Contemporary Clinical Trials Communications 14 (2019) 100316

Contents lists available at ScienceDirect

Contemporary Clinical Trials Communications

journal homepage: www.elsevier.com/locate/conctc

Sample size estimation and re-estimation of cluster randomized controlled trials for real-time feedback, debriefing, and retraining system of cardiopulmonary resuscitation for out-of-hospital cardiac arrests

Akiko Kada[a,*,1], Akihiro Hirakawa[a,b,1], Fumie Kinoshita[c], Yumiko Kobayashi[c], Toshihiro Hatakeyama[d,e], Daisuke Kobayashi[e], Chika Nishiyama[f], Taku Iwami[e]

[a] Department of Clinical Research Management, Clinical Research Center, National Hospital Organization Nagoya Medical Center, Nagoya, 460-0001, Japan
[b] Department of Biostatistics and Bioinformatics, Graduate School of Medicine, The University of Tokyo, Tokyo, 113-8654, Japan
[c] Department of Advanced Medicine Nagoya University Hospital, Nagoya, 466-8560, Japan
[d] Department of Emergency and Critical Care Medicine, Emergency and Critical Care Center, Dokkyo Medical University Saitama Medical Center, Koshigaya, 343-8555, Japan
[e] Kyoto University Health Service, Kyoto, 606-8501, Japan
[f] Department of Critical Care Nursing, Kyoto University Graduate School of Human Health Science, Kyoto, 606-8507, Japan

ARTICLE INFO

Keywords:
Cluster randomization
Emergency medical services
Intracluster correlation coefficient
Sample size re-estimation

ABSTRACT

Background: In cluster randomized controlled trials (RCTs) for emergency medical services (EMS) system, we encounter the situation that the actual cluster size and ratio of allocated patients between two groups eventually differ from those used for sample size estimation because of the nature of patient enrollment. In such trials, estimations of effect size of test intervention and intra-cluster correlation coefficient (ICC) used for sample size estimation are also difficult. To improve efficient management on clinical cluster RCTs, we need to understand the effect of such inconsistencies of the design parameters on the type I error rate and statistical power of testing.
Methods: We planned the trial which evaluated the 1-month favorable neurological survival of out-of-hospital cardiac arrest patients with or without real-time feedback, debriefing, and retraining system by EMS personnel. Under the conditions that we possibly encountered in this trial, we examined the effect of inconsistencies in the actual ICC, cluster size, and ratio of patient allocation with those expected for sample size estimation on the type I error rate and power, using simulation studies. We further investigated the contribution of incorporating sample size re-estimation, based on the results of interim analysis of the trial, on the power increase.
Results: This simulation study showed that the inconsistencies of cluster size and patient allocation ratio decreased the power by 5–10% in some cases. In addition, the power decreased by 3–4% when the actual ICC was larger than that expected for sample size estimation. Furthermore, the use of a generalized estimating equation method to evaluate the difference in the 1-month favorable neurological survival between two groups caused inflation of type I error rate. Finally, the increase in power by incorporating sample size re-estimation was limited.
Conclusions: We identified remarkable effects of sample size estimation and re-estimations in a cluster RCT for real-time feedback, debriefing, and retraining system of cardiopulmonary resuscitation for out-of-hospital cardiac arrests. The estimation of design parameters for sample size estimation is generally challenging in cluster RCTs for EMS system; therefore, it is important to conduct a trial simulation that assesses the statistical performances under sample sizes based on the various expected values of the design parameters before beginning the trial.

1. Introduction

A cluster randomized controlled trial (RCT) is performed as an intervention in a group of subjects (collectively labeled as a 'cluster'), rather than for an individual. This design is used in emergency medicine when researchers would like to conduct clinical trials focusing on

* Corresponding author. Department of Clinical Research Management, Clinical Research Center, National Hospital Organization Nagoya Medical Center, 4-1-1 Sannomaru, Naka-ku, Nagoya, 460-0001, Japan.
E-mail address: akiko.kada@nnh.go.jp (A. Kada).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.conctc.2019.100316
Received 26 September 2018; Received in revised form 28 December 2018; Accepted 3 January 2019
Available online 04 January 2019

"cluster" such as fire department and emergency medical services (EMS) station [1–6]. Sample size estimation in cluster RCTs requires the expected effect size of the test intervention, the intra-cluster correlation coefficient (ICC), the cluster size (i.e., the number of treated patients per EMS station), and the ratio of patient allocation between two groups, in addition to the type I error rate and the statistical power of testing [7–9]. However, we often encounter the situation that the actual values of the cluster size and the ratio of allocated patients between two groups eventually differ from those used for sample size estimation because of the nature of patient enrollment [6]. Furthermore, in cluster RCTs for EMS system, the difficulty lies in estimating the effect size of the test intervention and ICC used for sample size estimation, because few estimates of the effect size and ICC from existing research have been published, leaving investigators with little data on which to base their expectations in some cases [10]. To improve efficient management on cluster RCTs, we need to understand the effect of the uncertainty of these design parameters on the type I error rate and power.

Recently, we started an open-label cluster RCT that enrolled 3000 to 5000 patients with out-of-hospital cardiac arrest (OHCA) and included 120 EMS stations from February 1, 2017 in Japan (UMIN000021431) to examine the clinical benefit of real-time feedback, debriefing, and re-training system of cardiopulmonary resuscitation (CPR) for OHCAs [11]. The intervention group (high performance CPR group) in this trial refers to CPR that makes use of a real-time feedback device, which quantifies and monitors the quality of CPR (real-time feedback), combined with post-event debriefing and physical retraining. The control group (standard CPR group) is the standard CPR without real-time feedback, debriefing, and retraining using the use of a real-time feedback device. The 120 EMS stations were randomly allocated to high performance CPR group or standard CPR group on a 1:1 ratio. Therefore, the EMS personnel belonging to the same EMS station treat cardiac arrest patients using the allocated procedure. The primary endpoint was the proportion of patients who survive with 1-month favorable neurological survival, defined as cerebral performance category score 1 or 2, for one month after the event.

Before beginning the trial, under the conditions that we might encounter, we comprehensively examined the effect of inconsistencies in the actual ICC, cluster size, and ratio of patient allocation with those expected for sample size estimation on the type I error rate and power using simulation studies. With respect to the uncertainty of the effect size of the test intervention, we investigated the contribution of incorporating sample size re-estimation based on the observed difference in the 1-month favorable neurological survival between the two groups at an interim analysis of the trial for the power increase.

In this study, we share and discuss the remarkable effects of sample size estimation and re-estimations based on our cluster RCT for EMS system, with and without sample size re-estimation. In the following section, we briefly summarize the sample size estimation for cluster RCT, the definition of ICC, and the sample size re-estimation we used in our trial, respectively. Subsequently, we conducted comprehensive simulation study and discussed our findings.

2. Materials and methods

2.1. Sample size calculation for a cluster RCT

In the standard cluster RCT, the sample size is often calculated by,

$$N = Jm = \frac{(Z_{\alpha/2} + Z_{\beta})^2[p_t(1-p_t) + p_c(1-p_c)][1 + (m-1)\rho]}{(p_t - p_C)^2}$$

where p_t and p_c are the expected proportion of 1-month favorable neurological survival in high performance CPR group and standard CPR group, respectively, J is the number of EMS stations, m is the cluster size for each station, and ρ is the ICC [9]. $Z_{\alpha/2}$ and Z_{β} are the upper $(\alpha/2)$th and βth percentiles of the standard normal distribution, respectively.

In our trial, the p_t was set to 0.061914 based on data from Osaka prefecture of past 5 years and p_c was set to 0.038490 using odds ratio = 1.65, which was expected by previous research of real-time feedback or debriefing [11]. To calculate the sample size required to achieve an 80% power ($\beta =$ 20%) at 5% level of significance ($\alpha =$ 5%) under this assumption, we further needed to specify the values of J and ρ, respectively. The specification of ICC (ρ) is described in the following section.

2.2. Intra-cluster correlation coefficient

In this section, we introduce the ICC definition used in our trial. We supposed that the clustered, binary outcome variable (e.g., 1-month survival) is denoted by Y_{ij} (= 0 or 1) for the jth patient (j = 1, 2, …, m_i) in the ith cluster (i = 1, 2, …, J), where m_i represents the cluster size. The intervention indicator is denoted by a dichotomous variable, X_{ij} (= 0 or 1) whose value depends only on the cluster; therefore, $X_{ij} = X_i$.

The mixed effect and marginal probability models based on the generalized estimating equation (GEE) developed by Liang and Zeger [12] take into account the correlation of patient survival outcome in the inferential processes. We considered the mixed effect model:

$$\text{logit}(p_{ij}) = \alpha_1 + \beta_1 X_i + U_i$$

where $p_{ij} = E(Y_{ij}|X_i, U_i)$, α_1 is a constant representing the baseline log-odds, and β_1 is population-average log-odds. The random variable, U_i is a random effect and is distributed with a mean of 0 and unknown variance σ^2, $U_i \sim$ Normal (0, σ^2). In the mixed effect model, the ICC (ρ_{lo}) is defined as $\sigma^2/(\sigma^2 + \pi^2/3)$.

In our trial, the ICC estimate was $\rho_{lo} = 0.003$, which corresponded to the value of $\sigma^2 = 0.01$. σ^2 was based on assuming random effect of about ± 0.3 relative to β_1. Thus, the required cluster size (or total sample size) when $\rho_{lo} = 0.003$ was 25 (or approximately 3000) when the number of EMS stations were ($J =$) 120.

In addition, we also considered the marginal probability model based on the GEE:

$$\text{logit}(p_{ij}) = \alpha_2 + \beta_2 X_i$$

The parameter of 'exchangeable' compound symmetry correlation structure could be used as the ICC. Furthermore, it should be noted that in the case where the probability of success in a binary endpoint is low, the ICC tends to be low [13]. Eldridge et al. [8] have provided several definitions of ICC in cluster RCTs.

2.3. Sample size re-estimation

As we described earlier, the difficulty lies in the determining the expected difference in the 1-month favorable neurological survival between two groups in the sample size estimation. Although we assumed an expected odds ratio of 1.65, a lower odds ratio (e.g., 1.60 corresponding to proportion of 1-month favorable neurological survival of 3.8% and 5.95% in standard CPR and high performance CPR groups, respectively) would also be clinically significant. Therefore, we planned to conduct an interim analysis to re-estimate the required sample size during the trial. Among the several useful methods for sample size re-estimation, we used the approach proposed by Mehta and Pocock [14] because it does not need to adjust the overall type I error rate of the trials, although careful inspection of the operating characteristics of the Mehta and Pocock (MP) method is required [15,16]. Although we do not explain the MP method in more detail in this paper, the sample size increases when interim analysis results fall within a 'promising zone' (e.g., 50–80% of conditional power), where it is deemed worthwhile to increase the conditional power to detect the difference in the primary endpoint between the two groups by adding more patients.

2.4. Simulation studies

2.4.1. Simulation settings

The above-mentioned sample size estimation of 3000 patients (i.e., 25 of cluster size for each EMS station) assuming that $p_t = 0.061914$, $p_c = 0.038490$, $\rho_{lo} = 0.003$, $\alpha = 0.05$, and $\beta = 0.20$ supposes an equal cluster size between the EMS stations, equal patient allocation between the groups, and a common expected ICC between the groups. These design parameters possibly differ from those observed in the actual trial; therefore, we investigated the type I error rate and statistical power of testing using the mixed effect model and GEE method under the following possible settings that we might encounter with respect to the cluster size, the ratio of patient allocation, and the ICC, with and without incorporating the sample size re-estimation through the simulation studies.

In the simulation studies, the 1-month survival probability was generated based on the mixed effect model such that $p_t = \exp(-3.231 + 0.497 + u_i)/\{1 + \exp(-3.231 + 0.497 + u_i)\}$ (or $p_c = \exp(-3.231 + u_i)/\{1 + \exp(-3.231 + u_i)\}$), where $u_i \sim \text{Normal}(0, \sigma^2)$. The value of σ^2 was set to 0.01 ($\rho_{lo} = 0.003$) or 0.11 ($\rho_{lo} = 0.032$ as conservative setting) for the two groups. Five possible situations for cluster size were considered (Table 1). Situation 1 is ideal wherein the cluster size of all the EMS stations are the same; however, it is apparently impractical. In situations 2–5, the cluster size is varied depending on the EMS stations. Additionally, under situation 1, the power was also evaluated when the ratios of sample sizes for the high performance CPR group and standard CPR group were (1200:1800), (1000:2000), (1750:1275), (1800:1200), (2000:1000), and (1275:1750), respectively.

Using the above-mentioned probabilities and situations, the number of 1-month survivors in each group was generated based on the Binominal distribution (e.g., Binomial (1500, p_t) and Binominal (1500, p_c)) and subsequently, the mixed effect model and the GEE method introduced in the previous section were applied to estimate the odds ratio and its p-value. We performed 10,000 repetitions (i.e., 10,000 simulated trials) for each setting and reported the proportion of simulated trials in which the p-value was less than or equal to 0.05 as the statistical power of testing (or the type I error rate under the null hypothesis) and average of estimated odds ratio. In the case of incorporating sample size re-estimation, we also calculated the proportion of observing the 'promising zone' result among 10,000 simulated trials, as well as the average of the final sample size after sample size re-estimation. We used SAS version 9.4 software (SAS Institute Inc., Cary, NC, USA) to implement the simulation studies.

3. Results

3.1. Mixed effect model and GEE

We found that the parameter estimation of the random effect in the mixed effect model did not converge because of the extremely low rates of patient survival in each cluster. The successful parameter estimation of the mixed effect model was less than 50% of 10,000 simulated trials, while it was 100% for the GEE method; therefore, we present the results using the GEE in the following section. We set the main analysis as GEE, and the secondary analysis as mixed effect model.

3.2. Power

Table 2 shows the results of simulation studies with situations of varying ICC and cluster sizes. In the case where the value of ρ_{lo} was identical to the expected value of 0.003 for both groups, and the cluster size was commonly 25 among the 120 EMS stations (i.e., Situation 1), the simulated power of 83% was almost same as the pre-specified target value of 80%. The average odds ratio of 1.68 was also close to the expected value of 1.65. The simulated power decreased by 3–4% when the ICC of the high performance and/or standard CPR groups increased to 0.032; however, those were between 76 and 85% in all the situations of cluster size we assumed in the simulation studies.

Table 3 shows the results of the simulation studies under conditions of varying the ICC and ratios of patient allocation between the two groups under situation 1 shown in Table 1. When the ICC was different in the two groups, the simulated power decreased by up to 10% from the target value of 80%. Overall, the difference between the target and simulated powers was approximately ± 5%.

Table 1
Five cluster-size situations for each group.

Cluster size	Number of EMS stations	Number of patients	Cluster size	Number of EMS stations	Number of patients	Cluster size	Number of EMS stations	Number of patients
Situation 1			Situation 2			Situation 3		
5	0	0	5	6	30	5	2	10
10	0	0	10	6	60	10	4	40
15	0	0	15	7	105	15	8	120
20	0	0	20	7	140	20	10	200
25	60	1500	25	8	200	25	12	300
30	0	0	30	7	210	30	10	300
35	0	0	35	7	245	35	8	280
40	0	0	40	6	240	40	4	160
45	0	0	45	6	270	45	2	90
Situation 4						Situation 5		
5	10	50	5	0	0			
10	8	80	10	0	0			
15	6	90	15	12	180			
20	4	80	20	10	200			
25	4	100	25	14	350			
30	4	120	30	14	420			
35	6	210	35	10	350			
40	8	320	40	0	0			
45	10	450	45	0	0			

EMS: emergency medical services.

Table 2
Results of simulation studies with situations of varying ICC and cluster sizes.

ICC of standard CPR group	ICC of high performance CPR group	Situation in standard CPR group	Situation in high performance CPR group	Power, %	Odds ratio
0.003	0.003	1	1	83	1.68
0.032	0.032	1	1	76	1.60
0.032	0.003	1	1	77	1.61
0.003	0.003	2	2	84	1.68
0.032	0.032	2	2	84	1.67
0.032	0.003	2	2	77	1.60
0.003	0.003	2	3	84	1.68
0.003	0.003	2	4	84	1.68
0.003	0.003	2	5	84	1.68
0.032	0.032	2	3	84	1.67
0.032	0.032	2	4	84	1.67
0.032	0.032	2	5	85	1.67
0.032	0.003	2	3	77	1.60
0.032	0.003	2	4	77	1.60
0.032	0.003	2	5	77	1.60
0.003	0.003	3	3	84	1.68
0.032	0.032	3	3	84	1.67
0.032	0.003	3	3	77	1.60
0.032	0.003	3	3	77	1.60
0.003	0.003	3	4	84	1.68
0.003	0.003	3	5	84	1.68
0.032	0.032	3	4	84	1.67
0.032	0.032	3	5	85	1.67
0.032	0.003	3	4	77	1.60
0.032	0.003	3	5	77	1.60
0.003	0.003	4	4	84	1.68
0.032	0.032	4	4	84	1.67
0.032	0.003	4	4	77	1.60
0.003	0.003	4	5	84	1.68
0.032	0.032	4	5	85	1.67
0.032	0.003	4	5	77	1.60
0.003	0.003	5	5	84	1.68
0.032	0.032	5	5	85	1.67
0.032	0.003	5	5	77	1.60

ICC: intra-cluster correlation coefficient, CPR: cardiopulmonary resuscitation.

Table 3
Results of simulation studies under conditions of varying ICC and ratio of sample size between the two groups.

ICC of standard CPR group	ICC of high performance CPR group	Sample size in standard CPR group	Sample size in high performance CPR group	Power, %	Odds ratio
0.003	0.003	1800	1200	82	1.68
0.032	0.032	1800	1200	83	1.67
0.032	0.003	1800	1200	75	1.60
0.003	0.003	1200	1800	81	1.68
0.032	0.032	1200	1800	83	1.68
0.032	0.003	1200	1800	74	1.61
0.003	0.003	2000	1000	80	1.68
0.032	0.032	2000	1000	81	1.67
0.032	0.003	2000	1000	73	1.60
0.003	0.003	1000	2000	78	1.70
0.032	0.032	1000	2000	79	1.69
0.032	0.003	1000	2000	70	1.62
0.003	0.003	1750	1275	83	1.68
0.032	0.032	1750	1275	83	1.67
0.032	0.003	1750	1275	76	1.60
0.003	0.003	1275	1750	82	1.68
0.032	0.032	1275	1750	84	1.68
0.032	0.003	1275	1750	75	1.61

ICC: intra-cluster correlation coefficient, CPR: cardiopulmonary resuscitation.

3.3. Type I error rate and power of design with sample size re-estimation

In this section, we describe the evaluation of the type I error rate under the null hypothesis (i.e., odds ratio = 1.0) and the power when the odds ratio was 1.60 (i.e., the true odds ratio is lower than the expected odds ratio of 1.65) when using the GEE method in the 10,000 simulated trials, with and without sample size re-estimation. The method of data generation was the same as that described in the previous section. The interim analysis was conducted when the sample size reached 1800 (60%) or 2400 (80%). For simplicity's sake, the cluster size was set to 25 for 60 EMS stations with equal patient allocation. The pre-specified, permitted maximum sample size (n_{max}) that is often determined by budgetary limitations and feasibility of recruitment was 3500 or 4000.

Table 4
Type I error rates for the GEE method with and without sample size re-estimation.

ICC of standard CPR group	ICC of high performance CPR group	n_1	n_{max}	Type I error rate without SSR, %	Type I error rate with SSR, %
0.003	0.003	1800	3500	5.41	5.25
0.032	0.032	1800	3500	5.09	5.33
0.032	0.003	1800	3500	6.04	5.98
0.003	0.003	1800	4000	5.41	5.48
0.032	0.032	1800	4000	5.09	5.57
0.032	0.003	1800	4000	6.04	6.38
0.003	0.003	2400	3500	5.41	5.19
0.032	0.032	2400	3500	5.09	5.26
0.032	0.003	2400	3500	6.04	5.90
0.003	0.003	2400	4000	5.41	5.54
0.032	0.032	2400	4000	5.09	5.67
0.032	0.003	2400	4000	6.04	6.26

GEE: generalized estimating equation, ICC: intra-cluster correlation coefficient, CPR: cardiopulmonary resuscitation, n_1: sample size at interim analysis, n_{max}: permitted maximum sample size, SSR: sample size re-estimation.

Table 5
Powers and average sample sizes for the GEE method with and without sample size re-estimation.

ICC of standard CPR group	ICC of high performance CPR group	n_1	n_{max}	Power without SSR, %	Power with SSR, %	Favorable, %	Promising, %	Unfavorable, %	Average of final sample size after SSR
0.003	0.003	1800	3500	80	79	58	16	27	3066
0.032	0.032	1800	3500	73	72	51	16	33	3066
0.032	0.003	1800	3500	72	72	51	15	33	3065
0.032	0.032	1800	4000	80	80	58	15	27	3101
0.032	0.003	1800	4000	73	73	50	17	33	3111
0.003	0.003	2400	3500	72	73	51	16	33	3108
0.032	0.003	2400	3500	80	79	64	11	25	3042
0.003	0.003	2400	4000	73	72	56	12	31	3047
0.032	0.032	2400	4000	72	72	56	12	31	3048
0.032	0.003	2400	4000	80	80	65	11	24	3068

GEE: generalized estimating equation, ICC: intra-cluster correlation coefficient, CPR: cardiopulmonary resuscitation, n_1: sample size at interim analysis, n_{max}: permitted maximum sample size, SSR: sample size re-estimation.

Tables 4 and 5 show the type I error rates and powers of trials performed with and without sample size re-estimation, along with the average sample size of trials with sample size re-estimation. As shown in Table 4, the type I error rate exceeded the nominal value of 5% by approximately 1% when the ICC between the two groups was different. Furthermore, the inflation was not corrected even if the trial included sample size re-estimation. We also found that the power did not sufficiently increase when using the sample size re-estimation based on the MP method (Table 5). This was likely because the probability that the interim results falls in a 'promising zone' was only approximately 15%. Furthermore, even if the sample size re-estimation was performed, increasing the sample size by 500 (or 1000) patients did not result in a satisfactory increase in the power.

4. Discussion

The power decreased by 5–10% in some cases when the cluster size and patient allocation ratio observed in the actual trial were different from those expected in the sample size estimation. The power also decreased by 3–4% when the actual ICC was larger than that expected for sample size estimation. In addition, the use of the GEE method to evaluate the difference in the 1-month favorable neurological survival between two groups yielded inflation of the type I error rate by up to 1.0%. This tendency was also observed in other research studies [13]. According to the results of the simulation studies conducted by Heo and Leon [13], the GEE method resulted in the highest type I error rates when the ICC = 0. Finally, the increase in power produced by incorporating sample size re-estimation was limited.

5. Conclusions

We determined the remarkable effects of sample size estimation and re-estimations in a cluster RCT for real-time feedback, debriefing, and retraining system of cardiopulmonary resuscitation for out-of-hospital cardiac arrests. We used of the GEE method for cluster RCT, and the inflation in type I error rate was adjusted. The misspecifications of the design parameters for sample size estimation possibly decreased the statistical power of testing by 5% on average in our simulation studies; however, the effect may vary depending on the situations that the investigators encounter. The precise estimation of design parameters for sample size estimation is generally challenging in cluster RCT for EMS system; therefore, it is important to conduct a trial simulation that assesses the statistical performance under a sample size based on the various expected values of the design parameters before beginning the trial.

References

[1] M. Ferri, A. De Luca, P. Giorgi Rossi, G. Lori, G. Guasticchi, Does a pre-hospital emergency pathway improve early diagnosis and referral in suspected stroke patients?–Study protocol of a cluster randomised trial [ISRCTN41456865], BMC Health Serv. Res. 5 (2005) 66 https://doi.org/10.1186/1472-6963-5-66.
[2] D. Hostler, S. Everson-Stewart, T.D. Rea, I.G. Stiell, C.W. Callaway, P.J. Kudenchuk, G.K. Sears, S.S. Emerson, G. Nichol, Resuscitation Outcomes Consortium Investigators: Effect of real-time feedback during cardiopulmonary resuscitation outside hospital: prospective, cluster-randomised trial, BMJ 342 (2011) d512 https://dx.doi.org/10.1136/bmj.d512.
[3] G. Nichol, B. Leroux, H. Wang, C.W. Callaway, G. Sopko, M. Weisfeldt, I. Stiell, L.J. Morrison, T.P. Aufderheide, S. Cheskes, J. Christenson, P. Kudenchuk, C. Vaillancourt, T.D. Rea, A.H. Idris, R. Colella, M. Isaacs, R. Straight, S. Stephens,

J. Richardson, J. Condle, R.H. Schmicker, D. Egan, S. May, J.P. Ornato, ROC investigators: trial of continuous or interrupted chest compressions during CPR, N. Engl. Med. 373 (23) (2015) 2203–2204 https://doi.org/10.1056/NEJMoa1509139.

[4] J. Taylor, S. Black, S.J. Brett, K. Kirby, J.P. Nolan, B.C. Reeves, M. Robinson, C.A. Rogers, L.J. Scott, A. South, E.A. Stokes, M. Thomas, S. Voss, S. Wordsworth, J.R. Benger, Design and implementation of the AIRWAYS-2 trial: A multi-centre cluster randomised controlled trial of the clinical and cost effectiveness of the i-gel supraglottic airway device versus tracheal intubation in the initial airway management of out of hospital cardiac arrest, Resuscitation 109 (2016) 25–32 https://doi.org/10.1016/j.resuscitation.2016.09.016.

[5] C. Axelsson, J. Nestin, L. Svensson, A.B. Axelsson, J. Herlitz, Clinical consequences of the introduction of mechanical chest compression in the EMS system for treatment of out-of-hospital cardiac arrest-a pilot study, Resuscitation 71 (1) (2006) 47–55 https://doi.org/10.1016/j.resuscitation.2006.02.011.

[6] G.D. Perkins, R. Lall, T. Quinn, C.D. Deakin, M.W. Cooke, J. Horton, S.E. Lamb, A.M. Slowther, M. Woollard, A. Carson, M. Smyth, R. Whitfield, A. Williams, H. Pocock, J.J. Black, J. Wright, K. Han, S. Gates, PARAMEDIC trial collaborators: Mechanical versus manual chest compression for out-of-hospital cardiac arrest (PARAMEDIC): a pragmatic, cluster randomised controlled trial, Lancet 385 (9972) (2015) 947–955 https://doi.org/10.1016/S0140-6736(14)61886-9.

[7] C. Roberts, S.A. Roberts, Design and analysis of clinical trials with clustering effects due to treatment, Clin. Trials 2 (2) (2005) 152–162 https://doi.org/10.1191/1740774505cn076oa.

[8] S.M. Eldridge, O.C. Ukoumunne, J.B. Carlin, The Intra-Cluster Correlation Coefficient in Cluster Randomized Trials: A Review of Definitions, Int. Stat. Rev. 77 (3) (2009) 378–394 https://doi.org/10.1111/j.1751-5823.2009.00092.x.

[9] M.J. Campbell, S.J. Walters, How to Design, Analyse and Report Cluster Randomized Trials in Medicine and Health Related Research, John Wiley & Sons, Chichester, 2014.

[10] J. Zhang, S.L. Pals, A. Medley, C. Nichols, P. Bachanas, D. van Zyl, F. Katuta, J. Juma, Parameters for Sample Size Estimation from a Group-Randomized HIV Prevention Trial in HIV Clinics in Sub-Saharan Africa, AIDS Behav. 18 (12) (2014) 2359–2365 https://doi.org/10.1007/s10461-013-0631-9.

[11] A. Hirakawa, T. Hatakeyama, D. Kobayashi, C. Nishiyama, A. Kada, T. Kiguchi, T. Kawamura, T. Iwami, Real-time feedback, debriefing, and retraining system of cardiopulmonary resuscitation for out-of-hospital cardiac arrests: a study protocol for a cluster parallel-group randomized controlled trial, Trials 19 (510) (2018), https://doi.org/10.1186/s13063-018-2852-8.

[12] K.Y. Liang, S.L. Zeger, Longitudinal data analysis using generalized linear models, Biometrika 73 (1) (1986) 13–22 https://doi.org/10.1093/biomet/73.1.13.

[13] M. Heo, A.C. Leon, Comparison of statistical methods for analysis of clustered binary observations, Stat. Med. 24 (6) (2005) 911–923 https://doi.org/10.1002/sim.1958.

[14] C.R. Mehta, S.J. Pocock, Adaptive increase in sample size when interim results are promising: a practical guide with examples, Stat. Med. 30 (28) (2011) 3267–3284 https://doi.org/10.1002/sim.4102.

[15] E. Glimm, Comments on ‘Adaptive increase in sample size when interim results are promising: A practical guide with examples’ by C. R Mehta and S. J. Pocock, Stat. Med. 31 (1) (2012) 98–99 https://doi.org/10.1002/sim.4424.

[16] S.S. Emerson, G.P. Levin, S.C. Emerson, Comments on 'Adaptive increase in sample size when interim results are promising: a practical guide with examples, Stat. Med. 30 (28) (2011) 3285–3301 https://doi.org/10.1002/sim.4271.