Common sense science for epidemics

Sourendu Gupta

TIFR Colloquium 09 September, 2020

Discussions, corrections, and lessons from/with DTP, TIFR, and ISRC. RPG and her network of microbiologists and virologists across the world.





Dimensional analysis and classical mechanics



Gloom and doom?



State/UT		Confirmed	Active	Recovered	Deceased	Tester
Maharashtra	0	8,63,062	2,10,978	6,25,773	25,964	190.4H
Andhra Pradesh	0	4,76,506	1,02,067	3,70,163	4,276	39.7
Tamil Nadu	0	4,51,827	51,633	3,92,507	7,687	51.3
Karnataka		3,79,486	99,101	2,74,196	ó,170	32
Uttar Pradesh		2,53,175	58,595	1,90,818	3,762	62
Delhi	0	1,85,220	18,842	1,61,865	4,513	17.1
West Bengal		1,74,659	23,654	1,47,553	3,452	20.71
Bihar		1,44,134	16,981	1,26,411	741	37.21
Telangana	0	+2,511 1,38,395	32,915	\$2,579 1,04,603	†11 877	162.1



Credulity about numbers

Coronavirus growth rate declines 40% post lockdown, giving India hope

Country's average growth of between March 15 and 31 was 2 per cent. Post lockdown, in the last two weeks the growth rate has come down to 1.2 per cent.

Topics

Coronavirus | Health Ministry

Ruchika Chitravanshi | New Deihi Last Updated at April 18, 2020 01:49 IST

Business Standard, 18 April 2020



Times of India, 21 April 2020

Home / Office / Mumbel / Maharashtra's Cavit growth rate falls below India's for the first ten

Maharashtra's Covid growth rate falls below India's for the first time

The seven-day compounded daily growth rate (CDGR) for Maharashtra has remained below the national average for three consecutive days now.

Indian Express, 3 June 2020

Science from Data

Natural science is not data science!

Science from Data

Natural science is not data science!

Natural sciences study objects and phenomena; what to ask in order to understand what we see; what to measure in order to answer a question.

Most TIFR seminars and questions are about not about the numbers, but about how you get to the numbers. Perhaps the most important step in experimental sciences is to understand errors; before you make a measurement you study the measuring instrument. Data is the last step.

Science from Data

Natural science is not data science!

Natural sciences study objects and phenomena; what to ask in order to understand what we see; what to measure in order to answer a question.

Most TIFR seminars and questions are about not about the numbers, but about how you get to the numbers. Perhaps the most important step in experimental sciences is to understand errors; before you make a measurement you study the measuring instrument. Data is the last step.

Richard Hamming

The purpose of computing is insight, not numbers.

Public health versus science

The goal of a public health crisis manager is not that of a scientist. Public health data may be intentionally biased to study the most important cases.

In a crisis, data collection is in the domain of public health. Its use for science needs great care.

Public health versus science

The goal of a public health crisis manager is not that of a scientist. Public health data may be intentionally biased to study the most important cases.

In a crisis, data collection is in the domain of public health. Its use for science needs great care.

Examples of bad science

When will the COVID-19 crisis end? Until now, many predictions: ranging from April 3 to September 10. Some even continually updated, so that model predictions have changed by several months over several months!

Incorrect understanding of input data. Insufficient attention to the meaning of models.

Precise is not accurate

Number of detected infections on day t is I(t). Suppose this is growing exponentially: $I \sim 2^{t/T}$.

Then doubling time is

$$rac{1}{T} = rac{1}{\log 2} imes \log \left[rac{I(t+1)}{I(t)}
ight]$$

One day the doubling time is 5 days, the next day it is 8 days. Breaking news!

Precise is not accurate

Number of detected infections on day t is I(t). Suppose this is growing exponentially: $I \sim 2^{t/T}$.

Then doubling time is

$$rac{1}{T} = rac{1}{\log 2} imes \log \left[rac{I(t+1)}{I(t)}
ight]$$

One day the doubling time is 5 days, the next day it is 8 days. Breaking news!

Scientific skepticism: can you distinguish a growth period of 5 from 8 days with one day's data? What accuracy do you need?

$$100 \times 2^{1/5} = 114.9$$
 $100 \times 2^{1/8} = 109.0$

Require errors in I(t) to be less than 5%. Does I have this accuracy?

What error on measurement do you expect from simple Poisson statistics?

What error on measurement do you expect from simple Poisson statistics?

In a city of 20 million, with (say) 10,000 people involved in disease surveillance, what are the chances of identifying every case?

What error on measurement do you expect from simple Poisson statistics?

In a city of 20 million, with (say) 10,000 people involved in disease surveillance, what are the chances of identifying every case?

What is the fraction of asymptomatic cases?

What error on measurement do you expect from simple Poisson statistics?

In a city of 20 million, with (say) 10,000 people involved in disease surveillance, what are the chances of identifying every case?

What is the fraction of asymptomatic cases?

Do swabs taken by health workers on the ground always reach RT-PCR labs without delay? Has employee efficiency become 100% during the epidemic?

What error on measurement do you expect from simple Poisson statistics?

In a city of 20 million, with (say) 10,000 people involved in disease surveillance, what are the chances of identifying every case?

What is the fraction of asymptomatic cases?

Do swabs taken by health workers on the ground always reach RT-PCR labs without delay? Has employee efficiency become 100% during the epidemic?

Are all RT-PCR labs working with equal efficiency every day? No employee takes leave now and then, or comes in late?

What error on measurement do you expect from simple Poisson statistics?

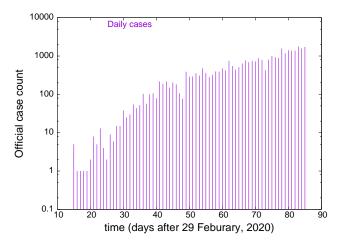
In a city of 20 million, with (say) 10,000 people involved in disease surveillance, what are the chances of identifying every case?

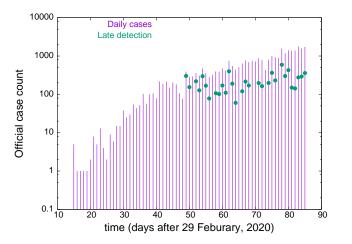
What is the fraction of asymptomatic cases?

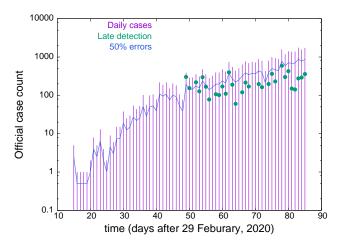
Do swabs taken by health workers on the ground always reach RT-PCR labs without delay? Has employee efficiency become 100% during the epidemic?

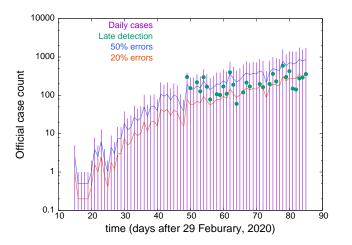
Are all RT-PCR labs working with equal efficiency every day? No employee takes leave now and then, or comes in late?

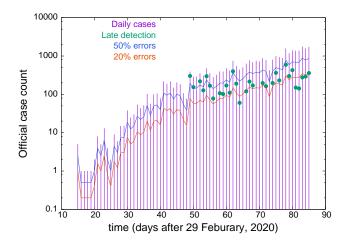
As the number of cases increases, the number of samples to be tested also increases? Is the testing process scaling up with perfect efficiency?



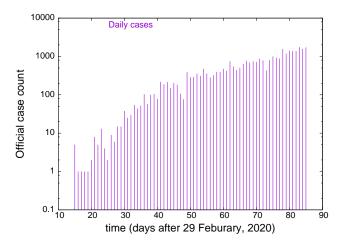


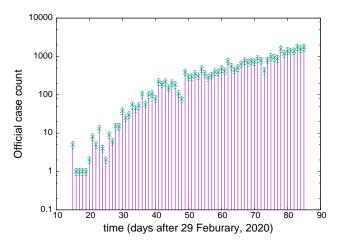


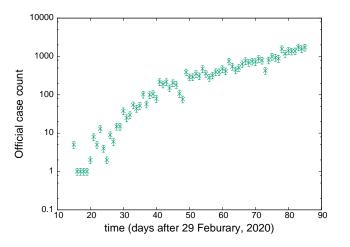


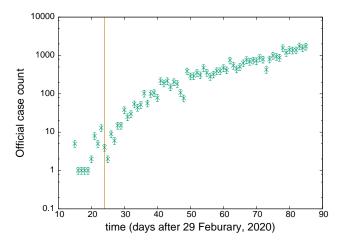


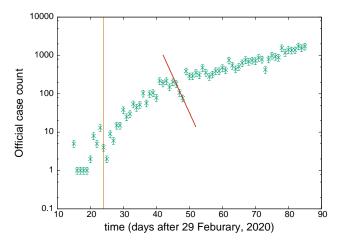
Errors due to delays, about 20-25% of the case numbers. Errors due to detection?





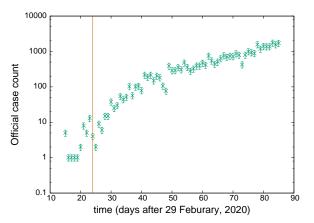






Odd behaviour

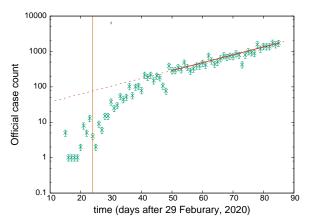
If lockdown causes flattening, then why does it come into play after 25 days? COVID-19 serial time is 3–5 days*. Why flattening after 6–8 generations of infections?



*: Nishiura, Linton, Akhmetzhanov, Int J Infect Dis (2020) 93:284

Odd behaviour

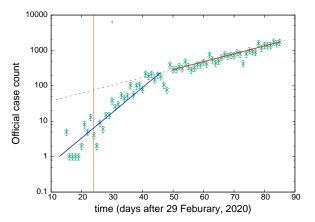
If lockdown causes flattening, then why does it come into play after 25 days? COVID-19 serial time is 3–5 days*. Why flattening after 6–8 generations of infections?



*: Nishiura, Linton, Akhmetzhanov, Int J Infect Dis (2020) 93:284

Odd behaviour

If lockdown causes flattening, then why does it come into play after 25 days? COVID-19 serial time is 3–5 days*. Why flattening after 6–8 generations of infections?



*: Nishiura, Linton, Akhmetzhanov, Int J Infect Dis (2020) 93:284

Resolving the puzzle

Compared 6 cities till mid-April: Ahmedabad, Chennai, Delhi, Indore, Mumbai, Pune. Cases rose in all cities with steadily changing doubling time:

$$I(t) \sim 2^{t/(T_0+tT_1)}$$

In all cases ${\it T}_0 \sim 1$ day, and ${\it T}_1 \sim$ 0.15–0.20 days. Why this adaptation?

Resolving the puzzle

Compared 6 cities till mid-April: Ahmedabad, Chennai, Delhi, Indore, Mumbai, Pune. Cases rose in all cities with steadily changing doubling time:

$$I(t) \sim 2^{t/(T_0+tT_1)}$$

In all cases ${\it T}_0 \sim 1$ day, and ${\it T}_1 \sim$ 0.15–0.20 days. Why this adaptation?

Another oddity: in the early days, Pune epidemic seeded by Mumbai. Most cases traced to contact in Mumbai or travel from Mumbai. Still cases rose faster in Pune than Mumbai.

Resolving the puzzle

Compared 6 cities till mid-April: Ahmedabad, Chennai, Delhi, Indore, Mumbai, Pune. Cases rose in all cities with steadily changing doubling time:

$$I(t) \sim 2^{t/(T_0+tT_1)}$$

In all cases ${\it T}_0 \sim 1$ day, and ${\it T}_1 \sim$ 0.15–0.20 days. Why this adaptation?

Another oddity: in the early days, Pune epidemic seeded by Mumbai. Most cases traced to contact in Mumbai or travel from Mumbai. Still cases rose faster in Pune than Mumbai.

Very likely to be a systematic error in the measurement. I attributed it to **disease surveillance network being built up**. SG, arxiv:2005.08499

Alternative explanations, with this change fitted to epidemic models gave predictions which were proved wrong by the end of May.

Remaining unknowns

- Sero-prevalence of antibodies (TIFR study + others) show that official disease surveillance catches one or two orders of magnitude less of infected than actual. Why?
- CFR (case fatality ratio) is the ratio of deaths due to COVID-19 and the number of detected cases. Observed CFR is currently around 5.8%. IFR (infection fatality ratio) is the ratio of deaths and the actual number of actual infections. Using official figure for deaths, and sero-data for infections, IFR is less than 0.1%.
- If cases underestimated by factor of 50, then how accurately are deaths due to COVID-19 estimated?

Remaining unknowns

- Sero-prevalence of antibodies (TIFR study + others) show that official disease surveillance catches one or two orders of magnitude less of infected than actual. Why?
- CFR (case fatality ratio) is the ratio of deaths due to COVID-19 and the number of detected cases. Observed CFR is currently around 5.8%. IFR (infection fatality ratio) is the ratio of deaths and the actual number of actual infections. Using official figure for deaths, and sero-data for infections, IFR is less than 0.1%.
- If cases underestimated by factor of 50, then how accurately are deaths due to COVID-19 estimated?

There are possible autocorrelations in the data of 3–4 days. Is this due to delays in the detection of new clusters of infections? Or is it due to clusters of infections in health workers?



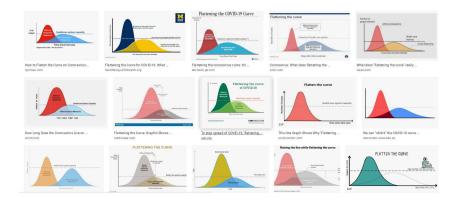




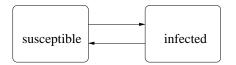


Puzzles

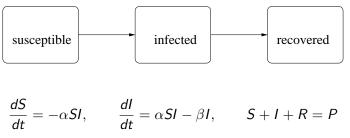
How should we maximize our own chances of survival?



What are these curves? What remains fixed, what is changed?







 α is the rate of infections, 1/ β is the mean time of recovery. For COVID-19 1/ $\beta\simeq$ 2 weeks.



$$\frac{dS}{dt} = -\alpha SI, \qquad \frac{dI}{dt} = \alpha SI - \beta I, \qquad S + I + R = P$$

 α is the rate of infections, $1/\beta$ is the mean time of recovery. For COVID-19 $1/\beta\simeq 2$ weeks.

Dimensional analysis: Choose unit of time as $1/\beta$. $\alpha \rightarrow \alpha/\beta$



$$\frac{dS}{dt} = -\alpha SI, \qquad \frac{dI}{dt} = \alpha SI - I, \qquad S + I + R = P$$

 α is the rate of infections, $1/\beta$ is the mean time of recovery. For COVID-19 $1/\beta\simeq$ 2 weeks.

Dimensional analysis: Choose unit of time as $1/\beta$. $\alpha \rightarrow \alpha/\beta$



$$\frac{dS}{dt} = -\alpha SI, \qquad \frac{dI}{dt} = \alpha SI - I, \qquad S + I + R = P$$

 α is the rate of infections, $1/\beta$ is the mean time of recovery. For COVID-19 $1/\beta \simeq 2$ weeks. Dimensional analysis: Choose unit of time as $1/\beta$. $\alpha \rightarrow \alpha/\beta$

Scaling: $P \rightarrow \lambda P$, $S \rightarrow \lambda S$, $I \rightarrow \lambda I$, $R \rightarrow \lambda R$.



$$\frac{dS}{dt} = -\alpha SI, \qquad \frac{dI}{dt} = \alpha SI - I, \qquad S + I + R = P$$

 α is the rate of infections, $1/\beta$ is the mean time of recovery. For COVID-19 $1/\beta \simeq 2$ weeks. Dimensional analysis: Choose unit of time as $1/\beta$. $\alpha \to \alpha/\beta$

Scaling: $P \rightarrow \lambda P$, $S \rightarrow \lambda S$, $I \rightarrow \lambda I$, $R \rightarrow \lambda R$.

Choose $\lambda = 1/P$. Then $0 \le S, I, R \le 1$. Also, in this unit, the rate constant α is called R_0



$$\frac{dS}{dt} = -R_0 SI, \qquad \frac{dI}{dt} = R_0 SI - I, \qquad S + I + R = 1$$

 α is the rate of infections, $1/\beta$ is the mean time of recovery. For COVID-19 $1/\beta \simeq 2$ weeks. **Dimensional analysis**: Choose unit of time as $1/\beta$. $\alpha \to \alpha/\beta$ **Scaling**: $P \to \lambda P$, $S \to \lambda S$, $I \to \lambda I$, $R \to \lambda R$.

Choose $\lambda = 1/P$. Then $0 \le S, I, R \le 1$. Also, in this unit, the rate constant α is called R_0



$$\frac{dS}{dt} = -R_0 SI, \qquad \frac{dI}{dt} = R_0 SI - I, \qquad S + I + R = 1$$

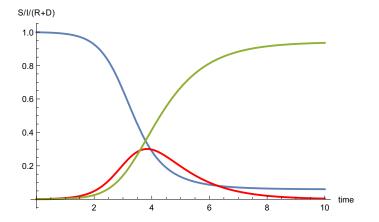
 α is the rate of infections, 1/ β is the mean time of recovery. For COVID-19 1/ $\beta\simeq$ 2 weeks.

Dimensional analysis: Choose unit of time as $1/\beta$. $\alpha \to \alpha/\beta$ Scaling: $P \to \lambda P$, $S \to \lambda S$, $I \to \lambda I$, $R \to \lambda R$.

Choose $\lambda = 1/P$. Then $0 \le S, I, R \le 1$. Also, in this unit, the rate constant α is called R_0

Terminology: R_0 is the basic reproduction rate. Sometimes $1/\beta$ is called the recovery time.

Numerical solution



$$\dot{S} = -R_0 SI, \qquad \dot{I} = (R_0 S - 1)I$$

$$\dot{S} = -R_0 SI, \qquad \dot{I} = (R_0 S - 1)I$$

Stable/unstable points: $\dot{S} = 0 = \dot{I}$. Only needs I = 0

$$\dot{S} = -R_0 SI, \qquad \dot{I} = (R_0 S - 1)I$$

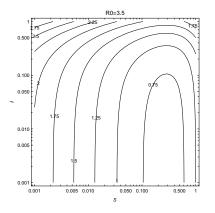
Stable/unstable points: $\dot{S} = 0 = \dot{I}$. Only needs I = 0

Maximum of *I*: $S = 1/R_0$.

$$\dot{S} = -R_0 SI, \qquad \dot{I} = (R_0 S - 1)I$$

Stable/unstable points: $\dot{S} = 0 = \dot{I}$. Only needs I = 0

Maximum of *I*: $S = 1/R_0$.



R0=3.5

0.050 0.100

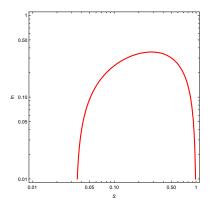
0.500

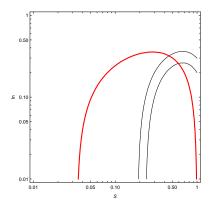
Phase space picture

$$\dot{S} = -R_0 SI, \qquad \dot{I} = (R_0 S - 1)I$$
Stable/unstable points:

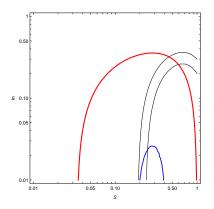
$$\dot{S} = 0 = \dot{I}. \text{ Only needs } I = 0$$
Maximum of I: $S = 1/R_0.$

The epidemic equations can be derived from a Lagrangian: classical mechanics. Go to Hamiltonian: symplectic structure, KAM tori. Many new and interesting questions, not asked before.

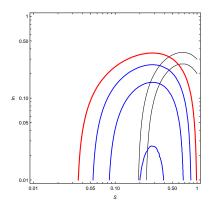




Your conceptual framework determines imaginable solutions

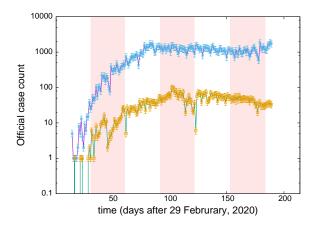


Your conceptual framework determines imaginable solutions When all you have is a hammer, everything looks like a nail



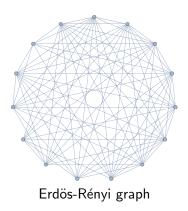
Your conceptual framework determines imaginable solutions When all you have is a hammer, everything looks like a nail

Something unexpected

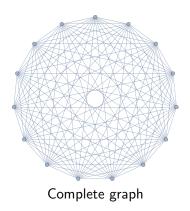


Three month long plateau. Serial time is 3–4 days, recovery time is 10–20 days. So major mismatch of scales. Dynamics at work, but plateau not expected in compartmental models.

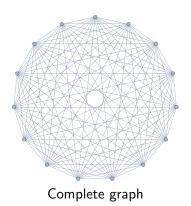
Mumbai in partial lockdown now: many housing complexes gate off visitors, meetings between people in different complexes restricted. Model this by incomplete mixing in the reaction equations. Difference from physical sciences: non-local interactions allowed. Resembles reaction-diffusion equations on a network.



Mumbai in partial lockdown now: many housing complexes gate off visitors, meetings between people in different complexes restricted. Model this by incomplete mixing in the reaction equations. Difference from physical sciences: non-local interactions allowed. Resembles reaction-diffusion equations on a network.

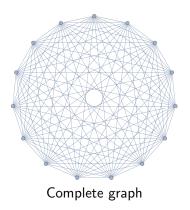


Mumbai in partial lockdown now: many housing complexes gate off visitors, meetings between people in different complexes restricted. Model this by incomplete mixing in the reaction equations. Difference from physical sciences: non-local interactions allowed. Resembles reaction-diffusion equations on a network.



Fixed points remain as before: I = 0. But stable long-term solution with slowly decreasing plateaus arise without any fine tuning of parameters.

Mumbai in partial lockdown now: many housing complexes gate off visitors, meetings between people in different complexes restricted. Model this by incomplete mixing in the reaction equations. Difference from physical sciences: non-local interactions allowed. Resembles reaction-diffusion equations on a network.



Fixed points remain as before: I = 0. But stable long-term solution with slowly decreasing plateaus arise without any fine tuning of parameters. Robust construction: describes Mumbai's plateau with 2 parameters. SG TIFR/TH/20-34

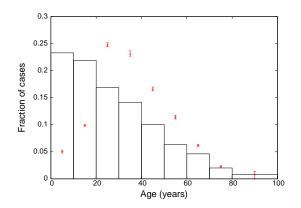




Dimensional analysis and classical mechanics



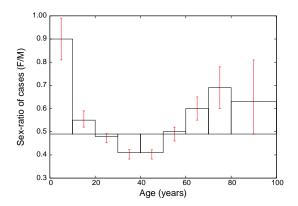
Age distribution of infections



Difference between population age structure and infections age structure: possibly due to different exposure rates, different infection rates, and different testing rates.

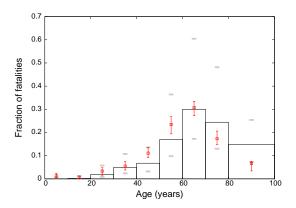
SG, MedRxiv-2020-153957

Sex and age distribution of infections



Women half as likely to be infected as men? Drop in infection rates of women between puberty and menopause! Not seen elsewhere in the world. Special Indian genetics, or differential access to health care? SG, MedRxiv-2020-153957

Fatalities due to COVID-19



Observed age distribution of deaths due to infections from COVID-19 agrees with the rest of the world. This predicts IFR of 0.4%, and implies unidentified deaths.

SG, MedRxiv-2020-153957



Death is one of the outcomes of COVID-19 in the short term. Complete recovery is another. There is a spectrum in between.

Death is one of the outcomes of COVID-19 in the short term. Complete recovery is another. There is a spectrum in between.

Damage to lungs (pulmonary effects) is known in severe cases, and similar damage has also been seen in mild cases. [Several publications]

Death is one of the outcomes of COVID-19 in the short term. Complete recovery is another. There is a spectrum in between.

Damage to lungs (pulmonary effects) is known in severe cases, and similar damage has also been seen in mild cases. [Several publications]

Damage to heart (cardiac effects) have been observed in severe and mild cases. [At least one moderately large study published]

Death is one of the outcomes of COVID-19 in the short term. Complete recovery is another. There is a spectrum in between.

Damage to lungs (pulmonary effects) is known in severe cases, and similar damage has also been seen in mild cases. [Several publications]

Damage to heart (cardiac effects) have been observed in severe and mild cases. [At least one moderately large study published]

Neurological effects (strokes, inability to concentrate, persistent extreme tiredness) after severe, mild, and asymptomatic cases reported. [Systematic studies?]

Death is one of the outcomes of COVID-19 in the short term. Complete recovery is another. There is a spectrum in between.

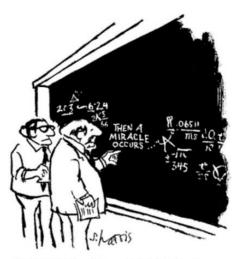
Damage to lungs (pulmonary effects) is known in severe cases, and similar damage has also been seen in mild cases. [Several publications]

Damage to heart (cardiac effects) have been observed in severe and mild cases. [At least one moderately large study published]

Neurological effects (strokes, inability to concentrate, persistent extreme tiredness) after severe, mild, and asymptomatic cases reported. [Systematic studies?]

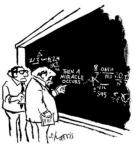
Damage to other organ systems reported.

A vaccine will solve everything



"I think you should be more explicit here in step two."

A vaccine will solve everything



"I think you should be more explicit here in step two."

India and the GAVI Foundation is betting heavily on the Oxford/AstraZeneca vaccine. The Serum Institute has stockpiled 10^6 doses, and can convert facilities to produce 60×10^6 doses a month, 50% for India. Funding from GAVI Foundation. Nature, news item 03 September, 2020

There are 10^4 hospitals in India. Assume 3×10^4 vaccination centers. With a monthly supply of 30×10^6 doses a month, average of 1000 doses per center.

There are 10^4 hospitals in India. Assume 3×10^4 vaccination centers. With a monthly supply of 30×10^6 doses a month, average of 1000 doses per center.

Plans to vaccinate susceptible persons first.

There are 10^4 hospitals in India. Assume 3×10^4 vaccination centers. With a monthly supply of 30×10^6 doses a month, average of 1000 doses per center.

Plans to vaccinate susceptible persons first.

Population of India is 1.3×10^9 people. Maybe 60% need to be vaccinated for herd immunity, *i.e.*, 10^9 people. The supply is 30×10^6 doses a month.

There are 10^4 hospitals in India. Assume 3×10^4 vaccination centers. With a monthly supply of 30×10^6 doses a month, average of 1000 doses per center.

Plans to vaccinate susceptible persons first.

Population of India is 1.3×10^9 people. Maybe 60% need to be vaccinated for herd immunity, *i.e.*, 10^9 people. The supply is 30×10^6 doses a month.

Immunization will require 34 months. Be prepared for a long wait.

Delivery of 405 million doses of measles-rubella vaccine took 3 years.

There are 10^4 hospitals in India. Assume 3×10^4 vaccination centers. With a monthly supply of 30×10^6 doses a month, average of 1000 doses per center.

Plans to vaccinate susceptible persons first.

Population of India is 1.3×10^9 people. Maybe 60% need to be vaccinated for herd immunity, *i.e.*, 10^9 people. The supply is 30×10^6 doses a month.

Immunization will require 34 months. Be prepared for a long wait.

Delivery of 405 million doses of measles-rubella vaccine took 3 years.

How effective is the vaccine? Is immunity life long?

Long term solutions

In the medium term, vaccine does not replace treatment. Burden of morbidity will last one generation. Understanding the biology of the disease and the virus will remain important.

Long term solutions

In the medium term, vaccine does not replace treatment. Burden of morbidity will last one generation. Understanding the biology of the disease and the virus will remain important.

Funding for science is important