

인공호흡기 사용 환자들에게 제공된 예비적 정보에 대한 내용분석의 측정자간 신뢰도

김 화 순*

I. 서 론

간호연구에서 관찰자 또는 측정자가 측정오류의 주된 요인으로 작용하는 경우에 대한 인식이 점차 고조되어 왔다. 그리하여 자료수집을 위한 관찰 및 측정에서 관찰자(측정자)간 차이에 대한 평가가 기본요건으로서 강조되고 있다. 간호연구에서 관찰자나 측정자가 측정오류의 주된 요인으로 작용하는 경우로는 환자기록에 근거하여 임상적 판단을 도출하는 내용 분석 연구, 직접적인 관찰을 통하여 간호현상을 측정하는 연구, 환자와 의료인의 상호작용에 관한 연구 등을 들 수 있다. 예를 들면 Banerjee & Fielding(1997)은 장기간 호시실에 거주하는 환자들의 임상간호기록에서 특정 간호진단의 존재유무를 조사하였고, 김과 김(1997)은 심도자 검사 환자에게 제공된 정보의 유형을 분석하였다. Garvin, Huston, & Baker(1992)는 스트레스를 유발하는 사건(심도자 검사)에 대한 정보제공 과정에서 환자와 간호사 사이의 상호작용을 분석하였다. 이러한 연구들은 측정자나 관찰자의 오류에 의해서 그 결과가 크게 달라질 수 있다. 그러므로 측정오류를 최소화하고 연구의 결과에 대한 신뢰성을 높이기 위해서는 측

정자간 또는 관찰자간 판단의 일치도에 대한 평가와 적절한 해석은 필수적이다(Topf, 1986).

내용분석연구나 상호작용 연구들의 측정자간 신뢰도를 확립하기 위한 방법으로는 두 단계의 과정이 있다(Brennan & Hays, 1992; Folger, Hewes, & Poole, 1984). 첫번째 단계는 전체 자료(예를 들면, 환자 기록지 또는 환자와 의료인의 상호작용에 관한 녹음 또는 녹화 자료)에서 부호화 할 단위(분석단위)를 확인하는 것이다. 부호화 단위를 확인하는 단계에서 관찰자들간의 적절한 일치가 필요한데, 이때 관찰자들 간에 분석단위를 일치되게 확인하는 정도를 분석단위 확인 신뢰도(unitizing reliability)라고 한다(Folge, Hewes & Poole, 1984). 이 단계에서 적절한 분석단위 확인 신뢰도를 확보하는 것은 최종 연구 결과의 신뢰성을 높이고 그 다음 단계의 신뢰도 평가를 위한 기본 전제조건이 된다.

그러나 선행연구들은 분석단위 확인 신뢰도에 대해서는 별로 주목하지 않았다. 즉, 대부분의 선행연구자들은 자신들의 자료를 분석 단위로 확인하는 과정이 신뢰성이 있다고 가정할 뿐 구체적인 검증 결과를 제시하지 않았다. 선행연구자들은 적절한 분석단위 확인 신뢰도를 확보하는 것이 전체

* 인하대학교 의과대학 간호학과

연구 결과의 신뢰도를 높이는데 얼마나 중요한가에 대해서 인식하지 못했다(Garvin, Kennedy, & Cissna, 1988).

두번째 단계는 확인된 분석단위를 해당되는 범주로 부호화하는 과정으로 측정자간에 적절한 일치도가 요구되며 이때의 측정자간 일치 정도를 해석적 신뢰도(interpretive reliability)라 한다. 명목자료(nominal data)들의 측정자간 또는 관찰자간 신뢰도를 구하는 절차들로는 agreement, phi, Cohen's kappa 등이 있다(Brennan & Hays, 1992; Topf, 1985). 일치도(agreement)는 단순히 관찰자간 일치정도를 백분율로 나타낸 것이다. Phi계수는 두 관찰자간의 판단에서 동등성을 알고자 할 때 유용하다. 일치도(agreement)와 phi는 관찰자가 두 명 이하인 경우에만 적용이 가능하다. 반면에 kappa는 관찰자가 두 명 이상인 경우에 적용이 가능하며 관찰자간 우연에 의한 일치를 교정한 신뢰도 추정치를 제공한다. 그리하여 간호연구에서 kappa는 측정자간 또는 관찰자간 신뢰도 측정에서 널리 사용되고 있다(Brennan & Hays, 1992). 그러나 Cohen's kappa의 이러한 유용성에도 불구하고 kappa는 몇 가지 문제점이 있는 것으로 지적되고 있다. 예를 들면, kappa는 관찰자간의 편견에 의해서나 측정 범주상 자료의 분포(prevalence)에 의해서 복합적인 방식으로 영향을 받기 때문에 때로는 역설적인 결과를 산출하거나 결과의 해석에 주의를 요하는 경우가 있다(Brennan & Hays, 1992).

따라서 이 논문의 주요 목적은 (1)내용분석이나 상호작용 분석시 적용되어야 하는 측정자간 신뢰도의 산출절차 및 그 유의사항에 대해서 살펴보고, (2)Cohen's kappa 계수의 올바른 사용에 대해 알아보고, (3)실제 적용사례를 통해서 측정자간 신뢰도 추정치의 올바른 해석에 대해 논의하고자 한다.

II. 관련 문헌에 대한 검토

관련 문헌 고찰은 분석 단위 확인 신뢰도, 해석적 신뢰도, Cohen's kappa의 산출에 영향을 미치

는 인자들에 대해서 알아보기로 한다.

1. 분석 단위 확인 신뢰도(Unitizing Reliability)

내용분석이나 상호작용분석에서 측정자간 신뢰도 확립의 첫 단계는 분석단위 확인 신뢰도를 산출하는 단계이다. 분석 단위 확인 신뢰도는 전체 자료중에서 부호화 할 것이 무엇인지를 측정자 또는 관찰자간에 일치되게 확인하고 있는 정도를 나타내는 지표이다(Brennan & Hays, 1992; Garvin, Kennedy, & Cissna, 1988). 분석단위(unit)의 확인이 대단히 구체적이고 명료한 경우에는 분석단위 확인 신뢰도의 보고가 반드시 필요하지는 않다. 예를 들면, Banerjee & Fielding(1997)의 연구에서처럼 분석단위가 간호기록에 존재하는 간호진단인 경우 분석단위의 확인은 객관적이고 구체적이므로 관찰자간에 일치하는 대단히 높을 것이다. 그러나 분석단위가 김과 김(1997)의 연구에서처럼 “분리될 수 있으면서 서로 다른 종류의 정보를 제공하는 단어, 구 또는 문장”인 경우에는 전자의 경우처럼 분석단위가 명료하거나 구체적이지 않다. 그러므로 측정자간 분석단위 확인에서 주관이 개입될 여지가 다소 클 수 있다. 또한 Holaday(1981)의 연구에서는 울음(crying)이 분석단위였는데 이 경우 역시 어느 시점을 울음의 시작으로 볼 것이고 어느 시점에서 울음이 끝난 것으로 볼 것인지를 결정하는 과정에서 관찰자간에 불일치가 다소 클 수 있다. 이러한 경우 관찰자간 분석단위 확인 신뢰도(unitizing reliability)를 검증하고 그 결과를 제시함으로써 전체 연구 결과에 대한 신뢰성을 높일 수 있다. 분석 단위 확인 신뢰도를 구하는 방법은 자료에서 확인된 분석단위의 수에 있어서 관찰자들간의 일치 정도를 백분율로 나타내어 평가할 수 있다.

분석 단위 확인 신뢰도는 여러 가지 요인들에 의해 영향을 받을 수 있다. Kerlinger(1973)에 의하면, 가장 중요한 요인은 분석단위를 확인하는데 있어서 연구자들에게 요구되는 추론의 정도이다. 측정자 또는 관찰자에게 요구되는 추론의 정도가 클수록 분석단위 확인에서 측정자의 주관이 개입

될 여지가 더 커지므로 적절한 수준의 분석 단위 확인 신뢰도를 확립하는데 어려움 또한 더 커질 수 있다. 그러므로 연구자들은 분석 단위 확인 신뢰도를 높이기 위해 더욱 세분화되고(예를 들면, 단어), 더욱 구체화된 분석단위를 선택한다. 분석 단위 확인 신뢰도는 자료의 형태나 유형에 의해서도 영향을 받을 수 있다(Garvin, Kennedy, & Cissna, 1988). 예를 들면, 오디오나 비디오로 녹화/녹음된 자료나 활자화된 자료의 분석은 생생한 현장에서 현상을 직접 관찰을 통하여 분석하는 것 보다는 더욱 용이하다.

일반적으로 관찰자들간에 분석단위 확인에서 80% 이상의 일치율을 보인 경우 신뢰할 만하다고 받아들여지고 있으나(Garvin, 1992), 이러한 기준점의 설정은 분석단위 확인에 있어서 요구되는 추론의 정도를 고려하여 해석하는 것이 더욱 타당할 것으로 보인다. 예를 들면, 분석단위가 비교적 구체적이고 객관적인 경우라면 분석자간 분석단위 확인 신뢰도는 기준점보다 훨씬 더 높아야 할 것이다.

2. 해석적 신뢰도(Interpretive reliability)

내용분석이나 상호작용분석시 측정자간 신뢰도 확립을 위한 다음 단계는 해석적 신뢰도를 산출하는 것이다. 해석적 신뢰도는 임상기록자료나 환자와 의료인의 상호작용이 녹화 또는 녹음된 자료에서 확인된 분석단위들을 해당되는 범주로 부호화시키는 과정이다. 해석적 신뢰도는 확인된 분석단위들을 해당 범주로 부호화시키는 과정에서 측정자간 또는 관찰자간의 일치 정도를 나타내며 Cohen's kappa가 널리 쓰이고 있다. 과거에는 이러한 판정에서 측정자들간의 일치 정도는 단순히 관찰된 일치도(agreement)를 계산하여 평가하였다. 그러나 이러한 통계치는 우연에 의해서도 관찰자들간에 어느 정도의 일치가 가능하다는 사실을 설명하지 못한다. 그러므로 Cohen(1960)은 우연에 의한 일치가 교정된 일치상수(chance-corrected coefficient of agreement)로서 kappa를 제안하였다. 즉, 일치도(agreement)는 부호화 과

정에서 측정자들간에 일치하는 정도를 나타내지만 우연에 의한 일치의 개입을 통제하지 못하기 때문에 얼마간 팽창된(inflated) 추정치를 제공한다.

해석적 신뢰도 산출의 단계에서 대부분의 연구들은 단지 전체적인 측정자간 신뢰도 추정치(global reliability)만을 보고하고 있다(Garvin et al., 1988). 그러나 연구결과의 신뢰성에 대한 올바른 해석을 위해서는 범주별 신뢰도(category-by-category reliability)의 보고가 필요하다. 전체 신뢰도(global reliability)는 전체 분석단위를 부호화하는 과정에서 관찰자들간의 일치정도를 반영한다. 그렇지만 실제로 많은 경우에 있어서 모든 범주가 선택될 확률이 동등하지 못한 경우가 있다. 즉, 어떤 범주는 사용에서 다른 범주보다 더욱 어려운 경우가 있는데 이러한 경우에는 범주별로 관찰자간의 부호화에서 일치정도를 검증하여 제시함으로써 전체 연구결과에 대한 올바른 해석을 도울 수 있다(Garvin et al., 1988). 범주별 신뢰도는 전체 해석적 신뢰도의 산출에서와 같이 kappa를 이용하여 구할 수 있다.

3. Cohen's kappa의 산출과 영향을 미치는 인자들

앞에서 진술한 바와 같이 kappa는 내용분석이나 상호작용분석에서 명목자료의 측정자간 신뢰도 산출에서 널리 이용되고 있다. Kappa를 구하는 방법에는 frequency table(빈도표)를 이용하는 방법과 비율을 이용하는 방법이 있다. 여기서는 편의상 빈도를 이용하여 kappa를 산출하는 방법을 살펴보자. 예를 들면 관찰을 통하여 자료를 수집하는 연구에서 '예' 또는 '아니오'의 두 범주중한 범주로 환자들의 행위를 부호화하는 경우에 두 측정자들간의 일치와 불일치는 2x2 matrix로 나타낼 수 있다(표 1).

이 때 두 관찰자의 판단에서 일치도는 수학적으로는 $p_o = (a+d)/N$ 이다. 이 경우에(a+d) 부분은 환자의 행동에 대한 관찰자들의 판단에서 관찰자들간에 일치정도를 나타내는 부분이다. 그러나 두 관찰자들이 각각 독립적으로 판단을 하는 경우에도 이러한 일치중의 일부는 단지 우연에 의해서

〈표 1〉 임상관찰 연구에서 관찰자간 일치와 불일치의 가상적 자료

		관찰자 1		
		예	아니오	합
관찰자 2	예	a	b	a+b
	아니오	c	d	c+d
	합	a+c	b+d	N=a+b+c+d

발생할 수도 있다. 단지 우연에 의해서 두 관찰자간에 환자의 행동을 일치되게 판단할 것으로 예측되는 정도를 p_e 로 표시하며 수학적으로는 $p_e = [(a+c)(a+b) + (b+d)(c+d)]/N^2$ 이다. 여기서 $p_0 - p_e$ 는 우연에 의한 일치를 배제한 경우의 두 관찰자 사이의 일치도이다. 관찰자들간에 완전한 일치에 있는 경우 즉, $p_0 = 1$ 인 경우, 우연에 의한 일치를 배제한 경우 획득 가능한 최대 일치도는 $1 - p_e$ 이다.

Cohen's kappa 계수는 우연에 의한 일치를 배제한 관찰자 사이의 일치도($p_0 - p_e$)와 우연에 의한 일치를 배제한 경우 획득 가능한 최대 일치도($1 - p_e$)와의 비율이다. 즉, kappa는 수학적으로는 $k = (p_0 - p_e)/(1 - p_e)$ 이다. 그러므로 kappa는 우연에 의한 일치가 제외된 일치도로 해석될 수 있다(Cohen, 1960). 두 관찰자 간에 판단이 완전히 일치하는 경우 kappa는 +1이다. 만약 두 관찰자 간의 판단에서 일치도가 우연에 의한 일치도와 같거나 더 클 때는 kappa는 0과 값이 같거나 더 크다. 두 관찰자간의 판단에서 일치도가 우연에 의한 일치와 같거나 작을 경우 kappa는 0과 같거나 더 작은 값이 된다. Landis & Koch(1977)에 의하면 대체로 kappa가 0.75 이상인 경우 관찰자간에 일치도가 아주 좋은(excellent) 것으로 평가하며 kappa가 0.40과 0.75 사이의 값인 경우 일치도가 좋은(fair or good) 것으로 평가한다. 그리고 kappa가 0.40 이하의 값인 경우 일치도가 낮은(poor agreement) 것으로 평가한다.

그러나 우연에 의한 일치를 배제한 측정치로서 kappa계수의 유용성에도 불구하고 해석과 사용에 있어서 몇 가지 문제점이 있는 것으로 지적되고 있다. Brennan & Hays(1992)에 의하면 kappa

는 관찰자들간의 일치양상에 의해 영향을 받는다. 일반적으로 관찰자들간에 일치양상에 있어서 상이한 차이는 관찰자들의 측정이나 판단에서 편견(bias)의 개입을 의미한다. 예를 들면, 두 관찰자중 한 관찰자가 판단에서 다른 성향을 갖고 있는 경우, 즉, 부정적인 범주보다는 긍정적인 범주로 부호화하는 성향이 큰 경우 관찰자들간의 판단에서 편견이 개입된 것으로 볼 수 있다(Hutchinson, 1993). 이처럼 부호화에서 편견이 개입된 경우 kappa는 편견의 개입이 적어 일치양상이 동일할 경우보다 팽창(inflated)된 추정치를 제공한다.

실제로 측정에서 관찰자들의 편견 개입 유무는 쉽게 알 수는 없지만 matrix 상의 가로 세로 가장자리에 있는 합의 분포(marginal distribution)를 살펴보면 어느 정도 파악할 수 있다. 가로와 세로에 있는 합의 분포에서 관찰자간에 일치를 보이지 않는 경우 편견이 개입되었음을 의미한다. Banerjee & Fielding(1997)은 편견지수(bias index=BI)를 이용하여 편견 개입 정도를 파악할 수 있다고 하였다. 예를 들면 〈표 1〉과 같이 2×2 matrix로 나타낼 수 있는 경우 편견지수(BI)는 수학적으로는 $(a+b)/N - (a-c)/N$ 즉, $(b-c)/N$ 이 된다.

관찰에 있어서 관찰자들의 편견의 개입이 어떻게 실제 kappa 계수에 영향을 미치는 가를 알기 위해 Byrt et al.(1993)에 제시된 예를 인용해보면 〈표 2〉와 같다. 가상의 예제인 A와 B의 경우 두 관찰자의 판단에서 일치도 p_0 는 똑 같이 60%이다. 그러나 A예의 경우는 가로와 세로에 있는 가장자리 합들이 두 관찰자 사이에 일치를 보이는 반면 B의 경우는 가로와 세로에 있는 합의 관찰자간에 크게 다르다. 그러므로 A의 경우는 편견지수 BI=0.0이고 B의 경우는 편견지수 BI=0.3이 되며, 관찰자의 편견이 더 많이 개입된 B의 경우 kappa 계수는 A의 경우보다 다소 증가된 0.24를 보이고 있다. 그러므로 올바른 kappa 계수의 해석을 위해서는 관찰자의 편견 개입 여부와 정도에 대한 평가가 이루어져야 하겠다.

Kappa 계수에 영향을 미치는 또 다른 중요한 요인으로는 부호화될 범주들에서 실제 측정자료들

〈표 2〉 관찰자들간에 편견(bias)이 kappa에 미치는 영향

(예제 A)

		관찰자 1		
		예	아니오	합
관찰자 2	예	40	20	60
	아니오	20	20	40
	합	60	40	100

$p_0=0.6$ $kappa=0.17$ $bias\ index=0$

(예제 B)

		관찰자 1		
		예	아니오	합
관찰자 2	예	40	35	75
	아니오	5	20	25
	합	45	55	100

$p_0=0.6$ $kappa=0.24$ $bias\ index=0.3$

의 분포, 즉 범주별 분포(prevalence)의 차이를 들 수 있다. 예를 들면, 관찰자들이 관찰한 결과를 “예” 또는 “아니오”의 두 범주 중 하나로 부호화 하는 경우 만약 ‘예’ 범주의 발생이 ‘아니오’ 범주의 발생보다 상대적으로 월등히 높은 경우에는 kappa는 작아지는(deflated) 경향이 있다. 즉, 한 범주의 모집단 내에서 발현이 상대적으로 높은 경우, kappa 공식($k = p_0 - p_e / (1 - p_e)$)에서 보는 바와 같이 우연에 의한 관찰자간의 일치율을 나타내는 p_e 부분이 커져 kappa는 모집단에서 범주별 분포가 고른 경우보다 작아지는 경향이 있다. Byrt et al.(1993)은 이와 같은 범주에 따른 분포 차이를 범주분포지수(prevalence index=PI)를 사용하여 사정할 수 있다고 주장하였다. 〈표 1〉에서 처럼 2×2 matrix로 표시할 수 있는 경우 전체 모집단에서 ‘예’ 범주의 발생 가능성의 가장 좋은 추정치(best estimate)는 $(a+b)/N$ 과 $(a+c)/N$ 의 평균으로 볼 수 있고 동일하게 ‘아니오’ 범주의 발생가능성의 가장 좋은 추정치는(best estimate)는 $(c+d)/N$ 과 $(b+d)/N$ 의 평균으로 볼 수 있다. 그러므로 이 경우 ‘예’ 범주와 ‘아니오’

범주의 발현 확률에서 차이를 범주분포지수(prevalence index)라 할 수 있고, 수학적으로는 $PI = (a-d)/N$ 이다 최대치는 +1 이며, 최소치는 -1이 될 수 있다.

Banerjee & Fielding(1997)의 연구 결과를 예로 들어 범주별 분포에서 차이가 kappa에 미치는 영향을 살펴보자. 이 연구의 목적은 장기 간호시설에 거주하는 환자들이 대한 두 간호사의 사정결과를 기록한 간호기록지를 분석하여 특정 간호진단의 빈도를 확인하는 것이다. 두 간호사는 각각 독립적으로 360명의 대상자를 사정하고 기록하였다. 그리고 두 간호사의 환자사정에 대한 기록자료에서 특정 간호진단이 있는지 또는 없는지에 따라 일치와 불일치의 범주로 두 간호사의 판단을 부호화하였다. 이 연구에서는 105개의 NANDA 간호진단 중에서 총 100개의 간호진단이 간호기록지에서 확인되었다. 이 중에서 ‘외상의 고위험성’은 이 연구의 대상자들에서 실제 발생이 대단히 잦은 간호진단중의 하나이다. 〈표 3〉에서 보는 바와 같이 두 간호사가 작성한 환자들의 간호기록에서 ‘외상의 고위험성’이란 간호진단의 존재 유무에서 일치양상은 유사하였다. 즉 편견지수 BI는 0.003으로 대단히 낮은 편이다. 그러나 이 간호진단이 대상자들에게서 발현될 가능성이 다른 간호진단에 비해 월등히 높아 범주분포지수(prevalence index)는 0.74로 상당히 높은 수준이었다. 그리하여 일치도(agreement)는 0.82로 높은 반면에 우연에 의한 일치를 배제한 일치도인 kappa는 0.21로 대단히 낮은 편이어서 일치도와 kappa 사이에 상당한 차이를 보여 일치도(agreement)의 정확한 해석에서 어려움이 있다. 그러므로 kappa 계수의 올바른 해석을 위해서는 kappa에 영향을 미치는 관찰자들의 편견 개입이나 특정 범주의 타 범주에 대한 상대적 발현 가능성의 정도에 대한 평가가 반드시 이루어 져야 한다. 또한 일치도(percentage agreement)를 같이 보고함으로써 내용분석이나 상호작용분석시 신뢰도 측정치의 올바른 해석을 도모할 수 있다.

〈표 3〉 의상의 고위험(High risk for injury) : 관찰 시간 일치자료

		간호사 1		
		존재	부재	합
간호사 2	존재	280	33	313
	부재	32	15	47
합		312	48	360

$p_0=0.77$ BI=0.003 PI=0.74

$p_0=0.82$ kappa=0.21

4. 선행연구에서 kappa 적용과 문제점

선행연구에서 상호작용을 분석한 연구들은 주로 부모와 자녀사이의 상호작용 또는 간호사와 환자사이의 상호작용 등에 초점을 맞추어 진행되어 왔다. 이러한 연구들에서 측정시간 또는 관찰자간 신뢰도 확립의 사례를 살펴보면, Onyskiw et (1997)은 성인의 과거 유년기시절 가족과의 경험과 현재 부모-영아 상호작용과의 관계를 조사하였다. 가정방문을 통해 66쌍의 부모와 영아들과의 상호작용을 개별적으로 관찰하였다. 부모-영아 상호작용에서 상호작용의 질은 표준화된 관찰을 통하여 점수화 되었다. 총 53가지 행위들이 관찰을 통하여 유·무로 점수화 되었다. 점수가 높을수록 더욱 책임있는 부모역할 상호작용을 하는 것을 나타낸다. 자료수집자들은 표준화된 비디오테이프와 가정 관찰을 통해 훈련되었다. 자료수집전에 최소 측정자간 신뢰도는 85% 였다. 녹화된 자료와 가정방문을 통한 자료수집의 10%에 대해서 관찰자간 일치도를 산출하였다. 산출된 관찰자간 일치도는 평균 88%였으며, 우연에 의한 일치를 교정한 global kappa는 0.61이었다. 일치도와 kappa 값 사이에 다소 차이가 있는 것으로 보아 범주별 분포(prevalence)의 차이로 인한 효과나 관찰자간 편견의 개입으로 인한 kappa 값의 팽창이나 감소에 대한 평가가 필요하다.

아버지와 영아 사이의 상호작용을 조사한 연구(김영희, 1998)에서는 5분 동안 비디오로 촬영한 내용을 2명의 관찰자가 아버지-영아 놀이 상호작용 척도를 적용하여 면밀하게 관찰한 후 분석하

여 점수를 내었다. 아버지-영아 놀이 상호작용 척도는 모두 16개 항목으로 이루어져있고 비디오의 내용을 관찰한 뒤 각 항목에 대해 다섯 개의 범주 중 한 범주로 아버지와 영아의 상호작용을 부호화 하였다. 관찰자간 일치도(agreement)는 96%였다. 그러나 이 연구에서는 관찰자들 간에 우연의 일치를 교정한 kappa 값을 보고하지는 않았다.

조수현(1990)의 연구에서는 어머니의 신생아에 대한 애착행위를 조사하였다. 어머니와 신생아의 첫 직접적인 육체적 접촉이 이루어지는 수유실에서 어머니의 신생아에 대한 애착행위를 연구자가 직접 관찰을 통하여 발현 유·무로 표시하였다. 총 30가지 애착행위에 대해서 부호화 되었다. 이 연구에서는 측정자내 신뢰도(intrarater reliability)를 평가·보고하지는 않았다. 그러나 kappa를 이용하여 우연에 의한 일치를 배제한 일치도를 산출하여 제시함으로써 전체 연구결과에 대한 신뢰성을 높일 수 있다.

Garvin & Kennedy(1986)는 간호사와 의사의 30분간 토론에서 confirming과 disconfirming의 사소통을 조사하였다. 각 토론은 비디오로 녹화되었고 훈련된 판정자에 의해서 confirming과 disconfirming의 두 범주로 부호화 하였으며 분석의 단위는 utterance 였다. 각 범주에 대한 관찰자간의 일치도는 각각 .95와 .82였으며 우연에 의한 일치도가 교정된 kappa 계수는 각각 .90과 .64였다.

Ⅲ. 적용 사례

내용분석 또는 상호작용 분석 등의 연구에서 관찰자간 또는 측정자간의 신뢰도 산출은 전체 결과의 신뢰성을 높이기 위하여 반드시 필요한 기본요건의 하나이다. 여기서는 환자에게 제공된 예비적 정보에 대한 정보유형분석에서 본 연구자가 적용한 측정자간 신뢰도 산출의 절차를 예로 들어 측정자간 신뢰도 절차의 올바른 사용과 해석에 대해 살펴보자. 여기서 예로 제시한 측정자간 신뢰도 분석 절차는 인공호흡기를 사용할 환자들에게 제공한 예비적 정보가 환자들의 정서적 대처와 행동적 대처에 미친 영향을 조사한 연구의 일부분이

다. 이해를 돕기 위해 원 연구에 대해 간단히 설명하면, 심장수술을 받은 환자들은 수술이 끝난 후 중환자실로 옮겨지고 즉시 인공호흡기를 부착한다. 환자들이 마취에서 깨어나서 혈액학적으로 안정이 되고 여러 가지 호흡역학 지표들이 안정된 범위에 있을 때 환자들로부터 인공호흡기의 이탈을 시도하게 된다. 이탈이 성공적일 때 인공호흡기는 환자에게서 제거되고 기관내관의 탈관이 이루어진다. 주로 12-24시간에 이 모든 과정들이 진행이 된다(허, 1993). 그러나 비교적 짧은 이 기간 동안 환자들은 여러 가지 요인들로부터 심리적, 신체적 고통을 당하고 있다.

그러므로 사전에 미리 환자들에게 수술후 인공호흡기 사용과 의사소통의 어려움에 대해 예비적 정보를 제공하여 준비시키면 이 기간 동안에 환자들이 겪게 될 심리적 고통은 감소할 것이다. 그리하여 본 연구에서는 수술 전에 환자들에게 인공호흡기가 부착되어 있는 동안 경험할 내용과 의사소통에 대한 예비적 정보를 제공하고 그 효과를 조사하였다. 선행연구들은 감각적 정보가 환자들의 위기 대처에 긍정적인 효과를 가졌다고 보고하였으나 어느 정도의 감각정보가 제공될 때 최대 효과를 얻을 수 있는지에 대해서는 보고하지 않았다. 그러므로 본 연구자와 Garvin이 Garvin(1992)에 의해 개발된 정보유형 분석도구를 이용하여 인공호흡기를 사용할 환자들에게 제공된 예비적 정보의 정보유형을 분석하고 그 속에 포함된 구체적 객관적 정보의 양을 측정하였다.

Garvin과 연구자에 의해서 이루어진 정보유형 분석 결과의 신뢰성을 높이기 위하여 위에서 진술한 측정자간 신뢰도 산출 절차를 본 분석에 적용하였다. 앞서 말한 바와 같이 내용분석시 측정자간 신뢰도 산출은 두 단계를 거쳐야 한다. 먼저 예비적 정보 속에 포함된 정보유형을 분석하기 위하여 두 연구자가 전체 예비적 정보를 분석단위로 나누었다. 연구자들에 의해서 확인된 분석단위는 분리될 수 있으면서 서로 다른 종류의 정보를 제공하는 단어, 구 또는 문장이다. 이 경우 분석단위가 다소 추상적이고 측정자들에게 요구되는 추론의 정도도 비교적 높아 측정자간 분석단위 확인

신뢰도에 대한 검증이 필요하다. 분석 단위 확인에서 측정자간 일치도의 평가는 확인된 분석단위의 수에서 일치 뿐만 아니라 분석단위의 시작과 끝의 일치도 포함한다. 분석단위의 시작과 끝의 일치만 한 단어 또는 두단어의 범위내에서 분석단위가 일치되게 확인되는 경우, 두 측정자간 분석 단위 확인에서 일치하는 것으로 분류하였다. 전체 연구결과와 신뢰성을 높이고 다음 단계의 신뢰도 평가를 위해 분석 단위 확인 단계에서 분석자 사이에 적절한 일치가 요구된다. 김과 김(1997)에 의해 분석된 내용중 한 예를 들어 보면, 심도자 검사 환자를 위한 예비적 정보 중 “(심도자 검사 중에)(귀하의 혈압을 측정하며)(심장 엑스레이 사진을 찍게 됩니다)”와 같은 문장에서는 두 분석자들은 괄호로 나뉜 것 처럼 일치되게 3개의 분석단위를 확인하였다. 본 연구에서는 분석단위를 확인함에 있어서 두 연구자간에 일치도를 증가시키기 위해 사전에 컴퓨터 단층촬영, colonoscopy, 간생검에 관한 예비정보를 이용하여 연구자들간에 일정한 사전 훈련을 거친 후 본 연구에 대한 분석을 시작하였다. 대체로 분석 단위 확인 신뢰도가 80% 이상이면 만족한 수준의 일치로 간주한다(Garvin, 1992).

두 연구자가 일치되게 확인한 분석단위의 수는 전체 132개 중 102개였으며 불일치가 있었던 분석단위는 30개로 분석 단위 확인 신뢰도는 만족한 수준에 가까운 77%였다. 분석단위 확인 신뢰도가 구해진 다음 연구자들간에 토의를 거쳐 부호화를 위한 최종 분석단위를 결정하였다.

다음 단계에서는 최종 확인된 132개의 분석단위를 두 연구자가 3가지 정보유형 범주로 부호화하였다. 정보유형의 세가지 범주는 구체적 객관적 정보, 절차정보, 그리고 기타 정보였다. 구체적 객관적 정보는 Johnson, Lauver & Neil(1989)의 개념정의를 그대로 적용하였다. 구체적 객관적 정보는 다음과 같은 특성을 갖는다: 잠재적으로 위협적인 사건의 경험자들의 50% 이상에 의해 오관을 통해 경험되는 구체적이고 객관적인 용어로 표현되는 신체감각들에 대한 서술이며, 사건이 일어나는 공간의 환경적 특성과 그 경험의 시차적 특

성에 대한 서술과 그러한 신체 감각들(보기, 듣기, 맛보기, 느끼기, 냄새)에 대한 이유 또는 원인을 설명하는 정보이다. 예를 들면, 이 연구에 포함된 구체적 객관적 정보는 흡인시 환자들은 일시적으로 기침과 구역질을 경험할 수 있고, 인공호흡기로부터 증기기관차의 엔진소리나 나무를 자를 때 나는 소리와 유사한 소리를 들을 것이라는 것을 알려주는 내용 들이다.

다음으로 절차정보는 경험이나 절차의 단계를 서술하는 것으로 예를 들면, 누가 무엇을 할지, 절차의 목적과 왜 그러한 절차를 밟아야 하는지, 사용되는 기구나 장비에 대한 설명과 환자가 절차에 대처하는 방법 등을 알려주는 내용들이다. 마지막으로 기타정보는 사전에 환자에게 주어진 정보들 중 절차정보나 구체적 객관적 정보의 범주에 포함되지 않는 정보를 말한다(Garvin, 1992). 두 연구자에 의해 3가지 정보유형의 범주로 부호화된 분석단위들은 <표 4>에서 처럼 3×3 matrix로 나타낼 수 있다.

<표 4> 정보유형분석 : 관찰자간 일치와 불일치

		관찰자 1			
관찰자 2	구체적 객관적 정보	절차정보	기타정보	합	
구체적 객관적 정보	41	3	4	48	
절차정보	2	71	10	83	
기타정보	0	0	1	1	
합	43	74	15	132	

<표 4>에서 보는 바와 같이 총 132개의 분석단위 중 41개의 분석단위가 연구자 사이에 일치되게 구체적 객관적 정보로 확인되었고, 총 71개의 분석단위가 연구자들 사이에 일치되게 절차정보로 분류되었다. 연구자들 사이에 일치되게 기타 정보의 범주로 부호화된 분석단위의 수는 1개 였다. 그러므로 부호화에서 연구자간의 일치도, ρ 는 3×3 matrix 상에서 대각선 상에 있는 숫자의 합을 전체 분석단위의 숫자로 나눈 값 $[(41+71+1)/132]$ 으로 85% 이다. 연구자 간에 우연에 의해서 일치

를 보일 수 분석단위의 비율은 ρ_e 로 표시하며 각각의 가로 세로의 가장자리에 있는 합을 각 범주끼리 곱하여 다 더한 후 전체 분석 단위의 제곱으로 나눈 값 $[(43)(48)+(74)(83)+(15)(1)]/(132)^2$ 으로 .47이다. 그러므로 우연에 의한 일치를 교정한 두 연구자 사이의 일치도 kappa(해석적 신뢰도)는 .72였다. Landis & Koch(1977) 의하면 kappa가 0.4와 0.75 사이의 값인 경우 측정자 간의 일치도는 좋은 것으로 간주한다.

그러나 어떤 범주는 사용에 있어서 다른 범주보다 어려워 타 범주에 비해 부호화에서 분석자간 일치도가 더욱 떨어질 수도 있다(Cissna et al., 1990 ; Garvin et al., 1988). 그러므로 범주별 신뢰도를 제시하여야 한다. 특히 분석단위가 주관적인 경우 범주별 신뢰도의 보고는 결과의 정확한 해석을 위해 더욱 중요하다. 본 분석에서 범주별 연구자간 일치도는 구체적 객관적 정보 85%, 절차정보 85%, 그리고 기타정보 6%였으며, 범주별 우연에 의한 일치를 교정한 신뢰도(kappa 계수)는 구체적 객관적 정보 .85, 절차정보 .80, 기타정보는 .14였다. 그러므로 인공호흡기를 사용할 환자에게 제공된 예비적 정보에 포함된 구체적 객관적 정보의 양의 분석에서 구체적 객관적 정보 범주로의 부호화에서 연구자 간의 일치도는 .85로 아주 좋았다.

Kappa는 관찰자들간의 일치 양상에 의해서 영향을 받는데(Brennan & Hays, 1992), 본 예제에서 관찰자간 일치 양상은 범주별로 가로 가장자리의 합과 세로 가장자리의 합을 비교해 봄으로써 알 수 있다. <표 4>를 살펴 보면, 구체적 객관적 정보의 경우 가로 세로의 합은 각각 빈도에서 43과 48로 큰 차이를 보이지 않고 있다. 절차정보의 경우도 마찬가지로 각각 74와 83으로 큰 차이를 보이지 않고 있다. 그러나 기타 정보의 확인에서는 연구자 사이에 큰 차이를 보이고 있어 연구자간에 다소 편견이 개입된 것으로 보인다. 그러므로 기타 정보 범주가 사용에 있어서 타 범주에 비해 어려웠음을 말해 주고 있는데 이는 기타 정보의 일치도가 6%에 불과하며 kappa 또한 .14인 것을 보아도 알 수 있다. 아마도 기타 정보의 확

인에서 관찰자의 편견 개입의 이유를 확인함으로써 이 범주의 확인에서 관찰자 간의 일치도를 더욱 증가 시킬 수 있을 것이다. 다음으로 kappa에 영향을 미칠 수 있는 요인은 범주별 분포(prevalence)인데, 본 분석에서는 절차범주의 발현 빈도가 타 범주의 발현빈도에 비해 다소 높은 것을 볼 수 있다. 그러므로 범주별 발현 빈도에서 차이가 전체 kappa 계수에 다소 영향을 주었을 것이다. 결론적으로, 본 예제의 경우에는 일치도(percentage agreement)와 kappa 계수가 각각 0.85와 0.72로 비교적 차이가 적고 역설적인 관계를 보이지 않고 있다. 그러므로 본 분석의 kappa 계수를 살펴보면 관찰자간 편견의 개입으로 인한 과도한 inflation이나 범주별 분포의 차이로 인한 지나친 deflation 효과는 비교적 작은 것으로 판단된다.

IV. 결 론

본 논문에서는 관찰자 또는 측정자가 측정오류의 주된 요인으로 작용할 수 있는 내용분석이나 상호작용 연구에서 측정자간 신뢰도의 산출과 해석에 대해 설명하였다. 측정자간 신뢰도를 산출하는 첫 단계에서 연구자간 분석단위 확인 신뢰도의 산출은 해석적 신뢰도 산출을 위한 전제조건이 될 수 있다. 특히 대상자들의 행위를 확인하기 위해 관찰자들에게 요구되는 추론의 정도가 클수록 이러한 전제조건은 반드시 충족되어야 할 것이다.

해석적 신뢰도를 산출하는 단계에서는 kappa가 널리 사용되고 있으나 kappa가 모든 상황에서 만족하게 사용될 수 있는 지표는 아니다. 특히 관찰자의 편견 개입이 크거나 범주별 상대적 발현 빈도에서 차이가 큰 경우 Cohen's kappa의 사용이 부적절 할 수도 있다. 때로는 kappa의 해석에 주의가 필요하다. 측정자간 신뢰도 검증에 kappa의 사용이 적절하지 못한 경우 몇가지 대안을 사용할 수 있다. 첫째로는 관찰자의 편견 개입이 큰 경우에는 관찰자간의 일치도를 산출하는 방법으로 Tetrachoric correlation을 사용할 수 있다. Tetrachoric correlation은 범주들의 가로 세로 가장자리 합의 일치양상에 의해 크게 영향을 받지 않는

다(Hutchinson, 1993). 둘째로는 범주별 분포(revalence)가 일치도에 미친 영향을 실질적으로 평가하기 위하여 Cicchetti & Feinstein(1990)이 제시한 긍정일치(positive agreement = ppos)와 부정일치(negative agreement = pneg) 지표를 사용할 수도 있다. 마지막으로는 prevalence에서 차이와 편견개입의 영향을 교정한 prevalence-adjusted, bias-adjusted kappa의 사용을 제안하고 있다(Byrt et al., 1993). 특히 일치도 연구들 간의 의미있는 비교를 위해서는 교정된 측정치를 함께 보고할 필요가 있다(Byrt et al., 1993). 뿐만 아니라, kappa의 정확한 해석을 위해서는 kappa와 함께 반드시 일치도(percentage agreement)를 같이 보고하여야 한다.

대부분의 연구들은 해석적 신뢰도의 산출에서 전체 신뢰도 추정치만을 보고하고 있다. 그러나 전체 신뢰도(global reliability)는 관찰자가 각각의 범주를 사용함에 있어서 얼마만큼의 일관성을 가졌는지에 대해서는 나타내지 못한다. 그러므로 범주별 신뢰도 추정치를 제시하여 각 범주를 사용하는 데 있어서 관찰자들이 얼마나 일관성을 유지하였는지를 보여줄 필요가 있다.

판정 결과가 이분적이 아니고 다원적(polytomous)인 명목자료인 경우 범주들의 특성을 고려하여 편의에 따라 수개의 2분적 조합을 만들어 개개의 kappa 추정치를 제시할 수도 있다(안 & 유, 1996).

참 고 문 헌

- 김영희 (1998). 아버지와 영아의 상호작용에 관한 연구. 이화여자대학교 대학원 박사학위논문.
- 김조자, 김화순 (1997). 심도자 검사환자를 위한 구체적 객관적(감각) 정보의 개발. 간호학회지, 27(2), 433-443.
- 안윤옥, 유근영 (1996). 의학통계론. 서울: 서울대학교 출판부.
- 조수현 (1990). 어머니의 신생아에 대한 지각과 애착행위와의 관계. 전남대학교 대학원 석사학위논문.

- Banerjee, M. & Fielding, J. (1997). Interpreting kappa values for two-observer nursing diagnosis data. Research in Nursing & Health, 20, 465-470.
- Brennan, P. F. & Hays, B. J. (1992). The kappa statistic for establishing interrater reliability in the secondary analysis of qualitative clinical data. Research in Nursing & Health, 15, 153-158.
- Byrt, T., Bishop, J., & Carlin, J. B. (1993). Bias, prevalence & kappa. Journal of Clinical Epidemiology, 46(5), 423-429.
- Cicchetti, D. V. & Feinstein, A. R. (1990). High agreement but low kappa: II. resolving the paradoxes. Journal of Clinical Epidemiology, 43, 551-558.
- Cissna, K. N., Garvin, B. J., & Kennedy, C. W. (1990). Reliability in coding social interaction: a study of confirmation. Communication Reports, 3(2), 58-69.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational & Psychological Measurement, 20(1), 37-46.
- Folger, J. P., Hewes, D., & Poole, M. S. (1984). Coding social interaction. in B. Dervin & M. Voight(Eds.), Progress in the communication sciences(pp.115-161). New York: Ablex.
- Garvin, B. J. (1992). Types of information instrument. Unpublished manuscript. The Ohio State University.
- Garvin, B. J., Huston, G. P., & Baker, C. F. (1992). Information used by nurses to prepare patients for a stressful event. Applied Nursing Research, 5(4), 158-163.
- Garvin, B. J. & Kennedy, C. W. (1986). Confirmation and disconfirmation in nurse/physician communication. Journal of Applied Communication Research, 14(1), 1-19.
- Garvin, B. J., Kennedy, C. W., & Cissna, K. N. (1988). Reliability in category coding system. Nursing Research, 37(1), 52-55.
- Holaday, B. (1981). Maternal responses to their chronically ill infants' attachment behavior of crying. Nursing Research, 30(6), 343-348.
- Hurley, P. M. (1981). Communication patterns and conflict in marital dyads. Nursing Research, 30(1), 38-42.
- Hutchinson, T. P. (1993). Kappa muddles together two sources of disagreement: Tetrachoric correlation is preferable. Research in Nursing & Health, 16, 313-315.
- Johnson, J. E., Lauver, D., & Neil, L. M. (1989). Process of coping with radiation therapy. Journal of Consulting & Clinical Psychology, 57, 358-364.
- Kerlinger, F. N. (1973). Foundations of behavioral research(2nd ed.), New York: Holt, Rinehart & Winston.
- Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33, 159-174.
- Onyskiw, J., Harrison, M., & Magill-Evans, J. (1997). Past childhood experiences and current parent-infant interactions. Western Journal of Nursing Research, 19(4), 501-518.
- Topf, M. (1986). Three estimates of interrater reliability for nominal data. Nursing Research, 35(4), 253-255.

Abstract

Key concept : Content Analysis, Preparatory Information, Interrater Reliability, Cohen's kappa

Interrater Reliability in the Content Analysis of Preparatory Information for Mechanically Ventilated Patients

Kim, Hwa Soon*

In nursing research that the data is collected through clinical observation, analysis of clinical recording or coding of interpersonal interaction in clinical areas, testing and reporting interrater reliability is very important to assure reliable results. Procedures for interrater reliability in these studies should follow two steps. The first step is to determine unitizing reliability, which is defined as consistency in the identification of same data elements in the record by two or more raters reviewing the same record. Unitizing reliability have been rarely reported in previous studies. Unitizing reliability

should be tested before progressing to the next step as precondition.

Next step is to determine interpretive reliability. Cohen's kappa is a preferable method of calculating the extent of agreement between observer or judges because it provides beyond-chance agreement. Despite its usefulness, kappa can sometimes present paradoxical conclusions and can be difficult to interpret. These difficulties result from the feature of kappa which is affected in complex ways by the presence of bias between observers and by true prevalence of certain categories. Therefore, percentage agreement should be reported with kappa for adequate interpretation of kappa. The presence of bias should be assessed using the bias index and the effect of prevalence should be assessed using the prevalence index.

Researchers have been reported only global reliability reflecting the extent to which coders can consistently use the whole coding system across all categories. Category-by-category reliability also need to be reported to inform the possibility that some categories are harder to use than others.

* Department of Nursing, College of Medicine, Inha University