Intraclass correlation coefficient in multicenter childhood trauma studies

Bahman Roudsari, Raymond Fowler, Avery Nathens

Injury Prevention 2007;13:344–347. doi: 10.1136/ip.2007.015313

See end of article for authors' affiliations

Correspondence to: Dr B Roudsari, Department of Epidemiology, University of Texas School of Public Health, Dallas Regional Campus, 5323 Harry Hines, V8.112, Dallas, Texas 75390-9128, USA; bahman.roudsari@utsouthwestern.edu

Accepted 25 May 2007

Objective: To calculate the intraclass correlation coefficient (ICC) for emergency department (ED) shock rate, early trauma death (ie, death during the first 24 h after arrival at hospital), and in-hospital trauma death rate for multicenter childhood injuries.

Methods: The National Trauma Data Bank (5th revision), the largest multicenter trauma registry in the US, was used. Data from 80 trauma centers were used to calculate the ICC for in-hospital trauma death rate. Thirty three states provided data for calculation of the ICC for ED shock and early trauma death rate.

Results: From 2000 to 2004, 13% of the 952 242 patients in the National Trauma Data Bank were <15 years old. Approximately 17 000 of these children had injuries with an injury severity score >15, of whom 84% (14 095 subjects) were hospitalized at level I or II trauma centers in 33 states. The ICCs for ED shock rate, early trauma death rate, and in-hospital death rate were 0.005 (95% CI 0.000 to 0.010), 0.014 (95% CI 0.004 to 0.024), and 0.023 (95% CI 0.013 to 0.033), respectively. These ICCs were calculated for boys and girls and also for blunt and penetrating injuries.

Conclusion: Clustered childhood trauma studies that aim to compare different aspects of pre-hospital and hospital trauma care should incorporate these ICCs for sample calculation. When cluster randomized clinical trials are mounted, if sample sizes are calculated without adjustment for ICC, then the planned trial is likely to be seriously underpowered.

Trauma registries are valuable tools for monitoring quality of care and can be used to evaluate the effectiveness of injury prevention programs.¹ For example, Gentilello *et al*² used data from a local trauma registry to evaluate the potential influence of brief alcohol intervention on hospitalization for alcohol-related injuries. Trauma registries have also facilitated design and conduction of multicenter trauma studies. Roudsari *et al*³ used data from the local or national trauma registry of eight developed countries to evaluate the association between type of pre-hospital care and patient outcome in developed countries.³

Multicenter studies, in general, have larger sample sizes which make detection of the small but important differences in outcomes more likely. Furthermore, they may have better generalizability, as subjects are often recruited from heterogeneous settings. However, what is often not considered in multicenter studies is that patients cared for in the same centers (or regions) share attributes related to exposures or outcomes that are different from those of patients cared for in other centers or regions. As a result, patient outcomes may correlate strongly within clusters (defined as a center or region).^{4–7} This can be attributed to measured and unmeasured characteristics such as demography of the patients, type and mechanism of injury, and quality of pre-hospital and in-hospital care.⁸ As a result of this correlation, there is a lack of statistical independence across subjects. As independence is a critical assumption for sample size calculations and most other statistical analyses, it is likely that, in many cases, this assumption is violated.^{4–6,9} If each cluster comprises exposed and unexposed subjects, stratification for clusters can alleviate the clustering effect.⁵ However, if clusters are either all exposed or all unexposed, special methods should be adopted for the sample size calculation and data analysis.^{4–7} Under these circumstances, standard sample size formulas that assume that observations are independent estimate sample sizes that are

artificially low^{10–15} and result in studies that are underpowered for detecting the desired level of difference between exposed and unexposed groups. Furthermore, pooling the data from different centers and using standard statistical methods—that is, ignoring the clustering design—may result in standard errors and p values that are spuriously low.^{5,9–30}

To demonstrate the potential influence of clustering design on sample size calculation and analysis in multicenter trauma research, Roudsari *et al*⁸ conducted a hypothetical study. They evaluated the association between policy on pre-hospital intravenous fluid therapy and trauma fatality.⁸ The exposure of interest was the policy, which varied at the system level. As a result, patients in each emergency medical services (EMS) system were either exposed or not exposed to the policy. Therefore, to accomplish the objectives of the study, data from several EMS systems, some with and some without the policy, were required.⁸ The authors showed that ignoring the clustering design resulted in a sample size that was about seven times smaller than the actual sample size required to accomplish the objectives of the study. Furthermore, they showed that, in the analysis phase, not taking into consideration the clustering issue resulted in artificially narrow confidence intervals and mistaken rejection of the null hypothesis.

Intraclass correlation coefficient (ICC) is a measurement that captures between-cluster and within-cluster variability in outcome and is required for sample size calculation in clustered studies that comprise exposed or unexposed clusters.^{4–7} Equation (1) shows the formula for sample size calculation for clustered studies and its relationship to standard sample size formula⁴:

Abbreviations: ED, emergency department; EMS, emergency medical services; ICC, intraclass correlation coefficient; ISS, injury severity score; NTDB, National Trauma Data Bank

$$n = \underbrace{\frac{(Z_{\alpha/2} + Z_{\beta})^2 [P_1(1-P_1) + P_2(1-P_2)]}{(P_1 - P_2)^2}}_I \times \underbrace{[1 + (m_0 - 1) \times \text{ICC}]}_{II}$$

where n is the number of subjects in the exposed or unexposed group, Z is the Z score, α is the type I error, β is the type II error, P_1 and P_2 are the expected proportion of outcome in the exposed and non-exposed group, respectively, and m_0 is approximately the average number of subjects per cluster.

Part I in equation (1) is the formula used for sample size calculation in standard studies. Part II is called the “design effect” or “variance inflator factor”.⁴ As presented, the design effect depends on the average number of subjects per cluster and the ICC. In multicenter studies with a large number of subjects per cluster, even a small ICC can lead to a large design effect.

In this analysis, we used the US National Trauma Data Bank (NTDB)³¹ to calculate the ICC for three of the most commonly used outcomes of interest in acute trauma research in children <15 years old: emergency department (ED) shock, early trauma fatality (ie, death during the first 24 h after hospital arrival), and in-hospital trauma fatality (death before discharge). We have hypothesized that, owing to higher variability of in-hospital care compared with pre-hospital care, the ICC for in-hospital trauma fatality should be higher than those for early trauma death and ED shock rate.⁸

METHODS

We used the American College of Surgeons’ NTDB (version 5.0 released in November 2005).³² The NTDB is the largest trauma registry in the US and is based on the voluntary registry of hospitalized trauma patients admitted to American College of Surgeons-verified, state-designated, or self-designated trauma centers.³² Hospitals submit their de-identified trauma registry data to the NTDB. NTDB authorities are responsible for controlling the quality of the submitted data. Because participating in this national data registry is voluntary, larger hospitals with greater resources are more likely to contribute data. As a result, the NTDB is not nationally representative of all trauma centers throughout the US. Although hospital and city level identifiers are submitted to the NTDB, this information is not released in order to protect hospitals’ identities. However, to allow researchers to differentiate between states and hospitals, the NTDB provides random identification numbers that can be used to cluster patients within hospitals and hospitals within states.

In 2005, 567 trauma care facilities from 43 states and US territories, including 123 level I and 123 level II trauma centers contributed data to the NTDB. Level I and II trauma centers are tertiary trauma care facilities that are capable of providing the highest level of care to trauma patients.³³ The main difference between level I and II trauma centers is the training responsibilities that level I trauma centers have. In the presence of a local level I trauma center, level II trauma centers do not have such a responsibility.³³

In general, all records of hospitalized trauma patients with ICD-9CM codes of 800–959 are eligible to be included in the NTDB, except those with late effects of trauma, such as wound complication and late limb amputation, with ICD-9CM codes of 905–909.³² Comprehensive information on demographic characteristics of patients and pre-hospital and hospital care is included in the NTDB. For this analysis, we focused on children (<15 years old) with severe injuries (injury severity score³⁴ (ISS) >15) who were directly transported from the scene of

injury to a level I or II trauma center. We used the following variables from the NTDB: age, gender, mechanism of injury, ISS, ED blood pressure (the first blood pressure recorded in the ED), mortality status at discharge, random identification numbers for the participating hospitals and states, and trauma center designation level.

Two sets of outcome were used for evaluation of pre-hospital or hospital trauma care: ED shock rate (blood pressure <90 mm Hg at the time of arrival in the ED) and early trauma death (death during the first 24 h after arrival at the hospital) are reflective of the quality of care that patients receive in the pre-hospital setting and are less confounded by the quality of hospital care.⁸ However, hospital fatality after the first 24 h of arrival at the hospital is mostly influenced by the quality of hospital care.

To estimate the ICC for in-hospital trauma fatality, we treated each trauma center as a separate cluster. This can be explained by the potential differences in the quality of care among trauma centers even within the same geographical area. To increase the precision of our estimates, we used information from level I and II trauma centers that had contributed data from more than 50 severely injured children. Therefore, only 80 of the 246 level I and II trauma centers were eligible to be included in this analysis.

ED shock and early trauma death, however, are affected to some extent by the quality of pre-hospital care that is provided by the local EMS systems. Hospitals within the same region are often served by the same EMS system, calling for clustering at city level. However, the NTDB does not provide any city level identifier that can be used for clustering trauma centers by cities. The only other variable that could be used for clustering trauma centers based on the region is the state. As a result, we considered hospitals within the same state as a single cluster. Because of this state-level clustering, we ended up with 33 clusters that provided data for calculation of the ICC for ED shock and early trauma death rate.

We used Stata 9.1 (Stata Corporation, College Station, Texas, USA) for data analysis. We used the “loneway” command in Stata, which calculates ICC for studies with equal and unequal numbers of subjects per cluster. Each ICC is presented with its 95% CI. The mathematical calculations behind the ICC calculation are summarized in the Appendix.

RESULTS

From 2000 to 2004, 13% of the 952 242 patients in the NTDB were children <15 years old. About 17 000 of these children had severe injuries (ISS>15), of whom 84% (14 095 subjects) were hospitalized at a level I or II trauma center.

Sixty percent of the children were male, and the proportion of male patients varied from 0.33 to 0.67 among different centers. The mean ISS varied from 18 to 27 at the state level and 16 to 38 at the hospital level. Blunt trauma comprised 78–100% of the injuries across trauma centers.

Data from 80 level I and II trauma centers were used to calculate the ICC for trauma fatality rate, and 33 states provided data for calculation of the ICCs for ED shock rate and early trauma death rate. Table 1 summarizes the ICCs for male and female patients and blunt and penetrating injuries. As presented, in general the ICC for ED shock rate (0.005, 95% CI 0.000 to 0.010) was less than the ICC for in-hospital trauma fatality rate (0.023, 95% CI 0.013 to 0.033). A similar pattern was observed when we compared the ICCs of ED shock and in-hospital trauma fatality for male and female patients, and also for blunt and penetrating injuries, separately. For these four categories, the ICC of ED shock rate was lower than the ICC of in-hospital trauma fatality. The ICC for early trauma death was

Table 1 Intraclass correlation coefficients for the outcomes of interest, National Trauma Data Bank, 2001–2004

	ED shock	Early trauma death	Hospital fatality
Male	0.003 (0.000 to 0.007)	0.015 (0.003 to 0.026)	0.022 (0.011 to 0.033)
Female	0.008 (0.000 to 0.016)	0.013 (0.002 to 0.024)	0.018 (0.007 to 0.031)
Blunt injuries	0.005 (0.004 to 0.009)	0.012 (0.003 to 0.020)	0.019 (0.010 to 0.028)
Penetrating injuries*	0.020 (0.000 to 0.051)	0.033 (0.000 to 0.09)	0.068 (0.018 to 0.119)
Total	0.005 (0.000 to 0.010)	0.014 (0.004 to 0.024)	0.023 (0.013 to 0.033)

ED, emergency department.

Values in parentheses are 95% CI.

*Based on the small number of subjects and clusters.

between the ICCs for ED shock and in-hospital fatality, regardless of sex or type of injury (table 1).

DISCUSSION

ICC reflects the variability in outcome within and between clusters. Ideally, to estimate the sample size for a clustered study, the ICC for the outcome of interest should be calculated on the basis of clusters in the same population. However, owing to the large number of clusters required for ICC calculation, researchers often borrow ICCs from studies in other communities or from studies that comprise data from several communities. Under these circumstances, researchers should be aware of the potential differences in study settings that might influence the size of the ICCs. For instance, if between-cluster and within-cluster variability in outcome of interest in community A is smaller than the observed variabilities in other communities, the borrowed ICC is most likely a conservative estimate. As a result, the study in population A will have an adequate sample size to evaluate the association between the exposure and the outcome of interest. However, if the between-cluster and within-cluster variability in community A is larger, the borrowed ICC might be an anti-conservative estimate. Therefore, the study in community A might not have enough samples to accomplish its objectives.

In this study, we provided real estimates of ICCs for three important outcomes of interest in research into acute childhood injuries. Even with the small size of an ICC, in the presence of a large number of subjects per cluster, small ICCs result in large design effects.

Comparing pre-hospital with in-hospital trauma care, we expected to observe a lower ICC for outcomes related to pre-hospital care compared with those related to in-hospital care. The primary reason for this is the highly protocolized care by

EMS regions and the relatively few treatment modalities offered⁴ compared with the wide variation in hospital care across centers.

The only other study that has reported the ICCs for acute trauma research is one conducted by Roudsari *et al*⁸ which calculated the same ICCs for patients ≥ 15 years old. Figure 1 compares the ICC of ED shock, early trauma death, and hospital fatality comparing children with patients ≥ 15 years old.

Study limitations

First, participation in the NTDB is voluntary for the trauma care facilities. Therefore, it is not nationally representative of all trauma care facilities in the US. Owing to the involvement of trauma centers from different parts of the country in the NTDB, the between-cluster variability in the outcomes of interest is probably larger than in any other study that uses trauma centers from a single state. As a result, this larger variability in the NTDB inflates the calculated ICCs. Therefore, we expect that the studies that use the calculated ICCs in our study for sample size calculation will have sufficient sample sizes. Secondly, often a single EMS system is responsible for pre-hospital care in a city, and all trauma patients hospitalized in that city are exposed to a similar level of pre-hospital care. Therefore, to calculate the ICC for ED shock rate, which is mainly influenced by the quality of pre-hospital care, we would prefer to cluster trauma centers at the city level. However, to protect hospitals' identities, no city-level identifier was included in the NTDB. As a result, considering state instead of city as a clustering level increases the variability within clusters, whereas variability among clusters that could inflate or deflate the ICCs of ED shock rate and early trauma fatality might increase or decrease. However, in spite of these limitations, the NTDB provides a unique opportunity to estimate ICCs for the outcomes of interest using a large number of clusters at both the hospital and state level.

CONCLUSION

We have provided a real estimate of the ICC for three important outcomes of interest in acute childhood injury research. These estimates should be incorporated into sample size calculations for any multicenter trauma research that uses exclusively exposed or unexposed clusters of patients.

Authors' affiliations

Bahman Roudsari, University of Texas School of Public Health, Dallas Regional Campus, Dallas, Texas, USA

Raymond Fowler, University of Texas Southwestern Medical Center, Dallas, Texas, USA

Avery Nathens, St Michael's Hospital, Toronto, Ontario, Canada

Funding: None.

Competing interests: None.

Figure 1 Comparison of intraclass correlation coefficients (ICCs) between children and adults, National Trauma Data Bank, 2001–2004. ED, emergency department.

REFERENCES

- 1 Hoyt DB, Coimbra R, Potenza BM. Trauma systems, triage, and transport. In: Mattox KJ, Feliciano DV, Moore EE, eds. *Trauma*. New York: McGraw-Hill, 2004.
- 2 Gentilello LM, Rivara FP, Donovan DM, et al. Alcohol interventions in a trauma center as a means of reducing the risk of injury recurrence. *Ann Surg* 1999;**230**:473–80; discussion 480–3.
- 3 Roudsari B, Nathens A, Cameron P, et al. International comparison of prehospital trauma care systems. *Injury*. In press.
- 4 Donner A, Klar NS. *Design and analysis of cluster randomization trials in health research*. New York: Oxford University Press, 2000.
- 5 Localio AR, Berlin JA, Ten Have TR, et al. Adjustments for center in multicenter studies: an overview. *Ann Intern Med* 2001;**135**:112–23.
- 6 Murray DM. *Design and analysis of group randomized trials*. New York: Oxford University Press, 1998.
- 7 Murray DM, Varnell SP, Blitstein JL. Design and analysis of group-randomized trials: a review of recent methodological developments. *Am J Public Health* 2004;**94**:423–32.
- 8 Roudsari B, Nathens A, Koepsell T, et al. Analysis of clustered data in multicenter trauma studies. *Injury* 2006;**37**:614–21.
- 9 Donner A, Klar N. Pitfalls of and controversies in cluster randomization trials. *Am J Public Health* 2004;**94**:416–22.
- 10 Xie T, Waksman J. Design and sample size estimation in clinical trials with clustered survival times as the primary endpoint. *Stat Med* 2003;**22**:2835–46.
- 11 Cosby RH, Howard M, Kaczorowski J, et al. Randomizing patients by family practice: sample size estimation, intracluster correlation and data analysis. *Fam Pract* 2003;**20**:77–82.
- 12 Baskerville NB, Hogg W, Lemelin J. The effect of cluster randomization on sample size in prevention research. *J Fam Pract* 2001;**50**:W241–6.
- 13 Campbell MK, Mollison J, Grimshaw JM. Cluster trials in implementation research: estimation of intracluster correlation coefficients and sample size. *Stat Med* 2001;**20**:391–9.
- 14 Reading R, Harvey I, McLean M. Cluster randomised trials in maternal and child health: implications for power and sample size. *Arch Dis Child* 2000;**82**:79–83.
- 15 Hayes RJ, Bennett S. Simple sample size calculation for cluster-randomized trials. *Int J Epidemiol* 1999;**28**:319–26.
- 16 Byar DP. Some statistical considerations for design of cancer prevention trials. *Prev Med* 1989;**18**:688–99.
- 17 Campbell MK, Grimshaw JM, Elbourne DR. Intracluster correlation coefficients in cluster randomized trials: empirical insights into how should they be reported. *BMC Med Res Methodol* 2004;**4**:9.
- 18 Chuang JH, Hripcsak G, Heitjan DF. Design and analysis of controlled trials in naturally clustered environments: implications for medical informatics. *J Am Med Assoc* 2002;**287**:230–8.
- 19 Dahmen G, Rochon J, Konig IR, et al. Sample size calculations for controlled clinical trials using generalized estimating equations (GEE). *Methods Inf Med* 2004;**43**:451–6.
- 20 Dobbins TA, Simpson JM. Comparison of tests for categorical data from a stratified cluster randomized trial. *Stat Med* 2002;**21**:3835–46.
- 21 Eccles M, Grimshaw J, Steen N, et al. The design and analysis of a randomized controlled trial to evaluate computerized decision support in primary care: the COGENT study. *Fam Pract* 2000;**17**:180–6.
- 22 Eldridge S, Cryer C, Feder G, Underwood M. Sample size calculations for intervention trials in primary care randomizing by primary care group: an empirical illustration from one proposed intervention trial. *Stat Med* 2001;**20**:367–76.
- 23 Hayes RJ, Alexander ND, Bennett S, et al. Design and analysis issues in cluster-randomized trials of interventions against infectious diseases. *Stat Methods Med Res* 2000;**9**:95–116.
- 24 Klar N, Gyorkos T, Donner A. Cluster randomization trials in tropical medicine: a case study. *Trans R Soc Trop Med Hyg* 1995;**89**:454–9.
- 25 Omar RZ, Thompson SG. Analysis of a cluster randomized trial with binary outcome data using a multi-level model. *Stat Med* 2000;**19**:2675–88.
- 26 Piaggio G, Carroli G, Villar J, et al. Methodological considerations on the design and analysis of an equivalence stratified cluster randomization trial. *Stat Med* 2001;**20**:401–16.
- 27 Watson L, Small R, Brown S, et al. Mounting a community-randomized trial: sample size, matching, selection, and randomization issues in PRISM. *Control Clin Trials* 2004;**25**:235–50.
- 28 Weir CJ, Lees KR, MacWalter RS, et al. Cluster-randomized, controlled trial of computer-based decision support for selecting long-term anti-thrombotic therapy after acute ischaemic stroke. *QJM* 2003;**96**:143–53.
- 29 Yudkin PL, Moher M. Putting theory into practice: a cluster randomized trial with a small number of clusters. *Stat Med* 2001;**20**:341–9.
- 30 Yudkin P. Cluster effects to the design and analysis of a randomised trial. *Fam Pract* 1998;**15**:289.
- 31 American College of Surgeons Committee of Trauma. National Trauma Data Bank Report 2006, version 6. Chicago: American College of Surgeons, 2006.
- 32 Clark DE, Fantus R. National Trauma Data Bank (NTDB) Report 2006, version 6. Chicago: American College of Surgeons, 2006.
- 33 American College of Surgeons Committee on Trauma. Resources for optimal care of the injured patients. Chicago: American College of Surgeons, 2006.
- 34 Baker SP, O'Neill B, Haddon W Jr, et al. The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. *J Trauma* 1974;**14**:187–96.

APPENDIX

STATISTICAL APPROACHES TO CALCULATING ICCS USING THE NTDB

We used equation (1) to calculate ICCs for the outcomes of interest⁴:

$$ICC = 1 - \frac{\left(\sum_{j=1}^k m_j \times \hat{p}_j (1 - \hat{p}_j) \right)}{\left(\bar{m} - 1 \right) \times \hat{p} (1 - \hat{p})}$$

where $\hat{p}_j$ denotes the proportion of the observed outcome in the j th cluster; $\bar{m}$ is the average number of subjects in each cluster, and m_j is the number of the subjects in cluster j . $\hat{p}$ is the average proportion of the observed outcome in the data and is calculated from the following formula:

$$\hat{p} = \frac{\sum_{j=1}^k m_j \hat{p}_j}{\sum_{j=1}^k m_j}$$

If the number of subjects per cluster varies, equation (2) is used to substitute by m_0 .

$$m_0 = \left[M - \sum_{j=1}^k m_j^2 / M \right] / (k - 1)$$