Karachi analysis

7/15/2020

Summary of results

Area	Area/Wave ID	Sample size	# positive	# households
Gulshan Town	91	500	2	89
Ibrahim Hyderi	92	500	0	110
Gulshan Town	291	500	100	93
Ibrahim Hyderi	292	504	64	114

Evidence of household transmission

In the Curmei et al. paper "Estimating Household Transmission of SARS-CoV-2", they discuss three metrics of household transmission: intra-household reproduction number (R_h), household secondary attack rate (SAR) and household conditional risk of infection (CRI). These quantities are defined below:

- The **intra-household reproductive number (R_h)** is the average number of new infections caused by an infected individual inside their household.
- The **household secondary attack rate (SAR)** is the probability an infected person infects a specific household member.
- The **household conditional risk of infection (CRI)** is the probability that an individual in the household is infected, given another household member is infected.

The first two measures require information about actual transmission (attribution) within households, which we do not have. However, the CRI is estimatable using data from a single time point.

Below are estimates of CRI based on the second wave of data for each area, along with 95% bootstrap confidence intervals.

Area	Estimate	95% lower bd	95% upper bd
291	0.410	0.277	0.515
292	0.312	0.155	0.466

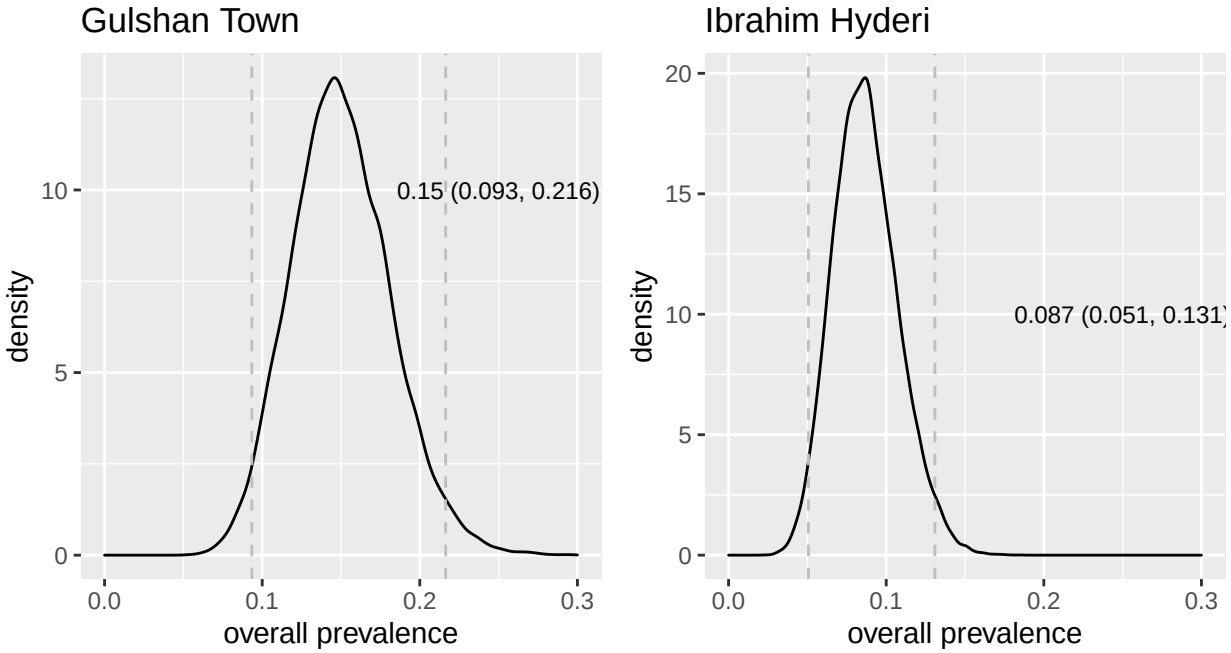
Now we partition households into those with at least one symptomatic individual and those without any. The formula for CRI only involves households with at least one seropositive individual. In the table below, I include the number of such households involved in each calculation. Interestingly, for GT, the estimate of CRI is higher in households without any symptomatic individuals, while in HI the estimate of CRI is higher in households with a symptomatic individual. Note that in this calculation, a household was classified as "symptomatic" if it contained at least one individual who reported feeling symptoms, regardless of whether that individual was seropositive or seronegative.

Symptoms?	Area	Num + households	Estimate	95% lower bd	95% upper bd
symp	291	11	0.378	0.195	0.549
symp	292	7	0.581	0.154	0.826
no symp	291	34	0.438	0.259	0.582
no symp	292	32	0.227	0.102	0.366

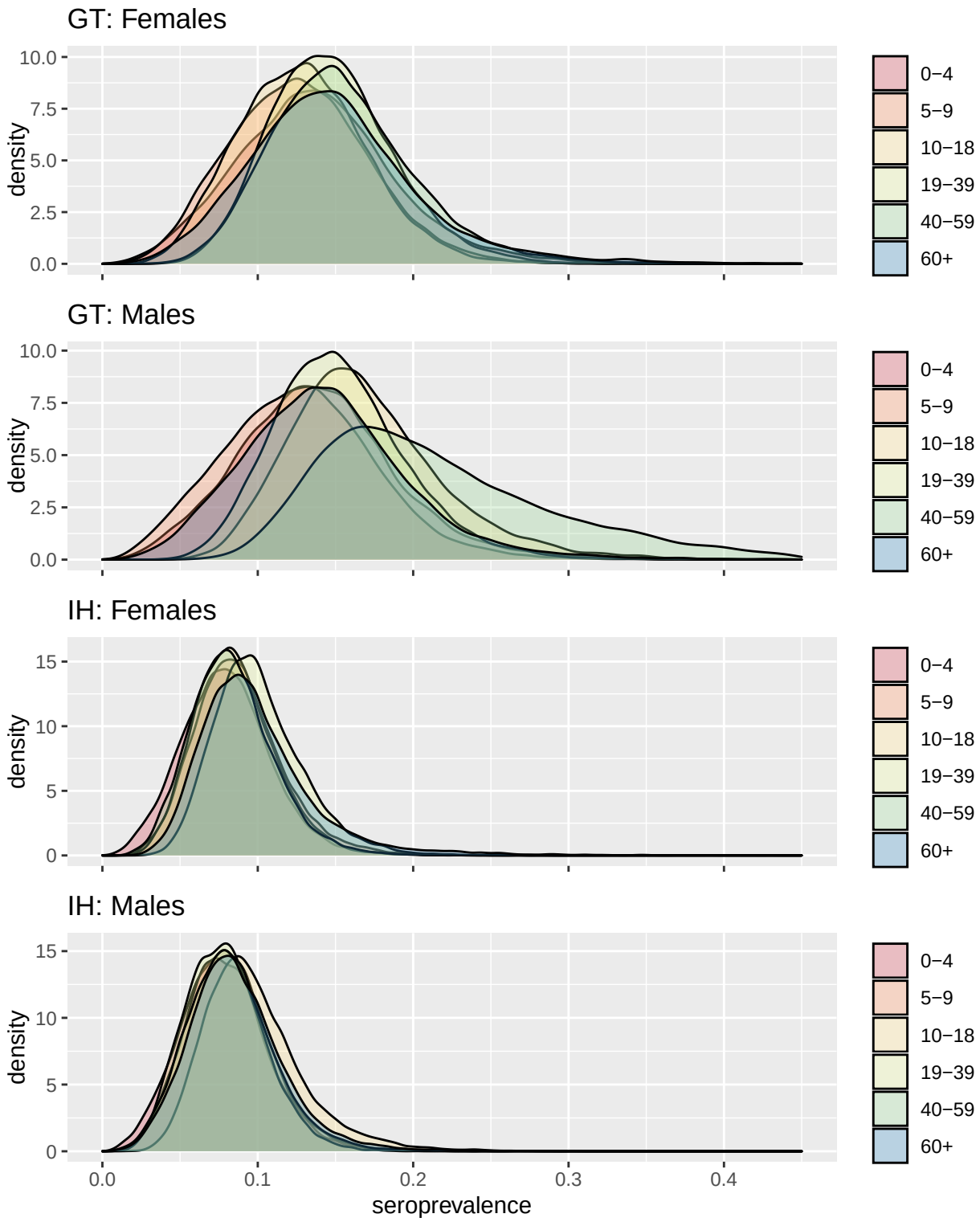
Estimating overall seroprevalence

We fit a Bayesian multilevel regression model and perform poststratification to obtain an overall estimate of seroprevalence in each area. We directly model the lab data reported by Roche for the Elecsys® Anti-SARS-CoV-2 assay to account for uncertainity in the test accuracy. We fit the model *separately* to the data from Gulshan Town and that from Ibrahim Hyderi and only consider the data from wave 2 since the number of positive tests were extremely low in wave 1.

Lab data: We include the results from all 5272 negative controls run by Roche (of which 5262 were negative) and include the results from the 88 positive control samples that were at least a week post PCR confirmation (of which 81 were positive). We omit the 116 samples run 0-6 days post PCR confirmation as it is probably fair to assume most of the seropositive indivdiuals sampled are at least 7 days post PCR confirmed. In addition, we omit the results from the 20 positive and 20 negative controls run specifically for this study since these were not run to estimate sensitivity and specificity, but rather to simply verify the lab procedure.



Prevalence estimates by age/sex



Below is a plot of the posterior mean estimates, along with 95% credible intervals for each gender/age group.

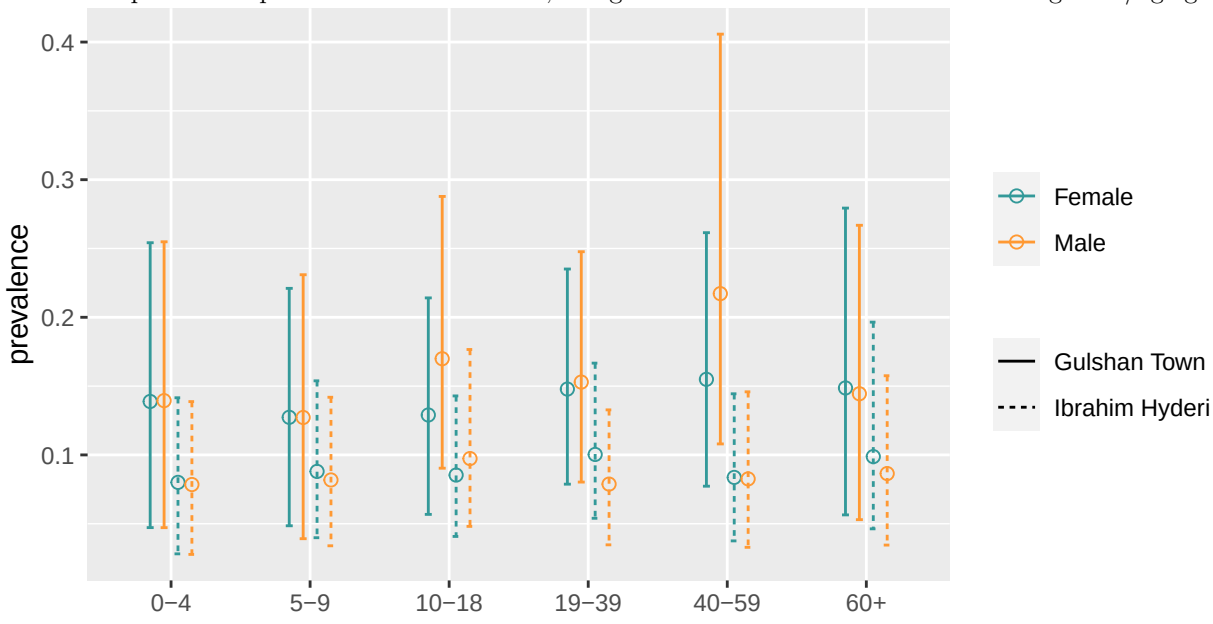


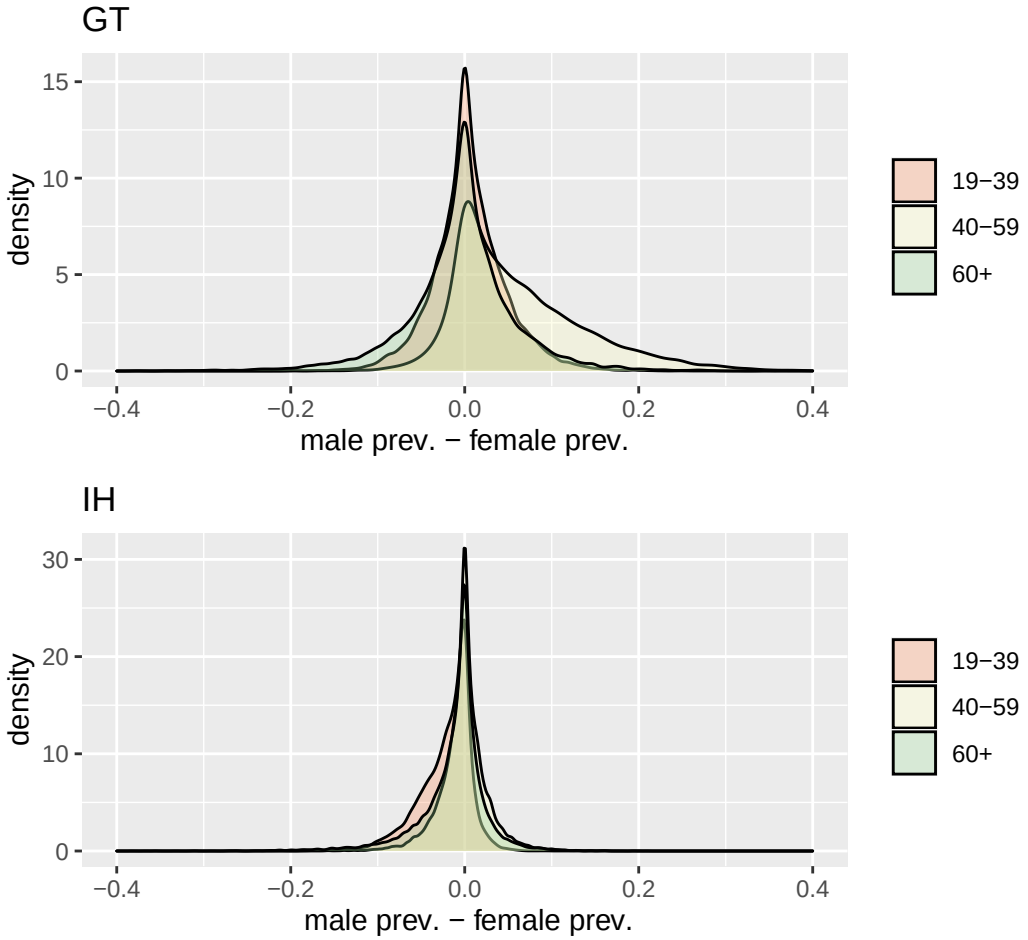
Table 4: Gulshan Town

Gender	Age	Prevalence	95% lower bd	95% upper bd
Female	0-4	0.139	0.047	0.254
Female	5-9	0.127	0.049	0.221
Female	10-18	0.129	0.057	0.214
Female	19-39	0.148	0.079	0.235
Female	40-59	0.155	0.077	0.261
Female	60+	0.149	0.056	0.279
Male	0-4	0.139	0.047	0.255
Male	5-9	0.127	0.039	0.231
Male	10-18	0.170	0.090	0.288
Male	19-39	0.153	0.080	0.248
Male	40-59	0.217	0.108	0.406
Male	60+	0.144	0.053	0.267

Table 5: Ibrahim Hyderi

Gender	Age	Prevalence	95% lower bd	95% upper bd
Female	0-4	0.080	0.028	0.142
Female	5-9	0.088	0.040	0.154
Female	10-18	0.085	0.041	0.143
Female	19-39	0.100	0.054	0.167
Female	40-59	0.084	0.038	0.144
Female	60+	0.099	0.046	0.196
Male	0-4	0.078	0.028	0.139
Male	5-9	0.082	0.034	0.142
Male	10-18	0.097	0.048	0.177
Male	19-39	0.079	0.035	0.133
Male	40-59	0.083	0.033	0.146
Male	60+	0.086	0.034	0.158

Estimate difference between prevalence by gender



Above are the posterior distributions for the difference between male and female prevalence for the adult age groups. Notice that they are mostly centered on zero indicating there is not evidence of a significant difference between genders.

Bayesian hierarchical model

Model for field data

Let y_i denote the test outcome of individual i. We model y_i as Bernoulli where probability the individual is seropositive is π_i and the probability the individual tests positive is p_i. Given a perfect diagnostic test, $p_i = \pi_i$, however we know there is non-zero probability of observing a false positive and false negative. Thus,

$$p_i = \pi_i * \text{se} + (1 - \pi_i) * (1 - \text{sp})$$

where se is the true sensitivity and sp is the true specificity of the test.

Recognizing that seroprevalence may vary by sex and age, we consider the following logistic regression model:

$$\pi_i = \text{logit}^{-1}(\beta_1 + \alpha^{ag}_{ag[i]} + \alpha^{hh}_{hh[i]})$$

where $ag[i]$ indexes which age/gender group person i belongs, and $hh[i]$ indexes which household person i belongs.

As in Gelman and Carpenter (2020), we place a uniform(0,1) prior distribution on the probability that an average person is positive by specifying a unit logistic prior for the intercept β_1. The effects for age/gender and household have hierarchical priors

$$\alpha^{ag}_j \sim \text{normal}(0, \sigma^{ag})$$

$$\alpha^{hh}_k \sim \text{normal}(0, \sigma^{hh})$$

where σ^{ag} and σ^{hh} are modeled using a $\text{normal}^+(0, 0.5)$.

Model for lab data

Notice that the model above requires knowledge of the true sensitivity and specificity. Instead of estimating these quantities and pretending they are known exactly, we directly model the lab validation data provided by Roche for the Elecsys® Anti-SARS-CoV-2 assay.

In general, consider a n_{se} positive control samples of which y_{se} test positive and n_{sp} negative control samples of which y_{sp} correctly test negative. We can model these results as binomial outcomes:

$$y_{se} \sim \text{binomial}(n_{se}, \text{se})$$

$$y_{sp} \sim \text{binomial}(n_{sp}, \text{sp})$$

We specify uniform(0,1) priors on the sensitivity se and specificity sp parameters.