IASS Webinar 3 September 2020

Statistical Design for Covid-19 Monitoring and Control

Stephen Haslett

Massey University, New Zealand
Systemetrics Research Ltd, New Zealand
Australian National University, Australia
University of Wollongong, Australia

s.j.haslett@massey.ac.nz

Abstract

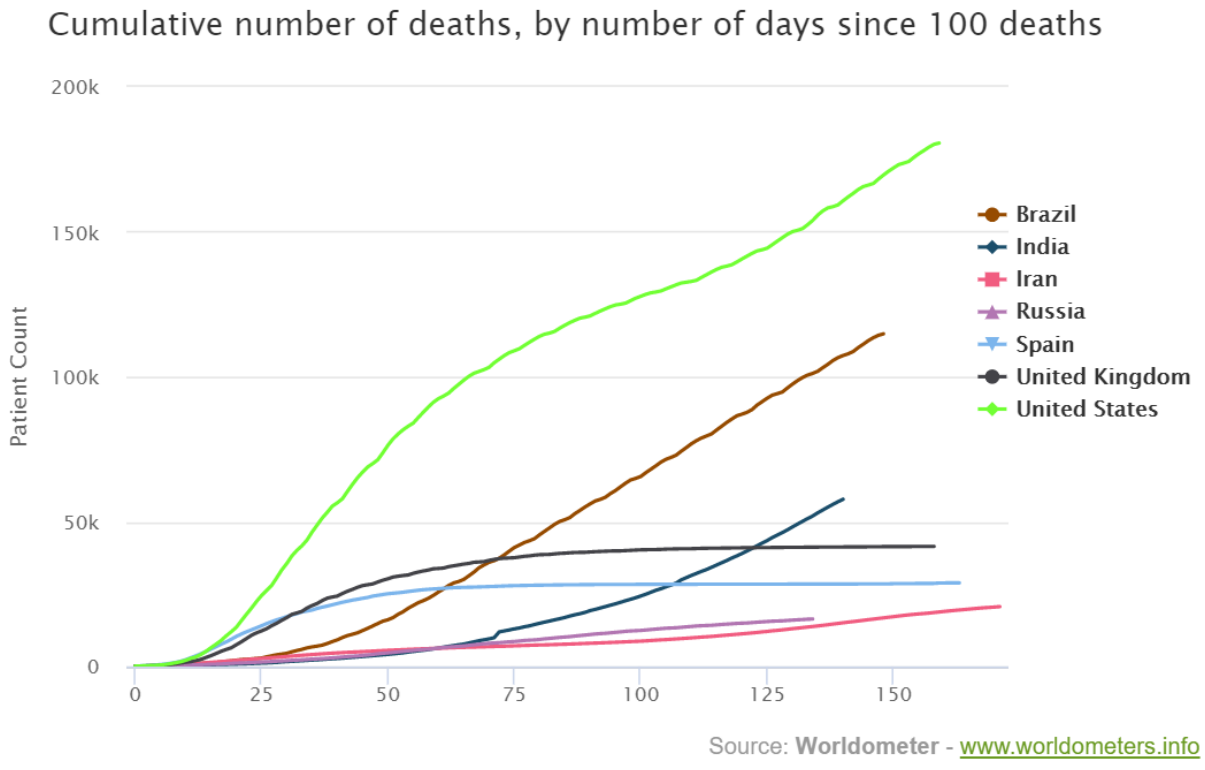
The Covid-19 pandemic has affected countries differently. In New Zealand all incoming international travellers are put in isolation for two weeks and any cases found are quarantined. There has been limited community transmission. Contact tracing of community cases remains exhaustive. There is a Statistical Advisory Group to the NZ Ministry of Health. The underlying strategy has been elimination rather than eradication, via a scale of alert levels that utilise lockdowns and bubbles.

Although the NZ situation is not always replicated elsewhere where Covid-19 prevalence is higher, there remain common underlying statistical themes and issues.

- The need for government Ministries and Departments of Health to prepare by commissioning design of prevalence surveys as soon as possible, even if implementation is delayed.
- Recognition of the potential to integrate sound sampling with contact tracing by using repeated adaptive cluster and network sample designs, which make it possible to track all contacts of known cases, have a chance of detecting unknown community cases, monitor special groups (such as those at the border) and get prevalence estimates with standard errors over time.
- The need to understand better the effect of the survey design on specificity of lab tests.
- To recognise how pooling laboratory tests using even simple experimental designs could improve test sensitivity.

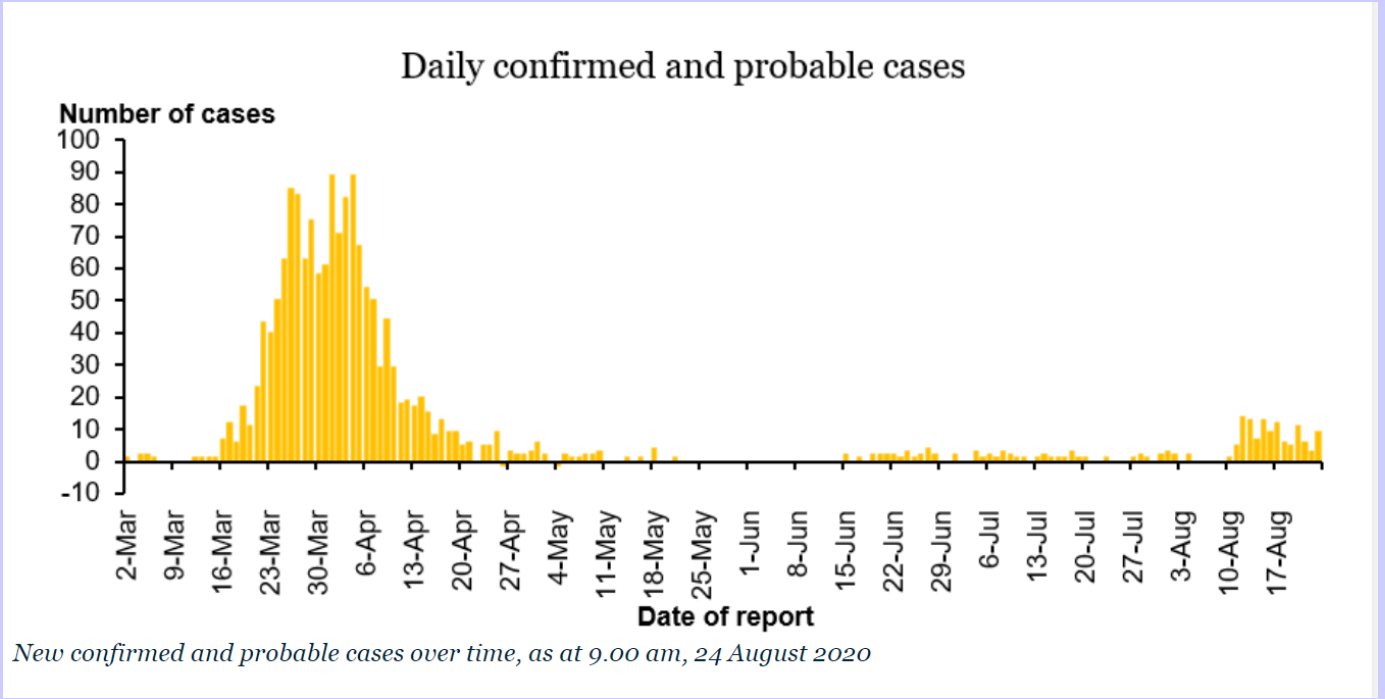
The Covid-19 pandemic has affected countries differently.

Cumulative number of deaths (by number of days since 100 deaths)



Source: https://www.worldometers.info/coronavirus/worldwide-graphs/ 24 August 2020
(See https://www.worldometers.info/coronavirus/?utm_campaign=homeAdUOA?Si for data)

In New Zealand, due to relatively late arrival of the virus and rapid government response, Covid-19 levels for the population of around 5,000,000 remain low.



New confirmed and probable cases over time, as at 9.00 am, 24 August 2020

https://www.health.govt.nz/our-work/diseases-and-conditions/covid-19-novel-coronavirus/covid-19-current-situation/covid-19-current-cases

Photo: Hilary Smith

In New Zealand

- All incoming international travellers are put in isolation for two weeks and any cases found are quarantined.
- There has been limited community transmission.
- Contact tracing of community cases remains exhaustive.
- The underlying strategy has been elimination rather than eradication, via a scale of alert levels that utilise lockdowns and “bubbles”.

There is a Statistical Advisory Group (SAG) advising the NZ Ministry of Health (MoH).

SAG has provided statistical advice to MoH on

- sensitivity and specificity of laboratory tests for Covid-19
- use of generic symptoms in diagnosis
- design and implementation of occasional simple, ad hoc sample surveys of at risk groups through the 20 District Heath Boards

NZ Statistical Advisory Group Membership

Dr Dean Anderson (Wildlife and Ecology Management, Landcare Research)
Prof Richard Arnold (Victoria University of Wellington, School of Mathematics and Statistics)
Prof Barry Borman (Massey University, Centre for Public Health Research)
Prof Nigel French (Massey University, School of Veterinary Science)
Alistair Gray (Statistics Research Associates Ltd)
Prof Steve Haslett (Emeritus Professor, Massey University, School of Fundamental Sciences and Centre for Public Health Research)
Prof Thomas Lumley (University of Auckland, Department of Statistics)
Dr Matt Parry (University of Otago, Department of Mathematics and Statistics)
A/Prof Patricia Priest (University of Otago, Department of Preventive & Social Medicine)
A/Prof Deborah Read (Massey University, Centre for Public Health Research)
Dr Lucy Telfar-Barnard (University of Otago, Department of Public Health)
A/Prof Robin Turner (University of Otago, Biostatistics Unit)

Although the NZ situation is not always replicated elsewhere where Covid-19 prevalence is higher, there remain common underlying statistical themes and issues.

The need for government Ministries and Departments of Health to prepare by commissioning design of prevalence surveys as soon as possible, even if implementation is delayed.

There have been surveys designed and implemented internationally e.g. UK https://www.ndm.ox.ac.uk/covid-19-infection-survey where the sample size intended was around 300,000 and the achieved sample size was around 100,000 using nasopharyngeal swabs rather than immunoassay.

However the UK survey is a standard stratified cluster design.

The need for government Ministries and Departments of Health to prepare by commissioning design of prevalence surveys as soon as possible, even if implementation is delayed *cont.*

In most other countries, no survey designs have been developed or implemented beyond occasional, short term, simple, ad hoc surveys usually of particular at-risk groups.

The lack of recognition in government that better, sound survey designs are needed, and that survey design cannot be instantaneous is widespread

In New Zealand the Ministry of Heath yesterday (2 September 2020) opened initial discussion on SAG's recommendation that survey design should begin as a matter of urgency.

A very recent general assessment of public health capacity in NZ can be found in

Crampton, P., Matheson, D. & Cotter, M. (2020) Assessing the Design and Capability of Our Public Health System in a Covid and Post-Covid New Zealand, *Policy Quarterly*, August 2020, 16, 3, 30-35.

Recognition of the potential to integrate sound sampling with contact tracing by using repeated adaptive cluster and network sample designs, which make it possible to:

- track all contacts of known cases, and integrating contact tracing into adaptive survey designs
- have a chance of detecting unknown community cases, by sampling at random points possibly with unequal selection probabilities based on risk or consequences
- monitor special groups (such as those at the border) by using a full coverage stratum
- get prevalence estimates with standard errors over time, since adaptive sampling while technically informative in using results from the survey to adapt selection probabilities, is nevertheless probability based and hence able to provide prevalence estimates and standard errors over time using repeated sample survey methods.

References:

Thompson, S.K (1992) *Sampling*, Wiley.

Thompson, S.K. & Seber, G.A.F. (1996) *Adaptive Sampling*, Wiley.

Loyal, J.D. & Chen, Y (2020) Statistical Network Analysis: A Review with Applications to the Coronavirus Disease 2019 Pandemic, *International Statistical Review*, 88, 2, 419–440 doi:10.1111/insr.12398

The need to understand better the effect of the survey design on specificity of laboratory tests.

		Has Covid-19	
		Yes	No
Test indicates Covid-19	Yes	a	c
	No	b	d

Sensitivity=a/(a+b) ***Specificity=d/(c +d)*** ***False positives=c*** ***False negatives=b***

Suppose to limit laboratory testing load using nasopharyngeal swabs or immunoassay tests, the test used is instead based a generic symptom such as a cough, headache, sore throat, fever, or on a set of generic symptoms. Then, if these symptoms are also associated with other diseases (for example influenza) when this other disease has higher prevalence (such as in winter) the number of false positives c increases relative to d and the generic screening test for Covid-19 becomes less specific. False positives increase pressure on medical services.

The nasopharyngeal swab test has sensitivity of around 75% and a specificity of over 99%.

In part this depends on the stage of infection, because in the earlier stages although the person is infectious the virus load in the nose may be insufficient for detection.

www.newshub.co.nz

This means sample design, choice of subpopulation, and selection of sample is important even at a point in time.

For example, if the subpopulation being tested is all symptomatic, or the generic symptom criteria for moving people to a nasopharyngeal test changes, then specificity of the lab test will depend on the particular subpopulation.

The effect of sample design on sensitivity (eg if all the nasopharyngeal tests are an early stage after onset) and specificity (if there is a pre-test based on generic symptoms) can be marked.

Understanding the survey design, whether it is implicit or explicit, is central, even critical.

To recognise how pooling laboratory tests using even simple experimental designs could improve test sensitivity and/or efficiency.

Pooling of tests to save on laboratory resources is not a new idea.

See for example:
https://www.cdc.gov/coronavirus/2019-nCoV/index.html,
https://www.fda.gov/medical-devices/coronavirus-covid-19-and-medical-devices/pooled-sample-testing-and-screening-testing-covid-19
and
Dorfman R. (1943) The detection of defective members of large populations. *Annals of Mathematical Statistics*. 14, 436–440.

However most designs for pooling are relatively simple and often individual specimens are in one pool only.

But these ideas can be extended.

	test 4	test 5	test 6	
test 1	N	**P**	N	**+**
test 2	N	N	N	**-**
test 3	N	N	N	**-**
	-	**+**	**-**	

Consider a three by three array, for nine people using six tests.

The tests allow the positive case P to be identified by the six results from the rows and columns.

There is a 33% reduction in the number of laboratory tests required.

There is also an increased probability of detection because each person is tested twice.

However this double testing does not necessarily translate to an increase in specificity for the individuals in the array because (with specificity of each test being less than one) it will for example only identify a row or a column in which there is an unidentifiable positive test (for an individual) if a single test is positive.

Larger arrays are possible, both in two and higher dimensions.

For example, for a 4*4 array, the number of tests is reduced from 16 to 8, ie a 50% reduction.

For a three dimensional 3*3*3 array, the number of tests required is now 9 tests for 27 people, a 66% reduction.

There is however a proviso related to identifiability and linked to prevalence, even if all the tests had sensitivity of one.

It can be proved that in a *p*-dimensional array for any arbitrary number of categories in each dimension, a person is identifiable as a positive case only if no more than one dimension of the array contains as its marginal counts two or more categories that are non-zero.

For example, a 2*2 array with four people and four tests (and hence no efficiency gain) the individuals are not identifiable if all four tests are positive, because to get this result either 2 or 3 or all four people may be positive.

Illustration of some non-identifiable patterns for 2*2 array with four positive tests as margins

	test 3	test 4	
test 1	N	P	+
test 2	P	N	+
	+	+	

	test 3	test 4	
test 1	P	P	+
test 2	P	N	+
	+	+	

	test 3	test 4	
test 1	P	P	+
test 2	P	P	+
	+	+	

For a 3*3 array, as an illustration of the method of proof, some sets of the six marginal test results do not identify an individual and some do, on the basis of whether or not they contain such a 2*2 subtable.

The proof can be extended to tables of any number of dimensions with any number of categories by considering all 2*2 subtables obtainable from interchange of categories within all pairs of dimensions.

This suggests that from a practical point of view not only are the logistics of how tests are set up in the laboratory important, but also that efficiency gain for this type of pooled laboratory testing depends on prevalence via the probability that more than one dimension of the table describing the testing regime contains more than a single non-zero count.

Alternatively, to increase sensitivity while at the same time improving identifiability and efficiency of laboratory tests, people's test material could be tested more than once within an array of more than 2 dimensions.

So that it has a name, let's call this:

Detection After Rotation Trace (DART) pooling.

An illustration is a 5*5*2 array (or two 5*5 layers) in which 25 people identified as A to Y are each tested four times.

The second layer displaces both each row and column by one more step with each increment, so that rows become (displaced) diagonals.

A	B	C	D	E	test 1
F	G	H	I	J	test 2
K	L	M	N	O	test 3
P	Q	R	S	T	test 4
U	V	W	X	Y	test 5
test 6	test 7	test 8	test 9	test 10	

A	H	O	Q	X	test 11
G	N	P	W	E	test 12
M	T	V	D	F	test 13
S	U	C	J	L	test 14
Y	B	I	K	R	test 15
test 16	test 17	test 18	test 19	test 20	

For this 5*5*2 array the theoretical efficiency gain is because only 4*5=20 tests are required for 25 people.

Each person is tested in the laboratory four times.

The corresponding sensitivity (assuming identifiability and that sensitivity of four tests for each individual are not correlated) is improved from *se* to $1-(1-se)^3$ (which for *se*=0.75 improves sensitivity to 0.98).

For larger arrays $k*k*2$, the efficiency gain is from the reduction of k^2 tests for k^2 individuals to $4k$ tests.

For a 10*10*2 array, this means 100 people could be tested four times each, using only 40 tests.

Identifiability is also improved because not all four tests in which an individual is involved need to be positive to identify the person; two are sufficient subject to sensitivity.

A result of diagonalisation is that nearest neighbours (which are all those in the same row or column in the first array) are now all in a different row and a different column in the second array. This removes much of the identifiability problem, subject to prevalence and sensitivity.

Arrays which are not square in any two dimensions, do not have one dimension with only two categories, or which have more than three dimensions are also possible. The *nxt* matrix of people by tests also has some interesting properties.

Nevertheless in pools, as the US Center for Disease Control notes, as for the original Dorfman design, dilution effects of pooling for positive individuals can require consideration.

https://www.cdc.gov/coronavirus/2019-nCoV/index.html

These design ideas need further development, and also need to consider sensitivity variation more carefully (for example by using a random effect for each individual for all of their tests).

But they do suggest that where laboratory tests are limited, the number of people who can be tested can be increased markedly and sensitivity of results improved without increasing the total number of laboratory tests required.

In pandemics, this would take pressure off testing regimes that can otherwise become inundated by testing of people with generic symptoms or the asymptomatic volunteers who have no symptoms at all.

It is the pandemic situation that makes pooling of tests important.

For other illnesses where there is not the same pressure on the maximum number of laboratory tests per day, testing individuals separately continues to have the advantage of operational simplicity.

https://twittercom/rob_thomas_nz/status/1296747602201870336?s=09