# Session 9
# Power and sample size

9.1 Measure of the treatment difference

9.2 The power requirement

9.3 Application to a proportional odds analysis

9.4 Limitations and alternative approaches

9.5 Sample size reviews

MRC Hubs for Trials Methodology Research
Edinburgh Hub

MRC Hubs for Trials Methodology Research
North West Hub

# 9.1 Measure of the treatment difference

Let $\theta$ measure the advantage of T over C

$\theta > 0$ T superior

$\theta = 0$ No difference

$\theta < 0$ T inferior

***Binary***

$p_T$: probability of success on T

$p_C$: probability of success on C

$$\theta = \log_e\left(\frac{p_T(1-p_C)}{p_C(1-p_T)}\right) \qquad \text{(log-odds ratio)}$$

***Ordered categorical data (assuming proportional odds)***

$Q_{kT}$: probability of being in $C_k$ or better on T

$Q_{kC}$: probability of being in $C_k$ or better on C

$$\theta = \log_e\left\{\frac{Q_{kT}(1-Q_{kC})}{Q_{kC}(1-Q_{kT})}\right\} \qquad \text{(log-odds ratio)}$$

## 9.2 The power requirement

**The null hypothesis of no treatment difference should be rejected at significance level $\alpha$ (*2-sided*), with probability $(1 - \beta)$, for a given magnitude $\theta = \theta_R$ of treatment difference**

*Prior to conducting the study, it is necessary to impose a model for the responses, in order to define the reference improvement*

# Information needed

Test of null hypothesis based on the assumption

$$\hat{\theta} \sim N\left(\theta,\ \frac{1}{w}\right)$$

Reject $H_0$ if

$$\left|\hat{\theta}\sqrt{w}\right| > c$$

where w will be a function of

1. **sample size**
2. **unknown parameters**

We need

$$P\left(\left|\hat{\theta}\sqrt{w}\right| > c; \theta = 0\right) = \alpha$$

i.e. $$P\left(\hat{\theta}\sqrt{w} > c; \theta = 0\right) = \alpha/2 \qquad \text{(i)}$$

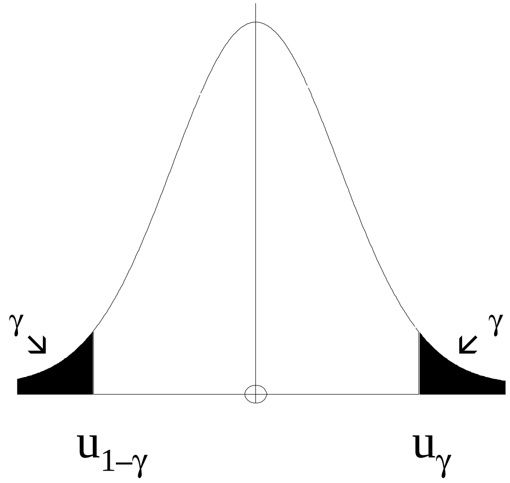
and $$P\left(\left|\hat{\theta}\sqrt{w}\right| > c; \theta = \theta_R\right) = 1 - \beta$$

i.e. $$P\left(\hat{\theta}\sqrt{w} > c; \theta = \theta_R\right) = 1 - \beta \qquad \text{(ii)}$$

as it is most unlikely that $\hat{\theta}\sqrt{w} < -c$ when $\theta = \theta_R$

*Standard normal density*

If $X \sim N(0, 1)$
then
$P(X > u_\gamma) = \gamma$




$u_{1-\gamma} = -u_\gamma$ *by symmetry*

Now $\hat{\theta}\sqrt{w} \sim N(\theta\sqrt{w}, 1)$

From (i) $P\left(\hat{\theta}\sqrt{w} > c; \theta = 0\right) = \alpha/2$

When $\theta = 0$ $\hat{\theta}\sqrt{w} \sim N(0, 1)$

So $c = u_{\alpha/2}$ (iii)

From (ii)

$$P\left(\left(\hat{\theta}-\theta_R\right)\sqrt{w} > \left(c-\theta_R\sqrt{w}\right); \theta = \theta_R\right) = 1-\beta$$

and when $\theta = \theta_R$, $\left(\hat{\theta}-\theta_R\right)\sqrt{w} \sim N(0,1)$

So $c-\theta_R\sqrt{w} = u_{1-\beta} = -u_\beta$ (iv)

Eq (iii) – Eq (iv) gives

$$\theta_R \sqrt{w} = u_{\alpha/2} + u_\beta$$

i.e.

$$w = \left( \frac{u_{\alpha/2} + u_\beta}{\theta_R} \right)^2$$

- This formula has general validity
- Can use $w = V$ (Fisher's information)

  *(Whitehead, 1996)*
- To obtain a sample size, $w$ must be related to $n$

  - This is the most approximate part of the procedure

## 9.3 Application to a proportional odds analysis

- Assume proportional odds
- Denote the log-odds ratio, measuring the advantage of T over C, by $\theta$
- Specify the difference sought (for which power is to be $1 - \beta$) as a value $\theta_R > 0$ of $\theta$

From Session 3

$$V = \frac{n_T n_C n}{3(n+1)^2}\left\{1 - \sum_{k=1}^{m}\left(\frac{n_k}{n}\right)^3\right\}$$

Suppose that it is intended that $n_T \approx n_C$, and anticipated that $n_k/n \approx \bar{p}_k,\ k = 1,\ldots, m$. Then

$$V \approx \frac{n}{12}\left(1 - \sum_{k=1}^{m}\bar{p}_k^3\right) \quad \text{so that} \quad n = \frac{12\left(u_{\alpha/2} + u_{\beta}\right)^2}{\theta_R^2\left(1 - \sum_{j=1}^{m}\bar{p}_k^3\right)} \qquad (9.1)$$

# Example: Head injury trial

- **Patients** – Head injury
- **Treatments** – Experimental drug vs placebo
- **Response** – Glasgow Outcome Scale at 3 months

  – Anticipated responses in placebo arm

| **Category** | **Good recovery** | **Moderate disability** | **Severe disability** | **Vegetative/ Dead** |
|---|---|---|---|---|
| Absolute Prob ($p_{kC}$) | 0.264 | 0.156 | 0.131 | 0.449 |
| Cumulative Prob ($Q_{kC}$) | 0.264 | 0.420 | 0.551 | 1 |

# Example: Head injury trial

- **Significance test** – 5% (two-sided)
- **Power** – 0.9
- **Clinically relevant difference** – proportion in Good recovery and Moderate disability categories to move from 0.42 on placebo to 0.52 on experimental drug

$$\theta_{\mathrm{R}} = \log_{\mathrm{e}}\left\{\frac{0.520(1-0.420)}{0.420(1-0.520)}\right\} = 0.403$$

Under the proportional odds model

$$e^{\theta_R} = \frac{Q_{kT}\left(1-Q_{kC}\right)}{Q_{kC}\left(1-Q_{kT}\right)}$$

so that

$$\frac{Q_{kT}}{1-Q_{kT}} = \frac{e^{\theta_R} Q_{kC}}{1-Q_{kC}}$$

that is

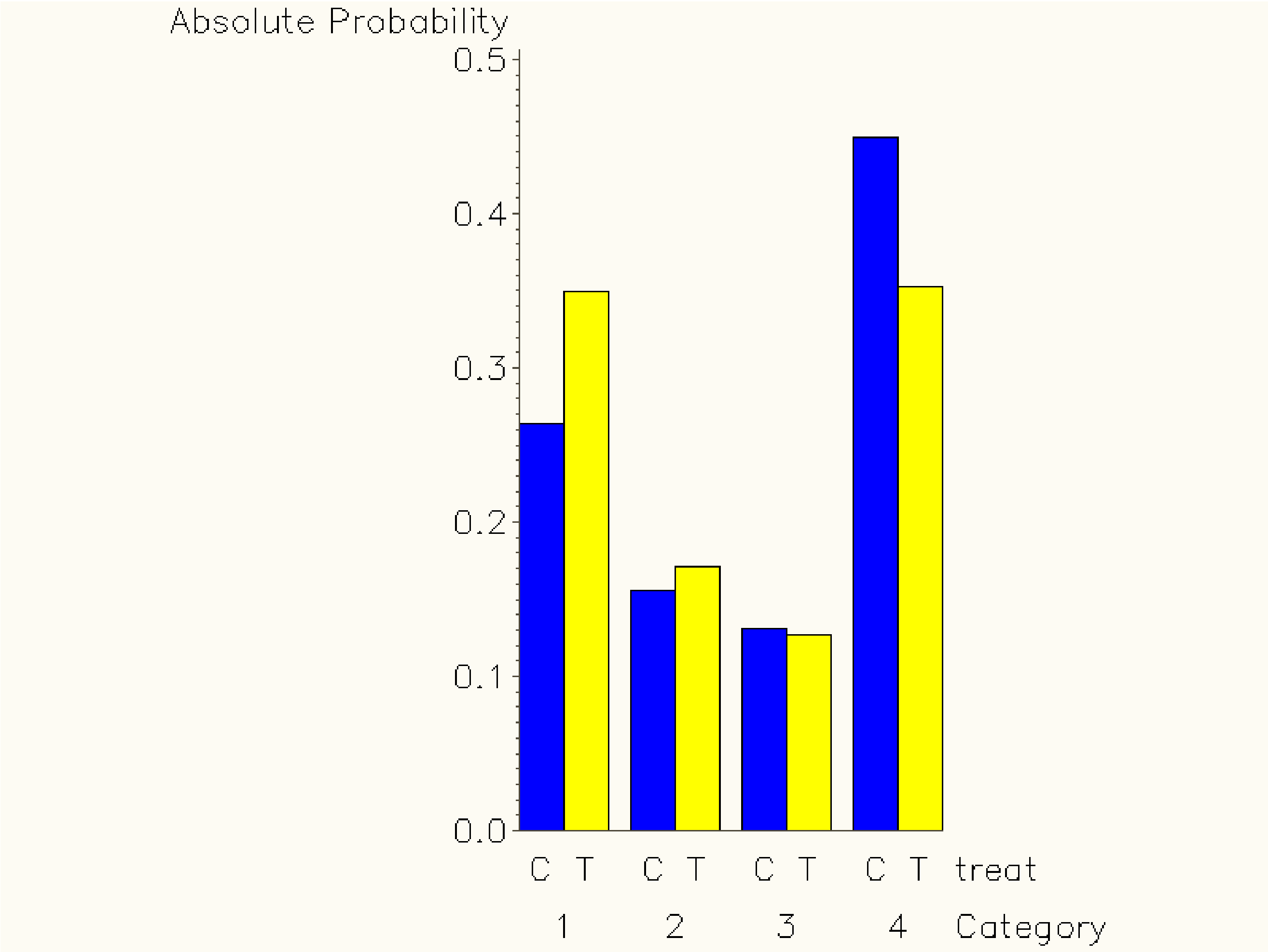
$$Q_{kT} = \frac{e^{\theta_R} Q_{kC}}{\left(1-Q_{kC}\right)+e^{\theta_R} Q_{kC}}$$

for $k = 1,2,3$

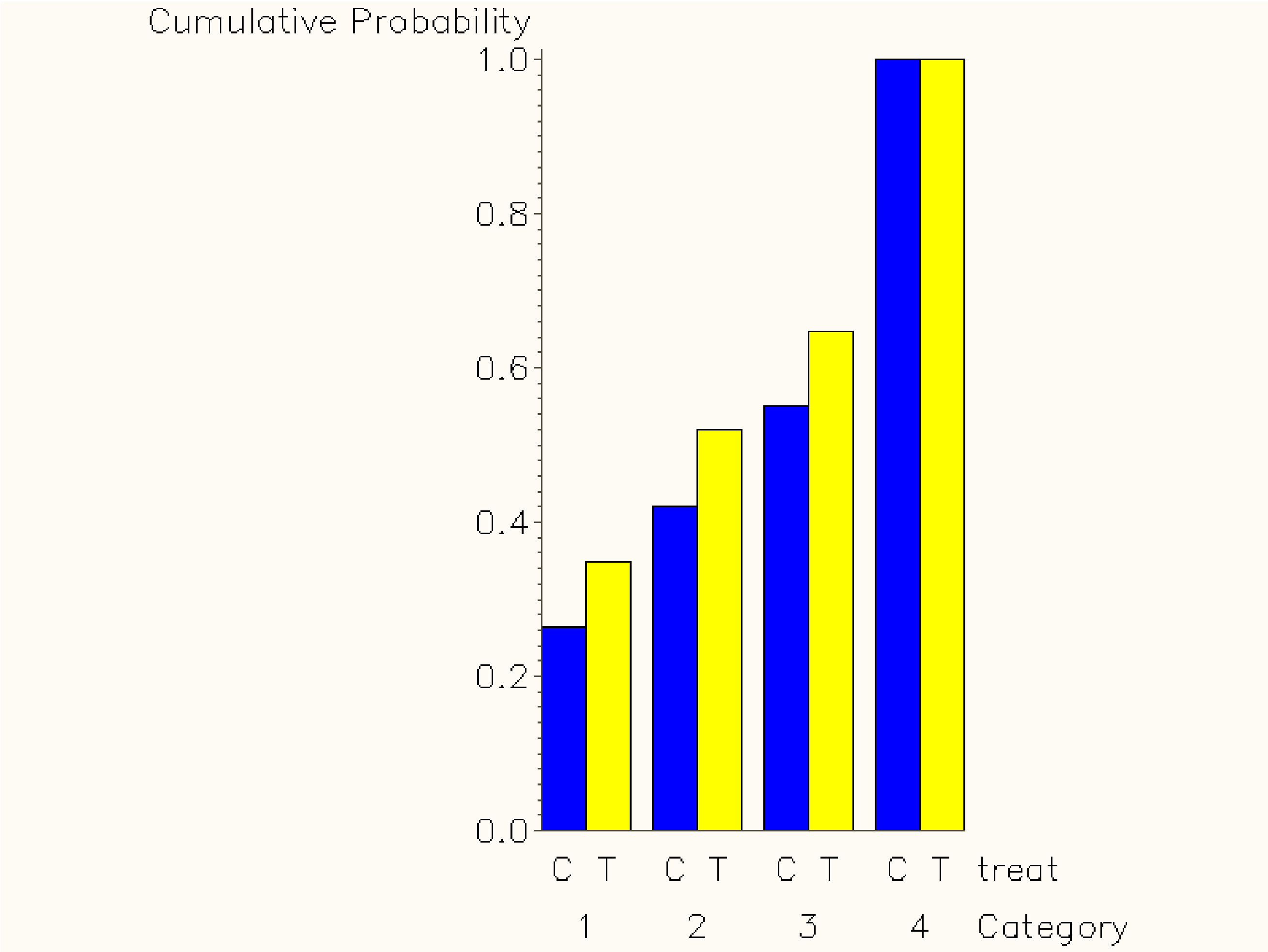
For $\theta_R$ = 0.403 and anticipated $Q_{kC}$ values

| Category | Good recovery | Moderate disability | Severe disability | Vegetative/ Dead |
|---|---|---|---|---|
| **Cumulative Prob ($Q_{kC}$)** | 0.264 | 0.420 | 0.551 | 1 |
| **Cumulative Prob ($Q_{kT}$)** | 0.349 | 0.520 | 0.647 | 1 |

| | | | | |
|---|---|---|---|---|
| **Absolute Prob ($p_{kC}$)** | 0.264 | 0.156 | 0.131 | 0.449 |
| **Absolute Prob ($p_{kT}$)** | 0.349 | 0.171 | 0.127 | 0.353 |
| **Average Absolute Prob ($\bar{p}_k$)** | 0.307 | 0.163 | 0.129 | 0.401 |




Is proportional odds assumption sensible?




Is proportional odds assumption sensible?

$$\alpha = 0.05 \qquad u_{\alpha/2} = 1.960$$

$$1 - \beta = 0.9 \qquad u_{\beta} = 1.282$$

$$1 - \sum_{k=1}^{4} \bar{p}_k^3 = 1 - 0.100$$

$$= 0.900$$

Hence

$$n = \frac{12\left(1.960 + 1.282\right)^2}{0.403^2 \times 0.900}$$

$$= 863$$

*That is the total sample size: 432 patients on each treatment*

# 9.4 Limitations and alternative approaches

Method is accurate if $\theta_R < 1$, and should be avoided if $\theta_R > 2$

*- in the example*

$\theta_R = 1 \Rightarrow n = 140$

$\theta_R = 2 \Rightarrow n = 35$

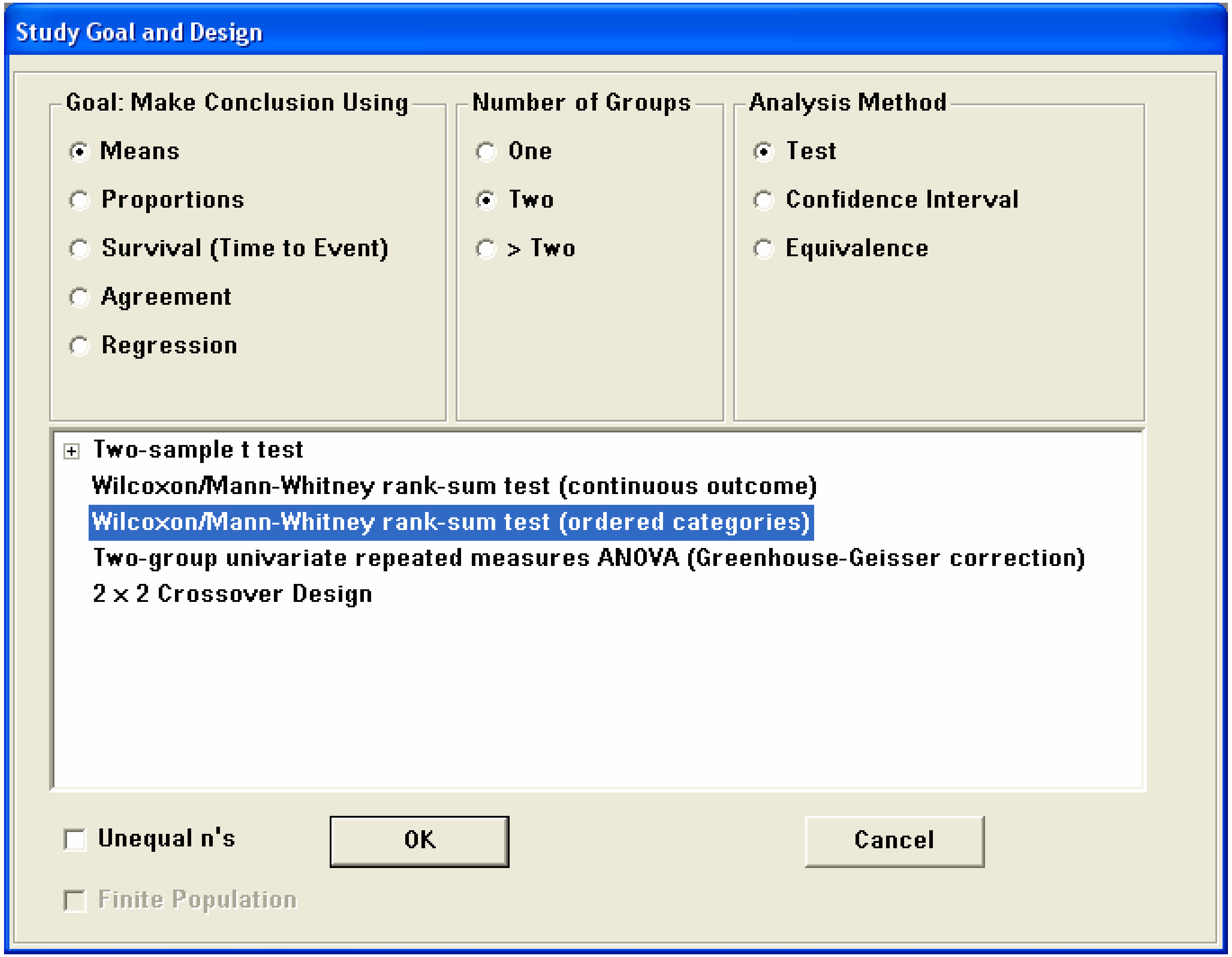
*To overcome*

- use an exact method *(Hilton and Mehta, 1993)*
- bootstrap
- simulate

Kolassa (1995) improves on equation (9.1), using a Cornish-Fisher approximation to the null distribution in place of the normal approximation

The method is implemented in the software nQuery Advisor

# nQuery Advisor: main menu



Entry of category probabilities for the two groups

MTT2-1: Aid table E1 -- Specification of probabilities in groups, categories

Compute | Transfer | Close | Restore | Clear | Cut | Copy | Paste | Print

| Category | Proportion in Group 1 (X) | Proportion in Group 2 (Y) |
|---|---|---|
| 1 | 0.264 | 0.349 |
| 2 | 0.156 | 0.171 |
| 3 | 0.131 | 0.127 |
| 4 | 0.449 | 0.353 |
| $\Sigma\pi_i$ | 1.000 | 1.000 |
| $p_1 = P(X<Y)$ | 0.440 | |

Calculation of power

| Wilcoxon (Mann-Whitney) rank-sum test that $P(X<Y) = .5$ (ordered categories) | | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| Test significance level, $\alpha$ | 0.050 | 0.050 | | | |
| 1 or 2 sided test? | 2 | 2 | | | |
| Number of categories, k | 4 | 4 | | | |
| Side table name | E1 | E1 | | | |
| $p_1 = P(X<Y)$ | 0.440 | 0.440 | | | |
| Power ( % ) | 89 | 90 | | | |
| n per group | 432 | 436 | | | |

Power is 0.89 for 432 patients per group - *as found from equation (9.1)*

Power is 0.90 for 436 patients per group

Lesaffre et al. (1993) present an alternative method based on simulation

Hilton (1996) evaluates the robustness of formula (9.1)

Julious and Campell (1996) examine (9.1) in the special case of binary data

Julious and Campell (1998) present formulae for the calculation of sample size for paired or matched ordered categorical data

# 9.5 Sample size reviews

Equation (9.1) is valid provided that:

- proportional odds hold
- $\bar{p}_k$'s are anticipated correctly

The latter can be checked at a *sample size review*

# Idea

1. Guess $\bar{p}_1, \ldots, \bar{p}_m$

2. Calculate n from equation (9.1) : denote value by $n_0$

3. Take $cn_0$ observations, $c \in (0, 1)$
   $\left(\text{e.g. } c = \tfrac{1}{2}\right)$

4. Estimate $\bar{p}_1, \ldots, \bar{p}_m$ from blinded data

5. Use estimates to recalculate n, denote value by $n_1$

6. Collect the remaining data needed to achieve this sample size

# Notes

- Final sample size must be $\geq cn_0$
- Can limit to values $\in (n_0, 2n_0)$, for example

*Gould (1992, 1995)* investigated the binary case, showed that type I error unaffected

## **Example in head injury** *(Bolland et al., 1998)*

Patients: suffering from severe head injury

Treatments: eliprodil vs placebo

Outcome: Glasgow Outcome Scale (GOS) six months after randomisation - ordinal

| | Proportion in each category | | |
|---|---|---|---|
| | GR | MD | SD/V/D |
| Placebo | 0.17 | 0.30 | 0.53 |
| Eliprodil | 0.274 | 0.346 | 0.38 |

Improvement to detect: GR + MD from 0.47 to 0.62

$$\theta_R = \log_e\left(\frac{Q_{kT}(1-Q_{kC})}{Q_{kC}(1-Q_{kT})}\right) = 0.610 \quad \text{for } k = 1, 2$$

$$n = \frac{12}{\left(1-\sum_j \bar{p}_k^{\ 3}\right)}\left(\frac{U_{\alpha/2}+U_\beta}{\theta_R}\right)^2 \qquad (9.1)$$

$$\alpha = 0.05,\ 1-\beta = 0.9,\ \bar{p}_1 = 0.222,\ \bar{p}_2 = 0.323,\ \bar{p}_3 = 0.455$$

giving n = 394

Total sample size rounded up to $n_0$ = 400

# Planned sample size review

- after responses from 100 patients
- timing at just beyond 9 months into trial after about 180 patients recruited (assuming entry rate of about 30/month)
- assessment of the need to adjust sample size for stratification
- new sample size to be used, n

$$n = \begin{cases} 400 \text{ if } n_1 + n_2 \leq 400 \\ n_1 + n_2 \text{ if } 400 < n_1 + n_2 < 600 \\ 600 \text{ if } n_1 + n_2 \geq 600 \end{cases}$$

# Actual sample size review

- responses from 93 patients
- 2 years into the trial
- stratification for Glasgow Coma Score at day 0 (4-5 vs 6-8)

$$n = \frac{12}{\sum_{h=1}^{2} S_h \left(1 - \sum_j \overline{p}_{kh}{}^3\right)} \left(\frac{U_{\alpha/2} + U_\beta}{\theta_R}\right)^2$$

- where $S_h$ is the proportion of patients in stratum h

| GCS at day 0 | Proportion of patients | Proportion in each category GR | MD | SD/V/D |
|---|---|---|---|---|
| 4-5 | 0.402 | 0.270 | 0.135 | 0.595 |
| 6-8 | 0.598 | 0.600 | 0.127 | 0.271 |

$n_1 + n_2 = 444$, rounded to 450

- recommendation of modest increase from 400 to 450 accepted by the Trial Steering Committee