portable wireless sensors can record unbiased proximity events between individuals facing each other, shedding light on pathways of infection transmission.

Design and methods: The aim is to characterize face-to-face contact patterns that may shape the transmission of respiratory infections in schools and households in Kilifi, Kenya. Two schools, one each from a rural and urban area, will be purposively selected. From each school, 350 students will be randomly selected proportional to class size and gender to participate. Nine index students from each school will be randomly selected and followed-up to their households. All index household residents will be recruited into the study. A further 3-5 neighbouring households will also be recruited to give a maximum of 350 participants per household setting. The sample size per site is limited by the number of sensors available for data collection. Each participant will wear a wireless proximity sensor lying on their chest area for 7 consecutive days. Data on proximal dyadic interactions will be collected automatically by the sensors only for participants who are face-to-face. Key characteristics of interest include the distribution of degree and the frequency and duration of contacts and their variation in rural and urban areas. These will be stratified by age, gender, role, and day of the week.

Expected results: Resultant data will inform on social contact patterns in rural and urban areas of a previously unstudied population. Ensuing data will be used to parameterize mathematical simulation models of transmission of a range of respiratory viruses, including respiratory

Open Peer Review

Reviewer Status ✓ ✓

	Invited Reviewers	
	1	**2**
REVISED **version 2** published 22 Aug 2019	✓ report	
version 1 published 22 May 2019	? report	✓ report

1 **Adam J. Kucharski**, London School of Hygiene and Tropical Medicine, London, UK

2 **Kin-on Kwok**, The Chinese University of Hong Kong, Hong Kong, China

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

syncytial virus, and used to explore the impact of intervention measures such as vaccination and social distancing.

Keywords
contact networks, contact patterns, wearable proximity sensors, infectious disease transmission, respiratory infections

This article is included in the KEMRI | Wellcome Trust gateway.

Corresponding author: Moses Chapa Kiti (mkiti@kemri-wellcome.org)

Author roles: Kiti MC: Conceptualization, Methodology, Software, Writing – Original Draft Preparation, Writing – Review & Editing; **Melegaro A**: Methodology, Writing – Review & Editing; **Cattuto C**: Methodology, Software, Writing – Review & Editing; **Nokes DJ**: Conceptualization, Funding Acquisition, Methodology, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by the Wellcome Trust [102975] and the DELTAS Africa Initiative [DEL-15-003]. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [107769] and the UK government. The views expressed in this publication are those of the author(s) and not necessarily those of AAS, NEPAD Agency, Wellcome Trust or the UK government.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Kiti MC, Melegaro A, Cattuto C and Nokes DJ. **Study design and protocol for investigating social network patterns in rural and urban schools and households in a coastal setting in Kenya using wearable proximity sensors [version 2; peer review: 2 approved]** Wellcome Open Research 2019, **4**:84 (https://doi.org/10.12688/wellcomeopenres.15268.2)

First published: 22 May 2019, **4**:84 (https://doi.org/10.12688/wellcomeopenres.15268.1)

REVISED **Amendments from Version 1**

1) I have updated the legend in Figure 2 to highlight that the dotted circles represent a classroom or a household.

2) I have included the definition of the term 'n' which is the number of nodes.

Any further responses from the reviewers can be found at the end of the article

Introduction

Background

In infectious disease epidemiology, contact networks consist of individuals (nodes) with connections (edges) between them representing interactions that may lead to infection transmission[1]. For respiratory and other infections that spread via close contact (such as influenza, severe acute respiratory syndrome (SARS), respiratory syncytial virus (RSV) measles, meningitis, ebola, etc.), social contact networks can be used to highlight potential transmission routes[2] and identify targeted intervention strategies through predictive mathematical models. Questionnaire surveys have been conventionally used to collect data on contact patterns[3–6] and networks[7–10], with increasing focus on studies in resource poor settings where disease burden is high[11–14]. Despite providing invaluable "who-contacts-whom" data that can be incorporated into models of infection transmission and control[4,6,15,16], the questionnaire method suffers several setbacks, key being recall bias and low participation rates[6,14,17]. It has been argued that current transmission dynamic models that omit important factors such as frequency, duration and location of contacts do not adequately capture the heterogeneity of transmission that has direct bearing on intervention measures[18–20]. Methods have been advanced to overcome these limitations of diary data, in particular automated data collection methods. These include wireless sensors embedded in portable devices such as mobile phones and customized wearable sensors that use Bluetooth and WiFi[21,22], or low power radio frequencies[23–25], to determine proximity and co-location of users.

Use of wireless proximity sensors to detect social networks

Proximity sensors (henceforth referred to as "sensors") using low-powered radio frequencies have been used in 'closed' settings such as households[26,27], schools[28–30], hospitals[25,31–34], work-places[24,35] and conferences[36] to characterize close contact social networks. The sensor platform in these studies has been designed to collect proximity data only from individuals facing each other while wearing the sensors, representing conversations or actual physical touch that can lead to infection transmission (www.sociopatterns.org). The majority of studies using this platform reported a high participation rate (≥75%), suggesting that an unobtrusive way of data collection requiring minimal participant intervention elicits better response rates compared to paper diaries[14], especially in settings with a high proportion of illiterate individuals[26]. Saturation of study populations to define full networks rapidly encounters boundaries due to logistic, time and cost constraints. However, methods have been developed to effectively impute missing data by generating synthetic networks structure given the underlying properties and also the demographic characteristics of the study population[37]. Most importantly, sensors provide a rich temporal data source, even for partial networks, that can be used to investigate plausible characteristics of infection spread on networks structures weighted by frequency and duration of contacts.

A feasibility study on acceptability and utility of using sensors within five households was conducted in Kilifi over three days of the week[26]. This pioneer study in Africa revealed three key points, particularly relevant to the design of similar studies in resource poor settings. First, individuals were willing to carry the sensors for extended periods of time because they were unobtrusive and did not require user or investigator intervention to collect data. However, the sensors were considered too big to be used on infants. Second, results suggested children spent more time in contact with other children at the household compared to other age groups, while adults appeared to act as bridges between households. Third, within household temporal contact patterns per day were stable across three days of observation, and contacts between individuals of different households were erratic. Residents aged ≥15 years were under-represented due to being away at school or work. This pilot study recommended elaborate community engagement strategies to ensure wider acceptance of study procedures, and proper training of participants on sensor use and storage (e.g. store them separately when sleeping). Current sensors being deployed in recent studies, such as by Ozella *et al.*[27] are smaller, lighter, round in shape and with an on-board flash memory that can store data over longer periods of time. These properties have made the sensors more suitable for use in larger population-based studies, suggesting that they can be deployed in more hard-to-reach populations particularly infants who bear the biggest brunt of respiratory infections.

Infection transmission in schools and households

Schools and households are locations where a high proportion of a population will spend most of their time and individual interactions are frequent and intense[4,6,11,13], potentially leading to a high propensity to spreading of respiratory infections[38–40]. These intense interactions are poorly understood, particularly in the very young infants. Studies have revealed the role of older siblings, mothers and other household members on the transmission of respiratory infections to infants. School-going children are notable introducers and transmitters of respiratory infections to same-household members[41–43] and to members of other households[44]. Infants too spend more time with their mothers compared to other household members[27,45,46], also suggesting the key role that mothers may have to play in the transmission of infection. In addition, other studies revealed that interactions between infants and non-household members were non-negligible but rarely captured[45], suggesting that targeting same-household members only to cocoon infants from infections may have a limited impact on transmission[45,46]. Measures such as school closure have been shown to be effective in reducing the magnitude of outbreaks by infections that spread via close contacts (e.g. influenza)[39,47], largely dependent on the transmissibility of the virus and the type of school closure (e.g. one class vs entire school[42]). As a side effect, school closure results in more age- and location-heterogeneous interactions, particularly with adults at home and with students from other schools[48]. This increases the potential for transmitting infections to students from other schools and to individuals

of working ages, thus suggesting that additional interventions such as vaccinating students from neighbouring schools and adults may need to be considered. Larger studies that may capture contact events within and between households (and school settings) are suggested to provide empirical data that is needed in mathematical models that simulate transmission and assess the impact of various control measures.

Significance and potential impact of the study

To provide greater insight into social network structures in resource poor settings, we propose to study social contact patterns within schools and households and compare and contrast patterns in the urban and rural setting exhibiting different demographic, economic, and socio-cultural characteristics. This will provide key data for use in transmission dynamic models for common respiratory viral and bacterial infections such as RSV and *S. pneumoniae* that are the leading cause of childhood morbidity and mortality in the SSA setting. We also seek to answer the question how we can optimize study design to capture individual and collective properties of networks that are representative of the community.

Objectives

The general objective of this work is to utilize radio frequency close-proximity sensors to describe and understand the nature of human networks within a low-resource population that have the potential to transmit respiratory infectious diseases. Specifically:

(i) To collect data on close-proximity interactions in schools and households in one rural and one urban location in Kilifi.

(ii) To characterize the number, duration, and temporal dynamics of social contacts and to define the network's properties and structure, as well as the underlying determinants within the household and school settings.

(iii) To investigate the potential effect of household and school network structures on the spread of respiratory infections using mathematical models.

Study design

Study design and site (geographical)

This is a cross-sectional study conducted in two locations within the Kilifi Health and Demographic Surveillance System (KHDSS) area[49], namely Matsangoni and Kilifi Township, categorized as rural and urban, respectively (Figure 1). These two sites, similar to the other 13 administrative locations covered by KHDSS within Kilifi County, have been under demographic surveillance from April 2002 onwards, and all household

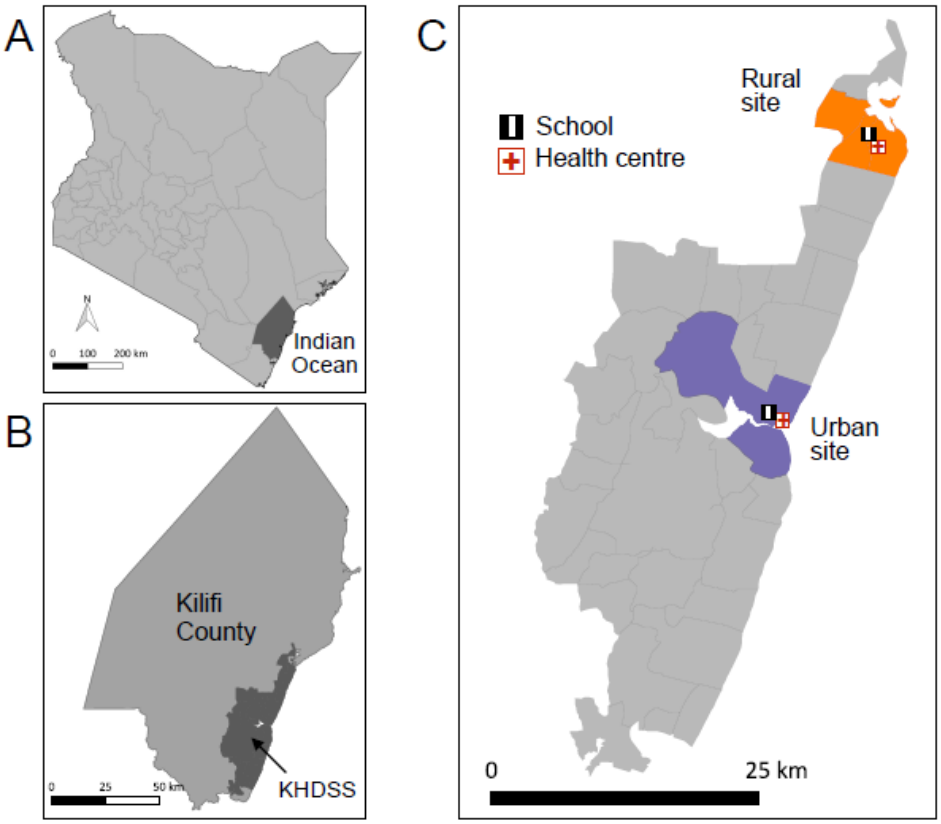

Figure 1. Map showing study locations. (**A**) The location of Kilifi County within Kenya. (**B**) The extent of the Kilifi Health and demographic Surveillance Site (KHDSS) within Kilifi county. (**C**) The locations of rural (orange) and urban (purple) sites, highlighting the location of each of the schools and the nearby health centre.

demographic information (births, deaths, migrations), and geographic location details have been linked to clinical surveillance data at the Kilifi County Hospital thus creating the KHDSS.

Community social network structure may be affected by structural differences in the social, demographic and economic profiles exhibited in the selected rural and urban locations. For example, rural residents are predominantly subsistence farmers and fishermen, while urban residents are formally employed or are small scale business owners. Households are bigger in the rural, compared to urban, area, with the former having several related families living within the same compound[49]. School size ranges from 600–900 in Kilifi, with average class sizes being larger and students being slightly older in the rural compared to urban areas. Box 1 contains definitions of terms used.

Box 1. Definition of terms

Household. A group of individuals eating from the same kitchen and referring to one of these individuals as the head.

Building unit. A dwelling in which individuals live. One or more building units can form a household.

Index household. The household in which an index student selected from the school resides.

Neighbouring household. One that shares a common fence with the index household (rural), or those that are co-located in a compound owned by one individual (urban).

Study populations

Study recruitment will be done in two phases targeting four sites: rural school and rural households, urban school and urban households. The first point of entry will be two primary schools as identified in Figure 1C. Immediately after collecting data from each school, household residents linked to a sub-selection of participating students will be recruited as depicted in Figure 2. For each site, the maximum number of participants is expected to be 350. This number is guided by the number of sensors available for deployment. Further details are given below.

School sampling

Each school will participate if:

(i) The County Education Officer (CEO) gives approval to engage the school. The CEO is responsible for all the administrative matters regarding education in the region. All engagement with the school requires express permission from the officer-in-charge.

(ii) The school includes both primary and early childhood development (ECD) or kindergarten (KG)

(iii) Approval from the Headteacher and school's Board of Management is given.

Each school will be further stratified into preschool (kindergarten), lower primary (grade 1–4) and upper primary (grade 5–8). Students in these grades are generally within the ages 3–5 years, 6–9 years and 10–15 years, respectively. From the school register, a fixed number (350) of students will be randomly selected proportional to grade and gender. The number of participants per school is limited by the number of sensors available for the experiment. All teachers, approximately 20 per school, will also be recruited into the study.

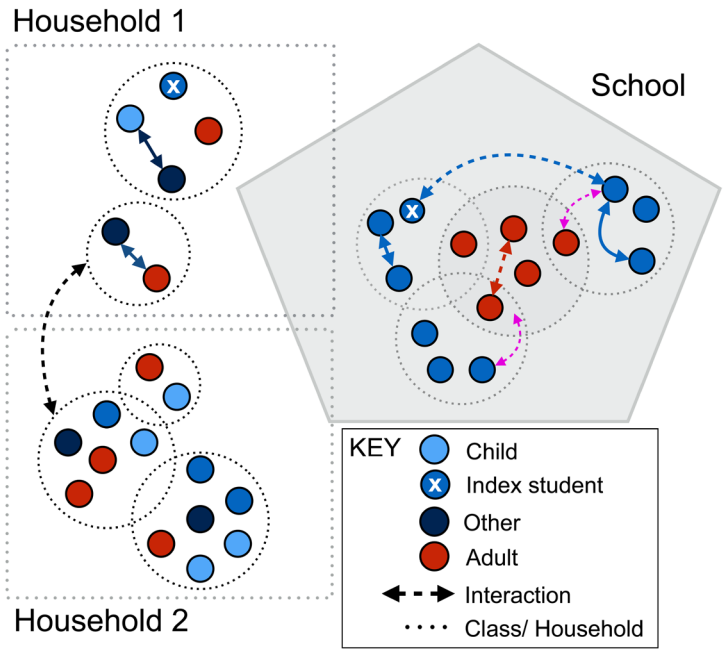

Figure 2. Conceptualization of an interaction framework within and between schools and households. In this scenario, there are children (not attending school), school students and adults. Interactions can occur within or between households, and within and between students and teachers in various classes.

Household sampling. From each of the defined school stratum (preschool, lower and upper primary), three index students will be selected by simple random sampling. This will give an initial 9 index students per school, who will be linked to their households through data available in the KHDSS. For each of the 9 households per setting, an additional 3–5 (minimum-maximum) neighbouring households will be recruited into the study to give an expected minimum of 396 and 324 residents in rural and urban setting, respectively (assuming average household size is 11 and 9 in rural and urban areas, respectively (unpublished KHDSS data). Individuals within households will be eligible to participate if:

(i) They are a member of a household in which the index student lives, or of a household immediately neighbouring the house of an index student.

(ii) Assent from head of household of an index student, or neighbouring household, is provided.

(iii) The household member provides written consent (teachers, adult, caregiver) or assent (child).

Should more than a third of the expected household members refuse (verbally) to participate, the particular household will be excluded from the study. However, this will not apply if members are away from the household for extended periods due to work or school.

Study procedures

Data collection infrastructure and type. Background socio-demographic data for each individual will be extracted from the KHDSS database and updated manually in case some details are missing. Proximity data will be collected using wearable proximity sensors (Figure 3A), henceforth called sensors. The sensors have been developed by the SocioPatterns project (a European consortium of institutions and investigators focused on social dynamics, www.sociopatterns.org). The sensors operate in the 2.4 GHz ISM band of the RF spectrum. The total weight of the sensor inclusive of a lithium coin battery (CR2032) is <6 grams. Sensors exchange ultra-low-power radio packets in a peer-to-peer fashion by transmitting and scanning their neighbourhood for packets sent by nearby tags on a specific radio channel. Sensors in proximity exchange a maximum of 1 data packet per second and can store over 1,000 hours of continuous data collection in an on-board memory. This exchange of low-power radio packets is used as a proxy for the close co-location of individuals wearing the sensors. The proximity between individuals and temporal resolution are estimated from the power levels and timestamps contained in the data packets, respectively. To estimate how close individuals are,

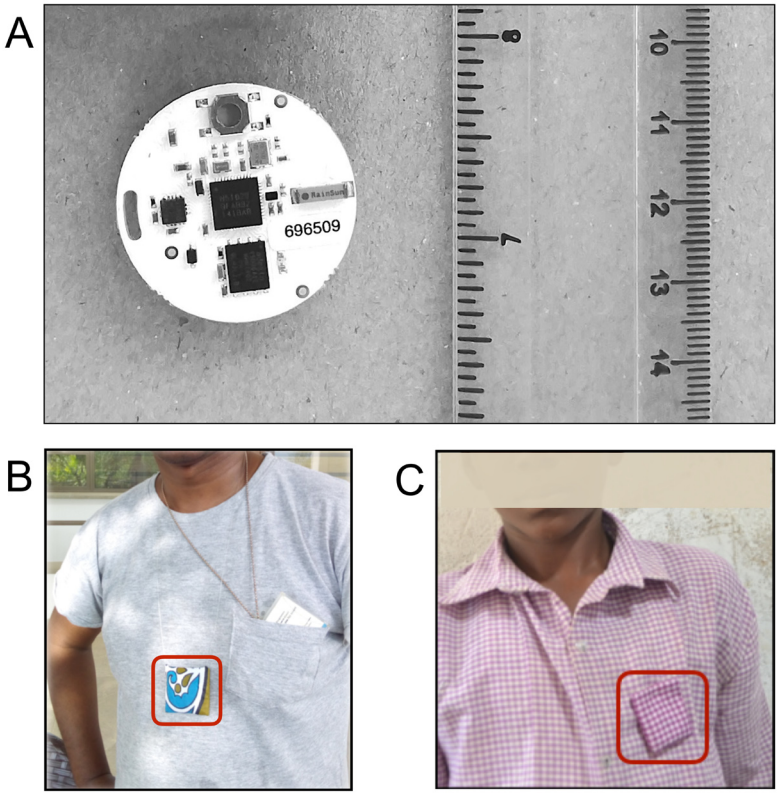

Figure 3. Wireless proximity sensors. (**A**) A sensor next to a ruler. (**B**) and (**C**) How a sensor will be worn by participants, either around the neck with a lanyard or pinned to the shirt pocket, respectively. Household participants will be able to choose the colour of the pouch (**B**). Students will have a pouch with colour similar to their school shirt/blouse.

the attenuation of the signals with distance is computed as the difference between the received and transmitted power. Proximity between individuals corresponding to a face-to-face interaction, such as during a handshake, is asserted when the median attenuation over a given time interval exceeds a specified attenuation threshold (in dBm). All individuals will wear a sensor attached to a lanyard around the neck so that it rests on the chest area (Figure 3B) or pinned to the front of a blouse/shirt especially for younger children (Figure 3C). In this manner, only face-to-face proximity relations will be detected; moreover, the low-power radio frequency in use cannot propagate through the human body[24].

Fieldworkers will ensure participants are properly trained on carrying and storage of devices. During the data collection, some simple measures will be put in place to minimize data loss through not carrying the sensors or deliberately tampering with the sensors. For instance, one class representative will be appointed to ensure that each student participating in the study wears the devices as expected. The head of the household will also be asked to ensure that the correct use of the devices will be maintained throughout the study. This is not expected to be a major role that would have affect the normal class/household routine.

Community engagement. The following community stakeholders will be consulted prior to field engagement: County Administration and local Chiefs and village elders, County Education Office, KCRs and household heads of identified households. A considerable amount of time may be spent at the school since this will be the first point of entry for the study. Where possible, general consent and assent of all household members of index households may be conducted at the same time if present.

(i) At school: the Headteacher in consultation with the Parent-Teacher Association (PTA) will be asked to give written consent for the school to participate in the study. Once the Headteacher has given approval, the students will be identified through the school register. For easier management, the engagement schedule will be broken down by grade. Parents of selected students will either (a) be requested to come to school for a group consenting exercise, or (b) be followed up to their household for consent by matching household records with those of the students. Only students whose parents give consent will be asked for individual assent to participate. Parents who do not attend the meetings will be traced to their homes by matching their records available in the KHDSS database with the students. Teachers and other staff will also provide individual consent. To minimize disruptions to the normal school routine, engagement sessions with the students will be arranged during their normal breaks, such as class recess, lunchtime or sports breaks. All school engagement will be conducted after obtaining requisite approvals from the relevant County Education Officers.

(ii) At home: approval will be requested from the household head to recruit household members. Other residents will provide individual informed consent (≥ 18 years) or assent (13–17 years) as appropriate. Neighbouring households will be identified during the home visits and consent sought from the household heads. Appointments will be scheduled with individual adult household members.

At the completion of the study, joint feedback meetings with study participants and other stakeholders will be organized per site. At each school on appropriate days, all parents, students and staff will be invited to a health awareness meeting. This will focus on basic hygiene such as hand washing, basic science principles of health research and informing the community on various research activities that KEMRI conducts.

Data analysis

The sensor firmware and data cleaning procedures have been developed as part of the custom SocioPatterns software developed for this and other studies. The publicly available version of the SocioPatterns tag firmware is a branch of the OpenBeacon firmware that has been developed, tested and verified by the ISI Foundation and the SocioPatterns project. A contact event is detected if one sensor records a radio packet from another sensor, and if the incoming radio power is higher than a given attenuation threshold, calibrated to corresponded to about 1.5 metres of separation distance[26]. Patterns of contact between participants will be analyzed by statistical distributions describing: a) the number of contacts in households and schools, b) the duration of the contacts, c) the cumulative time spent in contact, and d) the temporal evolution of the networks. Heterogeneity of the contacts and their statistical distributions will be assessed across five key variables: age group (0–4, 5–14, 15–19, 20–49, >50 years), gender (female and male), temporality (hourly, daily and weekly), grade (kindergarten, grades 1–8), and setting (rural/urban). Analysis will be conducted using various Python 2.7 libraries (Numpy v1.12 and Pandas v0.2) and custom and non-public data processing software by the SocioPatterns project (data cleaning and management), R 3.2.1 (statistical analysis and network visualization), Gephi (network visualization) and QGIS (cartography). Network data analysis and visualization will be aggregated at the school and household level with nodes representing students and household residents, respectively. Links between two individuals i and j in contact will be weighted by the cumulative duration of interaction between them. Temporal data will be aggregated into time windows of 10 minutes, hourly, daily and over the entire duration of the study (7 days).

Definition of terms. The primary outcome of interest is the median $\langle k \rangle$ degree and corresponding interquartile range (IQR, 25th and 75th percentiles). For a contact network, the following quantities, similar to a previous household study in Kilifi[26], are defined:

(i) A contact event occurs if at least one data packet is exchanged between two devices during a continuous 20-second time window. A contact is considered broken if a 20-second time window passes without data exchange.

(ii) The degree k_i of a node i is the number of other nodes to which it was linked during a contact event.

(iii) The weight n_{ij} of an edge between nodes i and j is the number of contact events recorded between these individuals during the time window (see Figure 4 for illustration). The mean number of contact events is computed as the sum of the individual contact events divided by the number of nodes, n, $\frac{\sum n_{ij}}{n}$.

(iv) The weight w_{ij} of an edge between nodes i and j is the total duration of contact events recorded between these individuals during a given time window (see Figure 4 for illustration). The mean contact duration was computed as the sum of individual contact duration divided by the number of nodes, $\frac{\sum w_{ij}}{n}$.

(v) The network density is the ratio of the number of observed edges formed in a network to the maximum number of expected edges[27].

(vi) The clustering coefficient measures the cohesiveness of local groups of nodes by calculating the probability of two different contacts of individual i also contacting each other.

(vii) The cosine similarity is defined as an individual's tendency to have repeated contacts with the same individual over two time points t_1 and t_2, taking into account the duration of contact (weight) $w_{ij,1}$ and $w_{ij,2}$ on the edge $i \leftrightarrow j$ measured at time t_1 and t_2[29], calculated as:

$$sim(i) = \frac{\sum_j (w_{ij,1})(w_{ij,2})}{\sqrt{\sum_j (w_{ij,1}^2)}\sqrt{\sum_j w_{ij,2}^2}} \quad (1)$$

Essentially, the cosine similarity calculates the changes in the neighbourhood of each node in each pair of daily networks, suggestive of whether a node i was in contact with and spent the same amount of time with the same nodes for each successive day pair. Cosine similarity takes values ranging from 0 and 1, with lower values suggesting that neighbouring edges are not the same at time t_1 and t_2, while accounting for time spent in contact. In order to assess the magnitude of the cosine similarities, these values will be compared to a null model. The null model reshuffles the weights of the networks among the edges but does change the topology of the network.

To visualize the networks, the Distributed Recursive Graph Layout (DrL) and Fruchterman-Reingold (F-R) force directed algorithms available in the igraph package in R will be used. In a force-directed algorithm, attractive forces act upon the edges while repulsive forces act between nodes. The F-R algorithm minimizes edge crossing and node overlap thus distributing nodes evenly in the visualization frame while ensuring that the lengths of edges are similar[50]. Due to this, nodes are clustered together as the density of the links among them increases. The DrL algorithm aims primarily to minimize the overlap of large clusters, or in other words, to emphasize dense clusters[51].

Ethical considerations

Ethical approval

Ethical approvals were issued by the Scientific and Ethical Review Unit, SERU (KEMRI, Kenya) C/025/3183, and the Biomedical and Scientific Research Ethics Committee, BSREC (University of Warwick, UK) REGO-2016-1738.

Safety concerns

The sensors and loggers have been used in a previous study in Kilifi that involved piloting the use of sensors in the community,

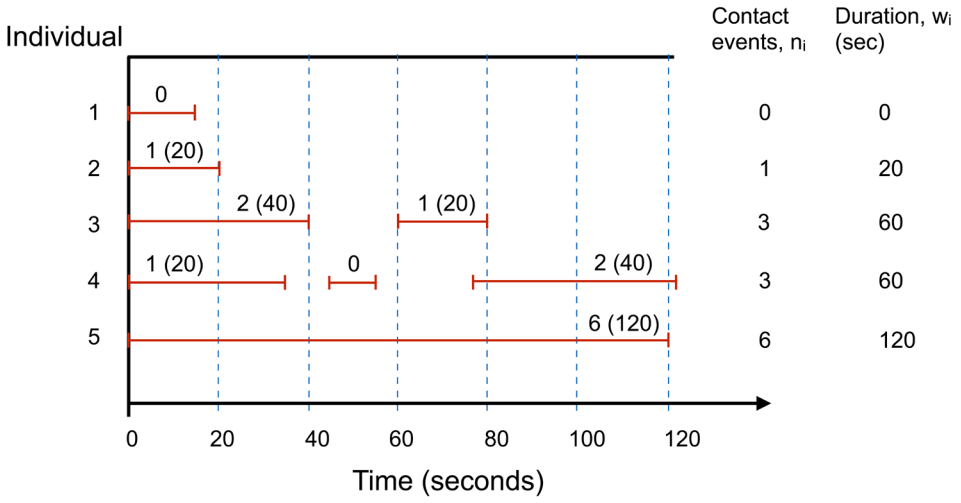

Figure 4. Schema demonstrating the definition of contact event and contact duration. There are 5 individuals. Each red horizontal line represents an interaction between two individuals i (1–5) and j lasting s-seconds. Individual 1 has zero contact events since the duration of contact is <20 seconds. Individual 2 has 1 contact event lasting exactly 20 secs. Individual 3 has three contact events with a 20-sec interruption between the contact events. Individuals 4 and 5 have three and six contact events lasting 60- and 120-seconds, respectively.

understanding community concerns and learning best practice methods for deployment[26], as well as in several other referenced studies. There are no known risks posed by the low power frequency signals emitted by the sensors. The devices will be inserted in a zipped pouch for personal safety, to minimize device loss by theft or misplacement, or data loss through tampering by the participants.

Informed consent

The informed consent process will be undertaken as described in the community engagement procedures. Consent and assent forms will be back-translated from English to two local languages, Swahili and Giriama. Participants will be free to choose the language in which they would like the information presented. Participants will be free to leave the study or to request the withdrawal of their data at any time and for whatever reason without explanation and without penalty.

Benefits to participants

Parents who attend the engagement sessions at the school will be refunded travel expenses. We anticipate that this will not exceed USD 2 (~KES 200.00). Participating households will also benefit from health talks that focus on prevention of communicable diseases such as pneumonia and diarrhoea. This will include talks on importance of washings hands (before and after visiting toilets, before handling food, before handling infants, etc), use of handkerchiefs or disposable tissues when sneezing or coughing, and a demonstration of proper hand washing techniques. Each household will then receive two bars of hand-washing soap at the end of data collection.

Schools will benefit from science and health talks from the research team after the data collection. Simple messages focusing on personal hygiene practices that prevent the spread of flu-like infections will be emphasized. Each class will also receive two bars of hand-washing soap. Further, all participating students will receive a stationery pack containing writing materials (writing books, pencils, ruler, mathematical set) and a storybook with information on how to prevent the spreading of flu-like infections.

Discussion

Respiratory infections and other diseases that are transmitted through close contacts are a predominant cause of morbidity, mortality and healthcare spending in developing countries. Social contact data are important to understand infection patterns since they underpin the transmission dynamics and are key input parameters in mathematical models that evaluate preventative and control measures against these diseases. To date, very few studies have been conducted in sub-Saharan Africa to elucidate the mechanisms of spread. This study proposed to use wireless proximity sensors to collect data from students in schools and household residents in rural and urban areas of coastal Kenya. Schools and households have been identified as the hubs of transmission for respiratory infections such as RSV[41,52,53] particularly due to prolonged and more intimate contacts at these settings. Results from this study will include the number and duration of contacts and how they vary by age, gender, day of the week, role, and location. With a vaccine against RSV imminent[54], predictive modelling can be used to support decision making at the national level for control of infectious disease and important to be based on locally collected data.

There are some limitations to this study. First, it is not possible to describe full social networks at participating school and households due to the limited number of sensors available. Nonetheless, this protocol will investigate contacts within schools and households, settings that are important for spread of respiratory infections. An attempt to minimize selection bias is through random selection of students in a school and linking the students to their entire households and neighbours. This study will also develop the body of knowledge on longitudinal patterns of social networks in rural and urban communities in a developing world setting. Due to the spatially restricted data collection sites, it will not be possible to generalize with certainty these results to other settings locally and globally because of differences in demography, social and cultural attributes. This suggests the need to collect additional national and regional studies, with the advantage that this study protocol can be adapted for use elsewhere since the sensors can be reused. In addition to reusability, the current sensors offer additional advantages when compared to a pilot study conducted in the same setting[26]: they are smaller and lighter, can be worn by people of all ages including infants, have a bigger memory and battery capacity meaning that data can be collected over longer periods. This may be effectively used to assess the effect of seasonality on changes in social contacts and disease transmission.

Dissemination

A lay summary of the results will be shared with the participating schools and households, as well as and communities surrounding the schools. Manuscripts will be submitted to appropriate journals discussing the methods, statistical analysis and output, and applications of the data in mathematical modelling of respiratory disease transmission and control.

Data storage and distribution

Anonymized data will be stored in a repository available online through the KWTRP Research Data Repository on Harvard Dataverse and the SocioPatterns website so as to be findable, accessible, interoperable and reusable (FAIR). Data in de-identified format will be open access to support future use by the wider research community or replication. Access to identifiable data by people outside the investigators and specific collaborators will require permission from the senior investigators, the Data Governance Committee in KEMRI-WTRP, and where necessary, National Ethics Committee. In future, we hope that information collected or generated during this study will be used to support new research by other researchers in Kenya and other countries on other health problems.

Study status

Data collection started in August 2016 and was completed in April 2017. Currently (20/03/2019), data analysis is ongoing.

Data availability

No data are associated with this article.

Grant information

This work was supported by the Wellcome Trust [102975] and the DELTAS Africa Initiative [DEL-15-003]. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [107769] and the UK government. The views expressed in this publication are those of the author(s) and not necessarily those of AAS, NEPAD Agency, Wellcome Trust or the UK government.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

The study group would like to thank the students and household residents who have participated in the study. We would also like to thank the school administration, data management team, and research assistants across the two sites.

References

1. Keeling MJ, Eames KT: **Networks and epidemic models.** *J R Soc Interface.* 2005; **2**(4): 295–307.
PubMed Abstract | Publisher Full Text | Free Full Text
2. Danon L, House TA, Read JM, *et al.*: **Social encounter networks: collective properties and disease transmission.** *J R Soc Interface.* 2012; **9**(76): 2826–2833.
PubMed Abstract | Publisher Full Text | Free Full Text
3. Kucharski AJ, Kwok KO, Wei VW, *et al.*: **The contribution of social behaviour to the transmission of influenza A in a human population.** *PLoS Pathog.* 2014; **10**(6): e1004206.
PubMed Abstract | Publisher Full Text | Free Full Text
4. Horby P, Pham QT, Hens N, *et al.*: **Social contact patterns in Vietnam and implications for the control of infectious diseases.** *PLoS One.* 2011; **6**(2): e16965.
PubMed Abstract | Publisher Full Text | Free Full Text
5. Hens N, Goeyvaerts N, Aerts M, *et al.*: **Mining social mixing patterns for infectious disease models based on a two-day population survey in Belgium.** *BMC Infect Dis.* 2009; **9**: 5.
PubMed Abstract | Publisher Full Text | Free Full Text
6. Mossong J, Hens N, Jit M, *et al.*: **Social contacts and mixing patterns relevant to the spread of infectious diseases.** *PLoS Med.* 2008; **5**(3): e74.
PubMed Abstract | Publisher Full Text | Free Full Text
7. Stein ML, Van Steenbergen JE, Buskens V, *et al.*: **Comparison of contact patterns relevant for transmission of respiratory pathogens in Thailand and The Netherlands using respondent-driven sampling.** *PLoS One.* 2014; **9**(11): e113711.
PubMed Abstract | Publisher Full Text | Free Full Text
8. Danon L, Read JM, House TA, *et al.*: **Social encounter networks: characterizing Great Britain.** *Proc Biol Sci.* 2013; **280**(1765): 20131037.
PubMed Abstract | Publisher Full Text | Free Full Text
9. Conlan AJ, Eames KT, Gage JA, *et al.*: **Measuring social networks in British primary schools through scientific engagement.** *Proc R Soc B Biol Sci.* 2011; **278**(1711): 1467–1475.
PubMed Abstract | Publisher Full Text | Free Full Text
10. Glass LM, Glass RJ: **Social contact networks for the spread of pandemic influenza in children and teenagers.** *BMC Public Health.* 2008; **8**(1): 61.
PubMed Abstract | Publisher Full Text | Free Full Text
11. Melegaro A, Del Fava E, Poletti P, *et al.*: **Social Contact Structures and Time Use Patterns in the Manicaland Province of Zimbabwe.** *PLoS One.* 2017; **12**(1): e0170459.
PubMed Abstract | Publisher Full Text | Free Full Text
12. Grijalva CG, Goeyvaerts N, Verastegui H, *et al.*: **A household-based study of contact networks relevant for the spread of infectious diseases in the highlands of Peru.** *PLoS One.* 2015; **10**(3): e0118457.
PubMed Abstract | Publisher Full Text | Free Full Text
13. Johnstone-Robertson SP, Mark D, Morrow C, *et al.*: **Social mixing patterns within a South African township community: implications for respiratory disease transmission and control.** *Am J Epidemiol.* 2011; **174**(11): 1246–1255.
PubMed Abstract | Publisher Full Text | Free Full Text
14. Kiti MC, Kinyanjui TM, Koech DC, *et al.*: **Quantifying age-related rates of social contact using diaries in a rural coastal population of Kenya.** *PLoS One.* 2014; **9**(8): e104786.
PubMed Abstract | Publisher Full Text | Free Full Text
15. Eames KT, Tilston NL, Brooks-Pollock E, *et al.*: **Measured dynamic social contact patterns explain the spread of H1N1v influenza.** *PLoS Comput Biol.* 2012; **8**(3): e1002425.
PubMed Abstract | Publisher Full Text | Free Full Text
16. You SH, Chen SC, Wang CH, *et al.*: **Linking contact behavior and droplet patterns to dynamically model indoor respiratory infections among schoolchildren.** *J Epidemiol.* 2013; **23**(4): 251–261.
PubMed Abstract | Publisher Full Text | Free Full Text
17. Edmunds WJ, O'Callaghan CJ, Nokes DJ: **Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections.** *Proc Biol Sci.* 1997; **264**(1384): 949–957.
PubMed Abstract | Publisher Full Text | Free Full Text
18. Smieszek T: **A mechanistic model of infection: why duration and intensity of contacts should be included in models of disease spread.** *Theor Biol Med Model.* 2009; **6**(1): 25.
PubMed Abstract | Publisher Full Text | Free Full Text
19. De Cao E, Zagheni E, Manfredi P, *et al.*: **The relative importance of frequency of contacts and duration of exposure for the spread of directly transmitted infections.** *Biostatistics.* 2014; **15**(3): 470–483.
PubMed Abstract | Publisher Full Text
20. Toth DJ, Leecaster M, Pettey WB, *et al.*: **The role of heterogeneity in contact timing and duration in network models of influenza spread in schools.** *J R Soc Interface.* 2015; **12**(108): 20150279.
PubMed Abstract | Publisher Full Text | Free Full Text
21. Eagle N, Pentland AS, Lazer D: **Inferring friendship network structure by using mobile phone data.** *Proc Natl Acad Sci U S A.* 2009; **106**(36): 15274–15278.
PubMed Abstract | Publisher Full Text | Free Full Text
22. Stopczynski A, Sekara V, Sapiezynski P, *et al.*: **Measuring large-scale social networks with high resolution.** *PLoS One.* 2014; **9**(4): e95978.
PubMed Abstract | Publisher Full Text | Free Full Text
23. Polastre J, Szewezyk R, Culler D: **Telos: enabling ultra-low power wireless research**. In: *Information Processing in Sensor Networks.* 2005; 364–369.
Publisher Full Text
24. Cattuto C, Van den Broeck W, Barrat A, *et al.*: **Dynamics of person-to-person interactions from distributed RFID sensor networks.** *PLoS One.* 2010; **5**(7): e11596.
PubMed Abstract | Publisher Full Text | Free Full Text
25. Hornbeck T, Naylor D, Segre AM, *et al.*: **Using sensor networks to study the effect of peripatetic healthcare workers on the spread of hospital-associated infections.** *J Infect Dis.* 2012; **206**(10): 1549–1557.
PubMed Abstract | Publisher Full Text | Free Full Text
26. Kiti MC, Tizzoni M, Kinyanjui TM, *et al.*: **Quantifying social contacts in a household setting of rural Kenya using wearable proximity sensors.** *EPJ Data Sci.* 2016; **5**(1): 21.
PubMed Abstract | Publisher Full Text | Free Full Text
27. Ozella L, Gesualdo F, Tizzoni M, *et al.*: **Close encounters between infants and household members measured through wearable proximity sensors.** *PLoS One.* 2018; **13**(6): e0198733.
PubMed Abstract | Publisher Full Text | Free Full Text
28. Fournet J, Barrat A: **Contact patterns among high school students.** *PLoS One.* 2014; **9**(9): e107878.
PubMed Abstract | Publisher Full Text | Free Full Text
29. Stehlé J, Voirin N, Barrat A, *et al.*: **High-resolution measurements of face-to-face contact patterns in a primary school.** *PLoS One.* 2011; **6**(8): e23176.
PubMed Abstract | Publisher Full Text | Free Full Text
30. Salathé M, Kazandjieva M, Lee JW, *et al.*: **A high-resolution human contact network for infectious disease transmission.** *Proc Natl Acad Sci U S A.* 2010; **107**(51): 22020–22025.
PubMed Abstract | Publisher Full Text | Free Full Text
31. Vanhems P, Barrat A, Cattuto C, *et al.*: **Estimating potential infection transmission routes in hospital wards using wearable proximity sensors.** *PLoS One.* 2013; **8**(9): e73970.
PubMed Abstract | Publisher Full Text | Free Full Text
32. Voirin N, Payet C, Barrat A, *et al.*: **Combining high-resolution contact data**

with virological data to investigate influenza transmission in a tertiary care hospital. *Infect Control Hosp Epidemiol.* 2015; **36**(3): 254–260.
PubMed Abstract | Publisher Full Text

33. Isella L, Romano M, Barrat A, *et al.*: **Close encounters in a pediatric ward: measuring face-to-face proximity and mixing patterns with wearable sensors.** *PLoS One.* 2011; **6**(2): e17144.
PubMed Abstract | Publisher Full Text | Free Full Text

34. Ozella L, Gauvin L, Carenzo L, *et al.*: **Wearable Proximity Sensors for Monitoring a Mass Casualty Incident Exercise: Feasibility Study.** *J Med Internet Res.* 2019; **21**(4): e12251.
PubMed Abstract | Publisher Full Text | Free Full Text

35. Génois M, Vestergaard CL, Fournet J, *et al.*: **Data on face-to-face contacts in an office building suggest a low-cost vaccination strategy based on community linkers.** *Netw Sci.* 2015; **3**(3): 326–347.
Publisher Full Text

36. Stehlé J, Voirin N, Barrat A, *et al.*: **Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees.** *BMC Med.* 2011; **9**(1): 87.
PubMed Abstract | Publisher Full Text | Free Full Text

37. Génois M, Vestergaard CL, Cattuto C, *et al.*: **Compensating for population sampling in simulations of epidemic spread on temporal contact networks.** *Nat Commun.* 2015; **6**: 8860.
PubMed Abstract | Publisher Full Text | Free Full Text

38. Viboud C, Boëlle PY, Cauchemez S, *et al.*: **Risk factors of influenza transmission in households.** *Br J Gen Pract.* 2004; **54**(506): 684–689.
PubMed Abstract | Free Full Text

39. Hens N, Ayele GM, Goeyvaerts N, *et al.*: **Estimating the impact of school closure on social mixing behaviour and the transmission of close contact infections in eight European countries.** *BMC Infect Dis.* 2009; **9**(108): 187.
PubMed Abstract | Publisher Full Text | Free Full Text

40. Baguelin M, Flasche S, Camacho A, *et al.*: **Assessing optimal target populations for influenza vaccination programmes: an evidence synthesis and modelling study.** *PLoS Med.* 2013; **10**(10): e1001527.
PubMed Abstract | Publisher Full Text | Free Full Text

41. Munywoki PK, Koech DC, Agoti CN, *et al.*: **The source of respiratory syncytial virus infection in infants: a household cohort study in rural Kenya.** *J Infect Dis.* 2014; **209**(11): 1685–1692.
PubMed Abstract | Publisher Full Text | Free Full Text

42. Gemmetto V, Barrat A, Cattuto C: **Mitigation of infectious disease at school: targeted class closure vs school closure.** *BMC Infect Dis.* 2014; **14**(1): 695.
PubMed Abstract | Publisher Full Text | Free Full Text

43. Cauchemez S, Carrat F, Viboud C, *et al.*: **A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data.** *Stat Med.* 2004; **23**(22): 3469–3487.
PubMed Abstract | Publisher Full Text

44. le Polain de Waroux O, Cohuet S, Ndazima D, *et al.*: **Characteristics of human encounters and social mixing patterns relevant to infectious diseases spread by close contact: a survey in Southwest Uganda.** *BMC Infect Dis.* 2018; **18**(1): 172.
PubMed Abstract | Publisher Full Text | Free Full Text

45. van Hoek AJ, Andrews N, Campbell H, *et al.*: **The social life of infants in the context of infectious disease transmission; social contacts and mixing patterns of the very young.** *PLoS One.* 2013; **8**(10): e76180.
PubMed Abstract | Publisher Full Text | Free Full Text

46. Campbell PT, McVernon J, Shrestha N, *et al.*: **Who's holding the baby? A prospective diary study of the contact patterns of mothers with an infant.** *BMC Infect Dis.* 2017; **17**(1): 634.
PubMed Abstract | Publisher Full Text | Free Full Text

47. Halder N, Kelso JK, Milne GJ: **Developing guidelines for school closure interventions to be used during a future influenza pandemic.** *BMC Infect Dis.* 2010; **10**(1): 221.
PubMed Abstract | Publisher Full Text | Free Full Text

48. Eames KT, Tilston NL, Edmunds WJ: **The impact of school holidays on the social mixing patterns of school children.** *Epidemics.* 2011; **3**(2): 103–108.
PubMed Abstract | Publisher Full Text

49. Scott JA, Bauni E, Moisi JC, *et al.*: **Profile: The Kilifi Health and Demographic Surveillance System (KHDSS).** *Int J Epidemiol.* 2012; **41**(3): 650–657.
PubMed Abstract | Publisher Full Text | Free Full Text

50. Fruchterman TMJ, Reingold EM: **Graph drawing by force-directed placement.** *Softw Pract Exp.* 1991; **21**(11): 1129–1164.
Publisher Full Text

51. Martin S, Brown WM, Klavans R, *et al.*: **DrL: Distributed Recursive (Graph) Layout**. Albuquerque; 2008.
Reference Source

52. Poletti P, Merler S, Ajelli M, *et al.*: **Evaluating vaccination strategies for reducing infant respiratory syncytial virus infection in low-income settings.** *BMC Med.* 2015; **13**(1): 49.
PubMed Abstract | Publisher Full Text | Free Full Text

53. Munywoki PK, Koech DC, Agoti CN, *et al.*: **Continuous Invasion by Respiratory Viruses Observed in Rural Households During a Respiratory Syncytial Virus Seasonal Outbreak in Coastal Kenya.** *Clin Infect Dis.* 2018; **67**(10): 1559–1567.
PubMed Abstract | Publisher Full Text | Free Full Text

54. Mazur NI, Higgins D, Nunes MC, *et al.*: **The respiratory syncytial virus vaccine landscape: lessons from the graveyard and promising candidates.** *Lancet Infect Dis.* 2018; **18**(10): e295–e311.
PubMed Abstract | Publisher Full Text

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 2

Reviewer Report 29 August 2019

https://doi.org/10.21956/wellcomeopenres.16859.r36298

✔ **Adam J. Kucharski**
Centre for the Mathematical Modelling of Infectious Diseases (CMMID), London School of Hygiene and Tropical Medicine, London, UK

The authors have addressed my main comments, and hence the article is suitable for approval, subject to one minor final point: I would encourage the authors to consider the logistics in my previous comment 1 (household selection). Their response did not fully address the issue of how they would choose the order in which neighbouring households are selected – presumably it will be random, but worth being clear about this.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Infectious disease epidemiology and modelling.

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 15 August 2019

https://doi.org/10.21956/wellcomeopenres.16665.r35915

✔

Kin-on Kwok
The Jockey Club School of Public Health and Primary Care, Stanley Ho Centre for Emerging Infectious Diseases, The Chinese University of Hong Kong, Hong Kong, China

The study design looks quite sound in this protocol. Data generated from this study can be used as the important input in mathematical models to understand transmission dynamics of respiratory diseases in Kenya.
I had few comments here:

1. Only 9 index cases were randomly recruited in each school in this study according to proportion of gender and grade. In total only 18 kids were recruited from a rural and an urban area in this study while a relative larger number of teachers (in total 40) were recruited. Is the sample size for the school-aged children large enough to give a good level of resolution of data to reflect kid's contact in school?

2. For the recruitment of household member of each index case, if not all, say more than 2/3 of household members, participated study e.g. mother is not willing to take part of this study. How does the data in the mathematical model reflect the role of those absentees of this study contributing to the disease transmission?

3. I tried to see how number of subjects recruited (396 and 324 residents) can be achieved based on my best understanding. For a recruited school in an urban area, all family members of 9 index cases, 20 teachers and 5 neighbouring households of the index will arrive at 374 residents (assuming 11 households members in each household). Are teachers recruited in this study in the same class as the selected index cases?

4. It is a cross sectional study. Will there be any other studies in Kenya or neighbouring countries to justify whether there are any seasonal effects on the contact patterns as described in longitudinal contact survey in Hong Kong study[1]. Will the 7 day period be the same for all individuals?

5. The author mentioned that this study has an advantage over the diary based data. It is worth to see the discrepancy of between two approaches in this study.

6. Some symbols are not defined (such as n) and consistent with each other (such as ni, nij, wij and wi, # in equation 1, sim (i) should be sim(ij)?).

References

1. Kwok KO, Cowling B, Wei V, Riley S, Read JM: Temporal variation of human encounters and the number of locations in which they occur: a longitudinal study of Hong Kong residents.*J R Soc Interface*. **15** (138). 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?
Yes

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

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Mathematical modelling, Infectious diseases epidemiology.

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 19 Aug 2019

Moses Kiti, KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya

1. Only 9 index cases were randomly recruited in each school in this study according to the proportion of gender and grade. In total, only 18 kids were recruited from a rural and urban area in this study while a relatively larger number of teachers (in total 40) were recruited. Is the sample size for the school-aged children large enough to give a good level of resolution of data to reflect kid's contact in school?
The actual number of students expected to be recruited randomly per school is 350. This sample size is limited by the number of sensors available for the study, and we expect to get a good resolution of data to reflect the students' contacts in school. Out of these 350 students, 9 index students will be randomly selected and followed up at their households.

2. For the recruitment of household member of each index case, if not all, say more than 2/3 of household members, participated study e.g. mother is not willing to take part in this study. How does the data in the mathematical model reflect the role of those absentees of this study contributing to the disease transmission?
If more than a third of household members are unwilling to participate, that household will not be recruited. However, this household will be replaced by another neighbouring household until the targeted number of households is achieved. In this protocol, we have not provided the full details of the individual-based model. In brief, the model will simulate a population of N-individuals based on the age profile of all residents enumerated in the KHDSS. Each individual will be assigned an expected number of contacts based on the age-specific probability of contact drawn from the data ensuing from this study.

3. I tried to see how the number of subjects recruited (396 and 324 residents) can be achieved based on my best understanding. For a recruited school in an urban area, all family members of 9 index cases, 20 teachers and 5 neighbouring households of the index will arrive at 374 residents (assuming 11 households members in each household). Are teachers recruited in this study in the same class as the selected index cases?
From each school, 9 index students will be selected. We will then recruit the entire households of the index students. We will also recruit all residents in 3-5 neighbouring households. Assuming we recruit 3 neighbouring households per index, this will result in {(9*11) + (9*3*11)} = 396 participants for the rural area, as given in the manuscript. The teachers' household members teachers will not be recruited and hence do not account for this number.

4. It is a cross-sectional study. Will there be any other studies in Kenya or neighbouring countries

to justify whether there are any seasonal effects on the contact patterns as described in longitudinal contact survey in Hong Kong study[1]. Will the 7-day period be the same for all individuals?
This protocol describes data that will be collected within two sites only (urban and rural) in coastal Kenya. There are no plans to collect data from other regions in Kenya or neighbouring countries in this protocol, but this can be considered for the future. The 7-day period will be the same for all participants in each of the 4 sites (rural school, rural households, urban school, urban households), but data from each site will be collected at separate time-points.

5. The author mentioned that this study has an advantage over the diary-based data. It is worth to see the discrepancy between two approaches in this study.
Yes, it is, but we do not plan to collect any diary-based data.

6. Some symbols are not defined (such as n) and consistent with each other (such as ni, nij, wij and wi, # in equation 1, sim (i) should be sim(ij)?).
***n* refers to the number of sensors, this has now been defined in the protocol. Equation 1 is sim(i) because it refers to the cosine similarity of individual *i*.**

Competing Interests: None.

Reviewer Report 07 June 2019

https://doi.org/10.21956/wellcomeopenres.16665.r35602

?

Adam J. Kucharski
Centre for the Mathematical Modelling of Infectious Diseases (CMMID), London School of Hygiene and Tropical Medicine, London, UK

The protocol describes the design of study looking at social interactions in different contexts in Kilifi. It is well motivated, and should generate some nice data relevant to disease transmission.

I had the following, mostly minor, comments:

- What is the selection process for neighbouring households? For example, if there are 6 neighbouring households, how will only 3-5 be selected? Or if there are only 2 neighbouring households, will a third be chosen? What will happen if some decline to participate?
- Will the 7 day collection period be the same for all individuals in the study? Or will it be site specific?
- Are there plans to collect data on the role within a family (father, mother etc.)? This might be useful for interpreting heterogeneity in contact patterns beyond just age.
- In Figure 2, it wasn't clear to me what the dotted circles represented.
- "the questionnaire method suffers several setbacks, key being recall bias and low participation rates" - there are some studies comparing diaries and sensors (e.g. Mastrandrea *et al.*, 2015[1]), so

it might be worth considering how the proposed proximity data might relate to existing diary-based studies.

References

1. Mastrandrea R, Fournet J, Barrat A: Contact Patterns in a High School: A Comparison between Data Collected Using Wearable Sensors, Contact Diaries and Friendship Surveys.*PLoS One*. 2015; **10** (9): e0136497 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: Infectious disease epidemiology and modelling.

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 19 Aug 2019

Moses Kiti, KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya

1. What is the selection process for neighbouring households? For example, if there are 6 neighbouring households, how will only 3-5 be selected? Or if there are only 2 neighbouring households, will a third be chosen? What will happen if some decline to participate?
All households within the Kilifi Health and Demographic Surveillance Site (KHDSS) have been mapped. The selected index households will be highlighted in a map, and the neighbouring households will be identified physically on the ground. Normally, these neighbouring households would share a boundary with the index household. A research assistant will visit the households in succession and recruit a maximum of five. If the head of a household declines participation, then that household will be skipped.

2. Will the 7-day collection period be the same for all individuals in the study? Or will it be site-specific?
There are four study sites: rural school, rural households, urban school and urban households. The 7-day period of data collection will be site-specific, and all individuals within the same site will collect data over the same period.

3. Are there plans to collect data on the role within a family (father, mother etc.)? This might be

useful for interpreting heterogeneity in contact patterns beyond just age.
Data on family roles will be available from the KHDSS data and will not be collected again.

4. In Figure 2, it wasn't clear to me what the dotted circles represented.
The dotted lines represent a classroom or a household. The figure legend has been updated to include this.

5. "the questionnaire method suffers several setbacks, the key being recall bias and low participation rates" - there are some studies comparing diaries and sensors (e.g. Mastrandrea *et al.*, 2015), so it might be worth considering how the proposed proximity data might relate to existing diary-based studies.
We considered this and decided not to collect any paper diary data. We have previously collected data using paper diaries *(Kiti et al., 2014)* and sensors *(Kiti et al., 2016)*, and the decision to collect data using sensors as per this protocol was guided by:
1) the ability of individuals to self-record their contacts (two-thirds of participants required shadows in the paper-diary study)
2) the overall success of the pilot study using sensors in a rural community.

Competing Interests: None.