The May-June national COVID-19 serosurvey: what does it tell us?

(Murad Banaji, 20/09/20)

(*This is a draft document which may be updated to correct any errors or as more information becomes available.*)

The Indian Council of Medical Research (ICMR) has now released the results of the May-June national COVID-19 seroprevalence survey. Some data and analysis appears in a paper in the Indian Journal of Medical Research (IJMR) entitled *Prevalence of SARS-CoV-2 infection in India: Findings from the national serosurvey, May-June 2020* (DOI: 10.4103/ijmr.IJMR_3290_20). This paper, with 74 authors, is from now on simply referred to as the "paper" or the "IJMR paper".

The considerable delay in releasing the survey results means that they may be of little use in framing policy decisions around COVID-19 today. But there may be broader lessons to learn from the conduct of the survey, including the errors and shortcomings in the analysis. These lessons could be relevant to the second national serosurvey currently underway.

This document can be seen as a follow up to my preliminary analysis of the serosurvey in June. Unfortunately, if I were to summarise my current view it would be:

> *The serosurvey raises interesting questions. But do not build any firm narratives around its results: there are too many errors, omissions and unquantified uncertainties, and any stories built on these foundations could collapse like a house of cards.*

Summary

- There is a high degree of **uncertainty** on the key claims. Some of the uncertainty is explicitly quantified, some of it is mentioned but does not enter into the given error bounds, and some of it is not even mentioned.
- There remain a number of **errors and ambiguities** to be corrected or clarified.
- There is not enough **transparency** around key assumptions and calculations. This could be improved with supplemental material or a technical appendix.
- As a result of the uncertainties and errors, all claims based on the survey should be taken as *highly* **speculative**.
- Once the uncertainties and errors are acknowledged, some of the survey results move closer to being **plausible and consistent** with other work.
- The paper seems to acknowledge **fatality undercounting**, an important step forward.

In the following two tables, some of the issues/errors are summarised, and details/possible interpretations given.

Table 1. Some issues and apparent errors in the paper. Further details, references and explanation are given in the text.

Issue/error	**Details/interpretation**
Mismatch between % prevalence and absolute prevalence estimates nationally.	The mismatch between the upper and lower limits on prevalence given as a percentage, and given in absolute terms, is quite large. This needs explanation. Possible explanations: different approaches to calculating % prevalence and absolute prevalence? Misinterpretation of adjusted % prevalence in sampled population as prevalence in national population?
Specificity of Kavach test as stated is not consistent with with number of raw positives from Kavach test.	With the stated specificity, a minimum of around 580 raw positives would be expected from the Kavach test applied to 28,000 samples. In fact only 290 raw positives were obtained. This definitely needs discussion. It is possible that the unexpectedly low number of Kavach raw positives prompted the authors to use serial testing? If so, this should be noted transparently along with an updated specificity estimate.
Only point estimates of sensitivity and specificity are used in calculations.	It seems that uncertainties associated with sensitivity and specificity of the tests are not factored into confidence interval calculations. This leads to unquantified – and possibly very large – additional uncertainties. There is some discussion around the key assumption of independence of the two tests, and the authors recognise that it creates an "*upward bias in our prevalence estimate*".
The reader is often left to infer the basis for calculations.	For example, several calculations rely on assumptions about the average time between infection and seroconversion, or infection and death recording, but these are not explicitly detailed or referenced. A lot of calculations cannot be fully reproduced because the required data – for example demographic data – is not given explicitly.
There is a mismatch between infection to case ratios based on case data in the paper, and based on MoHFW case data from the time.	The Ministry of Health and Family Welfare (MoHFW) figures on cases from the time were about 20% less than the figures from ICMR used in the paper. It is likely that there was a significant time-delay in MoHFW case numbers and the ICMR numbers are thus the "correct" ones. Nevertheless, as it is MoHFW's numbers that can be accessed publicly, it is important to note that MoHFW's numbers misrepresented the actual case-load at the time.
There is insufficient detail about the positive test results which could help to make sense of them.	Given the very low total number of positive tests (157 from a total 28,000 samples from 700 villages/wards) more detail could have been given: for example how they were clustered in villages/wards, and the dates when they were obtained. This would give some idea of whether these apparently positive results were likely to be a consequence of detected/undetected outbreaks, or were in fact false positives.

Table 2. Some notable features of the results in the paper, what they might mean if taken at face value, and how they might alternatively be interpreted. Details in the text.

Features of the results as presented in the paper	**Interpretation at face value**	**Alternative interpretation assuming prevalence overestimation**
Prevalence estimates are surprisingly high	High undetected early spread. Lockdown was very leaky leading to infection spreading out from hotspots.	Prevalence was overestimated as a consequence of incorrect estimation of combined test-kit properties. Especially, perhaps, if the assumption of independence of Kavach and Euroimmun tests was incorrect.
Very high infection to case ratio	Very poor early detection of infections. Disease surveillance in many areas had not caught up with the disease. If so, this has (most likely) subsequently improved significantly. Narratives at the time about effective control of COVID-19 were then entirely misleading.	Apparently poor detection is at least partly because prevalence is overestimated. Early case detection was still low, but closer to what we would estimate today from various later serosurveys.
Little variability in apparent prevalence between low and high case-load districts	Case detection was extremely variable. There were some regions where disease was spreading silently. For example, about 0.68% of people had been infected in districts where no cases had been reported by April 25.	There was relatively low prevalence in all strata, but there was variation by stratum. However, worse than estimated sensitivity and specificity led to a high signal to noise ratio: the testing strategy could not pick up differences in prevalence.
Low infection fatality rate, taking the data as a whole. Closer to other serosurvey values in high case-load stratum.	Authors interpretation: poor death surveillance in lower case-load strata – only the values from the high case-load stratum are reliable.	A mixture of fatality undercounting and prevalence overestimation. The latter would have been particularly high in the low and zero case-load strata if specificity was lower than assumed.

The headline prevalence claim, and a discrepancy between estimates

The headline claim of the study was – and remains – that an estimated 0.73% of the adult population had been infected with COVID-19 by the time of the survey. This figure was reported in June and reappears in the IJMR paper but with an important difference: a 95% confidence interval (CI) of 0.34%-1.13% is now given on this figure, indicating that it comes with high uncertainty. But this CI seems only to take into account certain kinds of uncertainty and so underestimates the uncertainty associated with this prevalence estimate.

Before discussing uncertainties in detail we'll examine the estimates at face value. But we immediately encounter a first difficulty. There is a **contradiction** between the following two lines in the IJMR paper:

1. "*The pooled adjusted prevalence of SARS-CoV-2 infection was 0.73 per cent (0.34-1.13%) at the national level*", and
2. "*A cumulative 6,468,388 adult infections (95% CI: 3,829,029-11,199,423) were estimated in India by the early May*".

If 0.73% prevalence equates to 6.47M adult infections, which seems about right, then 0.34% prevalence should equate to about 3M adult infections (not 3.8M); and 1.13% prevalence should equate to about 10M adult infections (not 11.2M). This discrepancy at least needs clarification.

From here on we'll use 3.8M and 11.2M as the authors' lower and upper estimates of adult prevalence nationally, even though these are inconsistent with the percentage estimates. It is possible that the percentage estimates in fact applied only to the population in the sampled districts and not nationwide, but because the calculations are not presented it is hard to know.

Estimates for the population as a whole, including children

In order to compare the ICMR estimates with other estimates and with data from other studies, we first need to recall that only adults (18 years and over) were surveyed. A prevalence of 0.73% by the time of the survey translates into an estimated 6.5 million adults nationwide if we assume that 66% of the national population are aged 18 or over.

There is little reason to believe that children are less likely to be infected than adults. Thus an estimate of 6.5M adults would imply almost 10M people across the country infected at the time. The lower estimate of 3.8M infected adults would translate to approximately 5.8M infected people across the country, while the upper estimate of 11.2M infected adults would translate as approximately 17M infected people across the country at the time.

At what time point do the prevalence estimates apply?

In the original press release it was claimed that the estimated prevalence was for late April. But, note that the survey was carried out between May 11 and June 4 with survey midpoint May 23. The authors of the IJMR paper suggest that their prevalence estimate applies to May 3 which appears to assume a mean time from infection to IgG seroconversion of 20 days. In fact, given high variability in the speed of IgG seroconversion, and the fact that the samples were being taken in the context of rapid growth in case numbers nationwide, the prevalence estimate might best apply to an even later date.

The question of when the prevalence estimates apply is important because recorded COVID-19 cases were doubling every 10-11 days or so at the time. So true infections could have doubled between April 30 and May 10. In this context it would be interesting to know the dates associated with the 157 positive samples.

Are the prevalence estimates plausible?

I'd argued earlier that the prevalence estimate of 0.73% nationally exceeded even the upper end of data-driven estimates of prevalence nationally for late April. This is in fact true even for the lower end of the confidence interval given in absolute terms, namely about 5.8M infections. But, this lower value just about lies within the range of values predicted for somewhat later. For example, if the estimate applies to May 3, then it is consistent with an early infection fatality rate of about 0.2% followed by about 60% of expected fatalities going "missing" by May 3, either as a result of a genuine drop in infection fatality rate (IFR) as disease moved to a younger population, or fatality undercounting, or both. If the estimate applies to May 10, then it is consistent with an early infection fatality rate (upto about April 10) of about 0.3% followed by about 55% of expected fatalities going "missing" by May 10.

These are, of course, just examples to indicate that the lowest estimates could be plausible, and this plausibility is quite strongly dependent on the dates to which they apply. We'll return to the question of plausibility when we discuss the detection of infections and fatality rates in greater detail. But first it is time to examine the uncertainties on the prevalence estimates a bit more carefully.

Uncertainties associated with the test kits

Accurately estimating a prevalence percentage as small as 0.73% is hard. Basic theory on confidence intervals for a proportion tells us that we need to take large sample sizes – and precisely how large depends on (i) how low the prevalence is, and (ii) how accurate we hope to be. Intuitively, if only 0.5% of people in a population have had COVID-19 and you sample 2000 people, you should "typically" find 10 who have had COVID-19. But it is fairly likely that you will actually find only 5, or indeed, 20 people in the sample to have had the disease. Thus the sample proportion may significantly underestimate or overestimate the true population proportion.

These kinds of uncertainties *are* taken into account in the IJMR paper, and appear to be the source of the wide confidence intervals on prevalence estimates. But there are other sources of uncertainty which are treated incompletely, and these concern the properties of the test-kits themselves.

Firstly, there is an **inconsistency** in the paper: the quoted value of specificity for the COVID Kavach-Anti-SARS-CoV-2 IgG Antibody Detection ELISA (henceforth just "Kavach" for short) of 97.9%, would imply a minimum of 0.021*28,000 = 588 raw positives from the Kavach test, even in a zero prevalence situation, whereas in fact only 290 raw positives were obtained. Such an occurrence would be highly unlikely: thus the measured data suggests that the test kit – in this context at least – had a higher specificity than the quoted value. In fact, an independent evaluation of the Kavach kit arrived at a value for specificity of 99.5%. This inconsistency is not even mentioned in the paper.

Setting aside this inconsistency, the study did recognise that in a low prevalence environment, minor inaccuracies in specificity can lead to major inaccuracies in the estimated prevalence. To mitigate against this, **serial testing** of possible positive samples was used: a sample was only accepted as positive if it tested positive using not just Kavach but, subsequently, another test kit, the

Euroimmun SARS-CoV-2 ELISA (IgG) (henceforth "Euroimmun" for short). It is not clear whether this serial testing strategy and the use of Euroimmun were part of the original study design. It is not mentioned in the original ICMR press release which refers only to Kavach: "*Sera... will be tested for presence of IgG antibodies using ELISA test developed by ICMR-National Institute of Virology (NIV), Pune*". Meanwhile the original protocol description states only that, "*Detection of SARS-CoV-2-specific IgG antibodies will be performed using an ELISA-based test...*"

Whether it was pre-planned or a response to the unexpectedly low number of raw Kavach positives, this serial testing of samples renders the specificity of each test-kit less important; it is now the *combined* specificity achieved by using them sequentially that matters. This value would be high whether, for example, Kavach had a specificity of 97.9% or 99.5% provided – and this is a very important caveat – the two test-kits were **independent**. If this is the case then the probability of a negative sample being incorrectly classified by Kavach as positive, and then again incorrectly classified by Euroimmun as positive, would be the product of the probabilities of each of these events, and so very small.

To be more specific, the final number of samples confirmed to be positive using both test kits was 157. If the test kits were truly independent, then with high probability no more than one or two of these would actually be negative samples incorrectly classified as positive. But if the independence assumption was incorrect, then many more of the 157 positives could in fact be false positives. The authors do note that if the assumption of independence is wrong, then prevalence estimates could be as low as 0.26%. But such considerations do not seem to enter into the calculation of confidence intervals presented upfront.

Turning to sensitivity, the quoted values of sensitivity used by the authors do not match other published estimates. In fact, if one scours the literature for values of sensitivity and specificity of commonly used antibody tests one often finds divergent values. For example, the sensitivity of the Kavach test is given in the IJMR paper as 92.4%, but another evaluation found it to be considerably lower at 75.7% (95% CI: 71.0% to 79.9%). The IJMR study used values for sensitivity and specificity of the Euroimmun assay as per the kit insert (93.8% and 99.6%), but these diverge somewhat from values given in another evaluation: 89.5% (95% CI: 75.3% to 96.4%) for sensitivity and 96.1% (95% CI: 90.1% to 98.8%) for specificity. Yet another study quotes the sensitivity of the Euroimmun IgG ELISA as only 65%. No doubt there are other evaluations which give different results again.

The point is that the values of these test-kit properties are uncertain, and often in fact seem to depend on details of the sample set used for testing. Using **point estimates for sensitivity and specificity** in calculations when there is actually high uncertainty can lead to errors whose magnitude is *a priori* hard to quantify.

Thus although the prevalence value of 0.73% was quoted with a wide 95% CI, the true uncertainty in this value could be much higher if uncertainty in sensitivity and specificity of the tests was factored in.

Case detection: infection to case ratios

The prevalence estimates lead to estimates for the proportion of infections which had been detected through testing. The authors find that for every case confirmed via RT-PCR there were 82 adult infections (using case data from May 11, namely 12 days before the survey midpoint) or 130 adult infections (using case data from May 3, namely 19 days before the survey midpoint).

Yet another **confusion/discrepancy** arises at this point. It has been noted that the infection to case ratio estimates use values for RT-PCR confirmed cases which differ signficantly from MoHFW's values made publicly available on their dashboard every day. Using MoHFW's numbers would increase the infection to case ratios by about 20%. We assume that the numbers quoted in the paper are the most reliable (although these are not publicly available). This does indicate, however, that the MoHFW dashboard at the time underplayed the level of spread.

What is the range of possible infection to case ratios that can be obtained from the ICMR study data? Changing "adult infections" to just "infections" would multiply the given ratios by approximately 1.5. Using prevalence estimates at the lower end of the 95% CI in the IJMR study would approximately halve these infection to case ratio, and conversely using the upper prevalence estimate would approximately double the ratio. Putting all this together, and using the data from the paper and values given for confirmed infections by May 11, the infection to case ratio across the population would be 124 (95% CI: 73-215). Even the most optimistic value at the lower end of the confidence interval tells us that disease surveillance was poor by early May.

What would these ratios mean today (mid-September) when there are about 5 million cases nationwide? An infection to case ratio of 124 today would mean that almost half the Indian population has been infected! Even taking the lowest value of 73 and applying it now would mean that more than a quarter of the population has been infected to date. These values are implausible based on more recent serosurvey results and current fatalities – even assuming large undercounting.

Are these detection levels plausible for early May? Perhaps. Generally, detection plays catch-up as infection spreads rapidly under the radar in the early days of an epidemic. Mumbai's serosurvey indicated detection of under 2%, but modelling indicated much poorer detection early on; and it is possible that detection in a major city would be significantly higher than the national average.

Thus the estimates of very poor infection detection do not, in themselves, discredit the prevalence estimates; but they provide another reason to be cautious about these estimates. We can say with some confidence that either the lower bound on prevalence from the serosurvey is an overestimate and/or infection detection has greatly improved since early May.

Variation in prevalence and detection between regions

Apart from generally poor disease detection, the serosurvey, taken at face value, indicates highly *variable* detection between different regions. One of the most surprising findings was very *little* variation in prevalence between areas with low and high case-loads.

The survey divides the districts surveyed into four "strata", with some variation in the numbers of people tested in each stratum. (The original design intended equal sampling in each stratum.) According to the original survey protocol these strata consist of districts with zero case-load, low case-load (0.1-4.7 cases per million population), medium case-load (4.8-10 cases per million), and high case-load (>10 cases per million). According to the IJMR paper, the decision about stratification was made based on reported cases by April 25.

The IJMR paper gives adjusted prevalence values for the four strata, with 95% CIs, as follows:

- zero case-load: 0.68% (0.42%-1.11%)
- low case-load: 0.62% (0.43%-0.89%)
- medium case-load: 1.03% (0.44%-2.37%)
- high case-load: 0.72% (0.44%-1.17%)

Note the insignificant difference in estimated prevalence in the zero case-load and high case-load strata!

If we accept these figures, then infection was spreading widely despite lockdown, and entirely undetected in some areas. Case detection was not just poor overall, but extremely *variable*. This is not entirely impossible: we know from Mumbai (and Pune) serosurveys that reported cases do not necessarily track infections: detection of infections in marginalised communities is likely to be poorer than the average. There could also plausibly be a rural/urban divide in detection.

Given how few samples tested positive in both tests (157 in a total of 700 villages/wards) more detail could have been provided as a sanity check. If, for example, they were clustered in a few wards/villages in the low- and zero-prevalence strata, this could indicate undetected outbreaks. The dates on which the positive samples were obtained in the different strata could also give clues about whether the positives were a result of infection spreading after the districts were classified into strata.

But while we shouldn't expect a strong correlation between cases and infections, the complete lack of correlation again reminds us to treat the prevalence estimates with scepticism. If serial sensitivity was lower than assumed, then a significant fraction of positive results could have been missed. If serial specificity was lower than estimated as a consequence of the two test-kits not being independent, a large number of negative results could potentially have been classified as positive. With such effects in play, the "**signal-to-noise ratio**" in the data could be very low, resulting in estimated prevalence in the different strata not correctly tracking actual prevalence.

Infection fatality rate: what were the estimates and are they plausible?

One number widely quoted in June, but which does not appear in the final IJMR paper, is a value of 0.08% for the IFR of COVID-19 nationally. It is impossible, without guesswork, to trace the origins of this figure. It could have been arrived at by taking all deaths reported in the medium and high prevalence strata by June 1 and dividing by the adult infections in these strata estimated from the serosurvey.

The IJMR paper in fact reports different IFR values for different strata. In the high prevalence stratum, assuming no undercounting of deaths, adult IFR estimates of 0.11% and 0.15% are obtained depending on the date chosen for recorded deaths. These values come with wide CIs associated with the CIs on prevalence.

The IFR estimates take as the numerator deaths measured on May 24, one day after the survey mid-point, and June 1, nine days after the survey midpoint. Given what we now know about extensive delays in death recording from various "reconciliations", it seems sensible to take the higher estimate to lessen "right censoring", namely taking fatality data too early, and hence missing a significant fraction of COVID-19 deaths associated with the estimated infections in the denominator.

Taking the estimated infected population as a whole, including children, in the denominator would lower the higher estimate of 0.15% to about 0.10%. This is roughly consistent with naive IFR values from a number of city serosurveys. For example: Delhi reported a naive IFR of about 0.07%, Mumbai of about 0.12%, and Ahmedabad of about 0.14% at the times of their serosurveys.

We know that two factors very seriously affect the reliability of IFR estimates: (i) fatality undercounting and/or reporting delays, and (ii) uneven spread in populations of different ages. For

example, at the time when the Mumbai serosurvey gave a naive IFR estimate of 0.12%, the true value, assuming that about 70-80% of the city's excess deaths not attributed to COVID were in fact COVID deaths, could have been roughly 0.22%. Taking into account variable spread in different subpopulations, we get plausible values of IFR varying between 0.18% and 0.32% over the duration of the epidemic so far, with the lower values occurring at the time when the majority of spread was amongst a younger slum-dwelling population.

Thus the higher naive IFR estimate for the high case-load stratum becomes plausible if coupled with fatality undercounting. But, once again, given the huge uncertainties in the prevalence estimates, the IFR estimates should also be taken with a pinch of salt.

Acknowledgement of significant fatality undercounting?

The IJMR paper suggests that only IFR values for the high case-load stratum where "*death reporting was more robust*" should be regarded as reliable. I had earlier noted that the seroprevalence and fatality data at face value gave an even lower naive IFR than the reported value at the time (0.08%), implying that the calculations could be factoring in some death undercounting. This is borne out by the IJMR paper.

In the districts with low or zero case-load, very low IFR values – down to about 0.002% – were calculated but disregarded by the authors. Such low values should not be surprising – after all, if surveillance was failing to pick up COVID infections, it would be unlikely to pick up COVID deaths. The low values could also, at least partly, be a consequence of considerable overestimation of prevalence in these strata.

If we were to use the deaths reported across all strata by June 1, then we get a naive adult IFR nationally of 0.055%, as compared to 0.15% in the high case-load stratum alone. By choosing the higher value, the ICMR appears to have acknowledged that almost two in three deaths could have been missed nationally by June 1. This is within the range of fatality undercounting for early June given by data-driven estimates.

Incidentally, the authors chose results from some problematic international studies to contextualise their fatality rate estimates: the deeply flawed Santa Clara study, and an estimate from Iran which is very likely to have greatly undercounted COVID-19 deaths. It would have been much more helpful to contextualise the estimates using meta-analyses of COVID-19 IFR and, in particular, analyses of COVID-19 fatality by age which could have been applied to give an estimated IFR for India based on India's age-structure.

Conclusions

The main conclusion must be that any claims made based on results of the national serosurvey are built on *very* shaky foundations. The multiple uncertainties, only some factored in by the authors, coupled with errors and omissions, mean that every claim needs to be accompanied by extensive caveats.

Because of the uncertainties not taken into account, the serosurvey does *not* provide compelling evidence to discount the hypothesis that lockdown largely confined disease to a few hotspots, and case detection nationally was low but comparable to, say, the values observed in city serosurveys. On the other hand, we can't discount that early spread was much wider than assumed, and disease and death surveillance were essentially absent from some areas.

Here are some broad conclusions and recommendations.

- More care and transparency was needed from the ICMR in reporting and analysing the data. This was especially true of the misleading press briefings in June, but remains true to some extent even after the IJMR paper. Different sources of uncertainty and error need to be taken into account. Assumptions need to be explained clearly and referenced.
- Various relatively minor, but confusing, discrepancies need to be fixed or explained, for example between the percentage prevalence values and the absolute prevalence values at national level.
- The lowest estimates of national prevalence are close to being consistent with modelling and data-driven work. This does not, of course, make them correct.
- The estimates at face value suggest very poor infection detection at best. If this held today it would imply more than 25% prevalence nationwide. If we believe the estimates, then it is likely that infection detection has much improved since the first national serosurvey: this should become clearer when the results of the second national serosurvey are released.
- The naive IFR estimates using the data as a whole are implausible; however those from the high prevalence strata alone are consistent with estimates from later serosurveys. All such estimates need to be taken alongside estimates of fatality underreporting.
- It is a welcome development that the ICMR appears to acknowledge likely COVID-19 death underreporting on a very significant scale.

Appendix. Technical summary of the serosurvey results as reported in IJMR article

1) Serosurvey carried out between 11/05 and 4/06 (mid-point May 23)
2) 28,000 adults tested from 700 villages/wards; 400 adults in each
3) Response rate: >90%
4) 290 tested positive with the Kavach ELISA, of whom 157 tested positive after further testing with Euroimmun ELISA.
5) Raw seroprevalence with both tests positive: 0.56% (95% CI: 0.48%-0.66%)
6) Little variation in raw prevalence: from 0.47% in strata with zero reported cases, to 0.74% in strata with medium reported cases. 0.59% in strata with high recorded cases.
7) Serial sensitivity and specificity (Kavach+Euroimmun) given as 86.67% and 99.99% respectively. The latter value assumes complete independence of the tests.
8) Seroprevalence nationally after adjusting for sampling weights and using serial sensitivity and specificity values: 0.73% (95% CI: 0.34%-1.13%).
9) An estimated 6.5M (95% CI: 3.8M to 11.2M) adults infected nationally. Discrepancy between these figures and the previous percentages is not discussed.
10) Naive adult infection fatality rate (IFR) – assuming no undercounting – from strata where "death reporting was more robust": 0.12% using a delay of 1 day from survey mid-point (95% CI: 0.07% to 0.19%) to 0.15% using a delay of 9 days from survey mid-point (95% CI: 0.09% to 0.24%). Using the later date and including children: naive IFR from the high case-load stratum: about 0.10%.
11) Naive adult IFR from all strata combined using deaths on June 1 (9 days after the survey mid point): about 0.055%.