



# Determining the size of a vaccine trial

Peter Smith

London School of Hygiene & Tropical Medicine

# SAMPLE SIZE IN VACCINE TRIALS



- **Phase 1** studies of vaccines are usually designed primarily to test safety in a small number of volunteers and, typically, formal sample size computations are not performed.
- **Phase 2** studies assess safety and immunogenicity (and sometimes efficacy) and may be used for dose finding/spacing/number optimisation – sample size computations may be done depending on the endpoints of primary interest.

# SAMPLE SIZE IN VACCINE TRIALS



- **Phase 3** (and some late stage **Phase 2**) studies of vaccines are usually designed primarily to measure **efficacy** where the main end-point is the occurrence of a specific event – usually a disease event.
- Such Phase 3 trials are often designated as “pivotal” trials, in that the results will be submitted to license a vaccine.
- The sample size of the trial is calculated in order that it will have a “high chance” (Power) of showing a “positive” (statistically significant) result if there really is a protective effect of the vaccine of some specified size.
- Safety is an important outcome in Phase 3 trials but is usually not a prime driver of study size – but sometime is – e.g. studies of rotavirus vaccines.



# SAMPLE SIZE IN VACCINE TRIALS

- Suppose we conduct a 2-arm placebo-controlled trial with a product that has zero efficacy – i.e. it is identical in effect to the placebo.
- Although we would expect the same disease rates in the two arms (vaccine and placebo), it is unlikely that the observed disease rates in the two arms would be identical.
- But any difference in the observed rates would be purely as a result of statistical fluctuations (if we toss a coin 10 times we would not necessarily see 5 heads and 5 tails).
- We can use statistical theory to specify the range in which we would expect the observed difference in rates to lie.



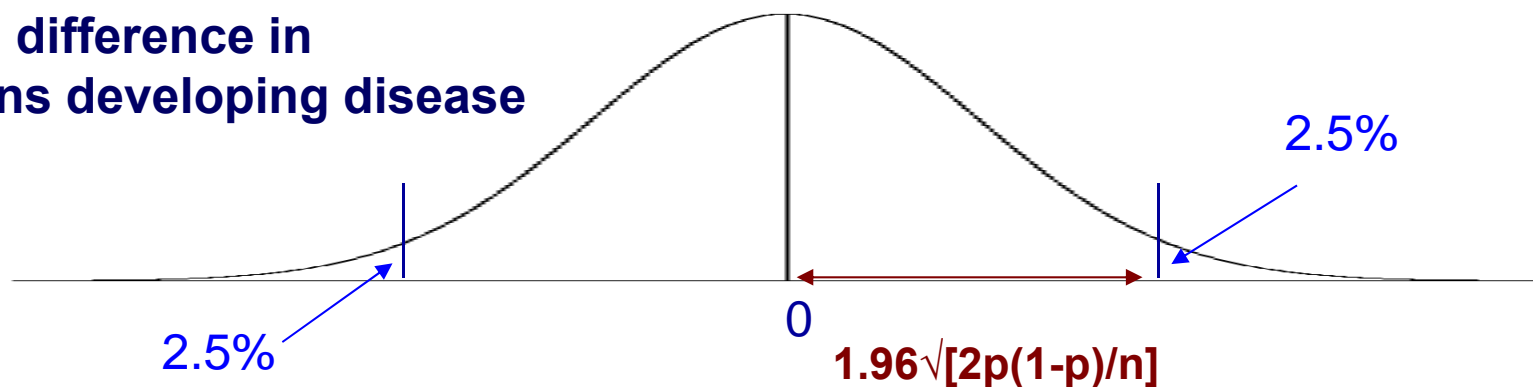
# SAMPLE SIZE IN VACCINE TRIALS

- Suppose we include “n” persons in each arm and the expected proportion who will develop disease in the study period is “p”
- In such circumstances there is a 95% chance that the observed difference in proportions with disease observed in the 2 arms will lie in the range:  
 $0 \pm 1.96 \times \text{Standard error of difference (SE)}$   
where  $SE = \sqrt{2p(1-p)/n}$

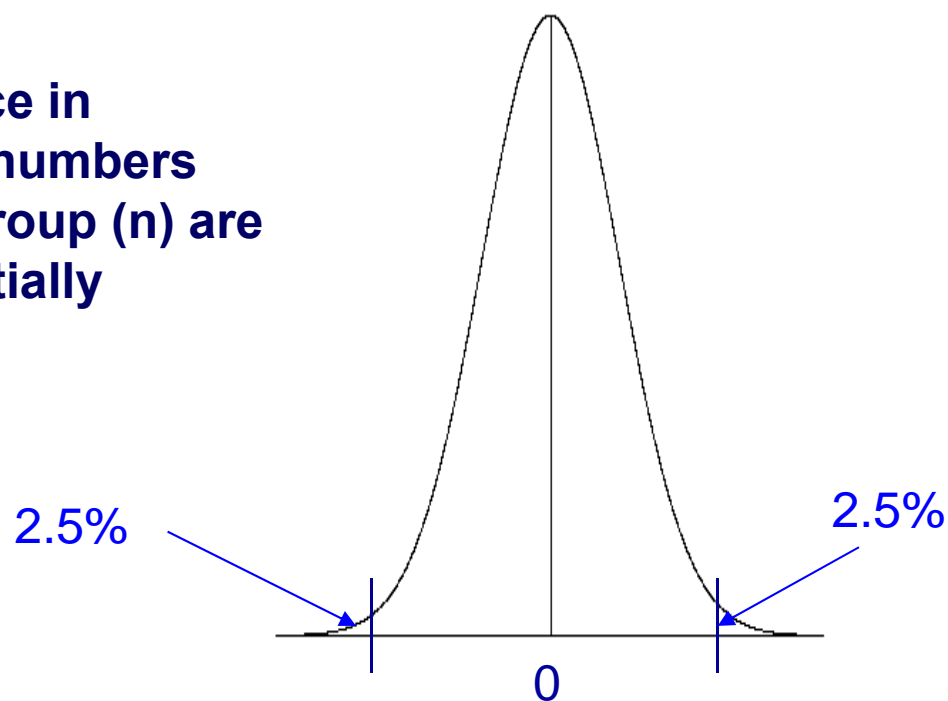


## Expected results of repeated trials when there is no true difference in disease rates in the 2 arms

Observed difference in proportions developing disease



Observed difference in proportions when numbers included in each group (n) are increased substantially

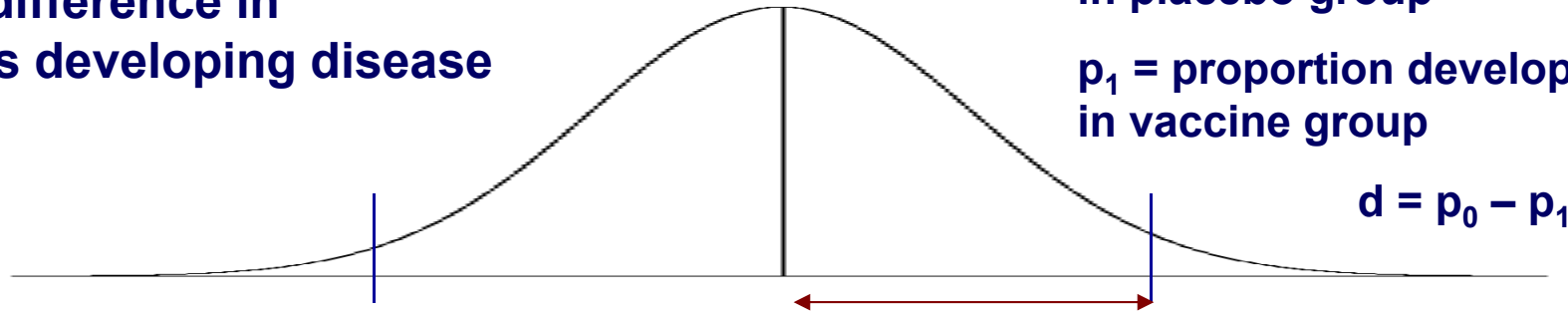


The larger the trial (2n) the narrower the range in which the observed difference would be expected to lie

# Expected results of repeated trials when there is a true difference (d) in disease rates between two arms



Observed difference in proportions developing disease



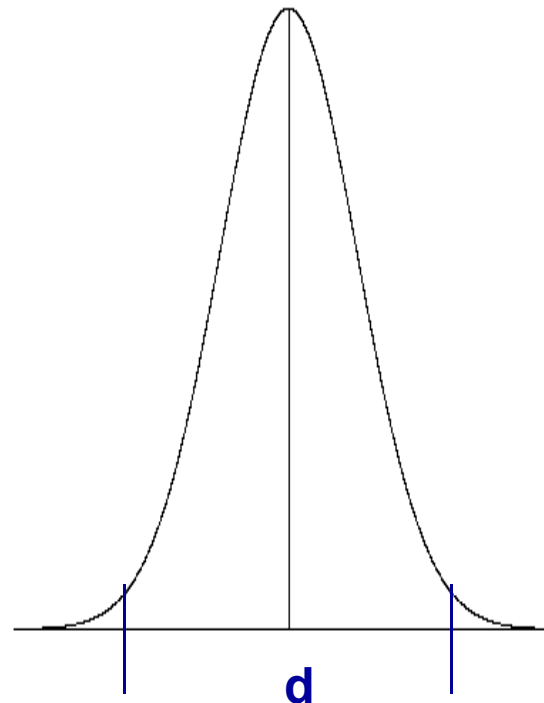
$p_0$  = proportion developing disease in placebo group

$p_1$  = proportion developing disease in vaccine group

$$d = p_0 - p_1$$

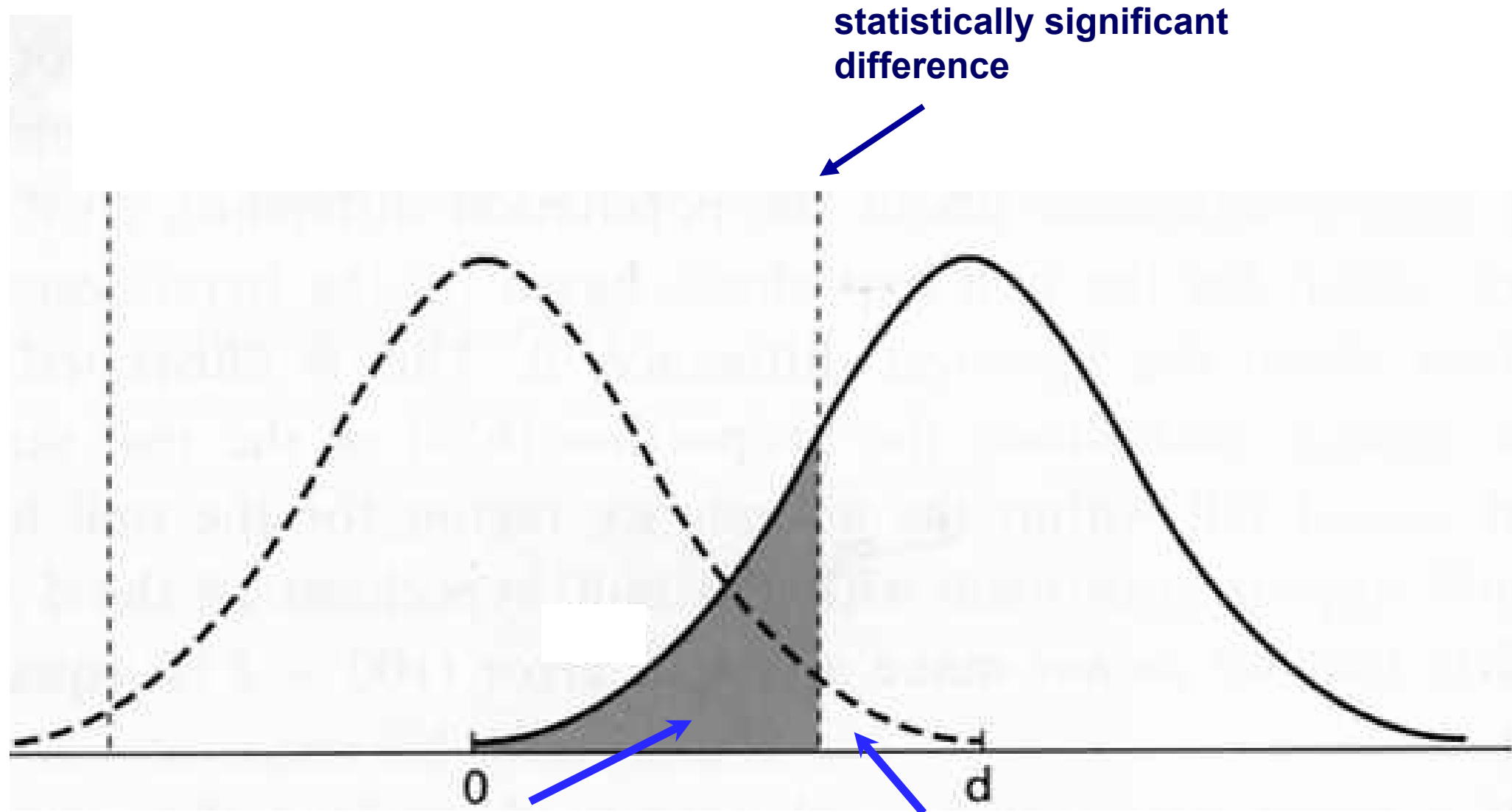
$$1.96\sqrt{[p_0(1-p_0)/n + p_1(1-p_1)/n]}$$

Observed difference in proportions when numbers included in each group are increased substantially



The larger the study the closer we would expect the observed difference to be to the true difference

# “Power” of a trial to detect an effect of a given size



Chance of “non-significant” difference when there is a true effect of size  $d$  ( $1 - \text{Power}$ )

Chance of “significant” difference when there is no true effect





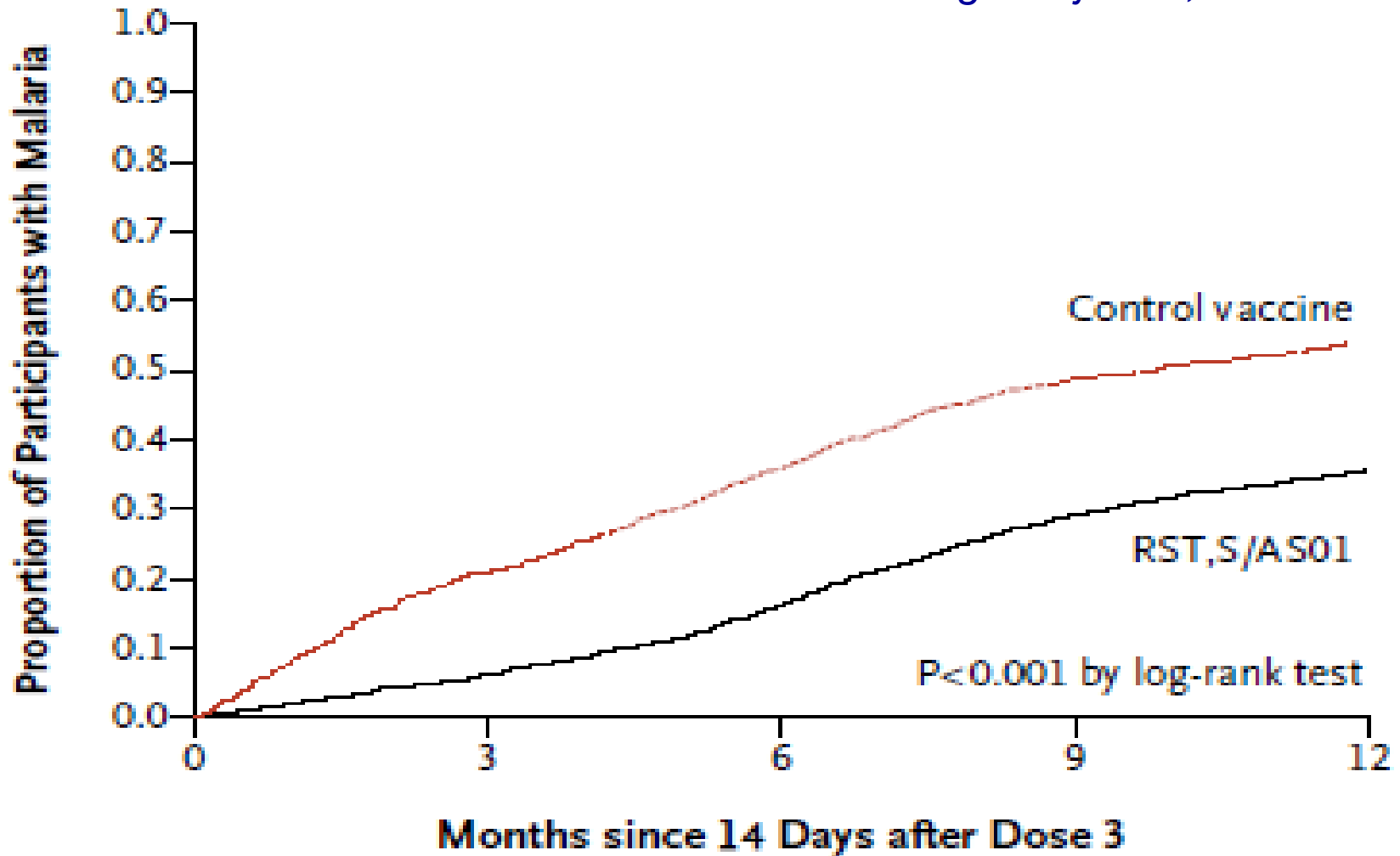
# Study size considerations

- What vaccine efficacy should a trial be designed to detect?
  - Proof of concept trial or pivotal trial?
  - Potentially licensable product?
  - What is a reasonable efficacy target?

# First Results of Phase 3 Trial of RTS,S/AS01 Malaria Vaccine in African Children aged 5-17 months



*Agnandji et al, 2011*

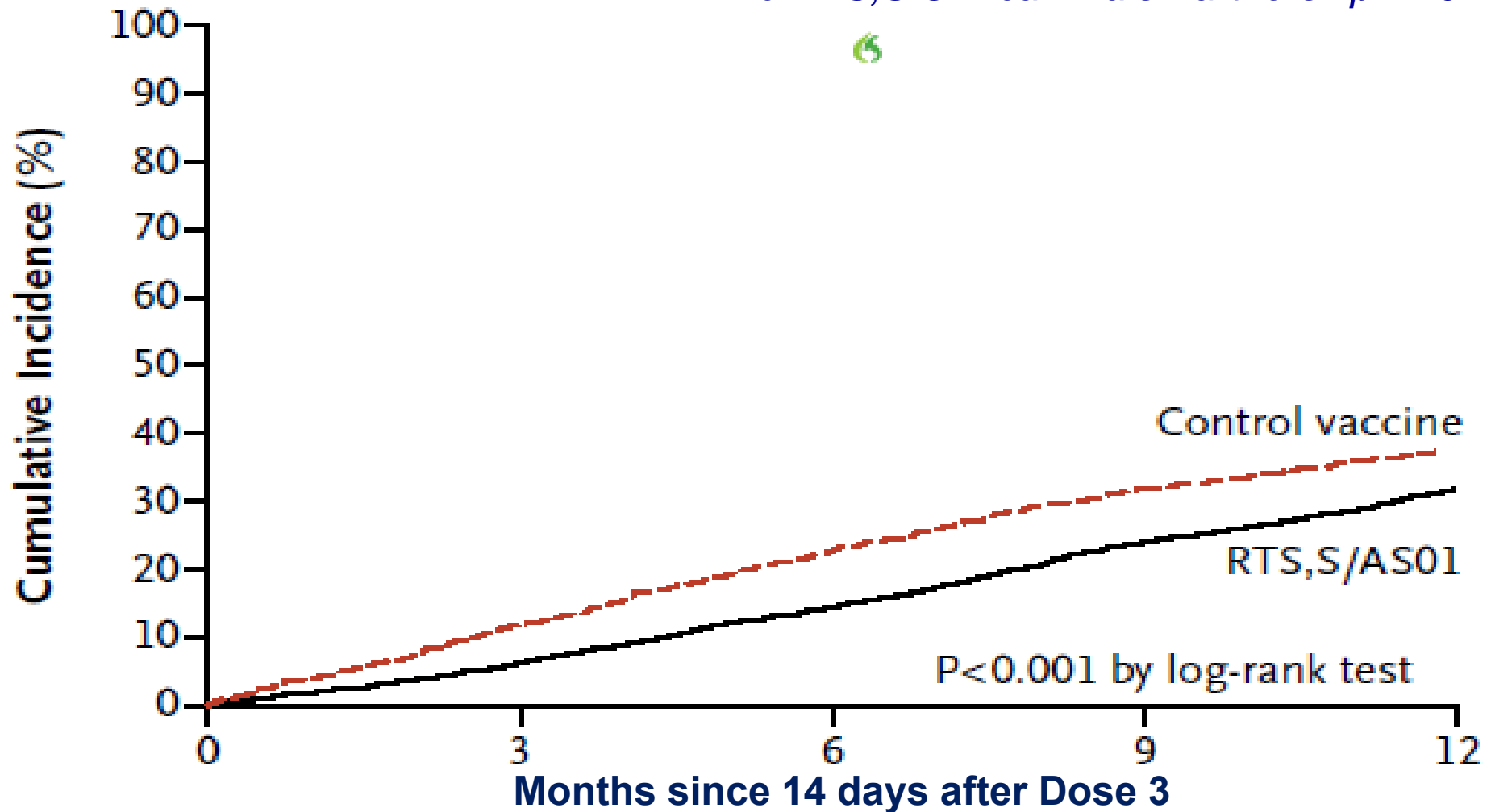


**Vaccine Efficacy against clinical malaria (per protocol analysis) = 47% (95% CI 22%, 64%)**

# RTS,S/AS01 Phase 3 trial – time to first malaria episode (infant vaccination)



*The RTS,S Clinical Trials Partnership NEJM 2012*



**Vaccine Efficacy against clinical malaria (per protocol analysis) = 30% (95% CI 24%, 36%)**

Efficacy of nine-valent pneumococcal conjugate vaccine against pneumonia and invasive pneumococcal disease in The Gambia: randomised, double-blind, placebo-controlled trial



*Cutts et al, 2005*

Endpoint	Vaccine (n=8189)		Placebo (n=8151)		Vaccine efficacy (95%CI)
	Number of cases	Rate/1000 years	Number of cases	Rate/1000 years	
Radiographic pneumonia	<b>207</b>	<b>22.1</b>	<b>323</b>	<b>35.1</b>	<b>37</b> (25 – 48)
Clinical pneumonia	<b>2172</b>	<b>231.5</b>	<b>2284</b>	<b>248.5</b>	<b>7</b> (1 – 12)
Hospital admissions	<b>1065</b>	<b>89.4</b>	<b>1216</b>	<b>104.7</b>	<b>15</b> (7 – 21)
Deaths	<b>330</b>	<b>25.2</b>	<b>389</b>	<b>30.1</b>	<b>16</b> (3 – 28)

# Vaccine Efficacy (95% CI) of GSK 2-dose Rotavirus Vaccine (2 weeks after dose 2 up to age 1 year)



	<b>Severe (requiring rehydration)</b>	<b>Hospitalization</b>
<b>Rotavirus gastroenteritis</b>	<b>85% (72%-92%)</b>	<b>85% (70%-93%)</b>
<b>Gastroenteritis from any cause</b>	<b>40% (28%-50%)</b>	<b>42% (29%-53%)</b>



## TRIAL SIZE TO ACHIEVE SPECIFIED POWER TO DETECT DIFFERENCE

Suppose the expected rate of the target disease is:

$r_0$  in the unvaccinated arm

$r_1$  in the vaccinated arm

i.e. Vaccine Efficacy (%) =  $100(1-r_1/r_0)$

Trial size to achieve the required statistical power

$$y = (z_1 + z_2)^2(r_0 + r_1)/(r_0 - r_1)^2$$

where:

$y$  = person-time in each arm

$z_1 = 1.96$  (for significance at  $P < 0.05$ )

=  $2.32$  (for significance at  $P < 0.01$ )

$z_2 = 0.84$  for 80% Power

=  $1.28$  for 90% Power

=  $1.64$  for 95% Power



## STUDY SIZE TO ACHIEVE SPECIFIED POWER TO DETECT DIFFERENCE

Comparison of 2 rates:

$$y = (z_1 + z_2)^2(r_0 + r_1)/(r_0 - r_1)^2$$

Another way of expressing this is:

$$yr_0 = (z_1 + z_2)^2(2 - v)/v^2$$

where  $yr_0$  = required number of cases in control group

$v$  = vaccine efficacy (measured as a proportion) =  $1 - r_1/r_0$

Study size is essentially determined by the number of endpoints required.

Total disease events required =  $y(r_0 + r_1)$

- Pneumococcal vaccine trial in The Gambia was designed to have 80% power to detect a vaccine efficacy against radiological pneumonia of 20% at the 5% significance level

$$yr_0 = (1.96 + 0.84)^2(2 - 0.2)/0.2^2 = 353$$

- The size of the study will depend upon how many child years of observation will be necessary to accumulate 353 cases of radiological pneumonia in the unvaccinated group



## Number of endpoints required in control arm for trials with specified Power to detect different vaccine efficacies

Vaccine Efficacy	P = 0.05 Power = 80%	P = 0.05 Power = 90%	P = 0.01 Power = 95%
20%	<b>353</b>	<b>472</b>	<b>824</b>
40%	<b>78</b>	<b>105</b>	<b>183</b>
60%	<b>30</b>	<b>41</b>	<b>71</b>
80%	<b>15</b>	<b>20</b>	<b>34</b>
90%	<b>11</b>	<b>14</b>	<b>25</b>



# Non-inferiority studies



- Suppose we have a vaccine with estimated vaccine efficacy of 85% and we wish to evaluate a new vaccine which we think is at least as efficacious as the existing vaccine. Proving equality of efficacy is impossible, but suppose we are prepared to consider it as “not inferior” to the existing vaccine if we can be reasonably sure that it is no more than, say, 5% less efficacious.

# Non-inferiority studies



If we intended to conduct a Phase 3 trial to evaluate a new vaccine compared to an existing vaccine, the number of disease event events that we would need to observe in the group allocated to the existing vaccine to be sure that it has “equivalent” efficacy – e.g. no worse than  $\delta$  less than the existing vaccine - is given by:

$$2(z_1 + z_2)^2(1 - v)^2/\delta^2$$

where:

$v$  = efficacy of existing vaccine (measured as a proportion)

$\delta$  = maximum difference in efficacy for new vaccine to be considered “non- inferior”

$z_1 = 1.64$  (for significance at  $P < 0.05$ ) – this time based on a one-sided test

= 2.05 (for significance at  $P < 0.01$ )

$z_2 = 0.84$  for 80% Power

= 1.28 for 90% Power

= 1.64 for 95% Power

# Non-inferiority studies



- Thus, suppose we have an existing vaccine that has 85% efficacy and we are planning to conduct a trial with a new vaccine which we think has similar efficacy.
- We want to be reasonable sure (say 80% Power) that, if the vaccines really are of equal efficacy, the 95% upper confidence bound on the difference in efficacy between the old vaccine and the new vaccine is less than 5%. We will need to expect to observe in the group receiving the old vaccine

$$2(z_1 + z_2)^2(1 - v)^2/\delta^2 = 2(1.64 + 0.84)^2(1 - 0.85)^2/.05^2 = \mathbf{111 \text{ cases}}$$

where:  $v$  = efficacy of existing vaccine = 0.85

$\delta$  = maximum difference in efficacy for new vaccine to be considered “non-inferior”  
=0.05

$z_1$  = 1.64 (for significance at  $P < 0.05$ ) – this time based on a one-sided test

$z_2$  = 0.84 for 80% Power

- This corresponds to an expected number of cases in the absence of vaccination of  $111/0.15 = 740$  and there would have to be a similar number of cases expected in the new vaccine group in the absence of vaccination – corresponding to a **total trial size of 1480** cases in the absence of vaccination.
- Because the numbers of cases required in such studies is very large they are rarely performed!



# Non-inferiority studies

- Non-inferiority trials are more usually conducted using some presumed surrogate measure of efficacy.
- Thus suppose the existing vaccine sero-converts 85% of those vaccinated and we think the new vaccine will be at least as good as this. When we do a sero-conversion trial we want to be reasonable certain that the upper confidence bound on the difference in sero-conversion rates between the old and the new vaccine will be less than, say, 5%.
- Then the sample size required in each group is

$$2(z_1 + z_2)^2 p(1-p) / \delta^2$$

Where:  $p$  = sero-conversion rate with existing vaccine (as a proportion)

$\delta$  = maximum difference in sero-conversion rates for new vaccine to be considered “non-inferior”

$z_1$  = 1.64 (for significance at  $P < 0.05$ ) – this time based on a one-sided test  
= 2.05 (for significance at  $P < 0.01$ )

$z_2$  = 0.84 for 80% Power  
= 1.28 for 90% Power  
= 1.64 for 95% Power

# Non-inferiority studies

- If the existing vaccine sero-converts **85%** of those vaccinated and we think the new vaccine will be at least as good as this. When we do a sero-conversion trial we want to be reasonable certain (say 80% Power) that the upper 95% confidence bound on the difference in sero-conversion rates between the old and the new vaccine will be less than, say, **5%**.
- Then the sample size required in each group is

$$2(z_1 + z_2)^2 p(1-p)/\delta^2 = 2(1.64 + 0.84)^2 0.85(1-0.85)/0.05^2 = \mathbf{627}$$

where:  $p$  = sero-conversion rate with existing vaccine (as a proportion) = 0.85

$\delta$  = max. difference in sero-conversion rates for new vaccine to be considered “non-inferior” = 0.05

$z_1$  = 1.64 (for significance at  $P < 0.05$ ) – based on a one-sided test

$z_2$  = 0.84 for 80% Power

- Thus we would **require a trial with 627 subjects in each arm, or 1254 in total**
- If we are prepared to use a  $\delta$  of 10% rather than 5% then the sample size required in each arm reduces to 156
- A web page where you specify the trial characteristics and the number is calculated for you is: [http://www.sealedenvelope.com/power\\_binary\\_noninferior.php](http://www.sealedenvelope.com/power_binary_noninferior.php)



# Non-inferiority studies

- It is relatively rare to conduct non-inferiority trials using disease end-points (as studies of very large size are required unless the disease is common)
- More usually based on comparisons of immunogenicity – even if the relationship between immunogenicity and protection is unclear.



# Other factors to consider when calculating study size

- Persons excluded because of protocol deviations (e.g. wrong age, wrong interval between vaccinations)
- Proportion of study subjects who may be lost to follow up
- Incidence of endpoints may be less than expected
  - “Natural” temporal variations
  - Preventive measures instituted in trial (e.g. distribution of bed-nets in malaria trial)
  - “Herd” effect because some in the population are vaccinated and will not transmit disease (may need cluster randomised trial to assess full effect of vaccination – individual protection plus herd immunity)
- Interest in duration of protection – though trials usually not powered to study whether efficacy changes with time since vaccination
- Lack of numbers *may* be compensated by increase in duration of follow up – but problems if protection wanes or endpoint is age-specific.



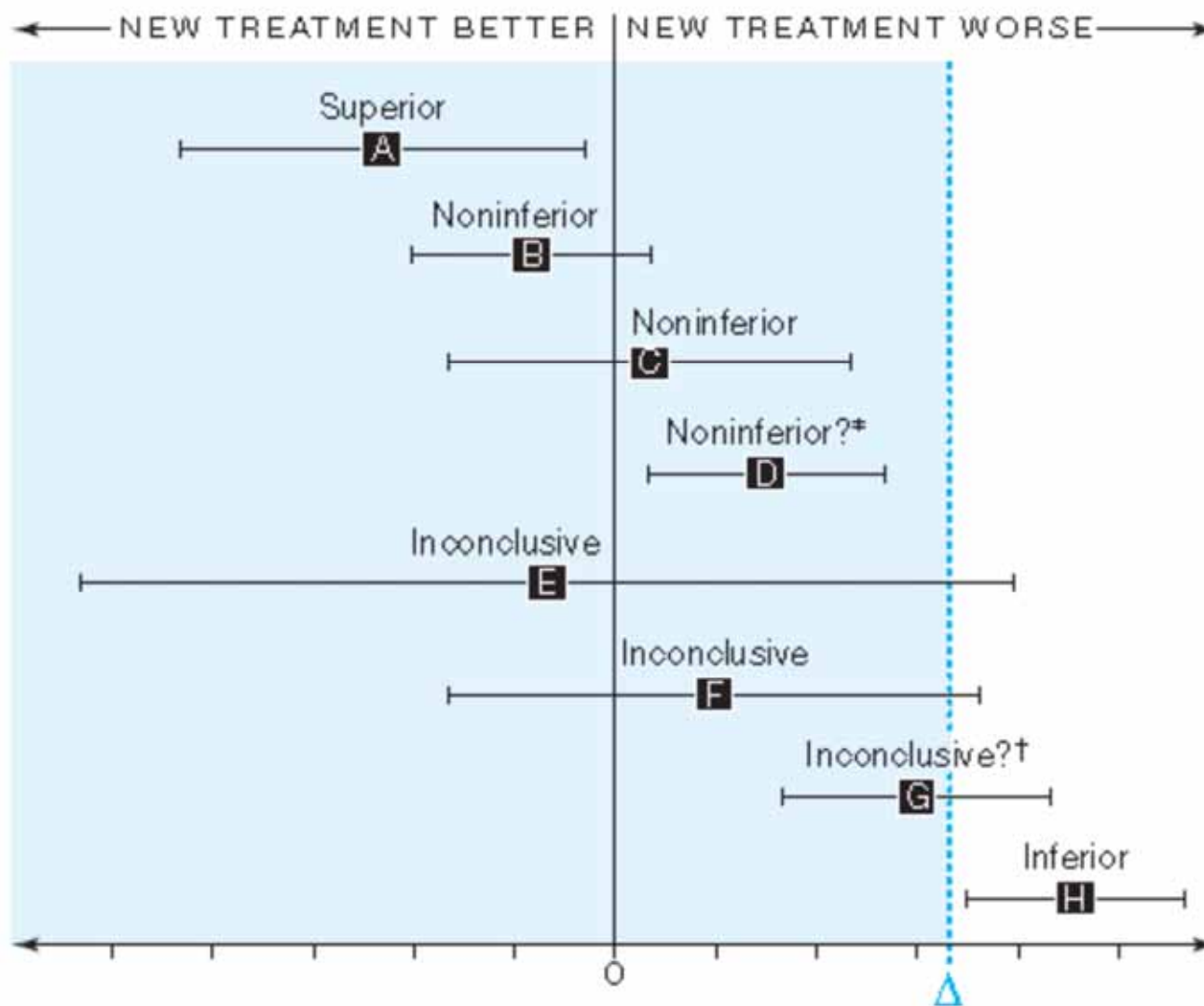
# References

- Blackwelder W, Chang M (1984). Sample sizes for “proving the null hypothesis” in clinical trials. *Controlled Clinical Trials* 5, 97-105. (*non-inferiority studies*)
- Cutts FT, Zaman SMA, Enwere G, et al (2005). Efficacy of nine-valent pneumococcal conjugate vaccine against pneumonia and invasive pneumococcal disease in The Gambia: randomised, double-blind, placebo-controlled trial. *Lancet* 365, 1139-1146.
- Kirkwood BR, Sterne JAC (2003). *Essential Medical Statistics* (2<sup>nd</sup> edition). Blackwell, Oxford.
- Smith PG, Morrow RH (Eds) (1996). *Field Trials of Health Interventions in Developing Countries: A Toolbox* (2<sup>nd</sup> Edition). MacMillan, London (3<sup>rd</sup> edition expected out in early 2014)





## Possible scenarios of observed treatment differences for adverse outcome



Treatment Difference for Adverse Outcome  
(New Treatment Minus Reference Treatment)