

[원저]

## 시간교환법을 이용한 EQ-5D 가치 평가 작업의 타당도 및 신뢰도 평가

조민우, 이상일<sup>1</sup>동국대학교 의과대학 예방의학교실, 울산대학교 의과대학 예방의학교실<sup>1</sup>

## - 요약 -

연구배경	우리나라에서는 아직까지 일반 인구집단을 대상으로 시간교환법을 이용하여 건강상태에 대한 가치 평가 작업이 이루어진 바 없다. 이 연구의 목적은 한국의 일반 인구집단을 대상으로 시간교환법을 이용하여 EQ-5D의 가치평가세트를 구하는 과정에 대한 타당도와 신뢰도를 평가하는 것이다.
방 법	서울과 경기 지역 일반 인구 500명을 대상으로 시각 아날로그 척도법과 시간교환법을 이용하여 가상의 EQ-5D 표준 건강상태에 대해 가치를 평가하였다. 총 41개의 EQ-5D 건강상태를 표준 건강상태로 하고, 이 중 11개를 중증도에 따라 무작위로 선택하도록 하였다. 가장 나쁜 건강상태인 '33333' 상태와 '무의식' 상태에 대해서는 모든 대상자가 가치를 평가하였다. 2차 설문조사에서는 500명 중 서울 지역의 97명이 응답하였는데, 1차 조사와 동일한 상태에 대해서 가치를 평가하였다. 이러한 수행 과정에 대한 타당도를 살펴보기 위하여 시각 아날로그 척도법으로 평가한 값과 시간교환법으로 평가한 값의 상관관계를 살펴보고, 또 다른 측면으로 모순값의 수를 파악하였다. 신뢰도 측면에서는 평가-재평가 신뢰도를 개인별 급내 상관계수를 이용하여 분석하였다.
결 과	시각 아날로그 척도법과 시간교환법을 이용하여 가치 평가한 값의 상관계수를 살펴본 결과 0.934로 매우 높게 나타났다 ( $p<0.001$ ). 시간교환법에서 4개 이상의 모순값을 가진 응답자는 12명(2.4%)이었다. 개인별 급내 상관계수의 평균은 0.857로 나타났다으며, 95% 신뢰구간은 0.836에서 0.878이었다.
결 론	한국의 일반 인구집단에서 가상의 EQ-5D 건강상태에 대해 시간교환법으로 가치 평가한 과정의 타당도와 신뢰도가 높은 수준으로 나타나, 한국의 일반 인구집단에서도 이러한 방법론을 적용할 수 있음을 확인하였다. (대한임상건강증진학회지 2007;7(2):96~103)
중심단어	EQ-5D, 시간교환법, 타당도, 신뢰도

## 서 론

보건의료에 대한 의사결정을 하고자 할 때, 건강상태에 대한 가치를 평가한 결과는 의사결정의 중요한 정보가 된다. 이러한 지표 중에 유용한 지표가 QALYs(quality adjusted life years)이다. QALYs는 생존기간으로 나타나는 삶의 양적인 부분과 삶의 질적인 측면을 함께 반영한 집단 건강의 요약 지표(summary measure of population health)의 하나이

며, 비용-효용 분석(cost-utility analysis)의 결과 측정치로 널리 사용되고 있다.<sup>1,2)</sup>

QALYs를 구하기 위해서는 각 건강상태에 대한 선호도(preference)를 반영한 질 가중치(quality weight)가 있어야 한다. 질 가중치를 구하는 접근법으로는 직접법과 간접법이 있다. 직접법은 표준도박법(standard gamble, SG)이나 시간교환법(time trade-off, TTO)을 이용하여 특정 건강상태의 질 가중치를 직접 평가하며, 간접법은 체계적으로 분류된 일반적 건강상태(generic health states)에 대해서 SG나 TTO 등을 이용하여 질 가중치 측정식 또는 표를 만들어 필요한 건강상태에 대한 질 가중치를 구하는 방법이다.<sup>3,4)</sup> 간접적인 접근법에 흔히 사용하는 일반적 건강상태 측정 도구로는 QWB<sup>5,6)</sup>, HUI<sup>7,8)</sup>, EQ-5D<sup>9,10)</sup>가 있다. 이 중 EQ-5D는 EuroQol group에서 개발한 일반적 삶의 질 측정 도구로 운동능력, 자기관리, 일상 활

\* 본 연구는 보건복지부 보건의료기술진흥사업의 지원에 의하여 이루어진 것임 (02-PJ1-PG10-21999-0001).

• 교신저자 : 이 상 일 울산대학교 의과대학 예방의학교실

• 주 소 : 서울시 송파구 풍납동 388-1

• 전 화 : 02-3010-4284

• E-mail : sleemd@amc.seoul.kr

• 접수일 : 2007년 6월 1일 • 채택일 : 2007년 6월 22일

동, 통증/불편, 그리고 불안/우울의 도메인(domain)에서 세 가지 수준(3 levels)으로 건강 상태를 분류하는데, 일반적으로 5개의 숫자로 건강상태를 나타낸다. 예를 들어 '12123'의 경우, 운동능력과 일상 활동에는 전혀 문제가 없으나 자기 관리와 통증/불편에는 다소 문제가 있고, 불안/우울 도메인에서는 심각한 문제가 있는 상태를 의미한다.<sup>11,12)</sup>

질 가중치 측정 방법에는 시각 아날로그 척도법(visual analogue scale, VAS), SG, TTO 등이 있는데, 크게 두 가지 관점에 따라서 나누어볼 수 있다. 첫 번째는 주어진 혹은 선택한 결과의 불확실성에 관한 문제이고, 두 번째는 건강상태 수준의 교환(trade-off)을 통한 선택 여부이다. VAS에서는 선택의 여지가 없이 주어진 건강상태에 대해서 평가하게 된다. 그리고 주어진 상태는 반드시 발생하는 사건이라고 가정한다. 대표적인 것인 EQ-VAS인데, 상상할 수 있는 최악의 상태를 0점으로 상상할 수 있는 최고의 건강상태를 100점으로 한 20cm 길이의 온도계 모양의 수직자를 두고, 이 사이에서 설문 당일 건강상태의 점수를 매기도록 한다. TTO의 경우에는 가치 평가하고자 하는 어떤 가상의 건강상태로 일정 기간 동안 사는 것과 그 보다는 짧지만 완전한 건강상태로 일정 기간 동안 사는 대안 중 한 가지를 선택하게 되며, 선택한 결과는 확실하게 발생하는 것으로 가정한다. SG의 경우에는 확률에 따라 사망과 최고의 건강상태 중 한 가지 상태로 결정되는 대안이 있고 특정 건강상태로 계속 살아가는 대안이 있는데, 확률에 따라 두 건강상태 중 한 가지 상태를 선택하게 된다. 원래는 SG가 효용의 의미를 가지는 질 가중치를 측정하는 가장 좋은 방법으로 알려져 있으나, 그 개념을 이해하기 어려워 그에 대한 대안으로 TTO가 제시되고 있다. TTO는 불확실성 하에 판단하지는 않으나 선택을 통해 가치를 판단하는데, 그 값이 SG와 유사하게 나타나는 것으로 알려져 있다. 그리고 SG보다 응답의 내적 일관성, 민감도, 그리고 신뢰도 측면에서 좀 더 나은 결과를 보이고 있는 것으로 발표된 바 있다.<sup>13)</sup> 이에 외국에서도 경제성 평가를 위해서 EQ-5D 건강상태의 질 가중치를 구하는 경우 TTO를 사용하고 있다.<sup>14,15)</sup>

한국에서도 EQ-5D를 이용한 QALYs 관련 연구가 여러 차례 이루어진 바 있으나<sup>16,17)</sup>, 이들 연구는 외국의 질 가중치를 이용하여 QALYs를 계산하였다. 하지만 이렇게 외국자료를 한국에 바로 적용하였기 때문에 타당도에서 문제가 있을 수 있다. 외국에서 구한 EQ-5D 건강상태에 대한 TTO 값과 한국의 값이 차이가 난다는 것은 이미 조민우와 이상일<sup>12)</sup>이 보고한 바 있다. 이렇게 차이가 날 수 있는 이유는 어떤 건강상태의 수준을 평가하는 과정에서 평가하는 사람의 사회문화

적인 가치관이 영향을 미칠 수 있기 때문이다. 따라서 많은 국가들이 EQ-5D에 대해서 건강 관련 삶의 질 수준에 대한 선호도를 나타내는 가치 평가 세트를 별도로 구축하고 있다.<sup>18)</sup> 또한 이러한 가치 평가에 방법론적인 측면에서 어려움이 있기 때문에 이러한 방법론을 적용할 수 있는지에 대한 연구를 같이 진행하기도 한다.<sup>19)</sup> 하지만, 아직까지 우리나라에서는 일반 인구집단에서의 TTO 적용에 대한 타당도 및 신뢰도 평가 연구는 아직 이루어진 바 없다.

이 연구의 목적은 한국의 일반 인구집단을 대상으로 EQ-5D 건강상태에 대해서 TTO 방법을 이용한 질 가중치 측정의 타당도 및 신뢰도 수준을 평가하는 것이다.

## 연구 방법

### 1. 연구 대상

서울 경기 지역의 만 20세 이상 성인 남녀를 설문 대상으로 하였다. 2004년 주민등록인구를 모집단으로 하여 비례할당추출을 통해 총 500명의 연령별 성별 수를 미리 선정하였다. 먼저 인구수에 따라 동 지역을 추출한 후, 해당 지역에서 조사 가구를 직접 방문하여 면접조사를 수행하였다. 조사는 2005년 3월 24일부터 4월 21일까지 수행되었다. 두 차례의 설문조사 교육을 받은 일정 교육수준 이상의 면접조사자 19명이 일반 인구집단을 대상으로 일대일 면접 조사하였다. 서울 지역 참가자 중 재조사 참가 의사를 밝힌 97명에 대해서는 평균 6주 후에 재조사를 시행하였다.

### 2. 가치 평가 대상 건강상태의 선택

가치 평가할 가상의 건강상태는 영국 MVH 연구<sup>14)</sup>와 동일하게 전체 243개 건강상태 중 일부 건강상태를 사용하였다. 두 가지 가치 평가 방법을 사용하였는데, VAS로 평가할 때에는 41개 EQ-5D 건강상태를 중증도(severity)에 따라 5개의 '아주 경한 상태(very mild)'에서 2개 건강상태를, 12개의 '경한 상태(mild)', '중등도 상태(moderate)'와 '중한 상태(severe)'에서 각각 3개 건강상태를 무작위로 선택하여 평가하도록 하였다(Figure 1). '11111', '33333', '무의식', 그리고 '사망'은 모든 사람들이 평가하도록 하였으므로 VAS로 평가한 상태는 15개 건강상태였다. TTO에서는 기준값으로 사용한 '11111'과 '사망'은 평가에서 제외하여 13개 상태만 평가하였다.

Each respondent valued 33333 and unconscious
plus
2 from 5 "very mild" states:
11112, 11121, 11211, 12111, 21111
plus
3 from 12 "mild" states:
11122, 11131, 11133, 21133, 21222, 21312, 12211, 11133, 22121, 12121, 22112, 11312
plus
3 from 12 "moderate" states:
13212, 32331, 13311, 22122, 12222, 21323, 32211, 12223, 22331, 21232, 32313, 22222
plus
3 from 12 "severe" states:
33232, 23232, 23321, 13332, 22233, 22323, 32223, 32232, 33321, 33323, 23313, 33212

**Figure 1.** EQ-5D health states used in this study for TTO (Dolan P, 1997)

EQ-5D 전체 243개 건강상태 모두를 평가할 수 없으므로, 그 중 이전 연구에서 빈번하게 표현되었던 상태, 상상하기 쉬운 상태, 그리고 5개 차원의 3개 수준의 건강상태가 잘 조합된 상태 41개 건강상태를 영국의 MVH 연구 그룹에서 표준 건강상태로 사전에 정하였음. 이 41개 건강상태를 그 상태에 따라 아주 경한 상태에서부터 심각한 상태로 4개 군으로 분류하고 이 군 안에서 아주 경한 상태군에서는 2개, 나머지 상태군에서는 3개 상태를 무작위로 선택하여 건강 상태에 대한 가치를 평가하도록 하였음. 각각의 응답자는 33333 상태와 무의식 상태에 대해서는 모두 가치 평가 작업을 수행하였음.

### 3. 설문조사 과정

응답자는 세 가지 과정을 통해 설문을 수행하였다. 첫 번째 과정에서는 응답자가 EQ-5D 설문지로 본인의 건강상태를 평가하도록 하였다. 두 번째 과정은 응답자가 VAS를 이용하여 가상의 건강상태를 평가하는 것이었다. 먼저 11개 건강상태를 선택하고, '11111', '33333', '무의식', '사망'을 포함한 총 15개 건강상태를 좋은 건강상태부터 나쁜 건강상태까지 순위를 매기도록 하였다. 그런 후 순서에 따라 각 건강상태를 VAS를 이용하여 가치를 평가하였다. 점수 부여 후 다시 한번 수정할 수 있도록 하였다. 세 번째 과정은 TTO를 이용하여 가상의 건강상태를 평가하는 것이었다. 먼저 VAS에서 평가하였던 건강상태 중 '11111'과 '사망'을 제외한 13개 건강상태를 죽는 것보다 나은 건강상태와 죽는 것이 더 나은 건강상태로 분류하였다. 그 후 각 집단 내에서 좋은 건강상태부터 나쁜 건강상태 순으로 순위를 매기고 그 순서에 따라서 각 건강상태를 TTO를 이용하여 가치를 평가하였다. '죽는 것보다 나은 건강상태'에서는 선택한 건강상태로 10년을 사는 것과 EQ-5D로 표현할 때 완전한 건강상태인 '11111'의 건강상태로 살지만 10년보다 짧은  $t$  년을 사는 것 중 하나를 선택하게 되는데, 두 대안의 선호도에 차이가 없는 무차별점(indifferent point)에서의  $t$  값을 구했다. 그리고 '죽는 것보다 못한 건강상태'에서는 당장 죽는 것과 10년을 반드시 살아야

하는데 '죽는 것보다 못한 건강상태'라고 분류한 그 건강상태로 (10- $t$ )년을 살다가 나머지  $t$  년을 '11111'의 건강상태로 사는 선택에서 무차별점  $t$  값을 구했다.

### 4. 시각 아날로그 척도법과 시간교환법의 점수 변환

최고의 건강상태에서는 1, 죽음은 0, 죽는 것보다 못한 건강상태는 음수인데 -1까지 값을 갖도록 다음의 두 식을 이용하여 변환하였다.

$$TTO_{Tx} = \frac{TTO_x}{10} \quad \dots \text{'죽는 것보다 나은 건강상태'}$$

$$TTO_{Tx} = \frac{(10 - TTO_x)}{10} - 1 = -\frac{TTO_x}{10} \quad \dots \text{'죽는 것보다 못한 건강상태'}$$

$TTO_{Tx}$ : 상태  $x$ 의 변환 후 TTO값

$TTO_x$ : 상태  $x$ 에 대해 TTO로 가치 평가한 값

TTO로 가치 평가한 값은 죽음보다 못한 상태를 평가하도록 하였고 이를 음수로 표현하였다. 따라서 VAS로 측정된 값도 TTO와 유사한 범위를 가지도록 다음의 식을 통하여 변환하였다.

$$VAS_{Tx} = \frac{VAS_x - VAS_{death}}{VAS_{11111} - VAS_{death}}$$

$VAS_{Tx}$ : 상태  $x$ 의 변환 후 VAS값

$VAS_{11111}$ : 상태 '11111'에 대해 VAS로 가치 평가한 값

$VAS_x$ : 상태  $x$ 에 대해 VAS로 가치 평가한 값

$VAS_{death}$ : 죽음에 대해서 VAS로 가치 평가한 값

### 5. 타당도

타당도는 연구자가 측정하고자 하는 것을 측정하는 정도라고 정의할 수 있다. 어떤 '황금기준(gold standard)'이 존재하여 그 기준에 따라 평가하는 것이 가장 이상적이지만, 질 가중치 도출 작업에서는 이러한 황금기준이 존재하지 않는다. 이 연구에서는 질 가중치 도출 작업의 타당도를 평가하기 위해서 수렴 타당도(convergent validity)와 자료의 질을 살펴본다. 수렴 타당도는 같은 상태에 대한 VAS와 TTO, 두 값의 상관관계의 크기로 판단하였다. 만일 시각 아날로그 척도법과 시간교환법의 두 작업을 모두 잘 이해하고 수행하였다면,

두 값의 상관성이 높게 나타날 것이다.

또 하나의 타당도의 기준으로 자료의 질을 살펴보았다. EQ-5D로 표현된 건강상태 중에는 논리적으로 비교할 수 있는 건강상태가 존재한다. 예를 들어 '11112'는 '11122'보다 좋은 건강상태이므로, '11112'의 질 가중치 값은 '11122'의 값과 같거나 커야 한다. 만일 그 반대라면 그것은 모순이 되는 것이다. Ohinmaa와 Sintonen<sup>20)</sup>은 한 사람에서 가치 평가한 질 가중치 값을 비교하여 모순값의 수가 4개 이상인 경우 모델링을 위한 자료에서 제외하도록 근거를 제시한 바 있다.

## 6. 신뢰도

이 연구에서는 1차 조사 후 재조사를 시행하였고, 이 자료에서 산출한 개인별 급내 상관계수(intra-class correlation coefficient, ICC)를 구하여 신뢰도 수준을 평가하였다. 만일 가치 평가 작업이 신뢰도 있게 수행되었다면, 두 값 사이의 ICC가 높게 나타날 것이다. ICC를 산출하기 위해서는 다음의 공식을 이용하였다.

$$ICC = \frac{\sigma_{jw}}{\sigma_{jw} + \sigma_{jb}} \quad \dots (1)$$

$\sigma_{jw}$ : j 번째 응답자에서 질 평가 도출 작업 내 변이

$\sigma_{jb}$ : j 번째 응답자에서 질 평가 도출 작업 간 변이

## 결 과

### 1. 연구 대상

본 조사와 재조사에서의 응답자의 일반적인 특성을 (Table 1)에 나타내었다. 본 조사와 재조사에 참여한 응답자의 일반

**Table 1.** Socio-demographic characteristics of study participants

Characteristics		Test(%)	Retest(%)	p*
Gender	Male	251(50.2)	48(49.5)	0.897
	Female	249(49.8)	49(50.5)	
Age group	20~29	112(22.4)	24(24.7)	0.940
	30~39	134(26.8)	27(27.8)	
	40~49	118(23.6)	20(20.6)	
	50~59	66(13.2)	14(14.4)	
	≥60	70(14.0)	12(12.2)	
Job	In or self employment	319(63.8)	64(66.0)	0.682
	Others (House work, student etc)	181(36.2)	33(34.0)	
Education	Middle school or below	58(11.6)	9( 9.3)	0.768
	High school	284(56.8)	58(59.8)	
	University or above	158(31.6)	30(30.9)	

\*Results from chi-square test

적인 특성을 비교한 결과 통계적으로 유의한 차이를 보인 변수는 없었다.

EQ-5D로 측정한 본인의 건강상태는 (Table 2)와 같다. EQ-5D의 5개 도메인에서 본 조사와 재조사 간 상태의 차이는 통계적으로 유의하지 않았다. 전체적인 건강상태를 나타낸 EQ-VAS를 살펴본 결과, 본 조사에서는 중앙값이 83.0이었고, 사분위수 범위가 15로 나타났고, 재조사에서는 중앙값이 85.0이었고, 사분위수 범위가 10이었다. 두 군을 월콕슨 부호화 순위 검정을 통해 비교한 결과, 통계적으로 유의한 차이는 없었다( $p=0.325$ ).

### 2. 타당도

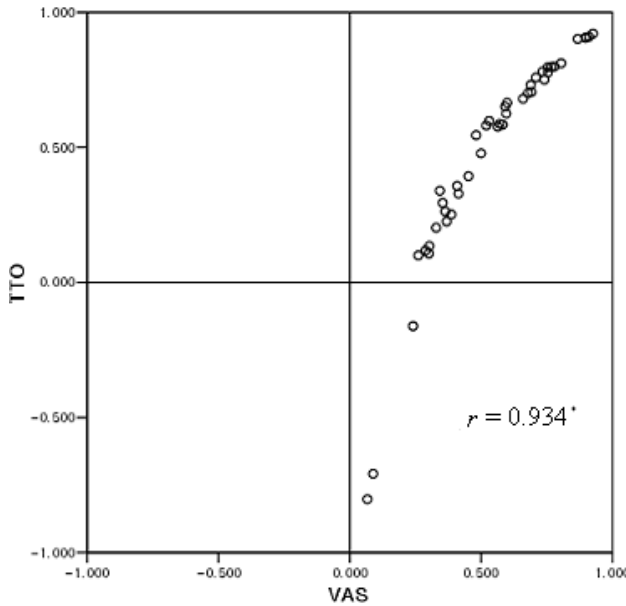
#### 1) 수렴 타당도

수렴 타당도를 살펴보기 위해 VAS 값과 TTO 값의 상관분석을 시행한 결과, 피어슨 상관계수가 0.934( $p<0.001$ )로 나타났다. 그리고 그 관련성을 회귀식으로 살펴본 결과,  $TTO=1-(1-VAS)^{1.028}$ 로 나타났고,  $R^2$ 은 0.973이었다(Figure 2). 재조사

**Table 2.** Self evaluated health states expressed by EQ-5D of study participant

Level	EQ-5D									
	Mobility N(%)		Self-care N(%)		Usual activities N(%)		Pain/discomfort N(%)		Anxiety/depression N(%)	
	Test	Retest	Test	Retest	Test	Retest	Test	Retest	Test	Retest
1	490 (98.0)	95 (97.9)	498 (99.6)	97 (100.0)	490 (98.0)	96 (99.0)	418 (93.6)	78 (80.4)	419 (93.8)	84 (86.6)
2	10 (2.0)	2 (2.1)	2 (0.4)	0 (0.0)	10 (2.0)	1 (1.0)	81 (16.2)	19 (19.6)	81 (16.2)	13 (13.4)
3	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.2)	0 (0.0)	0 (0.0)	0 (0.0)
p*		1.000		1.000		1.000		0.546		0.545

\*Results from chi-square test



**Figure 2.** Convergent validity measured by correlation measures between VAS and TTO

\* Pearson's correlation coefficients ( $p < 0.001$ )

같은 건강상태에 대해서 2가지 방법, VAS와 TTO를 이용하여 가치 평가 작업을 수행하였음. 만일 두 방법이 잘 수행된다면, 같은 상태에 대해서 비슷한 값으로 평가할 것이고, 그렇지 않다면 상관성이 작을 것임. 이러한 측면을 보는 것이 타당도 분석 방법 중 수렴 타당도(convergent validity)임. 전체적으로 상관성이 매우 높게 나타났다.

에서도 거의 유사한 결과가 나왔는데,  $TTO = 1 - (1 - VAS)^{1.085}$ 였고,  $R^2$ 은 0.969였다.

### 3. 자료의 질

타당도의 또 다른 측면으로 자료의 질을 모순값으로 살펴본 결과, 본 조사에서는 모순값의 수가 0인 응답자의 수가 333명(66.6%)이었고, 재조사에서는 56명(57.7%)이었다(Table 3). 두 조사에서 모순값의 수가 3이하인 응답자 수가 각각 488명(97.6%)과 95명(97.9%)으로 유사하였다.

**Table 3.** Numbers of inconsistencies

No. of inconsistencies	Test			Retest		
	N	%	Cumulative %	N	%	Cumulative %
0	333	66.6	66.6	64	66.0	66.0
1	109	21.8	88.4	24	24.7	90.7
2	30	6.0	94.4	4	4.1	94.8
3	16	3.2	97.6	3	3.1	97.9
4 or more	12	2.4	100.0	2	2.1	100.0
Total	500	100.0	-	97	100.0	-

### 4. 신뢰도

신뢰도 측정을 위해 조사-재조사한 자료에 대해서 개인별 ICC를 측정한 결과, 평균은 0.857이었으며, 표준편차는 0.102였다. 최소값은 0.522였고, 최대값은 0.999였다. 그리고 ICC를 일반적인 특성에 따라 나누어 살펴본 결과, 성별, 연령군, 직업, 그리고 학력에 따른 평균 ICC의 차이는 거의 없는 것으로 나타났다(Table 4).

**Table 4.** Comparison of intra-class correlation coefficients by sex, age, job, and education

		n	Mean	S.D.	p
Sex	Female	49	0.844	0.114	0.227*
	Male	48	0.869	0.089	
Age	20~29	24	0.849	0.091	0.531†
	30~39	20	0.848	0.110	
	40~49	17	0.885	0.090	
	50~59	14	0.874	0.102	
	60~	12	0.827	0.127	
Job	In or self employment	64	0.854	0.108	0.673*
	Others (House work, student etc)	33	0.863	0.092	
Education	Middle school or below	9	0.856	0.101	0.970†
	High school	58	0.859	0.095	
	University or above	30	0.853	0.118	

\*Result from t-test

† Result from ANOVA test

### 고 찰

이 연구는 한국에서 일반 인구집단을 대상으로 TTO를 이용하여 건강상태별 가치 평가 작업을 처음으로 수행한 연구이다. 수렴 타당도와 모순값의 수로 타당도 수준을 평가하였고, ICC를 이용하여 조사-재조사 신뢰도를 평가하였다.

연구 응답자 특성과 관련하여, 최고령 응답자가 69세였기 때문에 70세 이상의 고령 인구의 특성을 반영하지 못한 제한점이 존재한다. 연구 응답자의 특성이 가치 평가 작업에 작지만 영향을 미칠 수 있기 때문에 조심스러운 결과 해석이 필요하다. 다만, 2차 조사에 참여한 인구집단을 1차 조사에 참여한 집단과 비교한 결과, 통계적으로 유의한 차이가 없는 것으로 보아 재조사 결과와의 비교에 대한 타당도가 높은 것으로 생각할 수 있다.

타당도는 검사도구가 측정하려는 것을 얼마나 충실히 측정하였는지를 평가하는 것이다. 검사 내용이 측정하려는 것을 얼마나 잘 반영하는지를 주관적으로 판단하는 기준으로는 내용 타당도(content validity)나 대면 타당도(face validity) 등이

있는데, 이러한 기준은 가치 평가 작업의 타당도 평가에는 잘 이용하지 않는다.<sup>20)</sup> 일반적으로 가치평가 과정에 대한 타당도 평가에는, 여러 가지 가치 평가 방법을 이용하여 동일한 건강상태에 대해 평가하였을 때 그 상관성이 높게 나타난다면 비교적 잘 평가하였다고 할 수 있을 것이기 때문에, 수렴 타당도로 타당도를 평가한다. 더불어 타당도의 또 다른 측면에서 모순값의 수로 자료의 질을 살펴보았다. EQ-5D의 경우 건강상태를 체계적으로 나타내었기 때문에 더 좋거나 혹은 더 나쁜 건강상태가 존재할 수 있다. 더 나쁜 건강상태를 더 좋게 평가한다면 이러한 결과는 EQ-5D 건강상태에 대해서 잘 이해하지 못했거나 가치 평가가 적절하게 이루어지지 못하였음을 의미한다.

이 연구에서는 수렴 타당도를 보기 위해서 VAS와 TTO를 같은 건강상태에 대해서 평가하도록 한 후 상관관계를 살펴보았다. 두 질 가중치 값을 변환한 상태에서 상관분석을 한 결과 상관계수가 0.934로 매우 높게 나왔다. Badia 등<sup>19)</sup>의 연구에서 보고한 0.94와 비슷한 수준으로 볼 수 있다. 또한 기존에 알려져 있는 바와 같이 경한 상태에서는 VAS가 더 낮게 나타났고, 중한 상태에서는 TTO에서의 값이 더 낮았다. 이러한 연구 결과도 Badia 등<sup>19)</sup>의 연구 결과와 일치하였다. 이러한 결과에 대한 이유를 설명한 연구는 많지 않은데, VAS로 가치 평가할 때 종종 응답자들이 시간을 무시하는 경향이 있어 점수를 낮게 주었다가 시간을 교환하는 단계에서는 이를 심각하게 받아들이기 때문이라는 설명이 있다.<sup>21)</sup>

TTO와 VAS의 관계를 회귀모형으로 살펴보았다. 일반적으로 그 관계식은  $(1-TTO)=(1-VAS)^a$ 의 형태로 나타나는 것으로 알려져 있다.<sup>22)</sup> 이 연구에서의 TTO와 VAS의 관련식도 비슷한 양상으로 나타남을 확인할 수 있었다. 지수의 형태로 나타나는 회귀계수는 연구마다 다양하게 0.47에서 0.64 정도로 보고하였다.<sup>22-25)</sup> 건강상태의 범위를 '사망'을 0, '건강한 상태'를 1로 하는 범위를 취했는가, 아니면 '상상할 수 있는 최저의 건강상태'에서 '최고의 건강상태'까지의 범위를 설정하는가 하는 기준점에 따라 또는 독립변수가 VAS인가 혹은 TTO인가에 따라 결과가 다르게 나타날 수 있다. 이 연구에서는 회귀계수가 1.028이었는데 TTO를 독립변수로 하면 0.973으로 외국의 연구보다 크게 나타났다. 이는 다른 연구와는 달리 TTO와 VAS값이 거의 비슷함을 의미하는데 이러한 차이를 보이는 이유에 대해서는 추가적인 연구가 필요할 것이다.

또 다른 타당도의 측면에서 자료의 질을 살펴보기 위해 모순값의 수를 살펴보았다. 결과에 제시한 바와 같이 모두 정확하게 한 경우가 66.6%를 차지하였고, EuroQol group에서 제안하고 있는 3개까지 모순값을 허용하면 전체 중 97.6%의

자료를 포함하게 되는데 이 수준은 영국<sup>15)</sup>, Greiner 등이 수행한 독일 연구<sup>26)</sup> 및 Badia 등의 스페인 연구<sup>27)</sup> 결과보다 다소 높은 수준이라고 할 수 있다. 이러한 요인에는 몇 가지 인자가 영향을 미쳤을 가능성이 있다. 우선, 고려해 봐야 할 것은 면접자 효과이다. 2차례의 면접자 교육을 통해서 면접자가 간단한 설명만을 하도록 역할을 제한하였음에도 불구하고, 응답자에게 여러 가지 요인으로 영향을 미쳤을 가능성이 있다. 또한 응답이 잘 되지 않은 경우는 설문조사를 중단하였기 때문에 응답에서 불확실한 반응을 보인 응답자가 애초에 배제되었을 가능성도 존재한다. 그리고 연구 방법에서 처음에 순위를 부여하고 점수를 부여한 다음에도 다시 수정할 수 있는 기회를 충분히 준 것도 모순값이 적게 나타나 요인 중의 하나일 것이다. 이러한 측면에서 면접자가 영향을 미치는 것을 최소화하면서도 응답자가 건강상태를 인지하고 가치 평가 작업을 충분히 잘 수행할 수 있도록 기회를 준다면 보다 효과적으로 가치 평가 작업을 수행할 수 있을 것이다.

신뢰도의 부분에서도 큰 문제는 없는 것으로 나타났다. 신뢰도는 값이 일관되게 나타나는지를 평가하는 지표인데, 평가자 간 신뢰도(inter-rater reliability), 검사-재검사 신뢰도(test-retest reliability), 내적 일치도 신뢰도(internal consistency reliability) 등이 있다. 이 연구에서는 동일한 상태를 같은 방법으로 2회 동일한 사람이 평가하는 것이므로 검사-재검사 신뢰도를 가지고 평가하였는데, 평균 6주간의 차이를 두고 재검사를 수행하여 1차 검사 결과와 비교하였다. 이 연구에서 개인별 ICC를 구해 평균을 구한 결과 0.861로 나타났는데, Badia 등<sup>19)</sup>이 구한 스페인 연구에서의 ICC는 0.84로 이 연구의 결과와 비슷하였으며, 영국 MVH 연구<sup>15)</sup>의 ICC는 평균이 0.73으로 이 연구에서보다 낮게 나타났다. 이는 일정한 간격을 두고 가치 평가 작업을 하여도 유사한 값을 도출할 수 있다는 것을 의미하므로, 가치 평가 작업이 신뢰도 있게 진행되었다고 판단할 수 있을 것이다. 또한 이러한 ICC 값이 성별이나 연령, 혹은 직업이나 학력에 따라 크게 다르지 않게 나타났기 때문에 다양한 집단에서 TTO를 적용할 수 있을 것이다.

이 연구 결과, 한국에서도 TTO를 이용한 건강상태에 대한 가치평가 작업이 충분히 가능할 것으로 판단하였다. 현재 이 연구 결과를 바탕으로 전체 EQ-5D 건강상태에 대한 질 가중치를 추정하기 위한 연구가 진행되고 있다. 이러한 일들이 마무리된다면 한국의 질 가중치에 대한 요율표(tariff)가 완성될 것이고, 이를 이용한 질병부담 연구, 경제성 평가 연구, 혹은 임상 연구들이 보다 활성화될 수 있을 것이다.

## 참고문헌

1. Drummond MF, O'Brien B, Stodart G, Torrance G. Methods for the Economic evaluation of health care programmes. 3<sup>rd</sup> ed. New York: Oxford University Press, 2005.
2. Murry CJL, Salomon JA, Mathers CD, Lopez AD. Summary measures of population health: concepts, ethics, measurement and applications. WHO, Geneva. 2002.
3. Guideline for the economic evaluation of health technologies. 3<sup>rd</sup> ed. Ottawa, ON, Canada: Canadian Agency for Drugs and Technologies in Health (CADTH), 2006.
4. Gold MR, Siegal JE, Russell LB, Weinstein MC. Cost-effectiveness in Health and Medicine. New York: Oxford University Press, 1996.
5. Patric DL, Bush JW, Chen MM. Methods for measuring levels of well-being for a health status index. *Health Serv Res* 1973; 8:228-245.
6. Kaplan RM, Anderson JP. The general health policy model: update and applications. *Health Serv Res* 1988;23:203-235.
7. Torrance GW, Boyle MH, Horwood SP. Application of multi-attribute utility theory to measure social preferences for health states. *Oper Res* 1982;30:1043-1069.
8. Torrance GW, Feeny DH, Furlong WJ, et al. Multiattribute utility function for a comprehensive health status classification system: Health Utilities Index Mark 2. *Med Care* 1996;34:702-722.
9. EuroQol Group. EuroQol: a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199-208.
10. Nord. E. EuroQol: health-related quality of life measurement. Valuations of health states by the general public in Norway. *Health policy* 1991;18(1):37-48
11. Brooks RG. EuroQol: the current state of play. *Health Policy* 1996;37:53-72.
12. Jo MW, Lee SI. General population time trade-off values for 42 health states in South Korea. *J Prev Med Public Health* 2007;40 (2):169-176 (Korean)
13. Dolan P, Gudex C, Kind P, Williams A. Valuing health states: A comparison of methods. *J Health Econ* 1996;15:209-31.
14. Tsuchiya A, Ikeda S, Ikegami N et al. Estimating an EQ-5D population value set: the case of Japan. *Health Econ* 2002;11: 341-53.
15. Dolan P. Modeling valuation for EuroQol Health States. *Med Care* 1997;35(11):1095-108.
16. Park SM, Park MH, Won JH et al. EuroQol and survival prediction in terminal cancer patients: a multicenter prospective study in hospice-palliative care units. *Support Care Cancer* 2006;14(4):329-333.
17. Lee HY, Park EC, Kim HJ, Choi JY, Kim HN. Cost-utility analysis of cochlear implants in Korea using different measures of utility. *Acta Otolaryngol* 2006;126(8):817-823.
18. Brooks R, Rabin R, Charro F. The Measurement and valuation of health status using EQ-5D: A European perspective. Dordrecht. The Netherlands: Kluwer academic publisher. 2003.
19. Badia X, Monserrat S, Roset M, Herdman M. Feasibility, validity, and test-retest reliability of scaling methods for health states: The visual analogue scale and the time trade-off. *Qual Life Res* 1999;8:303-10.
20. Ohinmaa A, Sintonen H. Inconsistencies and modelling of the Finnish EuroQol (EQ-5D) preference value. In: Greiner W, Schulenburg J-M Graf v.d, Piercy J, editors. EuroQol Plenary Meeting, 1-2 October 1998. Discussion papers. Centre for Health Economics and Health Systems Research, University of Hannover, Germany. Uni-Verlag Witte, 1999:57-74
21. Robinson A, Dolan P, Williams A. Valuing health status using VAS and TTO: what lies behind the numbers? *Soc Sci Med* 1997;45(8):1289-97
22. Torrance GW. Measurement of health states: An empirical evaluation of three measurement techniques. *Socioecon Plann Sci* 1976;10:129-136.
23. Loomes G. Disparities between health state measures: An explanation and some implications. York, UK: Department of Economics, University of York, 1988.
24. Van Busschbach, J. De validiteit van QALY's(The validity of QALY's). PhD thesis. Rotterdam, The Netherlands.: Erasmus University Rotterdam, Sanders Institutes, 1994.
25. Stiggelbout AM, Eijkemans MJC, Kiebert GM et al. The utilities of the visual analog scale in medical decision making and technology assessment. *Int J Tech Ass Health Care* 1996;12(2) 291-298.
26. Greiner W, Claes C, Busschbach JJ, Graf von der Schulenburg JM. Validating the EQ-5D with time trade off for the German population. *Eur J Health Econ* 2004;6(2):124-30.
27. Badia X, Roset M, Herdman M, Kind P. A comparison UK and Spanish general population time trade-off values for EQ-5D health states. *Med Decis Making* 2001;21:7-16.

[ Abstract ]

## Validity and Reliability of Korean EQ-5D Valuation Study Using the Time-trade off Method

Min-Woo Jo, Sang-Il Lee<sup>1</sup>

Department of Preventive Medicine, College of Medicine, Dongguk University

Department of Preventive Medicine, University of Ulsan College of Medicine<sup>1</sup>

<b>Background</b>	There has been no study for testing the applicability of time trade-off method to general population in South Korea. This study aimed to examine the validity and reliability in obtaining EQ-5D value set with the time trade-off (TTO) method in general population of South Korea.
<b>Methods</b>	For the initial survey, 500 persons were recruited from general population in Seoul and Gyeonggi province. Ninety-seven among them took part in the second survey. Direct valuation was elicited for 43 health states (42 EQ-5D health states and 'unconscious' state) in face-to-face interviews using the visual analog scale (VAS) and TTO method. Validity was assessed with the number of inconsistent responses and the relationship between VAS and TTO scores. Intra-class correlation coefficient (ICC) was calculated to evaluate the test-retest reliability.
<b>Results</b>	Pearson's correlation coefficient between VAS and TTO values was 0.934 ( $p < 0.001$ ). The relational formula was $TTO = 1 - (1 - VAS)^{1.028}$ and $R^2$ was 0.973. The mean ICC was 0.857 (95% confidence interval: 0.836 - 0.878). Twelve persons (2.4%) had 4 or more inconsistent responses.
<b>Conclusions</b>	We confirmed that TTO method can be used as a valid and reliable tool for valuing EQ-5D health states in general population of South Korea. (Korean J Health Promot Dis Prev 2007; 7(2):96~103)
<b>Key words</b>	EQ-5D, Time trade-off, validity, reliability

• Address for correspondence : Sang-Il Lee  
Department of Preventive Medicine, University of Ulsan College of Medicine  
• Tel : 02-3010-4284  
• E-mail : sleemd@amc.seoul.kr