The probabilities of an outcome on intervention and control can be estimated by randomizing subjects to different testing strategies: A requirement when assessing the effectiveness of test, trace and isolation

Huw Llewelyn MD FRCP
Department of Mathematics
Aberystwyth University
Penglais
Aberystwyth
SY23 3BZ
Tel 01970622802
Fax: 01970622826

hul2@aber.ac.uk
Abstract

The efficacy of an intervention can be assessed by randomising patients to different diagnostic tests instead of directly to an intervention and control. This principle is applied by allocating an individual to intervention if the test result is ‘positive’ (or on one side of a threshold) but allocating that individual to a control if the result is ‘negative’ (or on the other side of the threshold). This can also be done with different dichotomising thresholds for one test. The frequencies of the outcome in those with each of the four resulting observations are then used to calculate the relative risk (RR) of the marginal probabilities by solving simultaneous equations. This assumes that the RR due to intervention compared to control is the same in both test groups created by randomisation. The calculations are illustrated by using data from a randomized controlled trial (RCT) that assessed the efficacy of an angiotensin receptor blocker (ARB) in lowering the risk of diabetic nephropathy in patients conditional on urinary albumin excretion rates (AERs). The calculations are also illustrated with simulated data for assessing the effectiveness of test, trace and isolation to reduce transmission of the SARS-Cov-2 virus by randomising to RT-PCR or LFD tests. This approach allows the probabilities of outcomes, their RRs and odds ratios (OR) conditional on the results of covariates (e.g. the RT-PCR test) to be determined. General conditions are specified for collapsibility and non-collapsibility regarding RR and OR with examples.

Keywords

Randomized intervention controlled trials, efficacy, effectiveness, albumin excretion rate, Covid-19, SARS-Cov-2 virus, RT-PCR test, LFD test, track and trace, self-isolation, viral spreader, sensitivity, specificity, predictiveness, relative risk, odds ratio, exchangeability, confounding, effect modification, collapsibility, non-collapsibility.
Randomizing to different diagnostic tests can predict treatment efficacy

1. Introduction

It is often not possible to randomize patients directly to intervention or control in clinical trials. This may happen when we wish to assess or to compare the performance of diagnostic tests for predicting response to a treatment or placebo when the latter’s efficacy has been established already in a previous randomised control trial (RCT). Such tests may have been invented by medical scientists, artificial intelligence researchers, mathematical modellers and medical statisticians. It is important to assess the performance of such diagnostic tests in order to avoid failing to give treatments to those who might benefit or to avoid giving treatments with possible adverse effects to those with little chance of benefit. Making this error has become known as ‘over-treatment’. ‘Over-diagnosis’ is another concern when a diagnostic label is attached to patients when there is little or no prospect of many patients benefiting from any of the treatments suggested by the label [1].

The variation in response to treatment in patients with different features is also known as the heterogeneity of treatment effect (HTE). This can be tackled by using regression based approaches to predictive heterogeneity of treatment effect analysis, including analyses based on risk modelling (such as stratifying trial populations by their risk of the primary outcome or their risk of serious treatment-related harms) and analysis based on effect modelling (which incorporates modifiers of relative effect) [2, 3]. However, the risk reduction due to a treatment for high blood pressure (BP), for example, will not reduce the overall risk added to by poor diabetic control as treatment for the BP will not also improve the diabetic control. The estimated risks arising from these models must therefore be regarded as test results in their own right and assessed in fresh studies to see how well they predict outcomes on individual treatments and controls. This gives rise to the same ethical issues as with single tests if efficacy has already been established in previous RCTs. In order to avoid the ethical issues of repeating RCTs, regression discontinuity design (RDD) might be used as an alternative [5, 6]. This is done by allocating patients to a treatment limb if the result of a test that predicts the outcome is on one side of a threshold and allocating them to a control limb if they are the other side of the threshold. A rough estimate of relative risk (RR) or odds ratio (OR) is obtained at the point of discontinuity by assuming that the RR or OR are similar or the same for a result just above or just below the threshold.

Pearl has pointed out the need for a logical framework for alternative approaches to RCTs of the kind described here based on concepts of causality, counterfactuals and collapsibility [7, 8]. Another approach to assessing how different findings predict outcomes with and without treatment might be to allocate subjects to two different diagnostic testing strategies. This approach is based on a traditional clinical view that a treatment will be more effective if given to patients based on the result of appropriate diagnostic information than if it given to those based on the result of inappropriate information. For example, if an inhaler is given to those with breathlessness and wheeze suggestive of asthma, then more will benefit than when the inhaler is given to those with breathlessness and audible crackles at the lung bases suggestive of left ventricular failure. If the inhaler is truly ineffective, no one will benefit from an inhaler whether they have wheeze (i.e. asthma) or crackles (i.e. heart failure). This principle suggests that the efficacy of a treatment could be assessed by randomizing patients to different diagnostic strategies instead of to a treatment and control, when the latter is difficult to justify ethically.
Randomizing to different diagnostic tests can predict treatment efficacy

2. Methods of modeling the link between diagnostic tests and treatment efficacy

The aim is to allow the outcome of a trial based on randomising to intervention or control to be predicted by randomising to different diagnostic testing strategies instead. The tests must have different predictive characteristics such as different sensitivities with respect to the outcome. The intervention is applied to a patient if the test result is on one side of a threshold or (when a test is positive) and to a control intervention if it is on the other side of the threshold (or if the same test is negative). This can be done for a pair of different tests or for one test with different thresholds of its numerical test results.

2.1 Rationale for methods

Consider that subjects are randomized to take part in two different randomised control trials Trial 1 and Trial 2 as shown in Figure 1. In Trial 1, the test T1 is performed on all subjects before they are randomised again a second time into those to be given a control or intervention. In those randomized to Trial 2, a test T2 is performed before randomisation again to control or intervention. The risk reduction due to the intervention in both trials is assumed to be the same and equal to \( x \) so that if the proportion with an adverse outcome on control in those who test T1 negative is \( a \), then the reduced risk with intervention is \( a \times x \). Similarly if the proportion with an adverse outcome on intervention in those testing T1 positive is \( b \), then the increased risk on control is \( b/x \). The same applies in Trial 2 when the outcomes are proportions \( c \), \( c \times x \), \( d \) and \( d/x \).

Figure 1: Diagram of randomisation to control or intervention in two trials after testing with test T1 in Trial 1 and testing with test T2 in Trial 2.
Randomizing to different diagnostic tests can predict treatment efficacy

We now perform a different study design as shown in Figure 2. We again randomize subjects to two groups, testing one group with test T1 and the other with test T2. However instead of randomizing again to control or intervention, we allocate subjects to a control if their test is negative and to intervention if the test is positive. In this design there are only 4 observed proportions, a, b, c and d as shown in Figure 2. However, these are the same proportions a, b, c and d shown in Figure 1. The relative risk of x is the same in Figures 1 and 2 also.

Figure 2: Diagram of randomisation to different tests and allocation to control if a test is negative or to intervention if the test is positive

Population to be studied

Randomisation

Test T1 based allocation

Test T2 based allocation

Allocation

Individuals who are T1 `\`

Proportion with outcome = a

Individuals who are T1 Pos => Intervention

Proportion with outcome = b

Individuals who are T2 Neg => Control

Proportion with outcome = c

Individuals who are T2 Pos => Intervention

Proportion with outcome = d

In Figure 2 the proportion a = the observed overall proportion with the adverse outcome and also having had a NEGATIVE result of test T1 and thus having been allocated to a CONTROL (see top line of Figure 2):

Again in Figure 2, x = is the relative risk so that a\*x = the calculated UNOBSERVED proportion having the adverse outcome and also having a NEGATIVE result of test T1 and thus having been allocated to the INTERVENTION (therefore calculated from knowing ‘a’ and x)

The proportion b = the observed proportion with the adverse outcome and also having had a POSITIVE result of test T1 and thus having been allocated to the INTERVENTION

The proportion b/x = the calculated UNOBSERVED proportion having the adverse outcome, also having a POSITIVE result of test T1 and having been allocated to the CONTROL (therefore calculated from knowing ‘b’ and x)

The proportion c = the observed proportion with the adverse outcome and also having had a NEGATIVE result of test T2 and thus having been allocated to a CONTROL

The proportion c\*x = the calculated UNOBSERVED proportion with the adverse outcome, also having a NEGATIVE result of test T2 and having been allocated to the INTERVENTION (therefore calculated from knowing c and x)

The proportion d = the observed proportion with the adverse outcome and also having had a POSITIVE result of test T2 and thus having been allocated to the INTERVENTION
Randomizing to different diagnostic tests can predict treatment efficacy

The proportion \(d/x\) = the calculated UNOBSERVED proportion having the adverse outcome, also having a POSITIVE result of test T2 and having been allocated to the CONTROL (therefore calculated from knowing \(d\) and \(x\))

Let \(a + a*x + b/x + b = y\), the probability of having the outcome when randomly allocated to Test 1

Let \(c + c*x + d/x + d = y\), the probability of having the outcome when randomly allocated to Test 2

As the overall prior probability ‘\(y\)’ of having the outcome is the same in the groups randomly allocated to test T1 and T2:

\[a + a*x + b/x + b = y = c + c*x + d/x + d\]  
Equation 1

Omitting \(y\) and rearranging Equation 1: \(a*x - c*x + b/x - d*x = c + d - a - b\)  
Equation 2

Rearranging Equation 2: \(x(a-c) - x(c + d - a - b) - (b-d) = 0\)  
Equation 3

Rearranging Equation 3: \((a-c)x^2 + (a-c)x + (b-d)x - (b-d) = 0\)  
Equation 4

Factorising Equation 4: \(((a-c)x + (b-d))(x-1) = 0\)  
Equation 5

From Equation 5 either: \((x+1) = 0\) and \(x = -(b-d)/(a-c)\)  
Equation 6

... or \(-(b-d)/(a-c)\) = 0 and \(x = 1\)  
Equation 7

Therefore \(x = -(b-d)/(a-c)\) = \((d-b)/(a-c)\) = the relative risk reduction  
Equation 8

For example, when \(a = 0.028\), \(b = 0.003\), \(c = 0.016\) and \(d = 0.006\), then

Relative risk is: \(x = (d-b)/(a-c) = (0.006-0.003)/(0.028-0.016) = 0.25\)  
Equation 9

The probability of the outcome conditional on T1 or T2 is: \(y = a + a*x + b/x + b = c + c*x + d/x + d\)  
\[= 0.028 + 0.007 + 0.012 + 0.003 = 0.016 + 0.004 + 0.024 + 0.06 = 0.05\]

3. Results based on real and simulated examples

3.1 Example based on real data

The following illustrative example is based on the result of a randomised controlled trial comparing the effect of placebo and irbesartan on the proportion of Type 2 diabetic patients who develop ‘Nephropathy’ in the form of severe proteinuria with an albumin excretion rate (AER) of over 200mcg/min within 2 years [9]. This AER range of >200mcg/min is regarded as one of the sufficient diagnostic criteria for the diagnosis of ‘Nephropathy’. This diagnosis suggests that the patient is in danger of suffering progressive renal impairment perhaps requiring renal dialysis and other support. The term ‘Nephropathy’ is also be used to indicate severe proteinuria within 2 years. The predicting test used was also the albumin excretion rate (AER) performed at the beginning of the trial. Note that randomisation was to 3 limbs. For the sake of simplicity the two intervention limbs are combined. The data in Table 1 show that the proportion developing nephropathy after 2 years on placebo was 30/196. However, the proportion developing nephropathy after 2 years on either dose of irbesartan was 29/379. This means that the relative risk reduction was \((29/379)/(30/196) = 0.499\).
Randomizing to different diagnostic tests can predict treatment efficacy

The pair of dichotomous test results T1 and T2 can be different tests such as a RT-PCR and Lateral Flow Device (LFD) or different dichotomising thresholds of a single numerical test such as an AER. This illustration will be based on thresholds of an AER of 40mcg/min and an AER of 80mcg/min. Thus a T1 positive was an AER >80mcg/min and T1 negative was an AER ≤ 80mcg/min. A T2 positive was an AER >40mcg/min and T2 negative was an AER ≤ 40mcg/min. If patients were randomised to T1 then the AER threshold would be 80mcg/min and if randomised to T2, the AER threshold would be 40mcg/min. Note that the results in shaded data in Table 1 would not have been seen by using this strategy.

Table 1 Proportion of patients developing nephropathy up to 24 months on different interventions after starting from different baseline urinary albumin excretion rates (AERs)

| Baseline AER     | Placebo | Irbesartan 150mg od | Irbesartan 300mg od |
|------------------|---------|----------------------|---------------------|
| 161 to 200 μg/minute | 2/7 = 28.57% | 4/13 = 30.77% | 1/2 = 50.00% |
| 121 to 160 μg/minute | 9/23 = 39.13% | 3/16 = 18.75% | 0/11 = 0.00%* |
| 81 to 120 μg/minute | 9/32 = 28.13% | 7/33 = 21.12% | 4/37 = 10.81% |
| 41 to 80 μg/minute | 9/57 = 15.79% | 5/66 = 7.58% | 4/74 = 5.41%† |
| 20 to 40 μg/minute | 1/77 = 1.30% | 0/59 = 0% | 1/68 = 1.47% |
| All: 20 to 200μg/minute | 30/196 = 15.30% | 19/187 = 10.16% | 10/192 = 5.21%# |

Relative risk for placebo and both doses of irbesartan = (29/379)/(30/196) = 0.499

The number of patients with an AER ≤40mcg/min allocated to placebo in Table 1 is 77. The number of patients with an AER>40mcg/min and allocated to treatment in Table 1 was 66 +74 + 33+ 37 + 16 + 11 + 13 + 2 = 252, which was 252/2 = 126 per limb. Therefore without having all the data in Table 1 available except for the un-shaded area, the estimated total number of patients in each limb is 77+126 = 203. This means that an estimated 203 patients were allocated to placebo and 406 were allocated to treatment with either dose of irbesartan. By performing the same exercise based on an AER threshold of 80mcg/min. the number of patients randomised to placebo <80mcg/min was 77 + 57 = 134. The number of patients allocated to treatment with an AER >80mcg/min was 33+ 37 + 16 + 11 + 13 + 2 = 112 or 112/2 = 61 patients per limb. The estimated total number of patients allocated to each limb based on a threshold of AER = 80mcg/min is therefore 134+ 61 = 195. The average of these two estimates is (195 + 203)/2 = 199 per limb. This means that the estimated number randomised to placebo was 199 and to treatment was 199 x 2 = 398.

3.2 Calculating estimates of the risk reduction and unobserved proportions

From Table 1, the estimated proportion of the outcome of nephropathy and having an AER ≤80mcg/min on placebo is 10/199 so that the estimated probability is 10/199 = 0.0503. This corresponds to probability ‘a’ in the above rationale. The estimated proportion of the outcome of nephropathy and having an AER >80mcg/min on treatment is 10/398 so that the estimated probability is 10/398 = 0.0477. This corresponds to probability ‘b’ in the above rationale. The
Randomizing to different diagnostic tests can predict treatment efficacy

estimated proportion of the outcome of nephropathy and having an AER ≤40mcg/min on placebo is 1/199 so that the estimated probability is 1/199 = 0.0050. This corresponds to probability ‘c’ in the above rationale. The estimated proportion of the outcome of nephropathy and having an AER >40mcg/min on treatment is 28/398 so that the estimated probability is 28/398 = 0.0704. This corresponds to probability ‘d’.

We are now in a position to calculate the estimated relative risk reduction. The probability a = 10/199 = 0.0503, b = 19/398 = 0.0477, c = 1/199 = 0.0050 and d = 28/398 = 0.0704. The calculated estimated relative risk is thus $x = (d-b)/(a-c) = (28/398-19/398)/(10/199-1/199) = (9/398)/(9/199) = 0.5$. This allows us to calculate the estimated unobserved proportions of nephropathy in those on treatment and control as shown in Table 2.

The proportion developing nephropathy on treatment and an AER≤80mcg/min is 10/199*0.5 = 10/398 = 0.0251. The calculated estimated proportion developing nephropathy on treatment and an AER>80mcg/min is 28/398*0.5 = 14/199 = 0.0704. The proportion developing nephropathy on control and an AER>80mcg/min is 29/398/0.5 = 29/199 = 0.1408.

The estimated observed and unobserved probabilities of nephropathy in those on treatment and control are shown in the upper row of Table 2. The estimated total proportion developing nephropathy on treatment in the top row is 10/199 + 19/199 = 29/199. The estimated total proportion developing nephropathy on treatment in the top row is also 10/398 + 19/398 = 29/398.

Table 2: Estimated observed and unobserved probabilities of nephropathy in those on treatment and control

| Threshold of AER | AER≤80mcg/min | AER>80mcg/min |
|-----------------|---------------|---------------|
| Control (a)     | a=10/199=0.0503 | a=19/398=0.0477 |
| Treatment (a*x) | (Calculated)   | (Calculated)   |
|                 | a*x=10/199*0.5=10/398=0.0251 | b/x=19/398/0.5=19/199=0.0954 |
| Treatment (b)   | (Calculated)   | (Calculated)   |
|                 | b=19/398=0.0477 |

| Threshold of AER | AER≤40mcg/min | AER>40mcg/min |
|-----------------|---------------|---------------|
| Control (c)     | c=1/199=0.0050 | c=29/398=0.0704 |
| Treatment (c*x) | (Calculated)   | (Calculated)   |
|                 | c*x=1/199*0.5=1/398=0.0025 | d=x=28/398/0.5=28/199=0.1408 |
| Treatment (d)   | (Calculated)   | (Calculated)   |
|                 | d=28/398=0.0704 |

Relative risk = x = (d-b)/(a-c) = (28/398-19/398)/(10/199-1/199) = (9/398)/(9/199) = 0.5

3.3 Some stochastic and other issues

The relative risk from Table 1 was (29/379)/(30/196) = 0.499. The calculations in Table 2 give an estimate of (9/398)/(9/199) = 0.5 happens to be identical to that using all the data in Table 1. This is clearly fortuitous in view of the small numerators of 9 in each case. The calculations summarised in Table 2 are estimating the result of an RCT with 196 subjects in the placebo limb and 379 subjects in
Randomizing to different diagnostic tests can predict treatment efficacy

the Irbesartan limb where the outcome was nephropathy AND a baseline AER between 40 and 80mcg/min. Table 1 shows that this result was 9/379 and 9/196 giving a relative risk of 0.5. When the observed proportions are 9/379 and 9/196 the P value for the difference is 0.002. However when the observed proportions are 29/379 and 30/196 the P value for the difference is 0.064. In order to achieve the same P value for the range 40 to 80mcg/min, about 3.6 times as many subjects would have to be recruited into the trial of the same proportions prevailed.

If there had been very large numbers of subjects, then the relative risks of nephropathy AND an AER in the other ranges would be expected to be the same. In the AER range 20 to 40mcg/min in Table 1 the relative risk point estimate was the same again at 1/379 and 1/196 = 0.5. However, for an AER between 80 and 200mcg/min the proportions are 19/379 and 20/196 giving a relative risk of 0.491. In order to conduct such a study subjects would have to be randomised into the 3 potential limbs of Placebo, Irbesartan 150mg or Irbesartan 300 mg but the substances would only be administered if the patient baseline AER were between 40 to 80mcg/min where it were considered that there was equipoise.

Subjects with baseline AER below 40mcg/min might be allocated to placebo and those with a baseline AER above 80mcg/min allocated to a treatment in order to construct curves that showed the probability of developing nephropathy on control and treatment for all baseline AERs from 20 to 200mcg/min. [4]. However although the strategy would be explained to subjects they would have to be ‘blinded’ to the result of their baseline AER and the subsequent nature of what was administered. In order to get sufficient statistical power and meaningful differences, the numbers randomised would have to be very large (about 4 times as many patients as in the IRMA2 study. This approach might be of value when monitoring the efficacy of treatments during day to day care and assessing newer diagnostic tests (e.g. the simpler albumin creatinine ratio as a possible replacement for the AER).

These point estimates from the overall proportions developing nephropathy from using all the data in Table 1 were 30/196 = 0.1530, and on treatment they were 29/379 = 0.0765. However from randomising to different diagnostic strategies the estimated overall proportion with nephropathy on control was 29/199 = 0.1457, and on treatment it was 29/398 = 0.0729. Clearly, precise results can be established only with a very large or infinite number of observations. However, the simplicity of randomising to different diagnostic tests instead of treatments means that it should be easier to recruit larger number of subjects that would reduce the width of the confidence intervals. The object of this paper is to demonstrate the principle of the approach. Placebo would be given to lower risk patients at lower risk of an adverse outcome and treatment given to those at higher risk. This might also be an advantage when it comes to assessing the effectiveness of a diagnostic and treatment strategy where randomisation of subjects to treatment or control would be problematic (e.g. during ‘test, trace and isolation’ (TT&I) for Covid-19).

4.1 Applications to TT&I for Covid-19 using simulated data from a suggested study design

Table 3 shows some simulated results from a suggested cluster design where people from different communities (e.g. schools) are randomised into 3 groups: (1) the RT-PCR group, (2) the LFD group with delay and (3) the LFD group with no delay. In Group 1, subjects testing positive for RT-PCR are asked to isolate 48 hours from when the test was performed (to ensure that all results were back) and those testing negative are asked not to isolate. In Group 2, those testing positive for a LFD test
Randomizing to different diagnostic tests can predict treatment efficacy

are asked to isolate 48 hours from when the test was performed (so that isolation was started after the same delay as for the RT-PCR group) but those with negative results are asked not to isolate. In Group 3, isolation is started immediately that LFD positive result becomes available (e.g. after 30 minutes). Both PRT-PCR and LFD tests are performed on all participants in the 3 groups as baseline. However, the decision in group 1 is based on the RT-PCR result and the decision in groups 2 and 3 is based solely on the LFD test result.

All participants in both groups testing positive and negative at day zero are asked to keep a record of contacts within two metres for more than 15 minutes for the next 10 days (perhaps with a smartphone app). After 10 days all the contacts of the 3 groups are tested with RT-PCR and LFD and those in the group who were tested negative originally but converted to be tested positive with either test at 10 days are designated ‘infected contacts’ and are ‘backward traced’ [10]. If they had been in contact within 2 metres for more than 15 minutes with a subject testing positive at the outset, the latter is designated a ‘positive spreader’ and the newly infected individuals termed ‘positive infected contacts’. If there are more ‘positive infected contacts’ (e.g. 75) linked to ‘positive spreaders’ (e.g.60) then some of the latter will have been ‘super-spreaders’ (e.g. up to (75-60)/75 = 0.2). The proportion of ‘positive spreaders’ infecting one or more would thus be 0.8, the number being 75*0.8 = 60.

The total number of ‘positive infected contacts’ (e.g. 75) is subtracted from the overall number of newly infected contacts at day 10 (e.g. 275) to give the total number of ‘negative infected contacts’ assumed to have been infected by those originally testing negative at day 0 (e.g. 275=75 = 200). The proportion of super-spreaders infecting these ‘negative infected contacts’ is assumed to be the same as for the ‘positive infected contacts’ (e.g. 0.2). The numbers of negative super-spreaders would therefore be estimated to be 200*0.2 = 40 and the number of ‘negative spreaders’ would be 200-40 = 160.

The reasons for estimating the number of viral spreaders is in order to provide meaningful estimate the sensitivity specificity, predictiveness etc of the RT-PCR and LFD tests. However, the efficacy of isolation in terms of relative risk and the effectiveness of isolation based on RT-PCR and LFD testing can be estimated from the numbers of infected contacts alone. The ratio of spreaders over infected contacts (e.g. 0.8) is the same for the positive and negative spreaders and infected contacts is assumed to be the same in all 3 groups and therefore has no bearing on the estimates of efficacy and effectiveness.

4.2 Simulated results from TT&I

The example ‘observed numbers’ per 100,000 used for the simulation of RT-PCR and LFD results are shown in Table 3. With these results of a = 160, b = 60, c = 280, d = 30, the estimated relative risk (RR) from Equation 9 is: (d-b)/(a-c) = (30-60)/(160-280)= 0.25. The ‘calculated’ numbers in Table 3 tell us that the overall proportion of Covid-19 spreaders without isolation is (160+240)/100,000 = 400/100000 = 0.004. The sensitivity of the RT-PCR test is 240/(240+160) = 240/400 = 0.6. As we would know the number of RT-PCR testing positive (e.g. 343 out of 100,000), the specificity can be calculated from the data in the P Map of Figure 3 where ‘A-y-P-x-B’ represents ‘given A, a proportion of X have B’ and ‘given B, a proportion of Y have A’. The presence of an arrow (e.g. ‘A-y-P-x->B’ indicates that A also has a causal effect on B as in a DAG diagram.
Randomizing to different diagnostic tests can predict treatment efficacy

Table 3: Estimated observed and unobserved numbers of Covid-19 in viral recipients in those isolated and not isolated

|                          | (60-30) / (280-160) | = 30/120             | = 0.25                     |
|--------------------------|---------------------|----------------------|---------------------------|
| OBSERVED number of       | CALCULATED number of | CALCULATED number of | OBSERVED number of        |
| spreaders per 100,000    | spreaders per 100,000| spreaders per 100,000| spreaders per 100,000     |
| in those RT-PCR test     | from RR=0.25 in those RT-PCR test | from RR=0.25 in those RT-PCR test | in those RT-PCR test |
| test negative and thus   | test negative &     | test positive &      | test positive and         |
| were actually allocated   | imagined allocated  | imagined allocated   | thus were actually         |
| to NO ISOLATION          | to ISOLATION        | to NO ISOLATION      | allocated to ISOLATION    |
|                          | a = 160             | b = 60               |                          |
| OBSERVED number of       | a*x = 160 x 0.25 = 40| b/x = 60 / 0.25 = 240|                          |
| spreaders per 100,000    |                     |                      |                          |
| in those LFD test        |                     |                      |                          |
| test negative and thus   |                     |                      |                          |
| were actually allocated   |                     |                      |                          |
| to NO ISOLATION          |                     |                      |                          |
|                          | c = 280             | d/x = 30 / 0.25 = 120|                          |
| OBSERVED number of       | c*x = 280 * 0.25 = 70|                      |                          |
| spreaders per 100,000    |                     |                      |                          |
| in those LFD test        |                     |                      |                          |
| test negative and thus   |                     |                      |                          |
| were actually allocated   |                     |                      |                          |
| to NO ISOLATION          |                     |                      |                          |
|                          |                      |                      |                          |

The overall proportion with no viral spread is: \((100000-400)/100000 = 99600/100000\). The proportion with no viral spread conditional on a negative PCR is \((99657-160)/99657 = 99497/99657\). The proportion overall with a negative PCR = \((100000-343)/100000 = 99657/100000\). From Bayes rule in Figure 1, the specificity is therefore \((99657/100000)*(99497/100000)/(99600/100000) = 99497/99600 = 0.998966\). The probability of viral transmission conditional on a positive RT-PCR without isolation is \(240/343 = 0.7\).

Figure 3: A P map of PCR positive / negative & viral spread /no spread with NO targeted isolation
Randomizing to different diagnostic tests can predict treatment efficacy

5.1 Discussion of initial simulation

This simulation shows that if no isolation were done then out of 100,000 subjects, 160+240 or 280+120 = 400 out of 100,000 would have resulted in transmission to at least one other individual. By isolating all those testing RT-PCR positive, 240-60 = 180 fewer or 400-220 = 310 out of 100,000 (instead of 400 out of 100,000) would have resulted in transmission to at least one other individual. However by applying TT&I using LFD, 120-30 = 90 fewer or 310-180 = 290 out of 100,000 would have resulted in transmission to at least one other individual. However, if in a third trial limb, when isolation occurred more rapidly as soon as the LFD result was known, only 10 would be found to have been spreaders (because the relative risk was 0.25 *5/15 = 0.0833). This would mean that 120-10 = 110 fewer spreaders would have occurred or 400-110 = 290 spreaders out of 100,000.

Figure 4: A P map of LFD positive / negative & viral spread / no spread with NO targeted isolation

![P-map diagram]

The sensitivity of the LFD test from Table 4 is 120/(280+120) = 0.3. As we would know the proportion of LFDs testing positive without isolation (e.g. 133 out of 100,000) and using the same reasoning with proportions as set out in the P Map in Figure 4, its specificity is 99587/99600 = 0.99987. The probability of Covid-19 transmission conditional on a positive LFD would be 120/133 = 0.9. As the relative risk is 0.25, the probability of viral spread conditional on a positive LFD WITH isolation is 0.9*0.25 = 0.225. The probability of ‘benefit’ conditional on a positive LFD with isolation is therefore 0.9-0.225 = 0.675. This means that the 0.675 probability of benefit conditional on a positive LFD is greater than the 0.525 probability of benefit conditional on a positive RT-PCR. However, fewer people would have a positive LFD (133/100,000) that would have a positive RT-PCR (343 out of 100,000). Therefore, the total number of people benefiting with a positive LFD (133 * 0.9 = 120 out of 100,000) is fewer than the total number benefitting with a positive PCR (343 * 0.7 = 240 out of 100,000).
Randomizing to different diagnostic tests can predict treatment efficacy

The superiority of the TT&I based on RT-PCR in this simulation is down to its greater assumed sensitivity of 0.6 compared to an assumed sensitivity of 0.3 of the LFD test. This is despite the probability of transmission conditional on a positive LFD (0.9) being higher than that for a RT-PCR (0.7). If a decision to isolate occurred only when both the LFD and RT-PCR tests were positive, then at best this combination would have a sensitivity of 0.3 so that the number of spreaders in those isolated would not change. However, if there was statistical independence between the likelihood of a positive RT-PCR and LFD results, the sensitivity of the combination would be 0.6*0.3 = 0.18. In this case the number of spreaders in those not isolated who were both LFD and RT-PCR positive would be lower at 18 so that with isolation of both LFT and PCR positive people, there would be 72-18 = 54 fewer spreaders. There would therefore be 400-54 = 346 spreaders instead of 400 out of 100,000. Thus isolating only those both LFD and RT-PCR positive would give the worst result. These results are summarised in Table 4.

The superiority of the TT&I based on RT-PCR in this simulation is therefore down to its greater assumed sensitivity of 0.6 compared to an assumed sensitivity of 0.3 of the LFD test. This is despite the probability of transmission conditional on a positive LFD (0.9) being higher than that for a RT-PCR (0.7). If a decision to isolate occurred only when both the LFD and RT-PCR tests were positive, then at best this combination would have a sensitivity of 0.3 so that the number of spreaders in those isolated would not change. However, if there was statistical independence between the likelihood of a positive RT-PCR and LFD results, the sensitivity of the combination would be 0.6*0.3 = 0.18. In this case the number of spreaders in those not isolated who were both LFD and RT-PCR positive would be lower at 18 so that with isolation of both LFT and PCR positive people, there would be 72-18 = 54 fewer spreaders. There would therefore be 400-54 = 346 spreaders instead of 400 out of 100,000. Thus isolating only those both LFD and RT-PCR positive would give the worst result. These results are summarised in Table 4.

Table 4: Effectiveness of different testing strategies for TT&I

| No TT&I | RT-PCR | LFD + delay | PCR & LFD + delay | LFD no delay |
|---------|--------|-------------|-----------------|-------------|
| 400     | 220    | 310         | 346             | 290         |
| No fewer| 180    | 90          | 54              | 110         |

5.2 A result if isolation was ineffective

If the following observations in Table 5 were made, this would indicate that isolation was ineffective with a relative risk of 1 but the performance of the PCR and LFT tests were the same as in Table 3. The same result could be obtained by performing the RT-PCR and LFD tests on the same patients, controversially (i.e. unethically) advising those testing both positive and negative for LFD and RT-PCR not to isolate at all and then observing the proportion of patients who went on to transmit to contacts of the positive and negative groups for both tests.

If the PCR and LFD tests were both useless because their sensitivities and false positive rates were the same and there was no risk reduction (i.e. the relative risk was 1), then all four observed outcomes and four calculated outcomes would be the same. If the following observations in Table 6 were made, this would indicate that both LFD and RT-PCR were highly predictive and that isolation highly effective so that there was a major impact on reducing transmission.
Table 5: Simulated data that suggest completely ineffective isolation

| OBSERVED number of spreaders per 100,000 in those RT-PCR test negative and thus were actually allocated to NO ISOLATION | CALCULATED number of spreaders per 100,000 from RR=1 in those RT-PCR test negative & imagined allocated to ISOLATION | CALCULATED number of spreaders per 100,000 from RR=1 in those RT-PCR test negative & imagined allocated to NO ISOLATION | OBSERVED number of spreaders per 100,000 in those RT-PCR test positive and thus were actually allocated to ISOLATION |
|---|---|---|---|
| 160 | $160 \times 1 = 160$ | $240/1 = 240$ | 240 |

| OBSERVED number of spreaders per 100,000 in those LFD test negative and thus were actually allocated to NO ISOLATION | CALCULATED number of spreaders per 100,000 from RR=0.1 in those LFD test negative & imagined allocated to ISOLATION | CALCULATED number of spreaders per 100,000 from RR=0.1 in those LFD test negative & imagined allocated to NO ISOLATION | OBSERVED number of spreaders per 100,000 in those LFD test positive and thus were actually allocated to ISOLATION |
|---|---|---|---|
| 280 | $280 \times 1 = 280$ | $120 / 1 = 120$ | 120 |

5.3 An example result if TT&I were highly effective

Table 6 tells us that the sensitivity of the RT-PCR test is $120/(120+80)=0.6$. As we know that the observed PCR positive tests was 343 out of 100,000, its specificity is $(50000*[(100000-300)/100000]-120+80)/(50000-120)=0.998597$.

Table 6: Simulated data that suggest highly effective TT&I

| OBSERVED number of spreaders per 100,000 in those RT-PCR test negative and thus were actually allocated to NO ISOLATION | CALCULATED number of spreaders per 100,000 from RR=0.1 in those RT-PCR test negative & imagined allocated to ISOLATION | CALCULATED number of spreaders per 100,000 from RR=0.1 in those RT-PCR test negative & imagined allocated to NO ISOLATION | OBSERVED number of spreaders per 100,000 in those RT-PCR test positive and thus were actually allocated to ISOLATION |
|---|---|---|---|
| 80 | $8 \times 0.1 = 8$ | $12 / 0.1 = 120$ | 12 |

| OBSERVED number of spreaders per 100,000 in those LFD test negative and thus were actually allocated to NO ISOLATION | CALCULATED number of spreaders per 100,000 from RR=0.1 in those LFD test negative & imagined allocated to ISOLATION | CALCULATED number of spreaders per 100,000 from RR=0.1 in those LFD test negative & imagined allocated to NO ISOLATION | OBSERVED number of spreaders per 100,000 in those LFD test positive and thus were actually allocated to ISOLATION |
|---|---|---|---|
| 40 | $4 \times 0.1 = 5$ | $16 / 0.1 = 160$ | 16 |

The sensitivity of the LFD test from Table 6 is $60/(160+40)=0.8$. As we know that the observed LFD positive tests was 133 out of 100,000, its specificity is $(50000*[(100000-323)/100000]-160+40)/(50000-160)=0.997562$.

Table 7 shows the result of using different LFD strategies when isolation is highly effective.

Table 7 the number of spreaders per 100,000 after different testing strategies for TT&I

| No TT&I | RT-PCR | LFD + delay | LFD no delay |
|---|---|---|---|
| 400 spreaders | 184 spreaders | 112 spreaders | 96 spreaders |
| No fewer | 216 fewer | 288 fewer | 304 fewer |
Randomizing to different diagnostic tests can predict treatment efficacy

By determining the numbers of spreaders carefully, it is possible to estimate the performance of T, T & I. In order to be solvable, the simultaneous equations must be mathematically independent. This depends on the tests used being different in terms of their mathematical characteristics such as sensitivity, specificity or predictiveness with respect to ‘viral spread’. It must be emphasised that the predictiveness (e.g. of 90 or 70%) of these tests applies to ‘spread’ and not to diagnosis. These tests are assumed by convention to be sufficient criteria for the diagnosis of Covid-19 and therefore have 100% predictiveness by circular argument. However, they are not definitive because although their positive tests are assumed to identify only those with Covid-19 (because they are assumed by circular reasoning to be 100% specific), they do not identify all those with Covid-19 (because they are not also assumed by the same circular reasoning to be 100% sensitive).

Instead of setting up simultaneous equations using a pair of different tests such as RT-PCR and LFD, it has already been shown using the AER that this can be done using a pair of different thresholds of a single test. The same principle can also be applied to the RT-PCR test by using two different Cycle thresholds (Ct) to report the result as positive or negative. For example, a positive RT-PCR T1 might be based on a Ct threshold above 25 cycles and a positive RT-PCR-T2 based on a Ct threshold above 35 cycles. The availability of these numerical results can also be used to estimate the probability of spread conditional on individual Ct threshold results by creating conditional probability curves based on ORs or RRs. However, this depends on the collapsibility of ORs or RRs regarding RT-PCR results with respect to the outcome of SARS-Cov-2 virus spread to others or the diagnosis of Covid-19 infection in an individual [Pearl et al].

6.1. The conditions for collapsibility

The set C is the control set (e.g. those not subjected to intervention such as isolation) and the set T is the set of those subjected to a treatment or other active intervention that differs from control. Ea is the initial finding before intervention or control that is common to set C and set T that establishes their exchangeability. Hc is an outcome on control and Ec is the finding after initiating control (that may be the same value as Ea). Ht is an outcome in set T after intervention and Et is the finding in set T after intervention (that may be different to Ea). E\not t is the complement of the finding Et.

6.2 Collapsibility of marginal relative risks

The RR for the marginal probabilities conditional on the sets C and T are collapsible if it is assumed that the effect of intervention compared to control is to reduce the marginal probabilities of the outcome by a ratio x such that

\[
p(H_t^{\not E_a | T}) / p(H_c^{\not E_a | C}) = x
\]

and that

\[
p(H_t^{E_a | T}) / p(H_c^{E_a | C}) = x
\]

It follows that

\[
\{p(H_t^{E_c | T}) + p(H_t^{\not E_c | T})\} / \{p(H_c^{E_a | C}) + p(H_c^{\not E_a | C})\} = x
\]

But as

\[
p(H_t | T) = p(H_t^{E_c | T}) + p(H_t^{\not E_c | T})
\]

and

\[
p(H_c | C) = p(H_c^{E_a | C}) + p(H_c^{\not E_a | C})
\]
Randomizing to different diagnostic tests can predict treatment efficacy

then
\[ p(H_T|T) / p(H_c|C) = x \]

and so that the RRs for these marginal probabilities are collapsible.

Figure 3 represents each of the exchangeable sets if no control or active intervention were applied. Thus \( p(H_c|C) \) is represented by the total marginal probability of ‘viral spread’ equal to 400/100000. \( p(H_c^AE|C) \) is represented by the marginal probability of ‘viral spread and a positive PCR’ equal to 240/100000. \( p(H_c^EA|C) \) is represented by the marginal probability of ‘viral spread and a negative PCR’ equal to 160/100000. Figure 3 also represents the result of implementing control by assuming that it does not change the status quo so that \( p(E_C) = p(E_A) \) and \( p(H_c^E|C) = p(H_c^E|A) \).

### 6.3 Collapsibility of conditional relative risks

Implementing an active intervention could have a number of effects as well as making \( p(H_T|T) \) different to \( p(H_c|C) \). An intervention may result in \( p(E_A) = p(E_A) \) or \( p(E_A) \neq p(E_A) \). Isolating everyone will reduce the probability of positive PCRs in potential contacts by contracting the virus from those already tested but isolation would not reduce the probability of positive PCRs already done in those already tested. The probability of a positive PCR in those isolated can be assumed to be the same as those on control (i.e. \( p(E_I) = p(E_A) = p(E_C) \). This situation results in the P map in Figure 5. When a PCR positive is represented by \( E_I = E_C \) then also \( p(E_I|T) = p(E_C|C) = 343/100000 \). Viral spread in Figure 5 is represented by \( H_T \) so that \( p(H_T|T) = 100/100000 \) and \( p(E_I|H_T) = 60/100 \).

Figure 5: P map of PCR positive / negative & viral spread /no spread after targeted isolation
Randomizing to different diagnostic tests can predict treatment efficacy

From Figure 5:
(1) \( p(H_T|T) = \frac{100}{100000} = 0.001 \)
(2) \( p(H_C|C) = \frac{400}{100000} = 0.004 \)

From Bayes rule:
(3) \( p(H_T|ET) = \frac{p(H_T|T) \times p(ET|H_T)}{p(ET)} \)
   i.e. \( p(H_T|ET) = \frac{100/100000 \times (60/100)}{343/100000} = \frac{60}{343} \approx 0.175 \)
(4) \( p(H_T|ÊT) = \frac{p(H_T|T) \times p(ÊT|H_T)}{p(ÊT)} \)
   i.e. \( p(H_T|ÊT) = \frac{100/100000 \times (40/100)}{99657/100000} = \frac{40}{99657} \approx 0.0004 \)
(5) \( p(H_C|EC) = \frac{p(H_C|C) \times p(EC|H_C)}{p(EC)} \)
   i.e. \( p(H_C|EC) = \frac{400/100000 \times (240/400)}{343/100000} = \frac{240}{343} \approx 0.7 \)
(6) \( p(H_C|ÊC) = \frac{p(H_C|C) \times p(ÊC|H_C)}{p(ÊC)} \)
   i.e. \( p(H_C|ÊC) = \frac{400/100000 \times (160/400)}{343/100000} = \frac{160}{99657} \approx 0.0016 \)

Therefore
(7) \( \frac{p(H_T|T)}{p(H_C|C)} = \frac{0.001}{0.005} = 0.25 \)
(8) \( \frac{p(H_T|ET)}{p(H_C|EC)} = \frac{0.175}{0.7} = 0.25 \)
(9) \( \frac{p(H_T|ÊT)}{p(H_C|ÊC)} = \frac{0.0004}{0.0016} = 0.25 \)

In general terms when \( x \) represents a RR then
(10) \( \frac{p(H_T|T)}{p(H_C|C)} = x \) and \( \frac{p(H_T|ET)}{p(H_C|EC)} = x \) and \( \frac{p(H_T|ÊT)}{p(H_C|ÊC)} = x \)
so that the RRs for these conditional RRs are collapsible (if and only if \( p(E_T) = p(E_C) \)).

6.4 Collapsibility of conditional odds ratios

In section 6.3 and Figure 5, \( p(E_T|T) = p(E_C|C) = \frac{343}{100000} \). However in Figure 6, \( p(E_T|T) \neq p(E_C|C) \).
Instead of being \( \frac{343}{100000} \), in Figure 6, \( p(E_T|T) = \frac{163.3}{100000} \). This probability was arrived at by first fixing the sensitivity and FPR in Figure 6 to make them identical to those in Figure 3, and then calculating the appropriate value of \( p(E_T|T) \). This is done because conditional predictive odds will be collapsible if and only if the likelihood ratios for the control and intervention sets are identical.

Figure 6: P map of PCR positive / negative & viral spread /no spread after viral eradication
Randomizing to different diagnostic tests can predict treatment efficacy

This situation might pertain if some ant-viral drug were used that instantly eradicated the virus in a proportion of those who are the source of the spread in the Set T so that the relative risk of spread for the marginal probabilities was also x (i.e. 0.25) as in Figure 5. However if the PCR test was repeated after giving the antiviral drug, then the proportion testing PCR positive reduced from 343/100000 to 163.3/100000. Note that there are many possible values for \( p(E_T|T) \) that are not constrained by the observed marginal probabilities with a RR of 0.25, 343/100000 and 163/100000 being merely 2 special cases of these many possibilities.

From Figure 6:
(11) \( p(H_T|T) = 100/100000 = 0.001 \)
(12) \( p(H_C|C) = 400/100000 = 0.004 \)

From Bayes rule:
(13) \( p(H_T|E_T) = p(H_T|T) \times p(E_T|H_T) / p(E_T) \)
   i.e. \( p(H_T|E_T) = (100/100000) \times (60/100) / (163.3/100000) = 60/163.3 = 0.367 \)
(14) \( p(H_T|Ê_T) = p(H_T|T) \times p(Ê_T|H_T) / p(Ê_T) \)
   i.e. \( p(H_T|Ê_T) = (100/100000) \times (40/100) / (99836.7/100000) = 40/99836.7 = 0.0004 \)
(15) \( p(H_C|E_C) = p(H_C|C) \times p(E_C|H_C) / p(E_C) \)
   i.e. \( p(H_C|E_C) = (400/100000) \times (240/400) / (343/100000) = 240/343 = 0.7 \)
(16) \( p(H_C|Ê_C) = p(H_C|C) \times p(Ê_C|H_C) / p(Ê_C) \)
   i.e. \( p(H_C|Ê_C) = (400/100000) \times (160/400) / (343/100000) = 160/99657 = 0.0016 \)

Therefore
(17) \( \text{odds}(H_T|T)/\text{odds}(H_C|C) = (0.001/(1-0.001))/(0.004/(1-0.004)) = 0.249 \)
(18) \( \text{odds}(H_T|E_T)/\text{odds}(H_C|E_C) = (0.367/(1-0.367))/(0.7/(1-0.7)) = 0.249 \)
(19) \( \text{odds}(H_T|Ê_T)/\text{odds}(H_C|Ê_C) = (0.0004/(1-0.0004))/(0.0016/(1-0.0016)) = 0.249 \)

In general when y is an OR then
(20) \( \text{odds}(H_T|T)/\text{odds}(H_C|C) = y; \text{odds}(H_T|E_T)/\text{odds}(H_C|E_C) = y; \text{odds}(H_T|Ê_T)/\text{odds}(H_C|Ê_C) = y \)
so that the ORs for these conditional ORs are collapsible.

This will be so if and only if the likelihood ratios
\( p(E_T|H_T)/p(E_T|Ê_T)/p(E_C|H_C)/p(E_C|Ê_C) = 1 \)
and the likelihood ratios
(21) \( p(Ê_T|H_T)/p(Ê_T|Ê_T)/p(Ê_C|H_C)/p(Ê_C|Ê_C) = 1 \)
Therefore:
(22) \( \text{odds}(H_T|E_T)/\text{odds}(H_C|E_C) = \text{odds}(H_T|T)/\text{odds}(H_C|C) \times p(E_T|H_T)/(p(E_T|Ê_T)/p(E_C|H_C)/p(E_C|Ê_C)) = \text{odds}(H_T|T)/\text{odds}(H_C|C) \times 1 = \text{odds}(H_T|T)/\text{odds}(H_C|C) = 0.249 \)
and
(23) \( \text{odds}(H_T|Ê_T)/\text{odds}(H_C|Ê_C) = \text{odds}(H_T|T)/\text{odds}(H_C|C) \times p(Ê_T|H_T)/(p(Ê_T|Ê_T)/p(Ê_C|H_C)/p(Ê_C|Ê_C)) = \text{odds}(H_T|T)/\text{odds}(H_C|C) \times 1 = \text{odds}(H_T|T)/\text{odds}(H_C|C) = 0.249 \)
and of course:
(24) \( \text{odds}(H_T|T)/\text{odds}(H_C|C) = 0.249 \)

Then because
(25) \( \text{odds}(H_T|E_T)/\text{odds}(H_C|E_C) = \text{odds}(H_T|Ê_T)/\text{odds}(H_C|Ê_C) = \text{odds}(H_T|T)/\text{odds}(H_C|C) \)
the odds are collapsible (NB if and only if of course the likelihood ratios for the control and intervention sets are identical).
Randomizing to different diagnostic tests can predict treatment efficacy

6.5 The theoretical nature of strict conditional collapsibility

The precise conditions of prior probabilities of findings being equivalent in the control and intervention set for conditional RRs and the likelihood ratios being equivalent for ORs to be collapsible can only be confirmed or refuted after the true probabilities are known after an infinite number of observations. If results based on some limited data satisfy these conditions, then this is probably fortuitous. They are therefore theoretical conditions for use in mathematical modelling. If the model’s output is calibrated against real limited data, then the calibrated model is clearly provisional to be updated with subsequent data. The most convenient model appears to be based on the odds ratio [4].

7. Conclusion

It is possible to estimate the overall relative risk of in the outcomes of a clinical trial by randomising subjects to two different testing strategies instead of randomizing them directly to an intervention or control. When the outcome on control (e.g. nephropathy as indicated by heavy proteinuria on placebo or a contact converting from RT-PCR or LFD negative to positive) is regarded as the outcome, it is possible to assess a test’s ability to predict this outcome. This would give the test’s positive predictiveness, sensitivity and specificity regarding the adverse outcome (not diagnosis). The assumption of equivalent likelihood distributions and constant odds ratios could be used to create model curves that after calibration against the latest data would display provisional probabilities of the outcome on intervention and control to be updated as new data become available.

Acknowledgments

I am grateful for the support of the past employees of Sanofi-Synthelabo and Bristol-Myers Squibb, and the investigators in numerous countries who participated in the IRMA2 trial for providing the data that helped me to develop these methods.

References

1. Moynihan R. Too much medicine? BMJ 2002;324:859
2. Kent DM, Steyerberg E, van Klaveren D. Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects. BMJ 2018 363 k4245
3. Kent DM, Paulus JK, van Klaveren D, D’Agostino R, Goodman S, Hayward R, et al. The Predictive Approaches to Treatment effect Heterogeneity (PATH) Statement. Ann Intern Med. 2020;172(1):35-45. DOI: 10.7326/M18-3667
4. Llewelyn H. The scope and conventions of evidence-based medicine need to be widened to deal with “too much medicine”. J Eval Clin Pract. https://doi.org/10.1111/jep.12981.
5. O’Keeffe AG, Geneletti S, Baio G. Regression discontinuity designs: an approach to the evaluation of treatment efficacy in primary care using observational data.BMJ 2014;349:g5293.
6. Nikki van Leeuwen Hester F. Lingsma, Simon P. Mooijaart, Daan Nieboer, Stella Trompet, Ewout W. Steyerberg, Regression discontinuity was a valid design for dichotomous outcomes in three randomized trials, JCE, 2018, 98, 70-79.
7. Pearl J, Mackenzie D. The Book of Why: The New Science of Cause and Effect. Penguin Books, 2018.
Randomizing to different diagnostic tests can predict treatment efficacy

8. Greenland S, Robins JM, Pearl J. Confounding and Collapsibility in Causal Inference. Statistical Science, 1999, Vol. 14, No. 1, 29–46.

9. Llewelyn H, Garcia-Puig, J. How different urinary albumin excretion rates can predict progression to nephropathy and the effect of treatment in hypertensive diabetics. JRAAS 2004; 5; 141-5.

10. Endo A, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Leclerc QJ et al. Implication of backward contact tracing in the presence of overdispersed transmission in COVID-19 outbreaks [version 3; peer review: 2 approved]. Wellcome Open Res 2021, 5:239 (https://doi.org/10.12688/wellcomeopenres.16344.3)