The Population Attributable Fraction (PAF) of cases due to gatherings and groups with relevance to COVID-19 mitigation strategies

Ellen Brooks-Pollock[1,2*], Jonathan M Read[3], Thomas House[4], Graham F Medley[5], Matt J Keeling[6], Leon Danon[7]

[1] Bristol Veterinary School, University of Bristol, Bristol, BS40 5DU, UK.
[2] Population Health Sciences, Bristol Medical School, Bristol, BS8 2BN, UK.
[3] Lancaster Medical School, Lancaster University, Lancaster LA1 4YW, UK.
[4]School of Mathematics, University of Manchester, Manchester M13 9PL, UK.
[5]Centre for Mathematical Modelling of Infectious Disease, London School of Hygiene and Tropical Medicine, London, WC1H 9SH, UK.
[6]Mathematics Institute and Department of Life Sciences, University of Warwick, Coventry, CV4 7AL, UK.
[7] Department of Engineering Mathematics, University of Bristol, Bristol, BS8 1TW, UK.

*Corresponding author: Ellen Brooks-Pollock, University of Bristol, Ellen.Brooks-Pollock@bristol.ac.uk.

Summary

Methods

- We used data from a survey of social contact behaviour that specifically asked about contact with groups to estimate the Population Attributable Fraction (PAF) of COVID-19 cases due to groups as the relative change in the Basic Reproduction Number when groups are prevented.

Results

- Under normal circumstances (with pre-COVID contact patterns), groups of 50+ individuals accounted for 0.5% of reported contact events, and we estimate that the PAF due to groups of 50+ people is 5.5% (95%CI 1.4%, 11.4%). The PAF due to groups of 20+ people is 18.9% (12.7%, 25.7%) and the PAF due to groups of 10+ is 25.2% (19.4%, 31.4%)
- With social distancing in place, the contribution of groups increases.

Conclusions

- Without social distancing and assuming normal behaviour, large groups of individuals have a relatively small epidemiological impact; small and medium sized groups between 10 and 50 people have a larger impact on any epidemic.
- The observed overdispersion in reproduction numbers derived from social contact data broadly matches phylogenetic studies, for example

METHODS

Social contact data

The Social Contact Survey (SCS) [1,2] collected data on social contacts from 5,388 participants between 2009 and 2010 in the UK. Participants were asked to enumerate other people with whom they had had contact over the course of a single day. Contacts were defined as those with whom participants had a face-to-face conversation within 3 metres and/or physically touched skin-on-skin. Participants were able report individual contacts and up to five groups of contacts, for instance church groups, weddings, large work functions or multiple contacts at work. The 'groups' question was designed to aid participants in reporting multiple similar contacts. Group contacts were defined in the same way as individual contacts, i.e. if a person attended a concert with 1,000 people, but only spoke to 5 people, the number of recorded group contacts would be 5. Participants were asked whether members of the group knew each other.

As well as the number of contacts, participants were asked to estimate the length of time spent with each contact or group of contacts as either: less than 10 minutes, 11-30 minutes, 31-60 minutes or over 60 minutes, the distance from home, the frequency with which the contact took place and whether it involved physical contact.

The SCS data are available to download at http://wrap.warwick.ac.uk/54273/.

Reproduction Numbers and Population Attributable Fraction

We calculate the Basic Reproduction Number with and without groups of various sizes. For each participant j, we use their j_k contact reports to calculate their individual Reproduction Number,R_{ind}^{j}. We assumed that R_{ind}^{j}. is proportional to the number of individuals reported in the contact, n_i, ($n_i = 1$ for single contacts, $n_i > 1$ for groups) multiplied by the duration of each contact, d_i:

$$R_{ind}^{j} \propto \sum_{i=1}^{j_k} n_i d_i$$

The duration of each contact is taken as the mid-point of each time interval, i.e. 5 minutes, 20 minutes, 45 minutes and 6 hours, as recorded by the participant. The interpretation of contact duration is different for individual versus group contacts, as there is a limit to the number of face-to-face contacts that one person can make in a finite time. We observe a saturation of contact duration for individuals with large numbers of contacts (figure 1B). The saturation occurs between 20 and 30 contacts per individual. We adjust for this by dividing the duration of group contacts by the number of individuals in the group, when the number of group contacts is greater than a random number between 20 and 30.

The population-level reproduction number is the average number of secondary cases caused by an average infectious person. Individuals with higher R_{ind}^{j} with contribute more to the population-level R_t because they are more likely to get infected than individuals with lower R_{ind}^{j}. Therefore, we estimate R_t as a bootstrap resample (random sample with replacement) of the individual reproduction numbers weighted by the individual reproduction numbers:

$$R_t^{*} = Boot\left(\left(R_{ind}^{j}\right)^2\right) \qquad j = 1 \dots N \tag{2}$$

where N is the number of participants in the SCS. The mean and 95% Confidence Intervals were calculated from the bootstrapped sample.

We calculated the PAF for groups of size G or greater as the percentage change in the Basic Reproduction Number[3]:

$$PAF_G = 1 - R_0(without\ groups \geq G)/R_0(with\ groups).$$

We investigate the PAF for groups of greater than 10, and up to groups greater than 100, in increments of 10. We investigated differences between groups that knew each other and groups that did not know each other.

We explored the impact of groups in the context of other social distancing measures by varying the proportion of other social contacts that took place.

RESULTS

Impact of groups on numbers of contacts per person

48,001 unique contacts were reported by 5,388 participants. Of those, 42,945 (89%) were individual contacts and 5,056 groups were reported (accounting for 11% of reported contacts). The median and mean number of contacts per person was 11.5 and 27.0, range 1 to 3,011 (fig 1A).

2,427 (45%) of participants reported group contacts. The majority of groups reported (3,860; 76%) were groups of people who knew each other. 2,979 (59%) groups had 10 or fewer members; the median and mean reported group size was 9 and 20.3 individuals respectively.

Restricting contacts to groups of size 50 or less, reduces the median and mean number of individual contacts per person to 11.0 and 18.8; restricting contacts to groups of size 20 or less, reduces the median and mean number of contacts per person to 10.0 and 14.1; restricting contacts to groups of size 10 or less, reduces the median and mean number of contacts per person to 9 and 11.0. Figure 1 shows the degree distribution (number of contacts) per person with and without contacts associated with groups of size greater than 10.

We observe that the total contact time, calculated as the sum per individual of all reported contact durations, increases with number of contacts for low numbers of contacts. For individuals with greater than approximately 20 contacts, there is a saturation of total contact time at 29.5 hours contact hours in a 24-hour day (fig 1B).

The frequency with which groups are reported decreases with group size (fig 1C). Nearly 60% of groups contain 7 or more individuals; 20% contain more than 20 individuals; 5% contain more than 50 individuals. It is likely that large events are reported as multiple smaller groups, but we cannot identify this from the data.

We observe that the presence of groups increases an individual's total number of contacts (fig 1D). Individuals reporting no groups have a median number of contacts of 7. Individuals reporting groups have a median of 28 contacts.

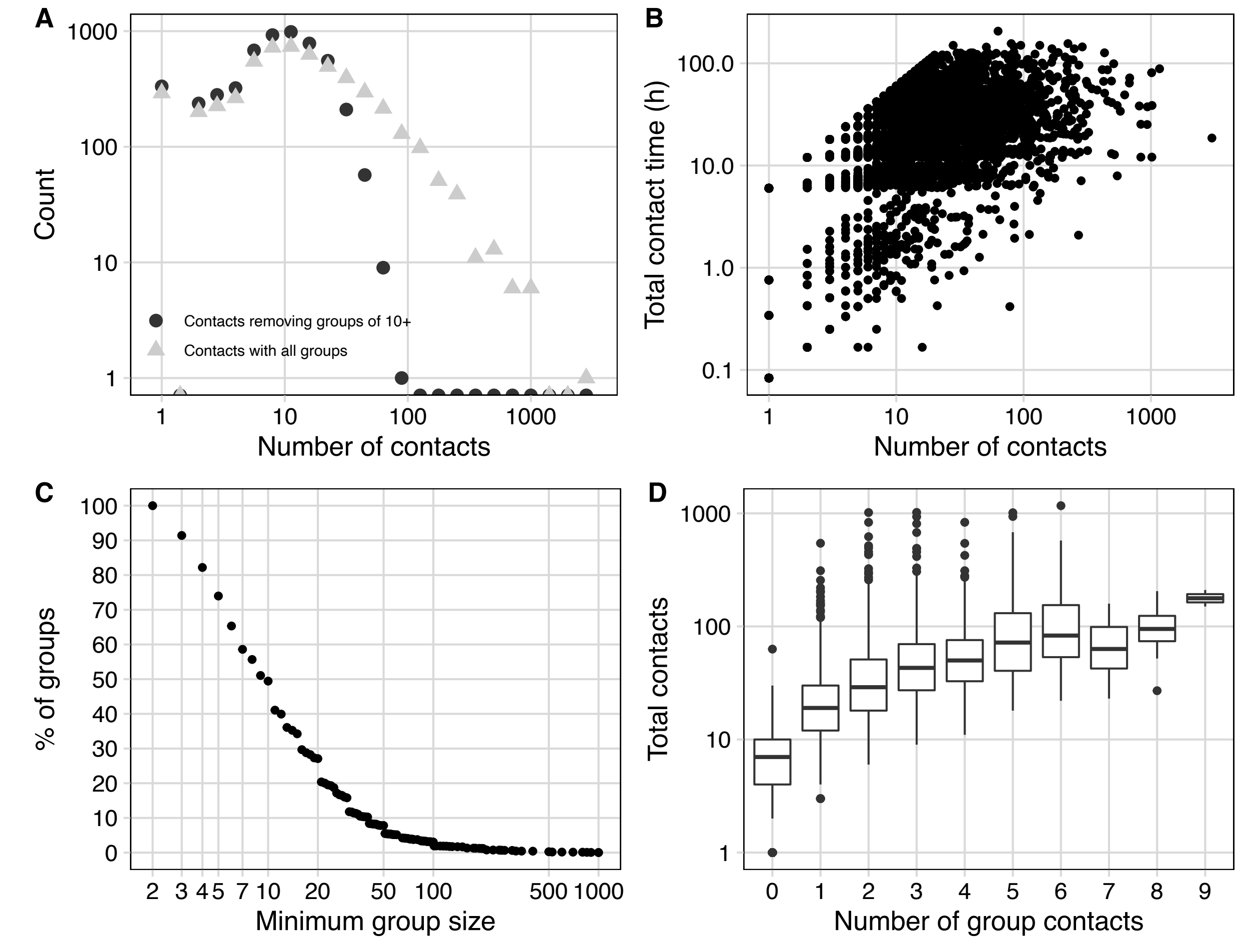

Figure 1: A) The distribution in the number of social contacts per participants from the Social Contact Survey (n=5,388) with and without groups of ten and greater. Even without groups of ten or more, individuals can have more than ten other contacts. B) The relationship between number of contacts and total contact duration. C) The percentage of groups reported with at least a given number of individuals. All data from the Social Contact Survey.

Compared to individual contacts, group contacts were more likely to be more than 2 miles from the participants home (64% versus 51%), less likely to involve physical contact (25% versus 44%) and more likely to involve new individuals (21% versus 15%) (table 1).

Table 1: Comparison of individual and group contacts in the Social Contact Survey (2010).

	Individual contacts	**Group contacts**
Frequency:		
4 or more times a week	31%	26%
2 or 3 times a week	17%	18%
Once a week	16%	19%
Less than once a week	22%	17%
Met for the first time	15%	21%
Distance:		
<2 miles from home	49%	36%
3 – 10 miles from home	29%	40%
11 – 50 miles from home	15%	18%
51+ miles from home	7%	7%
Involving touch:		
Yes	43%	25%
No	57%	75%

Population Attributable Fraction (PAF)

The PAF due to groups decreased with increasing group size. For the largest groups with more than 100 individuals the PAF_{100} is estimated at 0.6% (95% CI: 0.4%, 0.8%). The PAF_{50} is estimated at 5.5% (95% CI: 1.4%, 11.4%); the PAF_{20} is 18.9% (95% CI: 12.7%, 25.7%); the PAF_{10} is 25.2% (95% CI: 19.4%, 31.4%) (fig 2, left panel).

The pattern of decreasing PAF with increasing group size is seen for both groups of individuals who are known to each other and groups of individuals who are unknown to each other. The PAF due to groups of 10+ known to each other is estimated at 20.6% (95% CI: 15.4%, 26.3%) and due to groups of 50+ known to each other is estimated at 2.9% (95% CI: 0.8%, 6.9%). The remaining contribution to R_t is due to contact with individuals.

The estimated impact of large groups on R_t is due to the relative frequency with which they are reported in the Social Contact Survey, which took place under normal circumstances with no social distancing. If the number of other types of contacts are reduced, then the relative importance of group contacts increases (fig 2, right panel).

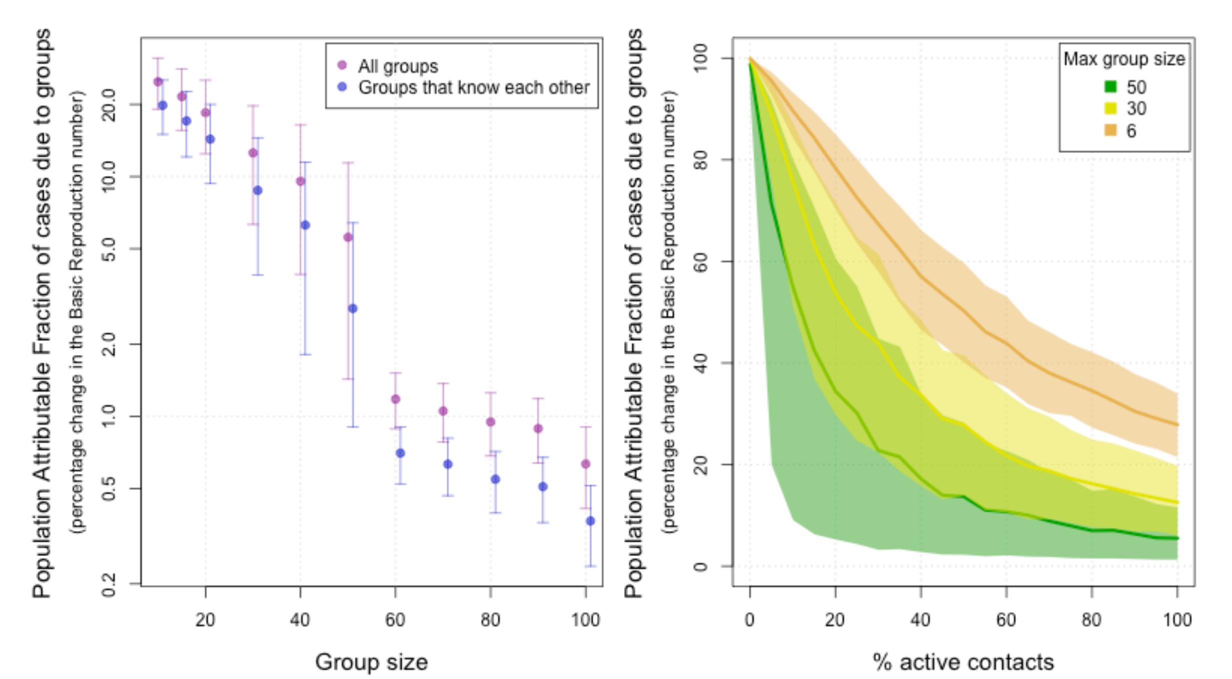

Figure 2: Left: The Population Attributable Fraction (PAF) of cases due to groups of various sizes. The purple circles are all groups, the blue circles are groups of people who are known to each other. The error bars are 95% confidence intervals. Right: The PAF due to groups of size 50, 30 and 6 against the number of other (non-group) contacts that are active. s

COMPARISON WITH OTHER STUDIES

This analysis was conducted before genomic sequencing could be used to quantify the role of superspreading and large events.

In a phylogenetic analysis of cases in the Boston area, USA[4], 29% of cases were reported to be responsible for 85% of secondary infections. In our analysis of the SCS, 29% of participants reported 77% contacts. After accounting for contact duration and risk of infection (as we do in the paper), 29% of individuals contribute 84% of secondary cases, suggesting that using social contact data is able to capture the overdispersion of superspreading. Other features of large events, such as the increased distanced travelled to group contacts, are not captured in our analysis.

Social contact surveys have been conducted during the COVID-19 pandemic[5–7]. A social contact survey conducted in 2020 in University of Bristol staff and students (CON-QUEST [7,8]) shows that smaller groups were reported in 2020 (median size of groups in CON-QUEST = 7, median size of groups in SCS = 9), see figure 3.

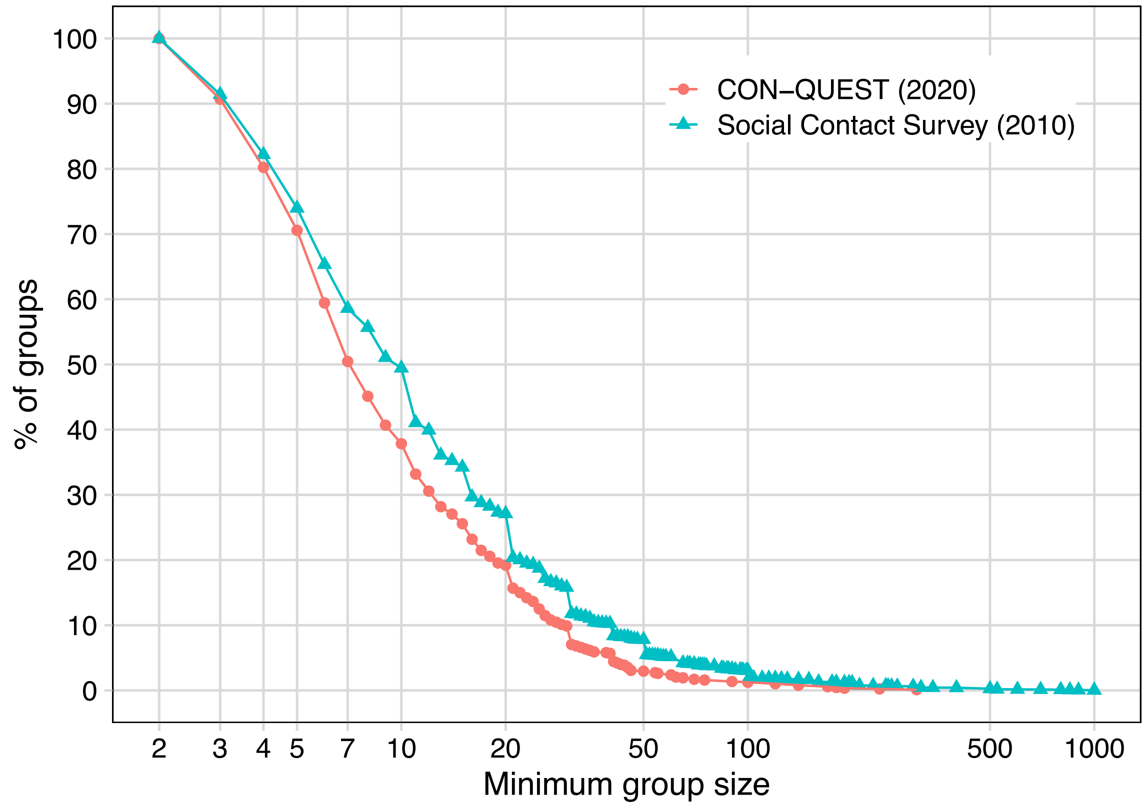

Figure 3: Comparison of group sizes reported in the Social Contact Survey (data collected in 2010) and CON-QUEST (data collected in 2020).

DISCUSSION

- In this paper, we analysed social contact data in the context of infectious disease transmission and gatherings.
- Our findings suggest that large groups of individuals have a relatively small impact on an epidemic, under normal circumstances with pre-COVID contact patterns.
- The relatively small contribution of groups is due to the relative rarity of large-scale gatherings and the sub-linear scaling between number of contacts and infectivity.
- This analysis included all groups, including leisure, work and home-based gatherings.

LIMITATIONS

- The data used in this analysis were collected in 2010. Social contact surveys conducted during 2020 have shown that contact patterns changed substantially in 2020[6–8].
- We are not able to differentiate between indoor and outdoor contacts.
- The groups reported in the Social Contact Survey are not necessarily public or mass gatherings and represented groups that both knew each other and those that did not.

- The group sizes reported in the SCS were not necessarily the same size of an event where contacts may have taken place. Therefore, this analysis should be considered in terms of contacts per person, rather than to guide the acceptable size of organised events.
- We found that group contacts were associated with additional characteristics not captured in our analysis, such as increased distance travelled, and likelihood of meeting a new contact.

Acknowledgements.
Thanks to Adam Trickey for extracting the CONQUEST data on group sizes.
EBP was partly supported by the NIHR Health Protection Research Unit (HPRU) in Behavioural Science and Evaluation. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. The NIHR had no role in writing the manuscript or the decision to publish. EBP, LD, MJK, JMR and TH are funded via the JUNIPER consortium (MRC grant MR/V038613/1). EBP, LD and MJK are funded by MRC grant MC/PC/19067. LD is funded by EPSRC grants EP/V051555/1 and EP/N510129/1. JMR is supported by EPSRC (EP/N014499/1) and MRC (MR/S004793/1, MR/V028456/1).

References

[1] Danon L, House TA, Read JM, Keeling MJ. Social encounter networks: collective properties and disease transmission. J R Soc Interface 2012;9.

[2] Danon L, Read JM, House TA, Vernon MC, Keeling MJ. Social encounter networks: Characterizing great Britain. Proc R Soc B Biol Sci 2013;280. https://doi.org/10.1098/rspb.2013.1037.

[3] Brooks-Pollock E, Danon L. Defining the population attributable fraction for infectious diseases. Int J Epidemiol 2017;46:976–82. https://doi.org/10.1093/ije/dyx055.

[4] Lemieux JE, Siddle KJ, Shaw BM, Loreth C, Schaffner SF, Gladden-Young A, et al. Phylogenetic analysis of SARS-CoV-2 in Boston highlights the impact of superspreading events. Science (80-) 2021;371. https://doi.org/10.1126/science.abe3261.

[5] Jarvis CI, Van Zandvoort K, Gimma A, Prem K, Klepac P, Rubin GJ, et al. Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK. BMC Med 2020;18:124. https://doi.org/10.1186/s12916-020-01597-8.

[6] Jarvis CI, Gimma A, van Zandvoort K, Wong KLM, Abbas K, Villabona-Arenas CJ, et al. The impact of local and national restrictions in response to COVID-19 on social contacts in England: a longitudinal natural experiment. BMC Med 2021;19:52. https://doi.org/10.1186/s12916-021-01924-7.

[7] Nixon E, Trickey A, Christensen H, Finn A, Thomas A, Relton C, et al. Contacts and behaviours of university students during the COVID-19 pandemic at the start of the 2020/21 academic year n.d. https://doi.org/10.1101/2020.12.09.20246421.

[8] Trickey A, Nixon E, Christensen H, Finn A, Thomas A, Relton C, et al. University students and staff able to maintain low daily contact numbers during various COVID-19 guideline periods. MedRxiv 2021:2021.01.19.21250097. https://doi.org/10.1101/2021.01.19.21250097.

Addendum on group sizes with vaccination