ScHARR
School of Health and Related Research

Designs in cluster trials: Problems of recruitment bias and checking some simple adjustments to sample size formulas

Mike Campbell

Professor of Medical Statistics

University of Sheffield

Review of talk

- Design of a study to reduce recruitment bias
- Checking out common sample size formulas using simulation

Recruitment bias

Diabetes Care for Diagnosis
(Kinmonth et al BMJ 1998)

Trial to train General Practitioners (GPs)in 'patient centred care' for people with Type II diabetes

Setting: 41 practices (21T, 20C)

Subjects 250/360 individuals (30-70 yr.) newly diagnosed with Type 2 diabetes and completing follow up at one year .

Intervention: Aimed at GPs -1.5 days group training, introducing evidence for, and skills of, patient-centred care and a patient held booklet encouraging questions.

Main outcomes: quality of life, HbA1c, lipids, body-mass-index

Cannot blind study

Diabetes Care Trial

Outcome

142 patients in intervention arm , 108 in control

GPs who were trained in patient centred care more likely to diagnose Type II diabetes

DESMOND Trial (Diabetes Education and Self Management for Ongoing and Newly Diagnosed) Davies et al BMJ 2008

- Trial of structured group education in Type II diabetics
- 207 practices and 824 patients
- Treatment practice based
- Outcome HbA1c%, weight loss, depression score

Outcome from DESMOND

- 437 patients in intervention arm and 387 in control
- Same problem as Diabetes Care Trial!
- We reanalysed data using only patients in intervention who were matched to patients in the control for disease severity (HbA1c%)
- Conclusions unchanged (original result showed intervention to be ineffective)

Third Attempt!
REPOSE: Relative Effectiveness of Pumps Over MDI and Structured Education (Heller et al, NIHR 2010)

- Trial of insulin pumps versus multiple dose injections (mdi) in patients with Type 1 diabetes. About to start.
- 288 patients required – about 6 courses in 7 centres
- Patients trained in groups of 6 or 8 – thus clustered

Design of REPOSE

Principle: recruit and consent patients before randomisation

We took advantage of fact that can randomise clusters to treatment

- Patients recruited and consented, and given a choice of two dates for their course.
- Courses work best when there is synergy between participants – need at least 6 people and not more than 8 per course
- When both dates have recruited 8 subjects each THEN , randomise dates to either pump or mdi.

Problems

- Withdrawal after consent but before randomisation – substitute another patient.
- Withdrawal after randomisation but before treatment starts –Have a reserve list of people who could come on at least one of the dates and choose first person on reserve list but treatment is predetermined. If they refuse, then chose next on list.

Balancing baselines

- Would like to try and balance baseline HbA1c% between treatments
- Solution - if patient can come on either date we will use a minimisation algorithm to allocate patient to date. After 4th course, we will use a minimization algorithm to decide allocation to 5th and 6th courses

Conclusion

- Cluster trials in general are not blind
- If the clinician doing the recruitment is the same as the one giving the intervention, biases may occur
- REPOSE is an attempt to minimise these biases
- Somewhat complicated design

Sample size formula

Simple formula

Suppose
n is sample size for individually randomised design
n_c is sample size for cluster design

Then
n_c = n x design effect

Design effect= 1+(m'-1)xICC

m' is some 'averaged' cluster size –we will consider fixed m

$$ICC = \frac{\sigma_B^2}{\sigma_W^2 + \sigma_B^2}$$

σ_B^2, σ_W^2 are the between and within cluster components of variance

Given fixed cluster size ,m

$$Number\ of\ \ clusters\ \ per\ arm = k = ceiling\,(n/m)$$

$$n_{c,adj} = 2*k*m$$

$n_{c,adj}$ is the sample size rounded up to allow for complete clusters

Thus to allow for complete clusters , will have to recruit slightly more than minimum for given power

Practical difficulties with sample size estimation

1. Need to know *m* at analysis stage , not design– dropouts will reduce *m* and so reduce the design effect and thus the required sample size. How does this affect the power?
2. Often will use baseline covariates in the analysis. How does this affect the ICC and the power?

Analysis Method

- Used ‘gee’ in R – generalised estimating equations
- Uses a ‘working’ correlation to allow for correlation within clusters
- Used ‘exchangeable correlation structure’
- Conventional to require 40 clusters in total to use gee
- For this talk, restrict to Gaussian outcomes

Mancl's corrections

Mancl (2001) suggested

- estimate the standard error of the estimate by multiplying the robust se by √(2k/(2k-1))
- Estimate p-value from the z-statistics using the quantiles of a t distribution with 2k-p df where p is number of parameters in model, rather than the Normal distribution.
- 'gee' does not routinely give Mancl's corrections, so used R to derive them

Simulation exercise

Given effect size (1,3,5), s^2_B=1 and s^2_W=19 (ICC 0.05),

Standardised effect 0.2, 0.45, 0.67

Generate a sample size n, (2 sided 5% sig and 80% power) according to a standard formula and simulate

$$y_{ij} = \alpha + z_i + \delta d_i + \varepsilon_{ij} \qquad (1)$$

where i=1...2k indexes clusters,

j=1...m indexes patients within clusters

d_i=0 for i<=k, d_i=1 for i>=k+1,

δ is treatment effect

z_i is the cluster random effect N(0, s^2_W)

ε_{ij} is residual random effect N(0, s^2_B)

Power &Type 1 error

Power

Find n, and then simulate model (1) 5000 times

Find proportion of times p<0.05

Type 1 error

Use model

$$y_{ij} = \alpha + z_i + \varepsilon_{ij} \qquad (2)$$

using same n, and simulate model (1) 5000 times

Find proportion of times p<0.05

If this proportion is not 0.05 then cannot interpret power.

Can't 'fix' Type 1 error to determine power

Simulation result
number of subjects per cluster, m=2

α=0.05, power=0.8, ICC=0.05

Stand'd Effect size	n	Rounded n	Number clusters 2xk=n/m	Type 1 error	Mancl's correction	Power	Mancl's Correction Power	Mean Working Correlation
0.22	658.6	660	330	0.0568	0.0490	0.7958	0.7826	0.0455
0.45	164.6	168	84	0.0600	0.0490	0.8252	0.7948	0.0361
0.67	73.1	76	38	0.0614	0.0506	0.8150	0.7998	0.0253

Conventional GEE overestimates Type 1 error and power
Mancl's corrections perform well
Working correlation increasingly underestimated
as sample size drops

Simulation result
m, number of subjects per cluster=20

α=0.05, power=0.8, ICC=0.05

Stand'd Effect size	n	Rounded n	No. Clusters	Type 1 error	Mancl's Correction Type 1 error	Power	Mancl's Correction power	Mean Working correlation
0.22	1223.2	1240	62	0.0582	0.0432	0.8078	0.7932	0.047
0.45	305.8	320	16	0.0924	0.0476	0.8424	0.7802	0.038
0.67	135.9	160	8	0.1420	0.0468	0.8918	0.7690	0.025

Conventional GEE overestimates Type 1 error and power

Mancl's corrections slightly underestimates Type 1 error and as sample size drops, underestimates power

Working correlation increasingly underestimated as sample size drops

Allowing for missing values

Suppose the proportion missing was mv

Then find $n_{c,\ adjusted}=n/(1-mv)$

In the simulation, we allow missing values to occur at random thus the number of subject per cluster will vary, reducing power

Simulation 20% missing
m, number of subjects per cluster=2

α=0.05, power=0.8, ICC=0.05

Stand'd effect size	Unadjusted n	Unadjusted no clusters 2k	Adjusted n	Adjusted number of clusters	Type 1 error	Power	Mean Working correlation
0.22	658.6	330	828	424	0.0500	0.8080	0.0465
0.45	164.6	84	208	108	0.0518	0.7984	0.0402
0.67	73.1	38	92	46	0.0522	0.8038	0.0203

The adjustment procedure works well

Simulation 20% missing
m, number of subjects per cluster=20

α=0.05, power=0.8, ICC=0.05

Standard'd effect size	Unadjusted n	Unadjusted 2k	Adjusted n	Adjusted 2k	Type 1 error	Power	Mean Working correlation
0.22	1223.2	62	1560	78	0.0514	0.8452	0.0469
0.45	305.8	16	400	20	0.0504	0.8238	0.0393
0.67	135.9	8	200	10	0.0636	0.8290	0.0270

Except for design with 8 clusters Type 1 error well controlled , but greater than expected power

Allowing for covariates

Model

$$y_{ij} = \alpha + z_i + \beta x_{ij} + \delta d_i + \varepsilon_{ij} \quad (3)$$

where

$$x_{ij} = z_i + \eta_{ij}$$

and let η_{ij} have same distribution as ε_{ij}

Thus x and y share the same random effect

We used $\beta=0.5$

Allowing for covariates(2)

Revised sample size $n_{adj}=n*(1-r^2)$

Reduction in sample size based on

$$Var(y \mid x) = Var(y) - \beta^2 Var(x) = Var(y)(1-r^2)$$

Under model 3

$$Var(y) = (1+\beta)^2 \sigma_B^2 + (1+\beta^2)\sigma_w^2$$

$$Var(x) = \sigma_w^2 + \sigma_B^2 < Var(y) \quad if \quad \beta > 0$$

$$Var(y \mid x) = \sigma_w^2$$

Allowing for covariates (3)

If we use $\text{var}(y \mid x) = (\sigma_B^2 + \sigma_W^2)(1 - r^2)\ (4)$ in the sample size calc

and in fact $\text{var}(y \mid x) = \sigma_W^2$

Then $(1 - r^2)(\sigma_w^2 + \sigma_B^2) < \sigma_W^2$

means that the variance is reduced by too great an amount

This occurs when $r^2 > ICC$

Thus if we use (4) in the sample size calc
the sample size will be too small unless $r^2 < ICC$

Allowing for covariates(4)

In simulation r^2 is about 0.25, ICC is 0.05,

Sample size based on variance of $\sigma_w^2 + \sigma_B^2 = 20$ and r=0.5 which gives a reduced variance of 15 in adjusted formula.

However conditional variance is in fact $\sigma_w^2 = 19$, so formula assumes a variance that is too small and the power is underestimated.

Simulation allowing for x covariate m, number of subjects per cluster=2

α=0.05, power=0.8, ICC=0.05, r=0.5

d	Un-adjusted n	Unadj No clusters	Adj N	Adj. No clusters	Type 1 error	Power	Unadj. Working correl.	Adj. Working correl.	Correl. y and x
0.22	658.6	330	496	248	0.0452	0.6872	0.0455	0.0422	0.479
0.45	164.6	84	124	62	0.0578	0.6678	0.0361	0.0269	0.471
0.67	73.1	38	56	28	0.0554	0.6700	0.0253	0.0056	0.460

The adjustment procedure inflates the Type 1 error slightly but loses power.

The working correlation is reduced relative to that estimated without the covariate.

Simulation allowing for x covariate m, number of subjects per cluster=20

α=0.05, power=0.8, ICC=0.05, r=0.5

Standard'd effect size	Un-adjusted n	Unadj No. clusters	Adj. n	Adjusted No. clusters	Type 1 error	Power	Unadj Working Correl.	Adj. Working correl.	Correl. y and x
0.22	1223.2	62	920	46	0.0568	0.6876	0.047	0.0436	0.479
0.45	305.8	16	240	12	0.0606	0.6422	0.038	0.0318	0.469
0.67	135.9	8	120	6	0.0648	0.6088	0.025	0.0144	0.456

The adjustment procedure inflates the Type 1 error but reduces power.

Conclusions

- GEE performs well with continuous data and only slightly underestimates power when number of clusters=8 (4 per arm) and m=20, using small sample corrections

- The correction for missing values performs reasonably wellwell , perhaps because although design effect is smaller (leading to greater power, cluster size varies, which will reduce power). With larger clusters overestimates samples size (missing values less important with large clusters).

- The simple adjustment for a covariate resulted in lower than expected power, because adding a covariate did not reduce variance as much as expected. However this depends on the expected relationship between y and x.