The Delphi Project

Using Design of Experiment methodology to make better vaccines faster

The Delphi Project
Better Vaccines, faster

Aim:

- Evaluate the use of "Design of Experiment" Methodology (DoE) for optimized vaccine clinical development

Potential:

- More efficient Phase 1 /Phase 2 trials leading to
- Shorter time to Phase 3
- Optimal vaccine composition going into Phase 3
  - Better immunogenicity
  - Lower reactogenicity
  - Better manufacturability
  - Reduced COGs
  - Faster scale up
- Lower risk of failure

How:

- Computer simulations of vaccine trials using DoE approaches compared to conventional design and based on existing vaccine trial design

This project was undertaken by Allan Saul while a GSK employee

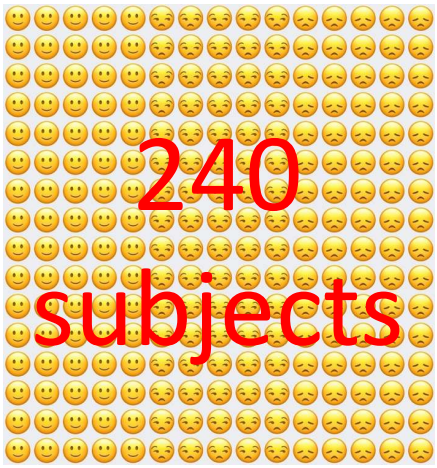
Summary

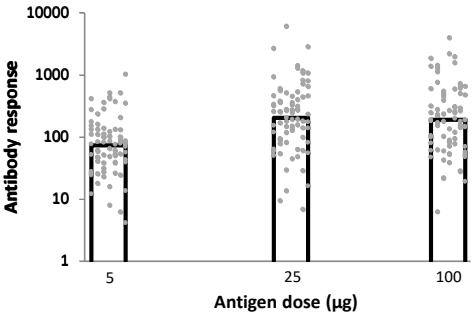
- Brief intro to Design of Experiment (DoE)
- The concept that “More is not always better”
- The concept of a “response surface”
- How to stretch or shrink data to give a optimum fitted response surface
- Best possible fit for antibody response for a example of antigen+ adjuvant vaccine (no person to person variation)
- “Real fits” how many people?
- What else we can do with a response surface?
- Next steps

On the next set of slides, I would like to introduce some of the concepts of DoE and how these differ from the normal discrete way we conventionally design vaccine trials. I will do this with a simple case – a dose escalation study with 3 different doses of a single antigen with a single vaccination with the aim to find the best dose. Later I will extend this to a more complex situation with two components – an antigen and an adjuvant were the dose of both are varied.

Conventional study

Dose response with a single antigen



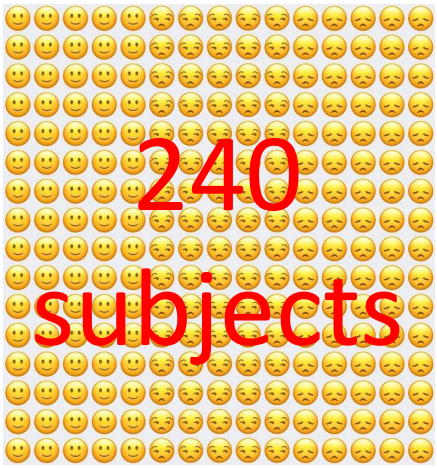


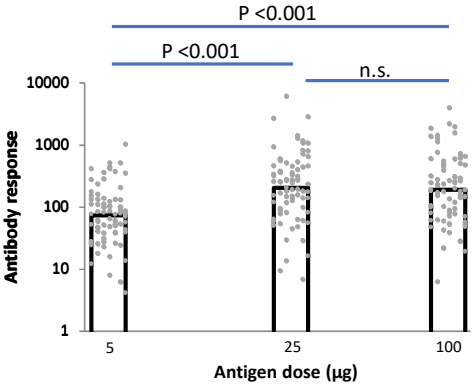
Based on *S. sonnei* GMMA study Launay et al EBioMedicine. 2017 Aug;22:164-172

In a typical dose escalation study, we may have three vaccine groups: low, medium and high doses. In this simulation I have 80 subjects per group to give a total of 240 subjects. This example is based on a small exploratory dose escalation study of the *S. sonnei* GMMA vaccine 1790 and uses the GMC antibody seen in the groups and the observed standard deviation of the log transformed antibody within each group. This simulation differs from the original to use a lot more subjects to enable the study to be powered to see a difference between the different vaccine doses. This slide shows the simulated antibody responses at the low, medium and high dose group and the geometric mean concentration (GMC) of antibody in the simulated groups.

Conventional study

Dose response with a single antigen





Based on *S. sonnei* GMMA study Launay et al EBioMedicine. 2017 Aug;22:164-172

After "vaccination", the simulated antibody responses were tabulated and an ANOVA performed to see if there is a difference between the GMC across the trial, and if so, then pairwise comparison was done to see which groups differ. In this simulation there was no significant difference between the medium and high dose groups but a significant difference between the low and medium, and the low and high doses. This shows the last simulation done in a series of simulations. Some gave a significant difference between the medium and high doses, in others like this example, the difference was not statistically significant. The important outcome is whether or not the groups differ. This approach does NOT provide direct information about the antigen dose that gives the best response, only which group was "best".

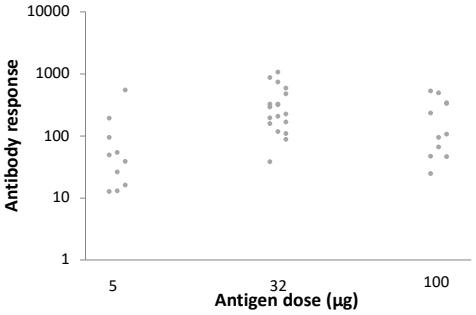
To recap – the trial design is for discrete groups and the outcome is if the groups differ

Design of Experiment study

Dose response with a single antigen

40

subjects

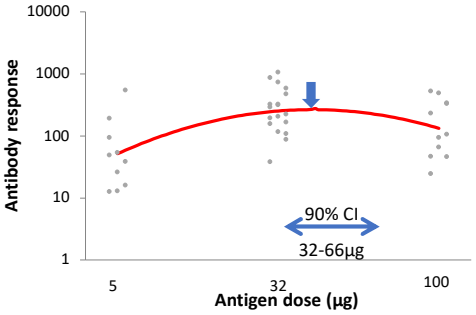


Based on *S. sonnei* GMMA study Launay et al EBioMedicine. 2017 Aug;22:164-172

On this slide I am showing a simulation of the same study based on "design of experiment" principles.
First you will note that there are lot fewer subjects but also there different numbers in the groups – twice as many in medium dose group compared to the high dose group for a total of 40 subjects. The slide shows the simulated antibody responses, but note – no GMC was calculated.

Design of Experiment study
Dose response with a single antigen

40 subjects



Based on *S. sonnei* GMMA study Launay et al EBioMedicine. 2017 Aug;22:164-172

After "vaccination", the simulated antibody levels are tabulated and an ANOVA done - In this case NOT to compare differences between groups, but to test if there is an overall fit to an equation that describes the dose-response. In this case the model chosen is a second order polynomial or a Quadratic equation or you may know it as a parabolic equation. This the simplest fit that allows a curvature. With 40 subjects there is a highly significant fit to a quadratic equation. Since there is only one variable in the trial (the dose of antigen) we have a "response curve" – the red line. Later in this talk, you will see cases where there are two variables –antigen and adjuvant and we will look at the response surface. (more complex cases are quite likely – several antigens and an adjuvant. We can't easily visualize this but we will still talk about a "response surface")

A major outcome from this study is the dose of GMMA that gives the maximum response. Theoretically, based on the data from the original trial used for the simulation, it is 44 µg and in this particular simulation, the predicted maximum antibody response occurs at 43 µg (at the arrow). When the same simulation was repeated many times, 90% of the predicted maximum responses lie between 32 and 66 µg. Unlike the "conventional" approach with 240 subjects, it is clear that the 100 µg dose is sub-optimal.

There is one other subtle difference between the two approaches. The DoE experiment has used a different "medium" dose 32 vs 25 µg. This is probably too small a difference

to make much overall impact- but it highlights a critical difference – (and in fact why this approach is called “Design of Experiment”) - the trial is design to maximize the ability to adequately fit the response surface not to try to use complex statistics to extract the information out of an arbitrary trial design.

To recap – in the DoE approach, the trial design is optimized to fit a continuous response to the dose escalation that enables the theoretical best response to be calculated even if this does not coincide with one of the groups.

All the way through this talk you will also hear the word “adequate”. The aim is NOT to perfectly fit a theoretical curve over the whole possible response surface but to get an adequate fit in the area of particular interest using the simplest possible equation for a response surface: in this case the area around the peak antibody response.

I will not detail the statistics underlying the DoE approach, but to summarize, DoE had its origin in the USA DoD in WWII with a pressing need to rapidly optimize weapon development. After WWII it was taken up by major chemical manufacturers (e.g. ICI and Du Pont) with substantial progress on designing the statistical and analytical basis for optimizing complex processes. Later, further advances took place in Japan where it was a key element underlying the huge expansion of quality manufacturing in that country. It is now widely used in many areas of engineering, science and even finance for process optimization. It forms a key component in “Quality by Design” and the 6 Sigma approach and is a key part of obtaining regulatory approval for manufacturing. It is widely used in vaccine industry for pre-clinical studies and for optimizing analytics and production.

Response surface analysis has been used by GSK to find optimum dose of tocopherol in seasonal flu vaccine and this is a nice example of how multiple different responses can be combined into “desirability score”* but as far as I know the full DoE approach has not been for any vaccine trial leading to registration.

Why? I don’t know, but the aim of the Delphi project is to see if there is a role in vaccine optimization in Phase I and Phase II clinical studies.

*See Rümke et al 2013, BMC Infectious Diseases 13:348 ; Dewé et al 2016 Journal of Biopharmaceutical Statistics, 26:2, 352-364,

More is not nece13/03/2020ssarily better

Even before we consider

- Safety
- Cost
- Production capacity

More does not necessarily give a better immune response

And with DoE approaches considerable dose optimization can take place with small groups in a Phase I or Phase II

But first, here are a number of examples where more is not better.

As a word of orientation, in almost all of the following slides, I am showing the log of the antibody responses not the linear response. You are probably used to this – if you quote geometric mean concentrations – you are making an implicit assumption that the antibody responses are log – normally distributed and the log antibody responses are important for the surface response fitting so I will be consistent

More Antigen is not necessarily better

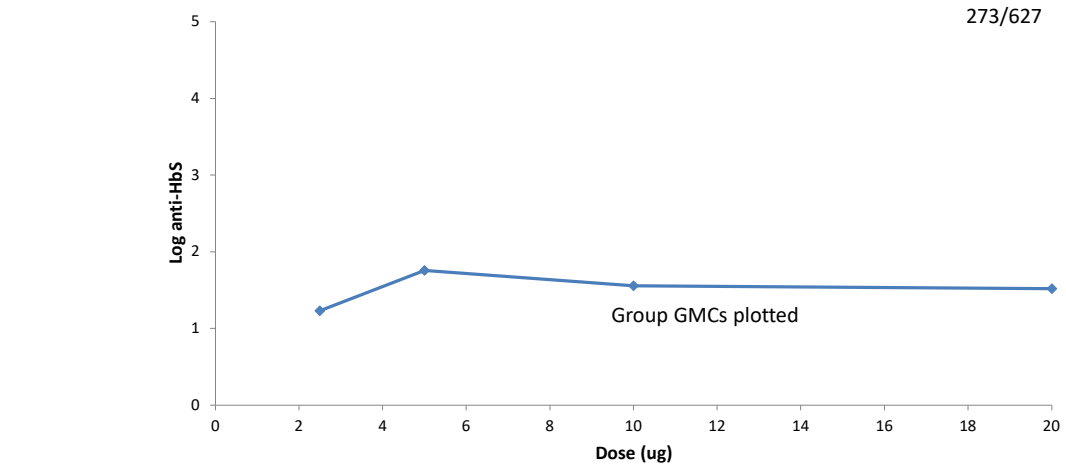
And with DoE approaches considerable dose optimization can take place with small groups in a Phase I or Phase II

But first, here are a number of examples where more is not better.

As a word of orientation, in almost all of the following slides, I am showing the log of the antibody responses not the linear response. You are probably used to this – if you quote geometric mean concentrations – you are making an implicit assumption that the antibody reponses are log – normally distributed and the log antibody responses are important for the surface response fitting so I will be consistent

Engerix

Anti-HBs after 2 vaccinations

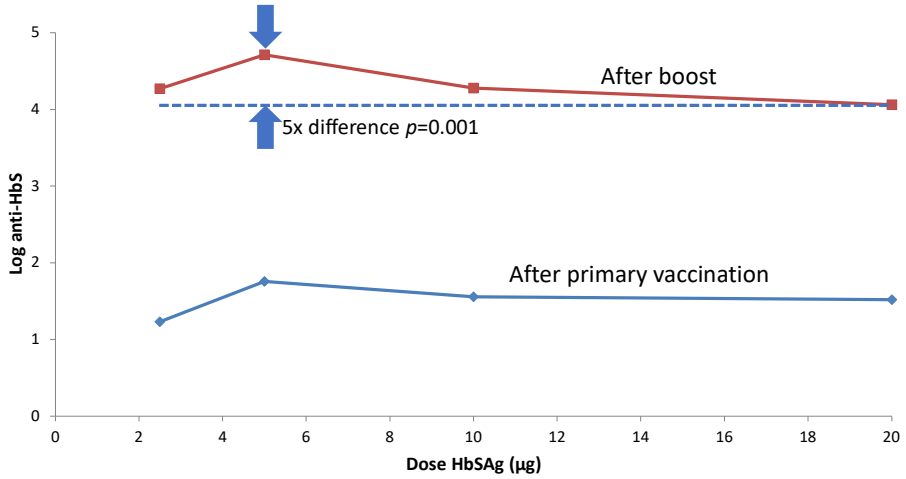


Weidermann et al 1987 Vaccine 5:179-183

Here is an example from the Smith Kline development of Engerix. This shows the antibody response from HBsAg after the second dose. You can see that the best antibody response was at 5 µg not the 20 µg that was used in the vaccine.

More antigen is often not better

Engerix



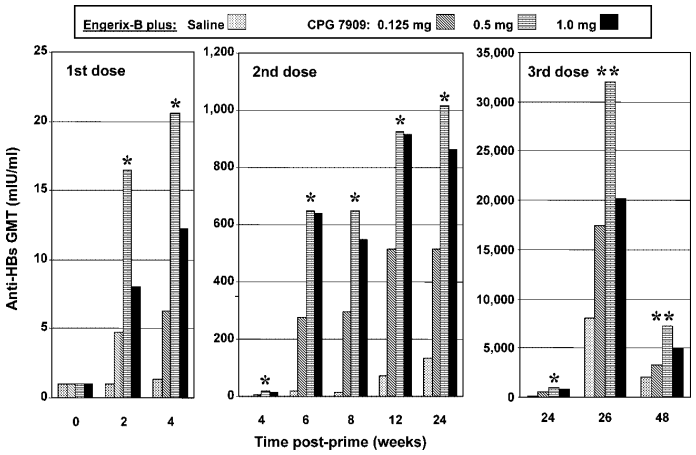
Weidermann et al 1987 Vaccine 5:179-183

This difference is more striking following a boost. In this case everybody was boosted with 20 µg but you can see that the people primed with 5 µg gave a highly significantly greater response than the people primed with 20 µg.

More Aduvant is not necessarily better

Adjuvant dose increases
HbSAg (Engerix) plus CpG7909

56

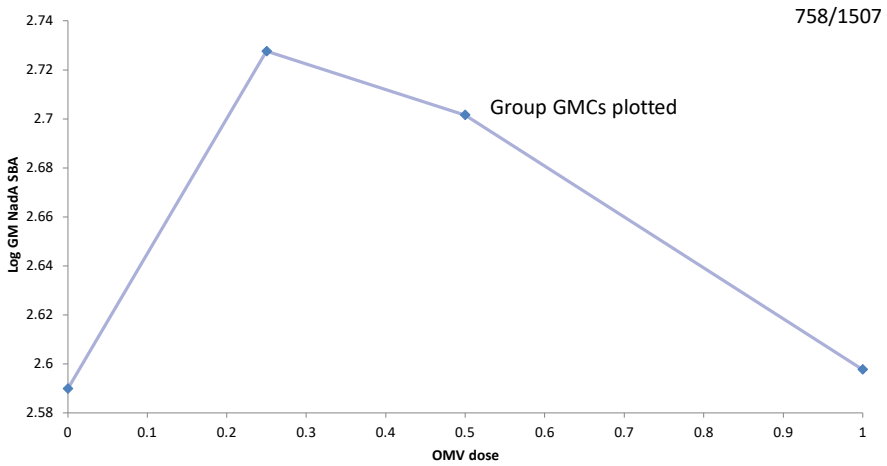


Cooper et al (2004) Journal of Clinical Immunology, 24:693-701

Here is Engerix mixed with an adjuvant – CpG7909 from Coley. In this example the Engerix dose was kept constant and the dose of adjuvant varied. The maximum response did not occur at the maximum dose of adjuvant. Note that the scale changes considerably in each panel.

Nad A response in infants as the OMV dose varies

NadA response in Bexsero adjuvanted by OMV



Esposito et al 2014 Human Vaccines & Immunotherapeutics 10:7 2005-2014

These are data from a Bexsero study in Infants where the dose of OMV was varied. This was a large study there were 758 infants contributing to this graph out of a total of 1507 overall. Here I am only showing the GMC for each group

This nicely illustrates the "Adjuvant" properties of OMV. With the addition of OMV the antibody response to NadA increases. The response to the other recombinant proteins (fHbp – by SBA and NHBA – by ELISA) was very similar. However as the dose was increased to the dose in Bexsero, the response dropped until it was the same as the antibody response in the absence of OMV. So the "best formulation" depends on what is the critical component – the response to PorA increased within increasing OMV, but the point here is that for the NadA, fHbp or NHBA in infants more OMV was not better

More vaccinations are not necessarily better

More vaccinations are not necessarily better

Pneumococcal vaccine

- 2 + 1 schedule (2, 4, 12 months) is at least as good as 3 + 1 (2,3,4,12 months)

- 1 + 1 schedule (3, 12 months) with PCV13
  - GMC µg/mL for serotypes 1, 4, 14, 19F significantly higher in 1+1
  - GMC µg/mL for serotypes 3, 5, 7F, 9V and 19F not significantly different
  - GMC µg/mL for 6A, 6B, 18C and 23F significantly lower in 1+1
  - No serogroup had significant difference in opsinophagocytic activity

Goldblatt et al 2006 Pediatr Infect Dis J, 25:312-319; Goldblatt et al 2018 Lancet Infect Dis 18: 171–79

DoE concepts: single component

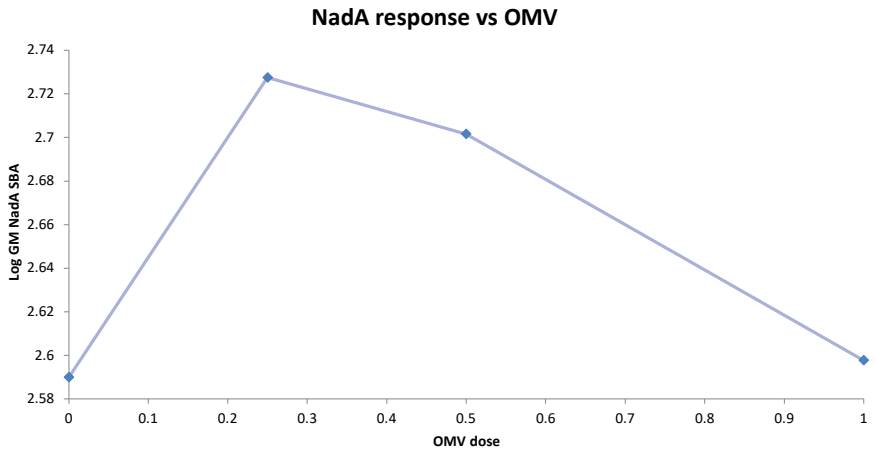
Shrinking and stretching the X axis to get a optimize the fit to the Response surface

Next I would like to introduce the concept that we can arrange the experiment to simplify obtaining an adequate fit to a response surface using one of the examples I have just showed.

Nad A response in infants as the OMV dose varies

NadA response in Bexsero adjuvanted by OMV

Linear dose scale

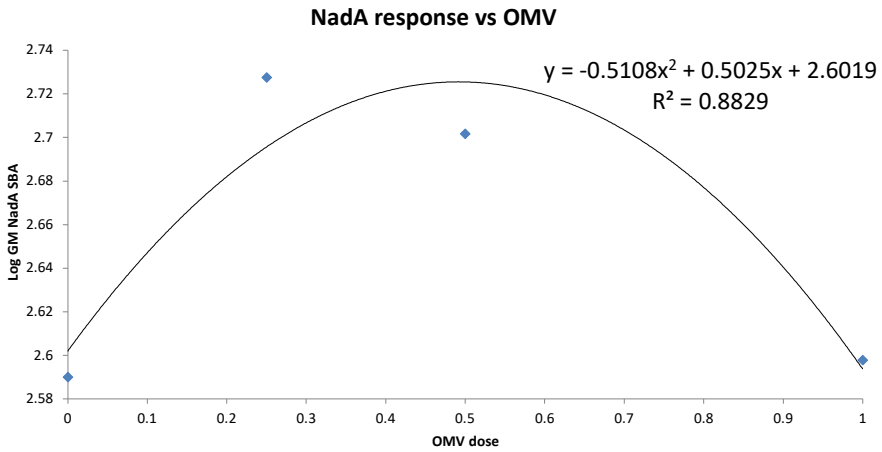


Here are the data on the impact of OMV with the log NadA response plotted as a linear function of the OMV dose

Nad A response in infants as the OMV dose varies

NadA response in Bexsero adjuvanted by OMV

Linear dose scale

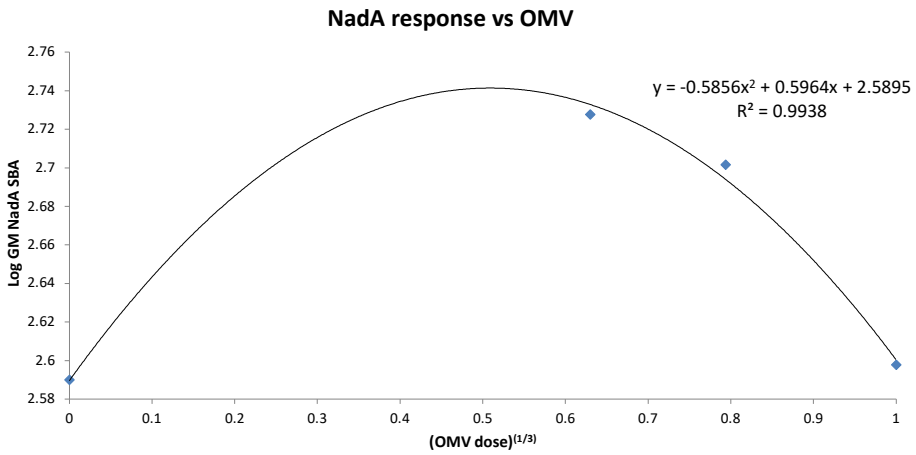


We would like to fit a quadratic equation to these data and you can see there is a reasonably good fit. R^2 of 0.88 means that most of the residual variance has been accounted for by the quadtratic equation, but you are probably not that impressed! There is a lot of residual scatter and the observed best response does not occur at the maximum of the fitted curve.

Nad A response in infants as the OMV dose varies

NadA response in Bexsero adjuvanted by OMV

Cube root dose scale

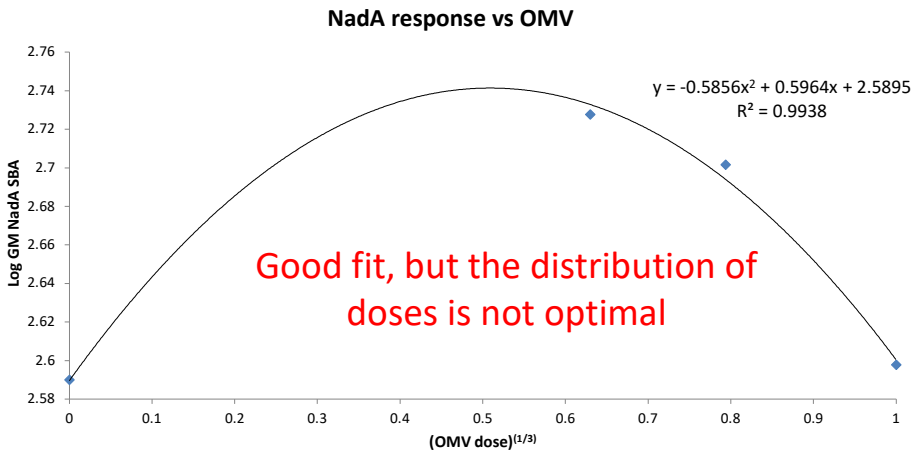


Here is a much better fit- This has been obtained by transforming the dose data to plot log antibody response as a function of the cube root of the dose, rather than the dose itself. This is a commonly used transformation in a DoE approach. However you can see that although the fit is very good, there is one drawback – the data points are not spread evenly and this impacts on the robustness of the fit.

Nad A response in infants as the OMV dose varies

NadA response in Bexsero adjuvanted by OMV

Cube root dose scale



Not shown here, a cube root transformation has worked with many other vaccine dose response curves and seems to be a general transformation that gives an adequate fit for a quadratic equation. This type of transformation is not uncommon, even for non DoE applications. Many dose response curves are plotted as a function of the log of the dose and that is a commonly used transformation in DoE experiments. In this particular case, as for other antibody dose responses examined that had many more dose levels (not shown), empirically the log transformation did not results in as good a fit.

Conclusions on stretching and shrinking

- For the vertical scale for antibody responses I want to only use log transformed data
  - Large data bases (not shown here) show that log of the antibody responses nearly always has a normal distribution but the standard deviation of the log responses vary (about 0.5 to 1)
  - The log transform lets us use parametric statistics for fitting the response surface and this is useful.
- For the horizontal dimensions I will use the cube root of the doses
  - This is an obvious transformation to try and is commonly used in DoE
  - It is not obvious that this would work so successfully. This is an important result!
  - Other transformations (e.g. Square root) have been tested on trial results but are not as useful.
  - Cube root avoids the problems of zeros – log of zero is negative infinity which makes the fit impossible!)

DoE modelling with vaccines with one antigen and one adjuvant

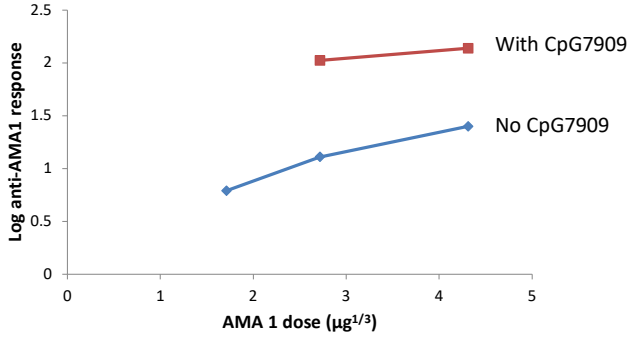
Three dimensional response surfaces

Lets move onto something more complicated – a two component dose response. Now we really will be looking at the three dimensional response surface.

Within the DoE universe there are multiple solutions to the most efficient design for a trials with 2 independent variants. In preliminary modelling, two designs stood out: a 3 x 3 or a 4x4 “Face Centred Design” Since there was no major difference in the 3x3 or 4x4, for the rest of this presentation I will use a 3x3 Face Centred design. This needs 9 combinations, preferably with doses of antigen and adjuvant equally spaced on the transformation used on antigen and adjuvant axes. I.e. the cube root of the middle dose is the average of the cube root of the high and low doses.

AMA1 malaria antigen + 568µg CpG7909 adjuvant

81

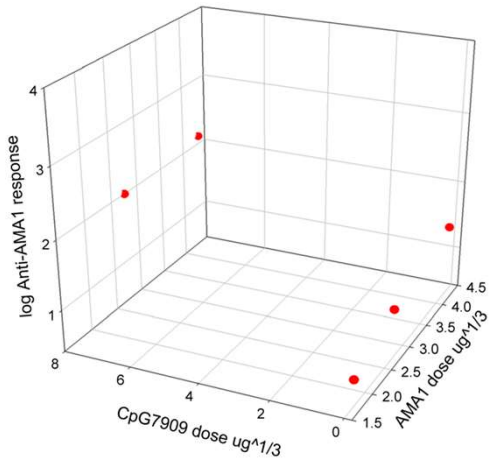


Mullen et al 2008 PLoS One 3:e2940; Malkin et al 2005 Infection and Immunity 73:3677–3685

Here is composite data from two malaria vaccine trials using an antigen called AMA1 and the same CpG7909 adjuvant used on an earlier slide with Engerix. The actual doses of AMA1 were 5, 20 and 100 µg and the 20 and 100 µg doses of AMA1 were also tested with 568 µg of CpG7909.

AMA1 + CpG 7909 response surface

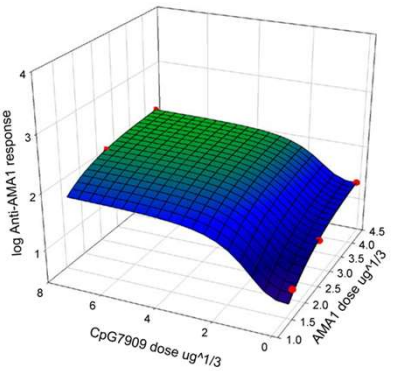
The back and front edge of a 3D response surface



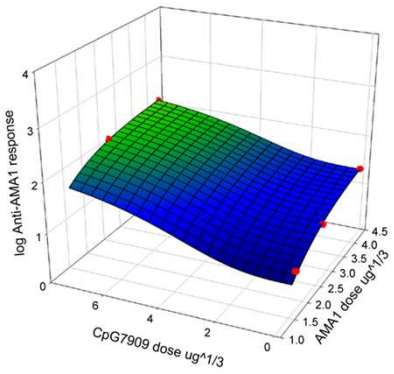
We have points on the “front” and “back” surface of a response curve

AMA1 + CpG 7909 theoretical response surface

Two hypothetical 3D response surfaces that fit



Antibody plateaux at low dose adjuvant



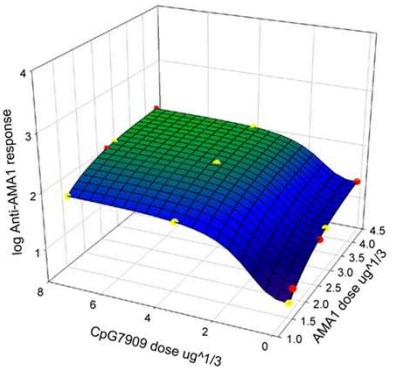
Antibody maximum requires high dose adjuvant

What we would really like to know is the actual (true) response curve in detail. Unfortunately all we have are the 5 points. However we can draw a series of hypothetical surfaces that exactly fit these 5 points. These hypothetical surfaces can then be used as the basis for modelling to see how a DoE approach would perform in picking the best combination of antigen and adjuvant.

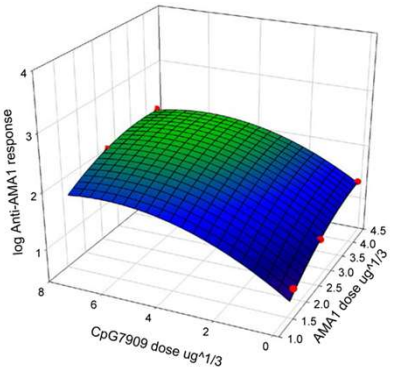
Here are two possible surfaces – how they are generated is not important but what is important is they give to fairly extreme examples of what may be possible. The response surface on the left assumes the adjuvant is AMA1 dose sparing and that the adjuvant increases the maximum response. The effect increases rapidly as low adjuvant doses are increased but that these effects rapidly plateau with increasing adjuvant. The surface on the right also assumes that the adjuvant is both AMA1 dose sparing and increases maximum response but requires much higher levels of adjuvant than the response surface on the left. Unlike the Engerix data, these simple hypothetical response surfaces do not assume that the response decreases at the highest adjuvant doses tested.

AMA1 + CpG 7909 theoretical response surface A

Adjuvant dose 0 to 500µg



Hypothetical surface with 9 groups for a 3x3 FCD (yellow dots)



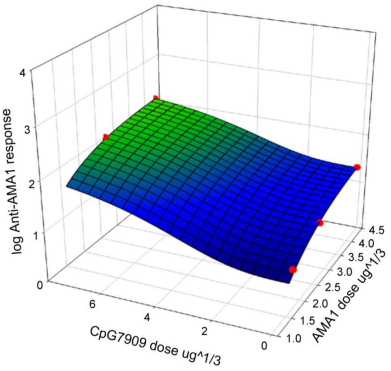
Best fit quadratic fit (zero variance in response)

So the first question: how adequate can a fit based on the 9 yellow points fit the hypothetical surface, before we start thinking about the natural variation we will see in the responses from individuals that get the same dose?

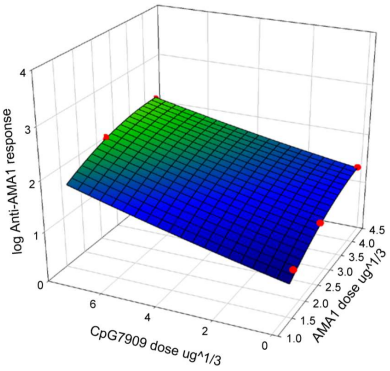
On the right is the best fit quadratic surface (i.e. Minimum sum of error squared). You can see it is not a perfect fit. For example, the lower edge of the hypothetical surface is sigmodal in shape but parabolic on the fitted surface. Not easily seen, but the maximum response on the fitted surface is not exactly at the highest dose of AMA1 and adjuvant. Never the less the fit over the plateaux is quite good. Is this "adequate"? Please bear with me and in a few slides we will look to answer that question a bit more quantitatively.

AMA1 + CpG 7909 theoretical response surface B

Adjuvant dose 0 to 500µg



Hypothetical surface



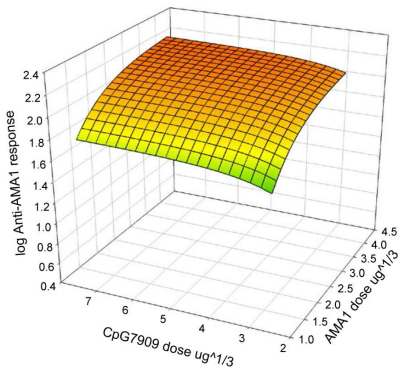
Best fit quadratic fit (zero variance in response)

Similarly, the fit to the second hypothetical response surface is also not perfect, but again “not bad” in the region of interest – the green area in the upper left

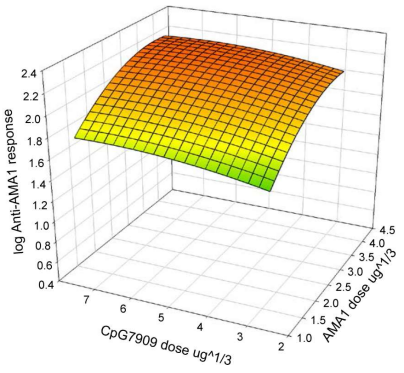
AMA1 + CpG 7909 response surface A

Theoretical vs best fit

Restricting range gives a very good fit: Adjuvant dose 25µg to 500µg



Hypothetical surface



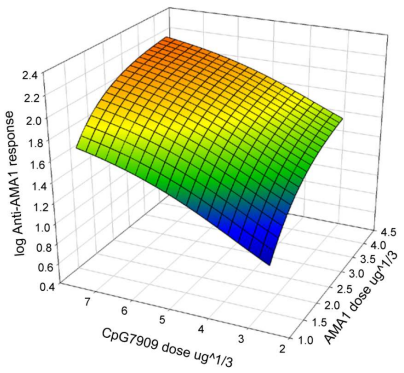
Best fit quadratic fit (zero variance in response)

DoE is all about pragmatisms A very small reduction in the range tested of the CpG7909 from a range of 0 to 500 µg to a range of 25 to 500 µg results in an even better fit with the fitted quadratic surface on the right and the hypothetical surface on the left for the first model

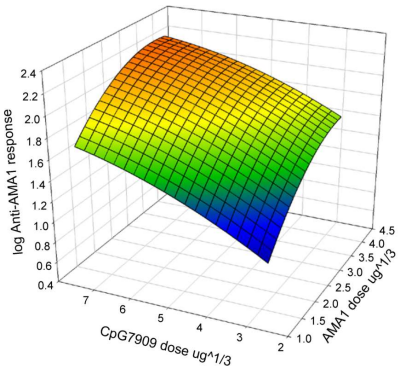
AMA1 + CpG 7909 response surface B

Theoretical vs best fit

Restricting range gives a very good fit: Adjuvant dose 25µg to 500µg



Hypothetical surface



Best fit quadratic fit (zero variance in response)

And again for the model that assumes that high doses of adjuant are needed.

Although this would be a simple way to improve the design of the trial, for the next section, I am going to stick to the 0 to 500 µg adjuvant range since that is a more stringent test of the modelling.

DoE modelling with vaccines with one antigen and one adjuvant

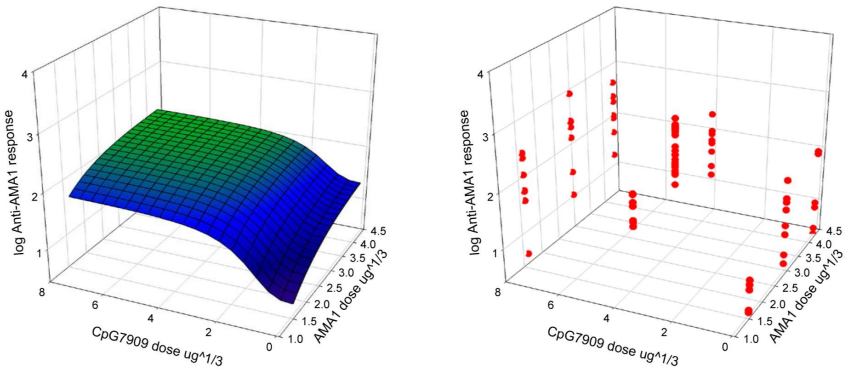
Using simulated data to find the combination giving the maximum antibody response

So it is one thing to fit a theoretical to a hypothetical surface in the absence of subject to subject variation, but in a real world this is not an option. The following slides used stochastic simulations of real trials. The simulation uses normally distributed random number generators that produce responses with the expected geometric mean and a standard deviation that matches that observed. Each time the simulation is run, there will be a slightly different set of responses. The number of subjects in each group is specified by commercially available DoE software (DX10)

In the following cases, the standard deviation is 0.5, close to the median of 0.51 observed in 110 published dose response vaccine trials.

AMA1 and CpG7909 A

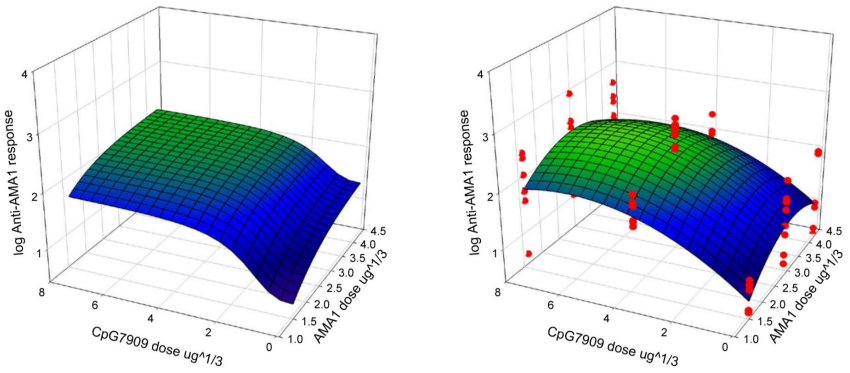
Simulated data (stdev of log antibody = 0.5) Total of 80 subjects



Again, hypothetical surface on the left. On the right is distribution of individual responses from one simulation with a total of 80 subjects. As is normal in a DoE setup, there are a lot more subjects in the central group (3X) than in the peripheral groups.

AMA1 and CpG7909 A

Simulated data (stdev of log antibody = 0.5) Total of 80 subjets

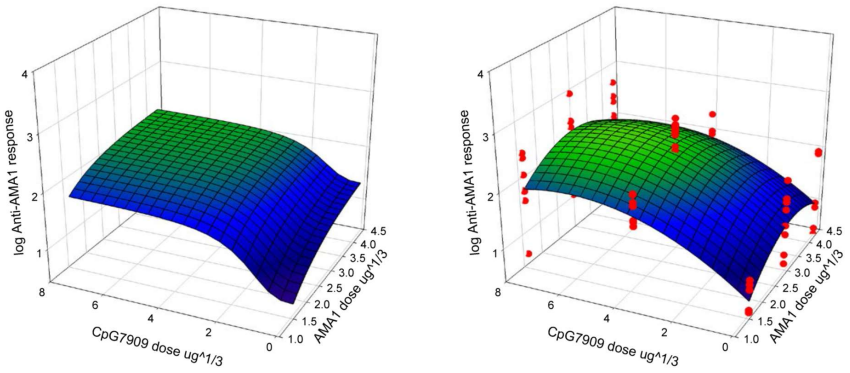


And here, on the right is the best fit quadratic surface to these scattered data. The fit is not as good as the fit in the absence of person to person variation, and again the critical question is if it is “adequate”. This surface will vary from simulation to simulation and so what we want to know is how the variation in this fit changes from simulation to simulation and as the number of subjects is changed.

AMA1 and CpG7909 A

Simulated data (stdev of log antibody = 0.5)

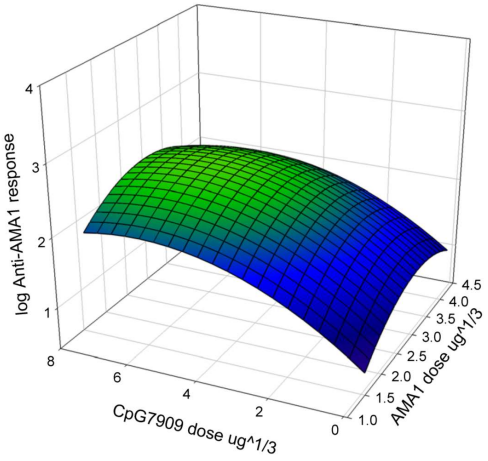
Variation in the fit in repeated simulations: 5 successive fits



And just so you don't think I am "Cherry Picking" I am going to show you the last 5 consecutive simulations done with this set of assumptions

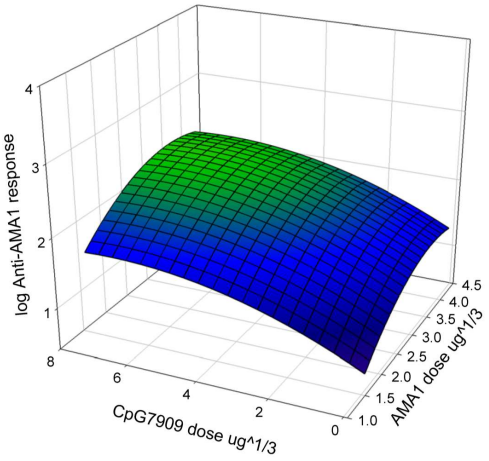
AMA1 and CpG7909 A

Simulated surface 1



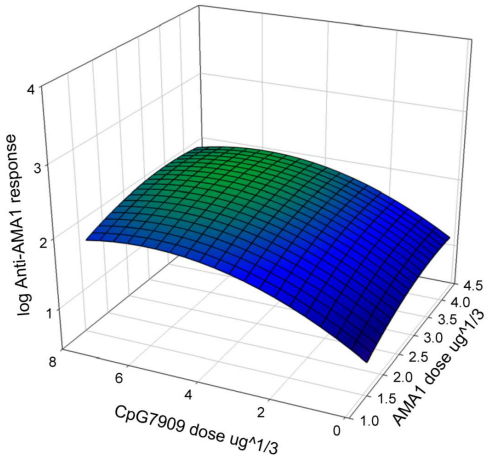
AMA1 and CpG7909 A

Simulated surface 2



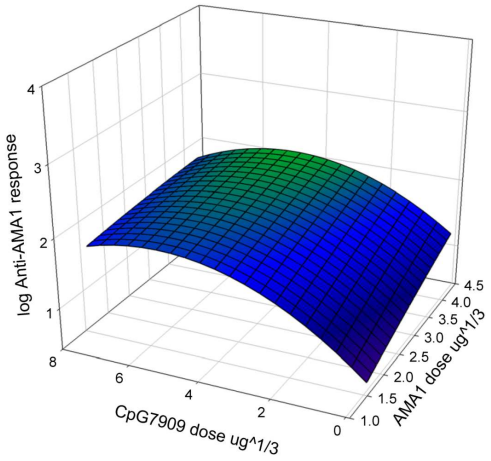
AMA1 and CpG7909 A

Simulated surface 3



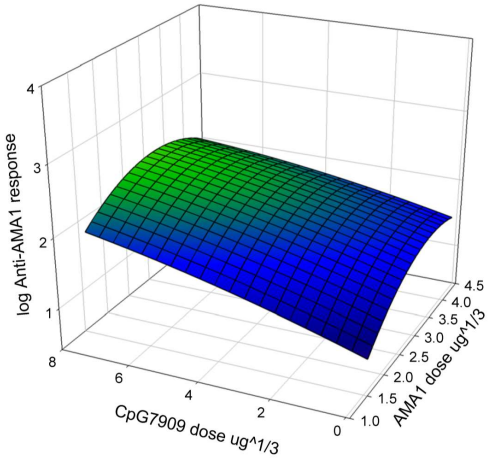
AMA1 and CpG7909 A

Simulated surface 4



AMA1 and CpG7909 A

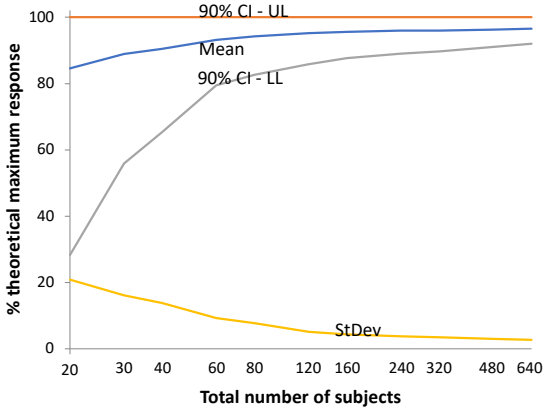
Simulated surface 5



AMA1 and CpG7909

Antibody response of predicted best combination vs theoretical maximum.

A total of 60 subjects spread over 9 groups has a 95% probability that the estimated best combination of antigen and adjuvant gives an antibody response >80% of the theoretical maximum



Here is a summary of the fit, as judged by maximum predicted antibody response of the combination of AMA1 and Adjuvant that gave the maximum antibody on the fitted surface. This "maximum predicted antibody" is the response judged from the hypothetical response surface at the predicted best combination from the fitted surface

There is a limit on how these data are determined that mimics a real life situation. If the fitted surface predicted that the maximum antibody response needed an antigen dose or an adjuvant dose higher than the maximum used in the toxicology study, then the model used a maximum "permissible" response on the edge of the hypothetical response surface. As a result, the average of these predicted "maximum responses" has to be less than 100% of the hypothetical maximum which lies on the exact corner of the hypothetical response surface.

This graphs shows the average, confidence interval and standards deviation of the predicted maximum response from 1000 simulations with the total subjects varying from 20 to 640 again assuming that the standard deviation of the log transformed antibody is 0.5.

So here is one definition of "adequate" – 60 subjects will give a response surface sufficiently accurate to give a 95% probability that the combination chosen will have a maximum GMC not less than 80% of the theoretical possible.

60 subjects total is a very small trial. This is an average of only 5.5 subjects in each of the peripheral combinations tested and 16 in the central group.

What else can we do with a response surface?

One outcome: the most economical antigen and adjuvant combination

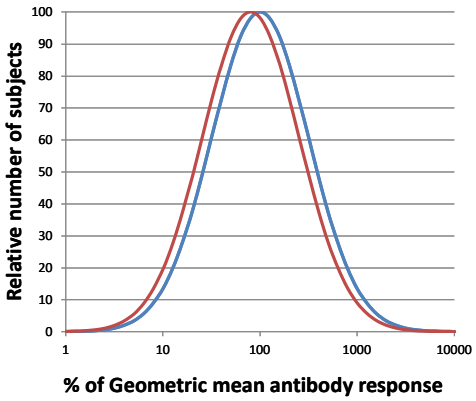
The trade-off between Immune response and resources

- How good does the response have to be?
  - Individual antibody responses typically vary over a factor of ±1000%
  - Would you be concerned if the average response was 20% lower than the theoretical best?
  - What if that meant you reduce the amount of production limited component by 4x and sell 4X as many vaccines?
  - Or alternatively in LMIC 4x more people could afford the vaccine?
  - Or you could roll out the vaccine 4X faster in the face of an epidemic?

Optimizing for what is “desirable” rather than only for maximum antibody

There are quite sophisticated ways of combining multiple responses as part of optomizing a DoE experiment. These could be antibody levels, reactogencity, cost etc. However in this talk I will use a very simple approach to address optimizing for a combination of antigen and adjuvant that gives an adequate response while minimizing cost.

What would be the impact of 20% decrease in maximum response?

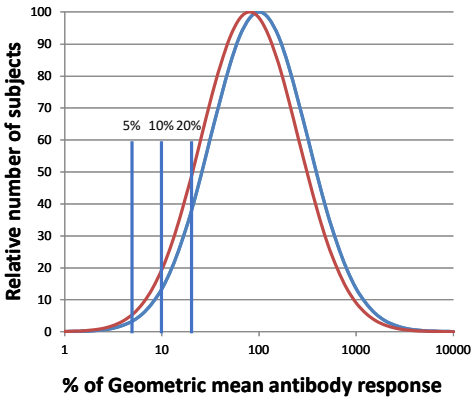


The distribution of antibody responses from a vaccine that was designed to give 80% of the maximum response will difficult to distinguish from a vaccine designed to give 100%.

What would be the impact of 20% decrease in maximum response?

Min. Protective level	% Protected	
	Vac A	Vac B
5% GMC	99.7	99.4
10% GMC	98.2	97.2
20% GMC	93.4	90.4

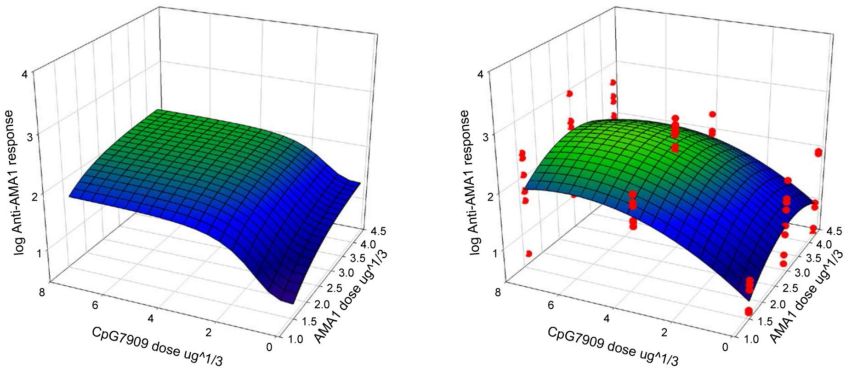
Vac A – Antigen and Adjuvant to give maximal response
Vac B – Antigen and Adjuvant to give 80% maximal response



By another measure – the % of people protected by the two vaccines will also be very similar depending on the minimum antibody level required for protection. Even for quite a "poor" vaccine where the minimum protection level is 20% of the GMC, the difference in % subjects protected by vaccines designed to give maximal or 80% responses will be very similar 93% vs 90%.

AMA1 and CpG7909 A

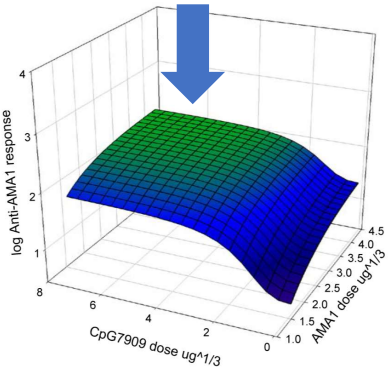
Simulated data (stdev of log antibody = 0.5)



The data you have now seen several time. However this time I am interested in the combination of antigen and adjuvant that gives a defined but less than maximal response. I have picked 80% . From the fitted surface we can predict this combination.

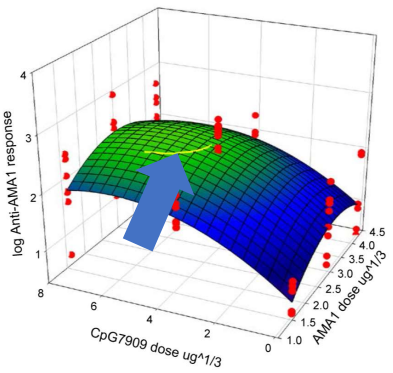
AMA1 and CpG7909 A

Simulated data (stdev of log antibody = 0.5)



On next slide we will see where the 80% contours of the 5 sucessive fits shown before, lie on the theoretical surface viewed from above

Yellow line is the 80% contour of the predicted surface



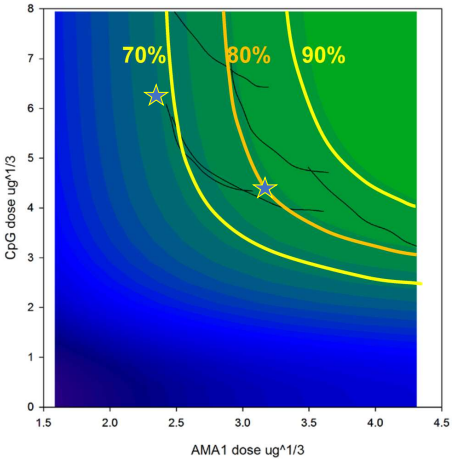
In fact there is not a single combination of antigen and adjuvant that will achieve this, but all combinations will lie on a contour line on the fitted surface – this yellow line shown in the right panel.

What we would really like to know is the combination of antigen and adjuvant that give the 80% response on the hypothetical surface but all we have available is the fitted surface.

However since this is a model we can ask the question “How accurately does the fitted 80% contour predict that hypothetical 80% contour?” and to do that through a series of simulations.

This gets a bit complicated to display so I am going to change views on the next slide to look down from above, i.e. from the perspective shown by the blue arrow in the left plot.

Combination of AMA1 and CpG to give predicted 80% response



Out of 1000 simulations with 80 subjects
Average antibody response: 80.0% max (90% CI 58% - 94%)
90% of low points between 47% and 90% of high points 68% to 93%

At an average antibody response of 80% max
Average AMA1: 30µg - 60% less than max
Average CpG: 155 µg – 70% less than max.

The stars mark the ends of the contour plot shown in the last slide projected on to this hypothetical surface.

The left end corresponds to AMA 1 and GpG doses of 14 and 225 µg (saving 83% and 55% of the antigen and adjuvant) to give 69% max response.

The right end corresponds to AMA 1 and GpG doses of 30 and 80 µg (saving 73% and 84% of the antigen and adjuvant) to give 77% max response.

In this contour plot the colours remain the same as on the 3D plots. Superimposed on this are the lines showing where the 70%, 80% and 90% of the hypothetical maximum response lie. Again, not to "Cherry pick" I am showing the last 5 simulations done with 80 subjects. The black lines show the position of the 80% contours from each of these 5 simulations, projected onto the hypothetical surface. As you can see for these 5 simulations, almost all of the lines lie between 70 and 90% contours

In 100 simulations, aiming at finding the combination that gave 80%, the midpoints of the lines on average predicted a combination of 83% (90% CI 72 to 93%) and even if you looked at the extreme values, i.e. The ends of these lines, the predicted combinations still give close to the 80% target.

At the average this would result in considerable savings of antigen and adjuvant. In a real world we don't have the ability to do 100 trials but as an illustration of what happens with just one trial: with the last simulation (shown with the stars), if we picked one end of that 80% fitted contour we would have saved 86% of antigen, 55% of adjuvant and delivered 69% max response or if we picked the other end we would have saved 70% of antigen, 84% of adjuvant and delivered 77% max response.

There is nothing magical about the 80% target. Since we know the equation to the response surface, AFTER the trial has been done, we could fit other contours (e.g. 50%

or 90%) to the fitted surface and do a similar analysis. It is hard to know how you would do this with a conventional study. Even if there was some way of guessing what combination to use, huge groups sizes would be needed to measure the appropriate efficacy.

Again this looks like an “adequate” fit of the theoretical response surface to address a very practical issue.

Conclusions and next steps

Conclusions

- Face centred DoE design with doses on a cube root scale can give adequate response surface for 2 component vaccines
- Number of subjects required to give an «adequate fit» is surpringly small
- Response surface can be used to determine minimum adequate doses thus filling aims of designing vaccines with high probability of efficacy and with minimum «cost»
- Simulation approach is both powerful and efficient
  - Provides a way of calculating group sizes.

Next steps

- Extension to more complex situations
  - Other response measures
    - Proportion of people “protected”
    - People with immune responses that cover a broad range of serotypes
  - Different vaccination schedules
  - Different age groups
  - More complex vaccines (Multiple antigens + adjuvant)
  - Extension to cover cost/manufacturability of the vaccines as part of the optimization
- More rigorous statistical analysis required
- Explore regulatory implications

This talk has just given a simple set of examples with a one antigen and one adjuvant dose response measuring antibody as the sole response. The DoE approaches is certainly not limited to this and in reality the utiltiy will increase as the study gets more complicated. The approach is also not limited to the numerical but categorical variables used in these examples. DoE can be applied to situations like looking at the impact on dose response with «Vaccine Schedule A» or «Vaccine schedule B»