METHODOLOGICAL DEVELOPMENTS OF CLUSTER RANDOMISED TRIALS (CRT) HOW TO (DESIGN AND) ANALYSE A CRT WITH REPEATED MEASURES?

B. Pereira, H Sancho-Garnier and A. Kramar

CRLC Val d'Aurelle – Biostatistics Unit (Montpellier, France)

bruno.pereira@valdorel.fnclcc.fr

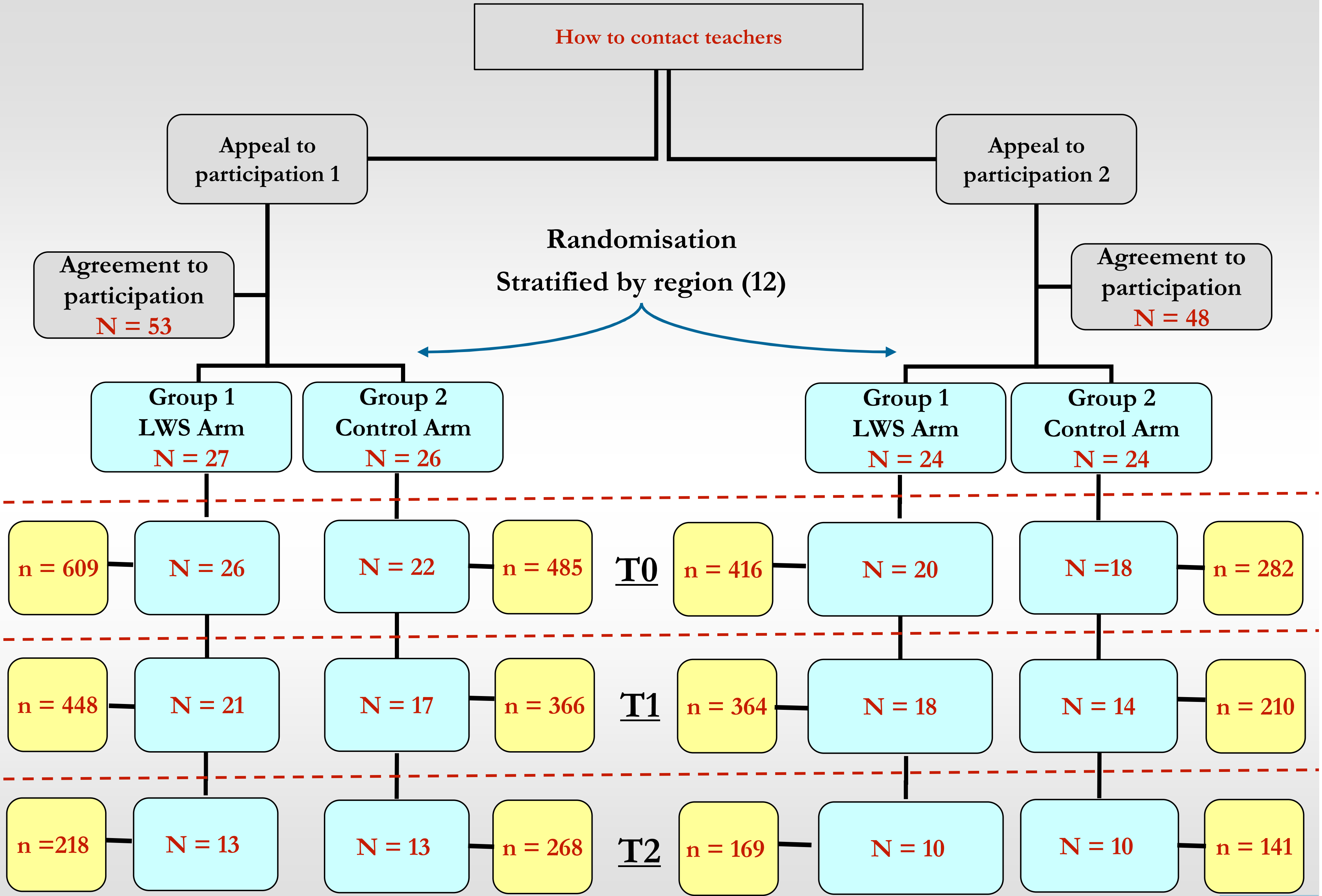
Context

- Cluster randomisation trials are an important tool in the evaluation of nontherapeutic interventions including lifestyle modification or educational programs
- A cluster randomized trial is a trial in which clusters of individual are randomized to different intervention groups rather than the individuals themselves
- **Objective:** To consider a Cluster-RT in which each subject is followed during several visits ⇨ The repeated observations (level 3) are nested within subjects (level 2) who are themselves nested within the unit of randomization (level 1)

Application – Trial *Living with sun*

<u>Objective</u>: Evaluation of an educational action

- To show that a process in health education involves:
  - A progression of knowledge
  - A modification of the attitudes and behaviors both in the children and their teachers as regards prevention against risks related to solar over-exposures
- To heighten and imply all members of the educational community
- To induce prevention reflexes in educators and to initiate scientific approaches to health problem


How to contact teachers
Appeal to participation 1
Appeal to participation 2
Randomisation
Stratified by region (12)
Agreement to participation
N = 53
Agreement to participation
N = 48
Group 1
LWS Arm
N = 27
Group 2
Control Arm
N = 26
Group 1
LWS Arm
N = 24
Group 2
Control Arm
N = 24
T0
n = 609
N = 26
N = 22
n = 485
n = 416
N = 20
N =18
n = 282
T1
n = 448
N = 21
N = 17
n = 366
n = 364
N = 18
N = 14
n = 210
T2
n =218
N = 13
N = 13
n = 268
n = 169
N = 10
N = 10
n = 141

GP Tobacco LR

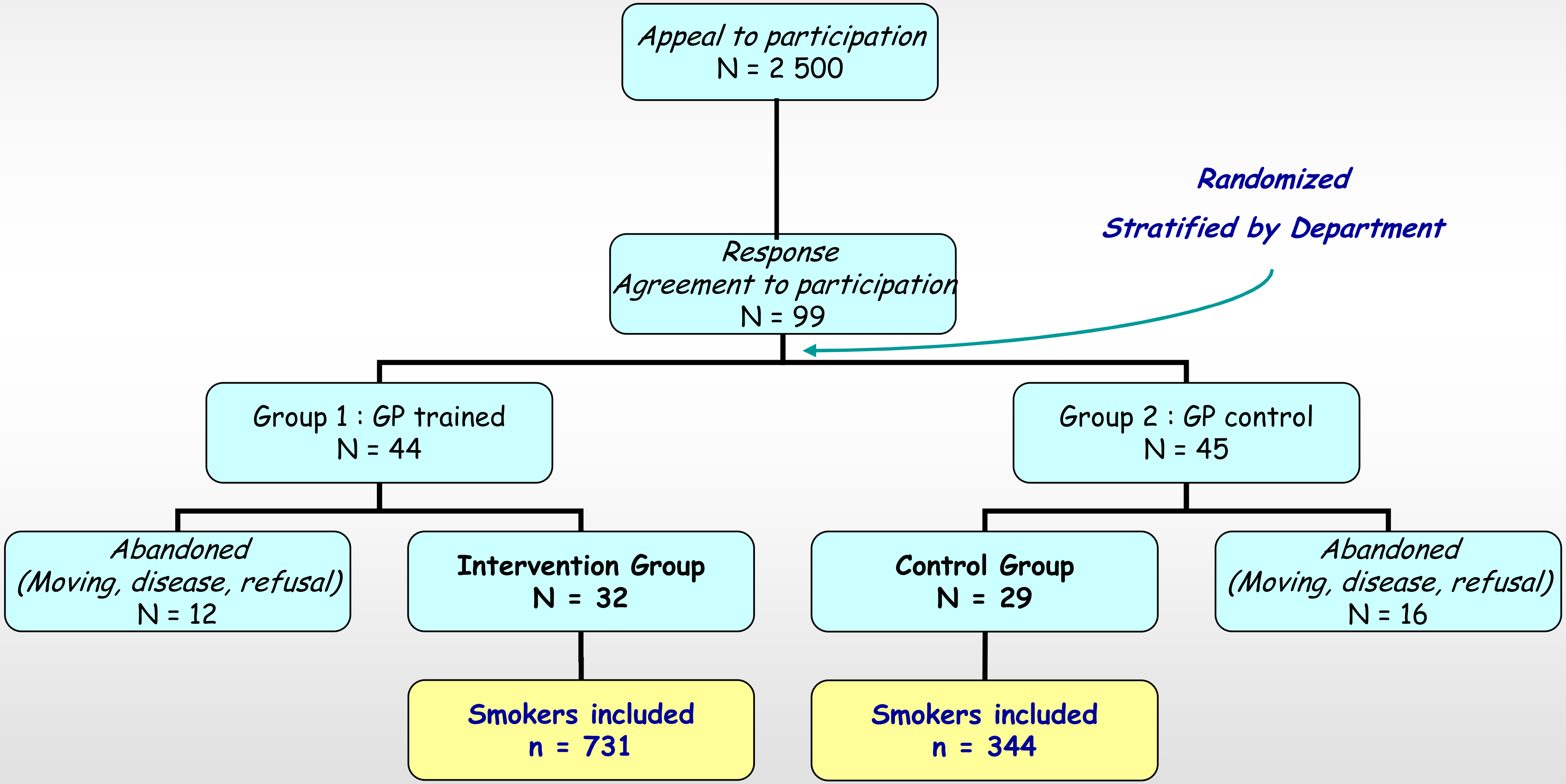
- Helping people to stop smoking is one of the major activities of tobacco control proposed by the Framework Convention on Tobacco Control.
- In France, about one in ten deaths is due to tobacco.

  More than 58% of smokers in France say they would like to stop smoking in the near future but few smokers stop spontaneously
- General practitioners (GP) are well-regarded by their patients and they have the opportunity to play an active role in providing cessation advice (only 37% of GPs in France provide routine or systematic advice to their smoking patients)

Objectives

- To show that a program of public health, based on an adapted training of GP, can modify the practices of assistance in stopping tobacco and increase the cessation rate
- To obtain predictive factors for tobacco cessation at 12 months
- To assess the performance of statistical estimation procedures for the analysis of binary repeated measures data
- To show a significant impact on smoking cessation rates during several visits which illustrate the dynamic of tobacco cessation

Chart of recruitment process



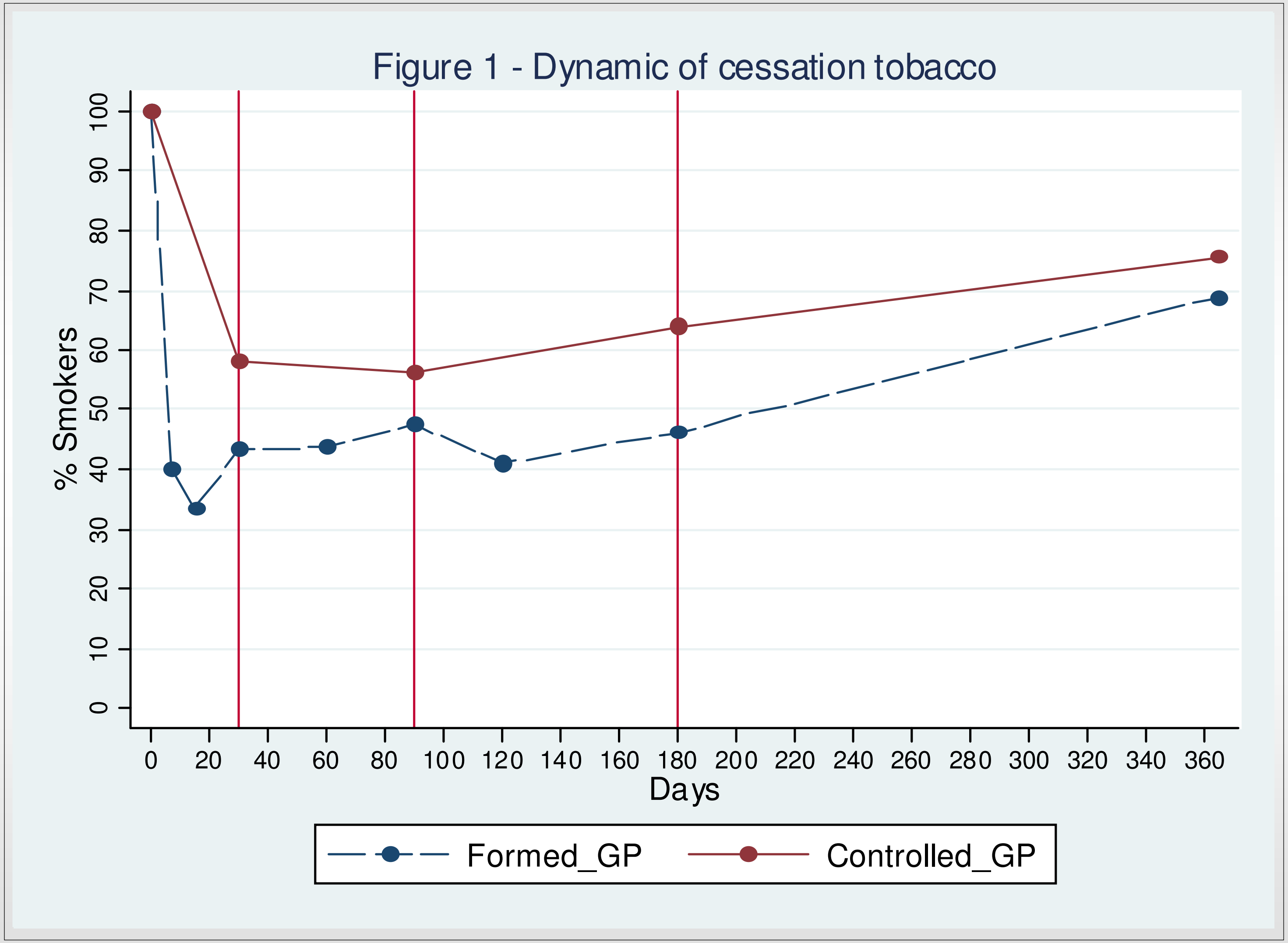


Figure 1 - Dynamic of cessation tobacco
% Smokers
0
10
20
30
40
50
60
70
80
90
100
0
20
40
60
80
100
120
140
160
180
200
220
240
260
280
300
320
340
360
Days
Formed_GP
Controlled_GP

Sample size (1)

- Suppose N clusters of m individuals are randomly assigned to each of arm i=1,2
- We denote the primary response variable for an individual by Y (normally distributed)
- **Sample size estimation** ⇨ Test hypothesis H0: $\mu_1=\mu_2 \Leftrightarrow \Delta=\mu_1-\mu_2=0$
  - Calculate sample size as for individually randomized trials
  - Multiply by variance inflation factor

$$n = mN = \frac{2\sigma^2\left(z_{\alpha/2} - z_\beta\right)^2}{\Delta^2}\left[1+\left(m-1\right)\rho\right]$$

$$IF = 1 + ((1+cv^2)\overline{m} - 1)\rho$$

Sample size (2)

- There are three sample size choices to be made:
  - Number of clusters and number of subjects per cluster
  - Number of repetitions per individual

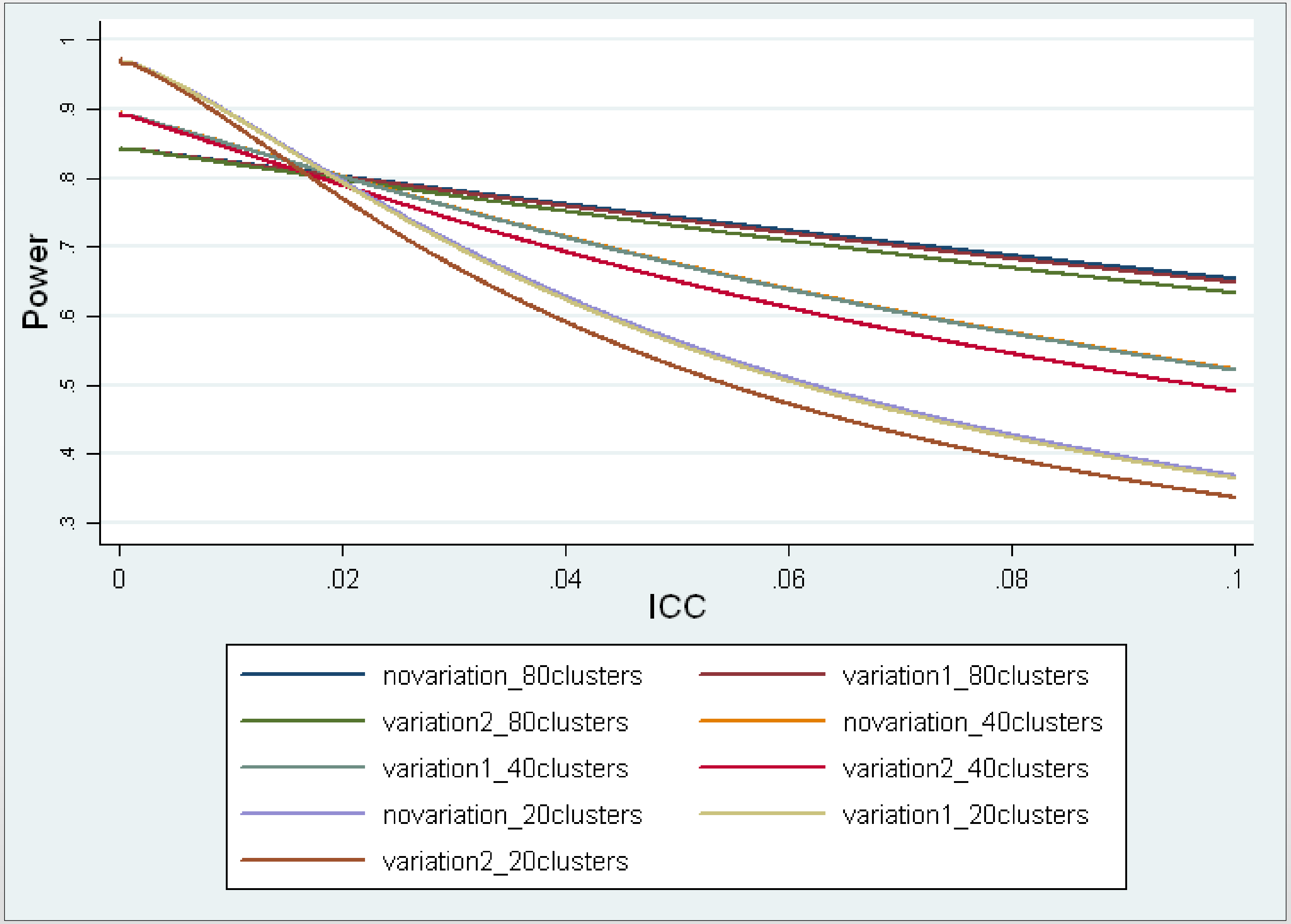
- <u>Several types of correlations</u>
  - Correlation among level 2 $$\rho_\lambda = Corr(Y_{jst}, Y_{js't'}) = \frac{\sigma_\lambda^2}{\sigma_\lambda^2 + \sigma_\psi^2 + \sigma_\varepsilon^2}$$
  - Correlation among level 1 $$\rho_\psi = Corr(Y_{jst}, Y_{jst'}) = \frac{\sigma_\lambda^2 + \sigma_\psi^2}{\sigma_\lambda^2 + \sigma_\psi^2 + \sigma_\varepsilon^2}$$

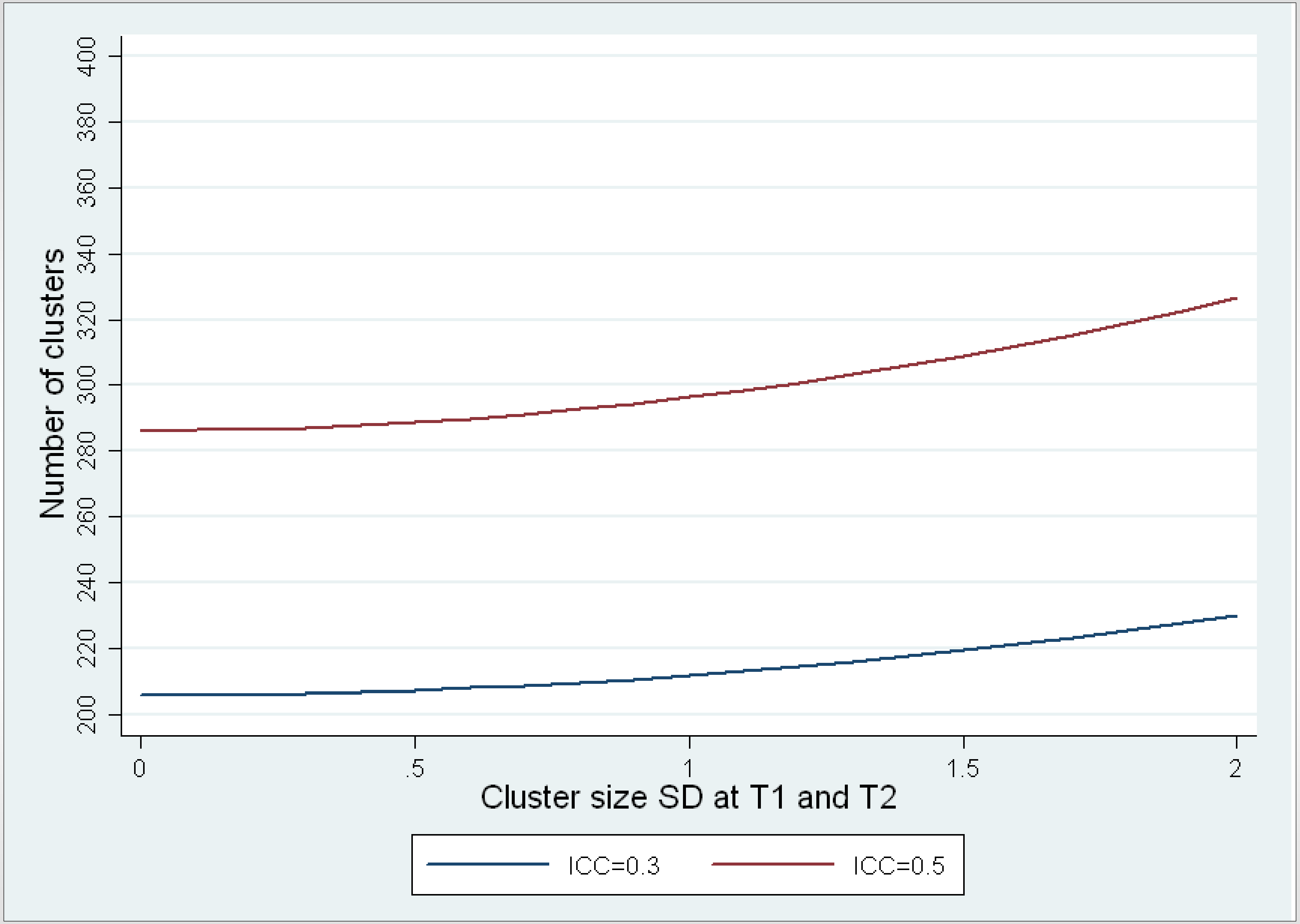
$$\left[1 + (m-1)\rho\right]$$

$$IF_{TE} = 1 + \bar{t}(\bar{m} - 1)\rho_\lambda + (\bar{t} - 1)\rho_\psi \Rightarrow N = \frac{2n}{mt} IF$$

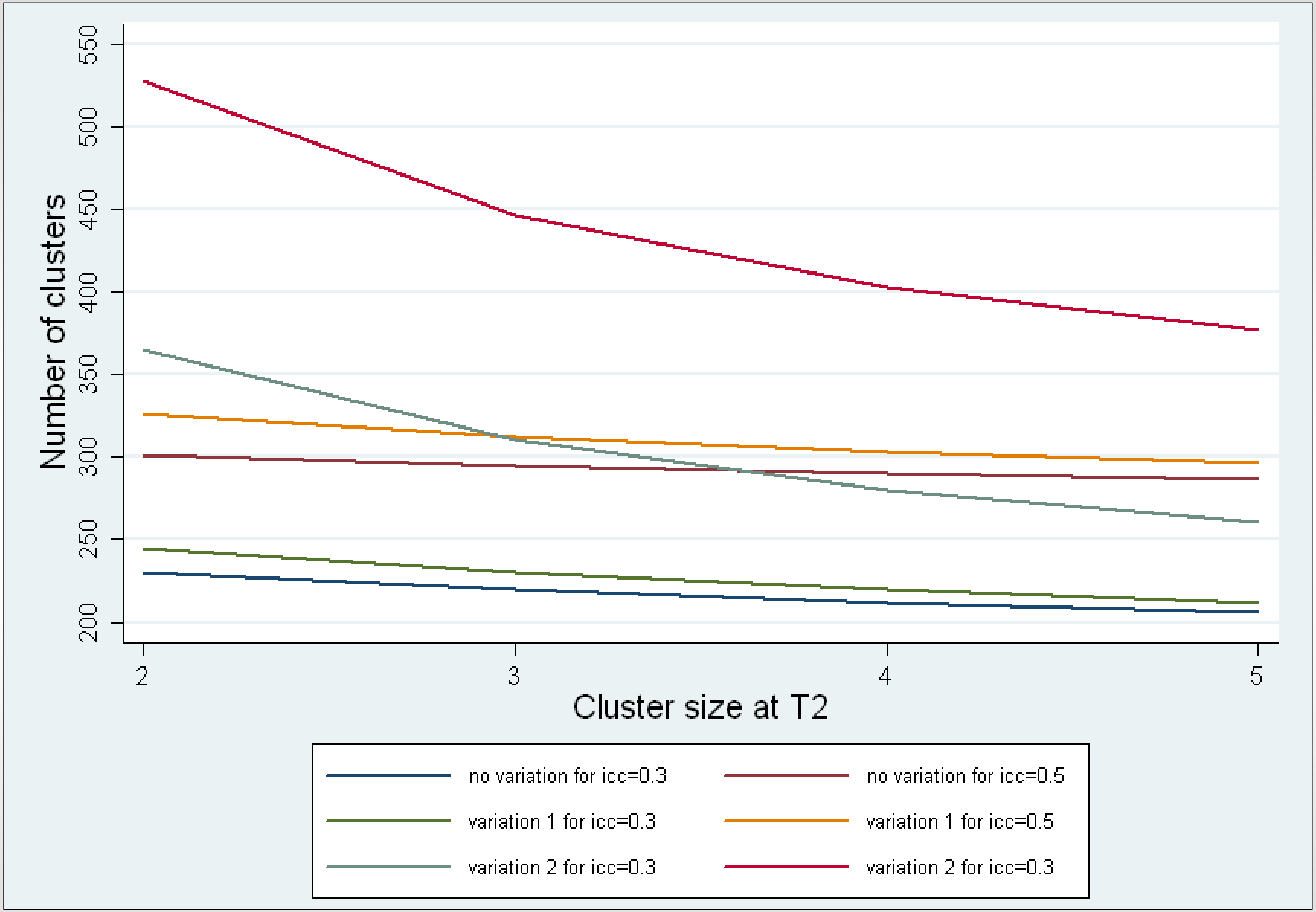
$$IF_{TE} = 1 + \bar{t}(\overline{m} - 1)\rho_{\lambda} + (\bar{t} - 1)\rho_{\psi}$$

$$IF = 1 + \Big((cv_{T1}^2 + 1)m_{T1} - 1 \Big)\rho_{\lambda} + \Big((cv_{T2}^2 + 1)m_{T2} - 1 \Big)\rho_{\lambda} + (t-1)\rho_{\tau}$$





Number of clusters
200
220
240
260
280
300
320
340
360
380
400
0
.5
1
1.5
2
Cluster size SD at T1 and T2
ICC=0.3
ICC=0.5


Number of clusters
200
250
300
350
400
450
500
550
2
3
4
5
Cluster size at T2
no variation for icc=0.3
no variation for icc=0.5
variation 1 for icc=0.3
variation 1 for icc=0.5
variation 2 for icc=0.3
variation 2 for icc=0.3

Statistical analysis – Individual-level

- Marginal model fitted using Generalised Estimating Equations
- Random effects model estimated by different approaches:

$$\log\left[\frac{P_{ijs}}{1-P_{ijs}}\right] = \beta_0 + Z_{ij}\beta_Z + X_{ijs}\beta_1 + U\lambda_{ij} = X\beta + U\lambda$$

  - Numerical approaches (GH and adaptative quadrature, MCMC, MC-EM)
  - Conditional approaches (Schall, Penalized Quasi-Likelihood, HGLM)
  - Marginal approaches (GAR, Marginal Quasi-Likelihood)

⇨ PQL underestimates fixed effects and their SE when average cluster size was small

⇨ MCMC performed better than PQL estimation given many small clusters but not as well as PQL estimation with a smaller number of larger clusters

⇨ Schall is better than GAR for highest σ_λ^2 values

Comparison of 2 and 3-level RE

$$\log\left[\frac{P_{ijst}}{1-P_{ijst}}\right] = X_{ijst}\beta + U\lambda_{ij} + T\psi_{ijs}$$

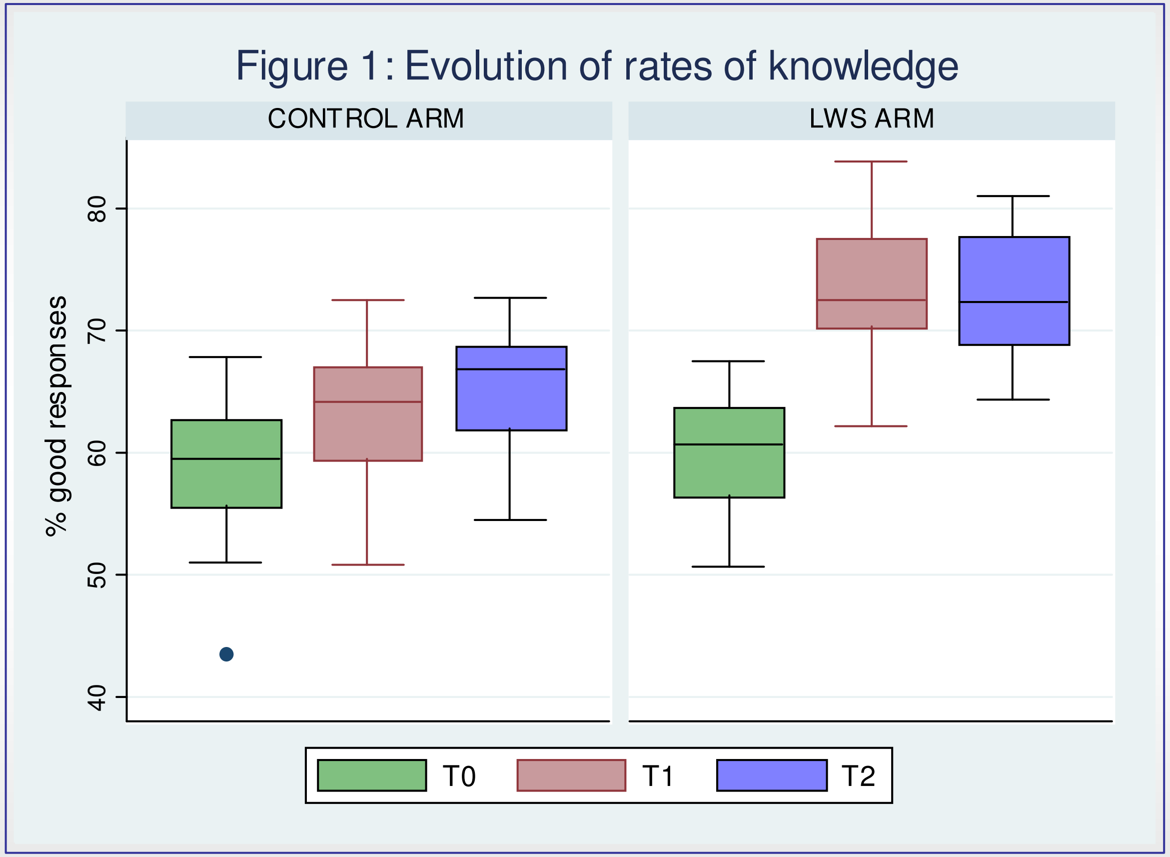
	Logistic model	GLMM with one random effect									
		Random effect λ					Random effect τ				
		GQ	AQ	MCMC	PQL	HGLM	GQ	AQ	MCMC	PQL	HGLM
β	1.376	1.608	1.604	1.604	1.596	1.604	1.493	1.494	1.490	1.489	1.491
(sd)	(0.102)	(0.319)	(0.314)	(0.314)	(0.311)	(0.314)	(0.231)	(0.231)	(0.228)	(0.228)	(0.229)
P-value	0.0001	0.017	0.019	0.017	0.019	0.018	0.010	0.010	0.011	0.014	0.010
σ_λ	--	0.671	0.668	0.671	0.674	0.671	--				
σ_τ	--	--					1.746	1.742	1.749	1.754	1.745
L	- 2 464	- 2 361	- 2 359	- 2 360	- 2 368	- 2 360	- 2 221	- 2 220	- 2 221	- 2 221	- 2 220

Comparison of 2 and 3-level RE

$$\log\left[\frac{P_{ijst}}{1-P_{ijst}}\right] = X_{ijst}\beta + U\lambda_{ij} + T\psi_{ijs}$$

	GLMM with 2 random effects					Marginal model	
						Random effect λ	Random effect τ
	GQ	AQ	MCMC	PQL	HGLM	Estimation by GEE	
β	1.510	1.722	1.583	1.59	1.680	1.499	1.283
(sd)	(0.431)	(0.470)	(0.435)	(0.447)	(0.468)	(0.251)	(0.141)
p-value	0.151	0.046	0.078	0.070	0.047	0.016	0.023
σ_λ	0.865	0.843	0.861	0.861	0.848		
σ_τ	1.552	1.548	1.551	1.552	1.550		
L	- 2 188	- 2 182	- 2 188	- 2 186	- 2 184		

Cluster-level analysis



Statistical analysis confirms that the **rate of knowledge is significantly higher for the children in the LWS classes.**

Individual-level analysis (1)

$$\text{M1} \quad : \quad y_{ijs} = \mu + \beta g_i + \lambda_{ij} + \varepsilon_{ijs}$$

$$\text{M2} \quad : \quad y_{ijs} - x_{ijs} = \mu + \beta g_i + \lambda_{ij} + \varepsilon_{ijs}$$

$$\text{M3} \quad : \quad y_{ijs} = \mu + \beta g_i + \gamma(x_{ijs} - \overline{x}_{...}) + \lambda_{ij} + \varepsilon_{ijs}$$

$$\text{M4} \quad : \quad y_{ijs} = \mu + \beta g_i + \gamma(x_{ijs} - \overline{x}_{ij.}) + \gamma_S(\overline{x}_{ij.} - \overline{x}_{...}) + \lambda_{ij} + \varepsilon_{ijs}$$

$$\text{M5} \quad : \quad y_{ijs} - x_{ijs} = \mu + \beta g_i + \gamma(\overline{x}_{ij.} - \overline{x}_{...}) + \lambda_{ij} + \varepsilon_{ijs}$$

$$\text{M6} \quad : \quad y_{ijs} = \mu + \beta g_i + \gamma x_{ijs} + \lambda_{ij} + \varepsilon_{ijs}$$

Individual-level analysis (2)

	Fixed effect (group)			σ_λ	Log-likelihood
	β	IC(95%)	p-value		
Model M1	3.43 (0.66)	2.14 , 4.73	0.0001	1.81 (0.28)	- 2 056
Model M2	3.17 (0.63)	1.94 , 4.41	0.0001	1.60 (0.28)	- 2 146
Model M3	3.33 (0.55)	2.24 , 4.41	0.0001	1.46 (0.24)	- 2 001
Model M4	3.28 (0.54)	2.22 , 4.34	0.0001	1.41 (0.23)	- 2 000
Model M5	3.28 (0.54)	2.22 , 4.34	0.0001	1.21 (0.27)	- 2 139
Model M6	3.33 (0.52)	2.24 , 4.41	0.0001	1.46 (0.24)	- 2 001

Individual-level analysis (3)

		Linear model	LMM		LMM
			Class effect	Child effect	Two random effects
Effect group	LWS Class	2.79 (0.22)	2.92 (0.64)	2.79 (0.29)	2.90 (0.63)
Time effect	Evaluation at T1	4.36 (0.26)	4.36 (0.24)	4.36 (0.20)	4.35 (0.20)
	Evaluation at T2	4.69 (0.26)	4.69 (0.24)	4.69 (0.20)	4.69 (0.20)
σ_λ		--	1.94 (0.24)	--	1.85 (0.25)
σ_τ			--	3.16 (0.13)	2.62 (0.13)
Log-likelihood		- 6 356	- 6 237	- 6 186	- 6 138

Conclusion

Objective: To consider a Cluster-RT in which each subject is followed during several visits ⇨ The repeated observations (level 3) are nested within subjects (level 2) who are themselves nested within the unit of randomization (level 1)

⇨ **Design and analysis of cluster randomized trials are two important challenges for epidemiologists and statisticians… (Murray 2008)**

⇨ **And it's even more difficult for "longitudinal Cluster-RT"!**

⇨ **Missing data…**