Hi so my talk today is on visual field, as you probably know by now, but specifically in the context of detecting glaucoma progression.

A series of visual fields can help clinicians decide if glaucoma has progressed over time. However, variability or the noise inherent in visual fields means it is not a good idea to detect progression by basic arithmetic operations alone.

As a result, different statistical analyses have been developed to help us detect and measure progression.

These analyses can either be trend-based such as linear regression, where we measure rates of change over time

or it also can be event-based, where we compare the most-recent measurements to baseline measurements to decide if progression has occurred.

Regardless of the different types of analyses, a common question we always want to ask is how good a particular analysis is in teasing out the signal from the noise. Suppose we have a patient with glaucoma and we measure their visual field once every year for four years.

We run the progression analysis and it says that the disease has progressed over the years.

What would then be of real interest to us is whether or not this is a false alarm. This is because we do not want to intensify the treatment and stress the patient out unnecessarily when there’s no progression in reality. In other words, we are really concerned with specificity here. In the past, researchers have successfully measured specificity by taking a series of visual field measurements in a short period of time during which no measurable deterioration can plausibly occur. However, what’s also important but has not really been explored is the following scenario:

suppose now that the analysis suggests no progression.

The question that interests us now is how confident are we that this is not a false negative result? This is perhaps all the more important because we do not want to be given a false sense of security and miss serious disease progression. To help us estimate false negative rates, why don’t we measure power? Power in this context is the likelihood of us detecting a given amount of progression.

Let me give an illustration in the next few slides to show why we want to bother with measuring power in individuals.

We have a patient here with very low variability as you can tell from this depicted visual field series when there’s no disease progression.

As a result, it would be relatively easy for a progression analysis to sound the alarm if the disease were to progress.

What happens if we, on the other hand, have another patient with a lot of noise in their visual fields. Well, it becomes much more difficult for the analysis to detect progression when it does occur. And it is these patients in a clinical setting that we really want to identify because there’s a higher likelihood of us missing progression if we over-rely on visual field results.

But here is the question. How do we measure power? Or put it in a different way, how do we measure the detectability of a given amount of visual field change?

Well, we can simply simulate progression thousands of times and compute how many times our progression analysis successfully detects the progression.

In fact, the idea of using simulation to measure power is nothing new. However, previous work is very limited in number and progression was expressed purely in terms of the overall change in visual field, which doesn’t really reflect real-world progression because glaucoma tends not to affect individual visual field locations to the same extent.

To test our idea, here’s what we did. We used a pre-existing dataset comprising 30 patients with stable glaucoma. Because each patient underwent 12 visual field measurements over a short period of time, no meaningful disease progression could have taken place, so this effectively means each visual field series represented the inherent measurement variability we would expect in an experimental setting.

Considering that our simulation was very computationally intensive, we took advantage of Azure cloud computing and ran our simulation 24/7 in the background on a virtual machine.

Let me talk about the design of our simulation.

First, the test intervals are adapted to 6 months, so each visual field series is assumed to span 5.5 years.

The visual field series is randomly reordered before

5 visual field locations are arbitrarily selected.

These locations are then injected with a specific rate of negative progression before

a progression analysis called Permutation of Pointwise Linear Regression or PoPLR is run and we get a p-value. If this p-value is smaller than 0.05, progression is said to have occurred, which in turn is recorded as a true positive result.

After this, the original, clean series is recalled and

randomly reordered again before the second cycle begins.

The cycle runs 100 times for each patient and for each rate of progression. Power is finally derived by computing the proportion of true positives.

Now I know that might sound really confusing but here is a very visual example. Each series has 12 visual fields coming from the same patient.

First, 5 locations are randomly selected.

Let’s use location 14 for the purpose of demonstration. The values underneath the visual field plots represent the sensitivity at this location.

Now we randomly reorder the series. Let’s call this the clean slate because we have not added any negative linear trend and the random reordering effectively removes any pre-existing systematic trend unknown to us.

We now inject a pre-determined rate of progression into this reordered series. The red values underneath each visual field plot represent the pointwise rate of progression, whereas the value above the blue arrow is the overall equivalent, roughly one tenth of the pointwise rate of progression.

We then run the progression analysis and we get our first p-value. If this p-value is smaller than 0.05, the analysis has correctly identified progression, so we regard this is a true positive result.

We recall our clean slate by randomly reordering the series again before injecting the same rate of progression. We run the analysis and get our second p-value.

This is repeated 100 times, so we get 100 p-values in total. The proportion of true positives is taken as the power

and the blue point on the graph here shows what the power is for this specific overall rate of progression.

We repeat the whole process all over again but for another rate of progression and we get another blue point.

The same process is repeated for all other rates of progression, and we finally get our first power curve for this patient.

To generate another power curve, everything is repeated with another randomly sampled set of locations.

15 different sets of locations are used in total so we have 15 power curves for each patient.

!!!!PAUSE!!!

Let’s look at the result from one patient. On the X-axis we have the rates of progression in terms of the overall value of visual field loss known as mean deviation. The Y-axis shows the power. Each faint gray line represents the power curve associated with a unique set of locations.

The average of these individual power curves is represented by this green line. We can see that the average power curve has a sigmoid shape, starting at around 5% power when the rate of progression is effectively zero, and this is consistent with the theoretical false positive rate of our progression analysis, given that the level of significance used by PoPLR is 5%. We can see that the power curve then ascends before converging and plateauing at 100%.

To achieve 80% power, or in other words, to be detectable 8 times out 10 by our analysis, this patient will have to progress by at least -0.18 dB/y. We can also observe a large spread between the individual curves. This suggests that a similar rate of change is more difficult to be detected at certain locations than other locations.

Let’s compare this to patient 25. What we can see here is that the average curve is now shifted to the left, which means progression is easier to be detected in this patient than patient 3. As a result, to achieve 80% power, patient 25 will only have to progress by -0.09 dB/y, which is half the rate of what is required in patient 3. But why do we get this discrepancy in power between patient 3 and 25?

Well, let’s have a look at the grayscale plots for patient 3 on the left and patient 25 on the right. It is clear that patient 3 has a much more damaged visual field than patient 25 and this results in higher variability in patient 3 as

you can see from the Mean Deviation time series in the top left corner. This increased amount of variability in patient 3 ultimately gives rise to lower power we see in this patient.

Here is a summary plot depicting the average power curves between 30 patients. What is interesting here is that all of the curves have a similar sigmoid shape, but just laterally shifted in relation to one another. This makes comparison between them relatively easy.

For example, let’s use 80% power as our cutoff again, we can see that the least powered and most powered curves differ by a factor of nearly 6. To achieve 80% power, the patient represented by the power curve furthest to the right will have to progress by at least -0.4 dB/y compared to a mere -0.07 dB/y for another patient represented by the curve furthest to the left. In conclusion, what we can potentially achieve with this simulation is to look at these power curves from different patients and tell how likely a specific rate of change can be detected in which patient. This could help us assess the quality of our patients as a visual field test taker and help us decide, for example, how much weight we really want to give to visual field measurements in patients not likely to be flagged by visual field tests alone.