Analysis of SARS-CoV-2 transmission clusters and superspreading events

Gwen Knight, Quentin Leclerc and Adam Kucharski, on behalf of CMMID working group
Report for SPI-M, 3rd June 2020

Evidence of overdispersion in SARS-CoV-2 transmission
Modelling the offspring distribution as a negative binomial with mean R and overdispersion parameter k, it has been previously estimated that k=0.16 (95% CI: 0.11–0.64) for SARS-CoV (Lloyd-Smith et al, 2005) and k=0.26 (95% CI: 0.09–1.24) for MERS-CoV (Kucharski et al, 2015). Note: as k tends to infinity, the offspring distribution converges to a Poisson process. The following are available estimates for SARS-CoV-2:

Data source	Estimated k	Reference
Early international clusters	0.04–0.2	Endo et al, Wellcome Open Research 2020
Early cumulative cases in China	0.54 (90% CI: 0.014–6.95)	Riou et al, Eurosurveillance, 2020
Contact tracing in Shenzen	0.58 (95% CI: 0.35-1.18)	Bi et al, Lancet ID, 2020
Outbreak investigation in Austria	0.16	https://www.ages.at/service/service-presse/pressemeldungen/epidemiologische-abklaerung-am-beispiel-covid-19/
Contact tracing in Hong Kong	0.45 (95% CI: 0.30 – 0.72)	Adam et al, Research Square, 2020
Rapid review of 49 superspreading reports (defined as at least 4 secondary cases from single primary case)	0.1–0.2 (assuming 2<R<4)	CMMID unpublished data, fitted using left-truncated negative binomial.

Implications of overdispersion for proportion of new cases in given cluster size
We estimated that the following proportions of new cases occur in clusters of at least a given size for different values of k and R.

Cluster proportions when R=0.8

Dispersion parameter	Proportion new cases in cluster of at least 5	Proportion new cases in cluster of at least 10	Proportion new cases in cluster of at least 25
0.1	67%	38%	7%
0.2	49%	18%	0.8%
0.3	38%	9%	0.08%
0.4	30%	5%	0.03%
0.5	25%	3%	0.01%

Cluster proportions when R=1.2

Dispersion parameter	Proportion new cases in cluster of at least 5	Proportion new cases in cluster of at least 10	Proportion new cases in cluster of at least 25
0.1	77%	53%	17%
0.2	62%	32%	4%
0.3	52%	20%	0.8%
0.4	45%	13%	0.2%
0.5	40%	9%	0.07%

Cluster proportions when R=3

Dispersion parameter	Proportion new cases in cluster of at least 5	Proportion new cases in cluster of at least 10	Proportion new cases in cluster of at least 25
0.1	90%	79%	49%
0.2	84%	65%	28%
0.3	79%	55%	16%
0.4	76%	48%	9%
0.5	73%	41%	6%

For context, one recent study estimated a 24% reduction in R from limiting gatherings to 10 people or less, compared to 9% reduction from limiting to 100 or less (Brauner et al, MedRxiv 2020).

Summary of features of transmission clusters (full details in appendix)
- A database of settings linked to COVID-19 clusters was initiated by researchers at LSHTM [Leclerc et al] and is now being used as a peer-supported resource (google sheets) for all cluster outbreaks of COVID-19.
- The initial aim was to detail *settings* in which transmission of SARS-CoV-2 was known to have occurred through the detection of more than one infection or symptomatic "cases" (= a cluster). We record only the first generation of cases linked to transmission in the setting.
- As of 26/05/2020, we have included 201 transmission events in our database, which we classified into 22 setting types. The vast majority of these setting types are either "indoor" or "indoor/outdoor" (21/22). Almost all clusters involved less than 100 cases in the first generation of transmission (i.e. linked directly to this setting) (181/201). Outlier settings with maximum cluster sizes of more than 100 cases were hospitals, elderly care, worker dormitories, food processing plants, prisons, schools, shopping, religious events and ships. Other settings with maximum cluster sizes between 50-100 cases were weddings, sport, bar, shopping, conference and work.

Conclusions and suggestions for future modelling directions

- Evidence that large proportion of cases in countries with high detection rates are part of clearly defined transmission clusters:
    - "Clusters now account for 68% of South Korea's COVID-19 cumulative cases nationally, while individual case-based contacts account for only 11%" (Imperial College Covid-19 response team, Report 25)
    - New Zealand: 617 cases in "significant clusters" out of 1504 "confirmed and probable" = 41% linked to "significant" clusters.
- Rapid identification of clusters may be disproportionately important for control, e.g. large-scale testing and quarantine based on event/location could be contributing larger reduction than reconstruction of individual-level transmission chains via contact tracing.
- Identification and prevention of clusters would be easier if most transmission occurs in predictable settings/situations.

Suggestions for future data analysis:

- Probe sources in the growing LSHTM database for more detailed data and hence compare attack rates by setting.
- Link analysis to outbreak investigation organisations (e.g. national health bodies such as PHE) to determine settings of transmission and "risk" (i.e. attack rates) from these data, as well as to assign proportion of transmission by setting.

Suggestions for future modelling:

- Investigate how under-detection of infections and the level of social distancing influences estimates for k. This parameter may be constrained to higher values during lockdowns (less potential for superspreading events).
- Identify the extent to which setting-specific restrictions on interactions (e.g. limit on gathering size) combined with targeted testing could reduce overall transmission.
- Estimate relative effectiveness of case vs cluster-based tracing methods for plausible assumptions about detection and adherence.

Appendix: Full description of features of transmission clusters

Data sources: The initial search was composed of a structured PubMed based search for "clusters", followed by an online search of media reports and then setting specific follow-up in online search engines. The current updated dataset consists of follow up online searchers as well as the inclusion of peer contribution through an online spreadsheet. Note that we restrict cluster size to only include individuals infected within a specific setting, and exclude secondary infections which occurred outside the settings. We also looked for evidence of the first 100 transmission events in a country but found very little open access data.

Biases: As this data is not taken from structured outbreak investigations there are several limitations. The crucial one is the inability to assign importance to these settings: we do not know what proportion of transmission occurs in the settings reported. This is compounded by recall bias and a reliance on media reports which will both bias towards the inclusion in our database of big/interesting events and associated settings.

What is missing:
- We identified few clusters linked to schools (8/201).
- Most of the clusters had insufficient data / evidence to determine secondary or final attack rates (we could estimate at least one of these rates for 40 clusters only).
- While there were several reports of outbreaks in choirs, there were no clusters as yet reported in our database of other regular meetings of individuals (e.g. chess club / mother&baby class / local rugby teams etc.). This could be due to the population attending these meetings (e.g. age distribution) or to the activity (see below).
- Most of the data comes from Asia (130/201). We had little data from the UK and other European countries due to the lack of published contact tracing or outbreak investigation data
- Linked to the above lack of tracing, we had no clusters linked to large football matches (e.g. the one supposedly before the large Northern Italy outbreak), races (e.g. Cheltenham) - some settings that we might expect, and anecdotally know, are linked to transmission, are not in our database.

Shared characteristics / signals in the data:
- Usual suspects: Care settings as well as religious events and cruise ships were identified as would be expected from transmission of other infectious diseases
- Indoor: The vast majority of the settings linked to COVID-19 clusters were indoor spaces. This may be due to the time of year (many came from Asia during winter) or to the recall bias which we mentioned above.
- Noise: Many of the settings were loud places (food processing plants / bars)
- Temperature: the high number of food processing plants (9) may suggest that cooler settings, in which the virus can survive for longer, may be linked to increased transmission
- Activity matters: not just time in shared space, but the activity being performed. E.g. many choirs, from different countries / transmission from an instructor giving a Zumba but not a subsequent Pilates class

Implications for modelling: We did not find evidence for homogeneous transmission - only a sub-sample of settings is included in this data. This is unlikely to be only driven by the current limited level of outbreak investigation and instead implies that there are some settings which are "more risky" either due to the (a) environment (indoor / lack of ventilation / population density) or (b) activity occurring (singing / loud) or (c) duration of contact (long shifts).

This suggests that transmission dynamic models of SARS-CoV-2 spread should be stratified by low and high risk settings. Further data and analysis to determine secondary and final attack rates, as well as the dispersion factor would help to support this structuring.

Next steps for the database: We are continually integrating new clusters / settings into the database with an approximately weekly update. The next step is to work with Prof. Noakes (University of Leeds) and her team to retrieve information from these clusters on the environment and route by which transmission occurred (fomite / aerosol / droplet) where possible.

setting	count	Secondary cases			Total cluster size			total	countries	Indoor / outdoor
		Min	Med	Max	Min	Med	Max			
Bar	12	2	9	16	3	13	80	319	Germany, Austria, Italy, Singapore, Japan, USA, Australia, New Zealand, Brazil	Indoor / outdoor
Building site	4	/	/	/	5	20.5	49	95	Singapore	Outdoor
Conference	5	/	/	/	3	10	89	148	Canada, Singapore, Japan, USA, New Zealand	Indoor / outdoor
Elderly care	17	/	/	/	5	19	167	638	UK, Canada, Scotland, France, Germany, Italy, USA, Japan, New Zealand, Luxembourg	Indoor
Food processing plant	9	2	2	2	3	84	518	1207	USA, Germany, Canada, Netherlands	Indoor
Funeral	1	3	3	3	4	4	4	4	USA	Indoor / outdoor
Hospital	9	1	3	14	2	10	118	224	China, Singapore, Italy, Taiwan, South Korea, Japan	Indoor
Hotel	2	/	/	/	3	5	7	10	Singapore	Indoor
Household	36	1	3	11	2	4	12	168	China, Italy, Vietnam, Taiwan, South Korea, Hong Kong, France	Indoor
Meal	17	1	3	10	2	5	47	134	Singapore, USA, Vietnam, China, South Korea, Japan	Indoor

Setting									Location	Type
Prison	4	351	351	351	66	226	353	871	USA, Ethiopia	Indoor
Public	4	/	/	/	10	10	27	57	China, Japan	Indoor / outdoor
Religious	15	1	18	52	2	23	130	570	USA, Singapore, South Korea, US, China, India, Netherlands, Germany	Indoor / outdoor
School	8	1	1	131	2	22	133	349	Singapore, France, USA, New Zealand, Australia, Sweden	Indoor / outdoor
Ship	5	619	619	619	78	662	1156	3597	Grand Princess, Diamond Princess, Ruby Princess, USS Theodore Roosevelt, Charles de Gaulle aircraft carrier	Indoor
Shipyard	1	/	/	/	22	22	22	22	Singapore	Indoor / outdoor
Shopping	9	5	10	19	7	20	163	361	China, Singapore, Peru, Mexico	Indoor / outdoor
Sport	6	1	1	1	2	7.5	65	95	South Korea, Singapore, Italy, Japan	Indoor / outdoor
Transport	1	1	1	1	3	3	3	3	China	Indoor
Wedding	3	/	/	/	13	43	98	154	Australia, New Zealand	Indoor / outdoor
Work	12	6	7	11	4	8.5	97	198	China, Singapore, South Korea, Germany	Indoor
Worker dormitories	21	/	/	/	3	24	797	1702	Singapore	Indoor