8-28-2015 12:00 AM

On the estimation of intracluster correlation for time-to-event outcomes in cluster randomized trials

Sumeet Kalia, *The University of Western Ontario*

Supervisor: Dr. Allan Donner, *The University of Western Ontario*
Joint Supervisor: Dr. Neil Klar, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Epidemiology and Biostatistics

ON THE ESTIMATION OF INTRACLUSTER CORRELATION FOR TIME-TO-EVENT OUTCOMES IN CLUSTER RANDOMIZED TRIALS

(Thesis format: Monograph)

by

Sumeet Kalia

Graduate Program in Department of Epidemiology and Biostatistics

A thesis submitted in partial fulfillment
of the requirements for the degree of
Masters of Science

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

Abstract

Cluster randomized trials (CRTs) involve the random assignment of intact social units rather than independent subjects to intervention groups. Time-to-event outcomes often are endpoints in CRTs where the intracluster correlation coefficient (ICC) serves as a descriptive parameter to assess the similarity among outcomes in a cluster. However, estimating the ICC in CRTs with time-to-event outcomes is a challenge due to the presence of censored observations. The ICC is estimated for two CRTs using the censoring indicators and observed outcomes.

A simulation study explores the effect of administrative censoring on estimating the ICC. Results show that the ICC estimators derived from censoring indicators and observed outcomes are negatively biased for positively correlated outcomes. Analytic work further supports these results. Censoring indicators may be preferred to estimate the ICC under moderate frequency of administrative censoring while the observed outcomes may be preferred under minimal frequency of administrative censoring.

Keywords: Cluster randomized trials; intracluster correlation coefficient; correlated time-to-event outcomes; multivariate exponential distribution.

Acknowledgments

I am indebted to my supervisors, Dr. Allan Donner and Dr. Neil Klar. This thesis would not be possible without their guidance and support.

I am sincerely grateful to Dr. Guangyong Zou, Dr. Greta Bauer, Dr. Igor Karp and other faculty members for the invaluable lectures in Epidemiology and Biostatistics.

I would like to thank my friends and student colleagues for providing an intellectually stimulating environment. I would also like to thank my family for the support they provided me through my entire life.

This work was financially supported, in part, by an Ontario Graduate Scholarship and Dr. Donner's Natural Sciences and Engineering Research Council grant.

Contents

List of Figures

List of Tables

List of Acronyms

ANOVA analysis of variance
CRTs cluster randomization trials
ICC intracluster correlation coefficent
ML maximum likelihood
MSB between mean sums of squares
MSE mean square error
MSW within mean sums of squares
REML restricted maximum likelihood
SSB between sum of squares
SST total sum of squares
SSW within sum of squares
VIF variance inflation factor

Chapter 1

Introduction

Randomized trials are prospective studies designed to evaluate the effectiveness of a given intervention (Friedman et al., 2010). The random assignment of individual subjects ensures that, on average, the intervention groups are comparable with respect to both known and unknown baseline risk factors. In addition, random assignment helps to prevent selection bias that may originate from participants or investigators who may have a preference for one of the interventions being evaluated.

1.1 Cluster randomized trials

Investigators may at times choose to randomize intact social units (or clusters) instead of individual subjects (Donner and Klar, 2000, p. 5). Some possible benefits of randomizing clusters include reduced treatment contamination, increased administrative convenience, and improved participant compliance. In one example, investigators chose to randomize Hutterite colonies to evaluate the effect of influenza vaccine on infection rates (Loeb et al., 2010). The adoption of this design assured that possible treatment contamination of influenza vaccine effects under individual randomization is minimized in each colony.

The development of cluster randomization trials (CRTs) has independently arisen in var-

ious disciplines including medicine, social science, psychology and educational research. In the medical literature, there was an early recognition of decreased precision associated with CRTs, e.g. Mainland (1952, p. 114). Despite this early acknowledgment, many researchers continue to face difficulties with the design and analysis of CRTs. This is demonstrated in several methodological reviews including Simpson et al. (1995), Smith et al. (1997), and Walleser et al. (2011). In many instances, investigators may not be aware of the need to account for clustering and thus inappropriately choose standard statistical procedures to evaluate the intervention effect. However, it has been shown numerous times that the results derived from naïve methods ignoring clustering can often be misleading (Donner and Klar, 2000, p. 85, 113, 129). This is because the fundamental assumption of independence among outcomes no longer holds and therefore the use of standard statistical procedures is invalid. In particular, the estimated variance obtained from standard statistical procedures assuming all outcomes are independent underestimates the true variance and this leads to exaggerated effects of statistical significance. The design of CRTs can also be severely hampered if the sample size is estimated through standard statistical procedures because this may give rise to an inconclusive study with low power.

The most common experimental designs used in CRTs include the completely randomized, pair-matched and stratified designs. The completely randomized design randomly assigns clusters to the intervention groups without any matching or stratification. This design is most suitable when a large number of clusters are available since in this case effective randomization can ensure that both known and unknown baseline risk factors are, on average, balanced across the intervention groups. An example of a completely randomized design comes from a study of neonatal intensive care units where 114 hospitals were randomized to the intervention or control group (Horbar et al., 2004). The pair-matched design involves matching clusters on the basis of relevant baseline characteristics and then for each pair of matched clusters assigning

one cluster at random to the intervention group and the other cluster to the control group. As an example, 14 nursing homes were pair-matched on the basis of total number of beds and geographical proximity to evaluate an intervention program designed to prevent injuries among the residents (Ray et al., 1997). The stratified design is simply an extension of the pair-matched design where several clusters within each stratum are randomly assigned to the intervention group or control group. For example, each Hutterite colony within the seven health regions classified as a strata were randomized to one of the two treatment groups (Loeb et al., 2010). The rationale for adopting the pair-matched and stratified designs is that the probability of imbalance on important baseline characteristics can be substantially greater in trials randomizing clusters than for individually randomized trials with the same number of participants (Donner, 1992). Although an extensive literature exists on each of these designs, the focus here will be limited to the completely randomized design.

Many CRTs involve the collection of correlated data consisting of the time to a specified event. For example, Loeb et al. (2010) examined the time to laboratory confirmed influenza cases among the members of Hutterite communities. Outcomes were considered to be administratively censored if a given individual did not acquire influenza during the follow up of the trial. In addition, censored observations could arise if clusters or individual participants were lost to follow-up, chose to dropout or acquired a competing event which precluded observing the primary endpoint. For the sake of simplicity, the focus in this thesis is limited to administrative censoring where the assumption of non-informative censoring is satisfied. Non-informative censoring can be defined as independence between event times and the censoring mechanism possibly conditioned on a set of covariates (Wax et al., 1993). As a consequence, this assumption ensures that the probability of being censored does not depend on the prognosis of a subject or a cluster (Kleinbaum and Klein, 2012).

There are at least four types of accrual schemes in CRTs. Zou et al. (2005) noted three

types of accrual schemes in CRTs: (1) recruiting clusters sequentially while recruiting subjects simultaneously, (2) recruiting clusters simultaneously while recruiting subjects sequentially and (3) recruiting both clusters and subjects sequentially in relation to randomization. In an ideal design of CRT, the identification and recruitment of participants must be done prior to the random assignment (Eldridge et al., 2009). This design feature reduces the possibility of selection bias which may arise from differential rates of recruitment across intervention groups. Hence the focus in this thesis is limited to the accrual scheme where both the clusters and subjects are assumed to be available at the start of the trial. This focus in combination with limiting attention to administrative censoring simplifies the design and analysis of simulation study described in chapters 4 and 5. As an example, subjects within each Hutterite communities were recruited prior to the randomization of the study (Loeb et al., 2010). In addition, trial researchers (e.g. recruiters, assessors etc.) and subjects should also be blinded (whenever ethically possible) to the allocation status to reduce other sources of systematic error.

1.2 Intracluster correlation coefficient

The lack of independence among outcomes observed on cluster members creates statistical challenges that may be accounted for using an estimator of the parameter ρ, where ρ measures the degree of similarity among outcomes within a cluster and is known as the intracluster correlation coefficent (ICC). The ICC has a rich history of application in various fields of research (Haggard, 1958). For example, it may be used to measure familial resemblance in epidemiological studies, heritability of traits in genetic studies, and reliability of assessors in psychological studies (Donner, 1986). The ICC was first defined as a descriptive measure "to determine the resemblance for any series of characters of the individuals of sub-classes... as compared to the random pairs from the population (or universe) which they constitute" (Harris, 1913). Although the estimation of ICC's was first established for continuous endpoints (Donner, 1986), several researchers have extended it to alternative endpoints including binary

(Ridout et al., 1999) and time-to-event (Xie and Waksman, 2003).

1.2.1 Continuous outcomes

Pearson product-moment correlation can be used to estimate the ICC for continuous outcomes and it is defined in the present context as the pairwise correlation between any two members of the same cluster where each distinct pair of members is counted twice. However, as noted by Donner (1986), the pairwise correlation loses efficiency when substantial variation exists in the size of the clusters. For example, consider the Loeb et al. (2010) trial where the total number of subjects per cluster varied from 11 to 123. In this case, the group consisting of 123 subjects would be weighted approximately 136 times more than the group consisting of 11 subjects $\left(\text{i.e.}\frac{\binom{123}{2}}{\binom{11}{2}} \doteq 136\right)$. The magnitude of this weight is the ratio of two permutations that enumerate the total number of pairs. In order to overcome this disadvantage, a weighted pairwise correlation can be constructed as proposed by Karlin et al. (1981). The weights may be constructed so that (i) each pair of observations has equal weight, (ii) each group has equal weight regardless of their sizes, or (iii) each pair is weighted depending on the total number of pairs in which the given individual appears.

Fisher (1925) recognized that the ICC "merely measures the relative importance of two groups of factors causing variation" and hence introduced analysis of variance (ANOVA) to estimate the ICC based on what now is denoted a one-way random effects model. The ANOVA partitions the overall variance into two factors: (1) within-cluster and (2) between-cluster variance and directly estimates the ICC as the proportion of total variation accounted for by between-cluster variation. In the case of fixed cluster size (i.e. same number of subjects per cluster), the ANOVA estimator of variance components has many appealing properties. For example, the ANOVA estimators of variance components are unbiased and under normality have the minimum variance among all unbiased estimators (Searle et al., 1992, p. 129, 175). However, the ANOVA estimator of ICC is inherently biased given that the expected ratio

of variance components does not equal to the ratio of expected variance components (Ponzoni and James, 1978). Furthermore, in the case of variable cluster size where the between-cluster variance component is greater than the within-cluster variance component or equivalently ρ is greater than half, the performance of the ANOVA estimator of ICC is noted to be inadequate as compared to the maximum likelihood estimator (Swallow and Monahan, 1984).

Furthermore, maximum likelihood (ML) estimation can be used to obtain variance component estimators when the underlying continuous endpoints follows the multivariate normal distribution (Donner and Koval, 1980). Unlike the ANOVA estimator, the ML estimator ensures that the non-negativity constraints of between-cluster variance component is met when the likelihood function is maximized over its parametric space (Harville, 1977). However, the ML estimation of variance components does not take into account the degrees of freedom associated with fixed effects. Thus the restricted maximum likelihood (REML) estimation of variance components may be used to overcome this disadvantage (Searle et al., 1992, p. 41). The REML estimator first partitions the likelihood function to separate the fixed effects from random effects and subsequently maximizes the portion of the likelihood function containing the terms of the random effects. Computation of ML and REML estimator requires iterations when clusters are of variable size. The ML and REML estimators of variance components are asymptotically optimal and asymptotically equivalent for fixed and variable cluster sizes when the random errors are independent and normally distributed with fixed variance. In the case of fixed cluster size, the pairwise and ML estimator of ICC are asymptotically equivalent to the ANOVA estimator of ICC (Appendix A.1).

1.2.2 Binary outcomes

Several ICC estimators for binary endpoints have been reviewed by Ridout et al. (1999). Some of the estimators used in their study are directly adopted from methods used for continuous endpoints, including the ANOVA estimator and the unweighted and weighted pairwise es-

timators. Ridout et al. (1999) showed that the ANOVA, unweighted pairwise, and kappa-type estimators of ICC performed well as their bias and mean square error (MSE) was comparatively low in relation to the other estimators. As noted by Elston (1977), ANOVA provides a consistent estimator of ICC for dichotomous outcomes.

Kappa statistics are used in reliability studies to estimate the degree of agreement among pairs of raters corrected for chance. These measures of agreement may be defined so that raters have the same probability of an event (Ridout et al., 1999), i.e. marginal homogeneity. This version of kappa, known as Scott's π (Scott, 1955) is equal to an ICC applied to binary outcome data. Furthermore, for studies where there are more than two raters, Fleiss and Cuzick (1979) described an extension of Scott's π. In either case these statistics are equivalent to the Pearson pairwise estimator with constant weights in the case of fixed cluster size (Zou and Donner, 2004).

1.2.3 Time-to-event outcomes

A standard method of estimating the ICC for correlated time-to-event outcomes remains unclear (Jahn-Eimermacher et al., 2013). The current literature suggests that the ICC can be estimated from the binary indicators (i.e. censored: yes vs. no) and observed outcomes (i.e. omitting censored endpoints). For example, Xie and Waksman (2003) and Jahn-Eimermacher et al. (2013) used censoring indicators while Segal and Neuhaus (1993) and Williams (1995) used the observed outcomes to estimate the ICC.

The ICC estimator proposed by Xie and Waksman (2003) only uses the information available from censoring indicators and completely ignores the information available from the observed time-to-event outcomes. Similarly, Jahn-Eimermacher et al. (2013) defined an ICC parameter using the information available from only the censoring indicators. According to Xie and Waksman (2003) , an attractive feature of estimating the ICC from censoring indicators is its analogous relationship with the binary estimators of ICC which have been extensively

studied by many researchers.

In contrast, Gangnon and Kosorok (2004) noted that ICC parameter should not only depend on censoring indicators but also on the correlated time-to-event outcomes. Gangnon and Kosorok (2004) suggest performing the sample size calculation from conservative estimates of the ICC that are derived from a plausible event time distribution and the censoring distributions. Moreover, several researchers including Jung and Jeong (2003) and Su et al. (2011) have developed an expression for the ICC using information available from correlated time-to-event outcomes and censoring indicators. The relationship between estimating the ICC from censoring indicators and estimating it from observed time-to-event outcomes is further discussed in chapter 2.

1.3 Variance inflation factor

Even though the ICC can be interpreted as a descriptor of dependence within clusters, it also has considerable implication for the design and analysis of CRTs. For example, the variance inflation factor (VIF) is routinely used as a correction factor to adjust the estimated sample size (Donner et al., 1981) and test statistics for continuous (Donner and Klar, 1994a) and binary endpoints (Donner and Klar, 1994b). The early derivation of VIF in the context of clustered sampling was described by Hansen and Hurwitz (1942). The VIF is defined as the function of both the ICC and cluster size: $[1+(m-1)\times\rho]$, where m is the fixed cluster size. With this definition of VIF, it can be reflected that CRTs are less efficient than the individually randomized trials.

However, the derivation of VIF is complicated for correlated time-to-event outcomes. This is evident as Segal and Neuhaus (1997) note the absence of a VIF for the analysis of correlated time-to-event outcomes. Williams (1995) has shown that the Greenwood's formula underestimates the true variance of the survival function when the time-to-event outcomes are positively correlated within each cluster. More recently, Gangnon and Kosorok (2004) derived

a sample size formula for correlated time-to-event outcomes where the VIF was used as the adjustment factor. Unfortunately, limited attention has been given to the design and analysis of CRTs with time-to-event outcomes (Campbell and Walters, 2014, p. 63, 139). In spite of many challenges associated with the construction of confidence intervals, the estimation of ICC for time-to-event outcomes in CRTs is crucial to reduce the possibility of misleading inference from standard statistical procedures.

1.4 Objectives of the study

The focus of this study is limited to the estimation of ICC for correlated time-to-event outcomes. In particular, the ANOVA estimator of the ICC using the censoring indicators is compared with the ANOVA estimator of the ICC using the observed time-to-event outcomes. A Monte Carlo simulation study is conducted to compare the bias, variance, MSE, between-cluster and within cluster variance components of these two ICC estimators. A sign test is also conducted to compare the two ICC estimators with respect to how closely they estimate the ICC parameter.

1.5 Organization of the thesis

This thesis contains seven chapters. Chapter two provides more information on correlated survival data along with methods available for estimating the ICC from censoring indicators, observed outcomes and joint information. Chapter three provides some analytic results for the multivariate exponential model which is further used to design the simulation study. Chapter four describes the design of the simulation study and chapter five describes the results obtained from this investigation. Estimation of the ICC based on data from two published CRTs (Binka et al., 1996; Daly et al., 1995) is carried out in chapter six using the methods described in chapter four. Chapter seven summarizes key results and provides suggestions for future research.

Chapter 2

Literature Review

This chapter discusses the importance of estimating the ICC for correlated time-to-event outcomes. There are six sections in this chapter. Section 2.1 describes the key assumptions of the thesis. Section 2.2 introduces the notation that is used throughout the thesis. Section 2.3 describes the use of censoring indicators while section 2.4 describes the use of observed outcomes to estimate the ICC parameter. Section 2.5 describes the use of both censoring indicators and observed outcomes to estimate the ICC parameter. Section 2.6 summarizes the chapter.

2.1 Thesis assumptions

The focus of this thesis is limited to estimating the ICC using either censoring indicators or observed outcomes where the following assumptions are satisfied:

1. **Intervention arm:** control group of CRTs with time-to-event outcomes.

2. **Cluster size:** fixed number of subjects within each cluster at the start of the trial.

3. **Hazard rate:** constant event rate throughout the follow-up period of the trial.

4. **Accrual period:** all clusters and subjects are available at the start of the trial.

5. **Censoring:** all subjects who did not have the outcome by the end of the study follow-up are administratively censored.

2.2 Notation

For randomized trials, time-to-event or survival data are obtained by following subjects over time from a defined time origin (i.e. random assignment) to the occurrence of an event or until they are censored, e.g. end of study. Generally there are three required components for time-to-event data in CRTs: (1) time origin, (2) a positively valued random variable T_{ij} , and (3) a censoring indicator Δ_{ij}, where $i = 1,...,k$ denotes the i^{th} cluster and $j = 1,...,m$ denotes the j^{th} individual within the i^{th} cluster. The T_{ij} are continuous random variables recorded once for each subject over the specified follow-up period of the trial. If T_{ij} is not observed then the time-to-event outcome is administratively censored using the following censoring indicator

$$\Delta_{ij} = \begin{cases} 1 & \text{if } T_{ij} \leq t_c \\ 0 & \text{if } T_{ij} > t_c \end{cases} \tag{2.1}$$

where t_c denotes the total follow-up period of the study beyond which all the unobserved events are censored. The observed outcomes are defined as $T^*_{ij} = T_{ij}$ for m_i failure times $< t_c$. The censored observations only arise if the outcome of interest is not observed during the follow-up period of the study. This is analogous to the type I censoring mechanism in individually randomized trials (Kalbfeisch and Prentice, 2002, p. 5).

Let the cumulative distribution function of random variable T_{ij} be $F(\mathbf{t}) = Pr(\mathbf{T} \leq \mathbf{t})$ and the probability density function be $f(\mathbf{t}) = \frac{\partial}{\partial \mathbf{t}} F(\mathbf{t})$, where the boldface $\mathbf{t}$ denotes the correlated time-to-event outcomes within each cluster. The relationship between the survival function $S(\mathbf{t})$ and hazard function $\lambda(\mathbf{t})$ can be established as

$$S(\mathbf{t}) = exp\left(- \int_{\mathbf{0}}^{\mathbf{t}} \lambda(\mathbf{u}) \partial \mathbf{u} \right) \tag{2.2}$$

In this study, the hazard rate is restricted to be constant with respect to time and denoted as λ. Furthermore, the equivalence between the survival function $S(\mathbf{t}) = P(\mathbf{T} > \mathbf{t})$ and percentage of censored observations c can be noted with the assumption of no accrual period and administrative censoring. Without loss of generality, equation (2.2) can be expressed as

$$c = exp\left(-\int_0^{t_c} \lambda \partial u\right) = exp\left(-\lambda t_c\right). \tag{2.3}$$

There is no standard method of estimating the ICC for time-to-event outcomes (Jahn-Eimermacher et al., 2013). This is evident as several investigators used the censoring indicators, observed outcomes or the information from both sources to show the expression of the ICC. Furthermore, the literature review suggest that the ICC has been used to adjust for the effects of clustering for correlated time-to-event outcomes (see Table 2.1).

2.3 Censoring indicators

Censoring indicators, in this thesis, provide information about the occurrence or non-occurrence of an event within the study period. The proportion of dichotomized time-to-event outcomes may depend on the pre-specified follow-up period of the study. For example, increasing the follow-up period of the study may increase the proportion of recorded outcomes. Furthermore, this dichotomization only utilizes partial information because information about the exact occurrence of an event at time t is lost.

Lui (2000) used ANOVA to estimate the ICC from censoring indicators. With the one-way random effects model (see Appendix A.1), the ANOVA estimator of the ICC for fixed clusters of size m can be expressed as

$$\widehat{\rho}_A = \frac{\widehat{\sigma}_b^2}{\widehat{\sigma}_b^2 + \widehat{\sigma}_w^2} = \frac{MSB - MSW}{MSB + (m-1)MSW} \tag{2.4}$$

where $\widehat{\sigma}_b^2$ and $\widehat{\sigma}_w^2$ are between and within cluster variance component estimators, respectively. In the context of binary endpoints (i.e. censoring indicators), the between mean sums of squares

(MSB) and within mean sums of squares (MSW) are

$$\begin{aligned} MSB &= \frac{1}{k-1}\left(\sum_{i=1}^{k}\frac{\Delta_{i\cdot}^2}{m} - \frac{1}{N}\left(\sum_{i=1}^{k}\Delta_{i\cdot}\right)^2\right) \\ MSW &= \frac{1}{N-k}\left(\sum_{i=1}^{k}\Delta_{i\cdot} - \sum_{i=1}^{k}\frac{\Delta_{i\cdot}^2}{m}\right) \end{aligned}$$

where $\Delta_{i\cdot} = \sum_{j=1}^{m}\Delta_{ij}$ and $N = km$. The use of the ANOVA ICC estimator does not require any distributional assumptions to ensure consistency, at least in the absence of censoring.

The focus of this thesis is limited to continuous time-to-event outcomes. However, discrete time-to-event outcomes may arise in CRTs when the outcomes are recorded in time intervals (Meorbeek, 2012). For example, time intervals may be defined as academic sessions to record student graduation. Lui (2000) considered discrete time-to-event outcomes when comparing survival curves using the log-rank test. Furthermore, Lui (2000) claims that the log-rank test must account for ICC since the classical log-rank test may contribute to inflated Type I error rates.

Xie and Waksman (2003) used the following ICC estimator

$$\widehat{\rho}_1 = \frac{\sum_{i=1}^{k}\sum_{j\neq l}^{m}(\Delta_{ij}-\hat{p})(\Delta_{il}-\hat{p})}{\hat{p}(1-\hat{p})(km)(m-1)} \tag{2.5}$$

where $\hat{p}$ is the overall event rate observed in the study and it is defined as

$$\hat{p} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{m}\Delta_{ij}}{km}.$$

It is shown in Appendix (A.2) that $\widehat{\rho}_1$ is the binary analogue of the ML estimator derived from Gaussian outcomes in the case of fixed cluster sizes and thus it is equivalent to the ANOVA estimator in equation (2.4).

Jahn-Eimermacher et al. (2013) derived a sample size formula for correlated time-to-event outcomes where the ICC parameter was defined using the information available from censoring indicators. This ICC parameter was simplified using the method of moments and using the assumption of independence between Δ_{ij} and Δ_{il} (where $j \neq l$) conditional on the shared gamma

frailty (Z_i). The shared frailty induces the within-cluster correlation as it is shared among the members of the same cluster (Wienke, 2011).

2.4 Observed time-to-event outcomes

The observed time-to-event outcomes provide an alternative approach to estimate the ICC where censored observations are omitted. Segal and Neuhaus (1993) used the ANOVA estimator to compute the within-cluster correlation of observed outcomes. In the context of observed time-to-event outcomes, the ANOVA estimator of ICC must take into account the variable number of outcomes (m_i) recorded in the i^{th} cluster due to the omission of censored observations. Hence the ANOVA estimator for observed time-to-event outcomes (T^*_{ij}) is defined as in equation (2.4) by replacing m with

$$m_o = \left(\frac{1}{r-1}\right)\left(N^* - \sum_{i=1}^{r}\frac{m_i^2}{N^*}\right) = \overline{m} - \sum_{i=1}^{r}\frac{(m_i-\overline{m})^2}{N^*(r-1)} \tag{2.6}$$

where $\overline{m} = \frac{1}{r}\sum_{i=1}^{r} m_i$. Note that r is the total number of clusters in which at least one outcome is recorded and thus in this context the total number of observations (N^*) are defined as the sum of m_i outcomes within r clusters. The between and within mean sums of squares are

$$\begin{aligned} MSB &= \tfrac{1}{r-1}\textstyle\sum_{i=1}^{r}\sum_{j=1}^{m_i}\left(\overline{T}^*_{i\cdot} - \overline{T}^*_{\cdot\cdot}\right)^2 \\ MSW &= \tfrac{1}{N^*-r}\textstyle\sum_{i=1}^{r}\sum_{j=1}^{m_i}\left(T^*_{ij} - \overline{T}^*_{i\cdot}\right)^2 \end{aligned}$$

where $\overline{T}^*_{i\cdot} = \sum_{j=1}^{m_i}\frac{T^*_{ij}}{m_i}$ and $\overline{T}^*_{\cdot\cdot} = \frac{1}{N^*}\sum_{i=1}^{r}\sum_{j=1}^{m_i} T^*_{ij}$ are the cluster-specific and overall means of survival outcomes, respectively.

Segal and Neuhaus (1993) explored the limitation of transferring the dependence structure of an observed outcomes to censoring indicators. Segal and Neuhaus (1993) suggested that the censoring indicators will inherit the appropriate dependence structure if both the censoring and survival time of a given individual in the i^{th} cluster depends on the same frailty term (Z_i). Segal and Neuhaus (1997) were interested in estimating the design effect (i.e. variance

inflation factor). Williams (1995) showed that the Greenwood's formula underestimates the true variance of the survival function in the presence of correlated outcomes. Jung and Jeong (2003) and Jeong and Jung (2006) were interested in examining the validity of the log-rank test for correlated time-to-event outcomes. An expression of the ICC parameter is derived using different types of models (see Table 2.1).

2.5 Joint information

Gangnon and Kosorok (2004) noted that ICC parameters should not only depend on the censoring distribution but also on the dependence between the observed outcomes within the clusters. Gangnon and Kosorok (2004) used the event-driven design of CRTs which involves conducting a trial until pre-specified number of events are observed. Using this design, Gangnon and Kosorok (2004) defined an ICC parameter which is derived from the counting process of observed events. Furthermore, Jung (2007) described an ICC estimator that jointly uses the information from censoring indicators and observed outcomes. More recently, Jung (2008) noted that the ICC parameter proposed by Gangnon and Kosorok (2004) does not separate the dependence of paired survival outcomes from censoring indicators. Thus, Jung (2008) derived an ICC parameter using the joint information of censoring indicators and survival outcomes (see Table 2.1).

2.6 Summary

The literature review indicates that the ICC can be estimated from censoring indicators or observed outcomes. Regardless of which source of information is chosen, the estimation of ICC has considerable implications for the design and analysis of CRTs with time-to-event outcomes. This is evident as many researchers have used the ICC to adjust the sample size in the trial design (Xie and Waksman, 2003; Gangnon and Kosorok, 2004; Jahn-Eimermacher et al., 2013) and variance of the log-rank test in the analysis (Jeong and Jung, 2006; Lui, 2000).

Table 2.1: Literature review for the use of ICC

Censoring Indicators		
Author	**Model used**	**Motives**
Lui (2000)	Dirichlet-multinomial model	Validity of log-rank test
Xie and Waksman (2003)	Marginal proportional hazard model	Sample size estimation
Jahn-Eimermacher et al. (2013)	Conditionally shared gamma frailty model	Sample size estimation
Observed time-to-event outcomes		
Segal and Neuhaus (1997)	Marginally positive stable frailty model	Dependence estimation
Williams (1995)	Log-survival model	Variance of survival probability
Jung and Jeong (2003)	Marginally shared gamma frailty model	Validity of log-rank test
Su et al. (2011)	Clayton-Oakes model	Variance of survival probability
Joint information		
Gangnon and Kosorok (2004)	Marginal proportional hazard model	Sample size estimation
Jung (2007)	Extension of Lin and Ying (1993) model	Sample size estimation
Jung (2008)	Positive stable frailty model	Sample size estimation

Chapter 3

A multivariate exponential distribution

This chapter contains four sections. Section 3.1 describes a multivariate exponential distribution that is further used to generate the data as described in chapter 4. Section 3.2 derives the ICC parameter in the absence of censoring while section 3.3 derives the ICC parameter in the presence of administrative censoring using censoring indicators and observed outcomes, respectively. Section 3.4 summarizes this chapter.

3.1 Moran Algorithm

A multivariate exponential distributions can be used to model correlated time-to-event outcomes. However, as shown by Fréchet (1951), there are infinitely many multivariate exponential distributions. Gumbel (1960) presents one of the first such extensions of the bivariate exponential distribution. Some models (for e.g. positive stable frailty model (Segal and Neuhaus, 1997)) are ruled out from this simulation study since the dependence parameter is in the form of Kendall's τ (i.e. not the usual form of an ICC). However, marginally shared gamma frailty model (Jung and Jeong, 2003) may still be alternatively pursued as it meets the pre-specified criteria of this thesis (see section 2.1).

A marginal exponential model, first described by Moran (1967) using a simple algorithm,

has been used by several researchers (e.g. Kang and Koehler (1997), Jung (2008), Lee et al. (2007) and Su et al. (2011)) interested in correlated time-to-event outcomes. Furthermore, the dependence parameter in this model is in the form of an ICC (denoted as ρ) which is defined to be non-negative. For these reasons, Moran's algorithm is used to generate the correlated time-to-event outcomes in chapter 4.

The Moran (1967) algorithm is initiated from two mutually independent vectors:

$$\mathbb{A} = \begin{pmatrix} A_{i,1} & . & . & . & A_{i,m} \end{pmatrix}$$

and

$$\mathbb{B} = \begin{pmatrix} B_{i,1} & . & . & . & B_{i,m} \end{pmatrix}.$$

Both vectors are generated from multivariate normal distributions:

$$\mathbb{A}, \mathbb{B} \sim N \begin{pmatrix} 0 & 1 & \sqrt{\rho} & \dots & \sqrt{\rho} \\ 0 & \sqrt{\rho} & 1 & \dots & \sqrt{\rho} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \sqrt{\rho} & \sqrt{\rho} & \dots & 1 \end{pmatrix}.$$

The two vectors can be transformed to define the marginally exponential random variable T_{ij} with hazard rate λ where

$$T_{ij} = \frac{A_{ij}^2 + B_{ij}^2}{2\lambda} \qquad \text{for } i = 1,..,k \text{ and } j = 1,...,m. \tag{3.1}$$

The probability density funtion of T_{ij} derived from Moran's algorithm is described by Kotz et al. (2000, Chapter 47). Since the Moran-Downton multivariate exponential distribution is an extension of the bivariate case (Balakrishnan and Ng, 2001), the focus here is limited to bivariate exponential distribution for the sake of simplifying the presentation. Furthermore, the exchangeable correlation structure among outcomes within a cluster of arbitrary size can be described using the following bivariate probability density function:

$$P_{T_1,T_2}(t_1,t_2) = \frac{\lambda^2}{1-\rho} I_0\left(\frac{2\lambda\sqrt{\rho t_1 t_2}}{1-\rho}\right) exp\left(-\frac{\lambda(t_1+t_2)}{1-\rho}\right) \tag{3.2}$$

where $t_1, t_2 > 0$, $\lambda > 0$ and $0 \leq \rho \leq 1$. Note that $I_0(\cdot)$ is the modified Bessel function of the first kind of order zero and in this context it is defined as

$$I_0\left(\frac{2\lambda\sqrt{\rho t_1 t_2}}{1-\rho}\right) = \sum_{q=0}^{\infty}\left(\frac{1}{2q!}\frac{2\lambda\sqrt{\rho t_1 t_2}}{1-\rho}\right)^{2q}.$$

Furthermore, Balakrishnan and Lai (2009) derived the following cumulative distribution function of this bivariate probability density function:

$$F(t_1,t_2) = (1-e^{-t_1})(1-e^{-t_2}) + t_1 t_2 e^{-(t_1+t_2)}\sum_{q=0}^{\infty}\frac{\rho^{q+1}}{(q+1)^2}L_q^{(1)}(t_1)L_q^{(1)}(t_2) \tag{3.3}$$

where $L_q^{(1)}(\cdot)$ is Laguerre polynomial (Koekoek, 1990) and is defined as

$$L_q^{(1)}(t) = \sum_{s=0}^{q}(-1)^s\binom{q+1}{q-s}\frac{t^s}{s!}.$$

If $\rho = 0$ then the equation (3.2) simplifies to the product of two marginally exponential probability density functions. This is also reflected in the cumulative density function (equation 3.3) as it simplifies to $F(t_1,t_2) = (1-e^{-t_1})(1-e^{-t_2})$. With this framework, the ICC parameter can now be derived in the absence and presence of administrative censoring.

3.2 Absence of censoring

If the time-to-event outcomes follow Moran (1967) algorithm as described in section 3.1, then the variance of T_{ij} in the absence of censoring can be computed as

$$\begin{aligned}
Var(T_{ij}) &= Var\left(\frac{A_{ij}^2 + B_{ij}^2}{2\lambda}\right) \\
&= \frac{1}{4\lambda^2}\left(Var(A_{ij}^2) + Var(B_{ij}^2) + 2Cov(A_{ij}^2, B_{ij}^2)\right) \\
&= \frac{1}{4\lambda^2}\left(Var(A_{ij}^2) + Var(B_{ij}^2)\right) && \text{Since } A_{ij}^2 \perp\!\!\!\perp B_{ij}^2 \\
&= \frac{1}{4\lambda^2}(2+2) && \text{Since Var}(\chi_1^2) = 2 \\
&= \frac{1}{\lambda^2}.
\end{aligned}$$

Similarily the expectation of T_{ij} can be computed as

$$\begin{aligned} E(T_{ij}) &= E\left(\frac{A_{ij}^2 + B_{ij}^2}{2\lambda}\right) \\ &= \frac{1}{2\lambda}\left(E(A_{ij}^2) + E(B_{ij}^2)\right) \\ &= \frac{1}{\lambda} && \text{Since } E(\chi_1^2) = 1. \end{aligned}$$

The product expectation can also be computed as

$$\begin{aligned} E(A_{ij}^2 B_{il}^2) &= E(A_{ij}^2)E(B_{il}^2) && A_{ij}^2 \perp\!\!\!\perp B_{ij}^2 \implies Cov(A_{ij}^2, B_{ij}^2) = 0 \\ &= 1 && \text{Since } E(\chi_1^2) = 1. \end{aligned}$$

The following transformation can be used to find the $E(A_{ij}^2 A_{il}^2)$:

$$(A_{ij}, B_{ij}) = (A_{ij},\ \sqrt{\rho}A_{ij} + aC_{ij})$$

where $a^2 = 1 - \rho$ and (A_{ij}, C_{ij}) are independent and identically distributed standard normal random variables. With this transformation, the product moment of $A_{ij}^2 A_{il}^2$ can be computed as:

$$\begin{aligned} E\left(A_{ij}^2 A_{il}^2\right) &= E\left(A_{ij}^2\left(\sqrt{\rho}A_{ij} + aC_{ij}\right)^2\right) \\ &= E\left(\rho A_{ij}^4 + 2a\sqrt{\rho}C_{ij}A_{ij}^3 + a^2 A_{ij}^2 C_{ij}^2\right) \\ &= E\left(\rho A_{ij}^4 + a^2 A_{ij}^2 C_{ij}^2\right) && \text{Since } A_{ij} \perp\!\!\!\perp C_{ij} \text{ and } E(C_{ij}) = 0 \\ &= 3\rho + a^2 && \text{Since } E(\chi_1^2) = 1 \\ &= 2\rho + 1 && \text{Since } a^2 = 1 - \rho \end{aligned}$$

The results shown above can be used to find the expression of ICC parameter in the absence of censoring:

$$\begin{aligned} Corr(T_{ij}, T_{il}) &= \frac{Cov(T_{ij}, T_{il})}{\sqrt{Var(T_{ij})Var(T_{il})}} && j \neq l \\ &= \lambda^2\left(E(T_{ij}T_{il}) - E(T_{ij})E(T_{il})\right) \end{aligned}$$

$$= \lambda^2\left(\frac{1}{4\lambda^2}\left(E(A_{ij}^2A_{il}^2)+E(A_{ij}^2B_{il}^2)+E(B_{ij}^2A_{il}^2)+E(B_{ij}^2B_{il}^2)\right)-\frac{1}{\lambda^2}\right)$$
$$= \lambda^2\left(\frac{1}{4\lambda^2}(4\rho+4)-\frac{1}{\lambda^2}\right)$$
$$= \rho \qquad \text{as needed } \square$$

Since $Var(T_{ij}) = \frac{1}{\lambda^2}$ and $\rho = \frac{\sigma_B^2}{\sigma_B^2+\sigma_W^2}$, it can be shown that $\sigma_B^2 = \frac{\rho}{\lambda^2}$ and $\sigma_W^2 = \frac{1-\rho}{\lambda^2}$ in the absence of censoring.

3.3 Presence of censoring

The focus of this study is limited to administrative censoring where all subjects and clusters are available at the start of the study. The analytic expression for the ICC parameter using censoring indicators and observed outcomes, respectively, are described in the next two sections.

3.3.1 Censoring indicators

The censoring indicators are defined in equation (2.1), where $\Delta_{ij} = 0$ denotes censored observation and $\Delta_{ij} = 1$ denotes observed outcomes within the specified follow-up period of the study (t_c). Since the marginal probability of censoring is specified in advance to be c, it can be noted that

$$Pr(T_1 > t_c) = Pr(T_2 > t_c) = c.$$

Table 3.1: The joint and marginal probabilities of censoring indicators

	$\Delta_{ij}=0$	$\Delta_{ij}=1$	
$\Delta_{il}=0$	$Pr(T_1<t_c,T_2<t_c)+2c-1$	$1-c-Pr(T_1<t_c,T_2<t_c)$	c
$\Delta_{il}=1$	$1-c-Pr(T_1<t_c,T_2<t_c)$	$Pr(T_1<t_c,T_2<t_c)$	1-c
	c	1-c	1

As shown by Molenberghs et al. (1996) and by Klar et al. (2000), the ICC parameter for

binary outcomes (i.e. censoring indicators) can be defined as

$$\rho_c = \frac{Pr(T_1 > t_c, T_2 > t_c) - Pr(T_1 > t_c)Pr(T_2 > t_c)}{\sqrt{Pr(T_1 > t_c)(1 - Pr(T_1 > t_c))Pr(T_2 > t_c)(1 - Pr(T_2 > t_c))}} = \frac{Pr(T_1 > t_c, T_2 > t_c) - c^2}{c(1-c)}.$$

Using the cross-classification probabilities (Table 3.1) and cumulative distribution function (equation 3.3), the joint probability can be found as

$$\begin{aligned} Pr(T_1 > t_c, T_2 > t_c) &= 2c - 1 + Pr(T_1 < t_c, T_2 < t_c) \\ &= 2c - 1 + (1-c)^2 + (c \times ln(c))^2 \sum_{q=0}^{\infty} \frac{\rho^{q+1}}{(q+1)^2} \left[L_q^{(1)}(-ln(c)) \right]^2 \\ &= c^2 + (c \times ln(c))^2 \sum_{q=0}^{\infty} \frac{\rho^{q+1}}{(q+1)^2} \left[L_q^{(1)}(-ln(c)) \right]^2 \end{aligned}$$

Now the ICC parameter using the censoring indicators can be expressed as the function of the true ICC (ρ) and the probability of administrative censoring (c):

$$\begin{aligned} \rho_c &= \frac{c^2 + (c \times ln(c))^2 \sum_{q=0}^{\infty} \frac{\rho^{q+1}}{(q+1)^2} \left[L_q^{(1)}(-ln(c)) \right]^2 - c^2}{c(1-c)} \\ &= \frac{c^2 \times ln^2(c) \sum_{q=0}^{\infty} \frac{\rho^{q+1}}{(q+1)^2} \left[L_q^{(1)}(-ln(c)) \right]^2}{c(1-c)} \\ &= \frac{c \times ln^2(c) \sum_{q=0}^{\infty} \frac{\rho^{q+1}}{(q+1)^2} \left[L_q^{(1)}(-ln(c)) \right]^2}{1-c} \end{aligned}$$

Furthermore, the ICC parameter can be approximated using the third-order expansion of the Laguerre polynomial (where $q \geq 3$ is omitted) as:

$$\rho_c = \frac{c^2 \times ln^2(c) \left(\rho + \frac{\rho^2}{4}(2 + ln(c))^2 + \frac{\rho^3}{9} \left(3 + 3ln(c) + \frac{1}{2} ln(c) ln(c) \right)^2 \right)}{c(1-c)}. \tag{3.4}$$

If ρ is fixed in equation 3.4 and the amount of administrative censoring (c) is allowed to vary from 1% to 99% then ρ_c becomes negatively biased with respect to ρ. The disparity between the mean ANOVA ICC estimator using the censoring indicators ($\overline{\widehat{\rho_c}}$) and its analytic expression (ρ_c) is defined as $\overline{\widehat{\rho_c}} - \rho_c$. This disparity is further explained in chapter 4 and examined in chapter 5.

3.3.2 Observed outcomes

The time-to-event outcomes are recorded only if the event occurs within the length of the study follow-up (t_c). For the sake of simplicity, the hazard rate is assumed to be one event per year. The conditional expectation of observed time-to-event outcome T^*_{ij} can be found using the results of the marginal exponential distribution found in Moran (1967) algorithm:

$$E(T_{ij}|T_{ij} \leq t_c) = E(T^*_{ij}) = \frac{1}{F(t_c) - F(0)} \int_0^{t_c} T^*_{ij} exp(-T^*_{ij}) \partial T^*_{ij}$$

Using integration by parts and after some algebraic manipulation, the expectation of T^*_{ij} can be shown as:

$$E(T^*_{ij}) = \frac{1 - (1 + t_c) exp(-t_c)}{1 - exp(-t_c)}.$$

Similarly, the variance of T^*_{ij} can also be found using integration by parts:

$$\begin{aligned} Var(T^*_{ij}) &= E(T^*_{ij})^2 - \left(E(T^*_{ij})\right)^2 \\ &= \frac{2 - (t_c(t_c + 2) + 2) exp(-t_c)}{1 - exp(-t_c)} - \left(\frac{1 - (1 + t_c) exp(-t_c)}{1 - exp(-t_c)}\right)^2 \end{aligned}$$

The covariance of T^*_{ij} and T^*_{il} (where $j \neq l$) can be numerically integrated as:

$$\begin{aligned} Cov(T^*_{ij}, T^*_{il}) &= E(T^*_{ij} T^*_{il}) - E(T^*_{ij}) E(T^*_{il}) \\ &= \frac{1}{1 - \rho} \int_0^{t_c} T^*_{ij} exp\left(\frac{T^*_{ij}}{1 - \rho}\right) \int_0^{t_c} T^*_{il} exp\left(\frac{T^*_{il}}{1 - \rho}\right) I_0\left(\frac{2\sqrt{T^*_{ij} T^*_{il} \rho}}{1 - \rho}\right) \partial T^*_{ij} \partial T^*_{il} \\ &\quad - \left(\frac{1 - (1 + t_c) exp(-t_c)}{1 - exp(-t_c)}\right)^2. \end{aligned}$$

With this information, the ICC parameter using the observed outcomes can be found as:

$$\rho_t = corr(T^*_{ij}, T^*_{il}) = \frac{cov(T^*_{ij}, T^*_{il})}{\sqrt{var(T^*_{ij}) var(T^*_{il})}} \qquad \text{for } j \neq l. \tag{3.5}$$

The method of numerical integration to find the ICC parameter using observed outcomes is not pursued in this thesis. Instead, the small sample properties of the ICC estimator using observed outcomes ($\widehat{\rho_t}$) are examined by the simulation study alone.

3.4 Summary

The Moran (1967) algorithm is used to generate the correlated time-to-event outcomes in chapter 4. The small sample properties of ICC parameter using censoring indicators (equation 3.4) are further evaluated in chapter 5. However, the ICC estimators using observed outcomes are only examined by the simulation study alone.

Chapter 4

Design of the Simulation Study

There are five sections in this chapter. Section 4.1 describes and justifies the selected parameter combinations. Section 4.2 explains how the data are generated using the multivariate exponential model. Section 4.3 describes the ICC estimators used in this simulation study. The criterion used to assess the performance of the ICC estimators are outlined in section 4.4. Finally section 4.5 summarizes the key issues described in this chapter.

4.1 Parameters

According to Burton et al. (2006), the parameter values should ideally represent the most common circumstances. Small and positive values of ICC are most commonly observed in CRTs (Donner and Klar, 2000, p. 9) and thus the parameter values of ICC are chosen accordingly (Table 4.1). In addition, as noted in chapter 2, a balanced study design with only one follow-up group is considered. This would be akin to using data from the control group of a completed CRT to estimate the ICC for a new trial. Forty clusters are chosen to evaluate the performance of each ICC estimator. The choice for the total number of clusters (k) and cluster sizes (m) is motivated by the following trials: Binka et al. (1996); Daly et al. (1995). Data from these trials are used as examples in chapter 6.

In this simulation study, the assumptions of no accrual period and constant marginal hazard rate (λ) simplify the relationship between study follow-up time (t_c) and percentage of administratively censored observations (c) as shown in equation (2.3). The assumption of constant hazard rate is in agreement with the prior simulation studies (Jung and Jeong, 2003; Su et al., 2011). Furthermore, a change in constant hazard rate implies a linear transformation of time-to-event outcomes (see section 3.1). Since the estimation of ICC is unaffected by linear transformation of the outcomes (Solomon and Taylor, 1999), the hazard rate is a nuisance factor in this simulation study and thus restricted to be one event per person-year.

Table 4.1: Parameter combinations of the simulation study

Parameter	**Description**	**Values**
ρ	ICC	0, 0.01
k	Number of clusters	40
m	Cluster size	200, 2
λ	Hazard rate (per person-year)	1
c (†)	Percent censored (%)	1% to 99% in increment of 1%

† also investigate the scenarios where the censoring mechanism is ignored (see section 4.4.5).

Since all subjects enter the study at the same point in time and since only administrative censoring is imposed, the censored observations are determined by dichotomizing the correlated time-to-event outcomes at the termination point of the study. The parameter values of percent censoring encapsulate the entire domain. For example, the primary outcome of early childhood mortality in Binka et al. (1996) is rare (high percent censored) while the primary outcome of tube failure in Daly et al. (1995) is common (low percent censored). The marginal hazard rate and amount of anticipated censoring is specified in advance to fix the length of the study. The length of the study ranges from 3.6 days (with 99% censored observations) to 4.61 years (with 1% censored observations). A factorial design is used to generate 396 ($= 2\times1\times2\times1\times99$) scenarios, which are derived from the parameter values listed in Table 4.1. Each scenario involves the generation of 1,000 independent datasets.

4.2 Data generation

Moran's algorithm is used to generate the correlated time-to-event outcomes, T_{ij} . This algorithm is initiated with the k replicates of two independent vectors $\mathbb{A}$ and $\mathbb{B}$ as described in section 3.1. The simulation study is conducted using SAS V. 9.4 (SAS Institute Inc., 2012). The generation of random numbers for the multivariate normal distribution is performed using the RANDNORMAL command for the clusters of size 200 and 2. In addition, a unique starting seed is specified for each scenario to ensure its independence with respect to other scenarios and its replication for future research. The SAS code used to generate the data is provided in Appendix B.

4.3 Estimation of the ICC

The ANOVA ICC is estimated using (1) censoring indicators and (2) observed outcomes (i.e. omitting the censored observations). The censoring indicators dichotomize the occurrence of time-to-event outcomes as shown in equation (2.1). The observed time-to-event outcomes are recorded only if the event occurs within the follow-up period of the study.

4.3.1 Truncation of the ICC estimator

Even though the ICC parameter is non-negative, a negative estimate of ICC can arise due to sampling error. The probability of obtaining a negative ICC estimate is inversely proportional to the number of clusters, the cluster size and the ICC parameter (Wang et al., 1992). For example, a relatively small number of clusters coupled with a smaller ICC parameter may lead to a substantial increase in the probability of obtaining a negative ICC estimate. Hence, many researchers (e.g. Lui (2000); Chakraborty et al. (2009)) considered the truncated ICC estimator, which involves truncating negative ICC estimates to zero. Although the use of truncated ICC estimators is often not recommended (Wang et al., 1992), such estimates are reported in this simulation study for the purpose of comparison with non-truncated ICC estimates.

4.3.2 Removal of singletons

The ICC estimator derived from observed outcomes involves the removal of censored observations and this will give rise to clusters of variable size. At times there may be only one member per cluster especially when $m = 2$. Even though this singleton provides an extra degree of freedom, it does not provide any additional information with respect to the within-cluster variance component of the ICC estimator. Thus the effect of removing a singleton is also assessed for the ICC estimator derived from observed time-to-event outcomes. It must be noted that the presence of singletons may increase the estimated variance when the values of the ICC are specified to be small (Swiger et al., 1964). The ICC estimator obtained from censoring indicators uses the information from fixed cluster sizes ($m = 200$ and 2) and thus the omission of singletons is not then applicable.

The conditions of truncating and of omitting singletons give rise to six ICC estimators in the simulation study, as listed in Table 4.2. There are four ICC estimators using observed outcomes: $\widehat{\rho}_t, \widehat{\rho}^*_t, \widehat{\rho}_{t\sim}$ and $\widehat{\rho}^*_{t\sim}$. The asterisk ($*$) indicates truncation of the ICC estimator while tilda ($\sim$) indicates omission of clusters with a single observation. On the contrary, there are two ICC estimators derived from censoring indicators: $\widehat{\rho}_c$ and $\widehat{\rho}^*_c$.

Table 4.2: The ANOVA ICC estimators of the simulation study

ICC estimator	**Source** (c,t)	**Negative estimates** $(*)$	**Omitting singletons** $(\sim)$
$\widehat{\rho}_c$	Censoring Indicators	Included	N/A
$\widehat{\rho}^*_c$		Truncated	N/A
$\widehat{\rho}_t$	Observed time-to-event	Included	No
$\widehat{\rho}^*_t$	outcomes	Truncated	No
$\widehat{\rho}_{t\sim}$		Included	Yes
$\widehat{\rho}^*_{t\sim}$		Truncated	Yes

4.3.3 Undefined ICC estimators

The ICC estimator derived from censoring indicators is undefined when all subjects either have the event during follow-up or all subjects have censored outcomes because then MSB

and MSW are zero. Similarly, the ICC estimator derived from observed outcomes is undefined when all observations are censored. The probability of obtaining such results is increased when the percent censoring is close to zero or one. Such replicates are recorded as failed samples and thus discarded from this simulation study. Furthermore, the failed replicates may contribute to biased and imprecise results (Burton et al., 2006). Hence, the results are only reported for scenarios with less than 10% failed samples.

4.4 Performance of the ICC estimators

It is important to assess the performance of an estimator using more than one criterion (Burton et al., 2006). Hence, the performance of the six ICC estimators described in Table 4.2 is assessed using bias, variance, MSE and a sign test measurng the relative frequency with which one estimator is closer to the true value of the ICC as compared to the other.

4.4.1 Bias

Bias is assessed by measuring the deviation of the mean ICC estimate in relation to the ICC parameter: $\delta = \overline{\widehat{\rho}_v} - \rho$, where $\overline{\widehat{\rho}_v}\left(= \sum_{w=1}^{n} \frac{\widehat{\rho}_{v,w}}{n}\right)$ denotes the mean ICC estimate derived from censoring indicators or observed outcomes with appropriate conditions of truncation and removal of singletons (Table 4.2).The index w is used to denote, n, the number of successful replications (see section 4.3.3). Note that n is only reported for scenarios with less than 10% failure rate of estimating the ICC (i.e. $n \geq 900$).

This definition of bias is chosen because the results of relative bias are uninterpretable when the ICC parameter is specified to be zero. However, the bias is standardized to establish the 10% relative bias as a criterion for acceptable amount of bias in this simulation study when $\rho \neq 0$ (Wang et al., 1992). The biases of between $\left(\sigma_B^2\right)$ and within $\left(\sigma_W^2\right)$ cluster variance components are also assessed to gain more insight on the bias of ICC estimators. Since the hazard rate is restricted to be one event per person year, σ_B^2 and σ_W^2 are a function of the ICC parameter where $\sigma_B^2 = \rho$ and $\sigma_W^2 = 1 - \rho$ (see section 3.2). For the sake of simplicity, the biases

of only non-truncated variance components are assessed.

4.4.2 Variance

In this simulation study, the variance of the ICC estimators is computed as:

$$Var(\widehat{\rho_v}) = \sum_{w=1}^{n} \frac{(\widehat{\rho}_{v,w} - \bar{\widehat{\rho}}_v)^2}{n-1}.$$

4.4.3 Mean square error

MSE is used to measure the accuracy of ICC estimators. MSE is estimated as the sum of variance and square of the bias (Wackerly et al., 2008, pg. 393). Equivalently, MSE is also defined as:

$$\text{MSE}(\widehat{\rho_v}) = \frac{1}{n} \sum_{w=1}^{n} (\widehat{\rho}_{v,w} - \rho)^2.$$

4.4.4 Sign test

A sign test is used to construct a criterion for the within-run comparison of the ICC estimators (Rosner et al., 1977) derived from censoring indicators and observed outcomes. The purpose of this test is to determine which ICC estimator tends to be consistently closer to ρ. The non-truncated ICC estimator using the censoring indicators $(\widehat{\rho_c})$ is compared with the non-truncated ICC estimator using the observed outcomes $(\widehat{\rho_t})$. Hence only one test statistic of the sign test $(d_{c,t})$ is constructed since the other four ICC estimators (Table 4.2), involving truncation and omission of singletons, intrinsically depend on $\widehat{\rho_t}$ and $\widehat{\rho_c}$.

An indicator variable is defined using the following expression:

$$\eta = |\widehat{\rho_c} - \rho| - |\widehat{\rho_t} - \rho|$$

for each of the n replicates. If η is positive for a particular replicate then it is coded one; otherwise it is coded zero. Furthermore, the proportion $d_{c,t}$ is computed as the average of the indicator variables over n replicates. The two-sided P-value is obtained from the binomial

distribution testing for the deviation of the proportion $d_{c,t}$ from 0.5. It can be deduced that if $d_{c,t} < 0.5$ then $\widehat{\rho}_c$ is estimating the ICC parameter more closely than $\widehat{\rho}_t$.

4.4.5 Other exploratory measures

The consistency of the non-truncated ANOVA estimator is explored in the absence of censoring mechanism to ensure the validity of this simulation study. The disparity between the ICC parameter using the censoring indicators (equation 3.4) and the mean ICC estimator using the censoring indicators $\left(\overline{\widehat{\rho}_c}\right)$ is evaluated to show agreement between the analytic work and the results of the simulation study. In addition, the correlation between the non-truncated ICC estimators using censoring indicators and observed outcomes is assessed (in the presence of administrative censoring) to discern the relationship between the two methods of estimating the ICC.

4.5 Summary

The primary focus of this simulation study is to compare the performance of six ICC estimators (Table 4.2) under the factorial combinations of five parameters (Table 4.1). The bias, variance and MSE of ICC estimators are obtained for all the investigated scenarios where the failure rate of estimating the ICC is less than 10%. In addition, a sign test is performed to determine which ICC estimator tends to be closer to ρ.

Chapter 5

Results of the simulation study

There are five sections in this chapter. Section 5.1 discusses how the results are formulated. Section 5.2 explores the consistency of ANOVA ICC estimator in the absence of censoring. Section 5.3 describes the performance of ICC estimators in the presence of administrative censoring. The criteria of bias, variance, MSE and sign test are used to assess the performance of ICC estimators. Section 5.4 describes some exploratory results of this study. Section 5.5 summarizes this chapter.

5.1 Formulation of results

The results of this simulation study are divided into four cases based on the factorial combinations of ρ (= 0 and 0.01), m (=200 and 2) and k (=40) specified in Table 4.1. Each case consist of 99 scenarios where percent censoring is specified from 1% to 99% in the increments of 1%. The results of scenarios with more than 10% failure rate of estimating the ICC are excluded (see section 4.3.3).

Series plots are used to compare the performance of the ICC estimators over the entire domain of percent censoring. The truncated estimators are labeled with dotted lines while the non-truncated estimators are labeled with solid lines. The results of the sign test are also shown

using the series plot where the shaded region imply statistical significance at the nominal level of 0.05 and 0.001. In particular, the upper half of the shaded region indicates that the ANOVA ICC estimator using the observed outcomes estimated the true parameter more closely than the ANOVA ICC estimator using the censoring indicators. The lower half of the shaded region indicates the opposite.

5.2 Absence of censoring

The ANOVA ICC is estimated using the data generated from the Moran's algorithm (i.e. without imposing any administrative censoring). One thousand replicates are nested within each scenario (denoted as θ) and 99 scenarios are nested within each case. Hence, Table 5.1 provide the overall summary of estimating the ICC in the absence of censoring for 396 parameter combinations of this simulation study.

Table 5.1: Mean bias of ANOVA ICC estimator in the absence of censoring

	Parameter combinations				Mean bias		
Case	ρ	m	k	θ †	Minimum	Median	Maximum
1	0	200	40	99	-0.00007	0.0000007	0.0001
2	0.01	200	40	99	-0.0005	-0.0001	0.0003
3	0	2	40	99	-0.01	0.0007	0.01
4	0.01	2	40	99	**-0.01** ‡	-0.002	**0.02** ‡

† corresponds to the 99 scenarios in the presence of administrative censoring.
‡ violates the criteria of 10% relative bias.

5.3 Presence of administrative censoring

Administrative censoring is imposed for the four cases listed in Table 5.1. Only the non-truncated ICC estimators ($\widehat{\rho}_c$,$\widehat{\rho}_t$) are reported for the first two cases because the effects of truncating the ICC estimators and of omitting singletons are negligible when $m = 200$. However, the effects of truncating the ICC estimators and of omitting singletons are substantial when $m = 2$. Thus, the six ICC estimators listed in Table 4.2 are reported when $m = 2$.

5.3.1 Time-to-event outcomes in community trials ($m = 200$)

The mean bias of non-truncated ICC estimators using censoring indicators ($\widehat{\rho_c}$) and observed outcomes ($\widehat{\rho_t}$) is summarized in Table 5.2. Furthermore, the mean bias of $\widehat{\rho_c}$ and $\widehat{\rho_t}$ over the entire domain of percent censoring is shown in Figure 5.1 when $\rho = 0$; Figure 5.7 when $\rho = 0.01$. The mean bias of $\widehat{\rho_c}$ and $\widehat{\rho_t}$ is symmetrically distributed around zero when $\rho = 0$. However, the mean bias of $\widehat{\rho_c}$ follows a negative quadratic relationship with respect to administrative censoring when $\rho = 0.01$. In contrast, the gradual increase in the magnitude of negative bias is evident for $\widehat{\rho_t}$ with the increasing frequency of administrative censoring when $\rho = 0.01$.

Table 5.2: The summary of mean bias when $m = 200$

			Mean bias		
ρ	Source	ICC estimator	Minimum	Median	Maximum
0	Censoring indicators	$\widehat{\rho_c}$	-0.001	0.00001	0.0001
0	Observed outcomes	$\widehat{\rho_t}$	-0.004	-0.00001	0.001
0.01	Censoring indicators	$\widehat{\rho_c}$	**-0.01** ‡	**-0.006** ‡	**-0.003** ‡
0.01	Observed outcomes	$\widehat{\rho_t}$	**-0.01** ‡	**-0.01** ‡	**-0.002** ‡

‡ violates the criteria of 10% relative bias.

The following description of results is applicable to case where $\rho = 0.01$. The variance of $\widehat{\rho_c}$ remains fairly constant while the variance of $\widehat{\rho_t}$ increases with respect to administrative censoring (see Figure 5.8). The MSE of $\widehat{\rho_c}$ is greater than MSE of $\widehat{\rho_t}$ under low frequency of administrative censoring (see Figure 5.9). However, the MSE of $\widehat{\rho_c}$ is consistently lower than MSE of $\widehat{\rho_t}$ under moderate to high frequency of administrative censoring. The between-cluster and within-cluster variance components of $\widehat{\rho_c}$ and $\widehat{\rho_t}$ are negatively biased (see Figure 5.10 and Figure 5.11, respectively) . The results of sign test indicate that $\widehat{\rho_t}$ estimates the true ρ more closely than $\widehat{\rho_c}$ under minimal frequency of administrative censoring (see Figure 5.12). However, $\widehat{\rho_c}$ consistently estimates the true ρ more closely than $\widehat{\rho_t}$ under moderate to high frequency of administrative censoring.

Figure 5.1: Mean ICC estimate for case 1: $\rho = 0$, $m = 200$, $k = 40$

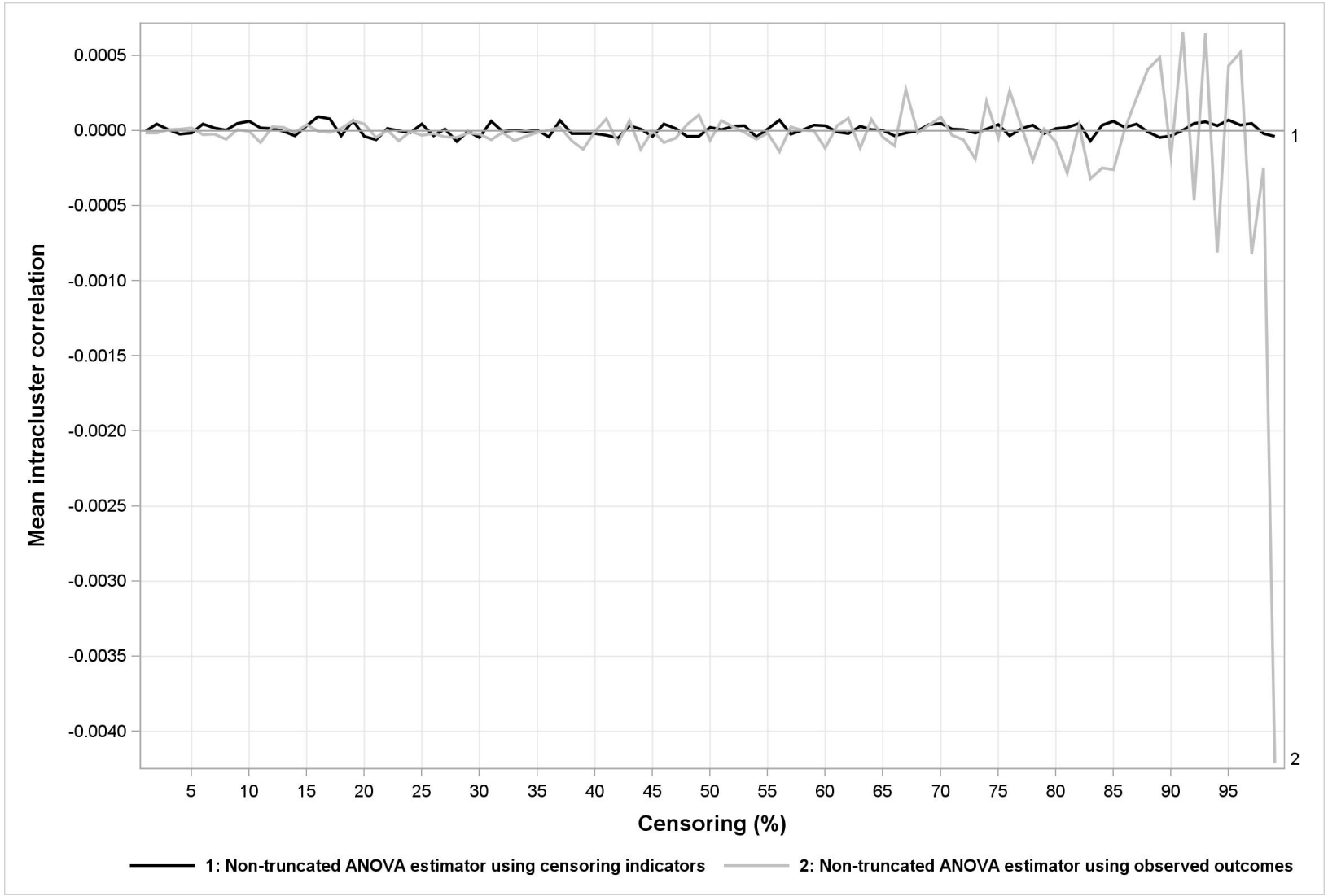


Figure 5.2: Variance of ICC estimators for case 1: $\rho = 0$, $m = 200$, $k = 40$

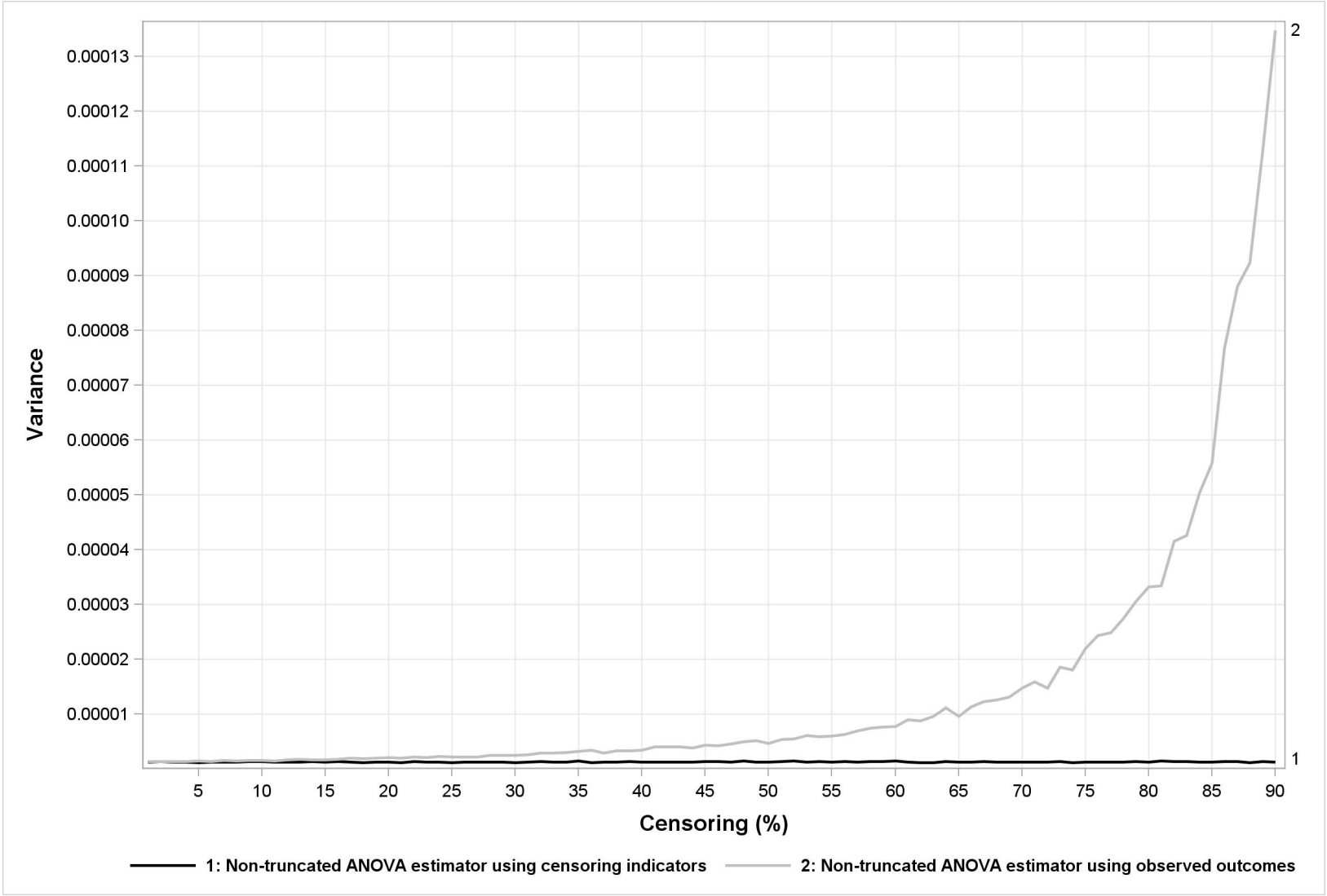


Figure 5.3: Mean square error of ICC estimators for case 1: $\rho = 0$, $m = 200$, $k = 40$

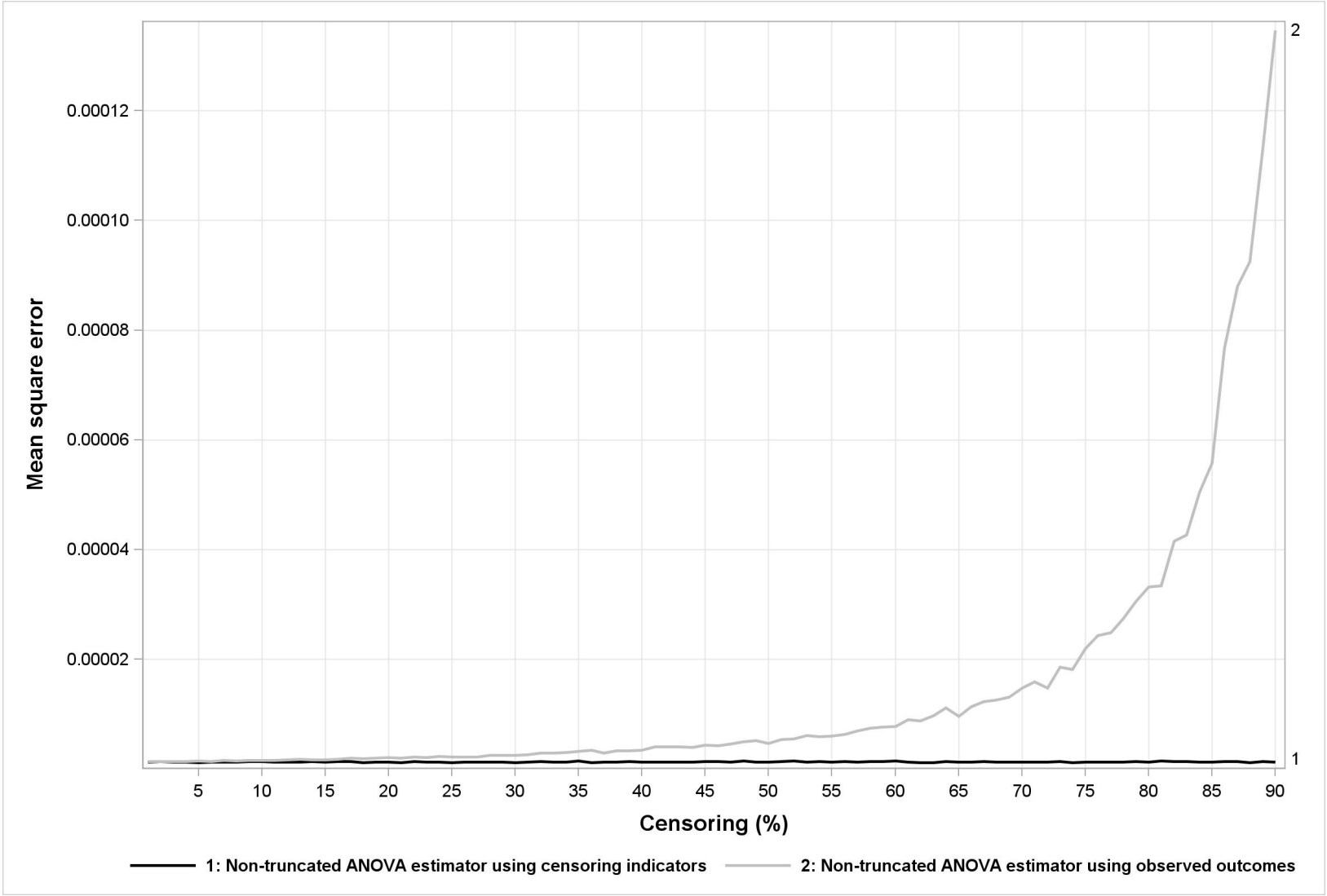


Figure 5.4: Mean between-cluster variance component for case 1: $\rho = 0$, $m = 2$, $k = 40$

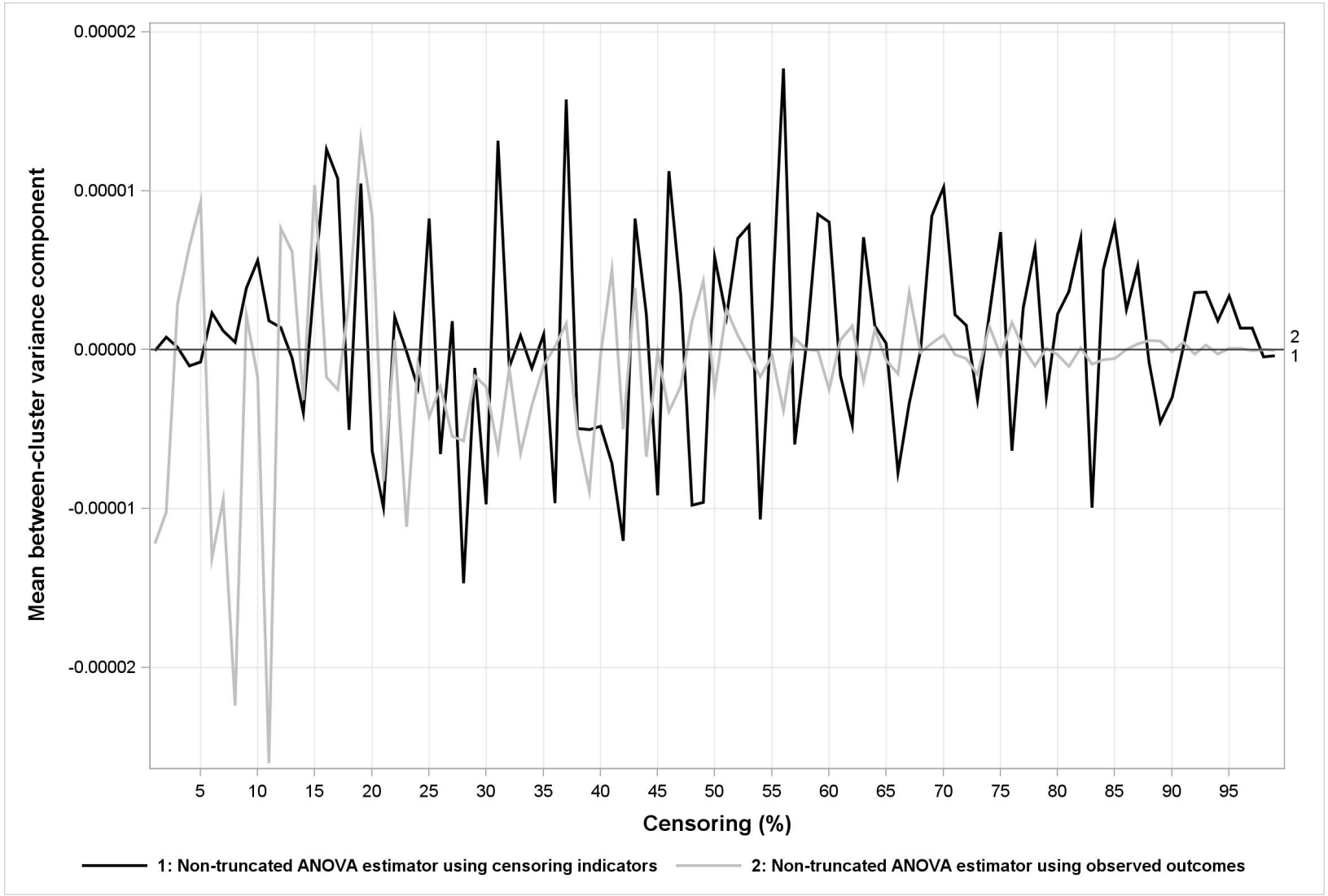


Figure 5.5: Mean within-cluster variance component for case 1: $\rho = 0$, $m = 200$, $k = 40$

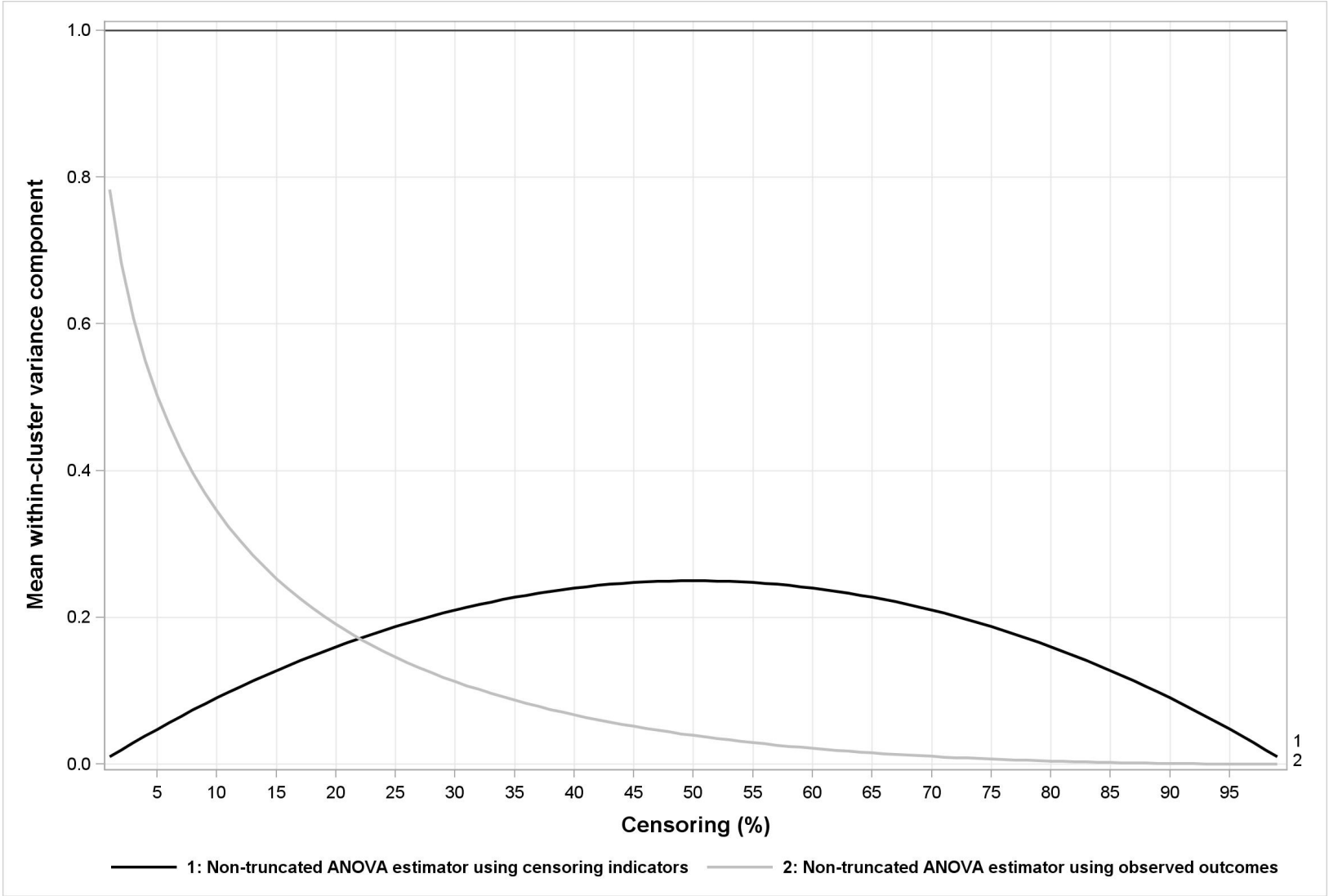


Figure 5.6: Sign test for case 1: $\rho = 0$, $m = 2$, $k = 40$

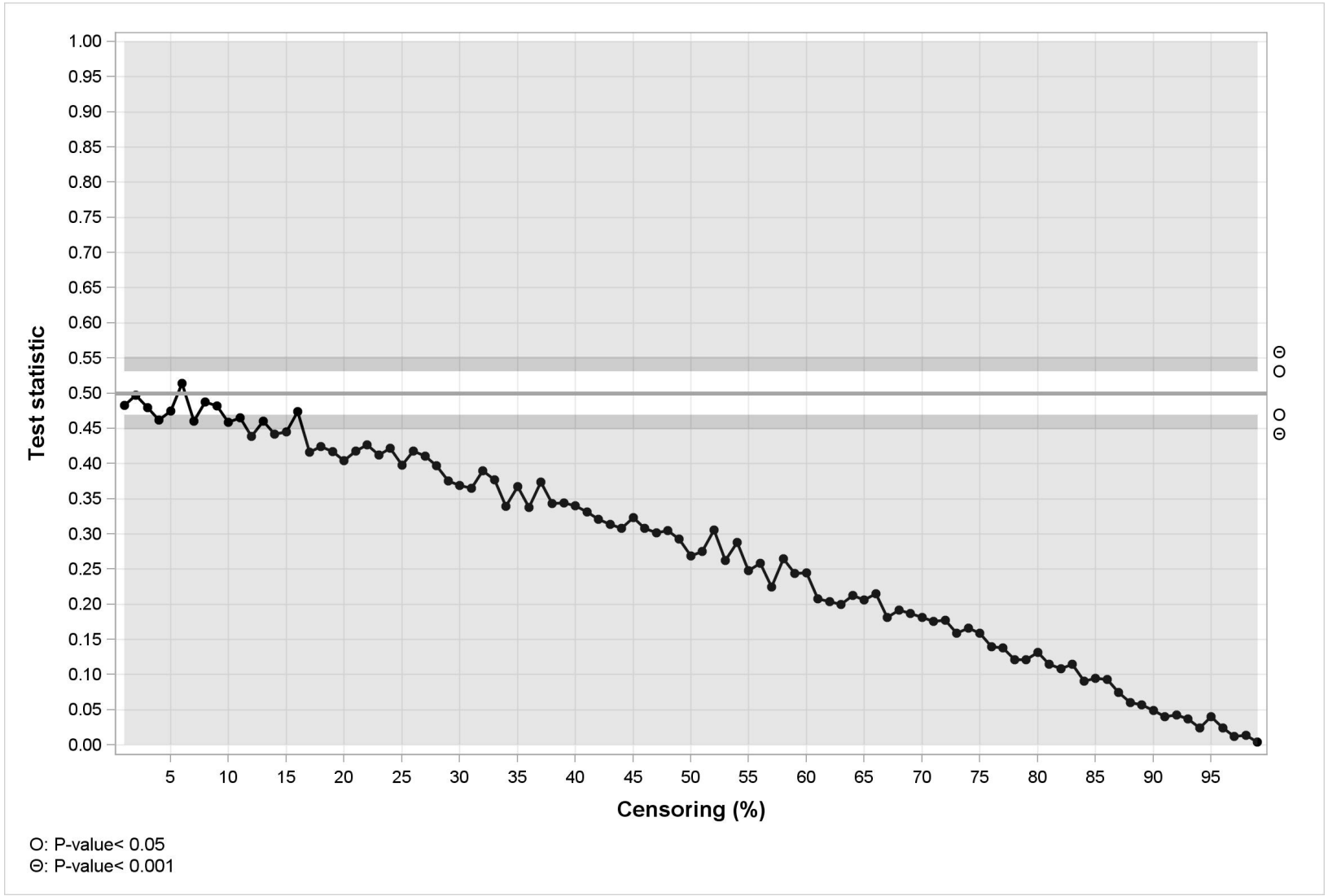


Figure 5.7: Mean ICC estimate for case 2: $\rho = 0.01$, $m = 200$, $k = 40$

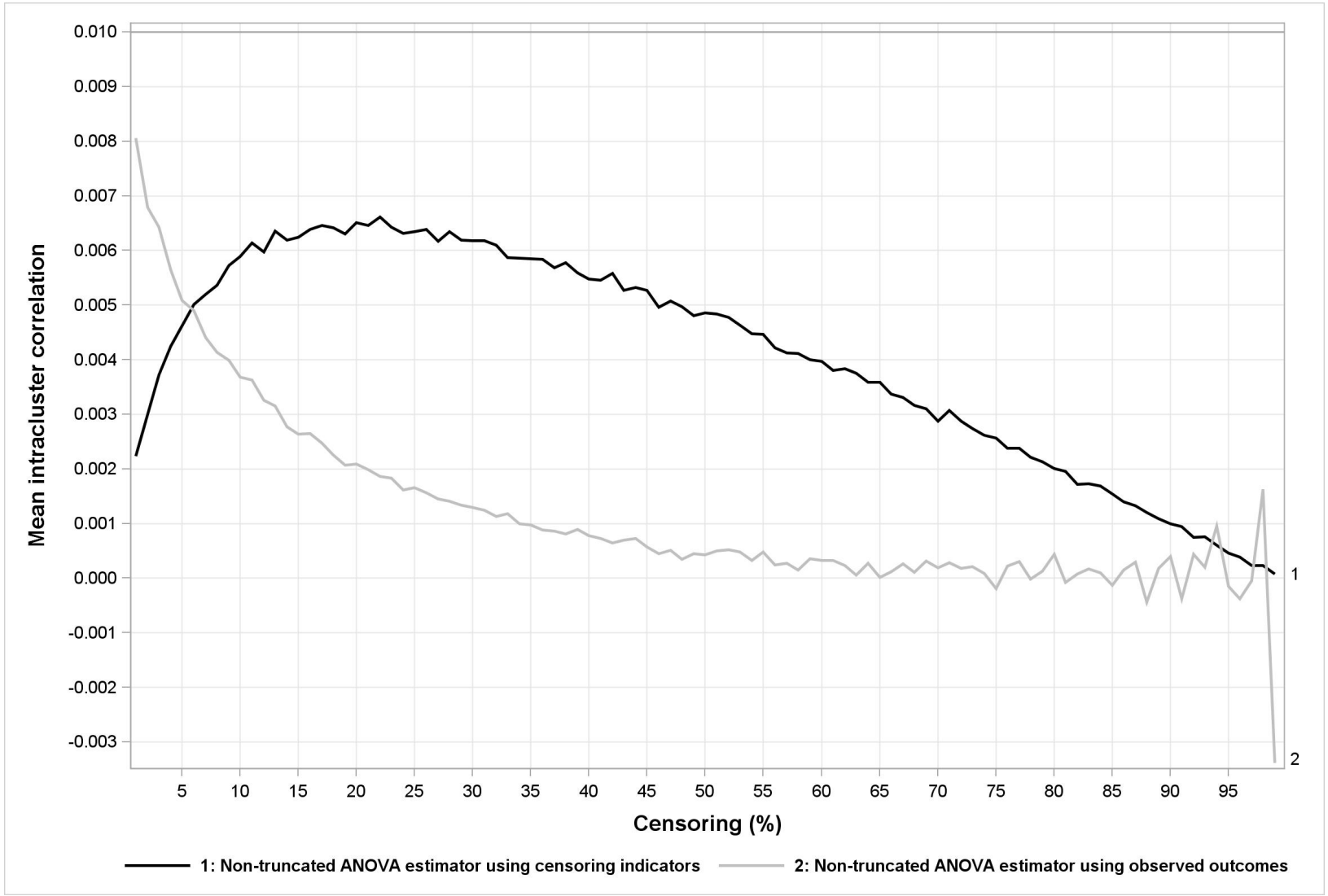


Figure 5.8: Variance of ICC estimators for case 2: $\rho = 0.01$, $m = 200$, $k = 40$

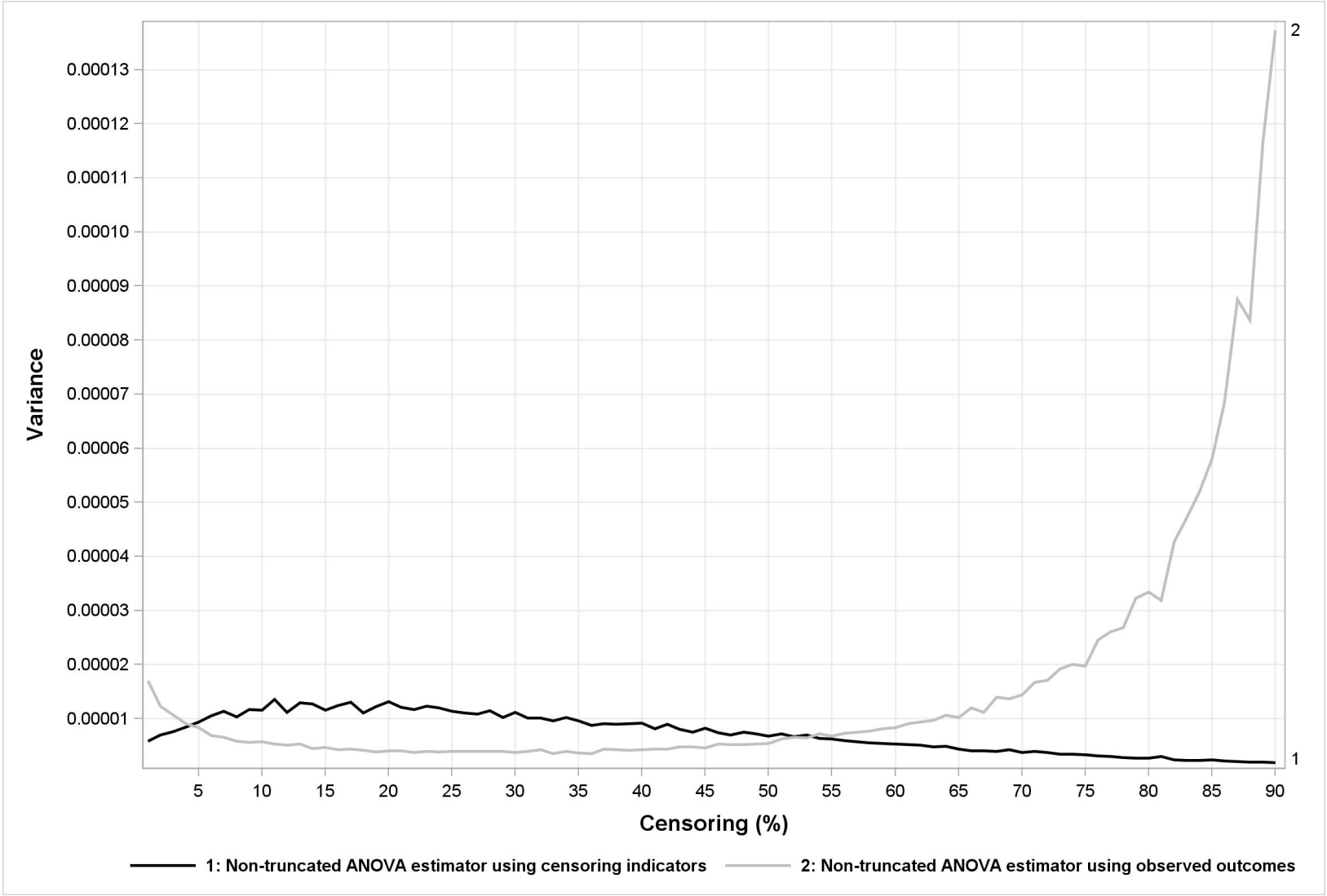


Figure 5.9: Mean square error of ICC estimators for case 2: $\rho = 0.01$, $m = 200$, $k = 40$

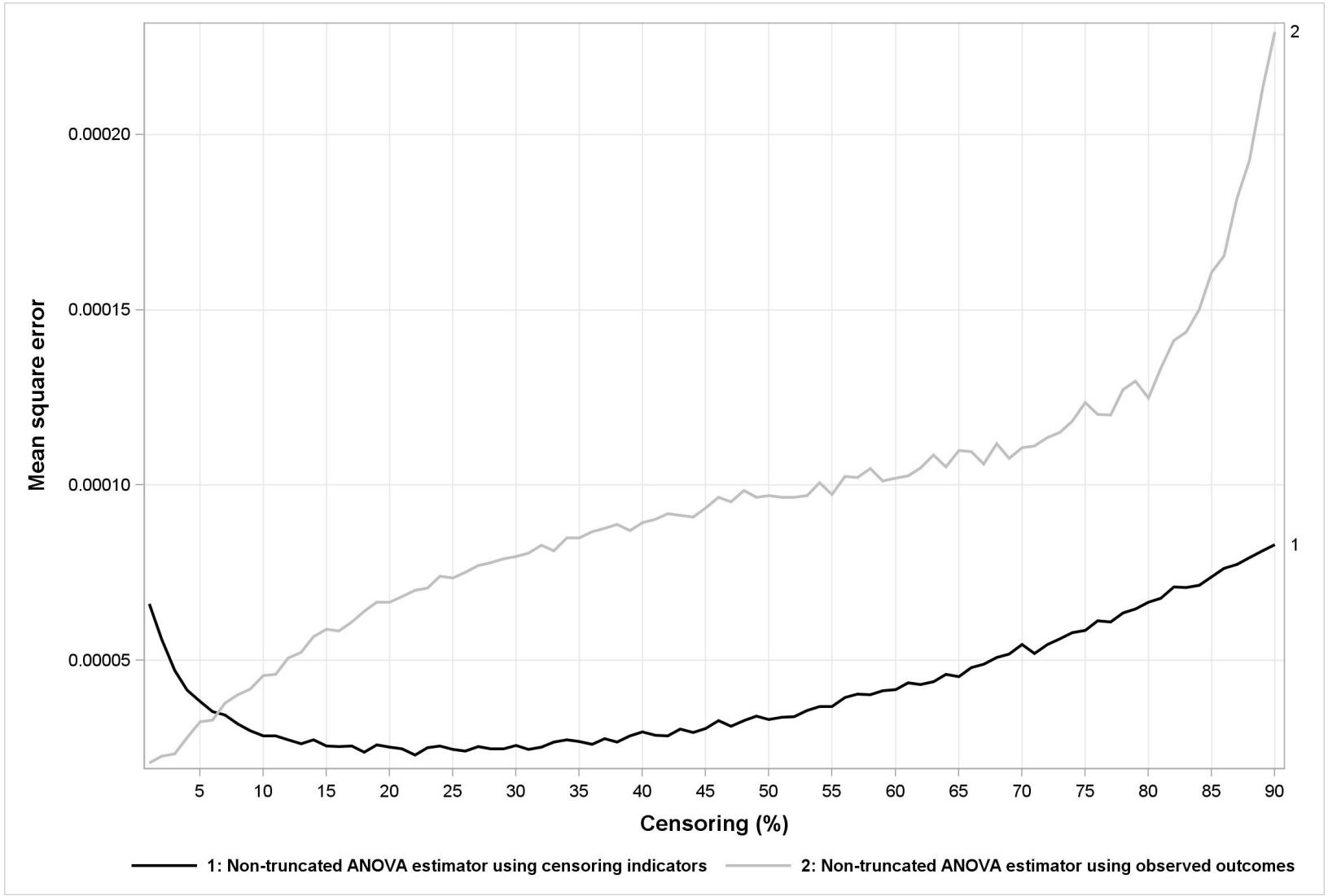


Figure 5.10: Mean between-cluster variance component for case 2: $\rho = 0.01$, $m = 200$, $k = 40$

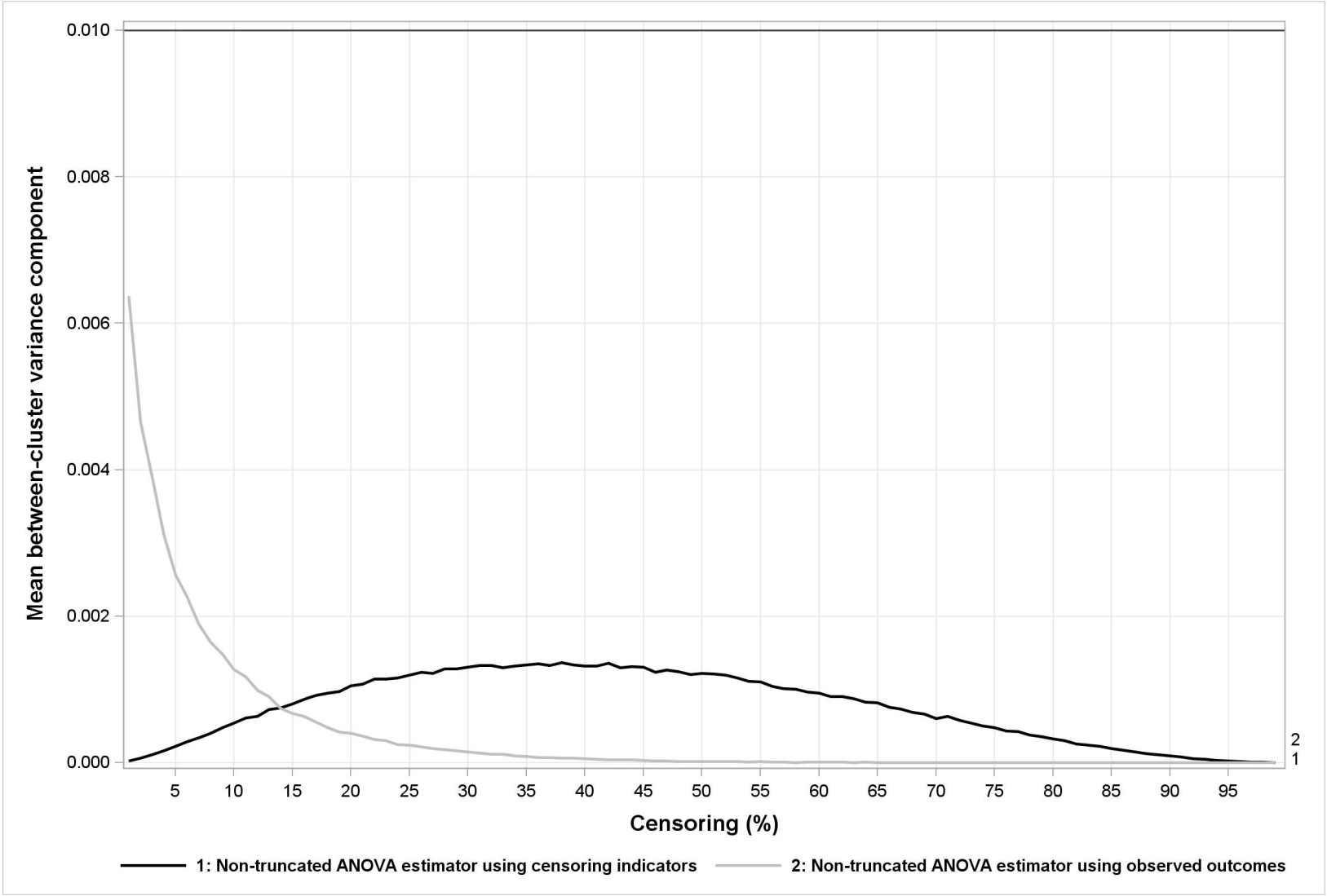


Figure 5.11: Mean within-cluster variance component for case 2: $\rho = 0.01$, $m = 200$, $k = 40$

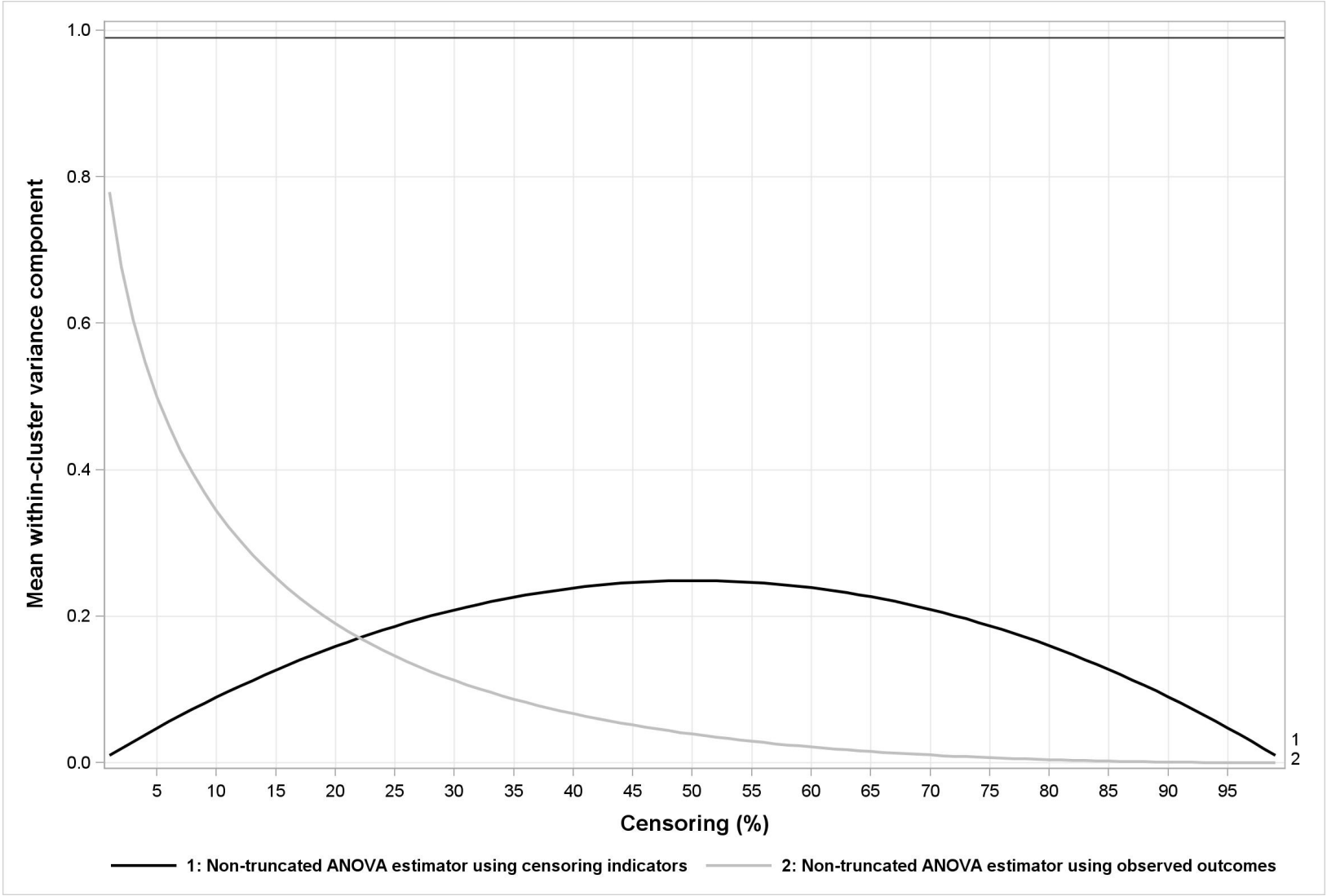
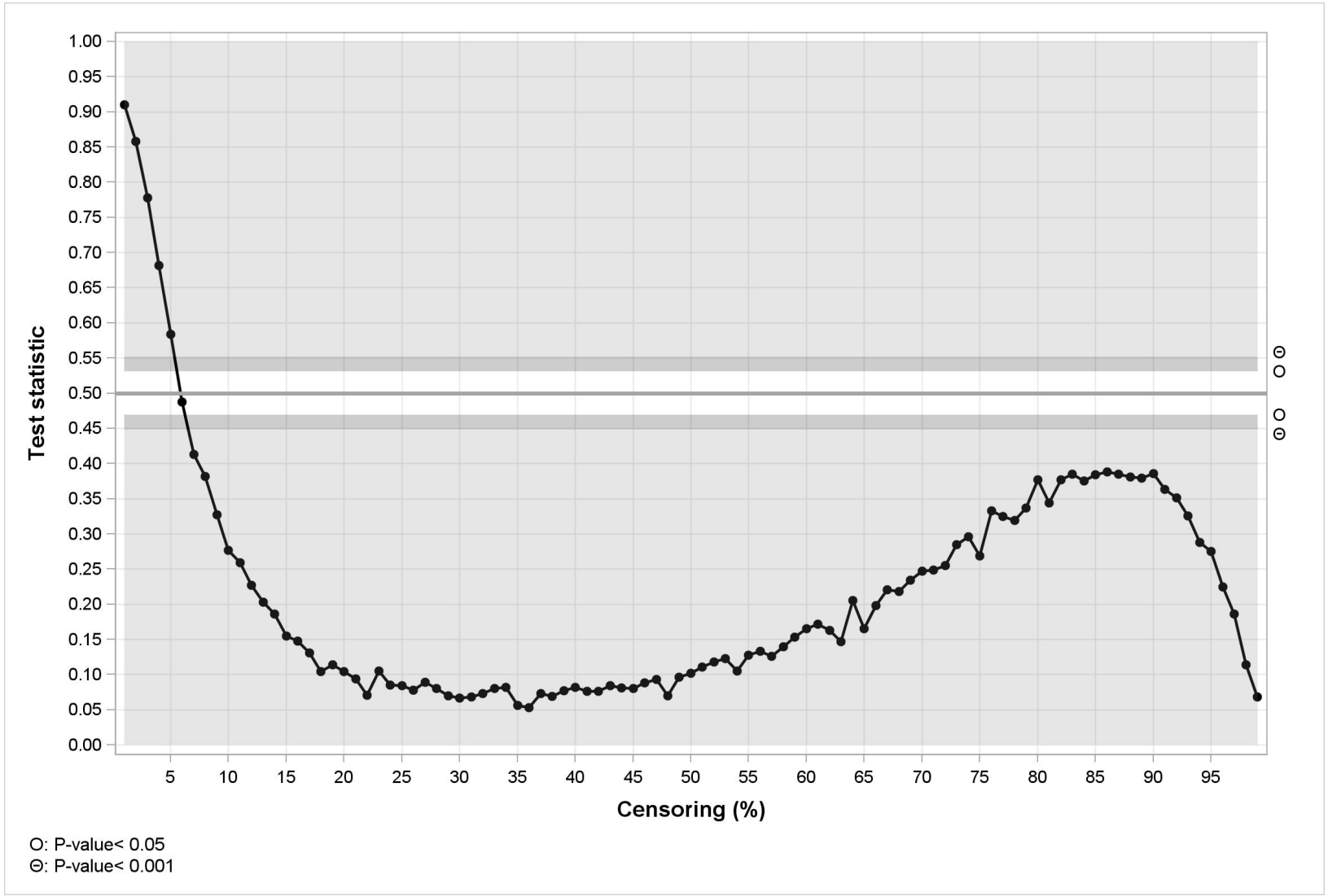


Figure 5.12: Sign test for case 2: $\rho = 0.01$, $m = 200$, $k = 40$



5.3.2 Bivariate time-to-event outcomes ($m = 2$)

The non-truncated ICC estimators using censoring indicators ($\widehat{\rho}_c$) and observed outcomes ($\widehat{\rho}_t$) are negatively biased when $\rho = 0.01$ (see Figure 5.19). However, $\widehat{\rho}_t$ and $\widehat{\rho}_c$ are symmetrically distributed around zero when $\rho = 0$ (see Figure 5.13). In general, the truncated ANOVA estimators $\{\widehat{\rho}_c^*,\widehat{\rho}_t^*,\widehat{\rho}_{t\sim}^*\}$ have greater bias but lower variance and MSE as compared to their non-truncated counterparts $\{\widehat{\rho}_c,\widehat{\rho}_t,\widehat{\rho}_{t\sim}\}$. The positive bias of truncated estimators generally increases with the increasing frequency of administrative censoring. The omission of singletons further increases the amount of negative bias as compared to $\widehat{\rho}_t$. The variance of $\widehat{\rho}_t,\widehat{\rho}_t^*,\widehat{\rho}_{t\sim},\widehat{\rho}_{t\sim}^*$ increases with the increasing frequency of administrative censoring (see Figure 5.14 and 5.20). However, the variance of $\widehat{\rho}_c$ and $\widehat{\rho}_c^*$ remains fairly constant with respect to administrative censoring. The MSE of the ICC estimators using the observed outcomes increases exponentially under high rate of administrative censoring while the MSE of ICC estimators derived from censoring indicators remained fairly constant (see Figure 5.15 and 5.21). The results of variance components and sign test are similar to the cases where $m = 200$.

Table 5.3: The summary of mean bias when $m = 2$

			Mean bias		
ρ	Source	ICC estimator	Minimum	Median	Maximum
0	Censoring indicators	$\widehat{\rho}_c$	-0.01	-0.00003	0.01
0		$\widehat{\rho}_c^*$	0.03	0.06	0.07
0	Observed outcomes	$\widehat{\rho}_t$	-0.02	0.001	0.03
0		$\widehat{\rho}_t^*$	0.06	0.1	0.3
0		$\widehat{\rho}_{t\sim}$	-0.05	-0.002	0.02
0		$\widehat{\rho}_{t\sim}^*$	0.06	0.1	0.2
0.01	Censoring indicators	$\widehat{\rho}_c$	**-0.02** ‡	**-0.006** ‡	**0.01** ‡
0.01		$\widehat{\rho}_c^*$	**0.008** ‡	**0.05** ‡	**0.07** ‡
0.01	Observed outcomes	$\widehat{\rho}_t$	**-0.04** ‡	**-0.007** ‡	**0.03** ‡
0.01		$\widehat{\rho}_t^*$	**0.06** ‡	**0.1** ‡	**0.2** ‡
0.01		$\widehat{\rho}_{t\sim}$	**-0.09** ‡	**-0.01** ‡	**0.005** ‡
0.01		$\widehat{\rho}_{t\sim}^*$	**0.05** ‡	**0.09** ‡	**0.2** ‡

‡ violates the criteria of 10% relative bias

Figure 5.13: Mean ICC estimate for case 3: $\rho = 0$, $m = 2$, $k = 40$

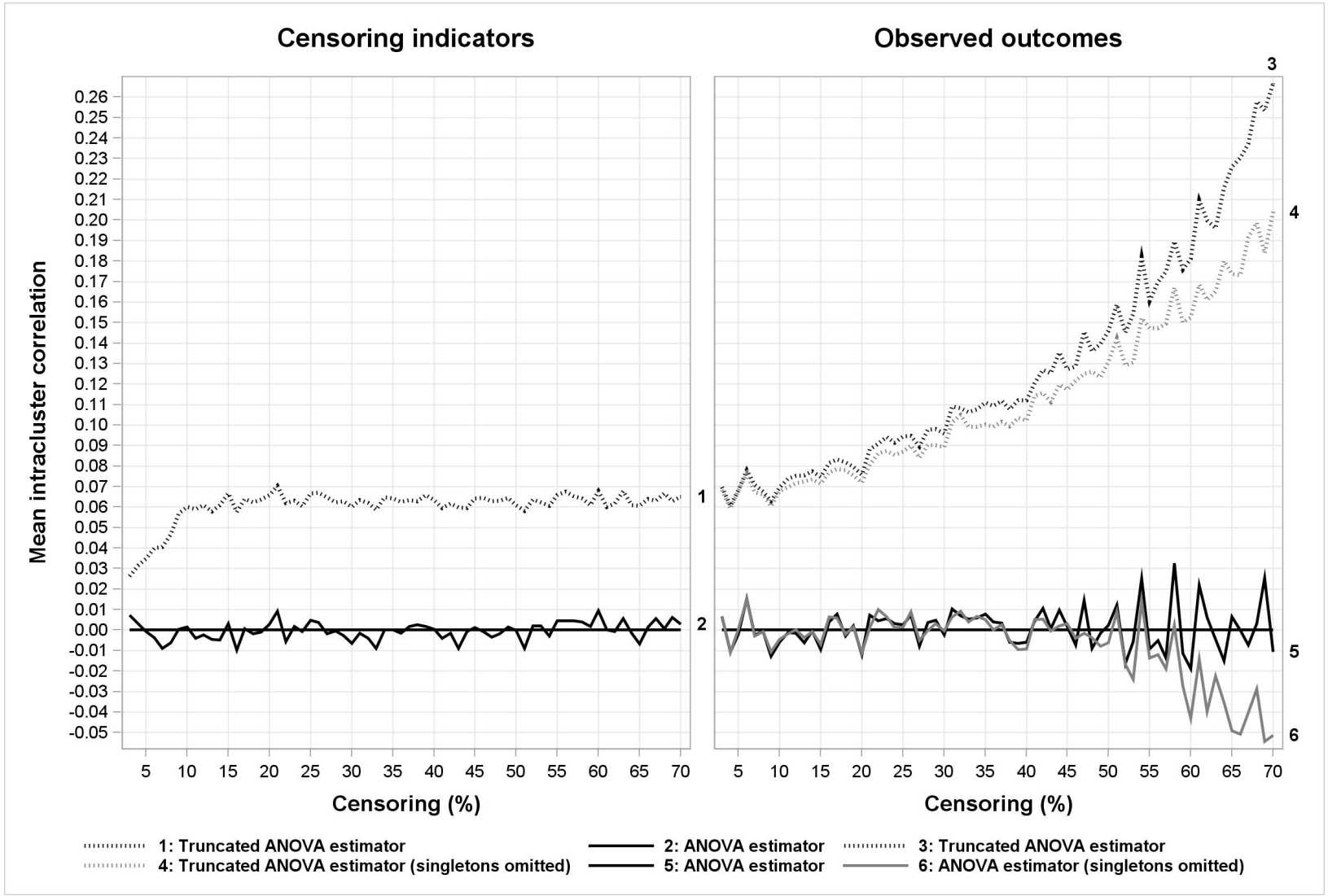


Figure 5.14: Variance of ICC estimators for case 3: $\rho = 0$, $m = 2$, $k = 40$

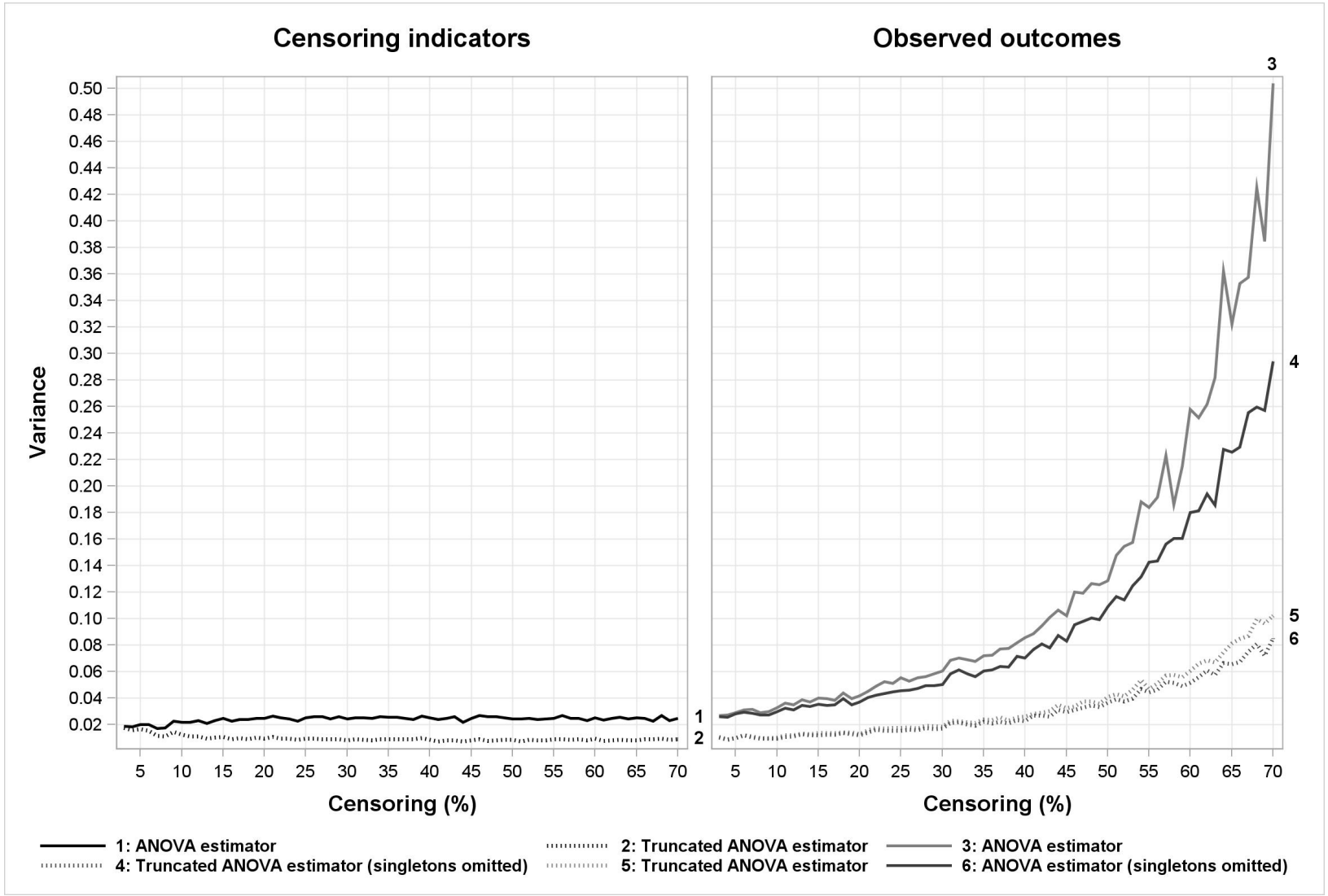


Figure 5.15: Mean square error of ICC estimators for case 3: $\rho = 0$, $m = 2$, $k = 40$

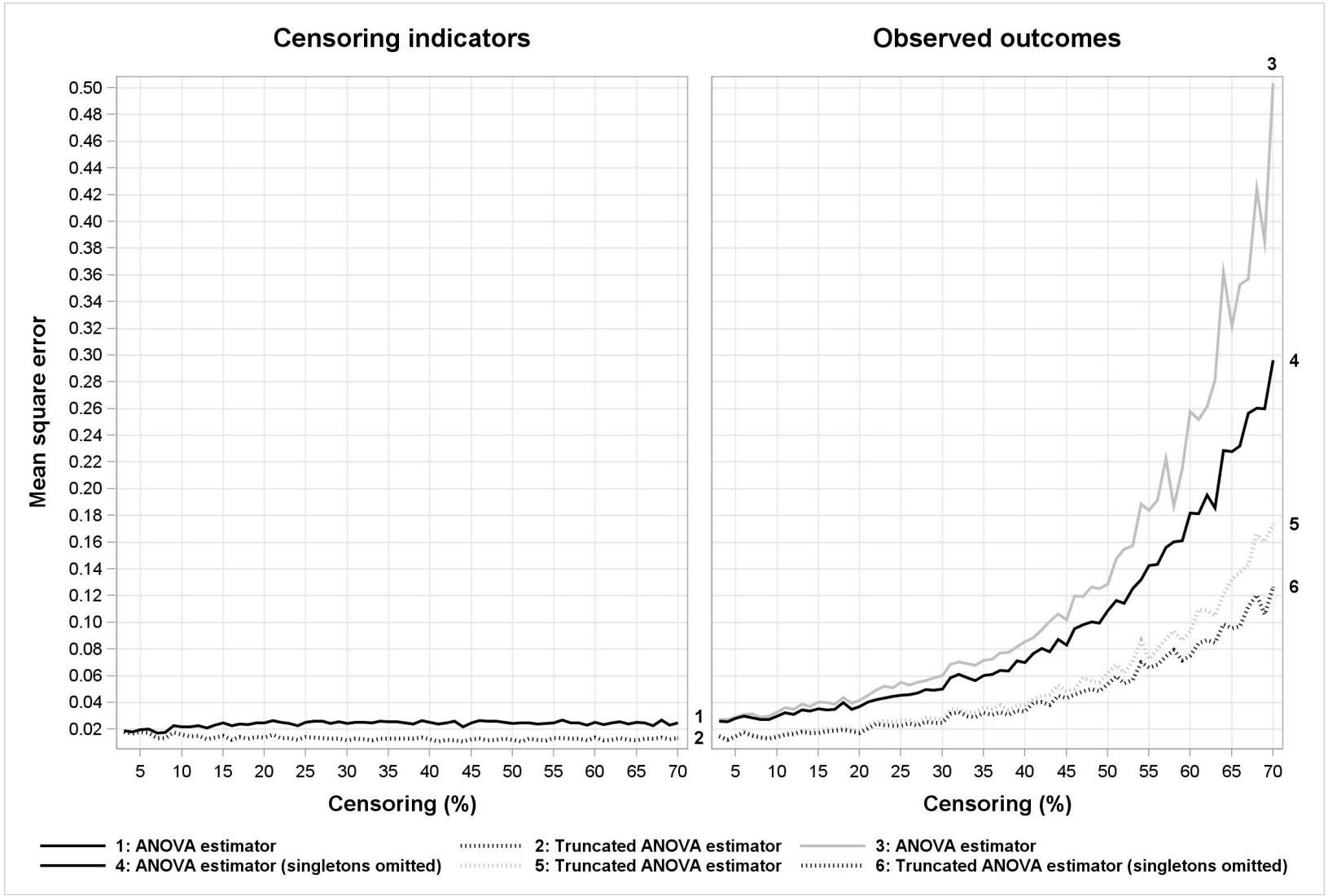


Figure 5.16: Mean between-cluster variance component for case 3: $\rho = 0$, $m = 2$, $k = 40$

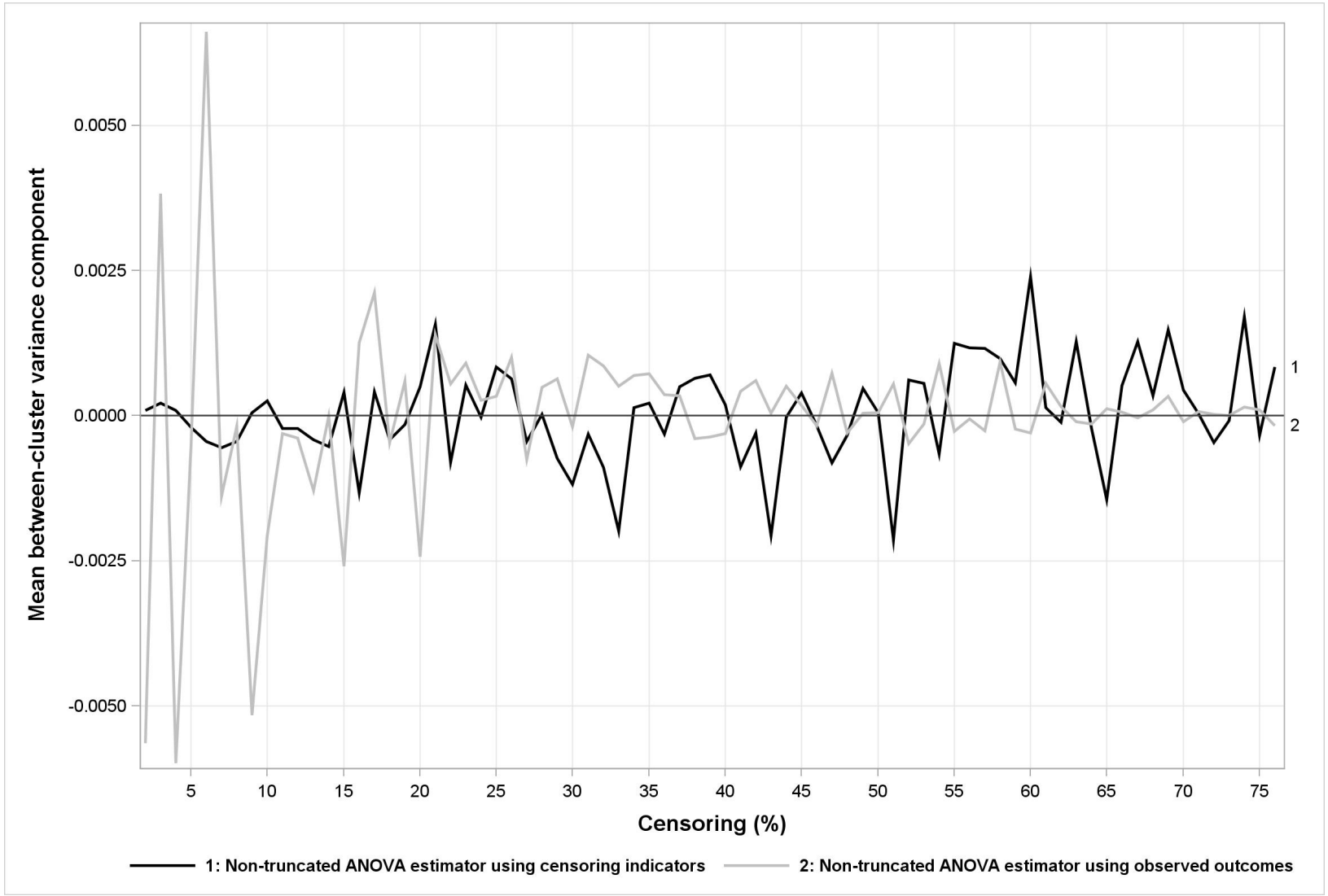


Figure 5.17: Mean within-cluster variance component for case 3: $\rho = 0$, $m = 2$, $k = 40$

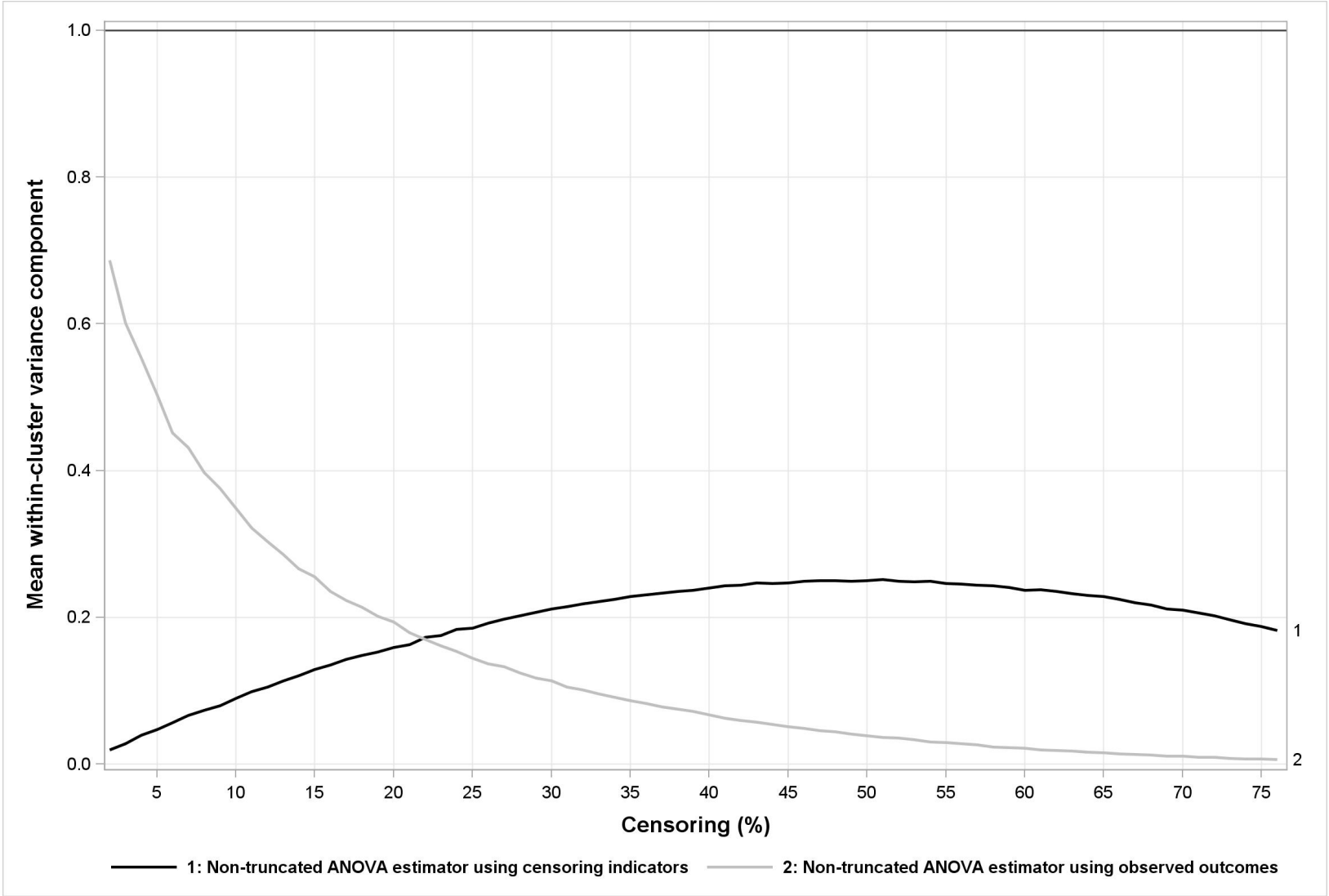


Figure 5.18: Sign test for case 3: $\rho = 0$, $m = 2$, $k = 40$

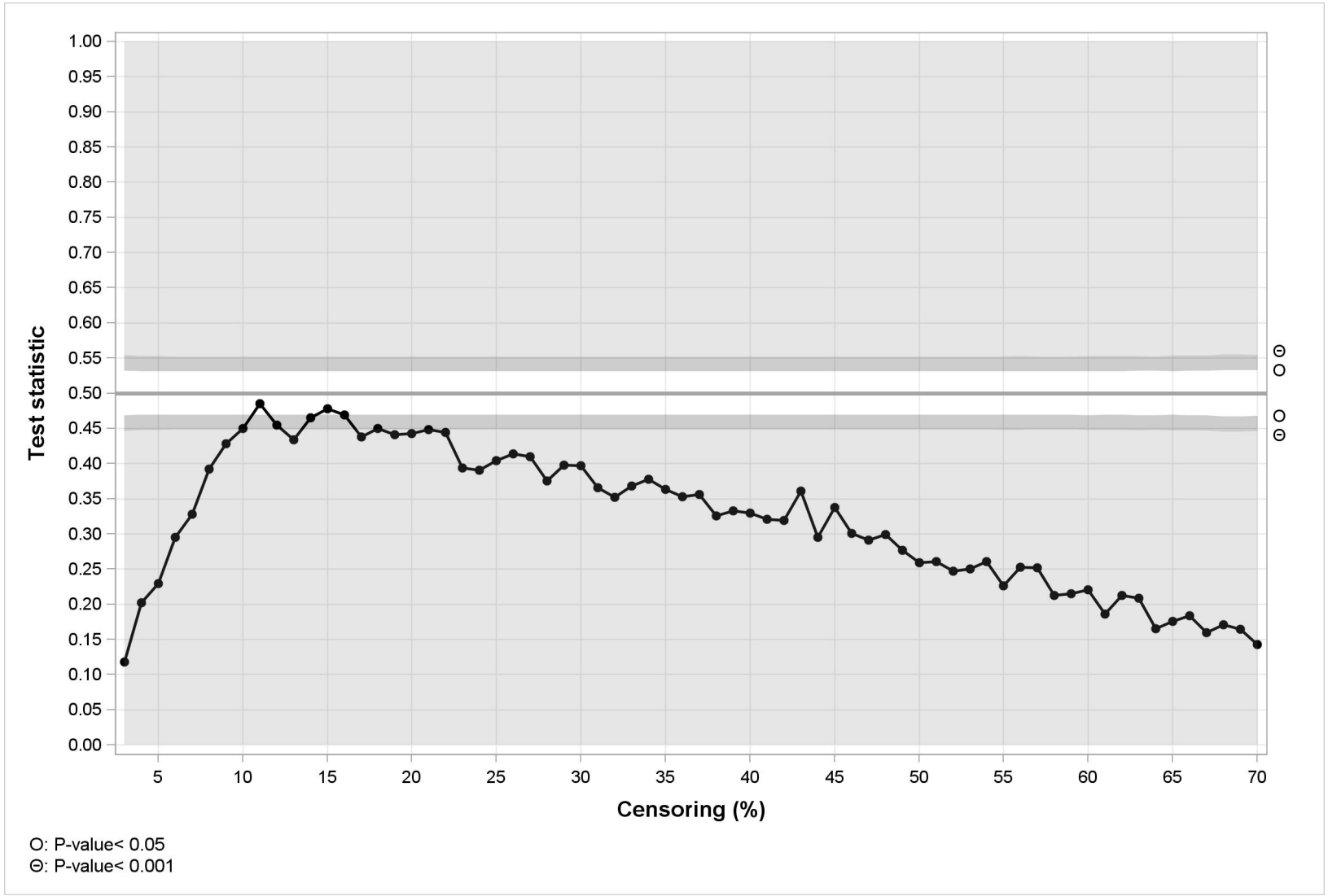


Figure 5.19: Mean ICC estimate for case 4: $\rho = 0.01$, $m = 2$, $k = 40$

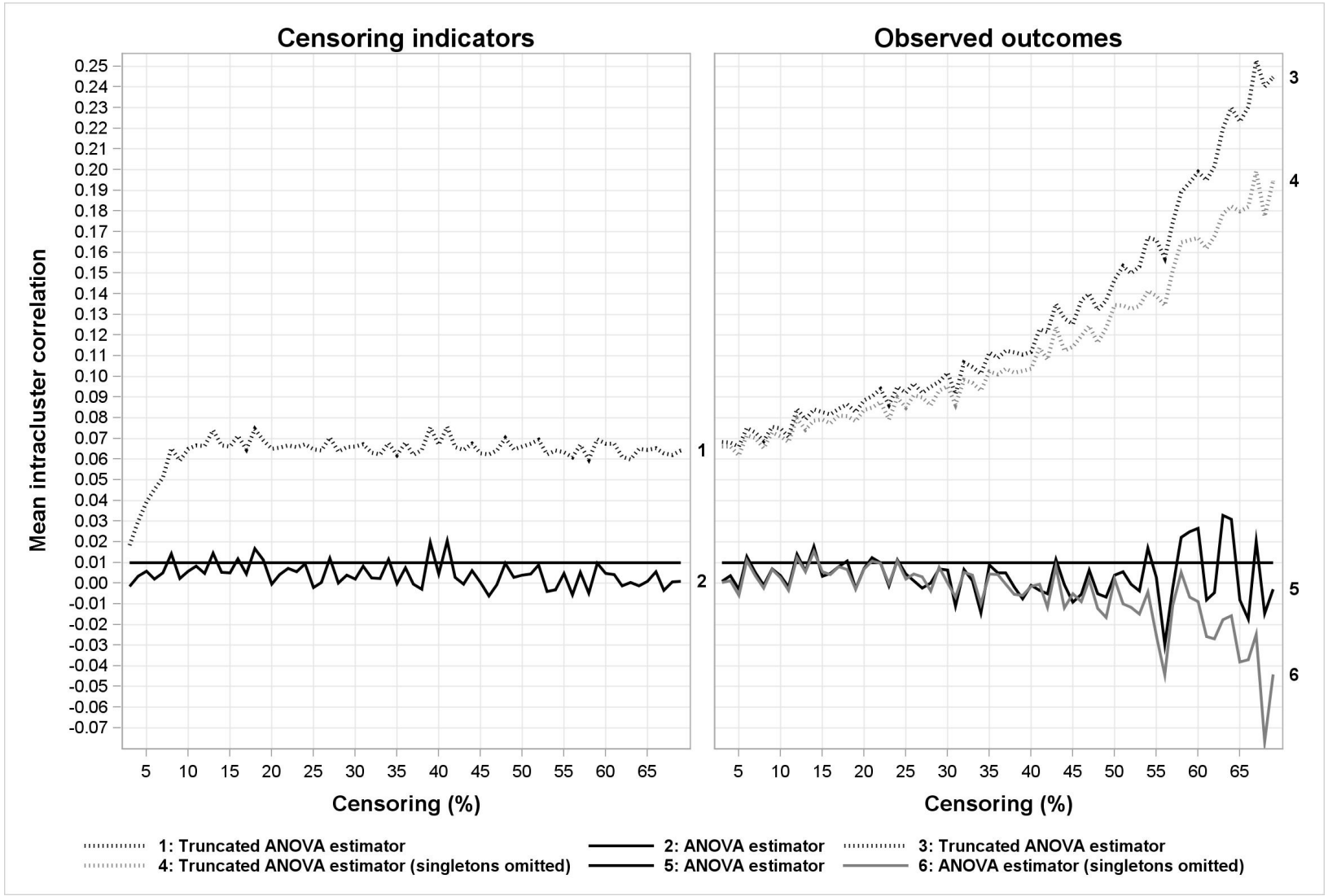


Figure 5.20: Variance of ICC estimators for case 4: $\rho = 0.01$, $m = 2$, $k = 40$

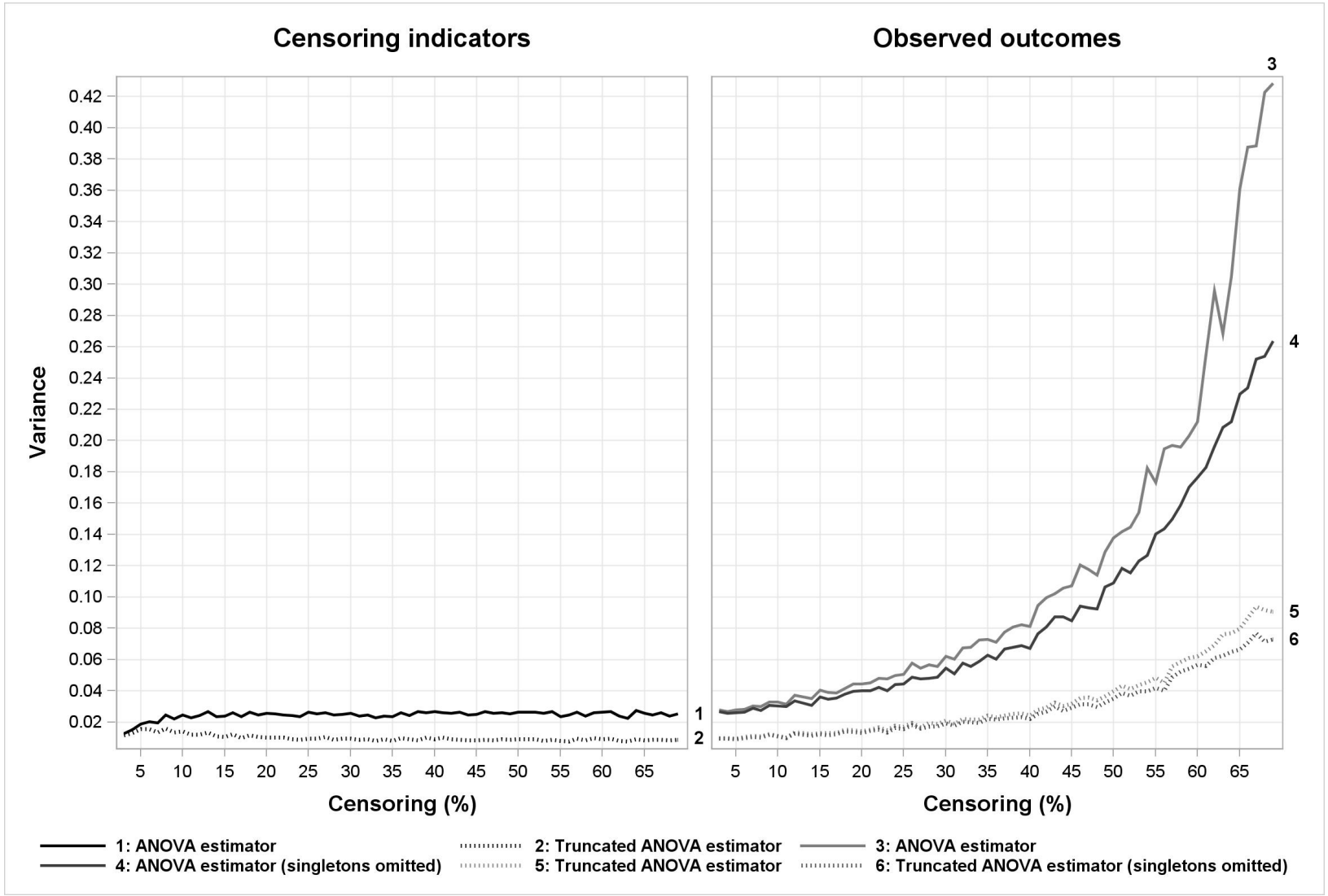


Figure 5.21: Mean square error of ICC estimators for case 4: $\rho = 0.01$, $m = 2$, $k = 40$

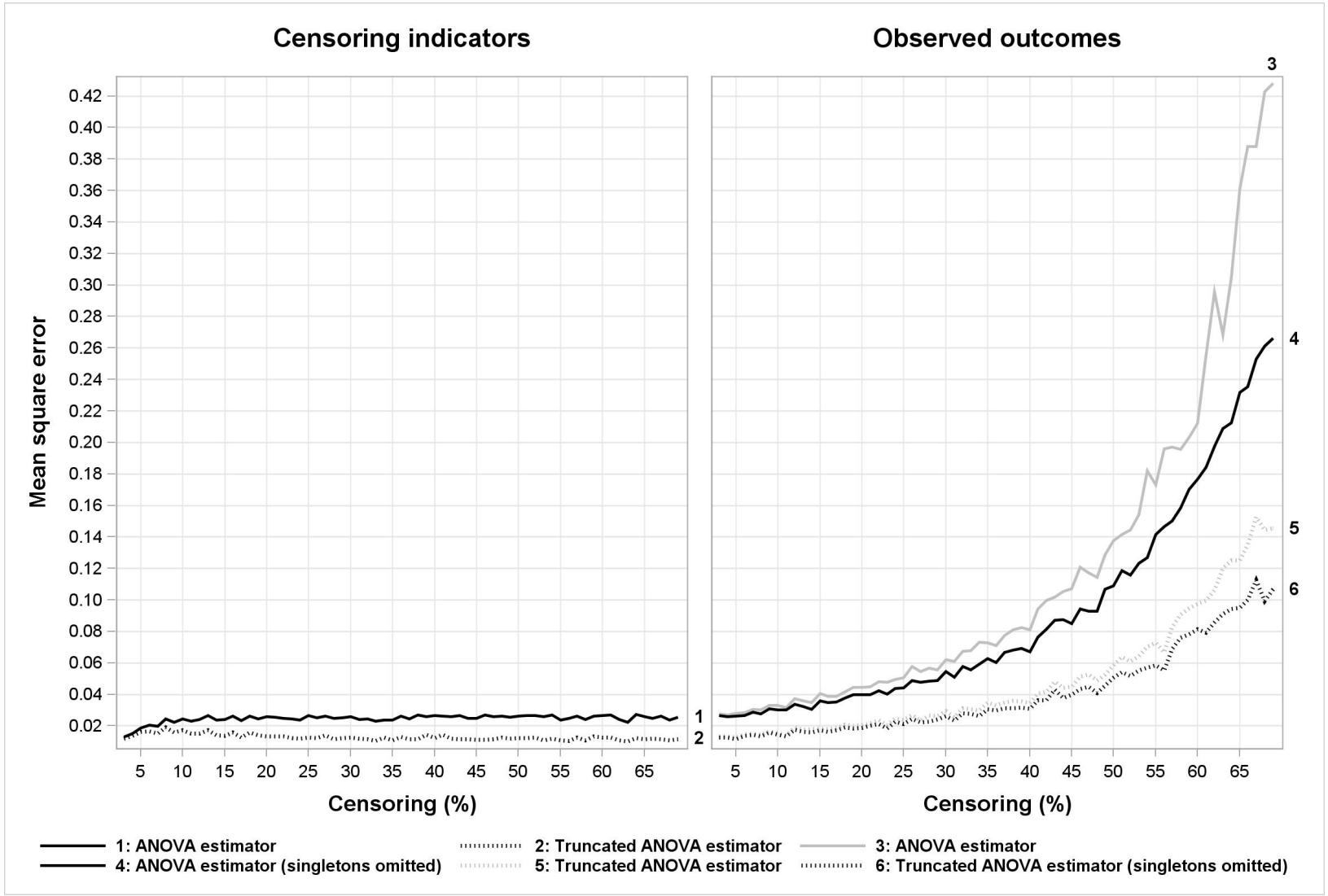


Figure 5.22: Mean between-cluster variance component for case 4: $\rho = 0.01$, $m = 2$, $k = 40$

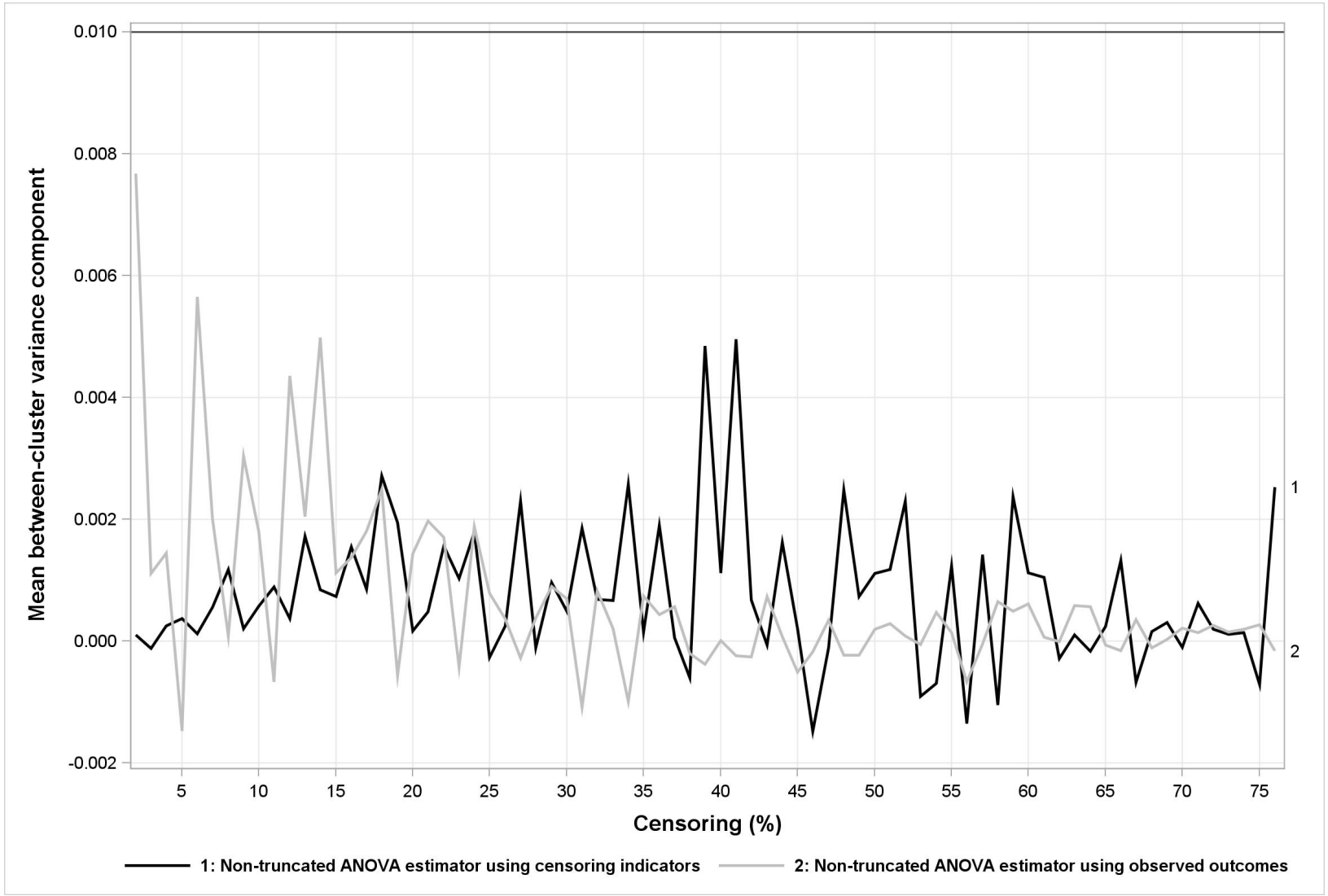


Figure 5.23: Mean within-cluster variance component for case 4: $\rho = 0.01$, $m = 2$, $k = 40$

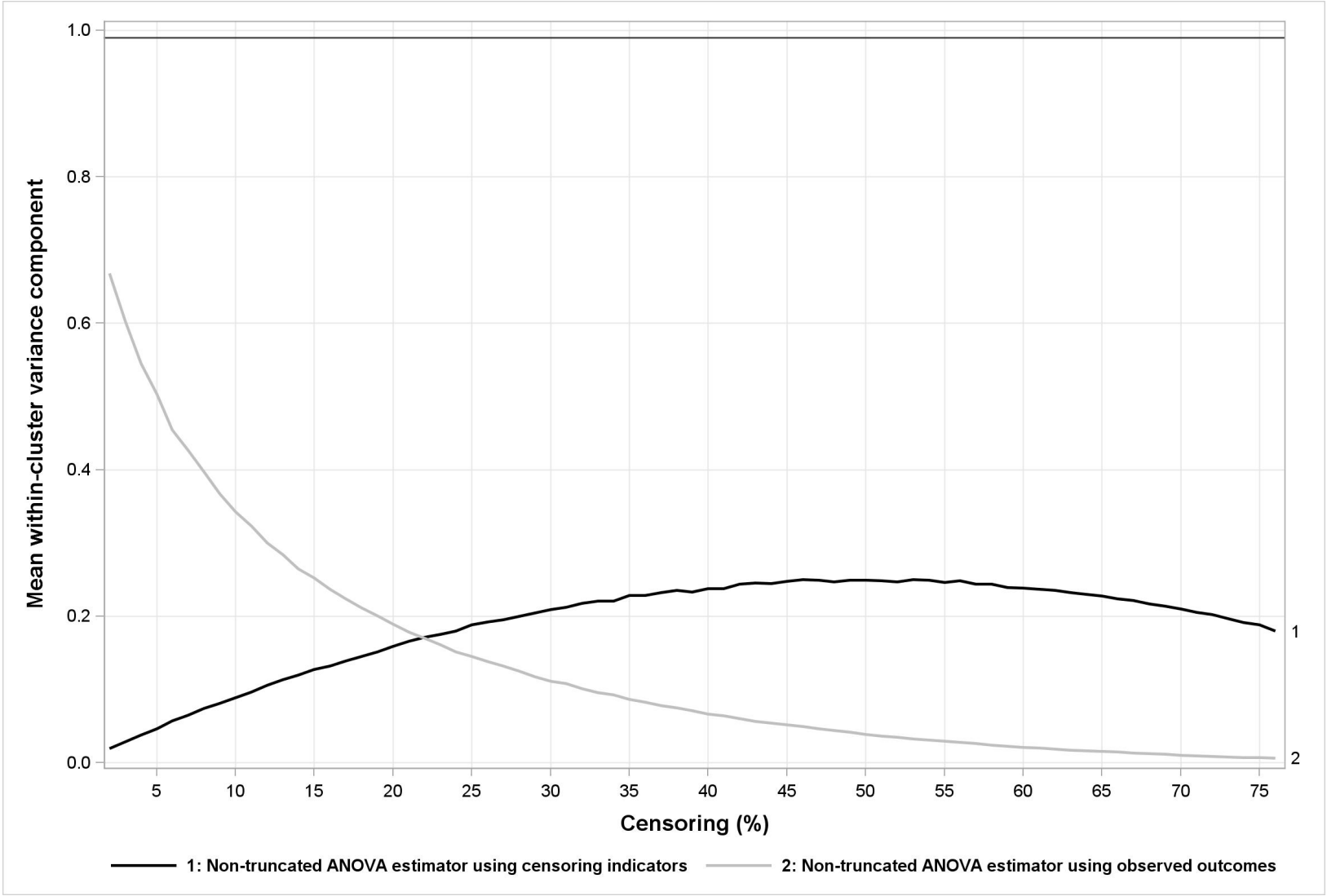
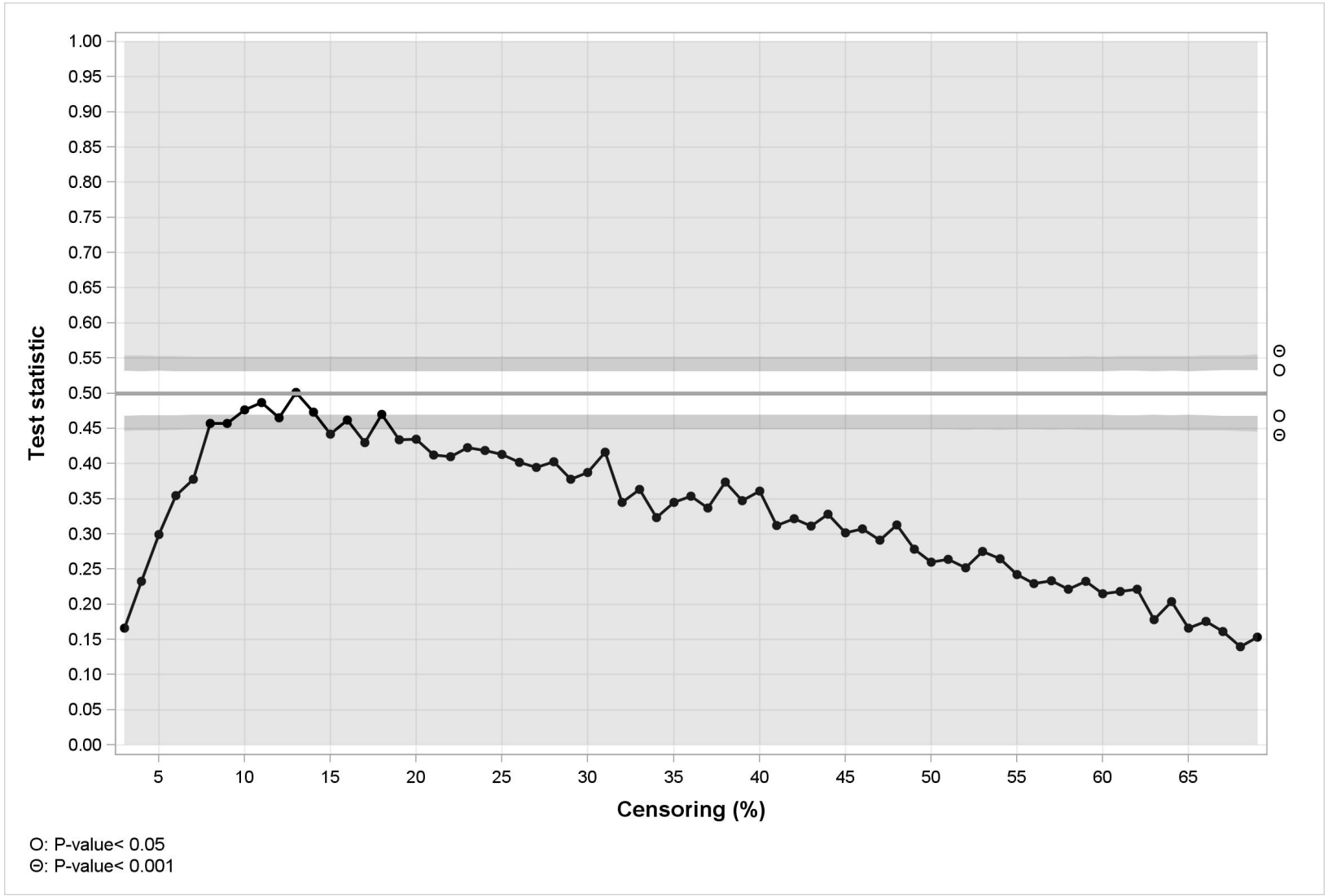


Figure 5.24: Sign test for case 4: $\rho = 0.01$, $m = 2$, $k = 40$



5.4 Exploratory results

The results of this simulation study are confirmed with the analytic work shown in chapter 3. In addition, the correlation between the non-truncated ICC estimators using censoring indicators and observed outcomes is assessed.

5.4.1 Agreement with analytic work

The mean of non-truncated ICC estimator derived from censoring indicators ($\overline{\widehat{\rho_c}}$) is compared with its analytic expression (ρ_c) over the entire range of administrative censoring. Figure 5.25 (a)-(d) and Table 5.5 summarizes the results of measuring the disparity between the $\overline{\widehat{\rho_c}}$ and ρ_c for the four cases considered in this simulation study.

Table 5.4: Summary of disparity ($=\overline{\widehat{\rho_c}}-\rho_c$) for the four cases.

	Parameters			Disparity ($=\overline{\widehat{\rho_c}}-\rho_c$)		
Case	ρ	m	k	Minimum	Median	Maximum
1	0	200	40	-0.00007	0.000007	0.0001
2	0.01	200	40	-0.0002	-.00001	0.0002
3	0	2	40	-0.01	0.00003	0.01
4	0.01	2	40	-0.01	-0.001	0.02

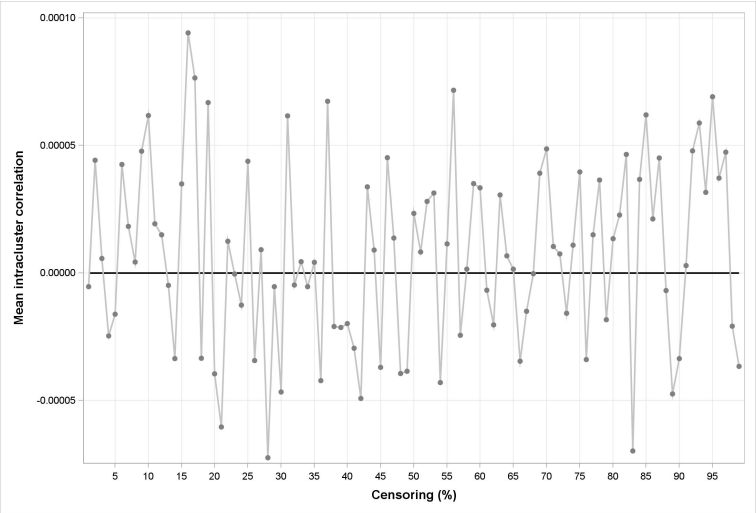
5.4.2 Corr($\widehat{\rho_c}$, $\widehat{\rho_t}$)

The correlation between non-truncated ICC estimator using censoring indicators ($\widehat{\rho_c}$) and observed outcomes ($\widehat{\rho_t}$) is assessed for the 396 parameter combinations of this simulation study. Figure 5.26 (a)-(d) and Table 5.5 summarizes the correlation between $\widehat{\rho_c}$ and $\widehat{\rho_t}$.

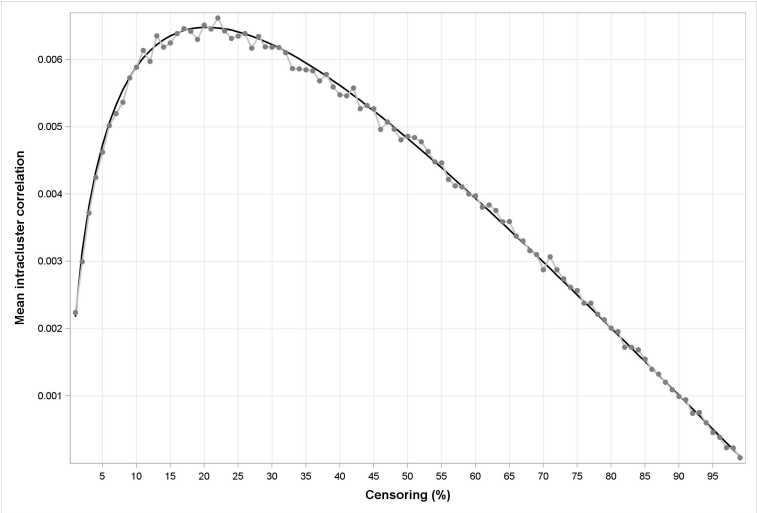
Table 5.5: Summary of Corr($\widehat{\rho_c}$, $\widehat{\rho_t}$)

	Parameters			$Corr(\widehat{\rho_c},\widehat{\rho_t})$		
Case	ρ	m	k	Minimum	Median	Maximum
1	0	200	40	-0.08	-0.007	0.09
2	0.01	200	40	-0.06	0.06	0.48
3	0	2	40	-0.2	0.0002	0.2
4	0.01	2	40	-0.2	-0.007	0.07

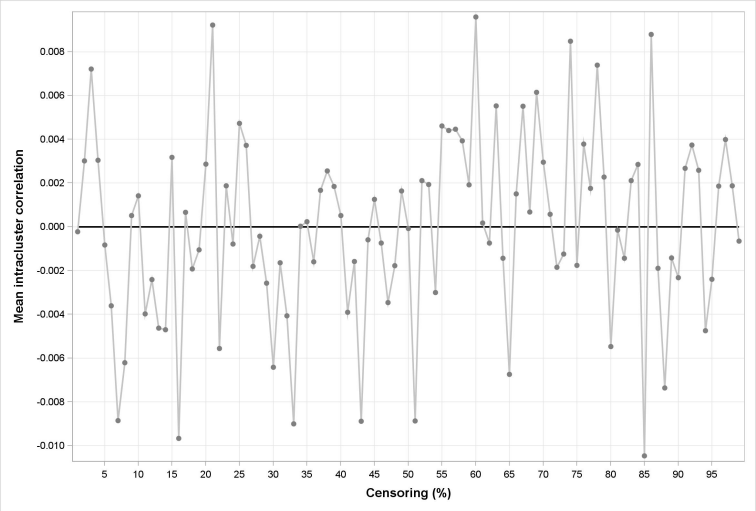
Figure 5.25: The mean of non-truncated ICC estimator (solid gray line) and ICC parameter (solid black line) derived from censoring indicators



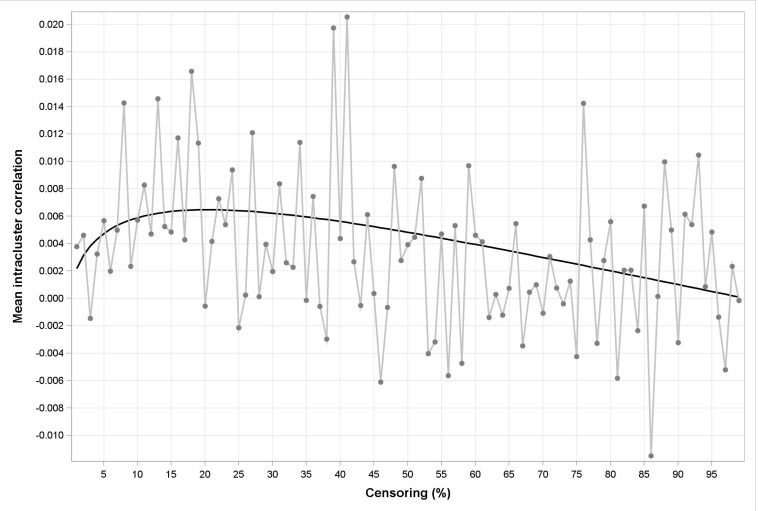
(a) Case 1: $\rho = 0$, $m = 200$, $k = 40$



(b) Case 2: $\rho = 0.01$, $m = 200$, $k = 40$

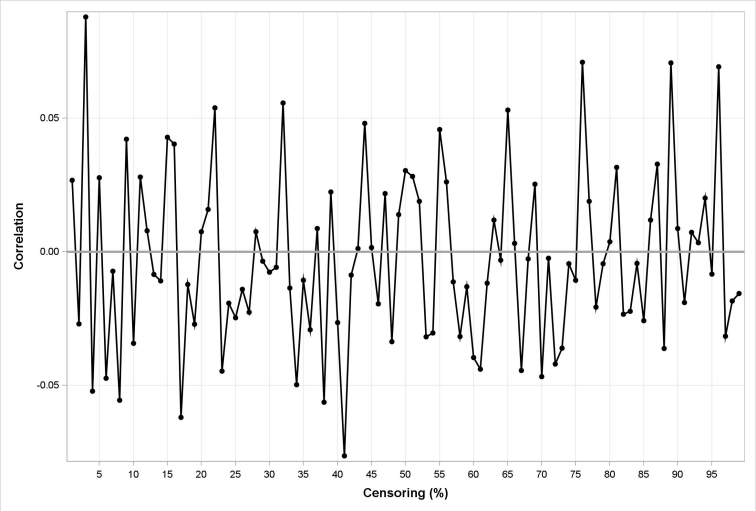


(c) Case 3: $\rho = 0$, $m = 2$, $k = 40$

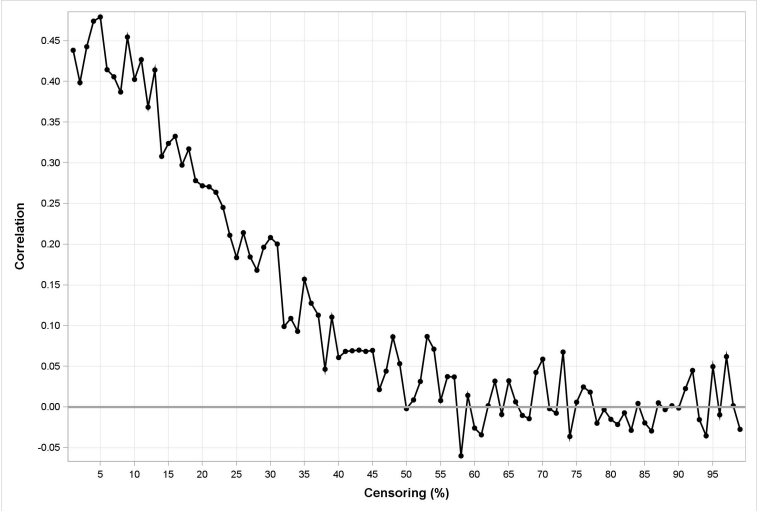


(d) Case 4: $\rho = 0.01$, $m = 2$, $k = 40$

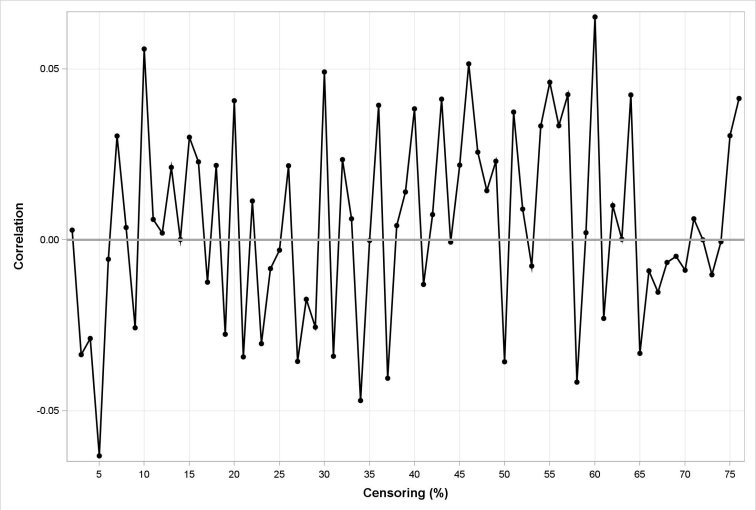
Figure 5.26: The correlation between non-truncated ICC estimators using censoring indicators and observed outcomes



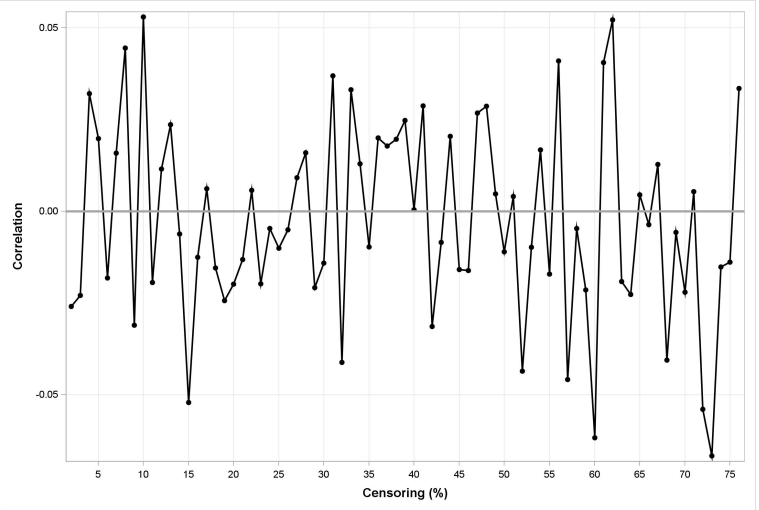
(a) Case 1: $\rho = 0$, $m = 200$, $k = 40$



(b) Case 2: $\rho = 0.01$, $m = 200$, $k = 40$



(c) Case 3: $\rho = 0$, $m = 2$, $k = 40$



(d) Case 4: $\rho = 0.01$, $m = 2$, $k = 40$

5.5 Summary

The results of this simulation study showed that the non-truncated ANOVA estimator of ICC derived from censoring indicators ($\widehat{\rho_c}$) and observed outcomes ($\widehat{\rho_t}$) are both negatively biased when $\rho \neq 0$. The negative bias is also evident for between cluster and within cluster variance components of $\widehat{\rho_c}$ and $\widehat{\rho_t}$. In addition, the sign test shows that $\widehat{\rho_t}$ estimates the ICC parameter more closely than $\widehat{\rho_c}$ under minimal frequency of administrative censoring when $\rho \neq 0$. However, $\widehat{\rho_c}$ estimates the ICC parameter more closely than $\widehat{\rho_t}$ under moderate to high frequency of administrative censoring. The analytic work in chapter 3 supports the results of this simulation study. The correlation between $\widehat{\rho_t}$ and $\widehat{\rho_c}$ is negligible when $\rho = 0$ or $m = 2$. However, substantial amount of correlation is observed when $\rho \neq 0$ and $m = 200$.

Chapter 6

Examples

This chapter describes the analyses of two CRTs with time-to-event outcomes. The first trial involved the randomization of large clusters of variable size while the second trial involved the randomization of small clusters each with a fixed number of observations. Background information and descriptive analysis are provided for each CRT in the first two sections. The purpose of this chapter is to estimate the ICC using the methods described in chapter 4. The estimation of ICC is carried out using observed outcomes and censoring indicators. Finally, section 6.3 provides some concluding remarks comparing ICC estimated using data from the two CRTs.

6.1 Malaria induced childhood mortality

Malaria is a mosquito-borne infectious disease caused by the protozoan belonging to the genus *Plasmodium* (Rodrigues et al., 2008). This infection is most common at the end of the rainy season when mosquitoes are breeding. According to the World Health Organization (2014), 90% of all global deaths due to malaria occur in Africa. Furthermore, children under the age of 5 account for 78% of all malaria deaths in Africa. Several studies have shown that transmission of malaria can be reduced by having children sleep under mosquito nets (Abdulla

et al., 2001; Binka et al., 1996).

The purpose of this trial was to assess the impact of insecticide inpregnated bednets on child mortality in relation to the control group (Binka et al., 1996). This trial was conducted in Ghana (Africa) from July 1993 to June 1995. The rural population in the study area lived in dispersed settlements of compounds which were composed of related individuals. Before the start of this trial, all members within the study area were enumerated and relevant information about their socioeconomic status was recorded. An open lottery system was used to randomly provide 48 of the 96 clusters (defined as geographic areas) with mosquito nets which were impregnated with permethrin every 6 months. During the study follow-up, each compound was visited every three months to record information on birth, migration and death. In addition, a field supervisor visited randomly selected clusters to evaluate compliance in the use of bednets. The primary outcome of child mortality due to malaria was ascertained by three independent physicians. Each death was classified to a particular category if at least 2 of the 3 physicians agreed; otherwise the cause of death was classified as undetermined.

6.1.1 Descriptive analysis

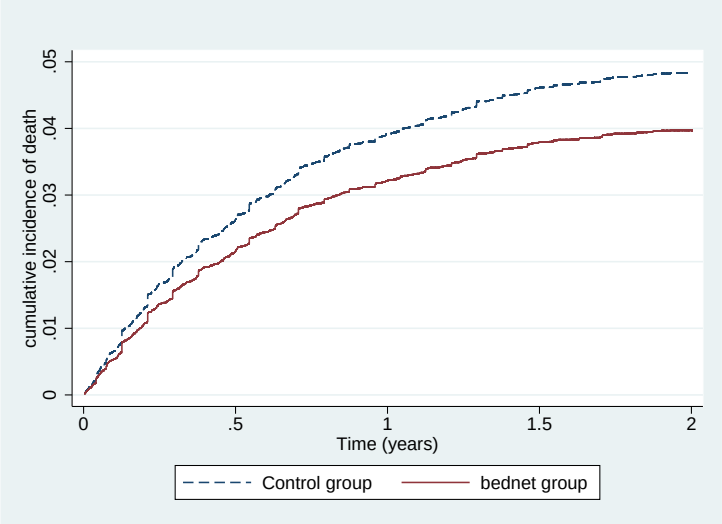
The greatest impact on child mortality was recorded for children under the age of three as they were most vulnerable to malaria (Binka et al., 1996). Hence attention is limited to children under 36 months of age when they started the trial. There were 8,816 children randomly assigned to the permethrin impregnated bednet group while 9,115 children were randomly assigned to the control group.

The minimum and maximum number of children per cluster assigned to the bednet group ranged from 119 to 268 and 102 to 298 in the control group. Heavy censoring was observed in both groups as mortality was rare. Figure 6.1 shows decreased childhood mortality in the bednet group as compared to the control group during the entire follow-up of the study.

Table 6.1: Summary statistics for the control and bednet group.

Summary statistics	Control group	Bednet group
Total children	8816	9115
Minimum cluster size	102	119
Mean cluster size	190	184
Maximum cluster size	298	268
Censoring rate (%)	95.1	96.0

Figure 6.1: Cumulative-incidence of childhood mortality in bednet and control group.



6.1.2 Estimation of the ICC

The ICC is estimated for both intervention groups using the observed time of death and censoring indicators. The ICC estimate in the control group was 0.004 using censoring indicators and -0.022 using observed time of child death. In contrast, the ICC estimate in permethrin impregnated bednet group was 0.001 using censoring indicators and 0.006 using observed time of child death. Furthermore, after the omission of a cluster where only one child death was recorded (cluster 39 in control group), the ICC estimate was -0.024. The negative ICC estimates may be truncated to zero as the ICC parameter is non-negative (Wang et al., 1992).

The ICC estimates in Table 6.2 may lead to different interpretations. For example, the outcome of child death within each cluster is positively correlated using censoring indicators while being negatively correlated using observed outcomes. These inconsistent results may be

explained, in part, due to the lack of correlation between $\widehat{\rho_c}$ and $\widehat{\rho_t}$ for time-to-event outcomes in community trials where high frequency of administrative censoring is imposed. This is shown in chapter 5.

Table 6.2: The ANOVA ICC estimates of control and bednet group.

Source	ICC estimator	Control group	Bednet group
Censoring indicators	$\widehat{\rho_c}$	0.004	0.001
	$\widehat{\rho_c^*}$	0.004	0.001
Observed outcomes	$\widehat{\rho_t}$	-0.022	0.006
	$\widehat{\rho_t^*}$	0	0.006
	$\widehat{\rho_{t\sim}}$	-0.024	0.006
	$\widehat{\rho_{t\sim}^*}$	0	0.006

* denotes truncated ICC estimator
~ denotes omission of singletons

6.2 Time-to-tube failure in children with otitis media

Inflammation of the middle ear (otitis media) is the most common illness among young children (Le and Lindgren, 1996). Many patients and physicians turn to surgical interventions such as placement of ventilating tube in the eardrum to reduce the incidence of otitis media episodes.

The purpose of this trial was to extend the life of ventilating tubes with a two week intervention of prednisone and sulfamethoprin as compared to the placebo group (Daly et al., 1995). From Feburary 1987 to January 1990, eighty children between the age of 6 months to 8 years were enrolled with documented history of otitis media. After the surgical placement of ventilating tubes in both ears, 41 children were randomized to the treatment group while 39 children were randomized to the placebo group. Subjects were initially examined before and two weeks after the surgical treatment. In addition, subjects were examined every three months for the entire duration of the study follow-up until the primary outcome of tube failure (blockage or extrusion) was observed.

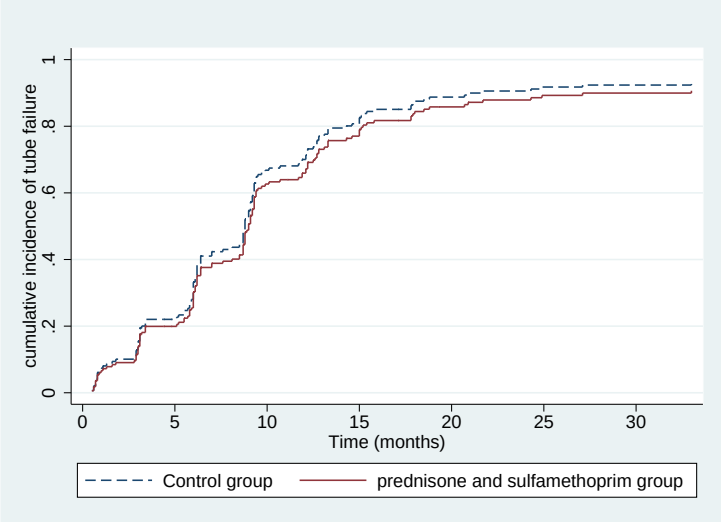
6.2.1 Descriptive analysis

Le and Lindgren (1996) studied the correlated time-to-tube failure for 78 children as two children were excluded from the analysis. One subject was omitted from the analysis as only one ear received a ventilating tube. The amount of censoring for 38 subjects in the placebo group and 40 subjects in the prednisone and sulfamethoprim group was similar and relatively low (Table 6.3). Furthermore, Figure 6.2 shows the decreased incidence of tube-failure in prednisone and sulfamethoprim group as compared to placebo group during the entire length of follow-up after randomization.

Table 6.3: Summary statistics for prednisone and sulfamethoprim, and placebo group

Summary statistics	Placebo group	Prednisone and sulfamethoprim group
Total children	38	40
Censoring rate (%)	9.2	6.3

Figure 6.2: Cumulative-incidence of tube failure in prednisone and sulfamethoprim, and placebo group.



6.2.2 Estimation of the ICC

The ICC is estimated for both groups using observed failure times of ventiliating tubes and censoring indicators. The ICC estimate in the placebo group was 0.54 using censoring

indicators and 0.48 using tube failure times. In contrast, the ICC estimate in prednisone and sulfamethoprim group was -0.05 using censoring indicators and 0.24 using tubal-failure times. After the removal of singletons, the ICC estimate was 0.49 in placebo group and 0.26 in prednisone and sulfamethoprim group. Similar to the previous example, the negative ICC estimates may be truncated to zero (Wang et al., 1992).

As shown in Table 6.4, the outcome of tube failure in prednisone and sulfamethoprim group is positively correlated using observed outcomes while being negatively correlated using censoring indicators. These inconsistent results of estimating the ICC using censoring indicators and observed outcomes may be explained due to lack of correlation between $\widehat{\rho}_c$ and $\widehat{\rho}_t$ for bivariate time-to-event outcomes (as shown in chapter 5). Furthermore, increased instability of estimating the ICC due to smaller cluster size may also contribute to these inconsistent results.

Table 6.4: The ANOVA ICC estimates of Placebo and prednisone and sulfamethoprim group.

Source	ICC estimator	Placebo group	Prednisone and sulfamethoprim group
Censoring indicators	$\widehat{\rho}_c$	0.54	-0.05
	$\widehat{\rho}_c^*$	0.54	0
Observed outcomes	$\widehat{\rho}_t$	0.48	0.24
	$\widehat{\rho}_t^*$	0.48	0.24
	$\widehat{\rho}_{t\sim}$	0.49	0.26
	$\widehat{\rho}_{t\sim}^*$	0.49	0.26

* denotes truncated ICC estimator
~ denotes omission of singletons

6.3 Summary

The ICC estimates computed from observed outcomes and censoring indicators were quite different in both examples. The inconsistency of estimating the ICC from observed outcomes and censoring indicators may further contribute to misleading description of results. The inconsistent reults may be described due to the lack of correlation between $\widehat{\rho}_c$ and $\widehat{\rho}_t$.

Chapter 7

Discussion

This chapter contains four sections. Section 7.1 highlights the discussion of estimating the ICC for time-to-event outcomes. Section 7.2 describes some limitations of the simulation study. Section 7.3 provides scope for future research. Section 7.4 summarizes this chapter and also provides some recommendations for estimating the ICC in CRTs time-to-event outcomes.

7.1 Key findings

Gao et al. (2015) recommended using the binary outcomes (i.e. censoring indicators) to estimate the ICC in CRTs because "ICC is difficult to estimate for survival data". Our study was primarily designed to gain more insight about the performance of ICC estimators in the presence of administrative censoring. The key findings are summarized below according to the performance criteria of bias, variance, MSE and sign test.

The ICC parameter using the censoring indicators (ρ_c) is a function of true ICC (ρ) and percent administratively censored (c) (see section 3.3.1). As a consequence, ρ_c remains unaffected with any changes in total number of clusters (k), cluster size (m) or hazard rate (λ). Hence, a substantial deviation in negative bias may not be observed if variable cluster size is considered in this simulation study. It may also be noted that the analytic expression of ρ_c simplifies to

zero when $\rho = 0$. However, in the presence of positively correlated outcomes, ρ_c displays an asymmetric quadratic relationship with respect to administrative censoring. This relationship can be explained with the following equation (i.e. in absence of Laguerre polynomial):

$$\frac{c^2 \times ln^2(c)}{c(1-c)}.$$

The Newton-Raphson method (Lachin, 2011) can be used to show that this equation is maximized at 20% censoring. The addition of Laguerre polynomial further attenuates the negative quadratic relationship depending on the value of ρ and c. However, more importantly, the analytic expression of ρ_c demonstrates that the censoring indicator approach leads to negatively biased estimation when $\rho \neq 0$. Exploratory results of this simulation study also confirm that an increase in the total number of clusters does not reduce the amount of negative bias of the ICC estimators using either the censoring indicators or observed outcomes.

Negative bias for non-truncated ICC estimators using the observed outcomes increases proportionally with the relative frequency of administrative censoring when $\rho \neq 0$. This is because administrative censoring systematically removes the outcomes with longer event times. As a result, the between-cluster ($\widehat{\sigma}_b^2$) and the within-cluster ($\widehat{\sigma}_w^2$) variance estimators become negatively biased. For example, consider the case where 1% administrative censoring is imposed for 40 clusters of size 200. On average, one might expect eighty out of eight thousand observations ($= 40 \times 200$) to be administratively censored. This corresponds to negligible change in the denominators of $\widehat{\sigma}_b^2$ and $\widehat{\sigma}_w^2$ (equation 2.4) which include the terms of total number of clusters (r) and average cluster size with the penalty factor (m_o). However, the omission of longer event times decreases the overall mean of observed outcomes $\left(\overline{T}_{..}^{*}\right)$. The cluster-specific mean of observed outcomes $\left(\overline{T}_{i.}^{*}\right)$ decreases only if a longer event time is censored within a particular cluster. This leads to substantial reduction in the sums of squared difference of MSB and MSW. The combined effect of omitting longer event times on r, k, m_o, $\overline{T}_{..}^{*}$ and $\overline{T}_{i.}^{*}$ ultimately attenuates $\widehat{\rho_t}$ towards zero as the amount of administrative censoring is increased.

The increase in cluster size to two hundred reduces the variance of all ICC estimators in relation to the cases where cluster size is two (Ukoumunne, 2002). The smaller variance further reduces the disparity between non-truncated and truncated ICC estimators as the occurrence of negative ICC estimates is reduced when $m = 200$. A decrease in variance of ICC estimator using the censoring indicators is recorded at the extreme values of percent censoring. This may be explained due to the ICC parameter using the censoring indicators being close to zero at the extreme values of percent censoring. Under moderate frequency of administrative censoring, the variance of all ICC estimators using the censoring indicators remains unchanged. However, the variance of all ICC estimators derived from observed outcomes increases with respect to administrative censoring. This may be explained due to the reduction in total number of events being recorded as the frequency of administrative censoring is increased. Furthermore, a decrease in variance is noted for the ICC estimators where singletons are omitted and this is in agreement with previous studies (Swiger et al., 1964).

Since MSE incorporate the information of variance and bias, similar results are also reflected in the performance criteria of MSE. For example, MSE is increased for all ICC estimators using observed outcomes with the increasing frequency of administrative censoring. The truncated ICC estimators have lower MSE than their non-truncated counterparts possibly due to reduction in estimated variance (Bellhouse, 2015).

The sign test is used to determine which ICC estimator tends to be closer to ρ over n replicates. The non-truncated ICC estimator using observed outcomes ($\widehat{\rho}_t$) is compared to the non-truncated ICC estimator using censoring indicators ($\widehat{\rho}_c$). As expected, $\widehat{\rho}_t$ estimate ρ more closely than $\widehat{\rho}_c$ under minimal frequency of administrative censoring ($\leq 5\%$). In contrast, $\widehat{\rho}_c$ estimates ρ more closely than $\widehat{\rho}_t$ under moderate to high frequency of administrative censoring ($> 5\%$). In spite of substantial amount of negative bias observed in both estimators, $\widehat{\rho}_c$ estimates ρ more closely than $\widehat{\rho}_t$ primarily due to $\widehat{var}(\widehat{\rho}_c) \leq \widehat{var}(\widehat{\rho}_t)$.

7.2 Limitations

The ANOVA estimator is negatively biased in the absence of censoring (as shown in section 5.2). However, the amount of negative bias is negligible as compared to the amount of bias observed in the presence of censoring for all six ICC estimators. Furthermore, the negative bias of the ANOVA estimator decreases with the increase in total number of clusters; thus implying that ANOVA is a consistent estimator of ICC (Lehmann and Casella, 1998).

The administrative censoring is imposed from 1% to 99% in increments of 1%. This gives rise to replicates where clusters of size one (singletons) are predominant, especially when $m = 2$. If a particular replicate consist of only singletons then the ANOVA ICC estimator using observed outcomes ($\widehat{\rho}_t$) would equate to one. In addition, $\widehat{\rho}_t$ is observed to be less than -1 when singletons are predominant within a replicate. The occurence of $\widehat{\rho}_t$ less than -1 may be explained using the following inequality which describes the distributional range of ICC estimates (Wang et al., 1992):

$$\frac{1}{m_o - 1} \leq \widehat{\rho}_t \leq 1.$$

If $1 < m_o < 2$ where m_o is defined on page 14 then the theoretical lower bound of $\widehat{\rho}_t$ is less than -1. An appropriate solution is to either use the truncated ICC estimator ($\widehat{\rho}_t^*$) or to use the ICC estimator where singletons are omitted ($\widehat{\rho}_{t\sim}$).

7.3 Future research

The focus of this thesis was limited to estimating the ICC for time-to-event outcomes using the multivariate exponential model as described by Moran (1967). Furthermore, attention was restricted to the following set of conditions: (1) accrual prior to randomization of the control group, (2) administrative censoring and (3) constant hazard rate with respect to time. The first two conditions allowed the time-to-event outcomes to be dichotomized at the termination point (t_c) of the study. Donner and Eliasziw (1994) evaluated the effect of dichotomizing

an inherently continuous outcome (under normality) on sample size estimation in reliability studies. A possible next step may be to explore the impact of using the ICC derived from observed outcomes and censoring indicators on sample size estimation in CRTs with time-to-event outcomes.

The future research may also encompass different type of censoring mechanism and accrual scheme. For example, He and Nagaraja (2011) and Lin et al. (2013) considered Type II censoring mechanism while Manatunga and Chen (2000) restricted attention to uniform censoring mechanism. Jung (2008) explored uniform accrual period to estimate the sample size for CRTs with bivariate survival outcomes.

A plausible next step may be to construct a new bias-corrected ICC estimator that accommodates censoring in time-to-event outcomes. This may share some similarity with the bias-corrected estimators for binary (Saha and Paul, 2005) and continuous outcomes (Atenafu et al., 2012). The use of ML estimation for Moran (1967) algorithm may utilize all the available information from time-to-event outcomes. Future studies may also examine the use of Kendall's τ to measure the degree of association among outcomes within a cluster (Lakhal et al., 2009).

7.4 Summary

This study showed that the use of censoring indicators or observed outcomes lead to negatively biased estimation of the ICC parameter. However, censoring indicators and observed outcomes may still be preferred to estimate the ICC under certain circumstances. For example, if moderate frequency of administrative censoring (e.g. $5\% < c < 25\%$) is recorded then censoring indicators are preferred to estimate the ICC. On the other hand, if minimal frequency of administrative censoring (e.g. $c \leq 5\%$) is recorded then observed outcomes are preferred to estimate the ICC. Ultimately, neither approach can be recommended in general due to substantial amount of negative bias in the presence of administrative censoring.

Bibliography

Abdulla, S., Schellenberg, J., Nathan, R., Mukasa, O., Marchant, T., Smith, T., Tanner, M., and Langeler, C. (2001). Impact on malaria morbidity of a programme supplying insecticide treated nets in children aged under 2 years in Tanzania: community cross sectional study. *British Medical Journal*, 22(7281):270–273.

Atenafu, E., Hamid, J., To, T., Willan, A., Feldman, B., and Beyene, J. (2012). Bias-corrected estimator for intraclass correlation coefficient in the balanced one-way random effects model. *BioMed Central Medical Research Methodology*, 12.

Balakrishnan, N. and Lai, C. (2009). *Continuous Bivariate Distributions*. Springer.

Balakrishnan, N. and Ng, H. (2001). On estimation of the correlation coefficient in Moran-Downton multivariate exponential distribution. *Journal of Statistical Computation and Simulation*, 71:41–58.

Bellhouse, D. (2015). Communication between Dr. Klar and Dr. Bellhouse.

Binka, F., Kubaje, A., Adjuik, M., Williams, L., Lengeler, C., Maude, G., Armah, G., Kajihara, B., Adiamah, J., and Smith, P. (1996). Impact of permethrin impregnated bednets on child mortality in Kassena-Nankana district, Ghana: a randomized controlled trial. *Tropical Medicine and International Health*, 1(2):147–154.

Burton, A., Altman, D., Royston, P., and Holder, R. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine*, 25(24):4279–4292.

Campbell, M. and Walters, S. (2014). *How to Analyse and Report Cluster Randomised Trials in Medicine and Health Related Research*. John Wiley and Sons.

Chakraborty, H., Moore, J., Carlo, W., Hartwell, T., and Wright, L. (2009). A simulation based technique to estimate intracluster correlation for a binary variable. *Contemporary Clinical Trials*, 30(2):71–80.

Daly, K., Giebink, G., Lindgren, B., Margolis, R., Westover, D., Hunter, L., Le, C., and Buran, D. (1995). Randomized trial of the efficacy of trimethoprim-sulfamethoxazole and prednisone in preventing post-tympansotomy tube morbidity. *The Pediatric Infectious Disease Journal*, 14(12):1068–1074.

Donner, A. (1986). A review of inference procedures for the intraclass correlation coefficient in the one-way random effects model. *International Statistical Review*, 54(1):67–82.

Donner, A. (1992). Sample size requirements for stratified cluster randomization designs. *Statistics in Medicine*, 11(6):743–750.

Donner, A., Birkett, N., and Buck, C. (1981). Randomization by cluster. sample size requirements and analysis. *American Journal of Epidemiology*, 114(6):905–914.

Donner, A. and Eliasziw, M. (1994). Statistical implications of the choice between a dichotomous or continuous trait in studies of intraobserver agreement. *Biometrics*, 50(2):550–555.

Donner, A. and Klar, N. (1994a). Cluster randomization trials in epidemiology: theory and application. *Journal of Statistical Planning and Inference*, 42(1):37–56.

Donner, A. and Klar, N. (1994b). Methods for comparing event rates in intervention studies when the unit of allocation is a cluster. *American Journal of Epidemiology*, 140(3):279–289.

Donner, A. and Klar, N. (2000). *Design and Analysis of Cluster Randomization Trials in Health Research*. Arnold, London, UK.

Donner, A. and Koval, J. (1980). The estimation of intraclass correlation in analysis of family data. *Biometrics*, 36(1):19–25.

Eldridge, S., Kerry, S., and Torgerson, D. J. (2009). Bias in identifying and recruiting participants in cluster randomised trials: what can be done? *British Medical Journal*, 339.

Elston, R. (1977). Query: Estimating "heritability" of a dichotomous trait. *Biometrics*, 33(1):231–233.

Fisher, R. (1925). *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh.

Fleiss, J. and Cuzick, J. (1979). The reliability of dichotomous judgements: Unequal number of judges per subject. *Applied Psychological Measurements*, (3):527–542.

Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Annales de l'Université de Lyon. Section A: Sciences Mathématiques et Astronomie*, 9(1):53–77.

Friedman, L., Furberg, C., and DeMets, D. (2010). *Fundamentals of Clinical Trials*. Springer, 4th edition.

Gangnon, R. and Kosorok, M. (2004). Sample-size formula for clustered survival data using weighted log-rank statistics. *Biometrika*, 91(2):263–275.

Gao, F., Earnest, A., Matchar, D., Campbell, M., and Machin, D. (2015). Sample size calculations for the design of cluster randomized trials: A summary of methodology. *Contemporary Clinical Trials*, 42:41–50.

Gumbel, E. (1960). Bivariate exponential distribution. *Journal of American the Statistical Association*, 55(292):608–707.

Haggard, E. (1958). *Intraclass correlation and analysis of variance*. Dryden, New York.

Hansen, M. and Hurwitz, W. (1942). Relative efficiency of various sampling units in population inquiries. *Journal of the American Statistical Association*, 37(217):89–94.

Harris, J. (1913). On the calculation of intra-class and inter-class coefficients of correlation from class moments when the number of possible combinations is large. *Biometrika*, 9(3):446–472.

Harville, D. (1977). Maximum likelihood approaches to the variance component estimation and to related problems. *Journal of the American Statistical Society*, 72(358):320–338.

He, Q. and Nagaraja, H. (2011). Correlation estimation in downton's bivariate exponential distribution using incomplete samples. *Journal of Statistical Computation and Simulation*, 81(5):531–546.

Horbar, J., Carpenter, J., Buzas, J., Soll, R., Suresh, G., Bracken, M., Leviton, L., Plsek, P., and Sinclair, J. (2004). Collaborative quality improvement to promote evidence based surfactant for preterm infants: a cluster randomised trial. *British Medical Journal*, (329):1004.

Jahn-Eimermacher, A., Ingel, K., and Schneider, A. (2013). Sample size in cluster randomized trials with time to event as the primary endpoint. *Statistics in Medicine*, 32(5):739–751.

Jeong, J. and Jung, S. (2006). Rank tests for clustered survival data when dependent subunits are randomized. *Statistics in Medicine*, 25(3):361–373.

Jung, S. (2007). Sample size calculation for weighted rank tests comparing survival distributions under cluster randomization: A simulation method. *Journal of Biopharmaceutical Statistics*, 17(5):839–849.

Jung, S. (2008). Sample size calculation for the weighted rank statistics with paired survival data. *Statistics in Medicine*, 27(17):3350–3365.

Jung, S. and Jeong, J. (2003). Rank test for clustered survival data. *Lifetime Data Analysis*, 9(1):21–33.

Kalbfeisch, J. and Prentice, R. (2002). *The Statistical Analysis of Failure Time Data.* John Wiley and Sons, 2nd edition.

Kang, S. and Koehler, K. (1997). Modification of the greenwood formula for correlated response time. *Biometrics*, 53(3):885–899.

Karlin, S., Cameron, P., and Williams, P. (1981). Sibling and parent-offspring correlation with variable family age. *Proceedings of the National Academy of Science USA*, 78(5):2664–2668.

Klar, N., Lipsitz, S., and Ibrahim, J. (2000). An estimating equations approach for modelling kappa. *Biometrical Journal*, 42(1):45–58.

Kleinbaum, D. and Klein, M. (2012). *Survival Analysis*. Springer, 3rd edition.

Koekoek, R. (1990). *Generalizations of the classical Laguerre polynomials and some q-analogues*. PhD thesis, Technische Universiteit Delft.

Kotz, S., Balakrishnan, N., and Johnson, N. (2000). *Multivariate Exponential Distributions, in Continuous Multivariate Distributions: Models and Applications*, volume 1. John Wiley & Sons, 2nd edition.

Lachin, J. (2011). *Biostatistical Methods: The Assessment of Relative Risks*. John Wiley & Sons, 2nd edition.

Lakhal, L., Rivest, L., and Beaudoin, D. (2009). IPCW Estimator for Kendall's Tau under Bivariate Censoring. *The International Journal of Biostatistics*, 5(1).

Le, C. and Lindgren, B. (1996). Duration of ventilating tubes: a test for comparing two clustered samples of censored data. *Biometrics*, 52(1):328–334.

Lee, J., Kim, J., and Jung, S. (2007). Bayesian analysis of paired survival data using a bivariate exponential distribution. *Lifetime Data Analysis*, 13(1):119–137.

Lehmann, E. and Casella, G. (1998). *Theory of Point Estimation*. Springer, 2nd edition.

Lin, D. and Ying, Z. (1993). A simple nonparametric estimator of the bivariate survival function under univariate censoring. *Biometrika*, 80(3):573–581.

Lin, Y., Lio, Y., and Ng, H. (2013). Bayes estimation of Moran–Downton bivariate exponential distribution based on censored samples. *Journal of Statistical Computation and Simulation*, 83(5):835–850.

Loeb, M., Russell, M., Moss, L., Fonseca, K., Fox, J., Earn, D., Aoki, F., Van-Caeseele, G. H. P., Chokani, K., Vooght, M., Babiuk, L., Webby, R., and Walter, S. (2010). Effect of influenza vaccination of children on infection rates in Hutterite communities: a randomized trial. *Journal of the American Medical Association*, 303(10):943–950.

Lui, K. (2000). A note on the log-rank life table analysis with correlated observations. *Biometrical Journal*, 42(4):457–470.

Mainland, D. (1952). *Elementary Medical Statistics; the Principles of Quantitative Medicine*. W.B. Saunders, Philadelphia.

Manatunga, A. and Chen, S. (2000). Sample size estimation for survival outcomes in cluster-randomized studies with small cluster size. *Biometrics*, 56(2):616–621.

Meorbeek, M. (2012). Sample size issues for cluster randomized trials with discrete time survival endpoints. *Methodology*, 8(4):146–158.

Molenberghs, G., Fitzmaurice, G., and Lipsitz, S. (1996). Efficient estimation of the intraclass correlation for a binary trait. *Journal of Agriculture, Biological and Environmental Sciences*, 1(1):78–96.

Moran, P. (1967). Testing for correlation of non-negative variates. *Biometrika*, 54(3):385–394.

Ponzoni, R. and James, J. (1978). Possible bias in heritability estimates from intraclass correlation. *Theoretical Applied Genetics*, 53(1):25–27.

Ray, W., Taylor, J., Meador, K., Thapa, P., Brown, A., Kajihara, H., Davis, C., Gideon, P., and Griffin, M. (1997). A randomization trial of a consolation service to reduce falls in nursing homes. *Journal of the American Medical Association*, 278(7):557–562.

Ridout, M., Demetrio, C., and Firth, D. (1999). Estimating intraclass correlation for binary data. *Biometrics*, 55(1):137–148.

Rodrigues, A., Schellenberg, J., Kofoed, P., Aaby, P., and Greenwood, B. (2008). Changing pattern of malaria in Bissau, Guinea Bissau. *Tropical Medicine and International Health*, 13(3):410–417.

Rosner, B., Donner, A., and Hennekens, C. (1977). Estimation of interclass correlation from familial data. *Journal of Royal Statistical Society*, 26(2):179–187.

Saha, K. and Paul, S. (2005). Bias-corrected maximum likelihood estimator of the intraclass correlation parameter for binary data. *Statistics in Medicine*, 24(22):3497–3512.

SAS Institute Inc. (2012). *SAS/STAT Software, Version 9.4*. Cary, NC.

Scott, W. (1955). Reliability of content analysis; the case of nominal scale coding. *Public Opinion Quarterly*, 19(3):321–325.

Searle, S., Casella, G., and McCulloch, C. (1992). *Variance Components*. John Wiley and Sons.

Segal, M. and Neuhaus, J. (1993). Robust inference for multivariate survival data. *Statistics in Medicine*, 12(11):1019–1031.

Segal, M. and Neuhaus, J. (1997). Dependence estimation of marginal models of multivariate survival data. *Lifetime Data Analysis*, 3(3):251–268.

Simpson, J. M., Klar, N., and Donner, A. (1995). Accounting for cluster randomization: A review of primary prevention trials. *American Journal of Public health*, 85(10):1378–1382.

Smith, P., Moffatt, M., Gilskey, S., Hudson, S., and Kaita, K. (1997). Are community health interventions evaluated appropriately? A review of six journals. *Journal of Clinical Epidemiology*, (50):137–146.

Solomon, P. and Taylor, M. (1999). Orthogonality and transformations in variance components models. *Biometrika*, 86(2):289–300.

Su, P., Chi, Y., Li, C., Yu, S., and Liao, Y. (2011). Analyzing survival curves at a fixed point in time for paired and clustered right-censored data. *Computational Statistics and Data Analysis*, 55(4):1617–1628.

Swallow, W. and Monahan, J. (1984). Monte Carlo comparison of ANOVA, MIVQUE, REML and ML Estimator of Variance components. *Technometrics*, 26(1):47–57.

Swiger, L., Harbey, W., Everson, D., and Gregory, K. (1964). The variance of intraclass correlation involving groups with one observation. *Biometrics*, 20(4):818–826.

Ukoumunne, O. (2002). A comparison of confidence interval methods for the intraclass correlation coefficient in cluster randomized trials. *Statistics in Medicine*, 21:3757–3774.

Wackerly, D., Mendenhall, W., and Scheaffer, R. (2008). *Mathematical Statistics with Applications*. Thomson Higher Education., 7th edition.

Walleser, S., Hill, S., and Bero, L. (2011). Characteristics and quality of reporting of cluster randomized trials in children: reporting needs improvement. *Journal of Clinical Epidemiology*, 64(12):1331–1340.

Wang, C., Yandell, B., and Rutledge, J. (1992). The dilemma of negative analysis of variance estimators of intraclass correlation. *Theoretical Applied Genetics*, 85(1):79–88.

Wax, Y., Baker, S., and Patterson, B. (1993). A Score Test for Non-Informative Censoring Using Doubly Sampled Grouped Survival Data. *Journal of the Royal Statistical Society*, 42(1):159–172.

Wienke, A. (2011). *Frailty Models in Survival Analysis*. Taylor and Francis Group, 1st edition.

Williams, R. (1995). Product-limit survival functions with correlated survival times. *Lifetime Data Analysis*, 1(2):171–186.

World Health Organization (2014). World Malaria Report 2014. Technical report.

Xie, T. and Waksman, J. (2003). Design and sample size estimation in clinical trials with clustered survival times as the primary endpoint. *Statistics in Medicine*, 22(18):2835–2846.

Zou, G. and Donner, A. (2004). Confidence Interval Estimation of the Intraclass Correlation Coefficient for Binary Outcome Data. *Biometrics*, 60(4):807–811.

Zou, G., Donner, A., and Klar, N. (2005). Group sequential methods for cluster randomization trials with binary outcomes. *Clinical Trials*, 2(6):479–487.

Appendix A

ICC estimators

Appendix A contains two section. Section A.1 limits the attention to clusters of fixed size to show the equivalence among ANOVA, ML and pairwise estimators of the ICC. Section A.2 derives the binary analogue of Gaussian ML estimator that uses censoring indicators to estimate the ICC.

A.1 Equivalence among ANOVA , ML and pairwise estimators of the ICC

Consider a one-way random effects model:

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \qquad i = 1, ..., k, j = 1, ..., m.$$

where $\alpha_i \sim N(0, \sigma_B^2)$ is the cluster-specific random effect and $\epsilon_{ij} \sim N(0, \sigma_W^2)$ is the residual error term. Moreover, α_i and ϵ_{ij} are assumed to be independent. Suppose also that there are k clusters each with a fixed size of m members. The ICC parameter can be expressed as

$$\begin{aligned} \rho_A &= Corr(Y_{ij}, Y_{il}) && \text{for } j \neq l \\ &= \frac{Cov(Y_{ij}, Y_{il})}{\sqrt{Var(Y_{ij})Var(Y_{il})}} \end{aligned}$$

$$
\begin{aligned}
&= \frac{E(Y_{ij}-\mu)(Y_{il}-\mu)}{\sqrt{E(Y_{ij}-\mu)^2 E(Y_{il}-\mu)^2}} \\
&= \frac{E(\alpha_i+\epsilon_{ij})(\alpha_i+\epsilon_{il})}{\sqrt{E(\alpha_i+\epsilon_{ij})^2 E(\alpha_i+\epsilon_{il})^2}} \\
&= \frac{E(\alpha_i^2)}{\sqrt{E(\alpha_i^2+\epsilon_{ij}^2)E(\alpha_i^2+\epsilon_{il}^2)}} \\
&= \frac{\sigma_B^2}{\sigma_B^2+\sigma_W^2}
\end{aligned}
$$

Using a one-way random effects model (Table A.1), the ANOVA estimator of ICC can be expressed as:

$$\widehat{\rho}_A = \frac{MSB - MSW}{MSB + (m-1)MSW} \tag{A.1}$$

The between sum of squares (SSB) and within sum of squares (SSW) are defined as:

$$MSB = \frac{1}{k-1}SSB = \frac{m}{k-1}\sum_{i=1}^{k}\left(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot}\right)^2$$

$$MSW = \frac{1}{k(m-1)}SSW = \frac{1}{k(m-1)}\sum_{i=1}^{k}\sum_{j=1}^{m}\left(Y_{ij} - \overline{Y}_{i\cdot}\right)^2$$

where $\overline{Y}_{\cdot\cdot} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{m} Y_{ij}}{km}$ and $\overline{Y}_{i\cdot} = \frac{\sum_{j=1}^{m} Y_{ij}}{m}$

Table A.1: One-way random effects model

Source	**Sum of squares**	**Degrees of freedom**	**Mean square**	**Expected mean square**
Between clusters	SSB	k-1	MSB	$m\sigma_B^2 + \sigma_W^2$
Within clusters	SSW	k(m-1)	MSW	σ_W^2
Total	SST*	mk-1		

* where total sum of squares (SST)

In the case of fixed cluster sizes, Rosner et al. (1977) derived the ML estimator of ICC under normality and also showed that it is equivalent to pairwise estimator of ICC. The asymptotic equivalence of ML estimator and pair-wise estimator to the ANOVA estimator can also be

established for a large number of k clusters:

$$\begin{aligned}
\widehat{\rho}_{PW} = \widehat{\rho}_{ML} &= \frac{\sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{l=1}^{m}\left(Y_{ij}-\overline{Y}_{..}\right)\left(Y_{il}-\overline{Y}_{..}\right)}{(m-1)\sum_{i=1}^{k}\sum_{j=1}^{m}\left(Y_{ij}-\overline{Y}_{..}\right)^2} && \text{for } j \neq l \\
&= \frac{\sum_{i=1}^{k}\sum_{j=1}^{m}\left(Y_{ij}-\overline{Y}_{..}\right)\sum_{l=1}^{m}\left(Y_{il}-\overline{Y}_{..}\right)-\sum_{i=1}^{k}\sum_{j=1}^{m}\left(Y_{ij}-\overline{Y}_{..}\right)^2}{(m-1)(SSB+SSW)} \\
&= \frac{\sum_{i=1}^{k} m^2\left(Y_{i\cdot}-\overline{Y}_{..}\right)-(SSB+SSW)}{(m-1)(SSB+SSW)} \\
&= \frac{m\sum_{i=1}^{k}\sum_{j=1}^{m}\left(Y_{i\cdot}-\overline{Y}_{..}\right)^2-(SSB+SSW)}{(m-1)(SSB+SSW)} \\
&= \frac{m(SSB)-(SSB+SSW)}{(m-1)(SSB+SSW)} \\
&= \frac{(m-1)SSB+SSW}{(m-1)(SSB+SSW)} \\
&= \frac{m-1}{m-1}\times\frac{(k-1)MSB+k(MSW)}{(k-1)MSB+k(m-1)MSW} \\
&= \frac{\frac{k-1}{k}MSB+MSW}{\frac{k-1}{k}MSB+(m-1)MSW} \\
&\approx \frac{MSB+MSW}{MSB+(m-1)MSW} \\
&= \widehat{\rho}_{A} && \text{as needed } \square
\end{aligned}$$

Hence, the equivalence among ANOVA, ML and the pairwise ICC estimator is shown for continuous outcomes where there are equal number of subjects per cluster.

A.2 Binary analogue of Gaussian ML estimator

In the context of binary outcomes (i.e. censoring indicators), the ML estimator is written as:

$$\widehat{\rho}_{ML} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{l=1}^{m}(\Delta_{ij}-\widehat{p})(\Delta_{il}-\widehat{p})}{(m-1)\sum_{i=1}^{k}\sum_{j=1}^{m}(\Delta_{ij}-\widehat{p})^2} = \frac{\sum_{i=1}^{k}\sum_{j\neq l}^{m}(\Delta_{ij}-\widehat{p})(\Delta_{il}-\widehat{p})}{(m-1)\sum_{i=1}^{k}\sum_{j=1}^{m}(\Delta_{ij}-\widehat{p})^2} \tag{A.2}$$

where Δ_{ij} is defined in equation 2.1 and $\widehat{p} = \frac{1}{mk}\sum_{i=1}^{k}\sum_{j=1}^{m}\Delta_{ij}$. Furthermore, equation A.2 is algebraically manipulated to show that the estimator proposed by Xie and Waksman (2003) is

equivalent to this binary analogue of the ML estimator. The denominator of equation A.2 is manipulated as

$$
\begin{aligned}
&= (m-1)\sum_{i=1}^{m}\sum_{j=1}^{k}(\Delta_{ij}-\widehat{p})^2 \\
&= (m-1)\sum_{i=1}^{m}\sum_{j=1}^{k}\left(\Delta_{ij}^2 - 2\Delta_{ij}\frac{1}{mk}\sum_{i=1}^{m}\sum_{j=1}^{k}\Delta_{ij} + \left(\frac{1}{mk}\sum_{i=1}^{m}\sum_{j=1}^{k}\Delta_{ij}\right)^2\right) \\
&= (m-1)\left(\sum_{i=1}^{m}\sum_{j=1}^{k}\Delta_{ij}^2 - \frac{1}{mk}\left(\sum_{i=1}^{m}\sum_{j=1}^{k}\Delta_{ij}\right)^2\right)
\end{aligned}
$$

Since

$$
\Delta_{ij} = \begin{cases} 1 & \text{if } T_{ij} \le t_c \\ 0 & \text{if } T_{ij} > t_c \end{cases}
$$

it can be noted that

$$
\sum_{i=1}^{m}\sum_{j=1}^{k}\Delta_{ij}^2 = \sum_{i=1}^{m}\sum_{j=1}^{k}\Delta_{ij}.
$$

With this assertion, the denominator can be further simplified as

$$
\begin{aligned}
&= (m-1)\left(\sum_{i=1}^{m}\sum_{j=1}^{k}\Delta_{ij} - \frac{1}{mk}\left(\sum_{i=1}^{m}\sum_{j=1}^{k}\Delta_{ij}\right)^2\right) \\
&= (m-1)\left(\sum_{i=1}^{m}\sum_{j=1}^{k}\Delta_{ij}\right)\left(1 - \frac{1}{mk}\sum_{i=1}^{m}\sum_{j=1}^{k}\Delta_{ij}\right) \\
&= mk(m-1)(\widehat{p})(1-\widehat{p}) \qquad \text{as needed } \square
\end{aligned}
$$

Therefore, the equivalence between the estimator proposed by Xie and Waksman (2003) and binary analogue of Guassian ML estimator is shown as:

$$
\widehat{\rho}_{ML} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{l=1}^{m}(\Delta_{ij}-\widehat{p})(\Delta_{il}-\widehat{p})}{(m-1)\sum_{i=1}^{k}\sum_{j=1}^{m}(\Delta_{ij}-\widehat{p})^2} = \frac{\sum_{i=1}^{k}\sum_{j\neq l}^{m}(\Delta_{ij}-\widehat{p})(\Delta_{il}-\widehat{p})}{mk(m-1)(\widehat{p})(1-\widehat{p})} = \widehat{\rho}_2
$$

Appendix B

Data generation

The data is generated using SAS V. 9.4 (SAS Institute Inc., 2012). The code of the SAS macro (generatedata) is shown below. Note that the time-to-event outcomes (labelled as 'Time') in dataset (&out) are not right truncated at the termination point of the study. Hence the dataset (&out) provides the user with the failure times simulated directly from Moran's algorithm. Furthermore, the time-to-event outcomes are dichotomized at the termination point of the study to determine the censoring indicators. The censoring indicators are labelled as 'censortime' in dataset '&out' where 0 denotes censored observations and 1 denotes observed outcomes.

B.1 SAS macro

```
/*************************************************************************
SAS macro for generating the data using Moran's algorithm:
1) seed= unique seed to generate the data
2) rep= total number of replications
3) rho= true ICC value
4) m= cluster size
5) k= number of clusters
6) hazard= constant hazard rate (per year)
7) percentcensor= administrative censoring (percent decimal: zero to one)
8) out= name of the dataset
*************************************************************************/
```

```
%macro generatedata (seed=,rep=,rho=,m=,k=,hazard=,percentcensor=,out=);
proc iml;
call randseed(&seed);

*censoring time;
censortime=(-1/&hazard)*(log(&percentcensor));
call symputx('censortime',censortime);

*mean of multivariate normal;
Mean = j(1,&m,0);

*variance-covariance matrix of multivariate normal distribution;
a = j(&m,&m,sqrt(&rho));
b= I(&m)*(1-sqrt(&rho));
Cov= a + b;

*Moran algorithm;
*Two independent vectors from multivariate normal distribution;
Aij = RandNormal(&k*&rep, mean, Cov);
Bij = RandNormal(&k*&rep, mean, Cov);

*indicators for replicates and cluster;
replicate = shape( repeat(T(1:&rep), 1, &k), &k*&rep );
cluster = shape( repeat(T(1:&k), &rep, 1), &k*&rep );

*Marginal exponential random variables;
Tij= (Aij##2 + Bij##2)/(2*&hazard);

*concatenation of indicators and Tij;
T= replicate|| cluster || Tij;
varNames = ( "replicate" )||( "cluster" )||
("T1":("T"+strip(char(ncol(Cov))))) ;

*generation of the dataset;
create wideform from T [colname=VarNames];
append from T;
close wideform;
quit;

*converting wideform to longform;
proc transpose data= wideform out=longform (RENAME=(_NAME_= Tij COL1=Time));
by replicate cluster;
run;
```

```
proc sort data=longform;
by replicate;
run;

*labelling the dataset with subject ID, cluster ID, censoring indicators;
data dataset;
set longform;
by replicate;
if first.replicate then ID=0; ID+1;
cluster= ceil(ID/&m);
if Time <= &censortime then censoring = 1;
else censoring = 0;
run;

proc sql;
create table &out as select
replicate, ID , cluster, Time, censoring
from dataset;
quit;

%mend;

%generatedata (seed=,rep=,rho=,m=,k=,hazard=,percentcensor=,out=);
```

Curriculum Vitae

Name:	Sumeet Kalia
Post-Secondary Education and Degrees:	University of Western Ontario London, ON 2013-2015 M.Sc. University of Toronto Mississauga, ON 2009-2013 B.Sc.
Honours and Awards:	Ontario Graduate Scholarship 2014-2015
Related Work Experience:	Teaching Assistant University of Western Ontario 2014 - 2015 Research Assistant University of Western Ontario 2013 - 2015
Conference Presentation:	Kalia, S. Estimation of intracluster correlation coefficient for survival data in cluster randomized trials. Oral presentation at the Canadian Society of Epidemiology and Biostatistics (CSEB) conference in Mississauga, Ontario, 2015.