응용통계연구 (2010)
**23**(3), 497–510

# 신뢰도 연구에서 급내상관계수와 관련한 표본수 결정 방법 비교

한수연[1] · 남정모[2] · 명성민[3] · 송기준[4]

[1]연세대학교 의학통계학과, [2]연세대학교 의학통계학과, [3]중원대학교 의료정보학과
[4]연세대학교 의학통계학과

(2010년 1월 접수, 2010년 3월 채택)

**요 약**

신뢰도 연구는 한 명의 평가자가 연구 대상을 반복 측정하거나 여러 명의 평가자가 한 대상을 평가할 때 평가자 내, 평가자 간 일치도를 알아보는 연구로 임상 의학 분야에서 빈번하게 쓰이고 있다. 이 같은 신뢰도 연구에서 적절한 표본수, 평가자수 및 반복수를 결정하는 것은 비용과 시간 측면에서 보다 더 효율적인 연구를 할 수 있게 해주는 중요한 요인이다. 본 연구의 목적은 신뢰도 연구에서 측정치가 정량적일 때 쓰이는 신뢰도 계수인 급내상관계수(ICC)와 관련한 기존의 표본수 산출 방법들을 비교분석하여 적절한 표본수나 반복수를 결정할 때 그 지침을 제공하는데 있다. 기존 논문에서 제시한 Walter 등 (1998), Giraudeau와 Mary (2001), Saito 등 (2006) 그리고 Bonett (2002)의 방법들을 비교하였다. 임의효과 일원배치 모형일 때 같은 조건에서 가장 적은 양의 정보를 필요로 하는 방법을 찾는 목적으로 요인을 변화시켜 가면서 표본수, 반복수, 신뢰구간 폭을 비교한다. 비교해 본 결과, 가장 작은 수의 표본을 필요로 하는 방법은 Giraudeau의 방법, 가장 작은 수의 반복을 필요로 하는 방법은 Saito의 방법으로 나타났다. 가장 많은 수의 표본과 반복을 필요로 한 방법은 Bonett의 방법이었다. 정도는 Giraudeau의 방법이 가장 높았고 Walter, Saito, Bonett 순으로 정도가 떨어졌다.

주요용어: 신뢰도 연구, 급내상관계수, 표본수, 반복수, 평가자수.

## 1. 서론

### 1.1. 연구 배경 및 목적

최근 임상 의학 연구에서 신뢰도 연구(reliability study)의 필요성이 점차 대두되고 있다. 신뢰도 연구는 새로운 검사 방법이 기존의 방법과 일치하는지 혹은 같은 표본에서의 측정이 일관된 결과를 제공하는지 알아보는 등의 경우에 유용하게 사용된다 (White와 Broek, 2004). 이러한 신뢰도 연구에서 측정치가 정량적일 때 쓰이는 급내상관계수(Intraclass Correlation Coefficient)를 이용하는 경우, 연구 설계 시점에서 중요한 관심사는 몇 명의 평가자(rater)가 몇 명의 표본(subject)을 몇 번 반복(replicate)하여 관찰해야 하는가의 문제이다. 위의 세 가지 요인을 적절하게 설정하는 것은 시간이나 비용 등의 여러 측면에서 보다 더 효율적으로 연구를 수행할 수 있게 해준다. 본 연구에서는 기존의 ICC와 관련한 표본수 산출 방법들을 비교하여 적절한 표본, 평가자 혹은 반복의 수를 결정하는데 일련의 지침을 제공하고자 한다.

[4]교신저자: (120-752) 서울시 서대문구 신촌동 134 연세대학교 의학통계학과, 연구조교수.
E-mail: biostat@yuhs.ac

표 **2.1.** Landis와 Koch가 제안한 신뢰도 계수 분류

| 값 | 신뢰도의 정도 |
|---|---|
| $\rho <= 0.00$ | Poor |
| $0.00 < \rho <= 0.20$ | Slight |
| $0.20 < \rho <= 0.40$ | Fair |
| $0.40 < \rho <= 0.60$ | Moderate |
| $0.60 < \rho <= 0.80$ | Substantial |
| $0.80 < \rho <= 1.00$ | Almost Perfect |

### 1.2. 연구 내용 및 방법

본 연구에서는 one-way 모형을 적용하여 기존의 방법들을 비교하고자 한다. 검정력(power)에 초점을 맞추어 one-way 모형에서 반복수가 고정일 때 요구되는 표본수를 계산한 Walter의 방법, 정도(precision)에 초점을 맞추어 각각 one-way, two-way 모형에서 반복수와 표본을 곱한 값이 정해져있을 때 표본수를 계산한 Giraudeau와 Saito의 방법 그리고 반복수에 따른 표본수 결정에 대해 두 모형에 모두 적용할 수 있는 근사 closed-form을 제시한 Bonett의 방법을 비교한다. 특정 상황에서 각 방법들에 대한 효율성을 비교하기 위해 고정된 유의수준에서 신뢰계수와 요인의 값을 변화시켜 가면서 확인하고자 한다.

### 1.3. 논문의 구성

제 1장에서는 연구의 배경 및 목적과 연구 내용 및 방법을 소개한다. 2장에서는 ICC와 관련한 표본수를 제공해주는 방법들을 포함하여 신뢰도 연구, ICC에 관한 이론적 배경을 언급한다. 3장에서는 실제적으로 요인을 변화시켜 가면서 방법들 간 차이를 다양한 각도에서 비교분석한다. 마지막으로 4장에서 결론 및 고찰에 관하여 논의한다.

## 2. 이론적 배경

### 2.1. 신뢰도 연구

신뢰도 연구란 한 명의 평가자가 연구 표본을 반복 측정하거나 여러 명의 평가자가 한 표본을 평가할 때 일치하는 정도를 알아보고자 하는 연구로, 새로운 검사 방법이 기존의 방법과 일치하는지 혹은 같은 표본에서의 측정이 일관된 결과를 제공하는지 등을 알아보는 경우에 주로 사용한다.

### 2.2. ICC의 정의

ICC란 신뢰도 연구에서 측정치가 정량적일 때 신뢰도를 나타내기 위해 쓰이는 상관 계수로, 한 표본에 대해 측정한 도구나 평가자 간 상관을 의미한다 (Shrout와 Fleiss, 1979). ICC가 클수록 높은 신뢰도를 나타내는데, 그 값에 대한 명확한 기준은 없으나 Landis와 Koch (1977)의 분류가 일반적이다 (표 2.1). 이 분류가 절대적이라고 할 수는 없지만 유용하다고 판단되어 널리 쓰이고 있다 (Donner와 Eliasziw, 1987). 표 2.2는 앞으로 비교분석을 위해 쓰일 변수의 표기법을 설명한다.

### 2.3. 모형

**2.3.1. 임의효과 일원배치 모형** One-way 모형은 가장 단순한 모형으로 표본 효과만을 고려한다. 모집단에서 임의 추출된 서로 다른 $k$개의 평가 도구로 각 표본을 평가하므로 $n$개의 표본에 대해 각각 $k$개

표 **2.2.** 변수의 설명

| 기호 | 의미 |
|---|---|
| $w$ | 신뢰구간 폭 |
| $n$ | 표본수 |
| $k$ | 반복수 |
| $N$ | 총 측정수$(= n * k)$ |
| $\rho$ | 급내상관계수 ICC |

표 **2.3.** One-way ANOVA 표

| 요인 | 자유도 | 평균제곱합(MS) | E(MS) |
|---|---|---|---|
| Between | $n-1$ | BMS | $k\sigma_t^2 + \sigma_e^2$ |
| Within | $n(k-1)$ | WMS | $\sigma_e^2$ |

의 관찰치가 얻어지고, 모형식은 다음과 같다 (Shrout와 Fleiss, 1979).

$$Y_{ij} = \mu + t_i + e_{ij}, \quad i = 1, \ldots, n, \ j = 1, \ldots, k,$$

여기서 $\mu$는 전체 평균이고 $t_i$는 표본 효과이다. 임의효과인 $t_i$는 평균이 0이고 분산이 $\sigma_t^2$인 정규분포를 따르고, 측정 오차 $e_{ij}$는 평균이 0이고 분산이 $\sigma_e^2$인 정규분포를 따르며 $t_i$와 $e_{ij}$는 서로 독립이라고 가정한다. 또한 위의 모형을 이용한 ANOVA 표는 표 2.3과 같다.

$$\bar{Y}_{i\cdot} = \frac{\sum_{i=1}^{n}(Y_{ij})}{k}, \quad \bar{Y}_{\cdot\cdot} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n}(Y_{ij})}{nk}, \quad \text{BMS} = \frac{\sum_{i=1}^{n} k\left(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}\right)^2}{n-1}, \quad \text{WMS} = \frac{\sum_{i=1}^{n}\left(Y_{ij} - \bar{Y}_{i\cdot}\right)^2}{n(k-1)}$$

BMS = Between Mean Square, WMS = Within Mean Square.

가설 $H_0 : \sigma_t^2 = 0$ *vs.* $H_1 : \sigma_t^2 > 0$을 검정하기 위한 통계량으로 $F = \text{BMS}/\text{WMS}$를 사용한다. 이 값이 $F > F_{n-1,n(k-1),1-\alpha}$이면 귀무가설을 기각한다.

One-way 모형에서 ICC는 전체 분산에 대한 관심효과 분산의 비 $\rho = \sigma_t^2/(\sigma_t^2 + \sigma_e^2)$로 정의되고, 표본 ICC는 $\hat{\rho} = (\text{BMS} - \text{WMS})/\{\text{BMS} + (k-1)\text{WMS}\}$가 된다. ICC의 점추정치는 $\rho = \max[\sigma_t^2/(\sigma_t^2 + \sigma_e^2), 0]$가 되고, 근사 양측 $100(1-\alpha)\%$ 신뢰구간의 하한과 상한은 각각 $\max[(F/F_{n-1,n(k-1),1-\alpha/2} - 1)/(k + F/F_{n-1,n(k-1),1-\alpha/2} - 1), 0]$, $\max[(F/F_{n-1,n(k-1),\alpha/2} - 1)/(k + F/F_{n-1,n(k-1),\alpha/2} - 1), 0]$이다 (Rosner, 2005).

### 2.4. ICC와 관련한 표본수 결정에 관한 방법들

**2.4.1. Walter의 방법** One-way 모형에서 반복수를 알고 있고 고정된 검정력으로 가설을 검정하려 할 때 요구되는 표본수에 대한 방법을 제공하였다. Donner와 Eliasziw (1987)가 제안한 ICC를 관심모수로 하는 신뢰도 연구에서 $H_0 : \rho \le \rho_0$ *vs.* $H_1 : \rho > \rho_0$를 검정하기 위해 요구되는 표본수와 반복수를 근사적으로 구하는 방법이다. 이를 정해진 유의수준과 검정력에서 위의 가설을 검정하려고 한다. $C = 1 + \{k\rho_0/(1-\rho_0)\}$이고 $H_0$하에서 특정 $\rho$인 $\rho_0$는 $\rho$의 최소값에 따라 바뀐다고 할 때 검정통계량의 기각역은 $CF_{\alpha;v_1,v_2}$가 된다. $C_0 = [1 + k\{\rho_0/(1-\rho_0)\}]/[1 + k\{\rho_1/(1-\rho_1)\}]$이고, $H_1$하에서 특정 $\rho$는 $\rho_1$이라고 할 때 검정력은 $1 - \beta = \Pr\{F \ge C_0 F_{\alpha;v_1,v_2}\}$이다. $F_{\alpha;v_1,v_2}$는 자유도 $n-1$, $n(k-1)$를 갖는 누적 $F$ 분포의 $100(1-\alpha)\%$인 점이다. Fisher의 방법을 사용하여 정규분포로 변환하면 $z = 0.5(\ln F) \sim N(\mu_z, \sigma_z^2)$가 되고 여기서 $\mu_z = 1/2(1/v_2 - 1/v_1) = (2n - nk - 1)/\{2n(k-1)(n-1)\}$,

$\sigma_z^2 = 1/2(1/v_2 + 1/v_1) = (nk-1)/\{2n(k-1)(n-1)\}$이다. 이에 따라 기각역 $z^*$는 $U_\alpha$가 누적 정규분포의 $100(1-\alpha)\%$ 지점일 때 $1/2(1/v_2 - 1/v_1) + U_\alpha[1/2(1/v_2 + 1/v_1)]$가 되고, 검정력은 $\Pr\{F \geq C_0 \exp(2z^*)\}$가 된다 (Johnson과 Kotz, 1970). 이를 조합하여 표본수 산출 방법이 식 (2.1)과 같이 도출된다. 이를 통해 귀무가설과 대립가설 하에서의 특정 $\rho$를 알 때, 주어진 유의수준과 검정력에서 반복수 $k$에 따른 표본수 $n$을 구할 수 있다. 다른 방법들과의 공평한 비교를 위해 $F$ 분포를 정규분포에 근사시켜 표본수, 반복수와 신뢰구간 폭을 얻었다.

$$n = 1 + \frac{2\left(U_\alpha + U_\beta\right)^2 k}{\left(\ln C_0\right)^2 (k-1)}. \tag{2.1}$$

**2.4.2. Giraudeau의 방법** One-way 모형에서 반복수와 표본수의 곱으로 표현되는 총 측정수가 고정되어 있을 때, 표본수와 반복수의 조합에 따른 신뢰구간 폭을 구하는 방법을 제시하였다. Shrout와 Fleiss (1979)가 제안한 $\hat{\rho}$에 대한 정확한 $100(1-\alpha)\%$ 신뢰구간에 대한 근사 방법으로, 정확한 식을 살펴보면 자유도 $n-1$, $n(k-1)$인 누적 $F$분포에서 $F_L$과 $F_U$이 각각 $100(\alpha/2)\%$, $100(1-\alpha/2)\%$인 지점이고 $F = \text{BMS}/\text{WMS}$일 때, 신뢰구간이 $[(F/F_U - 1)/\{F/F_U + (k-1)\}, (F/F_L - 1)/\{F/F_L + (k-1)\}]$가 된다. 근사식은 정확한 식에 테일러 급수(Taylor series)와 Fisher의 변환 방법을 적용하여 식 (2.2)와 같이 산출할 수 있다. 표본수 $n$과 반복수 $k$ 그리고 $\rho$를 알 때 신뢰구간 폭 $w$에 대한 식이다. $z$는 누적 정규 분포에서 $100(1-\alpha/2)\%$인 점이다.

$$w = 2\sqrt{2} z_{1-\frac{\alpha}{2}} \frac{\{1 + (k-1)\rho\}(1-\rho)}{\sqrt{nk(k-1)}}. \tag{2.2}$$

**2.4.3. Bonett의 방법** ICC의 근사 분산과 신뢰구간 폭을 구하는 방법과 목표 $\rho$가 있을 때 필요한 근사 표본수에 대한 방법을 제공한다. 얻고자 하는 $\rho$에 대한 구체적인 정보는 사전 조사나 전문가의 의견으로부터 얻을 수 있다. One-way와 two-way 모형에 모두 적용할 수 있는 반복수에 따른 표본수 결정 방법에 대한 근사 closed-form을 제시하였다. 여기서 제시한 근사 방법은 Fisher (1954)가 제안한 방식으로 얻었고, 표본수 $n$이 적당히 클 때($n \geq 30$) 적용할 수 있다 (Donner와 Koval, 1983). $\hat{\rho}$의 분산 $\text{Var}(\hat{\rho}) = [2(1-\rho)^2\{1 + (k-1)\rho\}^2]/\{k(k-1)(n-1)\}$을 $w = 2z_{1-\alpha/2}\sqrt{\text{Var}(\hat{\rho})}$에 대입하여 신뢰구간 폭 $w$를 구할 수 있다. 표본수 $n$과 반복수 $k$ 그리고 $\rho$의 정보가 있을 때 구할 수 있다. 또 표본수를 얻기 위한 식을 유도하려면 식에서 $\rho$대신 얻고자 하는 목표값 $\tilde{\rho}$를 넣어서 다음의 식 (2.3)과 같이 구한다.

$$n = 8z_{1-\frac{\alpha}{2}}^2 \frac{(1-\tilde{\rho})^2\left\{1 + (k-1)\tilde{\rho}\right\}^2}{k(k-1)w^2} + 1. \tag{2.3}$$

**2.4.4. Saito의 방법** Two-way 모형에서 단일 평가자인 ICC(single-rater ICC)를 가정할 때, 반복수에 따른 분산을 구하는 근사식 그리고 표본수와 반복수의 최적의 조합에 대한 방법을 제시하였다. 표본 ICC에 대해 단조적으로 증가하는, 로그를 취한 표본 ICC의 분산을 최소로 하는 반복수에 대한 방법을 델타 방법($\delta$-method)을 이용하여 도출하였다. 테일러 급수(Taylor series)를 사용하여 $\text{Var}(\log(\hat{\rho})) \approx \text{Var}(\hat{\rho})/\hat{\rho}^2$로 나타낼 수 있다. 표본수 $n$과 반복수 $k$의 곱인 총 측정수 $N$을 알 때 반복수 $k$에 대한 식으로 표현할 수 있다. $\phi = \sigma_r^2/(\sigma_t^2 + \sigma_r^2 + \sigma_e^2)$, $\epsilon = \sigma_e^2/\sigma_r^2$일 때, 식은 다음과 같다.

$$\text{Var}\left(\log\left(\hat{\rho}\right)\right) = 2\frac{ak^2 + bk + c}{k^2 - (N+1)k + N}, \tag{2.4}$$

$$a = -\phi^2 - \frac{N-1}{N}\phi^2\epsilon(2+\epsilon)$$

$$b = 2\phi^2\left(1 - \frac{1}{\rho}\phi\epsilon\right) - \frac{2}{N\rho}\phi^3\epsilon^2\left[(2+\epsilon)N - (1+\epsilon)\right]$$
$$c = -N\phi^2 + \frac{2}{\rho}\phi^3\epsilon(1+\epsilon) - \frac{1}{\rho^2}\phi^4\epsilon^2(1+\epsilon)^2.$$

Saito 원저에는 $\sigma_r^2/\sigma_e^2$로 정의되는 상대비 $R$이 0∼10000까지 여러 범주로 나와 있으나, one-way 모형에 해당되는 경우는 $\sigma_r^2 = 0$일 때이므로 본 논문에서는 이 경우만 고려한다.

## 3. 요인 변화에 따른 방법 비교

### 3.1. 개요

3장에서는 one-way 모형에서 2.4절에서 소개한 네 가지 방법을 세 가지로 구분하여 접근하였다. 표본수나 반복수에 대한 방법을 제시하거나 정도를 보기 위해 신뢰구간 폭에 대한 방법을 제시하였다. 표본수나 반복수 접근 방법은 가설검정에서 다른 조건을 통제했을 때 표본수나 반복수가 얼마나 필요한지가 주 관심사이고, 신뢰구간 폭 접근 방법은 고정된 유의수준에서 방법이 얼마나 정확한지를 보고자 한다. 신뢰구간 폭은 ICC의 분산과 비례 관계에 있는데, 분산이 작다는 것은 오차와 표준편차가 작은 것으로 정도가 높다고 할 수 있다. 즉 3장에서는 제한된 조건에서 더 작은 수의 표본이나 반복을 필요로 하는 방법 그리고 더 좁은 신뢰구간 폭을 갖는 방법을 요인을 변화시켜 가면서 찾는다. ICC는 표 2.1에 제시한 Landis와 Koch가 제안한 신뢰도 계수 분류에 따른다. 또한 방법 간 비교에서 ICC는 표본수를 구할 때의 신뢰도 정도가 'moderate' 이상을 고려하므로 0.5∼0.8로 한정하였다. 고정된 유의수준 외 변화시키는 변수의 범위는 기존 논문을 바탕으로 하는 것을 원칙으로 하고 현실성도 감안하여 설정하였다.

### 3.2. 표본수

같은 조건에서 더 작은 수의 표본을 필요로 하는 방법을 찾기 위한 목적으로, 비교 결과를 통해 반복수가 미리 정해져있을 때 효율적으로 표본수를 결정할 수 있다. 표본수를 구하기 위해서는 $w, k, \rho$ 그리고 $\alpha$값이 필요하다. 기존의 논문을 바탕으로 요인의 값을 설정하여 반복수는 2, 3, 4, 5, 10으로, 신뢰구간 폭은 0.2, 0.3, 0.4로 한정하여 각 경우에 대해 필요한 표본수를 조사하였다.

**3.2.1. 결과** 반복수가 많아지고 신뢰구간 폭이 넓어질수록 그리고 $\rho$가 커질수록 필요한 표본수는 작아지게 된다. 다른 방법들과는 다른 변화 양상을 보이는 Saito의 방법을 제외한 나머지 세 방법에서 얻어진 표본수의 비교는 그림 3.1∼3.5에 나타난다. 그림 3.1∼3.5는 반복수가 각각 2, 3, 4, 5, 10일 때 신뢰구간 폭이 넓어짐에 따른 표본수의 변화를 방법 간 비교한 것이다. Walter와 Bonett의 방법은 대체적으로 비슷한 값을 보이는데, 반복수가 작고 신뢰구간 폭이 큰 경우에 Walter의 방법이 Bonett의 방법보다 조금 더 작은 수의 표본을 필요로 하였다. 신뢰구간 폭이 넓어지고 반복수가 많아질수록 Walter와 Bonett의 방법에서 나온 결과가 같아진다. 방법 간 비교 결과를 살펴보면 Giraudeau의 방법이 같은 조건에서 가장 작은 수의 표본을 필요로 하고 Walter의 방법, Bonett의 방법 순으로 많은 수의 표본을 필요로 하였다.

### 3.3. 반복수

반복수는 같은 조건에서 더 작은 수의 반복수를 얻기 위함을 목적으로 표본수 외 조건들이 미리 정해져 있을 때 몇 번의 반복이 필요한지에 대한 정보를 제공한다. 신뢰구간 폭은 0.2에서 0.4까지, 표본수는 20부터 120까지 변하는 조건에서 반복수를 조사했다. $n$, $w$, $\rho$ 그리고 $\alpha$에 대한 사전 정보가 필요하다.

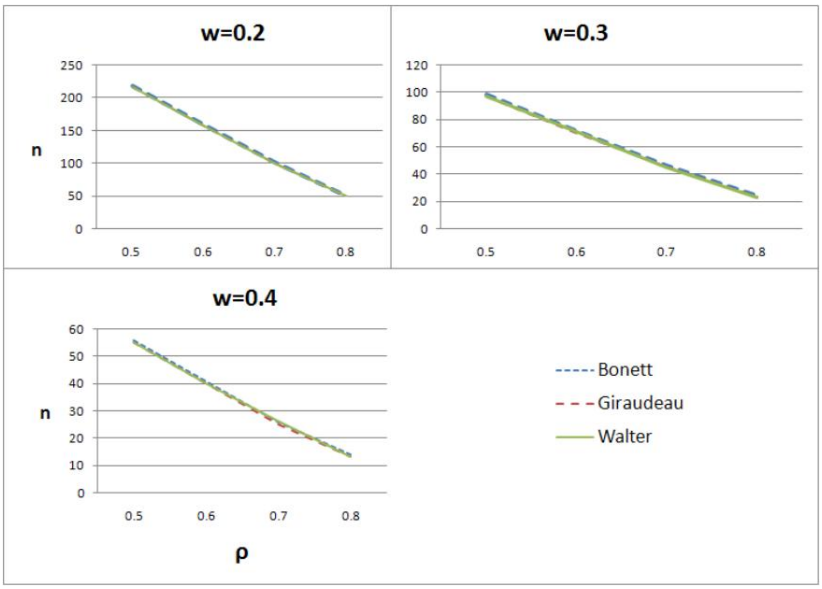


그림 3.1. 반복수가 2일 때, 방법 간 표본수 비교

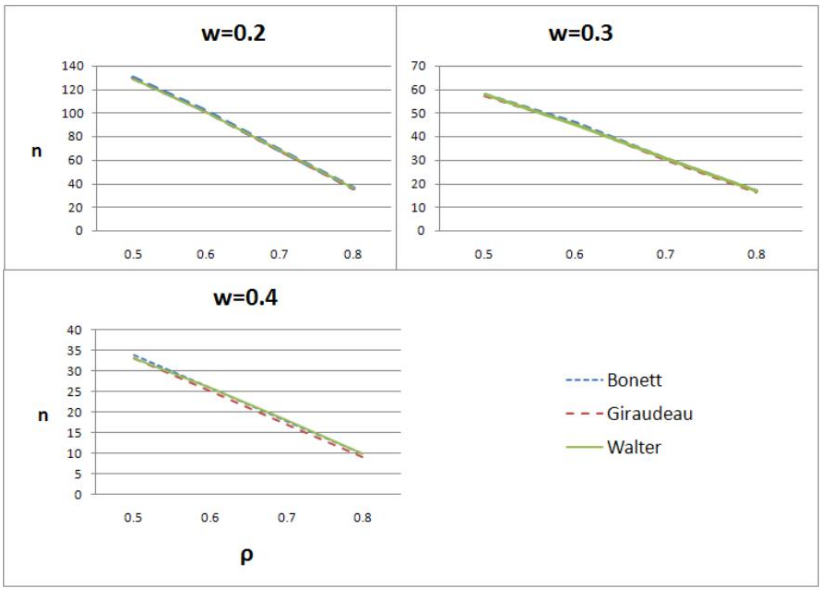


그림 3.2. 반복수가 3일 때, 방법 간 표본수 비교

**3.3.1. 결과** 반복수를 구할 때 표본수와 관련한 제한 조건이 생기는데 예를 들어, 표본수보다 많은 수의 반복을 필요로 하는 결과는 쓸모가 없으므로 타당한 반복수를 얻기 위해서는 표본수와의 관계를 고려해야 한다. 신뢰구간 폭이 좁고 표본수가 작은 경우에 얻어지는 반복수는 매우 큰 값으로 얻어지게 된다. 이러한 이유로 신뢰구간 폭에 따라 필요한 최소 표본수에 차이가 난다. 신뢰구간 폭이 0.2일 때 표

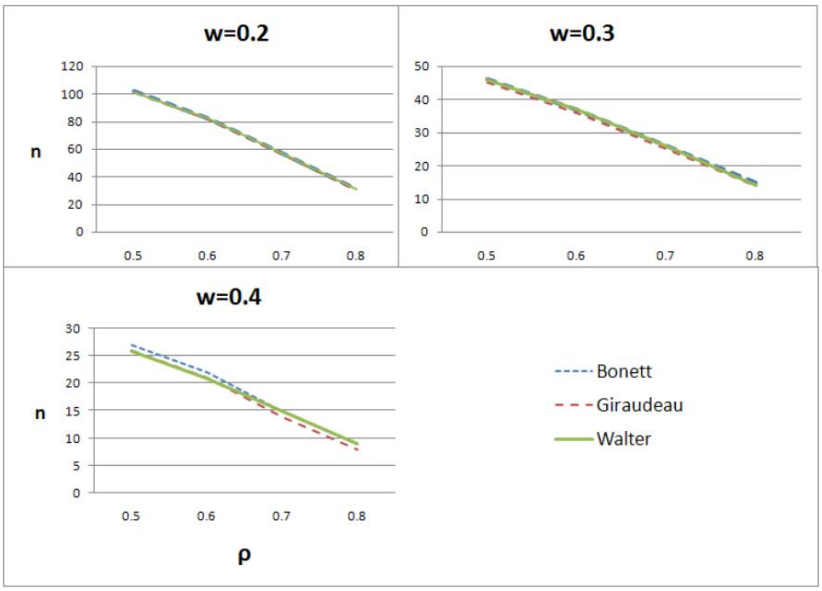


**그림 3.3.** 반복수가 4일 때, 방법 간 표본수 비교

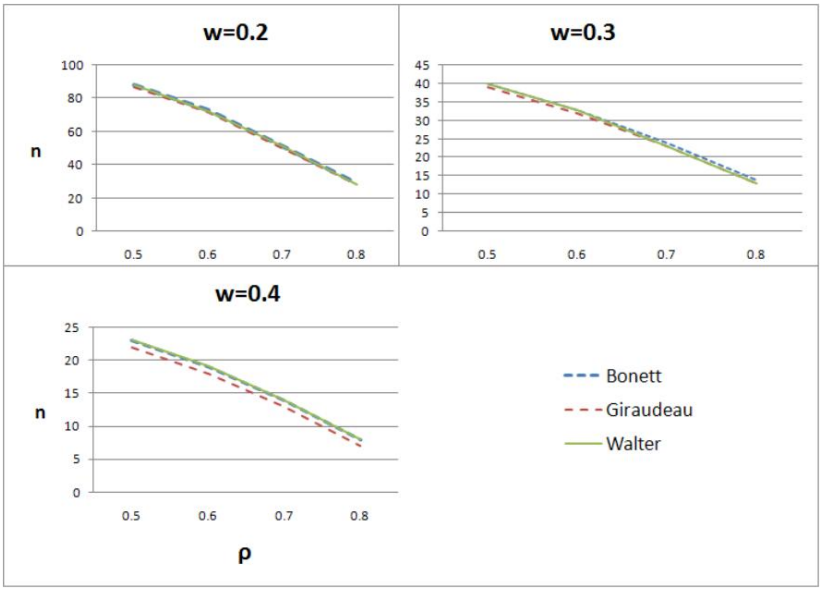


**그림 3.4.** 반복수가 5일 때, 방법 간 표본수 비교

본수는 60이상, 0.3일 때는 30이상이어야 실현 가능한 반복수를 구할 수 있다.

표본수가 많아지고 신뢰구간 폭이 넓을수록, 그리고 $\rho$가 커질수록 필요한 반복수는 작아진다. 그림 3.6~3.10은 표본수가 각각 30, 40, 60, 80, 100일 때 신뢰구간 폭과 $\rho$에 따른 방법 간 반복수 변화의 차이를 보여준다. 폭이 넓어지고 표본수가 많아질수록 방법들 간 차이는 감소하며, $\rho$에 따른 변화가 무

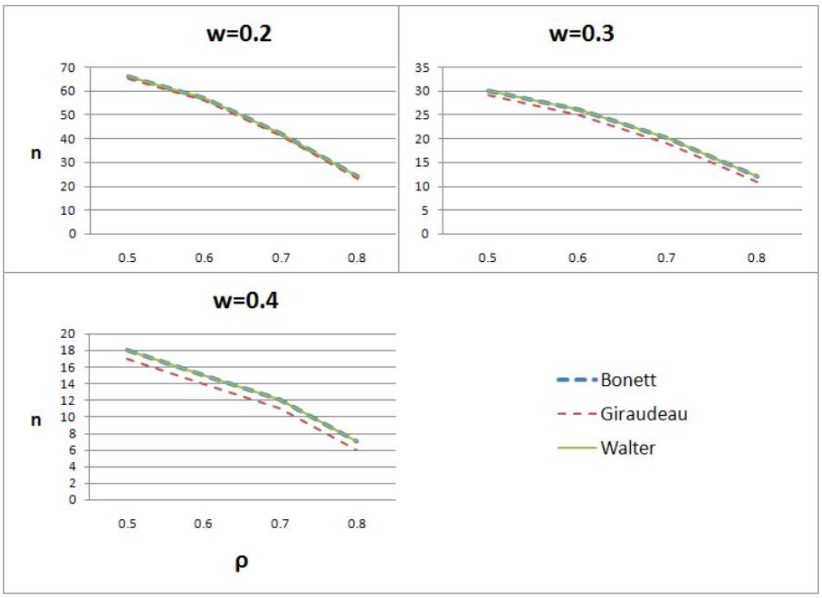


그림 **3.5.** 반복수가 10일 때, 방법 간 표본수 비교

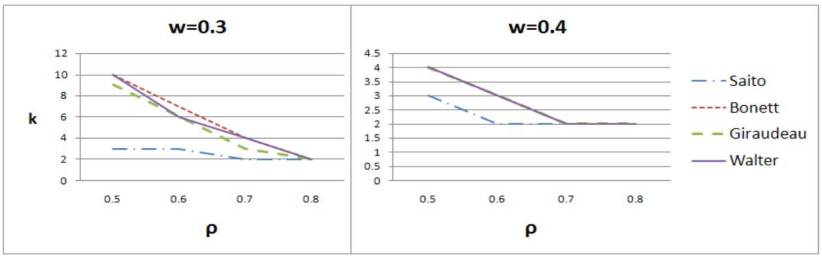


그림 **3.6.** 표본수가 30일 때, 방법 간 반복수 비교

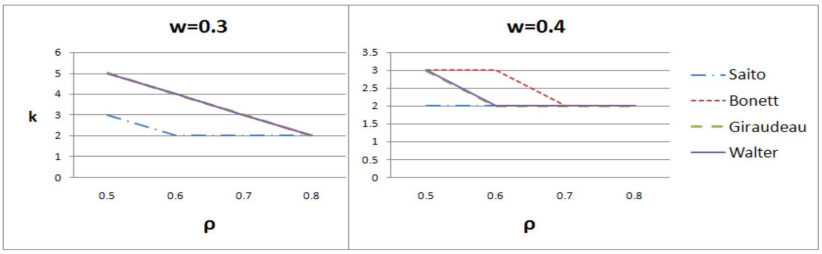


그림 **3.7.** 표본수가 40일 때, 방법 간 반복수 비교

더진다. 다른 방법들은 신뢰구간 폭, 표본수, $\rho$의 영향을 많이 받는 반면에 Saito의 방법은 신뢰구간 폭과 $\rho$에 크게 영향을 받지 않는다. 신뢰구간 폭이 작거나 표본수가 작은 경우 Saito의 방법을 제외한 나머지 방법에서 얻어진 반복수는 매우 큰 값을 나타내는 것을 볼 수 있다. Saito의 방법은 다른 방법에 비해 신뢰구간 폭과 표본수, $\rho$에 대해 견고하게 나타난다. Walter의 방법이 신뢰구간 폭, 표본수나 $\rho$에 가장 민감하게 반응하고 Bonett, Giraudeau, Saito의 방법 순으로 요인에 따른 변동이 줄어든다.

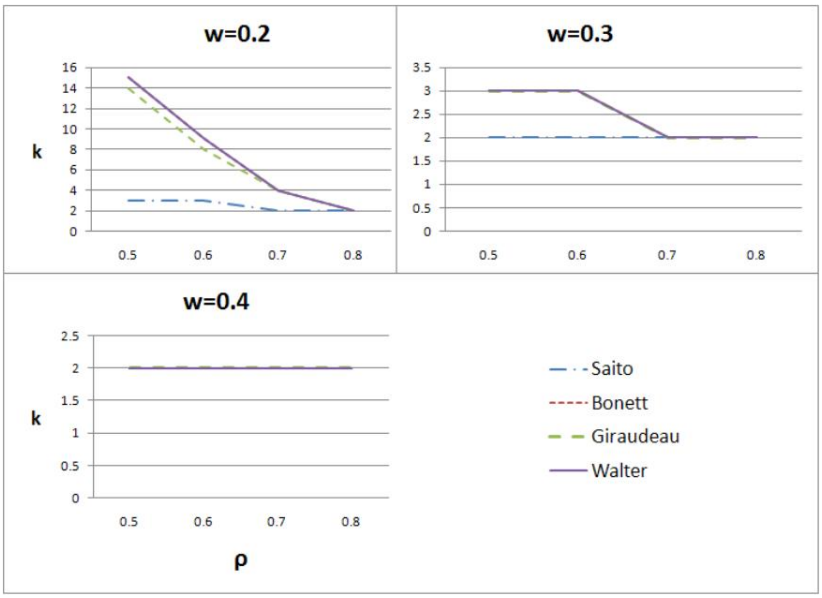


그림 3.8. 표본수가 60일 때, 방법 간 반복수 비교

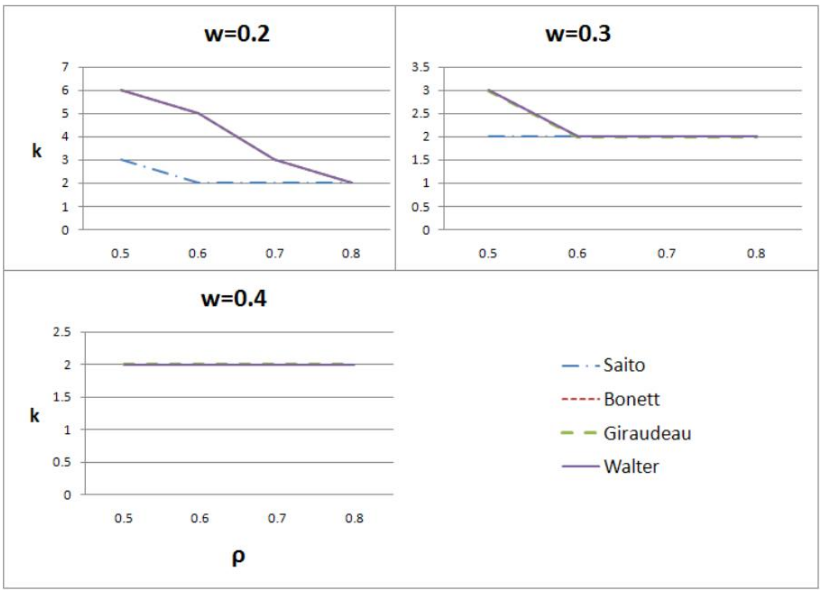


그림 3.9. 표본수가 80일 때, 방법 간 반복수 비교

### 3.4. 신뢰구간 폭

신뢰구간 폭은 어떤 방법이 보다 더 높은 정도를 갖는지 알아보기 위한 것으로, 같은 조건에서 더 좁은 폭을 가지는 방법이 더 높은 정도를 갖는다고 할 수 있다. 비교할 조건은 반복수와 표본수의 곱인 총 측정수가 20, 60, 120인 경우에 가능한 반복수와 표본수의 조합으로 설정하였다. 신뢰구간 폭에 대한 방법

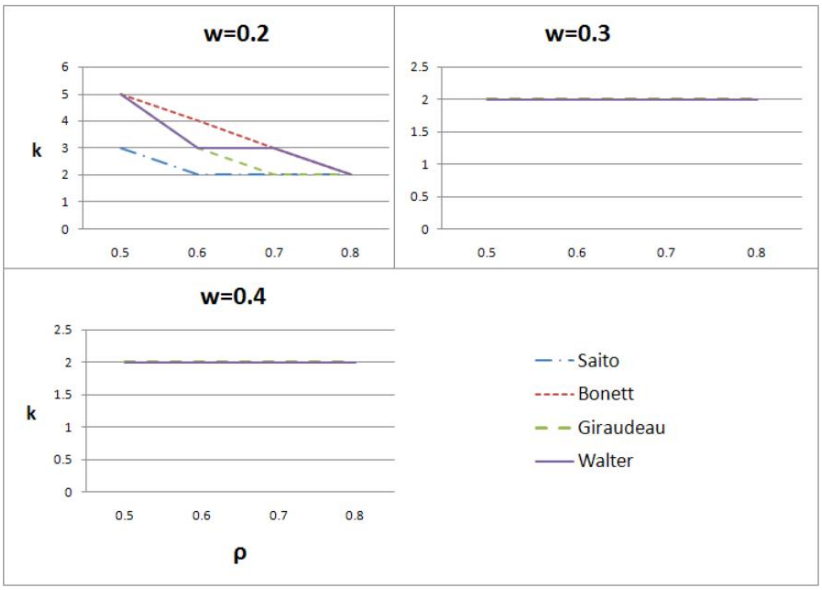


그림 **3.10.** 표본수가 100일 때, 방법 간 반복수 비교

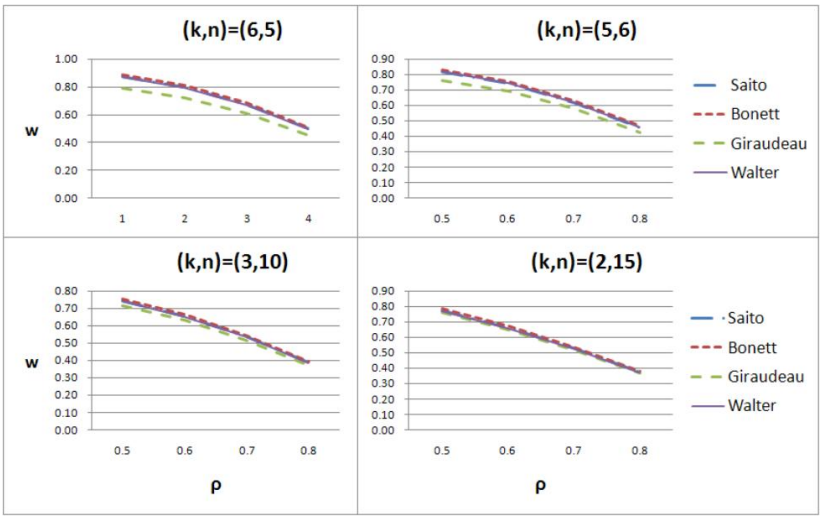


그림 **3.11.** 총 측정수가 30일 때, 방법 간 신뢰구간 폭 비교

을 계산하기 위해서는 네 방법에서 공통으로 $k$, $\rho$, $\alpha$에 대한 정보가 있어야 하고 추가적으로 Walter나 Bonett의 방법에서는 총 측정수 $N$에 대한, Giraudeau나 Saito의 방법에서는 표본수 $n$에 대한 정보가 필요하다.

**3.4.1. 결과** 총 측정수가 고정일 때 표본수와 반복수의 조합과 $\rho$의 증가에 따른 신뢰구간 폭의 변화를 보면 $\rho$와 반복수에 반비례하고 표본수에 비례하는 추세를 보인다. 네 가지 방법에서 얻어진 결과를 보면 총 측정수가 30처럼 작은 경우에는 신뢰구간 폭이 너무 넓게 나오는 한계점을 그림 3.11을 통해 알

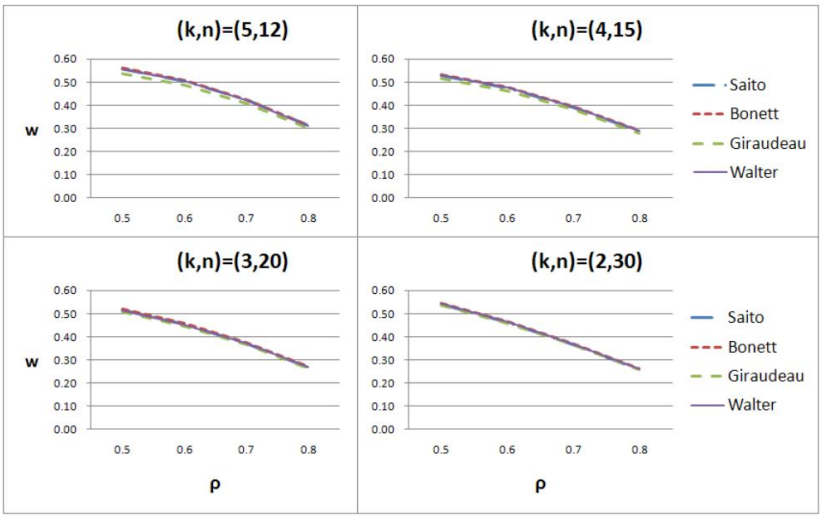


그림 **3.12.** 총 측정수가 60일 때, 방법 간 신뢰구간 폭 비교

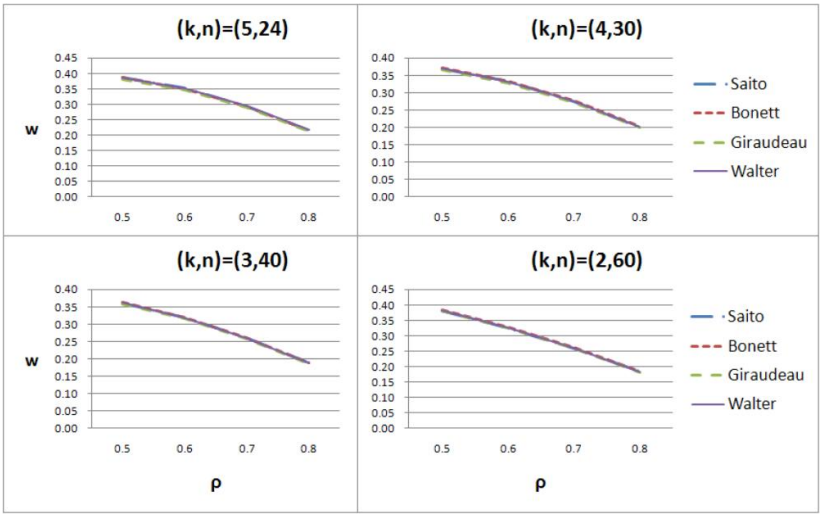


그림 **3.13.** 총 측정수가 120일 때, 방법 간 신뢰구간 폭 비교

수 있다. 네 가지 방법에서 얻은 반복수가 거의 비슷하게 나타나는 가운데, Giraudeau의 방법이 가장 좁은 폭의 신뢰구간을 제공해주어 가장 높은 정도를 나타내었다. Walter, Saito, Bonett의 방법 순으로 정도가 떨어진다. 조합 내에서 반복수가 많아질수록 혹은 $\rho$가 작을수록 Giraudeau의 방법이 다른 방법들에 비해 더 높은 정도를 가진다.

## 4. 고찰 및 결론

본 논문은 신뢰도 연구에서 ICC와 관련하여 표본수, 반복수 및 평가자수의 결정에 대해 연구하였다. 실제 임상연구에서 이에 관한 결정은 중요한 고려사항이 되는데 그 이유는 너무 작게 설정하면 검정력이

나 정도가 떨어져 신뢰도가 부족하게 되고 너무 크면 자원의 낭비를 초래할 수 있기 때문이다 (Bonett, 2002). 본 논문에서는 기존에 제시되어 있는 방법들을 다양하게 비교하여 연구자가 더 나은 방향으로 의사결정을 할 수 있는 지침을 제공하고자 하였다.

요인을 변화시키면서 방법들의 표본수와 반복수 그리고 신뢰구간 폭을 비교하였다. 원하는 정보를 얻기 위해 통제해야 할 다른 조건들은 기존의 논문을 바탕으로 설정하되 현실성도 또한 고려하여 반영하였다. 신뢰도 연구에서 몇 명이 몇 명을 몇 번 관찰해야 하는지에 대한 정보를 한 번에 모두 얻을 수 있는 방법은 없다. ICC와 관련하여 표본수나 반복수 및 평가자수에 대한 개별 정보를 얻으려면 나머지에 대한 가정이 필요하고 가정은 곧 제한 조건이 된다. 모든 방법에는 미리 정해야 하는 변수에 대한 수적 제한이 따르게 되며 이 외에도 방법이나 접근마다 장단점이 생기게 된다. Walter의 방법은 비교적 간단하게 표본수를 계산할 수 있지만 다른 방법들과 달리 귀무가설 그리고 대립가설 하에서의 ICC와 제 2종 오류를 미리 정해야하는 현실적 제약이 있고 $F$분포를 가정하였다. 다른 방법들과의 공평한 비교를 위해 $\beta$의 정보 없이도 구할 수 있도록 정규분포로 근사시켜 표본수와 반복수와 신뢰구간 폭을 구하였다. Giraudeau의 방법은 원래부터 one-way 모형을 가정하였고 별다르게 요구되는 조건이 없어 다른 방법들과의 비교가 수월하였다. Bonett의 방법은 closed-form을 제시해주어 어느 모형에나 적용할 수 있어 비교적 간단하고 쓰임새가 간편해 보였지만 실제로 비교를 해본 결과, 가장 많은 표본수와 반복수를 요구하였다. 마지막으로 Saito의 방법은 근사식인데도 불구하고 계산 방법이 복잡하였고 two-way 모형을 전제로 한 방법이어서 다른 방법들과 다른 변화 양상을 보였으며 표본수를 구하는 데 어려움이 있었다.

요인을 변화시켜 가면서 실험해 본 결과 연구의 목적에 따라 최선의 방법이 다르게 나타났다. 여기서 말하는 최선의 방법이란 표본수나 반복수의 측면에서는 가장 적은 수만 필요로 하는 방법이고 정도 측면에서는 신뢰구간 폭이 가장 좁게 나타나는 방법을 말하는 것으로 예를 들어, 임상 의학 연구에서 평가할 의사의 수 혹은 방법의 수가 정해져 있을 때 최소 몇 명의 환자를 대상으로 해야 일정 수준의 신뢰도를 갖는지는 3.2장의 표본수 비교를 통해 알 수 있고, 반대로 환자수가 미리 정해져있을 때 환자를 최소 몇 번 평가해야 일정 수준의 신뢰도를 갖는지에 대한 정보는 3.3장의 반복수에 대한 비교에서 정보를 얻을 수 있다. 전자의 경우인 반복수와 신뢰구간 폭이 고정일 때 표본수를 구하려는 목적의 연구를 한다면 Giraudeau의 방법에서 얻어진 값이 가장 작았다. 후자의 경우인 표본수와 신뢰구간 폭이 고정일 때 반복수를 구하는 목적의 연구라면 Saito의 방법에서 얻어진 값이 작게 나타났다. 특히 Saito의 방법은 신뢰구간 폭과 표본수가 작을 때도 크게 영향을 받지 않는 견고함을 보였다. 이는 Saito의 방법이 반복수를 구함에 있어 식 (2.4)에 언급한 것처럼 모수 $R$, $\epsilon$, $\phi$의 영향을 받아 다른 방법들보다 $\rho$의 영향을 덜 받기 때문이라고 판단된다. 아울러 이러한 결과는 $\rho$값이 작고 신뢰구간 폭이 좁은 경우 Saito의 방법을 이용해서 구한 반복수가 보다 더 현실적인 값이라고 볼 수 있다. 정도 면에서는 Giraudeau의 방법이 가장 좁은 신뢰구간 폭을 가졌고 Walter, Saito, Bonett의 방법 순으로 신뢰구간 폭이 넓어져 정도가 떨어졌다. 총 측정수 조합 내 반복수가 많거나 $\rho$가 작아지는 경우 Giraudeau의 방법의 정도가 우월하였다.

본 연구에서는 신뢰도 연구를 할 때 어떤 조건이 주어져있고 무엇을 구하려는지 상황과 목적에 따라 적절한 방법을 이용하여 보다 더 효율적인 연구를 설계할 수 있기 위해 방법들을 비교분석하였다. 기존 연구에서의 제한점은 먼저, 실제 연구에서 표본수가 작은 경우가 빈번한데 비해 이 때 적절한 반복수를 구하기 어렵다는 점이었다. 둘째로, 네 가지 방법 모두에서 신뢰구간 폭을 구할 때 총 측정수가 적당히 크지 않으면 폭이 너무 넓게 나타나는 한계점이 드러났다. 마지막으로, 어떤 하나의 요인에 대한 정보를 얻기 위해 사전에 가정해야 할 모수들이 많다는 점이다. 그러므로 향후에는 작은 표본수에 대하여 적절한 반복수를 구할 수 있는 방법과 더 적은 양의 정보로 원하는 요인을 구할 수 있는 방법에 대한 연구가 필요할 것으로 여겨진다.

## 참고문헌


Bonett, D. G. (2002). Sample size requirements for estimating intraclass correlations with desired precision, *Statistics in Medicine*, **21**, 1331–1335.

Donner, A. and Eliasziw, M. (1987). Sample size requirements for reliability studies, *Statistics in Medicine*, **6**, 441–448.

Donner, A. and Koval, J. J. (1983). A note on the accuracy of Fisher's approximation to the large sample variance of an intraclass correlation, *Communications in Statistics-Computation and Simulation*, **12**, 443–449.

Fisher, R. A. (1954). *Statistical Methods for Research Workers*, 12th ed. Hafner, New York.

Giraudeau, B. and Mary, J. Y. (2001). Planning a reproducibility study, how many subjects and how many replicates per subject for an expected width of the 95 per cent confidence interval of the intraclass correlation coefficient, *Statistics in Medicine*, **20**, 3205–3214.

Johnson, N. L. and Kotz, S. (1970). *Distributions in Statistics, Continuous Univariate Distrubution 2*, John Wiley & Sons, Inc.

Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data, *Biometrics*, **33**, 159–174.

Saito, Y., Sozu, T., Hamada, C. and Yoshimura, I. (2006). Effective number of subjects and number of raters for inter-rater reliability studies, *Statistics in Medicine*, **25**, 1547–1560.

Shrout, P. E. and Fleiss, J. L. (1979). Intraclass Correlation, uses in assessing rater reliability, *Psychological Bulletin*, **86**, 420–428.

Rosner, B. (2005). *Fundamentals of Biostatistics*, 6th ed. Thomson Brooks/Cole.

Walter, S. D., Eliasziw, M. and Donner, A. (1998). Sample size and optimal designs for reliability studies, *Statistics in Medicine*, **17**, 101–110.

White, S. A. and Broek, N. R. (2004). Methods for assessing reliability and validity for a measurement tool, a case study and critique using the WHO Haemoglobin Colour Scale, *Statistics in Medicine*, **23**, 1603–1619.

# A Comparison of Sample Size Requirements for Intraclass Correlation Coefficient(ICC)

Han Soo Yeon[1] · Nam Jung Mo[2] · Myoung Sung Min[3] · Song Ki Jun[4]

[1]Department of Biostatistics, Yonsei University College of Medicine
[2]Department of Biostatistics, Yonsei University College of Medicine
[3]Department of Medical Informatics, Jungwon University
[4]Department of Biostatistics, Yonsei University College of Medicine

(Received January 2010; accepted March 2010)

**Abstract**

In medical practice and research, the problem of assessing reliability between two or more quantitative measures is quite common. Intraclass correlation coefficient(ICC) is commonly used to scale of reliability. Some methods were developed to calculate the required number of subjects, raters or replicates in one-way or two-way random ANOVA models. This paper, studies and compares the performance of four methods such as Walter *et al.* (1998), Giraudeau and Mary (2001), Saito *et al.* (2006) and Bonett (2002). In order to compare the efficiency of methods we compare the number of subjects, replicates and the width of confidence interval of ICC needed for some specific ICC values. In the case of subject size, Giraudeau's method is the best. In case of the number of replicates, Saito's method was superior to others. The width of confidence interval of ICC was narrower for Giraudeau's method than any others.

Keywords: Reliability, intraclass correlation coefficient, sample size.

[4]Corresponding author: Research Assistant Professor, Department of Biostatistics, Yonsei University College of Medicine, Seoul 120-752, Korea. E-mail: biostat@yuhs.ac