ROYAL SOCIETY OPEN SCIENCE

royalsocietypublishing.org/journal/rsos

Research

Cite this article: Walker J, Aylett-Bullock J, Shi D, Kahindo Maina AG, Samir Evers E, Harlass S, Krauss F. 2023 A mixed-method approach to determining contact matrices in the Cox's Bazar refugee settlement. *R. Soc. Open Sci.* **10**: 231066. https://doi.org/10.1098/rsos.231066

Received: 24 July 2023
Accepted: 9 November 2023

Subject Category:
Mathematics

Subject Areas:
health and disease and epidemiology/computational biology

Keywords:
contact matrices, simulation, individual-based model

Author for correspondence:
Joseph Walker
e-mail: j.j.walker@durham.ac.uk

A mixed-method approach to determining contact matrices in the Cox's Bazar refugee settlement

Joseph Walker[1,2], Joseph Aylett-Bullock[1,3], Difu Shi[1,4], Allen Gidraf Kahindo Maina[5], Egmond Samir Evers[6], Sandra Harlass[7] and Frank Krauss[1,2]

[1]Institute for Data Science, Durham, UK
[2]Institute for Particle Physics Phenomenology, Durham, UK
[3]United Nations Global Pulse, New York, NY, USA
[4]Institute for Computational Cosmology, Durham, UK
[5]UNHCR Public Health Unit, Cox's Bazar, Bangladesh
[6]WHO Headquarters, Geneva, Switzerland
[7]UNHCR Public Health Unit, Geneva, Switzerland

JW, 0000-0001-8734-8293; JA-B, 0000-0001-7551-3423; FK, 0000-0001-5043-3099

Contact matrices are an important ingredient in age-structured epidemic models to inform the simulated spread of the disease between subgroups of the population. These matrices are generally derived using resource-intensive diary-based surveys and few exist in the Global South or tailored to vulnerable populations. In particular, no contact matrices exist for refugee settlements—locations under-served by epidemic models in general. In this paper, we present a novel, mixed-method approach for deriving contact matrices in populations, which combines a lightweight, rapidly deployable survey with an agent-based model of the population informed by census and behavioural data. We use this method to derive the first set of contact matrices for the Cox's Bazar refugee settlement in Bangladesh. To validate our approach, we apply it to the UK population and compare our derived matrices with well-known contact matrices collected using traditional methods. Our findings demonstrate that our mixed-method approach successfully addresses some of the challenges faced by traditional and agent-based approaches to deriving contact matrices. It also shows potential for implementation in resource-constrained environments. This work therefore contributes to a broader aim of developing new methods and mechanisms of data collection for modelling disease spread in refugee and internally displaced person (IDP) settlements and better serving these vulnerable communities.

THE ROYAL SOCIETY PUBLISHING

1. Introduction

Epidemics such as COVID-19 have led to devastating consequences for afflicted individuals and their societies. Understanding how such infectious diseases spread, anticipating future trajectories for transmission and gathering evidence to inform decision-making efforts to prevent, mitigate and respond to epidemics is therefore of vital importance. Mathematical and computational models to simulate disease spread are regularly used to support these efforts. Contact matrices are key to understanding social mixing patterns in populations, and a vital input to epidemiological models [1,2]. Despite renewed efforts to develop such models, additional work must be done to ensure they are available to all [3].

In this paper, we present a new method for determining contact patterns based on combining the information gained from increasingly sophisticated models of disease spread with that from lightweight surveys, which can be rapidly rolled out to populations of interest. We attempt to provide information on contact patterns without requiring the traditional, costly methods of contact data collection. Specifically, we will focus on the use case of the Cox's Bazar refugee settlement in Bangladesh. Epidemics in refugee and internally displaced person (IDP) settlements are commonplace and tend to spread rapidly [4], and only very few models have been designed to simulate outbreaks in these unique environments and to inform public health decision-making [3]. Given the application domain, we believe this is not just an important area in which to contribute knowledge about disease spread patterns, but also a challenging test case which demonstrates the strengths of our methodology.

Throughout this work, we will use the June-Cox model [5], an agent-based model built on the June framework [6]. The model constructs a virtual population at the level of individual residents within a digital twin of the Cox's Bazar settlement. Interactions are simulated between the agents—the virtual residents—in a number of 'venues' or 'locations' that include shelters, food distribution centres, market places and learning centres. We use the information from the lightweight survey to guide these interaction patterns based on the demographics of the agents attending the venues contemporaneously. The contact matrices encode information on the number and duration of contacts between people of one age group and another, and are usually specific to certain venues or locations in which people interact. There are various types of matrices which can be used both separately and in combination, including (i) one-directional, contact matrices NCM [7], which count the (normalized) number of contacts a person in category i has with a person in category j, (ii) bi-directional reciprocal matrices $\mathrm{NCM_R}$ [7], which also add the number of contacts people in category j have with persons in i, and (iii) venue contact matrices $\mathrm{NCM_V}$ [8,9] which assume that every person at venue L has contact with everybody else present. In this article, we will discuss an approach to estimating all three types of matrices.

Traditionally, contact matrices are derived using large-scale surveys in which participants record the number of contacts they have in different locations and the ages of the people they came into contact with. Additional metadata are sometimes collected, such as the intensity of the contact (e.g. physical or non-physical) and the duration of each individual contact. Surveys of these types have predominantly been run in the Global North, with comparatively few serving countries in which many particularly vulnerable communities reside [10]. Indeed, to date and to our knowledge only one work has published contact matrices for an IDP settlement [11], and no such work exists on contact matrices in refugee settlements. While such traditional diary-based methods of collecting contact data may be considered the gold standard, they are extremely resource-consuming to collect, and therefore cannot be easily run during an ongoing outbreak. As an alternative to these expensive direct means of contact data collection, several other methods have sought a more indirect approach. Using the information from existing contact surveys conducted in eight European countries [7], and knowledge of the underlying demographic structures in these populations, Prem *et al.* [12] used a Bayesian hierarchical model to project these matrices onto those of 144 of countries, given similar demographic data and underlying similarities between each of these countries and the original eight selected in the direct data collection. This has recently been expanded to 177 countries [13].

Similarly, census/demographic data have also been used to construct synthetic populations, which are then used to estimate contact matrices. Fumanelli *et al.* [8] use such data from 26 European countries to construct representative synthetic household, school, workplace and 'general community' environments and then assume that each individual in each setting has a single contact with every other member. This has been extended to 35 countries, while also incorporating finer-grained data to develop more representative virtual populations [14]. The same approach is used by Xia *et al.* [15] for the setting of Hong Kong. While such approaches are beneficial as they do not require the expensive collection of long-term contact survey data, they are limited by the assumption that different venues contain static populations and that within-venue mixing is homogeneous.

By combining demographic data with data sources such as time-use surveys [16,17] or transportation surveys [9], stochastic approaches—e.g. agent-based models—have been developed to capture a broader variety of mixing patterns in populations. These approaches expand on those described above by exploring many permutations of possible within-venue mixing patterns. Despite this, these methods still present similar limitations as those described above. Namely, in the absence of any prior information on interaction patterns, it is largely assumed that each agent contacts every other agent in those venues. As a partial remedy to this challenge, disease data are commonly used to fit arbitrary hyperparameter multipliers to these matrices. While this is generally a necessity to be able to forecast disease spread even when using directly collected contact data [18], due to differences between disease transmission routes this may not resolve the errors at the matrix element level. Indeed, the output of this process does not provide an understanding of the base level of contacts, but rather a set of contact matrices for each disease. This limits the usefulness and general applicability of such matrices in comparison with corresponding matrices from directly collected data.

In this paper, we seek to contribute at two levels: (i) we develop a methodology which addresses the challenges above by taking a mixed-method approach to deriving contact matrices. It combines techniques of extracting contact matrices from sophisticated agent-based models, with information derived from a lightweight survey designed to inform and validate the model-derived matrices, while being significantly less expensive to run than the traditional large-scale contact surveys. (ii) We use this new approach to present, to the best of our knowledge, the first contact matrices for a refugee settlement. Because of their use in different types of models, the matrices need different normalizations, either to the full population, as in the case of location-unspecific simple compartment models of the SEIR type, or to the part of the population actually visiting a venue. We will therefore present results for all three types of contact matrices, for a variety of locations, either normalized to the overall population 'P' type contact matrices (PNCM, $\mathrm{PNCM_R}$ and $\mathrm{PNCM_V}$) or to the actual users of a location 'U' type matrices (UNCM, $\mathrm{UNCM_R}$ and $\mathrm{UNCM_V}$).

This work therefore also contributes to the global call to action laid out in prior work, which aims, among others, to develop new methods and mechanisms of data collection for modelling disease spread in refugee and IDP settlements [3].

2. Methods

The goal of our method is to construct location-dependent social contact matrices with a high level of granularity without resorting to detailed contact surveys. We achieve this by fitting the (virtual) contact matrices of an individual-based model constructed from higher-resolution demographic data of the population to the real-world results from lightweight surveys with a much lower resolution. The resolution and accuracy implicit to the model allows us not only to infer the highly granular contact matrices, but also allows us to give a first estimate of the associated uncertainties. In the following, we further detail this procedure and exemplify it with the construction of social contact matrices for the residents of Cox's Bazar refugee settlement.

2.1. Cox's Bazar refugee settlement

In this work, we specifically focus on the Kutapalong–Balukhali Expansion Site of the Cox's Bazar refugee settlement. With over 600 000 persons of concern (PoC), the expansion site is one of the largest refugee settlements in the world [19]. The refugees are primarily Rohingya who have fled targeted violence and serious human rights violations in Myanmar since August 2017 [20]. A number of risk factors make the settlement vulnerable to epidemic outbreaks, including: high rates of global acute malnutrition and other comorbidities such as respiratory illnesses, which could lead to lower general immunity among camp residents [21]; high population density and communal facilities, which increase the risk of person-to-person transmission; and limited access to sources of reliable information as well as low levels of literacy, which make public health campaigns challenging.

The settlement is organized into 22 camps (regions) as depicted in figure 2. Each camp contains approximately 30 000 people each (figure 1). The camps are very densely populated, with on average seven people per shelter (cf. figure 20). Owing to the young demography of the settlement, the majority of households contain children and therefore a significant proportion are multi-generational households (cf. figure 19) which increases the risk factor of disease spread across all age groups.

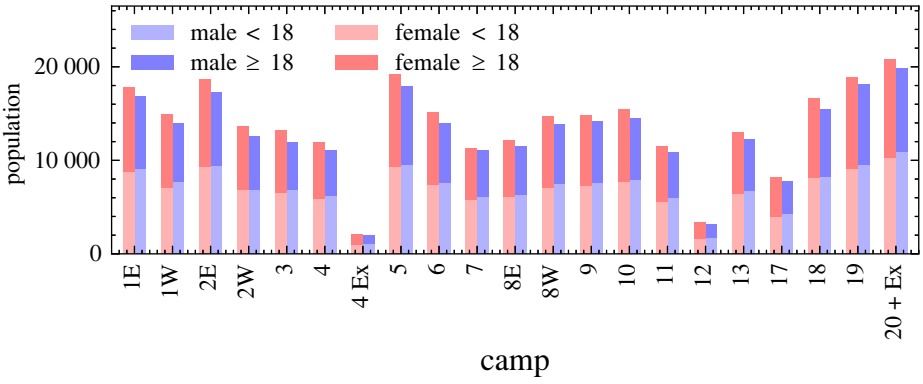

Figure 1. Demography of each camp modelled in June-Cox. The population of men is shown in blue, women in red and the proportion of adults to children is represented by the higher and lower portion of each bar, respectively. In June-Cox, we combine Camp-20 and the Camp-20 extension together forming camp '20 + Ex'.

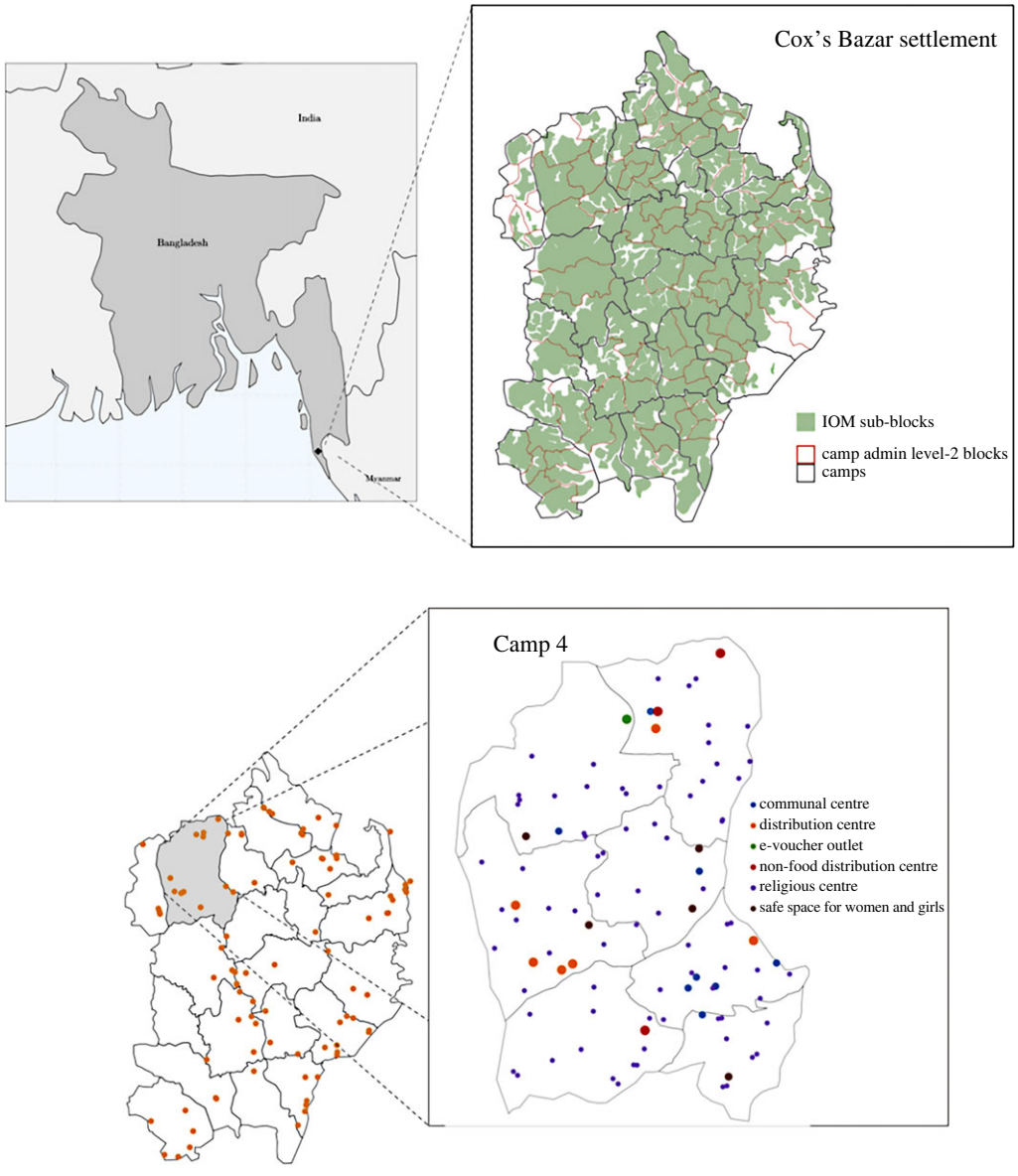

Figure 2. Kutapalong–Batukhali Expansion Site and digital twin geographical and location information. Upper left: map of Bangladesh showing location of the Cox's Bazar refugee settlement. Upper right: map of the modelled Expansion Site with three geographical layers. Lower left: modelled distribution centres in the Expansion Site. Lower right: detailed view of Camp 4 showing six types of modelled locations implemented in the June-Cox digital twin. Basemaps from [22,23].

The social spaces and amenities are typically spread evenly throughout Cox's Bazar according to population density. However, the number of venues of each type is highly variable, resulting in people possibly having to travel between camps to access amenities. In the camp, there exist 115 communal centres, 21 distribution centres, 7 e-voucher outlets, 67 female-friendly spaces, 1244 learning centres, 41 non-food distribution centres, 2032 pumps and latrines and 2065 religious centres. Owing to increased health risk factors and frequent use of shared spaces, it is therefore imperative that the social interactions across the households and various venues be understood.

2.2. The survey

The level of detail accessed by surveys in refugee camp settings is often heavily constrained by resource considerations (timing, number of enumerators, need for rapid results, etc.), and the highly aggregate contact survey we ran in the Cox's Bazar refugee settlement between October and November 2020 is no exception. During this period, the settlement continued to experience cases of COVID-19 [24]. However, reported case numbers were low, and the settlement activity had largely returned to pre-pandemic levels, with the exception that learning centres (schools) remained closed and masks were still being worn [25] (Cox's Bazar refugee operation Community Based Protection (CBP) team 2022, personal communication). The following demonstrates the ability to rapidly run a survey during a public health emergency, in a resource-light way, while producing representative results of the contact patterns which can be used in future studies and modelling works. Although a more intensive survey—such as a diary-based longitudinal study—would provide more precise and accurate data, the ability to perform such a survey may be limited by the number of researchers available or more practical concerns such a limiting social contacts between members of the community and enumerators during a public health crisis.

The survey underpinning our study was conducted by experienced enumerators from the United Nations High Commissioner for Refugees (UNHCR) Community Based Protection (CBP) team, following standard UNHCR practices [26,27]. Its objective was to collect information on the number of contacts people of different demographics have with others in different venues they attend during a typical day. The survey considered only three categories of residents, defined by their age: children (less than 18 years), adults (greater than or equal to 18 and less than 60 years) and seniors (greater than or equal to 60 years), and we constrained the set of surveyed locations to those contained in the digital twin, June-Cox. Data were collected from two male and two female residents in each of the three age brackets in each of the 22 camps in Cox's Bazar refugee settlement. In addition, two persons with disabilities were surveyed in each camp, resulting in a total of $22 \times 14 = 308$ respondents. Details of the survey can be found in appendix C and the accompanying metadata to the anonymized results [28]. The respondents were asked if they attended various venues and, if so, to estimate the number of adults and children they had come into contact with there. To avoid skewing results through uncharacteristically long or short times at a venue, the respondents were asked how much time they generally spend at those venues at any given visit such that the total contacts can be re-scaled to contacts per hour. Since the June modelling framework and many of the demographic data underpinning June-Cox do not distinguish adults (18–59) and seniors (greater than 59) we choose to combine the data in these two age bins into one 'adult' category, thereby arriving at highly aggregate 2×2 total contact matrices t_{ij}[1] for the various locations L[2]. We use the survey to calculate $\mathrm{UNCM}_{\mathbf{R}}$ type matrices for different locations. Here we present the methodology to calculate the different versions of the contact matrices:

(i) One-directional contact matrices [7], NCM, (UNCM and PNCM): following the notation in [29], the PNCM are denoted as M with elements m_{ij} defined by $m_{ij} = t_{ij}/n_j$ with t_{ij} the aggregate total number of contacts of n_j survey respondents in category j reported with people in category i.

There is a subtle difference to the UNCM with elements μ_{ij}, where the aggregate number of contacts t_{ij} is normalized to the number of actual users in the venue, η_j $\mu_{ij} = t_{ij}/\eta_j$. To make contact between the PNCM and UNCM, one therefore merely has to re-normalize to the overall number of respondents in category j, $m_{ij} = t_{ij}/n_j = \mu_{ij}\eta_j/n_j = \mu_{ij}\, a_j$, where a_j denotes the attendance rate to the venue in category j. This re-normalization can be performed for any conversion from population normalized 'P' to user 'U' normalized matrices.

[1]We interpret contact matrix Δ_{ij} such that person i contacts person j and graphically as subgroup on x-axis contacts subgroup on y-axis.

[2]To improve the readability of the manuscript we refrain, where possible, from explicitly indexing contact matrices, etc., with a location index.

(ii) Bi-directional, reciprocal contact matrices [7], $\mathrm{NCM}_{\mathbf{R}}$, ($\mathrm{UNCM}_{\mathbf{R}}$ and $\mathrm{PNCM}_{\mathbf{R}}$): following, again [29], the $\mathrm{PNCM}_{\mathbf{R}}$ are denoted by C and their elements are defined as

$$c_{ij} = \frac{1}{2}\left(m_{ij} + m_{ji}\frac{w_i}{w_j}\right) = \frac{1}{2w_j}\left(t_{ij}\frac{w_j}{n_j} + t_{ji}\frac{w_i}{n_i}\right), \tag{2.1}$$

where the $w_{i,j}$ are the overall population sizes in categories i and j. This motivates the notion of these matrices being normalized to the overall population. While using these matrices in compartment models, their application in individual-based models may lead to unwanted results. As an example, consider the case of contacts between adults and children in school settings, and assume that this is meant to primarily capture the contact of teachers and pupils. Normalizing the number of contacts to the overall adult population size would obviously lead to a massively reduced average number of contacts compared with a more correct normalization to the number of teachers in the respective age bins. We therefore define the user-normalized contact matrices $\mathrm{UNCM}_{\mathbf{R}}$ Γ with entries

$$\gamma_{ij} = \frac{1}{2\omega_j}\left(t_{ij}\frac{\omega_j}{\eta_j} + t_{ji}\frac{\omega_i}{\eta_i}\right), \tag{2.2}$$

where $\omega_{i,j}$ denote the actual users attending the venue, i.e. $\omega_i = w_i a_i$. In fact, since we resolve the random movement of individuals to distinct locations in June, we use the Γ instead of the C that are more relevant for compartment models. However, we also present results for the population-normalized $\mathrm{PNCM}_{\mathbf{R}}$, which can be obtained by simple rescaling by attendance factors a_i and a_j from the Γ.

(iii) Isotropic venue contact matrices, $\mathrm{NCM}_{\mathbf{V}}$, ($\mathrm{UNCM}_{\mathbf{V}}$ and $\mathrm{PNCM}_{\mathbf{V}}$): due to the lack of attendance data we cannot directly derive such matrices b_{ij} and β_{ij} from the survey. However, they can be determined virtually.

There is one further subtlety: the contact survey was conducted such that there is an equal number of men and women respondents. However the population of Cox's Bazar is 48% and 52% men and women by population, respectively, we therefore re-weight the contact matrices by sex. Finally, we comment on our treatment of the uncertainties in the survey results. Given the small survey sample size, we right-censor the data at the level of the 90th percentile and perform a bootstrap analysis [30] to determine the median number of contacts between subgroups, μ_{ij}. We assume the uncertainty of this value, $\Delta\mu_{ij}$, to be well estimated by the standard error of the bootstrap distribution of the median. We note that the contact survey is done retrospectively; hence, we expect an element of recall bias within the respondents' answers. We choose to right-censor the distribution of numerical answers to remove more extreme responses and, furthermore, we take the median statistic and apply a bootstrap method to estimate median contact patterns and their errors. These errors can be used within our model to apply further stochasticity to agents' contact patterns, i.e. increasing the variability in their behaviours. From $\Delta\mu_{ij}$, it is straightforward to derive the uncertainty, $\Delta\gamma_{ij}$, of the reciprocated matrices, we assume the error in the contacts are dominated by the error from reported number contacts per hour at a venue. We take $\omega_{i,j}$ in equation (2.2) as an exact quantity from the survey.

The lightweight survey in the camp was conducted across the following venues: 'community centres', 'distribution centres', 'e-voucher outlets' and 'formal education centres'. For the remaining two venues that exist in June-Cox—'play groups' and 'shelters'—we assume that everyone generally mixes with everyone else in that location given the assumed small groups of children who play together, as well as the dense shelter environments. Since certain shelters are shared between multiple families, we differentiate intra- and inter-family mixing with the latter being represented by the diagonal elements of the aggregate matrix (i.e. setting these to the number of contacts within each of the two families in the shelter, and with the off-diagonal elements set to the number of contacts between the families). As discussed in previous work [5], we set the number of contacts within the families or play groups to the average size of these respective groups assuming homogeneous mixing in these settings. In the case of the play groups, we disaggregate the population into three age groups 3–6, 7–11 and 12–17 which mix homogeneously to emulate children typically interacting with children of similar age (table 1).

Table 1. Summary of the CM notation and reported matrix types. The CM are time normalized, the various UNCM are further normalized by population at the venues, while the PNCM are instead normalized by the *total* population.

matrix	symbols	matrix	symbols
CM	t, t_{ij}		
UNCM	μ, μ_{ij}	PNCM	M, m_{ij}
$\text{UNCM}_{\mathbf{R}}$	Γ, γ_{ij}	$\text{PNCM}_{\mathbf{R}}$	C, c_{ij}
$\text{UNCM}_{\mathbf{V}}$	β, β_{ij}	$\text{PNCM}_{\mathbf{V}}$	B, b_{ij}
population venue	η_{ij}	population world	ω_{ij}
population survey venue	n_{ij}	population survey world	w_{ij}

2.3. The model

For the construct ion of the digital twin and simulator, we use an existing individual-based model, June-Cox [5], specifying the original June modelling framework [6] to the demographics of the Cox's Bazar refugee settlement. (Note that the original application of the June framework was to model the spread of COVID-19 in the UK and we will refer to this UK-specific specification as June-UK.)

2.4. The simulation framework: June

The June framework is broken down into several layers of digital world construction.

2.4.1. Geography and demography

Both June-UK and June-Cox use census data to create virtual populations at the individual level. The census data of its population is organized according to a geographical hierarchy; approximately 600 000 residents are distributed over the 21 camps[3] (regions) which make up the digital twin of Cox's Bazar, June-Cox. Each camp contains between two and seven UNHCR Admin level-2 blocks (super areas) comprising approximately 5000 people, which in turn are composed of sub-blocks (areas) with 90 households on average.

2.4.2. Household construction

The census data of the population provide the distribution of individuals' locations and demography in each area. The household is explicitly incorporated in the model through the geo-locations of the area centres. For a more complete description of how we distribute individuals into households, see appendix E and the original work describing June-Cox [5].

2.4.3. Construction of venues

June-Cox constructs different venues in the settlement given their latitude and longitude coordinates: food distribution centres, non-food distribution centres (including liquid gas distribution centres), e-voucher outlets, community centres, safe spaces for women and girls, religious centres, learning centres, and hand pumps and latrines.

2.4.4. Simulation of social interactions

June uses calendar days to distinguish weekday and weekend activity profiles where certain venues will be closed.[4] To simulate the movement of individuals in the settlement, we decompose each calendar day into discrete time-steps in units of hours. Many individuals have fixed, static activities, such as the 4 h at the learning centres for enrolled children and the adults specified as teachers. There is also a fixed 14 h

[3]In reality, there are 22; however, we combine Camp-20 and the Camp-20 extension together given data availability constrains.

[4]This can be further customized in June to specific daily closures or behavioural patterns; however, for this proof of concept we only specify weekday and weekend variations.

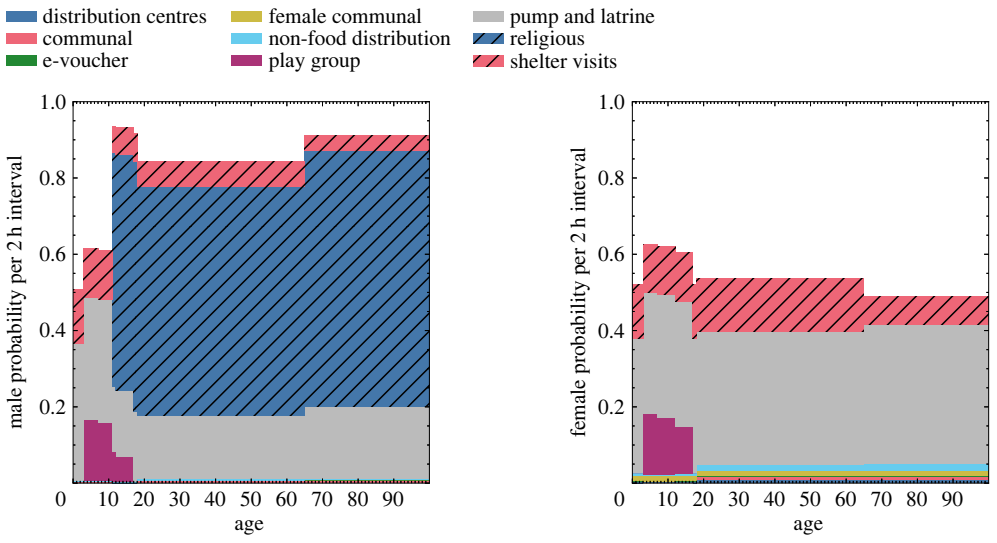

Figure 3. The mean probability P to attend certain venues in any weekday 2 h time-step interval Δt by age for left, men; right, women.

night-time period, during which everyone returns to their shelter. However, the remaining time is free and people are distributed dynamically. Each person not otherwise occupied (i.e. working, or at a medical facility) is assigned a set of probabilities for undertaking other activities in their free time in the model. These probabilities are part of our social interaction model, and depend on the age and sex of the person (figure 3). They are based on previously collected data capturing daily attendance rates and coarse estimates in proportions of adult/child and male/female attendance (see previous work for details on these calculations and associated data sources [5] and have been further augmented by a series of interviews with CBP officials as detailed in appendix D and probabilities tuned such that populations reflect the interview; see figure 16).

Given N possible activities with associated probabilities per hour given by $\lambda_1, \ldots, \lambda_N$, for a person with characteristic properties p, the overall probability, P, of an individual being involved with any activity in a given time interval Δt is modelled through a Poisson process,

$$P = 1 - \exp\left(- \sum_{i=1}^{N} \lambda_i(p)\Delta t \right). \tag{2.3}$$

If the individual participates in at least one of these activities, the specific activity i is selected according to

$$P_i = \frac{\lambda_i(p)}{\sum_{i=1}^{N} \lambda_i(p)}, \tag{2.4}$$

and the person is moved to the relevant location. If no activity is selected, the individual will stay in their shelter.

One of the outcomes of this procedure is condensed in figure 3, which shows the probabilities that men and women attend the different venues in the model as a function of their age.

It is important to stress that such census and demographic data is by default recorded by UNHCR and other non-governmental organizations (NGOs) operating in refugee and IDP settlements, and it can be further supplemented or clarified by the survey described above or by interviews with settlement staff. This implies that it is a relatively straightforward exercise to apply our procedure outlined here to other settlements.

2.5. A mixed-method approach

We have now set the stage to combine information about the aggregate contact patterns with our highly detailed model of interactions in a representative virtual population and to interrogate the model and extract detailed, survey-informed matrices. In this work, new code was developed within the June framework to implement contact tracking and contact matrix construction. The June contact tracking framework uses stochastic methods to simulate contacts between members of the virtual population, which can be used to construct synthetic CMs. In June, the random behaviour of the virtual

population is encoded in repeatedly sampling the γ_{ij} from a Poisson distribution, $\tilde{\gamma}_{ij} \sim \mathcal{P}(\kappa_{ij})$ with the argument κ_{ij} distributed according to a normal distribution,

$$\kappa_{ij} \sim \mathcal{N}(\bar{\mu}, \sigma) \quad \text{with } \mu = \frac{\Delta T}{T}\gamma_{ij} \text{ and } \sigma = \frac{\Delta T}{T}\Delta\gamma_{ij}, \tag{2.5}$$

with the γ_{ij} and their uncertainty ($\Delta\gamma_{ij}$) taken from the survey and re-scaled by the ratio of the typical time people attend a location, T (table 4), and the size of the emulation time-step in the model, ΔT. It is important to sample the argument κ_{ij} so that we encapsulate errors from the bootstrap methodology in the contact survey into our model. Finally, we statistically round the individual instances $\tilde{\gamma}_{ij}$ to integer values, since the model agents can only contact a discrete number of other agents within the model with discrete characteristics. The resulting emulated set of $\tilde{\gamma}_{ij}$ are normalized such that they represent an individual agent's integer contacts per time-step. Averaging generates the $\hat{\gamma}_{ij}$ which can be directly compared with the γ_{ij} obtained from the survey.

In the simulation, we aim to perform a virtual survey on the virtual population, as close as possible to the conditions in the real-world lightweight surveys. We sample individual behaviour over 28 virtual days to obtain individual $\tilde{\gamma}_{ij}$s every time an agent attends a venue. The venues are filled according to the probabilities described above (equations (2.3) and (2.4)), and we 'measure' the total raw contacts $\hat{t}_{ij}$ (see algorithm 1 in appendix F) in the simulation. To further ensure the correct total expected attendance time at the virtual venues compared with the real world, we proportionally close venues to approximate their possible fractional opening times.

This procedure allows us to directly compare resulting matrices $\hat{t}_{ij}$, $\hat{\gamma}_{ij}$ and $\hat{c}_{ij}$ with their real-world counterparts t_{ij}, γ_{ij} and c_{ij} above. Even more, we are not constrained to the creation of virtual 2×2 contact matrices only, but can infer matrices for any sub-classification i and j that our simulation allows—in the results we present here, the i and j are age brackets of size 1 year. The final type(s) of contact matrix, PNCM_V and UNCM_V, $\hat{b}_{ij}$ and $\hat{\beta}_{ij}$, can also be calculated with a minor modification to the contact tracking algorithm. Instead of generating a list of people p_j at the venue in contact with each person p_i, we allow 'democratic/isotropic' contacts of all people,

$$\hat{\beta}_{ij} = \hat{\eta}_j - \delta_{ij}. \tag{2.6}$$

For each entry, ij, this represents the total contacts the $\hat{\eta}_i$ people with characteristics i at the venue have with the population of the venue in each subgroup. The Kronecker δ corrects for 'self-contacts'. For clarity, a summary of the notation and the contact matrix types can be found in table 1.

2.6. Mixed-method validation

We begin by validating our method in the context of the UK where we compare our results against contact patterns directly collected by a traditional survey [29]. Recognizing that the UK and Cox's Bazar are very different settings, this validation step simply acts as a closure test and sanity check in the context of a well-understood and well-studied setting. Once our method has been validated, we present the matrices for the Cox's Bazar refugee settlement. Throughout, we use several key metrics to determine the similarity between any two sets of matrices:

(i) Normalized Canberra distance, D_C [31],

$$D_C(C, C') = \frac{1}{\text{Dim}(C) - Z}\sum_i \sum_j \frac{|C_{ij} - C'_{ij}|}{|C_{ij}| + |C'_{ij}|}, \tag{2.7}$$

where C and C' represent two contact matrices we wish to compare, Dim denotes the number of elements, $\text{Dim}(C_{n\times m}) = n \cdot m$, and Z is the number of zero elements of the difference $(C_{ij} - C'_{ij})$.

(ii) Q index as measure of assortativity [32],

$$Q = \frac{\text{Tr}(C/\sum_{ij} C_{ij}) - 1}{\text{Dim}(C) - 1}. \tag{2.8}$$

(iii) Dissimilarity index, I_s^2 [33],

$$I_s^2 = \frac{1}{2}\frac{\langle (X - Y)^2 \rangle_{F_c}}{\sigma_p^4}, \tag{2.9}$$

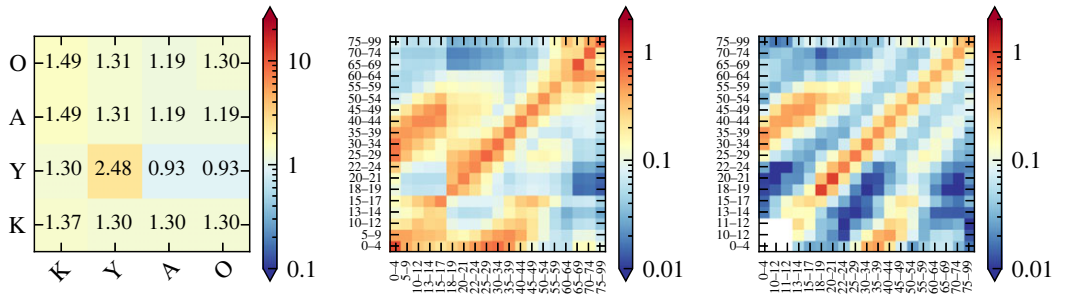

Figure 4. Contact matrices from the UK validation procedure. Left: the derived input interaction matrix, $\text{UNCM}_{\mathbf{R}}$ for 'households' for contacts between K, 'kids' less than 18; Y, 'young adults' 18–25; A, 'adults' 26–65; O, 'old adults' greater than 65. Centre: the simulated age-binned $\text{PNCM}_{\mathbf{R}}$ matrix with entries $\hat{C}_{ij}$ from June-UK. Right: the BBC Pandemic project [29] 'all home' contact matrix, C, with entries c_{ij}. The simulated matrix and the survey matrix share the same colour map for ease of comparison. The input interaction matrix has its own so that the colour map has suitable contrast over the full range of number of contacts.

where σ_p is the standard deviation of the ages of the population, and $\langle (X-Y)^2 \rangle_{F_c}$ represents the expectation age difference between contacts x and y of the function $F_c(x, y)$,

$$F_c(x, y) = \frac{f(x) C^L_{xy} f(y)}{\sum_x \sum_y f(x) C^L_{xy} f(y) \Delta x \Delta y}. \tag{2.10}$$

Here, Δx and Δy are the age bin sizes from the contact survey.

The normalized Canberra distance gives an estimation of the similarity between two matrices—approaching 0 when they are more similar and 1 when dissimilar. The remaining statistics measure the level of assortativity—the level of diagonal dominance and therefore the rate at which similar ages interact compared with dissimilar ages. The Q index ranges from 0—homogeneous, proportionate mixing—to 1—fully assortative. I_s^2 measures the deviation from perfect assortativity with a value of 0 when fully assortative, and 1 for homogeneous interactions.

3. Results

In this section, we present the results of the contact matrices derived from our mixed-method approach. We will start with evidence for the validation of the mixed-method approach by comparing June-UK with real-world survey-derived contact matrices. Then we will report the new Cox's Bazar-derived contact matrices.

3.1. UK validation

The first step of our virtual survey validation is to compare our results with that of real surveys conducted in far greater granularity. June-UK has been extensively tuned for COVID-19 modelling in the UK [6,18,34]. As a proof-of-concept, we focus on the most complex contact matrix—that of households—and compare results produced by the simulation with those from a traditional diary-based survey [29]. The input contact matrix is constructed from a combination of this data, the Office of National Statistics (ONS) census data of UK households [35,36] and UK population demographics [37]. Since the UK census for household types distinguishes children (kids, K, less than 18 years old), young adults such as students or other dependent resident (Y, assumed 18–25 years old in June-UK), adults (A, assumed 26–65 years old) and older adults (O, assumed greater than 65 years old), we aggregated the granular contact matrix derived from the survey into a significantly coarser 4×4 matrix mapping the census categories. We also corrected for different household types to better incorporate the details of the venue-specific heterogeneities in their demographic composition. For more details on this procedure, see specifically §4 and appendix C of the original description of the June-UK modelling set-up [6].

The results in figure 4 show the 4×4 input matrix derived from the aggregation process described above and a comparison of the output of the PNCM_{R} $\hat{\Gamma}$ from the June-UK model virtual contact survey with the results of the matrix C from the traditional survey. Corresponding results for workplace and school settings can be found in appendix B. This provides a closure test indicating that June-UK returns realistic contact matrices from coarse aggregate matrices. This closure test explicitly demonstrates that the agent movement and contact tracing infrastructure in June produces self-consistent results provided the virtual

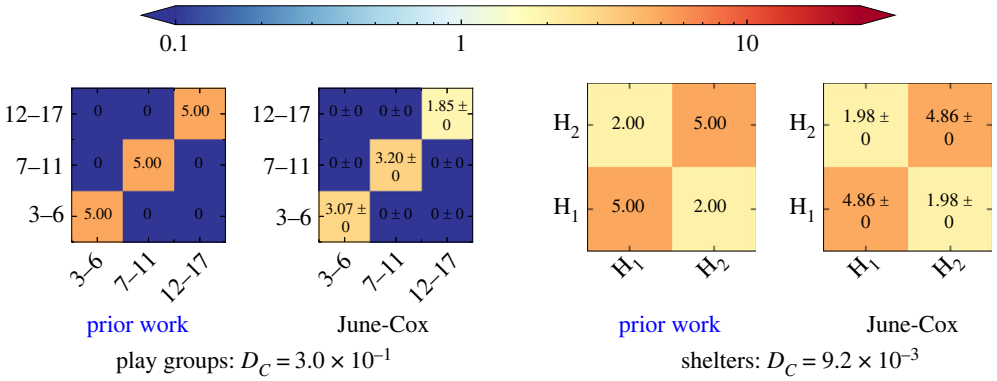

Figure 5. Two pairs of $UNCM_R$ for the virtual venues determined in prior work [5] (left) and the June-Cox virtual survey in the same coarse population bins (right), including the Canberra distance between them.

Table 2. Contact matrix statistics calculated for June-UK and the BBC Pandemic project [29] in figure 4 that of the household. These statistics are calculated for the UK demography reported by ONS in 2011 [37].

	Q	I_s^2	D_C
BBC Pandemic	0.14	0.36	
June-UK $PNCM_R$	0.12	0.30	0.32

setting is a realistic approximation of the real world and its social interactions. Clearly, our mixed-method approach is able to reproduce the broad structure of the real-world data—especially capturing the patterns of contacts between children and their parents represented in the off-diagonal structures. The original survey did not contain information on the contacts of younger children due to constraints on the data collection methodology; our method is able to fill this gap. To further validate our approach, we compare the Q, I_s^2 and D_C metrics of the two matrices. Table 2 shows that the first two metrics are in close agreement, with the overall Canberra distance being close to 0, thereby confirming the similarity of the matrices. Indeed, the difference between the measures of assortativity are comparable or better than those found in similar studies but which do not make use of the guiding input aggregate matrix as we do here [17]. Given these strong findings, together with the visual and structural similarities of the matrices, we consider our mixed-method approach to be reasonably validated for application to settings in which intensive survey-based approaches to deriving contact patterns are not feasible. As our methodology is clearly not exactly reproducing the original surveys, users will have to decide whether these deviations are acceptable in comparison with having little or no knowledge about contact patterns, or making necessary assumptions about these patterns. It is also worth noting that the virtual agent behaviours of June-UK are much better informed than those in June-Cox. This will become clear in the disparity between NCM, NCM_R and NCM_V type contact matrices. $PNCM_V$ matrices presented in figure 15 and $PNCM_R$ matrices in figures 4 and 14 have the same general shape and scaling of features. In the case of June-Cox derived matrices NCM, NCM_R and NCM_V types are less similar.

3.2. Contact matrices in Cox's Bazar refugee settlement

We report the results for the $UNCM_R$ γ_{ij} of the prior work namely, 'play groups' and 'shelters' and of the survey for 'community centres', 'distribution centres', 'e-voucher outlets' and 'formal education centres' in figures 5 and 6, respectively. We also perform a closure test by comparing them with the UNCM $\hat{\mu}_{ij}$ results from performing a similar survey in June-Cox with the same coarse population categories. In the two figures, we use the shorthand 'T' and 'S' for teachers and students in the learning centres, and 'H_x' for household x in a shared shelter.

Once we have determined the UNCM and confirmed that their stochastic uncertainties are within the uncertainties of the input interaction matrices, we can perform any custom binning for arbitrary group characteristics. Figure 7 shows the final fully disaggregated (by age and venue) set of matrices for the Cox's Bazar refugee settlement based on the input contact matrices from the lightweight survey, combined with our highly detailed agent-based model of the settlement. The results of the lightweight survey for each venue can be found in appendix C (cf. figure 6). The combination of these two techniques leads to interesting consequences in the structure of the derived contact matrices. Contact rates from the

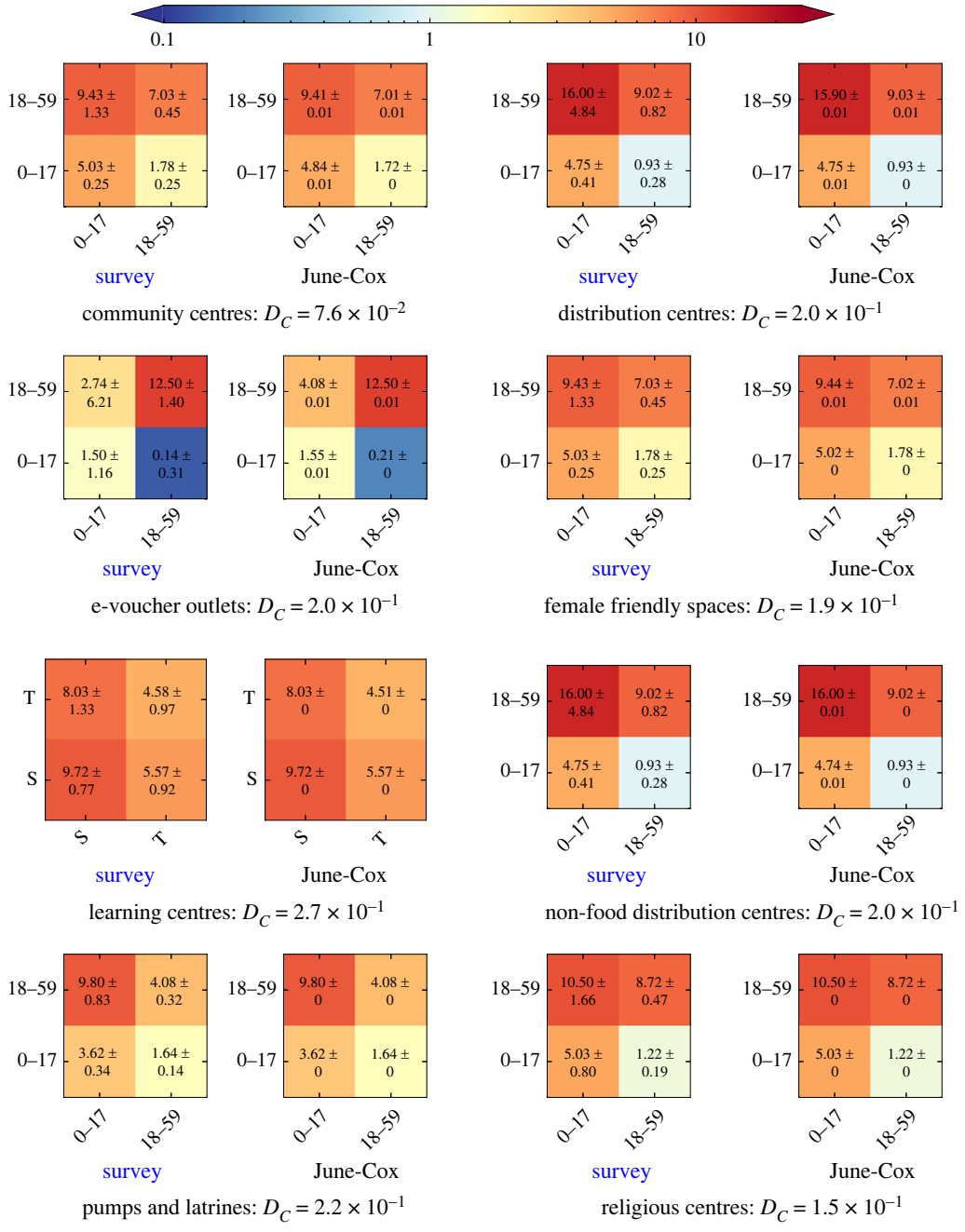

Figure 6. The UNCM_{R} from the lightweight survey data (left) and June-Cox virtual survey UNCM (right), with the relative Canberra distances. We set 'community centres' and 'distribution centres' identical to 'female friendly spaces' and 'non-food distribution centres', respectively.

lightweight survey provide the baseline coarse social interaction patterns between broad subgroups at a given venue, whereas the agent-based model embeds the dynamics from data on the social behaviour of individuals, connecting many independent venues within the model. In particular, we see bands due mainly to 11–18-year-olds for two reasons. Firstly, many behavioural patterns are defined differently for adults and children leading to attendance differences at 18. Secondly, at 11 years of age males are permitted to attend the religious centres. Owing to the high rate of attendance observed at the religious centres, there is a drop in attendance at other non-religious centre venues of this age group relative to other age groups. The corresponding UNCM and UNCM_{V} can be found in appendix A.

In figure 7, we can clearly see the effects of the different age groups and guiding contact rates. For example, we observe large differences in the number of contacts between all age groups with adults in the community centres relative to the distribution centres, with substructures based on the age profile of children attending these locations shown through the higher number of contacts in younger age brackets.

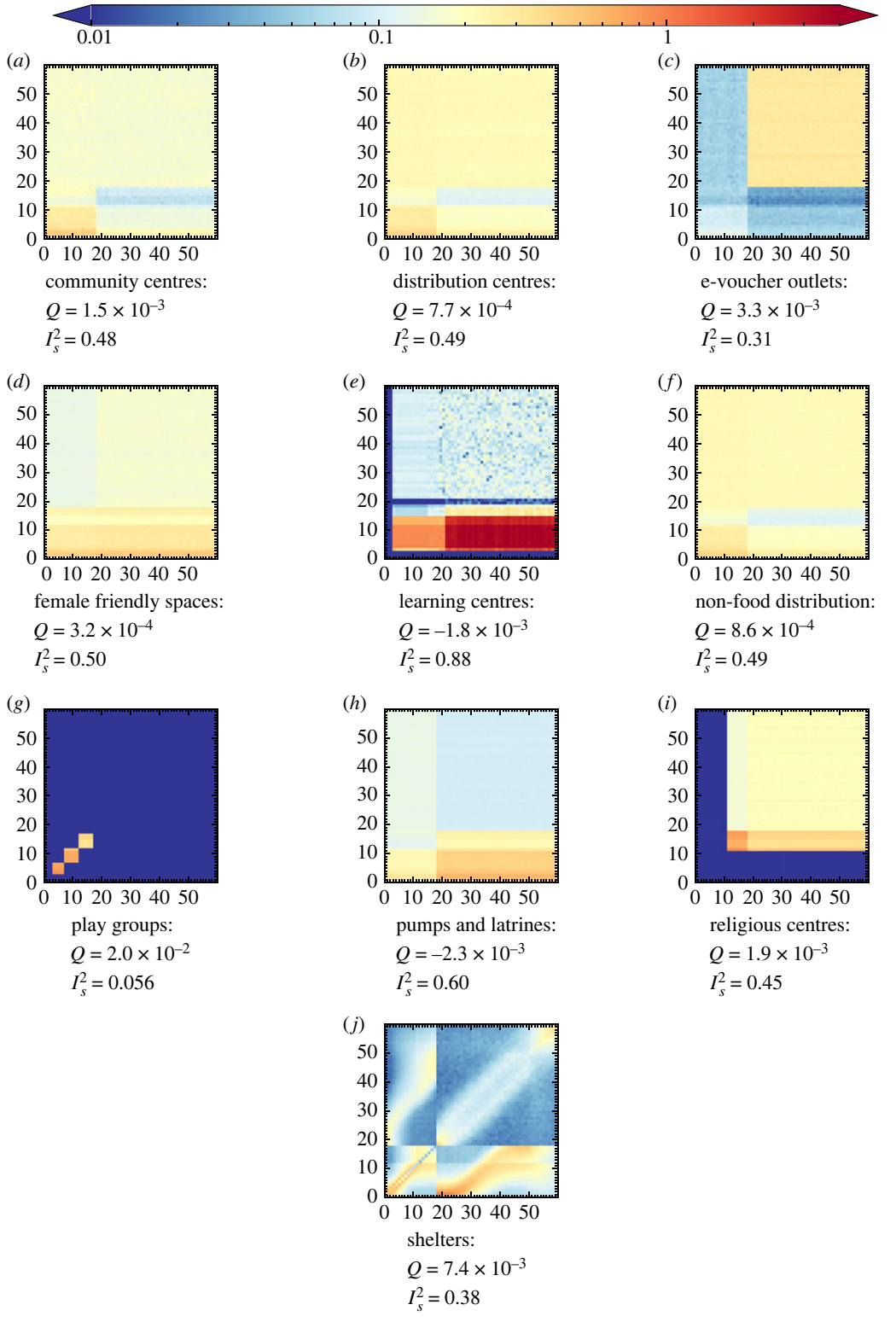

Figure 7. The reciprocal normalized contact matrices ($UNCM_R$) by age as simulated in June-Cox. Note that the data inputs in (g,j) stem from a previous survey.

In addition, the learning centre matrices show a clear mix of contacts between children in their mixed classes and their teachers—this matrix also encodes information on the enrolment rate of children in the education system, with lower enrolment rates as the age of children increase. Finally, the detailed information available on household and shelter composition appears in the shelter contact matrix, which contains a number interesting features. We reconstruct a strong leading diagonal which represents persons of similar ages living together; siblings, parents and grandparents of similar ages. The width of the band reflects spousal

age gaps and minimal age gaps between consecutive siblings. Using more detailed information about the average age of parents at the birth of their first child, we also develop off-diagonal structure in the upper left and lower right quadrants. There exists an almost linear structure corresponding to children and parents interacting and ageing together. This structure then tapers off, indicating interactions in multi-generational households before many children would leave home at around 18. The details of the household construction and the statistics that define it can be found in appendix E.

A simpler approach is to just assume that everyone contacts everyone else in these dense settings in the absence of other information—we also present the results for the corresponding UNCM and UNCM_V in appendix A, figures 10 and 11 and the survey binned UNCM_V in figure 9. However, clearly there is a significant loss of information in doing this, in comparison with the mixed-method approach, as can be seen in the absence of structural detail in many of the UNCM_V matrices.

Population-normalized matrices can be calculated from the user normalized matrices with a simple re-scaling as described above. We present these for completeness and the varied utility of each normalization in different model types in appendix A, figures 8, 12 and 13. Each of the contact matrix elements derived from June come with an associated standard error, these errors are typically very small due to the long run time (28 simulated days) and the small variance in overall attendance rates of difference demographic groups at each of the venues.

4. Discussion

To the best of our knowledge, the matrices presented in this paper are the first contact matrices derived for a refugee settlement. While not collected using traditional survey methods, we use a mixed-method approach for their calculation, which presents a new way to collect contact data. This is particularly useful in settings such as in refugee settlements, in which data collection can present many challenges, and therefore needs to be lightweight and integrated into existing data collection regimes and programming.

We are able to perform closure tests on the contact matrices we derive and show that they clearly demonstrate great potential for a lightweight survey and an agent-based model to provide deeper insights into social environments when combined together. The survey and June-Cox-derived contact matrices are initially validated by a comparison of their Canberra distances over the survey subgroups *ij*. These Canberra distances are found to be very close to zero with the exception of the e-voucher outlets in which child–child contacts are higher in June-Cox than reality. This discrepancy can be explained by considering that the survey has a high uncertainty in the expected child–child contacts, an error which June-Cox incorporates into the contact tracking algorithm.

Further validation is performed with June-UK-derived matrices on age-disaggregated contact matrices in which we are able to use other statistics such as I_S^2 and Q. These matrices were found to be in good agreement with other, more intensive contact surveys. This validation ensures that the combination of coarse input contact matrices and the attendance rates responsible for agent dynamics yield representative contact patterns over all ages.

In the case of refugee settlements, the derived contact matrices can be used to understand the social contact patterns using data already collected regularly by international organizations such as UNHCR, while being supplemented by data which can easily be collected by enumerators in a resource-efficient way. The highly detailed matrices derived for the Cox's Bazar settlement demonstrate clear inter-age mixing patterns, which are crucial inputs to other epidemic models to represent realistic social mixing patterns. In particular, clear features are present in the matrices due to differing attendance rates and household compositions.

From the technical perspective, there are several further considerations and limitations to this methodology that become apparent when analysing the full age-disaggregated contact matrices (figure 7 and appendix A). These pertain to the way in which the data are collected and the model is constructed, and can be used as ways to diagnose the performance of the method:

(i) **Subgroup classification:**
Subgroup classification refers to the broad definition of subgroups defined in the model. Throughout June-Cox and June-UK, we define 'adults', 'children', 'teachers', 'workers', etc., which all have unique parameters and rules governing their behaviour. Subgroups defined by age can lead to strong banding artefacts in the contact matrices. These effects can be mitigated by blurring the age cut-off with some finite probability—e.g. that a child of 17 may behave like an adult. This mitigation should only be implemented in situations in which we are certain that there should not be a discontinuity in behaviours in the real world. For example, only over-11-year-old males are

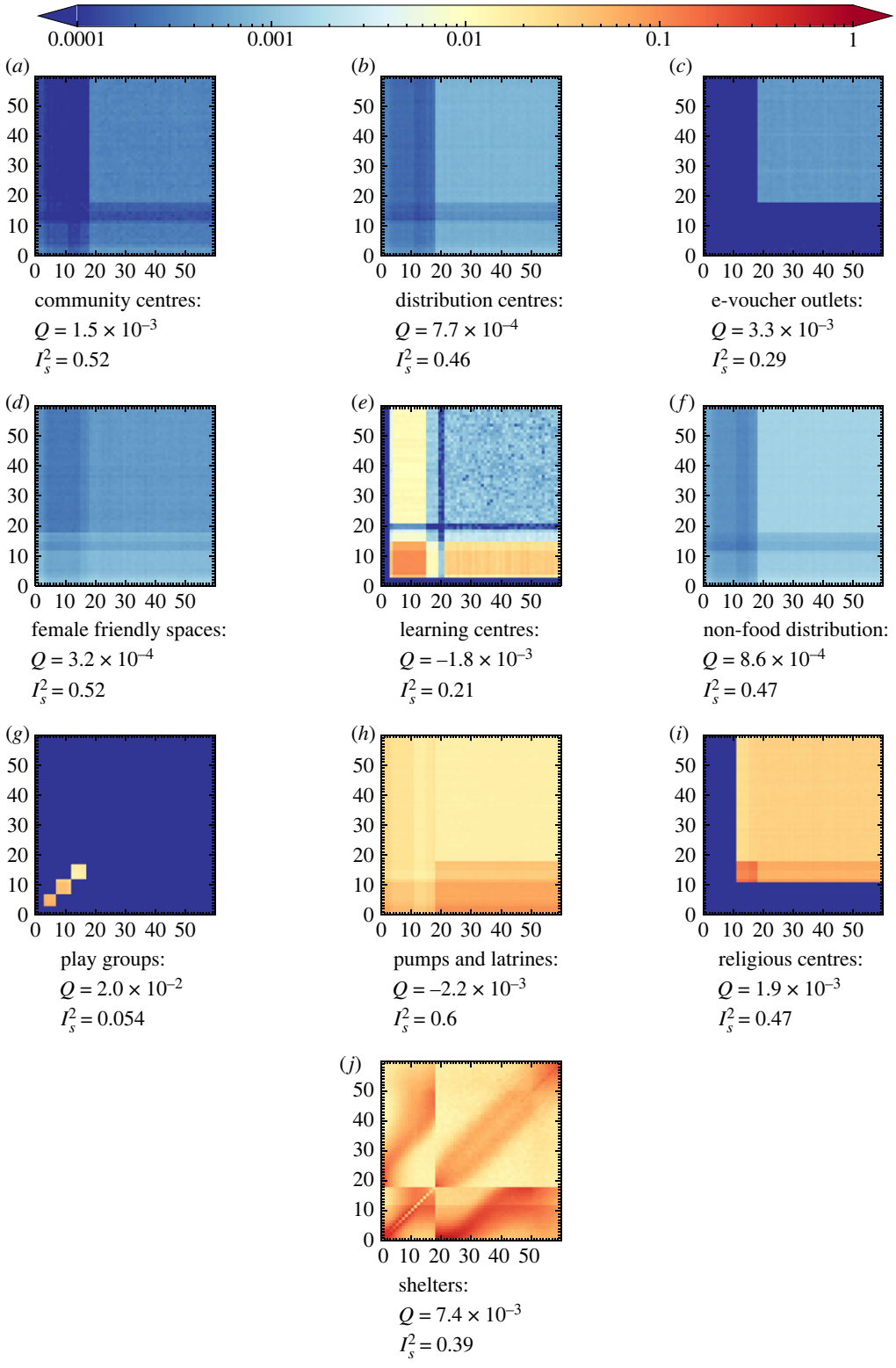

Figure 8. The population normalized contact matrices ($PNCM_R$) by age as simulated in June-Cox. Note that the data inputs in (g,j) stem from a previous survey.

permitted to attend the religious centres and hence we expect a cut-off in the contact matrices, whereas in many other venues we expect a gradual shift in behaviour as children move into adolescence and then adulthood. This can be a positive feature of the model—i.e. that the model represents the behavioural and movements patterns correctly and forces agents to make a choice between activities they perform as they would in real life—however, this relies on reasonable behavioural data, insights and assumptions. This is demonstrated most clearly in the *shelters* contact matrices in

Table 3. Contact matrix statistics calculated for June-UK and BBC Pandemic project reported for company and school mixing (cf. figure 14). These statistics are calculated for the UK demography reported by ONS in 2011 [37].

	company			school		
	Q	I_s^2	D_C	Q	I_s^2	D_C
BBC Pandemic	2.6×10^{-2}	0.31		0.13	0.14	
June-UK $PNCM_R$	2.6×10^{-3}	0.42	0.73	0.21	5.0×10^{-2}	0.63

which the household clustering places adults and children differently, based on fixed rules derived from survey and census data (see appendix E).

(ii) **Virtual venue demography:**
The dynamics of virtual spaces in the simulation are dictated by the probabilistic attendance rates (figure 3) and age cut-offs. The attendance rates are a function of age, sex, time and venue, which leads to different demographies across the virtual spaces and therefore different social mixing behaviours. Again, due to the nature of the simulation in which we have strict probabilistic rules which determine the attendance of different subgroups (children, adults, age or sex, etc.), we can obtain strong divisions between groupings. This is shown by the discontinuities in the heat-map representation of the contact matrices. In particular, only males over the age of 11 are permitted to attend the religious centres, leading to discontinuities in the religious centre contact matrices. In June-UK, there is no simulation of parent–teacher interactions at school that might occur during pick-up or drop-off times, and the virtual school setting is strictly modelling student–teacher interactions where any teacher–teacher interactions would be restricted to the classroom setting. Further, no children attend any workplace settings, and agents can only be employed or attend a workplace venue between the ages of 18 and 65. The contact matrices produced from June-UK therefore lack certain features shown by the BBC Pandemic project. However, this is a problem all such approaches that rely on an imperfect virtual representation of reality can experience.

(iii) **Virtual world rules and behaviour patterns:**
The combination of the above points leads to complex interconnected behaviours across the simulation. Considering the behaviour of coarser subgroups across all venues, we see more general behaviours emerge; for instance, children are less likely to attend any virtual venue than adults, and males are more likely to attend any venue than females due to the attendance at religious centres which increases the overall rate of males not staying in the shelters compared with females, leading to an asymmetry in the shelter contact matrix. An 11–18-year-old is more likely to see a 6–11-year-old than the converse. A 6–11-year-old is more likely to be home than a 11–18-year-old; therefore, on average, in any time-step a 6–11-year-old will not contact an 11–18-year-old in shelters, but when the 11–18-year-old is home they will probably contact the 6–11-year-old. The normalization of contacts by users (or population) and contact duration (as done throughout) makes this effect visible. There are other instances, such as the community centres, in which we see a banding effect which is an induced artefact from the movement criterion of the agents in the model (figure 3). The high attendance rate expected of 11+ males leads to a reduction in attendance of this group across all other venues, and many of the contact matrices show a banding effect between 11 and 18 due to this behaviour.

Given the level of detail contained within the model-derived contact matrices, they have the ability to reveal potential shortcomings in both the survey set-up as well as the modelling of the virtual world, as they reflect how sophisticated and well understood each venue type is. This means that the amount of resources needed to be expended on collecting more data on certain locations can be estimated in order to improve certain matrices. These can be traded off against the resources available and the relative expected gain from their expenditure. In this work, we validated our contact tracker in two very different models, June-UK and June-Cox. In the former, we demonstrated that the NCM_V and NCM_R agree well with data collected using traditional methods (cf. tables 2 and 3). In the latter, NCM_V-type contact patterns are not available, as our extracted contact matrices used coarse survey information on venue attendance to inform the simulation of contact patterns there, with the notable exception of the shelters, which are relatively precisely captured by the census data. Our mixed-method approach allows us to partially compensate for the gaps in detailed understanding of demographic structures at the lesser-known venues.

5. Conclusion

In this work, we demonstrate the complementary power of a lightweight contact survey, approximate details about venues and their attendance rates by different demographic groups, and an agent-based model to generate detailed social contact matrices. In the case of the Cox's Bazar refugee settlement, we use an existing model of the settlement developed using the June framework to perform a virtual contact survey, which is informed by the highly aggregated real-world survey, to produce more granular contact matrices which can be further interrogated. Our constructed contact matrices will provide an important input to future disease spread modelling or social dynamic studies in the settlement, and provide a baseline which can be translated to other settlements as well. These contact matrices could be further scrutinized by examining the basic reproduction number associated to them [38,39], thus providing a direct epidemiological comparison. However, this is beyond the scope of this particular work. Further, our method can easily be adapted to other settings for which detailed contact matrices are not available, thereby enabling the use of disease models in contexts where previously large assumptions would have had to have been made about contact patterns. Contact matrices form the backbone of many disease models, and so calculating them at a global scale, with the specific inclusion of those groups who are often most vulnerable to disease spread, is essential [3].

Ethics. The survey run as part of this study was approved by the ethics committee of Durham University, reference: PHYS-2020-09-04T10:24:22-gnvq71.

Data accessibility. June and June-UK: the current public release of the June simulation framework, and by extension the latest version of the June-UK model, can be found at https://github.com/IDAS-Durham/JUNE and have been archived within the Zenodo repository https://doi.org/10.5281/zenodo.7199142 [40]. June-Cox: the current public release of June-Cox epidemic model can be found at https://github.com/UNGlobalPulse/UNGP-settlement-modelling and have been archived within the Zenodo repository https://doi.org/10.5281/zenodo.10120763 [41]. Data: the data from the survey is available by application at https://microdata.unhcr.org/index.php/catalog/587. Contact Survey: details and calculation at https://github.com/UNGlobalPulse/UNGP-contact-survey. June-UK household contact matrix: details and calculation at https://github.com/IDAS-Durham/june_household_matrix_calculation and have been archived within the Zenodo repository https://doi.org/10.5281/zenodo.10091526 [42]. Contact matrix results: our contact matrices are reported at https://github.com/IDAS-Durham/june_mixed_method_CM_results formatted in excel documents for convenience and have been archived within the Zenodo repository https://doi.org/10.5281/zenodo.10091522 [43].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. J.W.: data curation, formal analysis, investigation, methodology, validation, visualization, writing—original draft, writing—review and editing; J.A.-B: conceptualization, data curation, funding acquisition, investigation, methodology, project administration, resources, software, supervision, writing—original draft, writing—review and editing; D.S.: investigation, methodology; A.G.K.M.: interpretation of results; E.S.E.: writing—original draft; S.H.: interpretation of results; F.K.: conceptualization, data curation, funding acquisition, investigation, methodology, project administration, resources, software, supervision, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. United Nations Global Pulse work is supported by the Governments of Sweden and Canada, and the William and Flora Hewlett Foundation. J.A.-B., D.S. and J.W. were supported by the Centre for Doctoral Training in Data Intensive Science under UKRI-STFC grant no. ST/P006744/1 for parts of this work. D.S. is also funded by STFC through a Data Innovation Fellowship (ST/R005516/1). J.W. and D.S. received financial support through the EPSRC IAA project 'Creating humanitarian impact through data modelling: Collaborating with WHO and UN Global Pulse'. F.K. gratefully acknowledges funding as Royal Society Wolfson Research fellow. This work used the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The equipment was funded by BEIS capital funding via STFC capital grant nos. ST/K00042X/1, ST/P002293/1, ST/R002371/1 and ST/S002502/1, Durham University and STFC operations grant no. ST/R000832/1. DiRAC is part of the National e-Infrastructure.

Acknowledgements. We would like to thank the UNHCR Cox's Bazar teams for their helpful comments on this work and for the support in setting up and running the survey in the settlement. In particular, we would like to thank Hussien Ahmad, Hosna Ara Begum, Mahfuzur Rahman, and all the members of the Information Management, Community Based Protection and Public Health teams. The authors would also like to thank Giulia Zarpellon and Miguel Luengo-Oroz for their helpful comments and suggestions throughout this work.

Disclaimer. The authors alone are responsible for the views expressed in this article, and they do not necessarily represent the views, decisions or policies of the institutions with which they are affiliated, including the United Nations.

Appendix A. Contact matrices

Here we present the remaining contact matrices derived from June-Cox.[5]

[5]We interpret contact matrix Δ_{ij} such that person i contacts person j and graphically as subgroup on x-axis contacts subgroup on y-axis.

A.1. UNCM$_V$ Interaction

See figure 9.

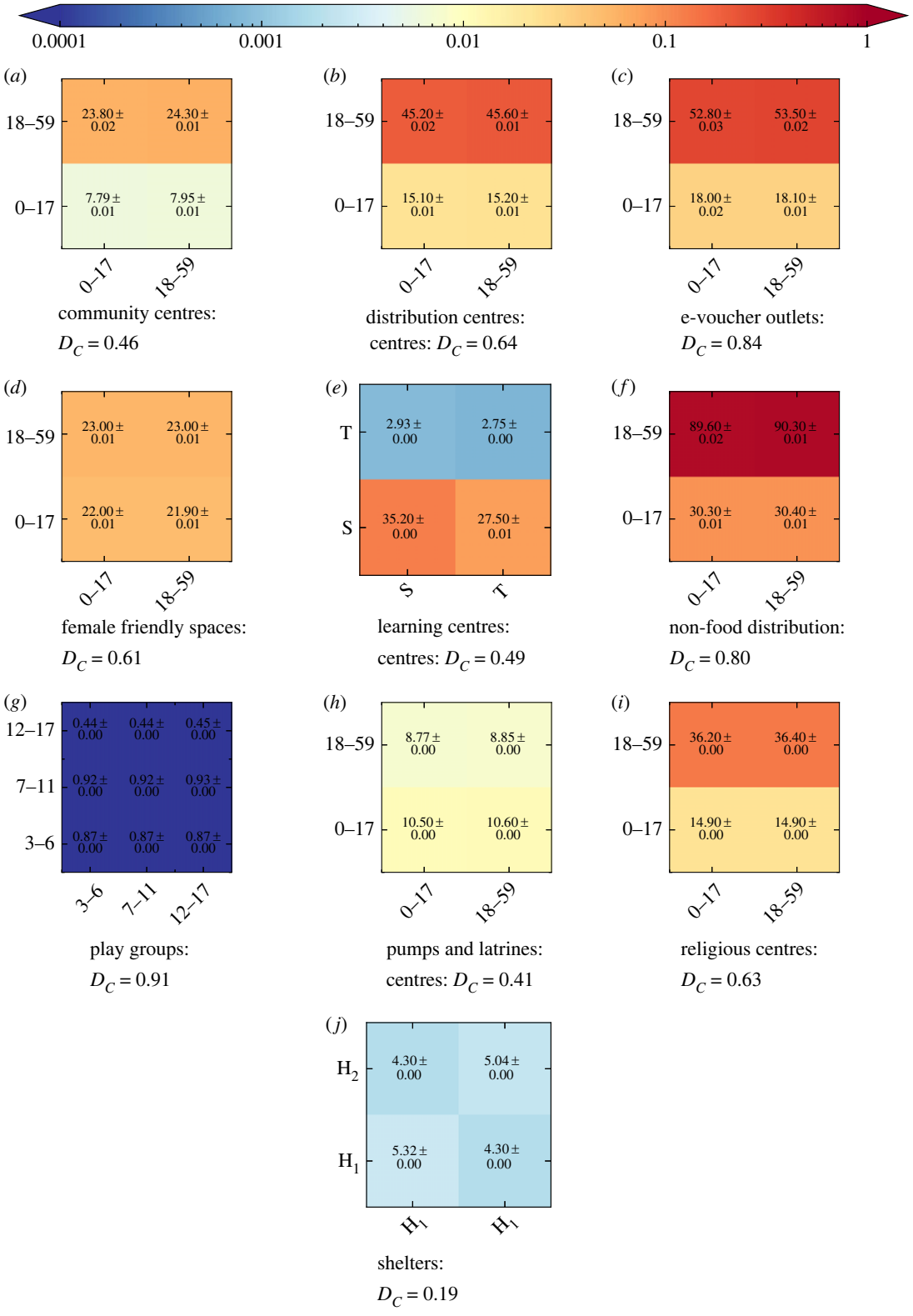

Figure 9. The normalized venue contact matrices (UNCM$_V$) by input survey subgroups as simulated in June-Cox. Note that the data inputs in (*g*,*j*) stem from a previous survey.

A.2. UNCM

See figure 10.

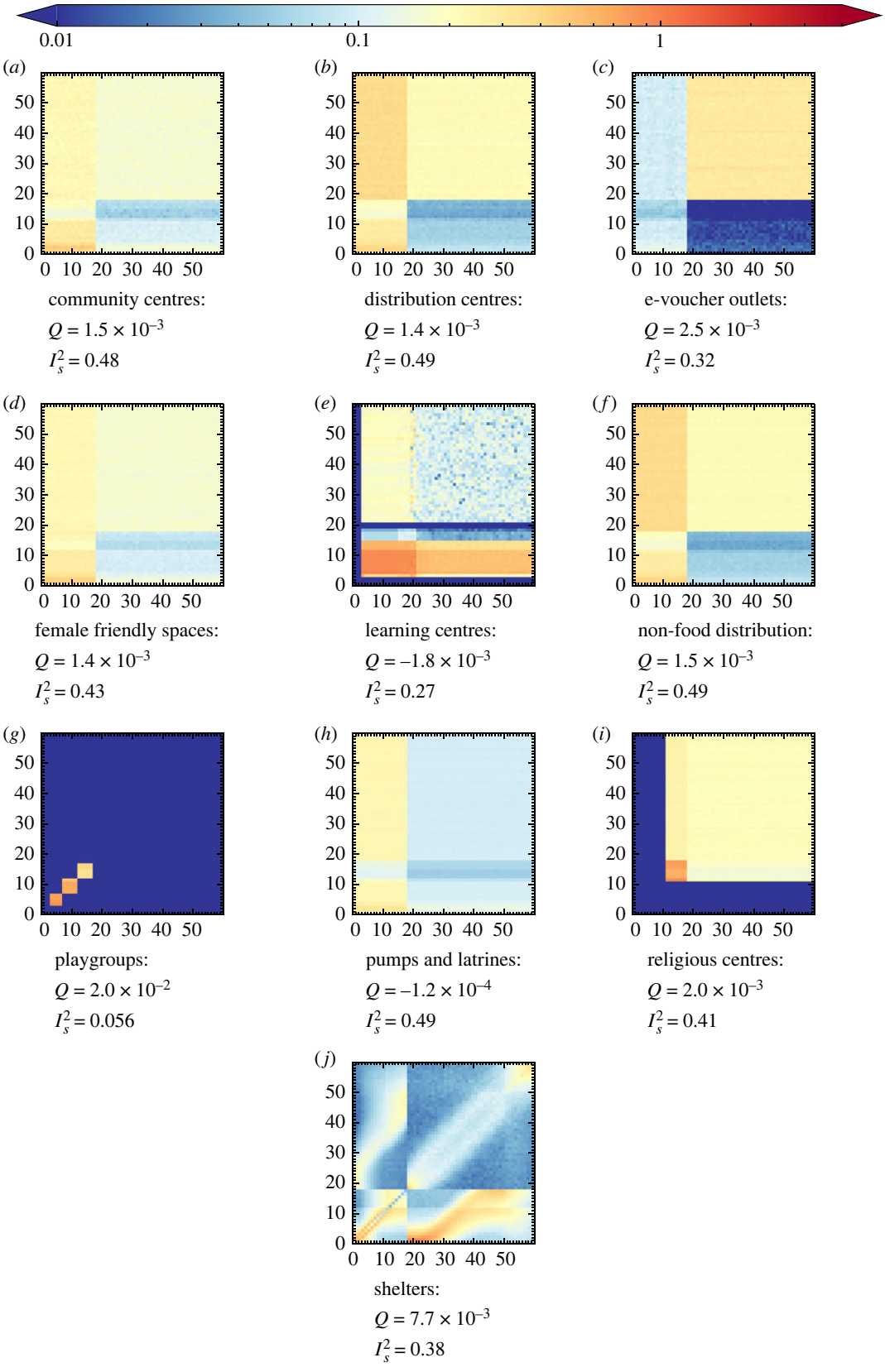

Figure 10. The normalized contact matrices (UNCM) by age as simulated in June-Cox. Note that the data inputs in (*g*,*j*) stem from a previous survey.

A.3. UNCM$_V$

See figure 11.

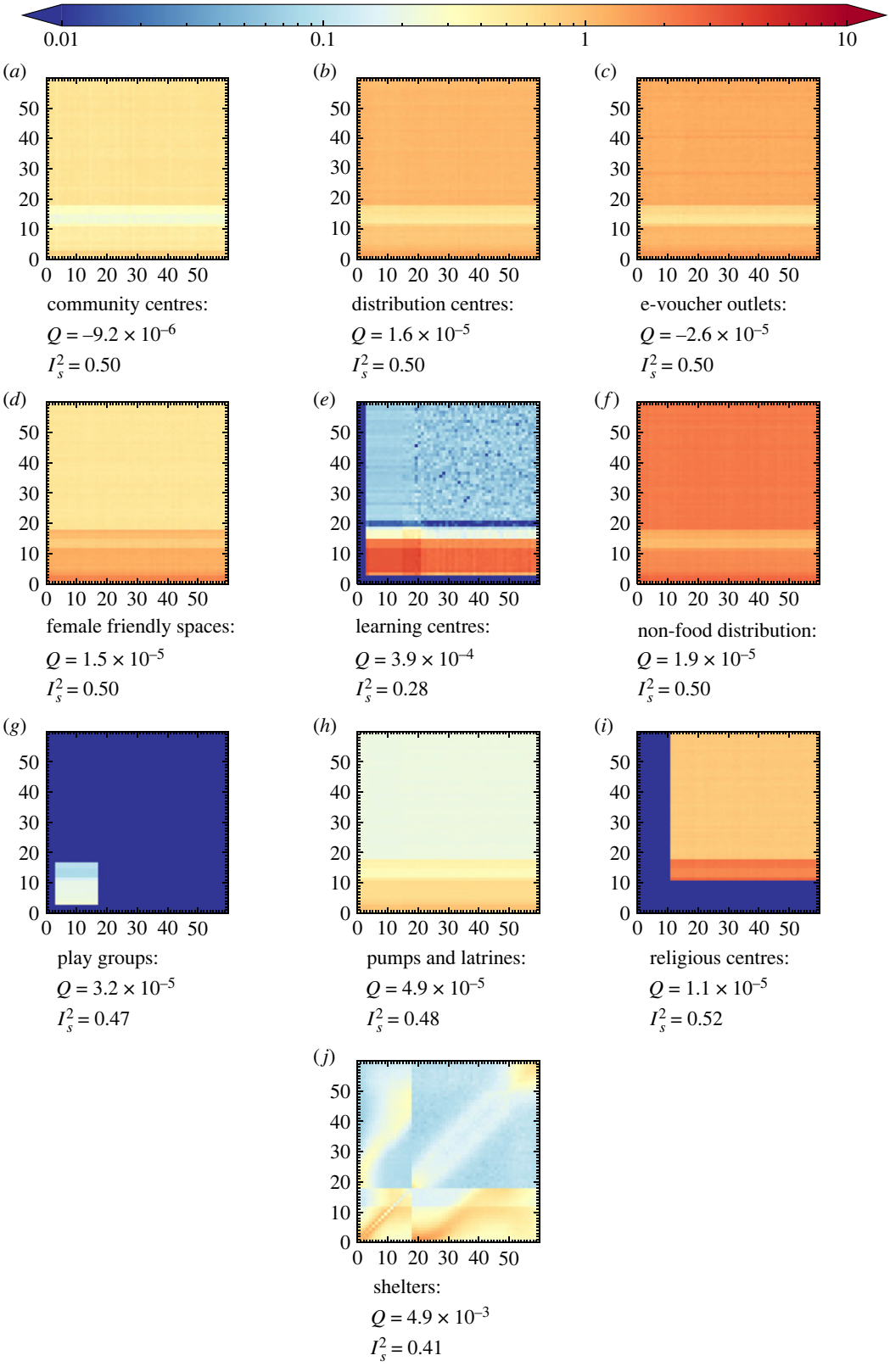

Figure 11. The normalized venue contact matrices (UNCM$_V$) by age as simulated in June-Cox. Note that the data inputs in (*g*,*j*) stem from a previous survey.

A.4. PNCM

See figure 12.

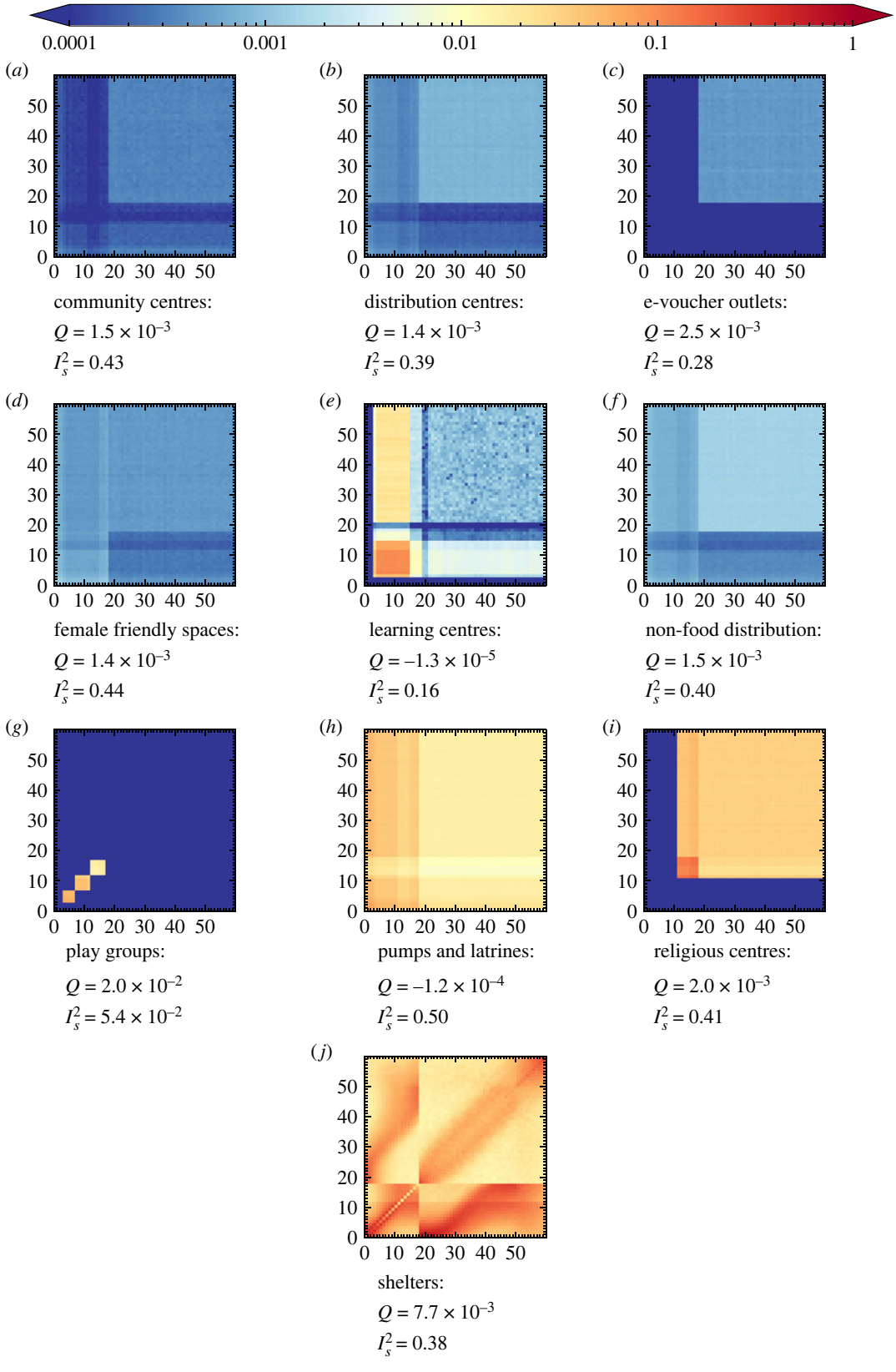

Figure 12. The normalized venue contact matrices (PNCM) by age as simulated in June-Cox. Note that the data inputs in (*g*,*j*) stem from a previous survey.

A.5. PNCM$_V$

See figure 13.

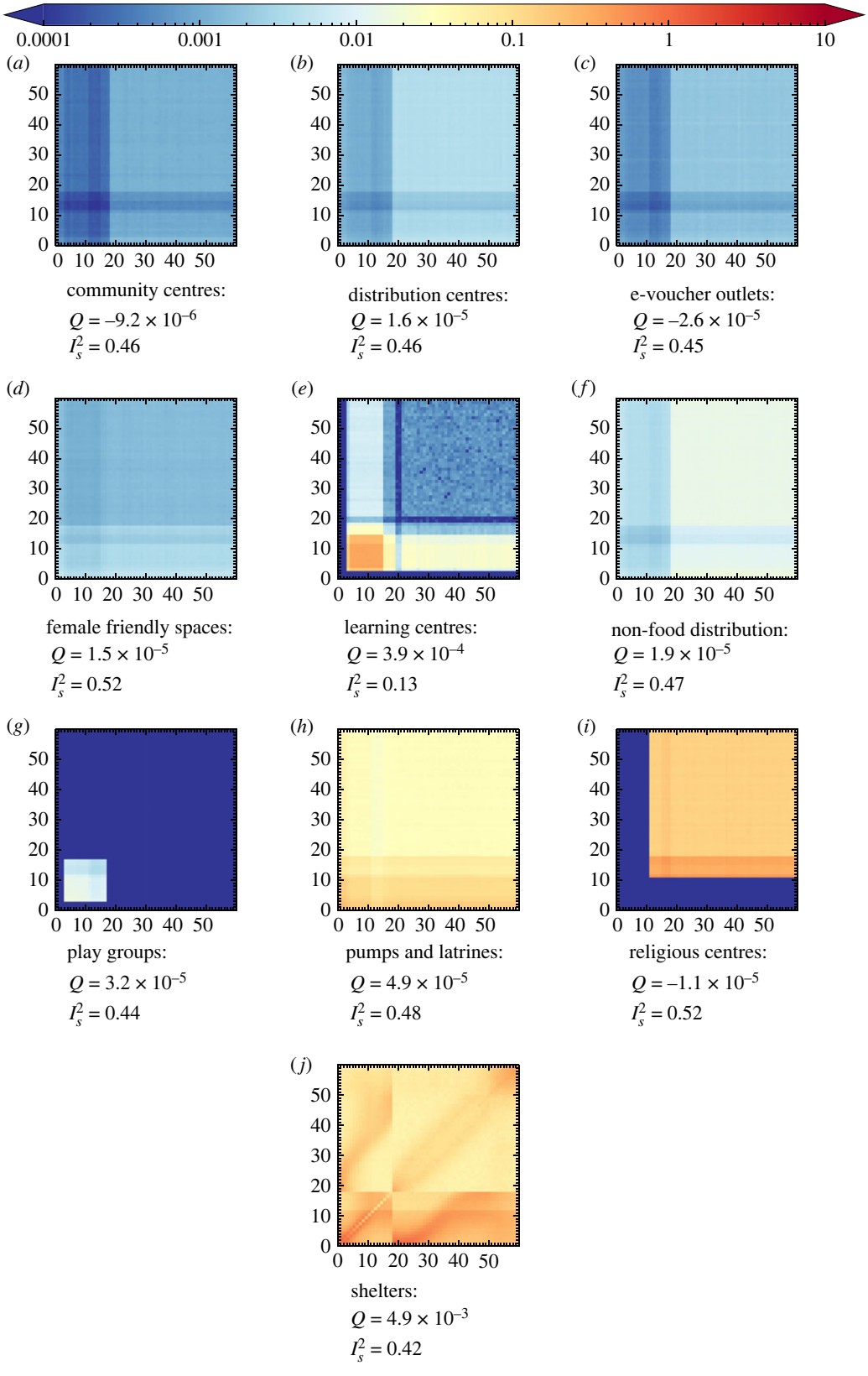

Figure 13. The normalized venue contact matrices (PNCM$_V$) by age as simulated in June-Cox. Note that the data inputs in (*g*,*j*) stem from a previous survey.

Appendix B. UK validation

See figures 14 and 15.

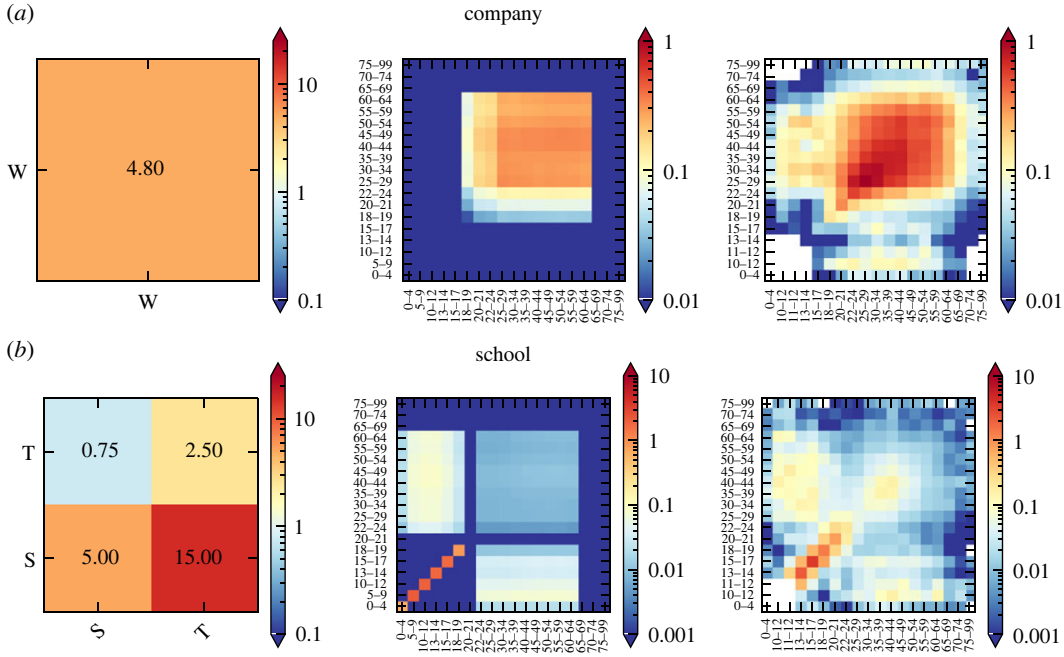

Figure 14. Contact matrices from the UK validation procedure. Left: the derived input interaction matrix, $\mathrm{UNCM_R}$ for 'companies' and 'schools', where the labels 'W' refers to 'workers', 'S' students, 'T', teachers. Centre: the simulated age-binned $\mathrm{PNCM_R}$ matrix with entries $\hat{C}_{ij}$ from June-UK. Right: the BBC Pandemic project 'all home' contact matrix, C, with entries c_{ij}. The simulated matrices and the survey matrices share the same colour map for ease of comparison. The input interaction matrix has its own so that the colour maps have suitable contrast over the full range of number of contacts.

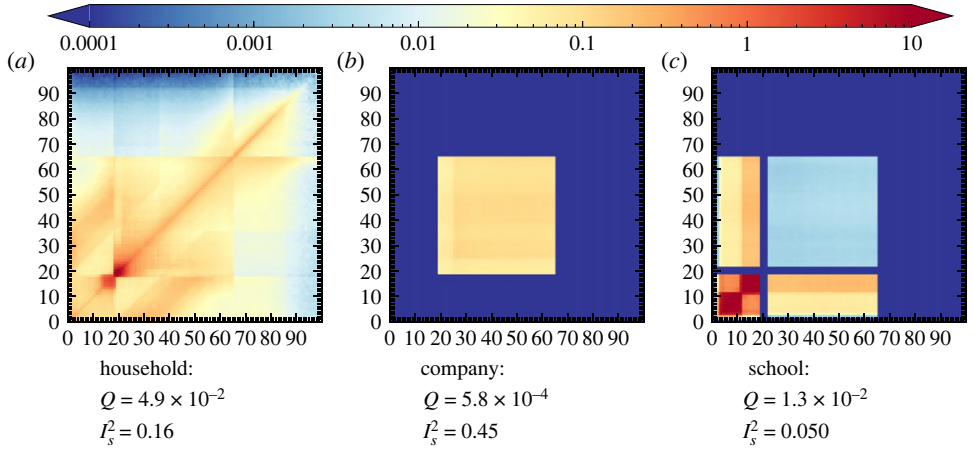

Figure 15. The normalized venue contact matrices ($\mathrm{PNCM_V}$) by age as simulated in June-UK.

Appendix C. Survey

The survey between October and November 2020 was conducted by enumerators from the UNHCR Community Based Protection team who regularly conduct surveys within the settlement following standard UNHCR practices [26,27]. Data were collected from 22 camps in the Kutapalong–Balukhali Expansion Site (part of the Cox's Bazar refugee settlement) consisting of two men and two women in each of the following categories: less than 18 years; greater than or equal to 18 years less than 60;

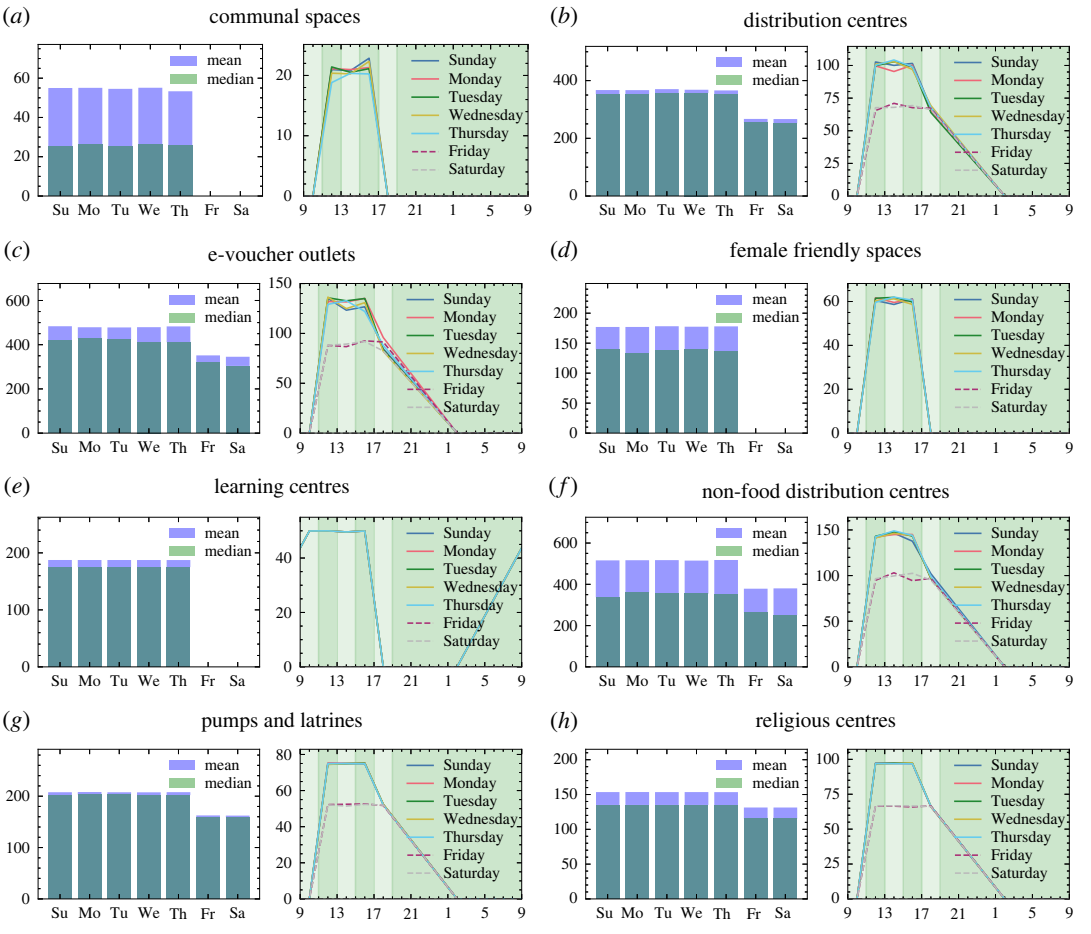

Figure 16. The unique person attendance rates per day (left) and by time of day (right) for the virtual venue. The green shading represents the discrete time-step bins of the simulation. June can in general distinguish between days of the week however in June-Cox we only model differences between weekend and weekdays.

greater than or equal to 60 years. In addition, two persons with disabilities were surveyed to make a total of 308 respondents. Anonymized results, and additional metadata can be accessed through UNHCR [28].

The survey was conducted by enumerators randomly sampling households in each camp and visiting them in person. Only one respondent per household was permitted and responses were collected using the Kobo Toolbox [44] based on the Open Data Kit [45]. The survey was formatted as follows (italicized text is spoken):

This questionnaire has been designed by teams from United Nations Global Pulse and UNHCR and is to inform efforts to better understand how people move around in the camp and interact with others to better understand how COVID-19 might spread in the camp to inform future COVID-19 protection measures.

Good day my name is __ from UNHCR and I am here to conduct a survey. This study is part of a scientific research project from United Nations Global Pulse and UNHCR. In this study, we will ask questions to better understand how people move around in the camp and interact with others. Your decision to complete this study is completely voluntary, and you may decline to answer at any time. Your answers will be completely anonymous. The results of the research may be presented at scientific meetings or published in scientific journals. For any questions or comments please contact: ___. The survey should not take longer than 30 minutes.

(i) — **If adult:** *Do you declare that you are at least 18 years of age and that you agree to complete this survey voluntarily?*

— **If child:**

(a) **To parent or guardian:** *Do you declare that you are at least 18 years of age, that you are the parent or guardian of this child and that you give consent for your child to complete this survey voluntarily?*

(b) **To child:** *Do you declare that this is your parent or guardian and that you give consent to complete this survey voluntarily?*

(ii) *Sex:* Female, Male, Other, Do not want to answer
(iii) *Location at the time:* __ (camp)
(iv) *Age:* under 18, over 18 but under 60, over 60
(v) *Disability:* Y/N
(vi) *Do you have access to a face mask?* Y/N
(vii) *When the learning centres were open, did you attend any formal education?* Y/N
(viii) — **If yes:**
  (a) *When you attended formal education, how much time do you spend there?* 30 min, 1 h, 1 h and 30 min, 2 h, other (please specify)
  (b) *When you attended formal education, approximately how many children do you come into contact with (for example, talk to)?*
  (c) *When you attended formal education, approximately how many adults do you come into contact with (for example, talk to)?*
(ix) *Do you ever go to a food distribution centre?* Y/N
(x) — **If yes:**
  (a) *When you go to a food distribution centre, how much time do you spend there?* 30 min, 1 h, 1 h and 30 min, 2 h, other (please specify)
  (b) *When you go to a food distribution centre, approximately how many children do you come into contact with at the centre (for example, talk to)?*
  (c) *When you go to a food distribution centre, approximately how many adults do you come into contact with at the centre (for example, talk to)?*
  (d) *When you go to the food distribution centre, do you wear a mask in the centre?*
(xi) *Do you ever go to an e-voucher outlet?* Y/N
(xii) — **If yes:**
  (a) *When you go to an e-voucher outlet, how much time do you spend there?* 30 min, 1 h, 1 h and 30 min, 2 h, other (please specify)
  (b) *When you go to an e-voucher outlet, approximately how many children do you come into contact with at the outlet (for example, talk to)?*
  (c) *When you go to an e-voucher outlet, approximately how many adults do you come into contact with at the outlet (for example, talk to)?*
  (d) *When you go to an e-voucher outlet, do you wear a mask in the outlet?*
(xiii) *Do you ever go to a community centre?* Y/N
(xiv) — **If yes:**
  (a) *When you go to a community centre, how much time do you spend there?* 30 min, 1 h, 1 h and 30 min, 2 h, other (please specify)
  (b) *When you go to a community centre, approximately how many children do you come into contact with at the centre (for example, talk to)?*
  (c) *When you go to a community centre, approximately how many adults do you come into contact with at the centre (for example, talk to)?*
  (d) *When you go to a community centre, do you wear a mask in the centre?*
(xv) *Do you ever go to a religious meeting?* Y/N
(xvi) — **If yes:**
  (a) *When you go to a religious meeting, how much time do you spend there?* 30 min, 1 h, 1 h and 30 min, 2 h, other (please specify)
  (b) *When you go to a religious meeting, approximately how many children do you come into contact with at the meeting (for example, talk to)?*
  (c) *When you go to a religious meeting, approximately how many adults do you come into contact with at the meeting (for example, talk to)?*
  (d) *When you go to a religious meeting, do you wear a mask in the meeting?*
(xvii) (a) *When you go to a water pump or latrine, how much time do you spend there?* 30 min, 1 h, 1 h and 30 min, 2 h, other (please specify)
  (b) *When you go to a water pump or latrine, approximately how many children do you come into contact with (for example, talk to)?*
  (c) *When you go to a water pump or latrine, approximately how many adults do you come into contact with (for example, talk to)?*
  (d) *When you go to a hand pump or latrine, do you wear a mask?*

Appendix D. Questions for the Community Based Protection team

To supplement our analysis, a series of informal interviews were conducted with members of the Cox's Bazar refugee settlement UNHCR Community Based Protection (CBP) team. In each of these interviews, a set of general enquiries into the behaviour and attendance rates were asked of members of the protection team which worked closely with those venue types.

This questionnaire has been designed by teams from United Nations Global Pulse and UNHCR and is to inform efforts to better understand how people engage with each venue in the camp and the demography of the venues.

For the following venues: Community centres, Female friendly spaces, Food distribution centres, E-voucher outlets, Non-food distribution centres—including LPG and blanket centres—Religious centres, and Learning centres. Where you are able and suitably informed please could you answer the following questions;

(i) — *Can you describe what a day looks like at* **venue***?*
   (a) *How many people do you expect at minimum and peak times?*
   (b) *How do these days and numbers of people vary by day, week, month/season?*
   (c) *Why do you think there are these variations?*

(ii) — *What is the makeup of multigenerational households—are there generally three generations or more?*
   (a) *Do these households include extended family?*
   (b) *Is this a cultural issue or a space constraint?*

(iii) — *What age do children typically move through the camp independently?*
   (a) *Move out from parents shelter?*
   (b) *Go to* **venues** *on their own? (e.g. collect items from the distribution centres for their shelter)*
   (c) *How many hours do they spend moving around in the camp independently?*
   (d) *Who do they mostly have contact with when they move around? (e.g. more children, teachers at school, other adults? all)*

(iv) — *What time do* **venues** *close?*

Here we outline the key findings from the interviews used to define the virtual venues in June-Cox.

— Community Centres and Female friendly spaces:
   (a) *Busiest in morning.*
   (b) *Typically 35–40 people per day.*
   (c) *Closed Friday and Saturday.*
   (d) *Less busy during rainy season and religious holidays but more busy in national holidays.*
— Distribution Centres (Food and Non-Food):
   (a) *Typically 300–400 people per day.*
   (b) *Sunday busiest day 500 people per day.*
   (c) *Families permitted to collect food every two weeks.*
   (d) *Children not permitted on their own.*
— E-voucher:
   (a) *Same as Distribution Centres.*
   (b) *No limit on frequency of attendance.*
— Learning Centres:
   (a) *Children (4–13) attend in morning or afternoon slots of 2–3 h.*
   (b) *Typically 80–100 children per group.*
   (c) *Closed Friday and Saturday*
   (d) *Less busy during rainy season.*
— Religious Centres:
   (a) *Attended only by men 13+ in age.*
   (b) *Five daily prayers throughout the day.*
   (c) *Typically 150–200 people per day.*
   (d) *Friday busiest day 200–300 people.*
   (e) *Less busy during rainy season but more busy in religious holidays 600–700.*

The attendance probabilities (figure 3) were tuned to achieve the desired attendance rates. These rates were chosen such that they represent an 'average' day of any particular day of the week in the camp ignoring any changes of behaviour from religious or national events or annual variations in climate and weather.

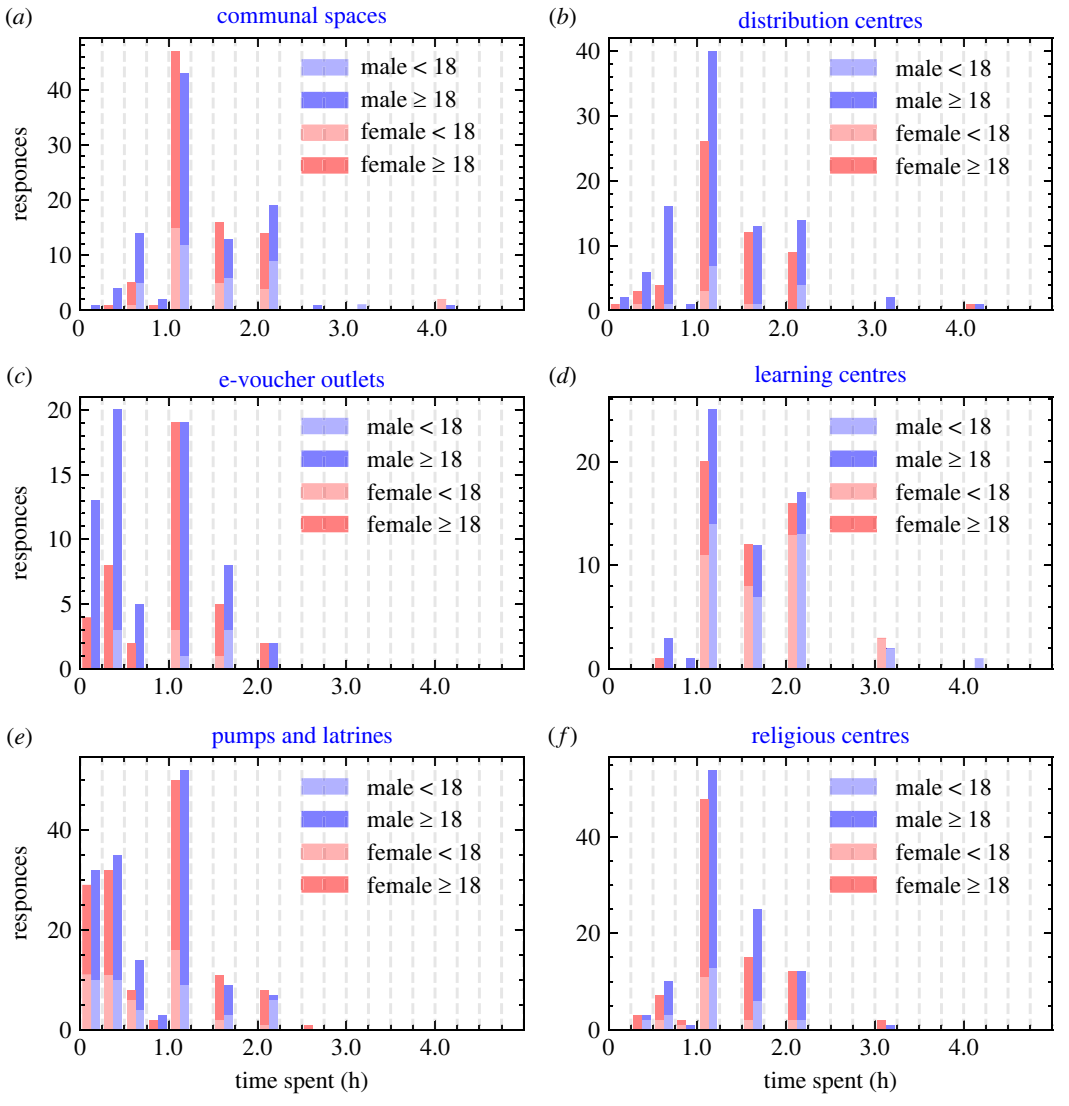

Figure 17. Quarterly hour binned histograms of the lightweight survey responses. The total response count of men is shown in blue, women in red and the proportion of adults to children is represented by the higher and lower portion of each bar respectively. The survey equates 'community centres' and 'distribution centres' to 'female friendly spaces' and 'non-food distribution centres', respectively.

The characteristic times spend at each venue T^L are extracted from the lightweight survey (cf. figure 17) and are reported below (table 4),

Appendix E. Demographic properties

Households are constructed stochastically by clustering individuals into households according to their age, sex and the following reported properties of the camp in order to create realistic demographic household structures:

— Macroscopic properties:
  (i) the distribution of household sizes in the camp (known at the region level);
  (ii) population demographics;
  (iii) the proportion of one-, two- and multi-generational households.
— Microscopic properties:
  (i) the probability of single parent;
  (ii) the mean spousal age gap;
  (iii) the mean age of mother at birth of first child.

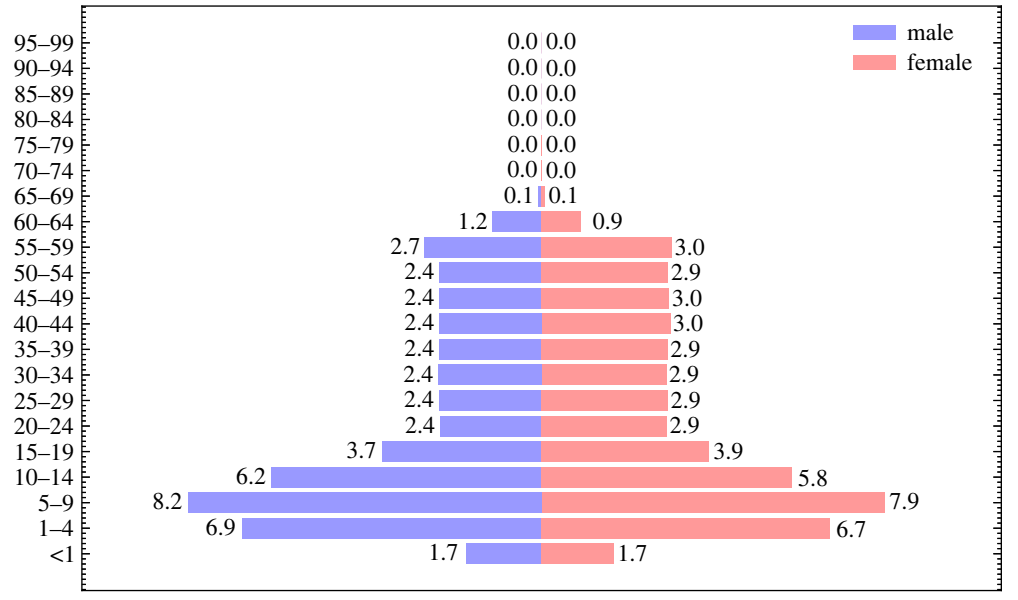

Figure 18. The population pyramid of Cox's Bazar across all camps [46,47]. Male population is shown on the left in blue and female on the right in red. The percentage of the population of each category is quoted.

Table 4. The characteristic time, T^L in hours for each venue reported in the survey. Value and error are determined using proportionate weighting between men and women using a median bootstrap method.

venue, V	characteristic time, T^L (h)
community centres	1.00 ± 0.02
distribution centres	1.00 ± 0.05
e-voucher outlets	0.8 ± 0.1
female friendly spaces	1.00 ± 0.02
learning centres	1.5 ± 0.1
non-food distribution centres	1.00 ± 0.05
pump and latrines	0.7 ± 0.1
religious centres	1.00 ± 0.01

These properties are all known at the super-area level unless specified otherwise. The population of Cox's Bazar is densely packed and particularly young (cf. figure 18), so the majority of households contain children, meaning a significant portion of multi-generational households exist in the camp. Therefore, as children can be key drivers in disease spread in the camp [5] it is important that the household structures in the model are representative of reality.

The resulting household demographic structures can be seen in figure 19 and shelter sizes in figure 20. The age brackets for each demographic are inferred from survey and data from the settlement. Children (0–18); 18 is the age at which marriage is legal for women (21 for men), Adults (18–49) (49 being chosen to provide a realistic age gap for potential grandparents, twice the average mother–child age gap, 22.43 years plus the average spousal age gap, 4.73 years). (49–100) for old adults, the remaining ages in the camp. June-Cox has an over-clustering of children with single parent housing, this due to any remaining children being randomly clustered into households with adults after the children with couples houses are constructed. The microscopic properties of the clustered households are summarized between figures 19 and 21.

Agent household clustering adapted from June-Cox [5] as outlined in appendix S1. The number of households are predetermined at the area level throughout the camp; however, the household demographic statistics are reported at the region level and the household size at the camp-wide level. Therefore, we have to assume that the camp to region to area statistics can be applied from large geographical areas to smaller ones where the information is unavailable without loss of generality.

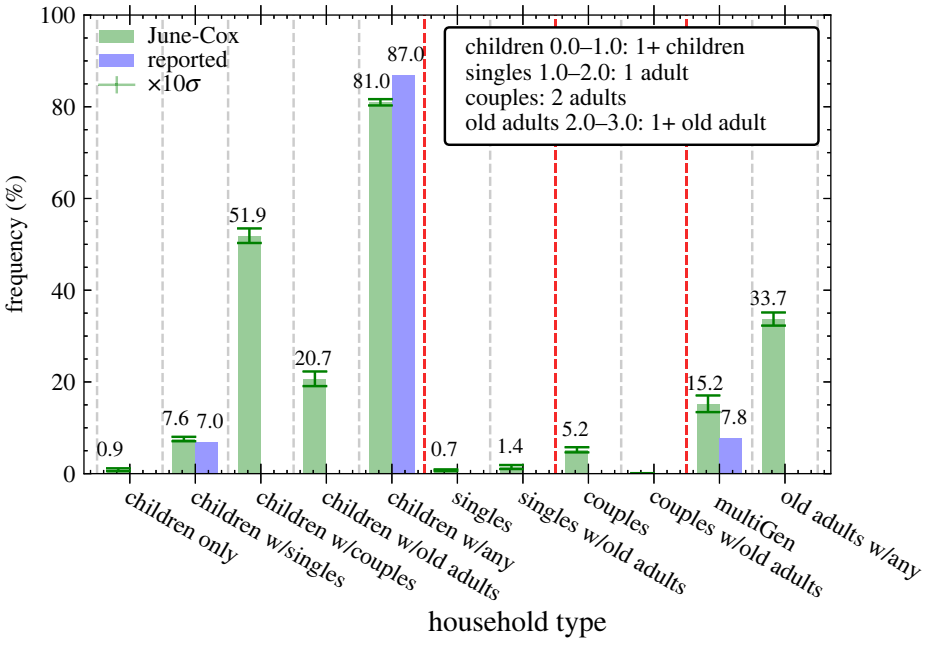

Figure 19. Figure of proportion of household types. Note that not all of these groups are mutually exclusive. Green represents the reconstruction in June and blue the reported data (if available). Those groups where data was unavailable are reported in the figure for completeness. June-Cox data reported with scaled error bars of the standard deviation over 20 independent household clustering.

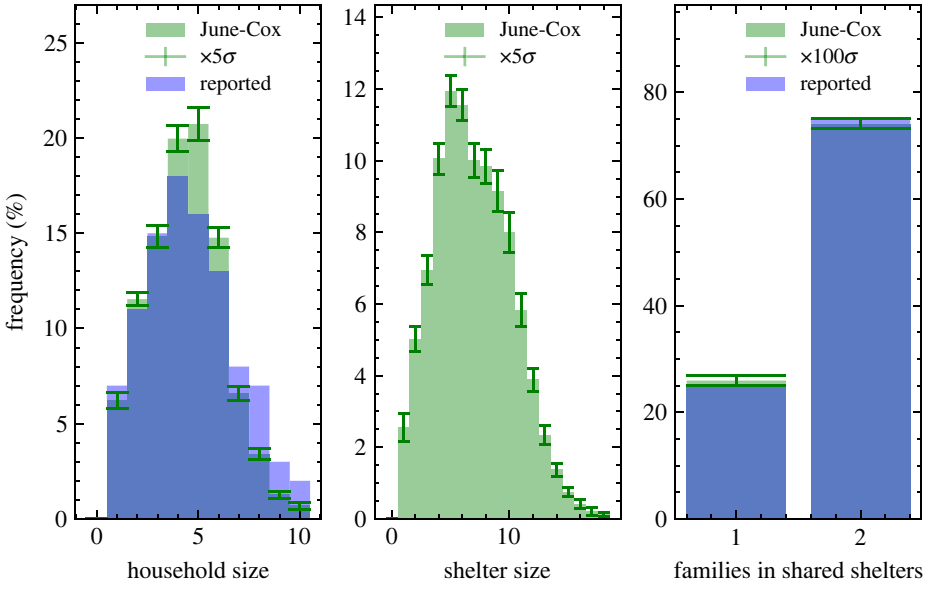

Figure 20. Figure of key shelter properties. Left: distribution of household family sizes. Middle: distribution of shelter sizes. Right: proportion of one- and two-household shelters. Green bars represent the reconstruction in June and blue bars are the reported data (if available). June-Cox data reported with scaled error bars of the standard deviation over 20 independent household clustering.

(a) We first create empty houses with sizes fitting the distribution shown in figure 20.
(b) We partition the houses in two groups containing subgroups of demographic properties:
    (i) households with children:
        — single parent,
        — multi-generational,
        — two parents;
    (ii) households without children.
(c) For all households, allocate one young adult (17–49) to each household, if available.

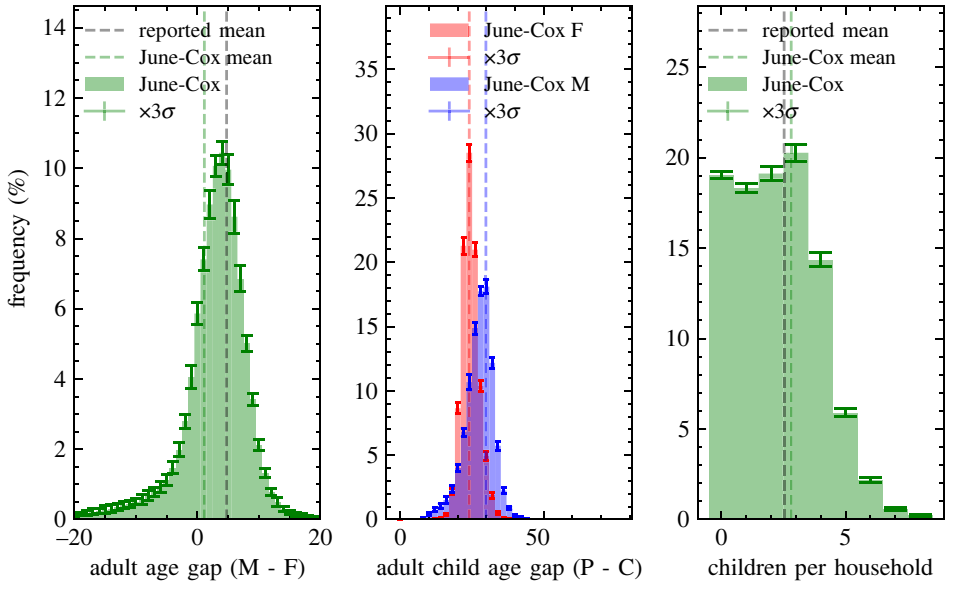

Figure 21. Figure of microscopic household properties. Left: distribution of household male female age gap in houses containing only two adults in the range 24–49. Middle: distribution adult child age gaps in households containing one adult (24–49) and the eldest child (0–18). Right: distribution of number of children by household. Green represents the reconstruction in June-Cox and black dashed lines the reported mean data (if available). June-Cox data reported with scaled error bars of the standard deviation over 20 independent household clustering.

(d) For single parent households, allocate one young adult (17–49) with an age gap drawn from the adult age gap distribution (figure 21) of opposite sex to current resident, if available.
(e) For all households, allocate all children into households randomly with children enforcing a 1 year age gap and adult child age gap drawn from distribution (figure 21), if available.
(f) Allocate older adults (50+) to multi-generational households with an adult age gap or adult child age gap drawn from distributions (figure 21) for opposite sex old adult current resident (grandparents age gap) and young adult current resident (parent—grandparent age gap), if available.
(g) Any remaining adults (17+) are allocated randomly to any remaining households with space under the following priority:

    (i) multi-generational households,
    (ii) households without children,
    (iii) households with children.

This methodology ensures that June-Cox households reflect the household demographics with a key emphasis on ensuring multi-generational households are represented proportionately.

Appendix F. Algorithm for the virtual survey

Algorithm 1: The virtual survey. Loop over all venues and people and simulate P_{contacts} between i and j subgroups from survey. The contacts can then be clustered into arbitrary subgroups k, l. We allow for multiple contacts between the same people at venue L.

Data: $\hat{t}_{ij}^{L} = [0]_{kl}$
$\hat{t}_{kl}^{L}$, number of contacts between subgroup k and l at venue L,
$\hat{\eta}_{i}^{L}$, population of subgroup i at venue L,
P^{L}, list of agents at venue L,
T^{L}, total time,
ΔT, simulation time step,
$\tilde{\gamma}_{kl}^{L}$, Stochastic Poisson sampled contacts between subgroup k and l at venue L,
for $L \in$ Venues **do**
  for $P_x \in$ People @ L, P^{L} **do**
    $i = \text{subgroup}(P_x)$
    $T^{L} = T^{L} + \Delta T$
    $\hat{\eta}_{i}^{L} = \hat{\eta}_{i}^{L} + 1$ **for** $j \in L_{\text{subgroups}}$ **do**
      Generate $\tilde{\gamma}_{ij}^{L}$ **if** $\tilde{\gamma}_{ij}^{L} = 0$ **then**
        continue
      else
        Generate randomly a list of P_{contacts} of $\tilde{\gamma}_{ij}^{L}$ people at L in subgroup j not including P_x
      end
    end
    for $P_c \in P_{\text{contacts}}$ **do**
      $k = \text{subgroup}(P_x)$
      $l = \text{subgroup}(P_c)$
      $\hat{t}_{kl}^{L} = \hat{t}_{kl}^{L} + 1$
    end
  end
end
$\hat{t}_{kl}^{L} = \hat{t}_{kl}^{L} / T^{L}$

References

1. Fine PEM, Clarkson JA. 1982 Measles in England and Wales—I: an analysis of factors underlying seasonal patterns. *Int. J. Epidemiol.* **11**, 5–14. (doi:10.1093/ije/11.1.5)
2. Anderson RM, May RM. 1985 Age-related changes in the rate of disease transmission: implications for the design of vaccination programmes. *Epidemiol. Infect.* **94**, 365–436. (doi:10.1017/S002217240006160X)
3. Aylett-Bullock J *et al.* 2022 Epidemiological modelling in refugee and internally displaced people settlements: challenges and ways forward. *BMJ Glob. Health* **7**, e007822. (doi:10.1136/bmjgh-2021-007822)
4. Altare C, Kahi V, Ngwa M, Goldsmith A, Hering H, Burton A, Spiegel P. 2019 Infectious disease epidemics in refugee camps: a retrospective analysis of UNHCR data (2009–2017). *J. Glob. Health Rep.* **3**, e2019064. (doi:10.29392/joghr.3.e2019064)
5. Aylett-Bullock J *et al.* 2021 Operational response simulation tool for epidemics within refugee and IDP settlements: a scenario-based case study of the Cox's Bazar settlement. *PLoS Comput. Biol.* **17**, e1009360. (doi:10.1371/journal.pcbi.1009360)
6. Aylett-Bullock J *et al.* 2021 June: open-source individual-based epidemiology simulation. *R. Soc. Open Sci.* **8**, 210506. (doi:10.1098/rsos.210506)
7. Mossong J *et al.* 2008 Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med.* **5**, e74. (doi:10.1371/journal.pmed.0050074)
8. Fumanelli L, Ajelli M, Manfredi P, Vespignani A, Merler S. 2012 Inferring the structure of social contacts from demographic data in the analysis of infectious diseases spread. *PLoS Comput. Biol.* **8**, e1002673. (doi:10.1371/journal.pcbi.1002673)
9. Del Valle SY, Hyman JM, Hethcote HW, Eubank SG. 2007 Mixing patterns between age groups in social networks. *Soc. Netw.* **29**, 539–554. (doi:10.1016/j.socnet.2007.04.005)
10. Hoang T, Coletti P, Melegaro A, Wallinga J, Grijalva CG, Edmunds JW, Beutels P, Hens N. 2019 A systematic review of social contact surveys to inform transmission models of close-contact infections. *Epidemiology* **30**, 723–736. (doi:10.1097/EDE.0000000000001047)
11. van Zandvoort K *et al.* 2022 Social contacts and other risk factors for respiratory infections among internally displaced people in Somaliland. *Epidemics* **41**, 100625. (doi:10.1016/j.epidem.2022.100625)
12. Prem K, Cook AR, Jit M. 2017 Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLoS Comput. Biol.* **13**, e1005697. (doi:10.1371/journal.pcbi.1005697)
13. Prem K, van Zandvoort K, Klepac P, Eggo RM, Davies NG, Cook AR, Jit M. Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group. 2021 Projecting contact matrices in 177 geographical regions: an update and comparison with empirical data for the COVID-19 era. *PLoS Comput. Biol.* **17**, e1009098. (doi:10.1371/journal.pcbi.1009098)

14. Mistry D *et al.* 2021 Inferring high-resolution human mixing patterns for disease modeling. *Nat. Commun.* **12**, 323. (doi:10.1038/s41467-020-20544-y)
15. Xia S, Liu J, Cheung W. 2013 Identifying the relative priorities of subpopulations for containing infectious disease spread. *PLoS ONE* **8**, e65271. (doi:10.1371/journal.pone.0065271)
16. Zagheni E, Billari FC, Manfredi P, Melegaro A, Mossong J, Edmunds WJ. 2008 Using time-use data to parameterize models for the spread of close-contact infectious diseases. *Am. J. Epidemiol.* **168**, 1082–1090. (doi:10.1093/aje/kwn220)
17. Iozzi F, Trusiano F, Chinazzi M, Billari FC, Zagheni E, Merler S, Ajelli M, Del Fava E, Manfredi P. 2010 Little Italy: an agent-based approach to the estimation of contact patterns-fitting predicted matrices to serological data. *PLoS Comput. Biol.* **6**, e1001021. (doi:10.1371/journal.pcbi.1001021)
18. Vernon I *et al.* 2022 Bayesian emulation and history matching of JUNE. *Phil. Trans. R. Soc. A* **380**, 20220039. (doi:10.1098/rsta.2022.0039)
19. UNHCR. 2019 UNHCR Bangladesh operational update, 1–31 December 2019. See https://reliefweb.int/report/bangladesh/unhcr-bangladesh-operational-update-1-31-december-2019.
20. UNHCR. 2018 Rohingya refugee emergency at a glance. See https://unhcr.maps.arcgis.com/apps/Cascade/index.html?appid=5fdca0f47f1a46498002f39894fcd26f.
21. ISCG. 2019 Bangladesh—2019 multi sectoral needs assessment #2. See https://data.humdata.org/dataset/bangladesh-coxsbazar-msna-october2019.
22. OCHA Bangladesh—subnational administrative boundaries. See https://data.humdata.org/dataset/cod-ab-bgd.
23. ISCG Bangladesh—Outline of camps of Rohingya refugees in Cox's Bazar. See https://data.humdata.org/dataset/outline-of-camps-sites-of-rohingya-refugees-in-cox-s-bazar-bangladesh.
24. WHO. 2020 Bangladesh—Rohingya crisis: early warning, alert and response system (EWARS). See https://www.who.int/bangladesh/emergencies/Rohingyacrisis/ewars.
25. Government of the People's Republic of Bangladesh Office of the Refugee Relief and Repatriation Commissioner. 2020 Rohingya refugee camp operations: essential programmes in light of COVID-19. See http://rrrc.gov.bd/sites/default/files/files/rrrc.portal.gov.bd/notices/c3aece34_0550_4b4d_b33c_e8864272ada9/2020-03-25-16-34-21d19f130456961e35a25dbd1e5ef780.pdf.
26. UNHCR. 2011 The 10-point plan, ch. 2: data collection and analysis. See https://www.unhcr.org/50a4c2b09.pdf.
27. UNHCR. 2018 Guidance on the protection of personal data of persons of concern. See https://www.refworld.org/docid/5b360f4d4.html.
28. UNHCR, WHO, UN Global Pulse, UN OCHA, Durham University. 2020 Bangladesh: COVID-19 exposure and protective measures. See https://microdata.unhcr.org/index.php/catalog/587.
29. Klepac P, Kucharski AJ, Conlan AJ, Kissler S, Tang ML, Fry H, Gog JR. 2020 Contacts in context: large-scale setting-specific social mixing matrices from the BBC Pandemic project. Technical report *medRxiv*. (doi:10.1101/2020.02.16.20023754)
30. Efron B, Tibshirani RJ. 1993 *An introduction to the bootstrap*. Monographs on Statistics and Applied Probability, vol. 57. Boca Raton, FL: Chapman & Hall/CRC.
31. Lance GN, Williams WT. 1966 Computer programs for hierarchical polythetic classification ('similarity analyses'). *Comput. J.* **9**, 60–64. (doi:10.1093/comjnl/9.1.60)
32. Gupta S, Anderson RM, May RM. 1989 Networks of sexual contacts: implications for the pattern of spread of HIV. *AIDS* **3**, 807–818. (doi:10.1097/00002030-198912000-00005)
33. Farrington CP, Whitaker HJ, Wallinga J, Manfredi P. 2009 Measures of disassortativeness and their application to directly transmitted infections. *Biom. J.* **51**, 387–407. (doi:10.1002/bimj.200800160)
34. Cuesta-Lazaro C *et al.* 2021 Vaccinations or non-pharmaceutical interventions: safe reopening of schools in England. *medRxiv*. (doi:10.21203/rs.3.rs-955495/v1)
35. Office for National Statistics. 2011 Table ID KS105UK. (Household composition). See https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/2011censuskeystatisticsandquickstatisticsforlocalauthoritiesintheunitedkingdompart3.
36. Office for National Statistics. 2018 Ref: 008855. (Families with dependent children by number of children, UK, 1996 to 2017). See https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/families/adhocs/008855familieswithdependentchildrenbynumberofchildrenuk1996to2017.
37. Office for National Statistics. 2011 KS102UK. (Age Structure). See https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/2011censuskeystatisticsandquickstatisticsforlocalauthoritiesintheunitedkingdompart3.
38. Chapa M *et al.* 2016 Quantifying social contacts in a household setting of rural Kenya using wearable proximity sensors. *EPJ Data Sci.* **5**, 21. (doi:10.1140/epjds/s13688-016-0084-2)
39. Hens N, Ayele GM, Goeyvaerts N, Aerts M, Mossong J, Edmunds JW, Beutels P. 2009 Estimating the impact of school closure on social mixing behaviour and the transmission of close contact infections in eight European countries. *BMC Infect Dis.* **9**, 187. (doi:10.1186/1471-2334-9-187)
40. Quera-Bofarull A *et al.* IDAS-Durham/June: June 1.2.0. *Zenodo*. (doi:10.5281/zenodo.7199142)
41. Aylett-Bullock J *et al.* UNGlobalPulse/UNGP-settlement-modelling: v2.0.1. *Zenodo*. (doi:10.5281/zenodo.10120763)
42. Walker J, Aylett-Bullock J. IDAS-Durham/june_household_matrix_calculation: First zenodo release. *Zenodo*. (doi:10.5281/zenodo.10091526)
43. Walker J, Aylett-Bullock J. IDAS-Durham/june_mixed_method_CM_results: First zenodo release. *Zenodo*. (doi:10.5281/zenodo.10091522)
44. Kobo Inc. KoboToolbox. See https://www.kobotoolbox.org.
45. Open Data Kit Open Data Kit. See https://opendatakit.org.
46. Joint Government of Bangladesh. 2020 UNHCR population factsheet as of 31st October 2020. See https://data2.unhcr.org/en/documents/details/82872.
47. Joint Government of Bangladesh. 2020 Population breakdown as of 30 April 2020. See https://data2.unhcr.org/en/documents/details/76158.