Notes on Mumbai's COVID-19 epidemic and the September resurgence

(Murad Banaji, 28/09/2020)

Estimating infection detection in slum and non-slum areas

Variation in the detection of infections is estimated as follows. We assume, based on the serosurvey, that citywide prevalence was 56.5% in the slums and 15.5% in the non-slum areas. Based on 2011 census data as reported here, wards are grouped into three roughly equally sized strata: slum-heavy, medium, and slum-light. Cases in each stratum on 08/07 were taken from MCGM bulletins. This date is chosen as being the approximate mid-point of the serosurvey. This is a generous choice, allowing for a considerable delay in infection reporting, and thus slightly pushing up the estimate of infection detection.

stratum	**wards**	**cases on 08/07**	**% in slums**	**total population (2011)**	**% in slums in whole stratum**
slum-heavy	H/E, K/E, L, M/E, P/N, R/N	24440	>=61	4463803	76%
medium	F/S, F/N, G/N, M/W, N, P, R/S, S	31526	>=33 & < 61	4422310	51%
slum-light	A, B, C, D, E, G/S, H/W, K/W, R/C, T	27990	<33	3556260	25%
city	all	83956		12442373	52.5%

Let d_s and d_n refer to the proportion of infections detected in the slums and non-slum areas respectively. Let a_s, a_m and a_n refer to the assumed number of infections in the non-slum areas of each stratum divided by the population of that stratum, and b_s, b_m and b_n refer to the number of infections in the slums of each stratum divided by the population of that stratum. Let c_s, c_m and c_n refer to the cases in each stratum divided by the population of that stratum. a_t, b_t and c_t refer to the values of these parameters for the city as a whole. a_s, a_m, a_n, a_t, b_s, b_m, b_n and b_t are estimated assuming 56.5% prevalence in the slums and 15.5% prevalence in non-slum areas. The population of each stratum is assumed to be 5% greater than its 2011 census value. Overall we get:

stratum	a_i	b_i	c_i
slum-heavy	0.0371044687	0.4297482269	0.0052144305
medium	0.076215015	0.2871839774	0.0067893843
slum-light	0.1161481305	0.1416213306	0.0074958363
city	0.0735974089	0.2967255742	0.0064262738

The goal is to minimise the difference between predicted and observed cases across the strata, namely $\Sigma_i(c_i-a_id_n - b_id_s)^2$ where i varies over s, m and n, while insisting that $c_t-a_td_n-b_td_s=0$, namely that predicted cases in the city as a whole match observed cases. This optimisation can be reduced to a simple regression. The process gives, to 3 significant figures, $d_s = 0.00762$ and $d_n = 0.0566$, corresponding to 0.76% and 5.66% case detection in the slums and non-slum areas respectively. The r^2 value of the regression was 0.99, and predicted cases in each stratum were within 5% of the recorded numbers (more precisely the method gave errors of 3.1%, -4.2% and 2.1% in predicted cases in the slum-heavy, medium and slum-light strata respectively).

Since cases in the daily summary bulletins were (around 08/07) about 4% higher than the totals for the wards, to correct for the mismatch, we increase the values of detection in the slums and non-slum areas by 4% to get infection detection rates of:

- slum areas: **0.79%**
- non-slum areas: **5.9%**
- city overall: **1.8%**.

Estimating infections on 25/09/2020

We can estimate infections in the city on September 25 under different assumptions. We vary assumed prevalence in the slums between 65% and 80%; we also allow cumulative infection detection in slum and nonslum areas to either (i) remain at serosurvey levels estimated above, or (ii) increase by 25%, or (iii) increase by 50%. The following table summarises the results of this process.

Assumed slum prevalence	**Nonslum prevalence, total prevalence, post serosurvey net case detection: at 0% increase in cumulative detection in slums/nonslums**	**… at 25% increase in cumulative detection in slums/nonslums**	**... at 50% increase in cumulative detection in slums/nonslums**
65%	46%, 56%, 4.3%	35%, 51%, 6.0%	27%, 47%, 8.2%
70%	45%, 58%, 3.9%	34%, 53%, 5.1%	26%, 49%, 6.6%
75%	44%, 60%, 3.5%	33%, 55%, 4.5%	26%, 52%, 5.6%
80%	43%, 62%, 3.2%	32%, 57%, 4.0%	25%, 54%, 4.9%

We find a minimum non-slum prevalence of 25% (80% slum prevalence, 50% increase in cumulative detection), rising to a maximum of 46% (65% slum prevalence, no increase in detection). Meanwhile the total prevalence ranges from 47% (50% increase in detection, 65% slum prevalence, 27% non-slum prevalence) to 62% (no increase in detection, 80% slum prevalence, 43% non-slum prevalence). Net infection detection post the serosurvey ranges from 3.2% to 8.2% with higher values corresponding to lower values of prevalence across the city.

Note that estimates of non-slum prevalence are somewhat insensitive to assumed slum prevalence, since the assumption is that the uneven detection holds. Even if we don’t assume that detection has improved in either slum or non-slum areas, there has been an increase in net detection as a consequence of the shift in infection to non-slum areas where detection is higher. In the most extreme scenario of a 50% increase in cumulative detection in both slum and non-slum areas and only a marginal increase in prevalence in the slums since the serosurvey, case detection since the serosurvey has averaged 8.2%.

Note that the estimates assume an even increase (or no increase) in detection between slum and non-slum areas. Of course, it is possible that increase has been uneven. If, for example, cumulative detection has increased by 50% in non-slum areas, but not at all in slums, then this would push up the values for non-slum prevalence and total prevalence from those in the final column of the table above, and push down the estimates of total detection.