

1 Comprehensive profiling of social mixing patterns in resource poor countries: a mixed methods research
2 protocol
3

4 Obianuju Genevieve Aguolu¹³, Moses Chapa Kiti³, Kristin Nelson³, Carol Y. Liu³, Maria Sundaram¹¹,
5 Sergio Gramacho³, Samuel Jenness³, Alessia Melegaro⁴, Charfudin Sacoor⁵, Azucena Bardaji^{5,6,12}, Ivalda
6 Macicame⁷, Americo Jose⁷, Nilzio Cavele⁷, Felizarda Amosse⁵, Migdalia Uamba⁷, Edgar Jamisse⁵,
7 Corssino Tchavana⁵, Herberth Giovanni Maldonado Briones⁸, Claudia Jarquín⁸, María Ajsivinac⁸,
8 Lauren Pischel², Noureen Ahmed¹, Venkata Raghava Mohan⁹, Rajan Srinivasan⁹, Prasanna Samuel⁹,
9 Gifta John⁹, Kye Ellington³, Orvalho Augusto Joaquim⁵, Alana Zelaya³, Sara Kim³, Holin Chen³, Momin
10 Kazi¹⁰, Fauzia Malik¹, Inci Yildirim², Benjamin Lopman^{3*}, Saad B. Omer^{1*}

11
12 Affiliations

13 ¹ Peter O'Donnell Jr. School of Public Health at UT Southwestern Medical Center, Dallas, Texas

14 ² Yale School of Medicine, Yale University, Connecticut, USA

15 ³ Rollins School of Public Health, Emory University, Georgia, USA

16 ⁴ DONDENA Centre for Research in Social Dynamics and Public Policy, Bocconi University, Italy

17 ⁵ Manhica Health Research Centre, Manhica, Mozambique

18 ⁶ ISGlobal, Hospital Clinic – Universitat de Barcelona, Barcelona, Spain

19 ⁷ Polana Caniço Health Research and Training Centre, CISPOC, Mozambique

20 ⁸ Centro de Estudios en Salud (CES), Universidad del Valle de Guatemala

21 ⁹ Christian Medical College Vellore, India

22 ¹⁰ The Aga Khan University, Karachi, Pakistan

23 ¹¹ Center for Clinical Epidemiology and Population Health, Marshfield Clinic Research Institute,
24 Marshfield, Wisconsin, USA

25 ¹² Consorcio de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Madrid,
26 Spain

27 ¹³ Division of Epidemiology, College of Public Health, The Ohio State University, Columbus, Ohio

28
29 *Contributed equally as co-Principal Investigators

30
31 Corresponding author

32 Obianuju Genevieve Aguolu, MBBS, MPH, PhD

33 Clinical Assistant Professor

34 Division of Epidemiology

35 College of Public Health

36 The Ohio State University

37 346 Cunz Hall, 1841 Neil Avenue, Columbus, OH 43210

38 aguolu.1@osu.edu

39
40
41
42
43
44
45
46
47
48
49

50 **Funding**

51 This work is supported by the National Institutes of Health (NIH) [grant number: 1 R01 HD097175-01].

52 The views expressed in this publication are those of the author(s) and not necessarily those of NIH or the
53 US government. The funders had no role in study design, data collection and analysis, decision to publish,
54 or preparation of the manuscript.

55

56 **Competing interests**

57 No competing interests were disclosed.

58

59 **Data Availability**

60 All data and code generated during the current study will be made available via our GitHub repository.

61 The data will also be made available to download on Zenodo.

62

63

64

65

66

67

68 **ABSTRACT**

69 **Background:** Low-and-middle-income countries (LMICs) bear a disproportionate burden of
70 communicable diseases. Social interaction data inform infectious disease models and disease prevention
71 strategies. The variations in demographics and contact patterns across ages, cultures, and locations
72 significantly impact infectious disease dynamics and pathogen transmission. LMICs lack sufficient social
73 interaction data for infectious disease modeling.

74 **Methods:** To address this gap, we will collect qualitative and quantitative data from eight study sites
75 (encompassing both rural and urban settings) across Guatemala, India, Pakistan, and Mozambique. We
76 will conduct focus group discussions and cognitive interviews to assess the feasibility and acceptability of
77 our data collection tools at each site. Thematic and rapid analyses will help to identify key themes and
78 categories through coding, guiding the design of quantitative data collection tools (enrollment survey,
79 contact diaries, exit survey, and wearable proximity sensors) and the implementation of study procedures.

80 We will create three age-specific contact matrices (physical, nonphysical, and both) at each study site
81 using data from standardized contact diaries to characterize the patterns of social mixing. Regression
82 analysis will be conducted to identify key drivers of contacts. We will comprehensively profile the
83 frequency, duration, and intensity of infants' interactions with household members using high resolution
84 data from the proximity sensors and calculating infants' proximity score (fraction of time spent by each
85 household member in proximity with the infant, over the total infant contact time) for each household
86 member.

87 **Discussion:** Our qualitative data yielded insights into the perceptions and acceptability of contact diaries
88 and wearable proximity sensors for collecting social mixing data in LMICs. The quantitative data will
89 allow a more accurate representation of human interactions that lead to the transmission of pathogens
90 through close contact in LMICs. Our findings will provide more appropriate social mixing data for
91 parameterizing mathematical models of LMIC populations. Our study tools could be adapted for other
92 studies.

93 **Keywords:** social mixing, low-and-middle-income countries, mixed methods study, mathematical
94 modeling

95

96

97

98

99

100

101

102

103

104

105 **Introduction**

106 Infectious diseases account for almost half of all child deaths worldwide.(1) Diarrhea, lower respiratory
107 infections, and malaria are the leading causes of death among the 5-14-year-old age group.(1) The
108 world's poorest billion bear a disproportionate burden of these diseases.(2) For example, 91% of all
109 deaths due to diarrheal diseases and 64% of all deaths due to respiratory illnesses occur in low-and-
110 middle-income countries (LMICs).(3) The recent increase in the incidence of emerging infectious diseases
111 and the recent COVID-19 pandemic have placed extreme stress on already fragile healthcare
112 systems.(4,5) Global structural inequalities further exacerbate COVID-19-related health disparities in the
113 LMICs, such as delayed access to vaccines and therapeutics.(6) During earlier phases of the pandemic,
114 nonpharmacological strategies such as quarantine, physical distancing, stay-at-home orders, and work-
115 from-home policies were the major tools used globally to mitigate the spread of the virus.(7)
116 Mathematical models of infectious diseases based on surveillance data can be used to address both
117 scientific hypotheses and infectious disease-control policy questions and are increasingly influential in
118 informing health policy and investment strategies globally.(8)

119
120 Direct data on social contacts collected for eight high-income European populations in the POLYMOD
121 study between May 2005 and September 2006 have informed many models with resulting policies.(9–20)
122 However, demographic structures and contact data differ between populations. The structure of human
123 contact networks varies by age, culture, and location with profound implications for pathogen dynamics
124 and transmission.(20) Social contact patterns may differ in different geographic areas due to a variety of
125 differences in social institutions, population density and family structure.(20–22) Despite these
126 differences, there is a lack of systematically collected data on social mixing in LMICs. Models
127 specifically parameterized to LMICs that incorporate localized social mixing patterns are essential to
128 inform public health policies.

129 While some studies have collected social mixing data in low-income settings,(11–20) they were
130 conducted in single populations using varying approaches. Therefore, it is difficult to infer whether

131 differences represent true variation between populations or are an artifact of study design. Furthermore,
132 data have traditionally been collected using paper-based contact diaries, which are subject to recall errors
133 and bias. Recently, automated methods of data collection utilizing portable technology such as mobile
134 phones and low-powered Bluetooth wearable sensors have become popular.(23) Thus, social mixing
135 study data are limited for parameterizing models of infectious diseases in LMICs. Information on changes
136 in contact patterns during times when physical distancing measures were implemented (for example, as
137 part of a public health response to the COVID-19 pandemic) can further inform mathematical models, by
138 noting the maximum reduction in social contacts that is realistically achievable in restrictive
139 environments.(24)

140
141 We propose a multicountry mixed methods (qualitative and quantitative) study with the overall goal of
142 using standardized methods (contact diaries and wearable proximity sensors) to collect social contact data
143 from urban and rural populations in LMICs. The contact diary will be used to collect self-reported data on
144 participants and their contact(s), as well as individual characteristics of the person they had contact with,
145 such as age, sex, and relationship to the contact person. The wearable proximity sensor is powered by a
146 battery and uses a low-power active-radio frequency identification (RFID) technology to transmit and
147 receive signals when hung around the neck or worn in a shirt/blouse pocket. This device will be used to
148 collect contact data among family members of infants younger than 6 months old. The study will include
149 individuals from all age groups. Furthermore, special focus will be given to studying the social
150 interactions of infants younger than six months of age. Data will be rigorously collected from four
151 different LMICs to capture a diverse cross-section of geographic locations: Guatemala (Central America),
152 India and Pakistan (South Asia) and Mozambique (Sub Saharan Africa). These sites were selected due to
153 the presence of a strong community-based infrastructure, existing disease and demographic surveillance
154 systems, and history of collaborating with the study team, and to ensure geographic diversity. The data
155 collected in this study will be made publicly available to be used in transmission dynamic models aiming

156 to accurately represent human interactions that lead to transmission of respiratory and gastrointestinal
 157 infections.

158 **Materials and Methods**

159 This is a multicountry cross-sectional survey. Our main study objective is to characterize social contact
 160 patterns in urban and rural settings in low-and-middle-income countries. Specifically, we aim to i)
 161 Explore the perceptions and acceptability of paper contact diaries and wearable proximity sensors to
 162 collect data on social contact and mixing patterns between individuals; ii) Characterize the patterns of
 163 social contact and mixing in urban and rural settings using standardized social contact diaries; and iii)
 164 Comprehensively profile the social contacts of infants with their household members using wearable
 165 proximity sensors. We will use qualitative (formative phase) and quantitative methods for our data
 166 collection and analysis. Data will be collected in four countries: Mozambique, Guatemala, India, and
 167 Pakistan (table 1, figure 1), one rural and one urban site, for a total of 8 study sites in each country.

168
 169 **Table 1.** Study site characteristics

Country	Mozambique		Guatemala		India		Pakistan	
Study site	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban
	Manhiça District, Maputo Province	Bairro Polana Caniço, Maputo City	Concepción Chiquirichapa: Telená, Tuilcanabaj (Centro/Bodega) San Juan Ostuncalco: Monrovia and La Esperanza in the Quezaltenango Department	San Juan Ostuncalco: Centro San Juan. Concepción Chiquirichapa: Centro Concepción	Kilarasampet, Kathalampet in Kaniyambadi Block, Vellore district, Tamil Nadu.	Ward 59, Vellore corporation	Matiari	Ali Akber Shah Goth in Karachi East
Source of participant data	Existing demographic surveillance system (DSS) in Manhiça and Polana Caniço		Randomized sampling from selected households in 57 clusters identified by aerial imaging		Existing DSS		Existing DSS for <5-year-old and 15–49-year-old women in Karachi;	

170
 171 **Figure 1.** Global map showing the location of Guatemala, India, Mozambique, and Pakistan.
 172 Each country has one rural and one urban site. There are separate rural (Manhica and Matiari) and urban
 173 (Maputo and Karachi East) sites in Mozambique and Pakistan, respectively. In Guatemala, rural and
 174 urban clusters will be identified from the Quetzaltenango and xxx. In India, rural and urban areas will be
 175 selected from Vellore.

176 **Data collection tools**

177 For the qualitative phase, data were collected through focus group discussions (FGDs) and psychometric
178 testing through cognitive interviews. For the quantitative phase, data will be collected using an enrollment
179 survey, paper and electronic contact diaries, an exit survey, and wearable proximity sensors.

180 **Enrollment survey:** The enrollment survey will be used to enroll participants after they provided
181 consent. It contains questions on demographic details of the participant including age, sex, school
182 enrollment, occupation, study site, as well as household characteristics such as household size and
183 composition, and sources of water.

184 **Contact diary.** We developed paper and electronic contact diaries to collect data on participants' social
185 mixing. We defined contact as:

- 186 • Nonphysical contact: This involves a two-way conversation with three or more words exchanged
187 in the physical presence of another person.
- 188 • Physical contact: This involves directly touching someone (skin-to-skin contact) or the clothes
189 they are wearing (for example a handshake, fist bump, elbow bump, foot bump, hug, kiss, etc.).
- 190 • Direct proximity: This involves being within 6 feet for 20 seconds or more.

191 The paper diary will be given to participants to self-report data on their contacts and characteristics of the
192 person with whom they had contact, such as age (approximate where appropriate), sex, relationship to the
193 contact person, location, frequency, and duration of contact. The electronic contact diary was designed in
194 REDCap and downloaded onto mobile tablets used by the field staff. For each site, we developed a set of
195 culturally relevant images representative of the 10 age groups that a contact may fall into. We classified
196 the age groups as follows: <6 months, 6-11 months, 1-4 years, 5-9 years, 10-14 years, 15-19 years, 20-29
197 years, 30-39 years, 40-59 years, and 60+ years. Participants will use the set of images to give estimates of
198 the age of their contacts. Each tool was translated to the major local languages for each study site
199 (Portuguese in Mozambique; Spanish in Guatemala; Tamil in India, and Urdu and Sindhi in Pakistan).
200 Each participant will be asked to randomly select the first day of contact diary keeping from a set of 7
201 cards, each symbolizing a day of the week. One day before the selected first contact diary day, the

202 participants will be trained in how to complete the diary. For adults or children unable to self-complete
203 the diary, they will select a third-party person (henceforth called a “shadow”) to assist them in recording
204 their contacts. Visit 3 will occur one day after the second diary day. The field staff will assess the paper
205 contact diary for completeness and inconsistencies and work with the participants to make any necessary
206 amendments. Once the verification process is complete, data in the paper diary will then be keyed into the
207 electronic diary by the field staff.

208 **Exit survey:** The exit survey will capture additional details such as sentiments regarding the accuracy of
209 the captured contacts, social and behavioral changes during the COVID-19 pandemic, and COVID-19
210 vaccination status.

211 **Wearable proximity sensors.** We purchased an open-source off-the-shelf proximity sensor that was
212 manufactured in the US, satisfies Federal Communications Commission (FCC) regulations, is relatively
213 cheap, customizable, and small enough to be carried by people of all ages (figure 2). We selected the
214 Adafruit Feather nRF52840 Sense sensor [<https://learn.adafruit.com/adafruit-feather-sense>] that uses
215 active unidirectional Bluetooth Low Energy operating at 2.4 GHz to periodically transmit and receive
216 data packets over short ranges relevant for respiratory infection spread.(25) It measures 51 mm x 23 mm x
217 7.2 mm and weighs approximately 6 grams. It has an on-board memory chip for data storage and is
218 powered by a 3.7 V 500 mAh. We redesigned and updated the firmware used in the SocioPatterns project
219 [www.sociopatterns.org] to be compatible with our new sensors. We also designed and 3D-printed
220 tamper-proof, weather-proof plastic cases to enclose the sensor and battery. A fully assembled sensor with
221 a battery and plastic case has a total cost of ~\$40 and weighs ~20-30 grams. Prior to deployment, we
222 conducted rigorous testing on case design, utility (ensuring radio communication between sensor dyads is
223 achieved), signal propagation (how far signals are propagated, effects of multipath on signal strength),
224 durability of battery charge and storage space capabilities (to establish optimum experiment duration).
225 When in proximity, sensor dyads exchange data packets consisting of the sensor identifier (ID),
226 timestamp, and transmission power level (received signal strength indicator, RSSI). Similar to previous
227 studies, a ‘contact’ was defined as occurring between two individuals if their sensors exchanged at least

228 one data packet over a 20-second interval. After a contact is established, its duration is defined as the
229 interval in which the devices continue to exchange at least one packet for every subsequent 20s interval.
230 Conversely, a contact is defined as broken if a 20-s interval lapses with no exchange of radio packets.(26)
231 When worn on the chest area, signal exchange occurs when individuals are face-to-face within 1-1.5
232 meters of each other, thus defining proximity that may be relevant for transmission of respiratory
233 infections. Assuming 8 hours of continuous data collection per day, the onboard sensor memory can log
234 data packets for up to 7 days. It is stressed that the sensors detect face-to-face proximity between
235 individuals by assessing the attenuation of the signals with distance but not an actual physical touch.
236 Previous sensor deployment suggested that a 20-second interaction can be considered as a reasonable
237 duration to describe “the fastest social interactions”, such as a quick conversation or handshake in a social
238 gathering that may lead to the transmission of respiratory infections through droplets from coughing or
239 sneezing.(26)
240 Once a battery has been inserted to activate the sensor, the sensor is sealed in a plastic case fastened with
241 screws and additionally inserted into a cover that is disposed of after use to prevent potential infection
242 transmission. We will request that the study participants hang the sensor at their chest by wearing it on a
243 lanyard around the neck or insert it into the front pocket of a shirt/blouse or attach it to clothing using a
244 clip. At the end of the study period, used sensors will be collected by study staff and the plastic covers
245 will be disinfected. Data from each sensor will be downloaded onto a password-protected laptop. After
246 data is downloaded, study staff will erase the data logs from the sensors and recharge them in readiness
247 for the next deployment.

248
249 **Figure 2.** Schematic representation of the sensors.
250 The sensor board consists of an antenna, flash storage device, and mini-USB port (A). A sensor will be
251 powered by a battery and placed in a plastic case with a sealable weather-proof cover that will be
252 disinfected in between each use (B). Sensors transmit signals only in the forward direction, thus only
253 detecting close face-to-face interactions when hung around the neck or worn in a shirt/blouse pocket (C).
254 The signals containing data packets are stored in the on-board memory of the sensor.

255 **Formative (Qualitative) Phase**

256 The goal of the formative phase is to understand contextual factors and attributes of the community
257 associated with the feasibility and acceptability of paper diaries and wearable proximity sensors in each
258 site. This understanding will inform the design of data collection tools and implementation of the survey
259 and other relevant study procedures.

260 **Selection criteria:** Residents of the communities who consented to participate in the study **Error!**

261 **Reference source not found.Error! Reference source not found.**were included in the qualitative study.

262 Residents who **Error! Reference source not found.Error! Reference source not found.Error!**

263 **Reference source not found.Error! Reference source not found.Error! Reference source not**

264 **found.Error! Reference source not found.Error! Reference source not found.Error! Reference**

265 **source not found.**did not consent to participate and non-permanent residents of the rural or urban

266 municipalities **Error! Reference source not found.**were not included.

267 **Training of field staff:** In-country study team members were trained in research ethics, data tools, and

268 data collection methods before the initiation of study activities. Due to travel and congregation

269 restrictions in place in the countries during the pandemic, trainings were conducted online via Zoom, an

270 online meeting platform. When restrictions were lifted, trainings were conducted in person. The training

271 focused on the objectives of the formative study, recruiting methods, data collection tools, techniques of

272 conducting focus group discussions, incorporating feedback on the FGD guide, data collection

273 procedures, data management, translation, and transcription. Quality control of qualitative data and

274 research ethics were also included in the training. After rapid analysis of the focus group discussions,

275 cognitive interview guides were finalized, and the field staff were trained to conduct cognitive interviews

276 (CI). The data were analysed on-site.

277 **Focus group discussions:** A focus group discussion guide designed with the assistance of the team

278 anthropologist was used to facilitate the discussions. The questions in the guide focused on identifying

279 community perceptions and assessing the acceptability and feasibility of our study procedures, and the

280 use of the contact diary and sensor. Participants also discussed potential barriers and suggested solutions

281 to overcome any challenges within their families or communities. Thirteen focus group discussions were
282 conducted per study site. The following groups were recruited to participate in the FGDs: 1) community
283 and religious leaders, 2) parents/caregivers of infants < 6 months old, 3) male and 4) female community
284 or household elders, and 5) teenagers.

285 *Sample size:* Between 1-2 FGD activities were conducted per site for each group, with 6-8 participants
286 per activity. To counter refusals and no-shows, we invited eight to twelve participants to each FGD.
287 However, we maintained a minimum of 6 participants per discussion group. The sessions were held at a
288 location convenient to participants in the community and conducted in a local language chosen by
289 participants, after obtaining their consent. During the FGDs, participants received an explanation and
290 demonstration of the use and operation of the proximity sensor and paper contact diaries through
291 illustrative images and simulations of use and function. Participants provided their feedback on the
292 design, language, ease of understanding, sociocultural barriers, and the study implementation process and
293 use of these tools. Each FGD lasted 60 minutes. All FGDs were recorded and translated to English for
294 analysis.

295 **Cognitive interviews:** During cognitive interviews, we clarified through a few carefully selected
296 questions on how community members understand and interpret survey questions and procedures,
297 specifically what these questions mean in the local context. The interviews helped to ensure that survey
298 items achieved their measurement purpose and elicited the intended meaning of the question. With the
299 assistance of the team anthropologist, we developed a cognitive interview guide that included select
300 survey and contact diary questions that were difficult to understand during FGDs. The cognitive
301 interviews identified comprehension and language problems, i.e., not knowing the meaning of
302 words/phrases, or having different local or cultural terms that are more appropriate or better understood;
303 inclusion/exclusion problems, i.e., determining whether certain concepts are to be considered within the
304 scope of an item, and whether caregivers include or exclude experiences within the intended scope of the
305 question; temporal problems, i.e., the period to which the question applies; logical problems, e.g., how
306 participants interpret certain phrases in survey questions, and computational problems, e.g., readability of

307 items, and guessing age of contacts. The interviewer starts by obtaining written consent from participants.
308 Participants were then asked to read each question, and before responding with an answer, they
309 paraphrased the question in their own words for the interviewer to ensure that they understood the
310 meaning correctly. While answering the question, they could also ‘think aloud’ to share what was going
311 through their mind when they are considering the response options. Participants were then asked to
312 complete a template of actionable recommendations based on the interviews. A separate note-taker helped
313 to document the interview. The cognitive interviews took place with 3 individuals per age group per site.
314 Six age groups were covered for at least 18 individuals in each study site.

315 **Recruitment procedures:** Appropriate strategies were used to recruit participants to each site. First, we
316 held meetings with local relevant stakeholders such as local authorities (chiefs of administrative posts and
317 localities, heads of the villages, chiefs of neighborhoods or any other authority), religious leaders, and
318 health practitioners to explain the purpose of conducting this study and identify potential participants.
319 Meetings were held according to the epidemiological situation and the COVID-19 alert system and
320 guidelines from the Ministry of Health. A strict protocol of facial masking, hand hygiene, social
321 distancing, and a well-ventilated room were followed. The dates and venues of meetings were defined
322 according to the availability of the participants and according to the epidemiological situation and
323 COVID-19 alert system of the Ministry of Health for the various sites. Where demographic surveillance
324 data were available, participants were selected at random based on 10 age groups identified for the
325 quantitative phase (i.e., <6 months, 6-11 months, 1-4 years, 5-9 years, 10-14 years, 15-19 years, 20-29
326 years, 30-39 years, 40-59 years, and 60+ years). In addition, we used the following methods of
327 recruitment for specific groups:

- 328 i. Community and religious leaders were identified with the support of the heads of localities and of the
329 administrative post, and religious leaders’ group with support of the religious affairs commission,
330 where large parts of the religions that were present in the study area are represented.
- 331 ii. Elders were recruited with the support of neighborhood chiefs and community leaders based on
332 selection criteria discussed with the study team.

- 333 iii. Parents of infants were identified by generating a list of all infants aged up to 6 months of age as of
334 data collection time, and identifying the names of their parents and household locations.
- 335 iv. Teachers were identified with the support of the principals or pedagogical heads of schools from the
336 study area, based on criteria to be discussed with the study team.

337 **Table 2.** Recruitment period for qualitative phase.

Study Sites	Recruitment Start Date	Recruitment End Date
Mozambique	04 August 2020	16 October 2020
Guatemala	19 July 2021	05 November 2021
India	16 February 2022	12 April 2022
Pakistan	14 Dec 2022	18 Feb 2023

338

339 **Qualitative data analysis:** Qualitative data were analyzed and suitable recommendations for updating the
340 survey questions were finalized with input from the principal investigators. Following each focus group
341 discussion meeting, the facilitator prepared a summary of impressions. The notes and audio recording
342 from the focus group discussion were transcribed and translated into English and study team members
343 reviewed the transcripts. Rapid analysis techniques (27,28) were applied to achieve results informing the
344 main study tools. Separate coding frameworks were developed for focus group discussions and cognitive
345 interview data. The thematic analysis method (29,30) was used to analyse the transcripts and identify and
346 organize themes into clusters and categories through coding. The analysis included reading the extended
347 transcript summaries several times to achieve an understanding of the meanings conveyed, identifying
348 important phrases, and articulating and confirming meanings through research team discussions to reach
349 agreement. Respondents' answers to each question were grouped, organized, and classified into
350 categories. Information obtained informed the development and refinement of the quantitative phase tools
351 and procedures for each respective site. Recommendations were made for modifications to the
352 quantitative phase study procedures and tools as appropriate.

353 **Qualitative data management:** At each study site, coinvestigators supervised the process of data
354 collection. All audios, interviews, and summaries were coded and saved in the local team's servers. All
355 data will be shared with Yale University and Emory University team members through a secure online

356 data sharing and storage platform. Only trained team members will be able to access the documents on
357 this platform. At all sites, all focus group discussions and cognitive interviews were audio-recorded, and
358 all texts or audio captured were transcribed in preferred local languages and then translated into English
359 verbatim. The languages used at each site were as follows: Portuguese in Mozambique; Spanish in
360 Guatemala; Tamil in India, and Urdu and Sindhi in Pakistan. Transcripts and field notes were anonymized
361 through the redaction of identifying details prior to translation. Digital recordings of FGDs and cognitive
362 interviews were destroyed after transcription; therefore, only the researchers conducting the FGDs, and
363 cognitive interviews and the transcriber had access to unredacted data. All quotations or descriptions used
364 in reports were anonymized, and additional potentially identifying details will be redacted or altered to
365 further protect anonymity. All photographs and video footage were only taken with consent and would be
366 appropriately anonymized. To ensure privacy, participants were requested to choose between the actual
367 name or nickname for the contact person or location. All signed consent forms were stored by the data
368 manager in a locked cabinet in the office.

369 **Quantitative Phase**

370 The goal of the quantitative phase is to collect and summarize data on the rates and characteristics of
371 social contacts using social contact diaries and electronic proximity sensors. Quantitative data collection
372 was scheduled in phases across the sites over two years.

373 **Table 3.** Recruitment period for quantitative phase

Study Site	Recruitment Start Date	Recruitment End Date
Mozambique	05 April 2021	05 April 2022
Guatemala	03 March 2022	22 December 2022
India	25 July 2022	04 February 2023
Pakistan	02 June 2023	Ongoing

374

375 **Contact diary study**

376 *Sample size:* Our study sample will be divided into 10 age groups (<6 months, 6-11 months, 1-4 years, 5-
377 9 years, 10-14 years, 15-19 years, 20-29 years, 30-39 years, 40-59 years, and 60+ years). We framed our
378 sample size calculations to detect sufficient precision in infants <6 months old, an age group of high

379 interest. We powered each age group to have a standard deviation consistent with the precision in the
380 estimate of the youngest age group (0-4 years) in the POLYMOD study (standard deviation = 7.65
381 contacts per day) (31). We chose to base our sample size calculation on the youngest age group because
382 infants less than six months of age represent the smallest fraction of the overall population size and whose
383 data are largely unavailable since they require a literate older person to keep records on their behalf (19).
384 We inflated the sample size by 10% for each age group to account for attrition due to study withdrawal
385 and loss to follow-up. This resulted in a sample size of 63 per age group per site, thus 630 for each site,
386 1260 per country, and 5040 participants overall.

387 **Electronic proximity sensor study**

388 *Sample size:* We will recruit 63 households per site (126 per country). At each site, households with a
389 child aged <6 months, 6-11 months, 1-4-years were selected randomly and proportional to these age
390 groups.

391 *Participant selection:* Once we locate a household, we will make an initial visit (visit 1) to seek
392 permission from the head of the household to allow us to engage the other household members. In
393 Guatemala only, where we do not have sampling frame, we identified household clusters around a 200 m²
394 radius using satellite imagery from Bing. From a central location in each cluster, we approached all
395 households in sequence until we attained our target sample size. We will obtain written informed consent
396 (adults ≥ 18 years old) or assent (for children aged <18 years).

397 *Study procedures:* To train all the participants in diary-keeping and sensor-wearing procedures, we will
398 request the head of the household to select a day of the week when all household members will be present
399 at the same time. On this selected day (start day), we will visit the household and train the members on
400 the diary and sensor study procedures (figures 3 and 4). Each participant will wear the sensor over 7
401 continuous days for data collection. Participants will be required to doff the sensor when not in use such
402 as when sleeping or leaving the house, or when encountering water. In addition, the head of the household
403 will select one day at random (diary day one) between the start day and end day (7 days after start day).
404 Each household member will be required to keep a record of their contacts in the paper diary on two

405 consecutive days. The research team will record the date and time when the sensors were issued and
406 collected, while participants will keep a daily log of when the sensors were worn or taken off. Data from
407 each sensor will be individually downloaded into a laptop at the site office.

408
409 **Figure 3.** Flow of field visits showing data collection procedures for the three visits (contact diary study)

410
411 **Figure 4.** Flow of field visit showing data collection procedures for the three visits for the (proximity
412 sensors study)

413

414 **Quantitative data analysis**

415 *Construction of age-specific contact matrices.* The primary outcome from this analysis will be age-
416 specific contact matrices containing the average (or expected) number of contacts made per day between
417 age groups combinations. We will create three age-specific contact matrices for each study site: a matrix
418 of physical contacts only, a matrix of nonphysical contacts, and a matrix of all contacts (both physical and
419 nonphysical contacts).

420 *Identifying predictors of number and duration of contacts.* We will investigate whether certain
421 demographic and biologic characteristics are predictors of the frequency and nature of contacts. This
422 regression analysis will result in the identification of key drivers of contact.

423 *Defining contacts using wearable sensors.* To comprehensively profile the frequency, duration and
424 intensity of social contacts of young infants with their household members, we will analyze high
425 resolution data collected using wearable proximity sensing devices. Data collected from each participant
426 will be obtained and cleaned separately. In sensitivity analyses, different time-interval definitions of a
427 ‘contact’ will be considered.

428 *Validation of diary-based collection (Sensor Study).* We will use the sensor data to validate the diary data
429 using two methods. First, using variables that characterize each contact such as age, sex, duration, and
430 relationship, we will identify contacts that were identified by [1] diary, [2] sensor, or [3] both. We will
431 then calculate the level of agreement between diary and sensor data using Cohen’s kappa coefficient (k).
432 Second, using sensor data, we will construct age-specific contact matrices containing the average number

433 of contacts made per day using the methods described in the contact diary study. We will compare these
434 matrices with those generated from diary data but restricted to household contacts using the coefficient of
435 accuracy ($C(a)$) and the coefficient of determination (r_c). $C(a)$ measures the accuracy and r_c measures the
436 precision of the relationship between the two data collection methods (32). We will then perform the
437 same comparisons for data on the duration of contacts.

438 *Characterizing the social contact of young infants.* We will generate a detailed characterization of the
439 social contacts of infants aged <6 months with their household members. To do this, we will calculate the
440 infant proximity score for each household member, which is defined as the fraction of time spent by each
441 household member in proximity with the infant, over the total infant contact time.(33) Certain
442 demographic and behavior characteristics (e.g., maternal age, sex, household size, and country) are
443 predictors of the infant proximity score. To evaluate the association between social contacts and these
444 predictors, we will use regression methods such as negative binomial regression where the dependent
445 variable is the infant proximity score.

446 *Estimating effects on total number and duration of contacts.* We hypothesize that certain demographic
447 and behavior characteristics (e.g., maternal age, sex, household size, and country) are predictors of the
448 frequency and nature of social contacts of young infants. Identifying these factors will aid in generalizing
449 our findings and serve as a guide for other researchers to select the right contact matrices and determine
450 what predictors are important to include in future analyses. To evaluate the association between predictors
451 and the total number or duration of contacts, we will use similar regression methods as above. In addition
452 to providing an analytical understanding of the determinants of contact frequency and types, the results
453 from this analysis will help to guide researchers on which contact matrix is best suited to a given
454 modeling application or generalizable to similar populations.

455 *Creation and analysis of an aggregated contact network.* We will create an undirected social contact
456 network at the household level, aggregated over the whole study period per site, draw network graphs,
457 and calculate descriptive network statistics. Nodes represent study participants and edges are the presence
458 of at least one recorded contact event between individuals during a contact window (figure 5.) Edge

459 attributes will reflect the type or intensity of the contact (i.e., cumulative duration of contact). The
460 attributes of the nodes will be included using the demographic data and relationship to participant as
461 collected during the initial interview of study participants. We will calculate [1] the degrees of each node
462 (k) and describe the distribution by degrees $P(k)$; [2] the distance (d) between nodes (e.g., the shortest
463 distance between individuals in households) and the average distance (\bar{d}) ; and [3] the intermediation
464 centrality defined as the proportion of the shortest paths that pass through a single node B_i . These
465 statistics can subsequently be used to explicitly model disease transmission in networks.[27] We will also
466 use an adjacency matrix (A_{ij}) to summarize connections made with our populations. Using the adjacency
467 matrix, we will calculate the average number of contacts according to the following formula: $\bar{n} = \frac{1}{N} \sum_{ij} A_{ij}$,
468 where N is the population size (i.e., size of the households).(35)

469 **Figure 5.** Household contact network.

470 Contact networks will be visualized in order to depict the presence, type and characteristics of contacts
471 (edges). Here, as an example, household members are represented as nodes color coded for relationship to
472 the infant. The thickness of the edges represents cumulative time in contact, with thicker/darker edges
473 representing more time spent in contact.

474

475 **Quantitative data management**

476 Study data will be collected and managed using REDCap electronic data capture tools hosted at Emory
477 University(36,37). REDCap (Research Electronic Data Capture) is a secure, web-based software platform
478 designed to support data capture for research studies, providing 1) an intuitive interface for validated data
479 capture; 2) audit trails for tracking data manipulation and export procedures; 3) automated export
480 procedures for seamless data downloads to common statistical packages; and 4) procedures for data
481 integration and interoperability with external sources. A different project will be created for each country,
482 such that country study teams will only have access to country-specific data.

483 *Security of the sensors.* Signals in the power range emitted by the sensors do not pose reported radiation
484 harm while in use by humans. Additionally, the radio system for these sensors is approved for use by the
485 United States Federal Communications Commission (US FCC). The sensor is FCC certified and licensed
486 under 47 CFR part 2, which means that it does not cause harmful interference to radio signals from

487 household appliances. Additionally, the beacon is certified to comply with rules limiting other policy
488 objectives - such as human exposure limits to radio frequencies and hearing aid compatibility (HAC) with
489 mobile wireless devices. The sensor and battery will be stored in a small weather resistant box that is
490 sealed with two small screws. This box will then be inserted into an additional outer protector that will be
491 discarded after each use. A strap will be attached to the outer protector, so that the participant can use it
492 hanging around the neck, and it will be adjusted so that it is kept at chest level. For use in young children,
493 an appropriate hook will be provided so that it can be attached to the participant's clothing. Parents will be
494 shown how to attach the sensor onto children's outer garments.

495 *Data security specifications.* The data exchanged between the sensors is automatically encrypted during
496 its internal storage. We have developed programs using Python software that data managers will run to
497 automatically download data from sensors. The sensors must be connected to the computer using data
498 cables before the data can be downloaded. Only data administrators will have access to these programs.
499 The decryption of the data will be done by the data analyst.

500 **Ethics approval and consent to participate**

501 The study protocol and all qualitative and quantitative survey tools were submitted to the Institutional
502 Review Boards (IRBs) of all the study sites as well as the national ethics committees of all the
503 international sites before initiation of study activities. This study addresses household and individual
504 information and all data collection shall follow institutional, national, and international policies and
505 procedures or guidelines for protecting the information provided by participants and maintaining the
506 confidentiality of participants and their contacts. To comply with this, the field staff were trained in the
507 need to ensure that data collection is done accordingly and ensure maximum confidentiality possible
508 under field conditions. Informed consent to participate was taken from parents/legal guardians of minor
509 participants. For all the sites, we developed the following consent and assent forms for both qualitative
510 and quantitative data collection activities: Consent for Adult Participants (applicable to individuals aged
511 18 years or older to give consent for themselves to participate in the study); Consent form for Parents and

512 Guardians of Individuals under 18 years old (applicable to parents and legal guardians of children under
513 the age of 18 years to give consent for them to participate in the study); Assent form for Adolescents aged
514 12 to 17 years (applicable to adolescents aged 12 to 17 years old to give assent for themselves to
515 participate in the study); and Consent for Household Heads (applicable to heads of households to give
516 consent for themselves and all members of their household to participate in the study). In Guatemala, we
517 obtained either written or verbal assent for children older than 7 years for focal group discussion and
518 cognitive interviews, upon request of the community. This was approved by the UVG IRB. Ethical
519 approval of the study was obtained from the Institutional Review Board for all the sites as well as at
520 Emory University before the onset of study activities. Ethical approval dates and numbers for all study
521 sites:

Site	Emory Approval date	Approval number	Site approval date
Parent (Emory University) protocol	March 4, 2020	00105630	3/04/2020
Yale University Protocol	Reliance Agreement: November 11, 2019	2000026911	6/12/2019 (Yale IRB)
Mozambique	August 9, 2020	CIBS-CISM/012/2021	3/26/2020 (CISM)
Guatemala	July 21, 2021	223-12-2020	1/25/2021 (National IRB) 3/03/2021 (UVG IRB)
India	June 26, 2019	IRB Min. No. 12065	2/02/2022 (CMC IRB)
Pakistan	December 22, 2022	2022-7912-23648	12/22/2022 (AKU ERC)

522

523 **Status and timeline of the study**

524 Qualitative data collection has been completed in three sites and is ongoing in the fourth site. IRB
525 approval has been obtained in all the sites. Quantitative data collection is expected to be complete by
526 March 2024.

527 **Discussion**

528 This multicountry cross-sectional study using qualitative (formative phase) and quantitative data from
529 Mozambique, Guatemala, India, and Pakistan is the first study to comprehensively characterize the
530 patterns of social contact and mixing in urban and rural LMIC settings using standardized social contact

531 diaries. We will also profile the social contacts of infants with their household members using contact
532 diaries and wearable proximity sensors. The deidentified data collected in this study will be made publicly
533 available for use in dynamic transmission models of infectious diseases. These empirical data will allow a
534 more accurate representation of human interactions that lead to the transmission of pathogens transmitted
535 through close contact.

536 The qualitative, formative phase of this study is unique and will generate useful information regarding
537 perceptions and acceptability of contact diaries and wearable proximity sensors to collect data on social
538 contact and mixing patterns between individuals in LMICs. This information will be used for developing
539 more complex studies that use contact diaries and wearable proximity sensors. Finally, conducting this
540 study in LMICs will generate results that may be more representative of contact patterns in such setting.
541 Settings. Currently, many modelers use European data from POLYMOD or POLYMOD data projected
542 onto LMIC populations.

543 Despite the use of rigorous mixed qualitative and quantitative methods, we anticipate some challenges.
544 Participants may only recall contacts that are recent or salient (those that are most important, such as
545 meeting family or friends compared to strangers). This may lead to an underestimation of contact
546 rates. We expect that children who are not in school and those aged less than 10 years old may not be able
547 to self-report their contacts. Additionally, adults who are unable to or have limited ability to read and
548 write may not be able to record their contacts in the paper contact diaries. For these groups, we will assign
549 parents/guardians as “shadows” to the children to help with recording contacts. Then adult participants
550 will be asked to select an individual to record their contacts on their behalf. This would require adequate
551 training of both participants and shadows to understand diary keeping procedures while respecting the
552 privacy of the participants. Furthermore, to profile the social contacts of infants with their household
553 members using contact diaries and wearable proximity sensors, only interactions between participants
554 wearing sensors will be recorded. Sensors do not record touch or other characteristics of interactions
555 between participants. Therefore, while the sensors provide accurate and time-resolved data on physical
556 proximity, they lack the richness of the characteristics of the contact. The site-specific contact matrices

557 and all underlying data will be made publicly available. Study findings will be published in scientific
558 journals and presented at national and international scientific and programmatic forums.

559
560
561
562
563
564
565
566
567
568
569
570

571 **Authors' contributions**

Contributor Role	Role Definition	Author
Conceptualization	Ideas; formulation or evolution of overarching research goals and aims.	Dr. Obianuju G. Aguolu, Dr. Moses C. Kiti, Dr. Saad B. Omer, Dr. Benjamin Lopman, Dr. Fauzia Malik
Data Curation	Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later reuse.	Dr. Moses C. Kiti, Holin Chen, Sara Kim, Carol Liu, Dr. Obianuju G. Aguolu, Dr. Saad B. Omer, Dr. Benjamin Lopman, Dr. Fauzia Malik
Formal Analysis	Application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data.	Dr. Moses C. Kiti, Holin Chen, Sara Kim, Carol Liu, Dr. Obianuju G. Aguolu, Noureen Ahmed, Dr. Saad B. Omer, Dr. Benjamin Lopman, Dr. Fauzia Malik
Funding Acquisition	Acquisition of the financial support for the project leading to this publication.	Dr. Saad B. Omer, Dr. Benjamin Lopman
Investigation	Conducting a research and investigation process, specifically performing	Dr. Moses C. Kiti, Alana Zelaya, Dr. Obianuju G. Aguolu, Noureen Ahmed, Dr. Saad B. Omer, Dr. Benjamin Lopman, Edgar Jammisse, Americo Jose, Charfudin Sacoor, Ivalda Macicame, Felizarda

	the experiments, or data/evidence collection.	Amosse, Nilzio Cavele, Corssino Tchavana, Migdalia Uamba, Dr. Inci Yildirim, Venkata Raghava Mohan, Prasanna Samuel, Dr. Fauzia Malik, Dr. Azucena Bardaji, Dr. Herberth Maldonado, Dr. Maria Ajsivinac, MPH, Claudia Jarquin
Methodology	Development or design of methodology; creation of models.	Dr. Moses C. Kiti, Alana Zelaya, Dr. Obianuju G. Aguolu, Dr. Saad B. Omer, Dr. Lauren Pischel, Noureen Ahmed, Dr. Inci Yildirim, Venkata Raghava Mohan, Prasanna Samuel, Carol Liu, Dr. Maria Sundaram, Dr. Benjamin Lopman, Dr. Fauzia Malik
Project Administration	Management and coordination responsibility for the research activity planning and execution.	Dr. Moses C. Kiti, Noureen Ahmed, Alana Zelaya, Dr. Obianuju G. Aguolu, Dr. Saad B. Omer, Dr. Inci Yildirim, Venkata Raghava Mohan, Rajan Srinivasan, Dr. Maria Sundaram, Charfudin Sacoor, Dr Ivalda Macicame, Dr. Benjamin Lopman, Dr. Fauzia Malik, Dr. Herberth Maldonado
Resources	Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools.	Dr. Moses C. Kiti, Dr. Obianuju G. Aguolu, Dr. Saad B. Omer, Noureen Ahmed, Dr. Benjamin Lopman, Dr Fauzia Malik
Software	Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components.	Dr. Moses C. Kiti, Sergio Gramacho, Corssino Tchavana, Nilzio Cavele, Kye Ellington, Dr. Obianuju G. Aguolu, Dr. Saad B. Omer, Dr. Benjamin Lopman
Supervision	Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team.	Dr. Moses C. Kiti, Dr. Obianuju G. Aguolu, Dr. Inci Yildirim, Venkata Raghava Mohan, Rajan Srinivasan, Gifta John, Dr. Saad B. Omer, Dr. Benjamin Lopman, Dr Fauzia Malik, Dr. Azucena Bardaji, Dr. Herberth Maldonado
Validation	Verification, whether as a part of the activity or separate, of the overall replication /reproducibility of results/experiments and other research outputs.	Dr. Moses C. Kiti, Dr. Obianuju G. Aguolu, Dr. Saad B. Omer, Dr. Benjamin Lopman, Dr. Fauzia Malik, Noureen Ahmed
Visualization	Preparation, creation and/or presentation of the published work, specifically	Dr. Moses C. Kiti, Dr. Obianuju G. Aguolu, Noureen Ahmed, Dr. Inci Yildirim, Dr. Saad B. Omer, Dr. Benjamin Lopman, Dr. Fauzia Malik

	visualization/data presentation.	
Writing – Original Draft Preparation	Creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation).	Dr. Obianuju G. Aguolu, Dr. Saad B. Omer, Dr. Benjamin Lopman
Writing – Review & Editing	Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision – including pre- or post-publication stages.	Dr. Moses C. Kiti, Dr. Obianuju G. Aguolu, Noreen Ahmed, Carol Liu, Dr. Saad B. Omer, Dr. Lauren Pischel, Dr. Inci Yildirim, Venkata Raghava Mohan, Rajan Srinivasan, Dr. Maria Sundaram, Dr. Benjamin Lopman, Charfudin Saco, Dr. Ivalda Macicame, Dr. Fauzia Malik, Dr. Azucena Bardají, Dr. Herberth Maldonado, Dr. María Ajsivinac, MPH Claudia Jarquín

572
573

574

575 **Acknowledgement**

576 We thank the many contributors to this study for their expertise and assistance throughout all aspects of
577 our study and for their help in writing the manuscript.

578 **Supporting information**

- 579 1. Survey:
580 a. Enrollment survey
581 b. Electronic contact diary
582 c. Exit survey
583 2. Paper contact diary
584 3. Places diary
585 4. Focus group discussion guide
586 5. Cognitive interview guide

587

588 **Figures**

- 589 1. Figure 1. Global map showing the location of Guatemala, India, Mozambique, and Pakistan.
590 2. Figure 2. Schematic representation of the sensors.
591 3. Figure 3. Flow of field visits showing data collection procedures for the three visits (contact diary
592 study)
593 4. Figure 4. Flow of field visit showing data collection procedures for the three visits for the (proximity
594 sensor device study)
595 5. Figure 5. Household contact network.

596

597 **List of abbreviations**

598 LMICs: low-and-middle-income countries

599 FGDs: focus group discussions

600

601

602

603

604

605

606

References

- 607 1. Liang J, Du Y, Qu X, Ke C, Yi G, Liu M, et al. The Causes of Death and Their Influence in Life
608 Expectancy of Children Aged 5–14 Years in Low- and Middle-Income Countries From 1990 to 2019.
609 *Frontiers in Pediatrics* [Internet]. 2022 [cited 2022 Nov 9];10. Available from:
610 <https://www.frontiersin.org/articles/10.3389/fped.2022.829201>
- 611 2. Jamison DT, Summers LH, Alleyne G, Arrow KJ, Berkley S, Binagwaho A, et al. Global health
612 2035: a world converging within a generation. *The Lancet* [Internet]. 2013 Dec 7 [cited 2022 Nov
613 9];382(9908):1898–955. Available from: [https://www.thelancet.com/journals/lancet/article/PIIS0140-
614 6736%2813%2962105-4/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736%2813%2962105-4/fulltext)
- 615 3. Department of Data and Analytics (DNA), Division of Data, Analytics and Delivery for Impact
616 (DDI), Department of Data and Analytics (DNA), Division of Data, Analytics and Delivery for
617 Impact (DDI), Department of Data Analytics (DNA) Division of Data, Analytics and Delivery for
618 Impact (DDI) WHO, Geneva. WHO methods and data sources for country-level causes of death
619 2000-2019 [Internet]. 2020 [cited 2022 Nov 9]. Available from: [https://www.who.int/docs/default-
620 source/gho-documents/global-health-estimates/ghe2019_cod_methods.pdf](https://www.who.int/docs/default-source/gho-documents/global-health-estimates/ghe2019_cod_methods.pdf)
- 621 4. Grimm CA. Hospitals Reported That the COVID-19 Pandemic Has Significantly Strained Health
622 Care Delivery [Internet]. USA: U.S. Department of Health and Human Services Office of Inspector
623 General; 2021 Mar [cited 2022 Nov 19] p. 62. Report No.: OEI-09-21-00140. Available from:
624 <https://oig.hhs.gov/oei/reports/OEI-09-21-00140.pdf>
- 625 5. Office of the Assistant Secretary for Planning, and Evaluation, U.S. Department of Health and
626 Human Service. Impact of the COVID-19 Pandemic on the Hospital and Outpatient Clinician
627 Workforce: Challenges and policy responses [Internet]. Washington, D.C.: U.S. Department of
628 Health and Human Services Office of Inspector General; 2022 May [cited 2022 Nov 11]. Report No.:
629 Issue Brief No. HP-2022-13. Available from:
630 [https://aspe.hhs.gov/sites/default/files/documents/9cc72124abd9ea25d58a22c7692dccb6/aspe-covid-
workforce-report.pdf](https://aspe.hhs.gov/sites/default/files/documents/9cc72124abd9ea25d58a22c7692dccb6/aspe-covid-
631 workforce-report.pdf)
- 632 6. Ho A, Dascalu I. Global Disparity and Solidarity in a Pandemic. *Hastings Center Report* [Internet].
633 2020 [cited 2022 Nov 9];50(3):65–7. Available from:
634 <https://onlinelibrary.wiley.com/doi/abs/10.1002/hast.1138>
- 635 7. Huang C, Liu X, Sun S, Li SC, Deng M, He G, et al. Insights into the transmission of respiratory
636 infectious diseases through empirical human contact networks. *Sci Rep* [Internet]. 2016 Aug 16 [cited
637 2022 Nov 9];6:31484. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4985757/>
- 638 8. Grassly NC, Fraser C. Mathematical models of infectious disease transmission. *Nature Reviews*
639 *Microbiology* [Internet]. 2008 Jun 1;6(6):477–87. Available from:
640 <https://doi.org/10.1038/nrmicro1845>
- 641 9. Campbell PT, McVernon J, Shrestha N, Nathan PM, Geard N. Who’s holding the baby? A
642 prospective diary study of the contact patterns of mothers with an infant. *BMC Infectious Diseases*
643 [Internet]. 2017 Sep 20 [cited 2022 Nov 16];17(1):634. Available from:
644 <https://doi.org/10.1186/s12879-017-2735-8>
- 645 10. Knight GM, Dharan NJ, Fox GJ, Stennis N, Zwerling A, Khurana R, et al. Bridging the gap between
646 evidence and policy for infectious diseases: How models can aid public health decision-making.

- 647 International Journal of Infectious Diseases [Internet]. 2016 [cited 2022 Nov 16];42:17–23. Available
648 from: <https://www.mendeley.com/catalogue/378149a0-66ce-3874-a459-0d4114f72c95/>
- 649 11. Leung K, Jit M, Lau EHY, Wu JT. Social contact patterns relevant to the spread of respiratory
650 infectious diseases in Hong Kong. *Sci Rep* [Internet]. 2017 Aug 11 [cited 2022 Nov 16];7:7974.
651 Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5554254/>
- 652 12. Johnstone-Robertson SP, Mark D, Morrow C, Middelkoop K, Chiswell M, Aquino LDH, et al. Social
653 mixing patterns within a South African township community: implications for respiratory disease
654 transmission and control. *Am J Epidemiol*. 2011 Dec 1;174(11):1246–55.
- 655 13. Fu Y chih, Wang DW, Chuang JH. Representative Contact Diaries for Modeling the Spread of
656 Infectious Diseases in Taiwan. *PLOS ONE* [Internet]. 2012 Oct 3 [cited 2022 Nov 16];7(10):e45113.
657 Available from: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0045113>
- 658 14. Ibuka Y, Ohkusa Y, Sugawara T, Chapman GB, Yamin D, Atkins KE, et al. Social contacts,
659 vaccination decisions and influenza in Japan. *J Epidemiol Community Health* [Internet]. 2016 Feb
660 [cited 2022 Nov 16];70(2):162–7. Available from:
661 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4752620/>
- 662 15. Stein ML, Steenbergen JE van, Chanyasanha C, Tipayamongkholgul M, Buskens V, Heijden PGM
663 van der, et al. Online Respondent-Driven Sampling for Studying Contact Patterns Relevant for the
664 Spread of Close-Contact Pathogens: A Pilot Study in Thailand. *PLOS ONE* [Internet]. 2014 Jan 8
665 [cited 2022 Nov 16];9(1):e85256. Available from:
666 <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0085256>
- 667 16. Dodd PJ, Looker C, Plumb ID, Bond V, Schaap A, Shanaube K, et al. Age- and Sex-Specific Social
668 Contact Patterns and Incidence of Mycobacterium tuberculosis Infection. *Am J Epidemiol*. 2016 Jan
669 15;183(2):156–66.
- 670 17. Ajelli M, Litvinova M. Estimating contact patterns relevant to the spread of infectious diseases in
671 Russia. *J Theor Biol*. 2017 Apr 21;419:1–7.
- 672 18. Blackwood JC, Cummings D a. T, Iamsirithaworn S, Rohani P. Using age-stratified incidence data to
673 examine the transmission consequences of pertussis vaccination. *Epidemics*. 2016 Sep;16:1–7.
- 674 19. Kiti MC, Kinyanjui TM, Koech DC, Munywoki PK, Medley GF, Nokes DJ. Quantifying Age-
675 Related Rates of Social Contact Using Diaries in a Rural Coastal Population of Kenya. *PLOS ONE*
676 [Internet]. 2014 Aug 15 [cited 2022 Nov 16];9(8):e104786. Available from:
677 <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0104786>
- 678 20. Grijalva CG, Goeyvaerts N, Verastegui H, Edwards KM, Gil AI, Lanata CF, et al. A Household-
679 Based Study of Contact Networks Relevant for the Spread of Infectious Diseases in the Highlands of
680 Peru. *PLOS ONE* [Internet]. 2015 Mar 3 [cited 2022 Nov 16];10(3):e0118457. Available from:
681 <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0118457>
- 682 21. Prem K, Cook AR, Jit M. Projecting social contact matrices in 152 countries using contact surveys
683 and demographic data. *PLOS Computational Biology* [Internet]. 2017 Sep 12 [cited 2022 Nov
684 18];13(9):e1005697. Available from:
685 <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005697>

- 686 22. Thindwa D, Jambo KC, Ojal J, MacPherson P, Dennis Phiri M, Pinsent A, et al. Social mixing
687 patterns relevant to infectious diseases spread by close contact in urban Blantyre, Malawi. *Epidemics*
688 [Internet]. 2022 Sep [cited 2022 Nov 18];40:100590. Available from:
689 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9176177/>
- 690 23. Vijayan V, Connolly JP, Condell J, McKelvey N, Gardiner P. Review of Wearable Devices and Data
691 Collection Considerations for Connected Health. *Sensors (Basel)* [Internet]. 2021 Aug 19 [cited 2023
692 Jul 5];21(16):5589. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8402237/>
- 693 24. Liu CY, Berlin J, Kiti MC, Del Fava E, Grow A, Zagheni E, et al. Rapid Review of Social Contact
694 Patterns During the COVID-19 Pandemic. *Epidemiology*. 2021 Nov 1;32(6):781–91.
- 695 25. Vanhems P, Barrat A, Cattuto C, Pinton JF, Khanafer N, Régis C, et al. Estimating Potential Infection
696 Transmission Routes in Hospital Wards Using Wearable Proximity Sensors. *PLOS ONE* [Internet].
697 2013 Sep 11 [cited 2023 Aug 9];8(9):e73970. Available from:
698 <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0073970>
- 699 26. Cattuto C, van den Broeck W, Barrat A, Colizza V, Pinton JF, Vespignani A. Dynamics of person-to-
700 person interactions from distributed RFID sensor networks. *PLoS ONE*. 2010;5(7):1–9.
- 701 27. Nevedal AL, Reardon CM, Opra Widerquist MA, Jackson GL, Cutrona SL, White BS, et al. Rapid
702 versus traditional qualitative analysis using the Consolidated Framework for Implementation
703 Research (CFIR). *Implementation Science* [Internet]. 2021 Jul 2 [cited 2023 Aug 9];16(1):67.
704 Available from: <https://doi.org/10.1186/s13012-021-01111-5>
- 705 28. Palinkas LA, Mendon SJ, Hamilton AB. Innovations in Mixed Methods Evaluations. *Annu Rev*
706 *Public Health* [Internet]. 2019 Apr 1 [cited 2023 Aug 9];40:423–42. Available from:
707 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6501787/>
- 708 29. Braun V, Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology*
709 [Internet]. 2008 Jul 21 [cited 2023 Aug 9]; Available from:
710 <https://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa>
- 711 30. Campbell K, Orr E, Durepos P, Nguyen L, Li L, Whitmore C, et al. Reflexive Thematic Analysis for
712 Applied Qualitative Health Research. *The Qualitative Report* [Internet]. 2021 Jun 25;26(6):2011–28.
713 Available from: <https://nsuworks.nova.edu/tqr/vol26/iss6/24>
- 714 31. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social contacts and mixing
715 patterns relevant to the spread of infectious diseases. *PLoS Medicine*. 2008;5(3):0381–91.
- 716 32. Lin L, Torbeck LD. Coefficient of Accuracy and Concordance Correlation Coefficient: New
717 Statistics for Methods Comparison. *PDA Journal of Pharmaceutical Science and Technology*
718 [Internet]. 1998 Mar 1 [cited 2023 Jul 17];52(2):55–9. Available from:
719 <https://journal.pda.org/content/52/2/55>
- 720 33. Ozella L, Gesualdo F, Tizzoni M, Rizzo C, Pandolfi E, Campagna I, et al. Close encounters between
721 infants and household members measured through wearable proximity sensors. *PLoS One*.
722 2018;13(6):e0198733.

- 723 34. Danon L, Ford AP, House T, Jewell CP, Keeling MJ, Roberts GO, et al. Networks and the
724 Epidemiology of Infectious Disease. *Interdiscip Perspect Infect Dis* [Internet]. 2011 [cited 2023 Jul
725 17];2011:284909. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3062985/>
- 726 35. Gregson S, Mushati P, White PJ, Mlilo M, Mundandi C, Nyamukapa C. Informal confidential voting
727 interview methods and temporal changes in reported sexual risk behaviour for HIV transmission in
728 sub-Saharan Africa. *Sexually Transmitted Infections* [Internet]. 2004 Dec 1 [cited 2023 Jul
729 17];80(suppl 2):ii36–42. Available from: https://sti.bmj.com/content/80/suppl_2/ii36
- 730 36. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research Electronic Data Capture
731 (REDCap) - A metadata-driven methodology and workflow process for providing translational
732 research informatics support. *J Biomed Inform* [Internet]. 2009 Apr [cited 2023 Aug 3];42(2):377–
733 81. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2700030/>
- 734 37. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O’Neal L, et al. The REDCap Consortium:
735 Building an International Community of Software Platform Partners. *J Biomed Inform* [Internet].
736 2019 Jul [cited 2023 Aug 3];95:103208. Available from:
737 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7254481/>
- 738
- 739



Figure 1

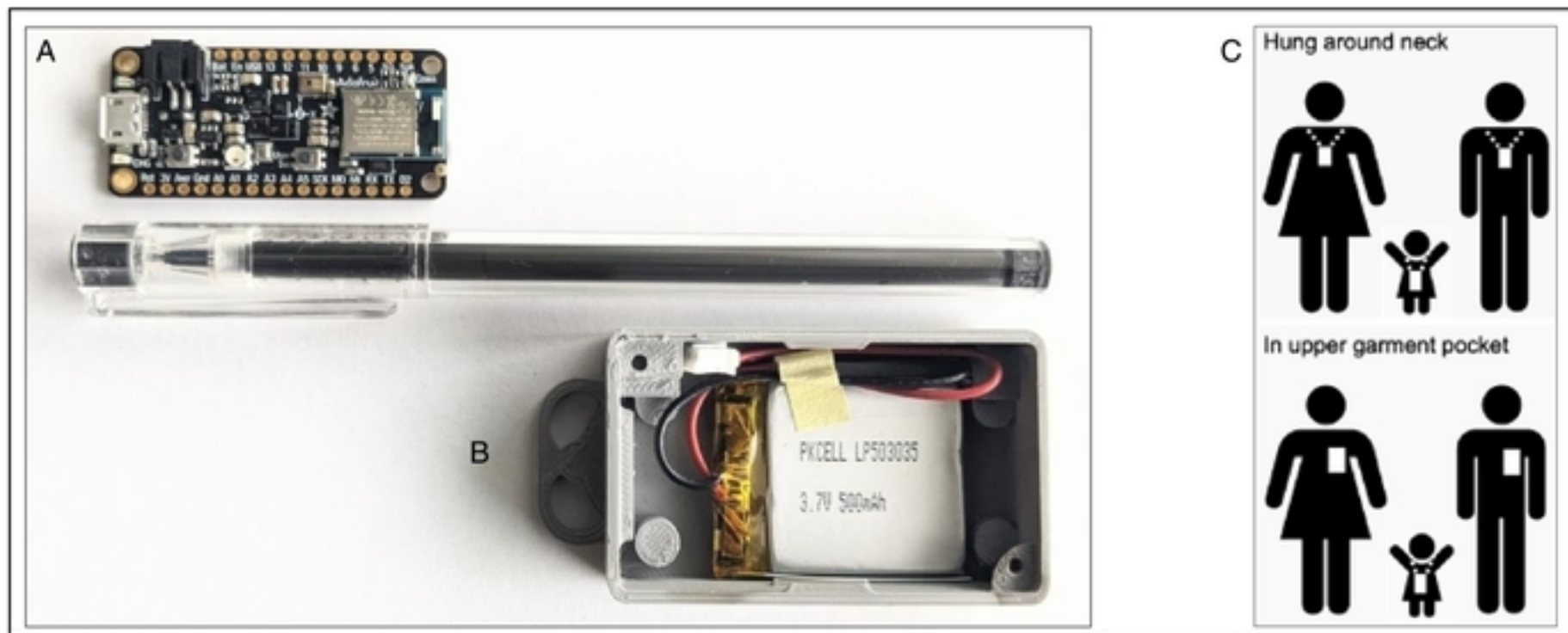


Figure 2

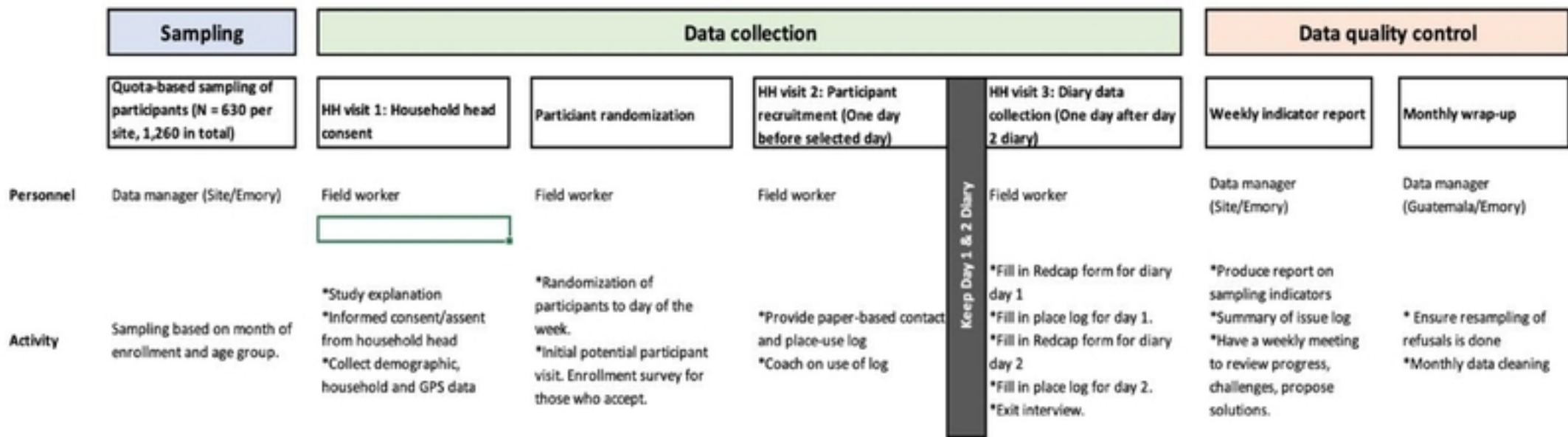


Figure 3

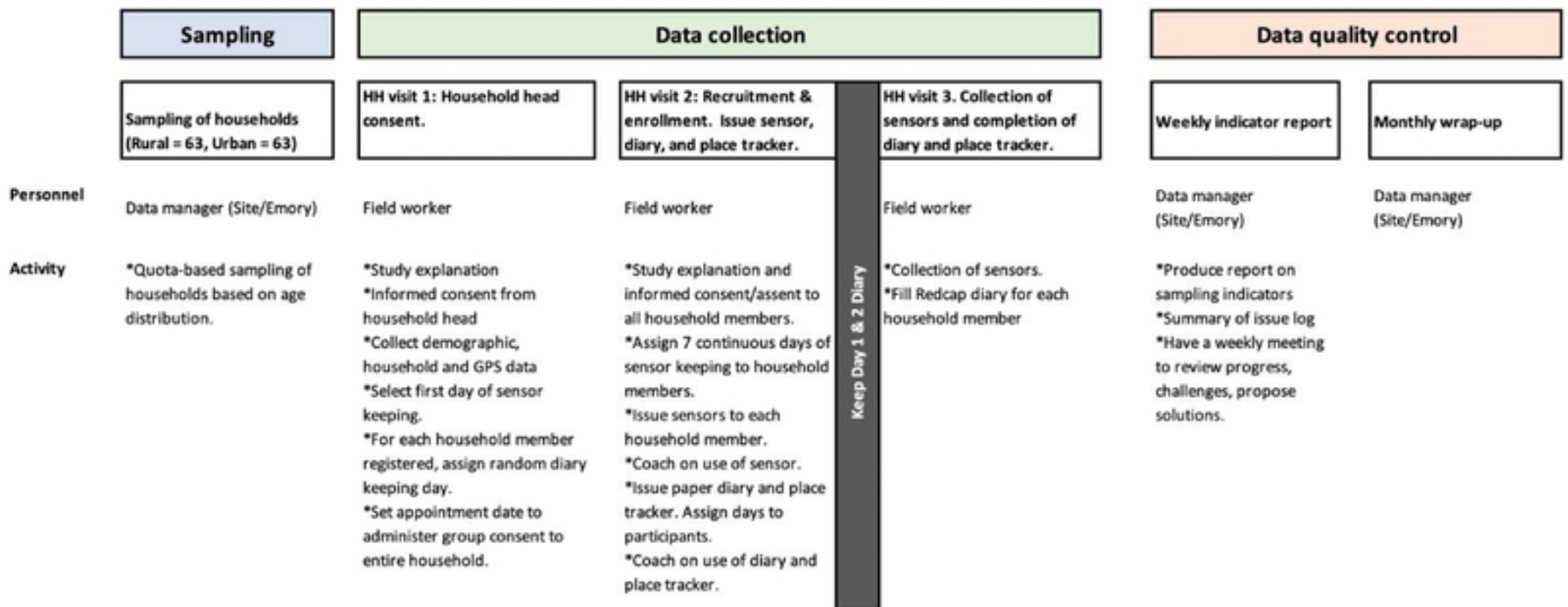


Figure 4

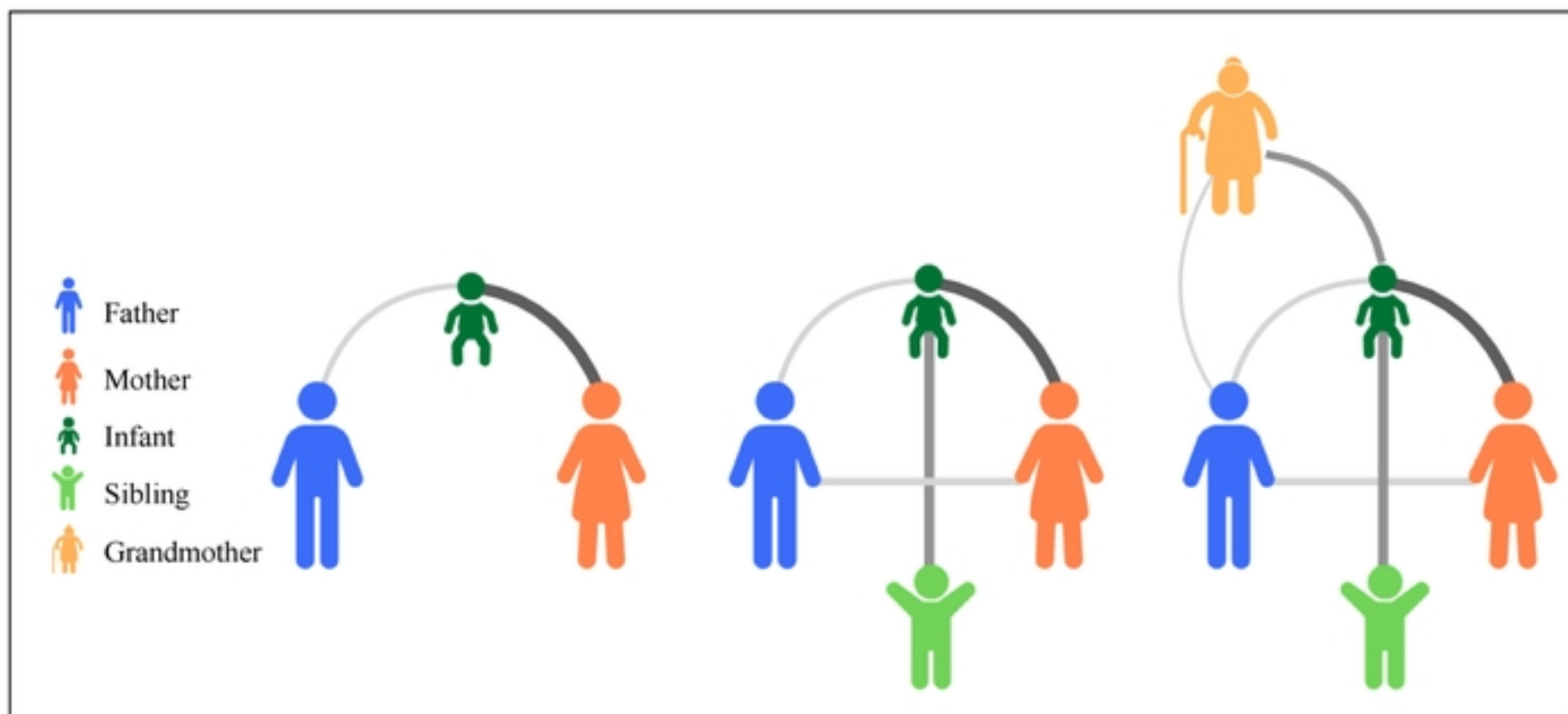


Figure 5