

한국인 질병의 장애가중치 측정 및 신뢰도 평가

이중규, 윤석준¹⁾, 도영경, 권영훈, 김창엽²⁾, 박기동, 김용익, 신영수

서울대학교 의과대학 의료관리학교실, 고려대학교 의과대학 예방의학교실¹⁾, 서울대학교 보건대학원²⁾

Disability Weights for Diseases in Korea

Jung-Kyu Lee, Seok-Jun Yoon¹⁾, Young Kyung Do, Young Hoon Kwon, Chang-Yup Kim²⁾,
Kidong Park, Yong-Ik Kim, Young-Soo Shin

Department of Health Policy and Management, Seoul National University College of Medicine,
Department of Preventive Medicine, Korea University College of Medicine¹⁾
Department of Public Health, Graduate School of Public Health, Seoul National University²⁾

Objectives : This study aimed to develop an evaluation protocol of disability weights using person trade-off, and to test the reliability of the developed protocol in a Korean context.

Methods : To develop the valuation protocol, the Global Burden of Disease(GBD) and the Dutch studies were replicated and modified. Sixteen indicator conditions were selected from the Korean version of disease classification, which was based on that of the GBD Study, and the person trade-off method referred to the Dutch method.

Results : The disability weights were valued in a two step panel study. The first step was a carefully designed group process by three panels, using person trade-off to establish the disability weights for sixteen selected indicator conditions. The second step consisted of interpolation of the remaining diseases, on a disability

scale, by the individual members of three panels. The members of three panels were all medical doctors, with sufficient knowledge of the consequences of a broad variety of diseases. The internal consistency of the Korean disability weights was satisfactory. Considerable agreement existed within each panel and among the panels.

Conclusions : It was feasible to use a modified evaluation protocol from those used in GBD and Dutch studies. This would provide a rational basis for an international comparative study of disability weights.

Korean J Prev Med 2003;36(2):163-170

Key Words: Cost of illness, Health status, Social value, Reliability, Resource allocation, Korea

서론

보건의료 분야에 있어서 정책의 우선 순위 결정(priority setting)은 매우 중요한 문제 중의 하나이다. 특히 이러한 우선순위 결정은 자원배분(resource allocation)의 문제와 밀접하게 연관이 되어 있기 때문에 더욱 그러하다. 보건학적으로 어떠한 질병이 자원배분의 우선순위가 높다는 것은 그만큼 그 질병이 사회적으로 부담이 높다는 것을 의미하며, 바로 이러한 사회적 부담을 줄이기 위한 공중보건학적 개입의 필요성을 반증하는 것이기도 하다.

그런데 수 백 개에 이르는 질병들 가운데서 어떠한 질병의 사회적 부담이 가장 높은지, 그리고 서로 다른 질병들 사

이에 어떠한 질병의 사회적 부담이 더 큰 것인지에 대한 비교를 어떻게 측정할 수 있을까?

일반적으로 서로 다른 질병들 사이의 중요성을 평가하는데 사망지표와 상병지표가 많이 사용되어 왔다. 그러나 전통적인 건강수준 측정지표인 사망지표 및 상병지표의 유용성의 한계가 제기되었고, 기존 건강지표들이 안고 있는 문제점 외에도 건강중재(health intervention)의 비용-효과를 분석하기 위해서는 중재의 편익을 건강상의 단일지표(single measure)로 합성할 필요성이 대두되었다. 이러한 문제의식을 바탕으로 세계보건기구(World Health Organization, 이하 WHO)와 세계질병부담연구(the Global Burden of Disease Study, 이하 GBD) 그

룹에서는 지난 10여 년 간 국제질병부담 연구를 수행하였다. GBD 그룹은 연구 과정에서 다양한 역학 지표와 인구학적 지표를 사용하여 결과를 제시하였는데, 그 중 장애보정생존년(Disability Adjusted Life-Year, 이하 DALY)은 GBD 연구를 위해 특별히 고안된 것으로 조기사망(Years of Life Lost, 이하 YLL)과 비치명적 건강 결과인 장애(Years Lost due to Disability, 이하 YLD)로 인한 질병부담을 종합한 측정지표이다 [1,2].

$$DALY = YLL + YLD$$

특히 DALY의 구성요소 중 하나인 YLD의 측정을 위해서는 각각의 질병들에 대한 장애가중치(disability weight) 측정이 필수적이다. 그러나 이 장애가중치 측정에 있어서 가장 큰 어려움은 동일한 질병에 대해서도 사람들이 느끼는 심각

접수 : 2002년 11월 19일, 채택 : 2003년 3월 31일

*본 연구는 보건복지부 보건의료기술 연구개발사업의 지원에 의해 이루어진 것임.(01-PJ1-PG1-01CH10-0007)

책임저자 : 이중규(서울시 종로구 연건동 28번지, 전화 : 02-740-8363, 팩스 : 02-743-2009, E-mail : jungkyu@snu.ac.kr)

성 혹은 경중의 정도에 차이가 발생한다는 것이다. 이는 각각의 사람들이 가지고 있는 사회적 가치(social value) 혹은 선호도(preference)의 차이에 기인한 것이다.

사람들이 서로 다르게 느끼는 사회적 가치 혹은 주관적인 선호도를 표준화하여 계량화할 수만 있다면, 해당 인구집단 전체의 사회적 가치, 즉 사회적 선호도 또한 측정 가능하게 될 것이다. 최근 이런 차이를 표준화하여 각각의 질병에 대한 사회적 가치를 측정할 수 있는 방법이 개발되었다 [3]. 이러한 표준화 도구로서 Person Trade-Off(이하 PTO), Time Trade-Off(이하 TTO), Visual Analog Scale(이하 VAS), 그리고 Standard Gamble(이하 SG) 등이 알려져 있다.

WHO와 GBD 그룹에서는 질병들의 장애가중치 측정을 위해서 PTO를 이용한 프로토콜을 개발한 바가 있다 [4,5]. 또한 네덜란드를 비롯한 전 세계 주요 국가들이 GBD 그룹이 개발한 프로토콜을 기반으로 하여 질병부담 측정연구를 수행 중에 있다 [6,7].

그러나 우리나라의 경우, 일부 암 질환에 대해서 텔파이법을 이용하여 장애가중치를 측정하는 바 있으나, 선진 외국의 경우와 같이 PTO를 비롯한 여러 방법론을 이용한 본격적인 장애가중치 측정은 이루어진 바가 없다 [8]. 따라서 GBD 그룹을 비롯한 세계 여러 나라에서 이루어지고 있는 장애가중치의 방법론별 측정결과와의 비교분석을 통해 우리나라 질병부담의 수준을 적절히 반영하고, 측정결과와의 타당도가 높은 방법을 찾기 위한 연구의 필요성이 대두되었다.

이에 본 연구는 장애가중치 측정 방법론 개발을 주도하고 있는 GBD 그룹이 개발한 PTO를 이용한 장애가중치 측정 프로토콜을 바탕으로 우리나라 현실에 맞는 새로운 프로토콜을 개발하여 질병들의 장애가중치를 측정하고, 개발된 프로토콜에 대한 신뢰도를 평가하고자 한다.

연구방법

1. 장애가중치 측정 방법론 선정

장애가중치 측정을 위해서는 사람들이 서로 다르게 느끼는 사회적 가치 혹은 주관적인 선호도를 표준화하여 계량화해야 한다. 그러한 방법으로 PTO, TTO, VAS, 그리고 SG 등이 있는데, WHO와 GBD 그룹에서는 질병들의 장애가중치 측정을 위해서 PTO를 이용한 프로토콜을 개발하였다. 본 연구에서는 GBD 그룹과의 비교연구 및 측정결과와의 타당도 분석을 위해서 장애가중치 측정 방법론으로 PTO를 채택하였으며, 프로토콜 개발을 위해서 WHO와 GBD 그룹에서 제시한 프로토콜과 네덜란드 장애가중치 그룹(Dutch Disability Weights Group, 이하 DDW)에서 사용한 방법론을 참고하였다 (Figure 1).

2. 장애가중치 측정 대상 질환 선정 및 지표질환 선정

GBD 그룹은 국제표준질병 및 사인분류체계(International Classification of Diseases, 이하 ICD)의 코드에 나와 있는 질병들을 바탕으로 하여, 107개 질병 및 상해 등으로 구성된 새로운 질병분류체계를 개발하여 사용하였다. 우리나라에서 이 분류체계의 적용가능성을 검토한 결과 질병분류에 있는 전염성질환 중에서

현재 국내에서 발생이 거의 없는 질병들이 다수 존재하고 있으며, 만성질환의 경우, 질병의 크기나 공중보건학적 중요성을 고려할 때 일부 질환의 분리의 필요성이 제기되었다. 따라서 본 연구에서는 한국질병부담 측정연구팀(Korean Burden of Disease Group)이 한국적 상황에 맞게 123개 질병으로 재분류한 한국형 질병분류체계를 따랐다 [9,10].

참고한 GBD 그룹과 DDW 그룹의 장애가중치 측정 방법들은 모든 질환에 대해서 PTO를 시행하지 않고, 우선 일부 지표질환들(indicator conditions)을 선정하여 측정한 후 이 지표질환들의 장애가중치를 기준으로 하여 내삽을 통해 나머지 질환들의 장애가중치를 구하였다. 본 연구에서도 우리나라 사정에 맞는 16개의 지표질환들을 다음의 세 가지 선정기준에 따라 선택하였다. 그 기준은 첫 번째 공중보건학적으로 의미가 있고, 두 번째 비교적 잘 알려진 질환이어야 하며, 세 번째 장애가중치 범위(0에서 1) 전체를 포괄할 수 있는 경중의 질환에서부터 중중의 질환들을 선택해야 한다는 것이었다.

의과대학생을 대상으로 한 예비조사(pilot test)를 통하여 본 연구에 적용 가능하도록 PTO 측정 프로토콜 및 측정양식을 수정하였다.

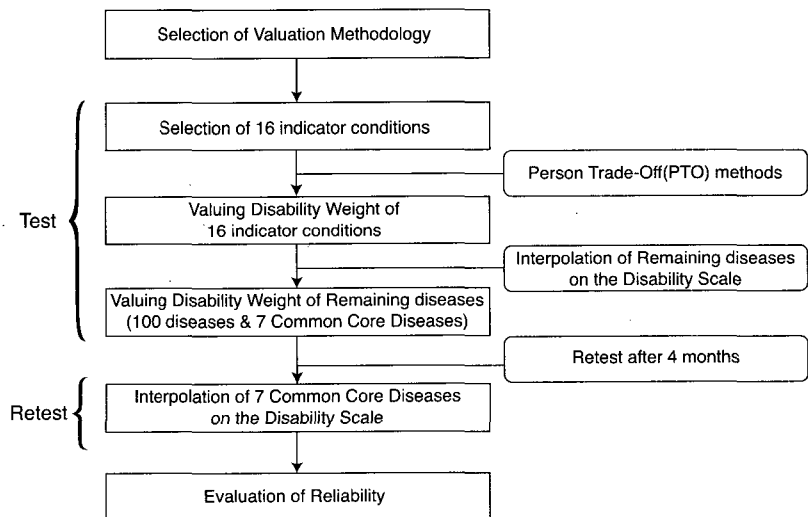


Figure 1. Schematic diagram of study for valuing disability weight in Korea.

3. 지표질환들에 대한 장애가중치 측정

선정된 16개 지표질환들에 대해서 다음과 같은 순서대로 측정을 하였다.

첫째, 보건의료분야의 전문가 10명씩으로 구성된 패널을 세 그룹으로 구성하였다. 본 연구에서는 30명의 패널 구성원들 모두가 의사였다. 성별 구성은 1차 측정 시 1명, 2·3차에 각 2명의 여성들이 참여하였다.

둘째, 각각의 패널에 대해서 서로 다른 날에 16개 지표질환들의 장애가중치를 측정하였다.

셋째, 16개 지표질환들은 개발한 PTO 측정양식 (Figure 2)을 이용하여, 질환별로 약 30분 씩 측정하였으며, 각각의 질환들에 대해서 진행자(facilitator)가 패널 구성원들에게 PTO1과 PTO2를 제시하고 PTO 측정양식에 적도록 하였다. 패널 구성원들은 이렇게 각자 제시한 자신들의 PTO1과 PTO2 값에 대해서 그 이유를 설명하고, 다른 구성원들과 의견개진을 하면서, 3차례 정도 같은 과정을 반복하면서, 각 패널 구성원들은 최종적인 PTO1과 PTO2 값을 제시하였다.

넷째, 질환별로 구해진 PTO1과 PTO2 값을 이용하여 아래 공식으로 장애가중치를 구하였다.

$$DisabilityWeight = 1 - \frac{1000}{\frac{PTO1}{PTO2}} = \frac{1000}{PTO2}$$

다섯째, 패널 구성원들이 측정한 16개 지표질환들에 대한 장애가중치 중앙값을 장애척도에 기록하였다 (Figure 3).

여섯째, 지표질환들 이외의 질환에 대해서는 각각의 패널 구성원들에게 서로 겹치지 않게 30개씩 계통추출한 질환들과 7개의 공통핵심질환들을 합쳐서 총 37개 질환들로 구성된 세트를 나누어주고, 장애척도에 표시된 지표질환들의 장애가중치를 참고하여, 내삽(interpolation)을 이용하여 지표질환을 포함한 123개 전체 질병들에 대한 장애가중치를 측정하였다 (Figure 4).

PTO 1 : Chronic Back Pain

현재 위의 장애를 가진 N명의 사람들이 정확하게 1년 수명을 연장하는 것과 1,000명의 건강한 사람들이 정확하게 1년 수명을 연장하는 선택 사이에 무차별하게 되는 사람들의 수 N

1st	PTO 1	
2nd revised	PTO 1	
3rd definitive	PTO 1	

PTO 2 : Chronic Back Pain

현재 위의 장애를 가진 N명의 사람들이 건강을 완전히 회복하여 정확하게 1년 수명을 연장하는 것과 위의 장애를 가진 N명의 사람들과 더불어 1,000명의 건강한 사람들이 정확하게 1년 수명을 연장하는 선택 사이에 무차별하게 되는 사람들의 수 N

1st	PTO 2	
2nd revised	PTO 2	
3rd definitive	PTO 2	

Figure 2. PTO form.

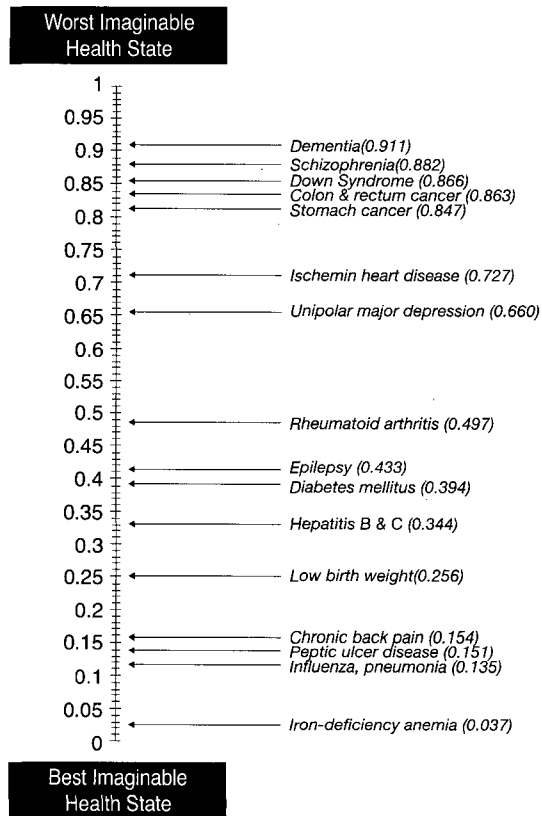


Figure 3. Disability Scale.

4. 신뢰도 평가

장애가중치의 신뢰도 검증은 상관분석을 이용한 패널 간 비교분석, 측정-재측정 결과 비교분석, 그리고 일반화가능도 이론을 이용한 신뢰도 측정으로 이루어졌다.

우선 상관분석을 이용한 패널 간 비교분석은 각 10명으로 구성된 세 패널들이

측정한 16개 지표질환들의 장애가중치의 중앙값(median)을 가지고 상관분석(Spearman rank order correlation)를 시행하였다.

그리고 측정-재측정은 시간변화에 따른 참가자들의 측정의 일관성(consistency)을 알아보기 위한 것이었는데, 이를

위해서 16개 지표질환들 이외에 나머지 107개 질환들 중에서 장애가중치 범위(0에서 1) 전체를 포괄하는 7개의 공통핵심 질환들(common core diseases)을 선정하였다. 이 공통핵심질환들은 16개 지표질환들 이외의 나머지 질환들에 대해서 내삽을 통한 장애가중치 측정을 위해 계통 추출한 30개의 질환들과 함께 세트를 구성하여 패널 구성원들에게 나눠주었다. 그리고 본 측정 4개월 후에 이 7개의 공통핵심질환에 대한 2차 장애가중치 측정 결과들에 대해서 비교분석을 시행하였다. 마지막으로 '행동측정의 신뢰성'에 대한 측정 이론인 일반화가능도 이론(Generalizability theory)를 이용하여, 측정 상황에서 발생하는 다중오차요인

(multiple sources of error)을 동시에 분석하고, 측정치에 대한 오차요인의 상대적 영향력을 산출(Generalizability study, 이하 G study)하여 측정의 일반화가능도 계수(Coefficient of Generalizability, 이하 G coefficient)와 적정수준의 측정조건(Decision study, 이하 D study)를 구하여 연구자들이 개발한 측정방법의 신뢰성을 검증하였다. 측정도구의 신뢰도 검증에 일반적으로 사용되는 Spearman의 고전검사이론(Classical test theory)는 진점수(true score)를 고정시킴으로써 오차요인을 구체적으로 설명하지 못하는 문제가 한계로 지적되어 왔다. 예를 들어 실기시험을 치르는 어떤 피험자가 선발에 관계된 채점자들에 의해 관찰되어질 때, 그 실

기 상황은 수많은 채점자 중 몇 사람에게 의해, 수많은 시간 중 한 경우에 한하여, 여러 종류의 실기행위 중 한 종류가 선택되어, 또는 여러 질문 중 몇 개의 질문형태를 통하여 측정되는 것이다. 이러한 각각의 선택 즉, 관찰대상, 관찰자, 관찰시기, 측정기간, 환경, 문항, 문항의 영역 등은 피험자에 대한 측정결과를 다르게 유도할 수 있는 오차요인으로 볼 수 있는데, 이러한 측정상황에서 발생할 수 있는 오차요인에 대한 설명이 기존의 고전검사이론으로는 불충분하였다. 따라서 연구자들은 단일 오차원(source of error)만을 고려하는 고전검사이론을 확대하여, 다중오차요인을 동시에 고려하는 측정모형에 분산분석(ANOVA)의 체계를 적용한 일반화가능도 이론을 측정도구의 신뢰도 평가 방법으로 채택하였다 [11,12].

Step	Example	Operation
1	the disability weight for indicator condition X is DW(X) = 0.70; and the disability weight for indicator condition Y is DW(Y) = 0.50.	-
2a	Evaluator A interpolated disease Z between indicator conditions X and Y	Disability weight for disease Z : $DW(Z)_A = (0.70+0.50) / 2 = 0.60$
2b	Evaluator B valued disease Z equal to indicator condition X	Disability weight for disease stage Z : $DW(Z)_B = 0.70$
3	-	Eventual disability weight for disease Z : $mean = [DW(Z)_A + DW(Z)_B] / 2 = 0.65$

Figure 4. Calculation of disability weights from the interpolations.

Table 1. Interquartile ranges of three panels' disability weights for sixteen indicator conditions

Indicator Conditions	1st panel median	IQR	2nd panel median	IQR	3rd panel median	IQR
Chronic back pain	0.163	0.156	0.215	0.133	0.082	0.112
Colon and rectum cancers	0.882	0.034	0.833	0.047	0.875	0.042
Down Syndrome	0.900	0.069	0.857	0.007	0.840	0.092
Diabetes mellitus	0.286	0.217	0.600	0.100	0.296	0.151
Peptic ulcer disease	0.078	0.109	0.296	0.119	0.078	0.039
Stomach cancer	0.879	0.025	0.833	0.035	0.829	0.192
HepatitisB and hepatitis C	0.258	0.213	0.574	0.150	0.199	0.329
Influenza, pneumonia	0.069	0.056	0.271	0.190	0.065	0.024
Unipolar major depression	0.667	0.133	0.826	0.063	0.487	0.129
Dementia	0.933	0.022	0.900	0.008	0.900	0.055
Low birth weight	0.167	0.141	0.562	0.132	0.041	0.060
Schizophrenia	0.871	0.042	0.900	0.037	0.875	0.026
Iron-deficiency anemia	0.010	0.005	0.091	0.006	0.010	0.005
Ischemic heart disease	0.750	0.076	0.764	0.028	0.667	0.115
Epilepsy	0.354	0.133	0.688	0.041	0.256	0.238
Rheumatoid arthritis	0.444	0.103	0.667	0.086	0.381	0.272

* IQR : Interquartile range

연구결과

1. 장애가중치 측정 결과

16개 지표질환별 장애가중치의 분포를 패널별 중앙값과 각 패널 구성원들이 측정한 장애가중치의 분포를 사분위 편차(Interquartile range)는 다음과 같다 (Table 1).

2. 신뢰도 평가

1) 패널 간 비교분석

각각 10명으로 구성된 세 패널 그룹들이 측정한 16개 지표질환의 장애가중치의 중앙값들(median)에 대한 상관분석 결과 모두 유의한 상관관계를 보였다 (Table 2). 이는 세 그룹의 패널들 간 장애가중치 결과의 순위(ranking)에 유의한 변동이 없음을 의미한다.

2) 측정-재측정 결과 비교분석

측정-재측정의 결과로 도출된 7개의 공통핵심질환들 간의 ranking correlation 또한 rs가 0.962로 매우 높았으나, 내삽을 통해 측정한 7개의 공통핵심질환들과 재측정을 통해서 나온 7개의 공통핵심질환들 각각의 rs는 낮은 편이었다 (Table 3).

3) 일반화 가능도 이론을 이용한 신뢰도 측정

G study 결과를 보면, 16개의 지표질환들의 기여도가 78.3%로 전체적인 분산 (variance)이 지표질환들에 의해서 설명이 된다는 것을 의미하며, 패널그룹과 패널 구성원들의 오차요인으로서 측정에 미치는 효과는 그다지 크지 않음을 알 수 있다. 나머지 분산 중의 5.2%는 오차원들 사이의 상호작용에 의해서 설명된다 (Table 4).

Table 2. Inter-panel rank correlations

	Panel 1	Panel 2
Panel 2	0.962	-
Panel 3	0.962	0.952

Table 3. Test-retest reliability of disability weights for 7 common core diseases

Diseases	Test		Retest		rs
	mean	median	mean	median	
Diarrheal disease	0.106	0.065	0.103	0.086	0.336
Substance use disorder	0.382	0.414	0.417	0.369	0.365*
Post-traumatic stress disorder	0.362	0.399	0.356	0.300	0.412*
Autism & Asperger's syndrome	0.691	0.708	0.665	0.693	0.348
Multiple sclerosis	0.716	0.748	0.714	0.757	0.404*
CVA	0.796	0.810	0.783	0.787	0.175
Asthma	0.504	0.556	0.401	0.369	0.233

* P-value < 0.05

Table 4. Estimates variance components(percent) of disease Diseases(16) × Occasions(3) × Panel members(30)

Sources of variance	Degree of Freedom	Variance	Contribution %
Disease(D)	15	0.090	78.3
Occasions(O)	2	0.008	7.0
Panel members(subjects) nested within panels(S(O))	27	0.001	0.8
Disease × panel members nested within Occasions(D×S(O))	405	0.010	8.7
All other interactions plus error	30	0.006	5.2
Total	479	0.115	100.0

Table 5. Summary of D-study results

	Sample size		panel member	G coefficient
	indicator number	measuring number		
1	16	1	10	0.92323
2	16	1	15	0.92650
3	16	1	20	0.92817
4	16	2	10	0.96008
5	16	2	15	0.96185
6	16	2	20	0.96273
7	16	3	10	0.97303
8	16	3	15	0.97424
9	16	3	20	0.97484

1) 전체 분산(Sources of variance) 중 지표질환들(Diseases(D))들에 의해서 설명되어지는 분산의 비율

Table 5의 D study 결과를 볼 때, 제시된 9개 연구모델 모두 일반화가능도 계수가 0.9이상으로 높음을 알 수 있었고, 특히 본 연구에서 채택한 연구의 설계(지표질환 16개, 측정횟수 3회, 패널 구성원 10명)의 일반화가능도 계수는 0.973으로 매우 적절하였음을 알 수 있다.

고찰

1. 연구 방법에 대한 고찰

사회적 선호도를 측정하기 위한 가장 최선의 방법에 대한 견해나, 혹은 어떤 사람들이 측정에 참여하는 것이 가장 적절한가에 대한 견해들은 매우 다양하다 [13-16].

이러한 견해들 중에서 일부는 Rating Scale이나 Magnitude Estimation이 가장 적절한 측정방법이라고 주장한 반면, Torrance 등은 von Neumann-Morgenstern에 의해 소개된 기대효용이론에 기반한 Standard Gamble이 가장 좋은 측정 방법이라고 주장하였다. 그러나 이들이 제시한 Standard Gamble은 아주 경미한 질병 혹은 매우 심각한 상태의 질병에 대해서 측정에 참가한 사람들이 편향된 확률을 제시하는 경향이 문제점으로 지적되고 있다. 한편 Standard Gamble에 기반한 또 다른 측정방법으로 Time Trade-Off(TTO)가 있는데, 이 방법 또한 측정에 참가하는 사람들이 가지고 있는 시간에 대한 선호도(time preference)의 차이로 인해 발생하는 문제점이 제기되고 있다.

본 연구에서 장애가중치 측정을 위해 이용한 PTO는 "equivalence of numbers

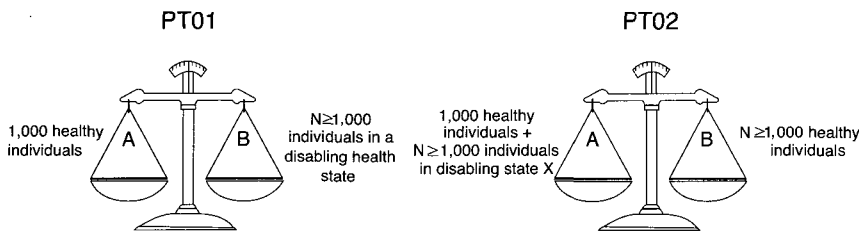


Figure 5. PTO1 and PTO2.

technique”이라는 이름으로 Patrick 등에 의해서 처음으로 사용되었고 [17], PTO 라는 명칭은 Nord에 의해 처음 사용되었는데, 이 측정방법론을 이용하여 GBD 그룹에서 장애가중치를 측정하였다. 즉 PTO의 개념은 Figure 5에 제시된 그림으로 요약될 수 있는데, 이를 조금 더 자세히 설명하면 PTO1은 건강한 사람의 1년 수명연장의 가치(A)와 질병 X에 걸린 사람의 수명을 1년 연장하는 가치(B)의 교환에 대한 것이다. A가 1,000명의 가치합일 때, 이것과 동등하게 교환하기 위해서 B는 몇 명의 가치가 되어야 하는가를 질문하는 것이다. PTO2는 PTO1에서 측정된 가치 A, B의 합(A')과 질병 X에 걸린 사람을 건강한 상태로 회복하여 1년 수명연장의 가치(B')과 마찬가지로 비교하는 것이다.

이렇게 PTO를 이용해서 구해진 장애가중치는 절대적인 가치가 아닌 상대적인 가치이므로 다른 질환에 대한 상대적인 경증도 비교와 다른 집단 혹은 국가의 질환들의 장애가중치와 비교를 할 때만이 그 의미가 있다고 할 수 있다. 따라서 현재 GBD 그룹을 비롯한 유럽지역(European Disability Weights Study)에서도 PTO방법론을 이용하여 각 질병들의 장애가중치를 구하고 있으며, 현재 국가간 지역간 비교연구가 활발하게 이루어지고 있다 [18,19]. 따라서 본 연구도 이와 같은 국제적인 비교연구의 필요성을 위해 PTO 방법론의 적용은 필수적이다.

한편 GBD 그룹에서 장애가중치 측정을 위해 사용한 PTO 방법론은 건강한 사람이 장애를 가진 사람보다 사회적 가치가 더 높은 것으로 판단하게 되어 발생하

는 윤리적인 문제(distributional preference)에 대한 비판과 함께 측정에 참가하는 사람들로 하여금 진행자가 묵시적으로 선택을 강요함으로써 선호도를 인위적으로 맞추려한다는 비판도 있다 [20,21].

그러나 이러한 비판이 PTO 방법론 자체의 근본적인 한계를 의미하는 것은 아니다. 이러한 근거로서 Nord는 2000년 European Disability Weights Study에서 PTO3를 제시하며, 기존의 PTO1과 PTO2 측정에서 발생하는 윤리적인 문제를 극복하고자 하였으며, 이는 Mansley 등도 PTO1과 PTO2에 의한 패널 구성원들 각자의 강제적인 내적 일치도(internal consistency)를 맞추기보다는 PTO를 한번만 측정할 것을 권하고 있다. 이처럼 몇몇 비판에도 불구하고 단일지표를 이용하여 건강수준을 측정하기 위해서는 PTO를 비롯한 여러 방법론을 이용한 장애가중치의 측정이 필수적이고 유용하다는 측면에서 공중보건학적 함의는 여전히 존재한다고 볼 수 있다. 결국 이 문제는 측정방법의 지속적인 개선을 통해서 극복해야 할 것이다 [22,23]. 또한 PTO를 비롯한 장애가중치 측정에 대한 다른 방법론에 대한 논의는 추후 TTO와 VAS와 같은 방법론을 적용한 장애가중치 측정을 통해 검토할 수 있을 것이다.

한편 장애가중치 측정 자체가 질병에 대한 것이기 때문에 일반인과 의료인들 사이의 측정결과에 유의미한 차이 발생한다는 견해가 있다. 즉 이와 같이 측정에는 어떠한 사람들이 참여해야 하는가에 대한 물음에 대해서는 GBD 그룹과 DDW 그룹의 연구 결과에 의하면 PTO를 이용한 질병의 장애가중치 측정에 참

여가능한 구성원들이 반드시 의료인일 필요는 없다는 것이 그들의 결론이었다. 다시 말해서 일반인들도 패널 구성원으로 참여가능하며, 이들을 통해서 얻게 되는 질병들의 장애가중치 결과와 의료분야 전문가들에 의해서 측정된 장애가중치 결과를 비교할 때 크게 다르지 않는 것이다. Froberg 등은 건강상태에 대한 사회적 선호도를 측정함에 있어서 측정에 참여하는 사람들에 따라 차이가 발생하는데, 이러한 차이가 발생하는 이유는 바로 사회적 가치 혹은 선호도를 측정하고자 하는 건강상태 혹은 질병에 대해서 사람들이 가지고 있는 지식이나 경험의 차이에서 비롯된다고 하였다. 질병에 대한 지식 혹은 경험의 많고 적음으로 인해서 사람들마다 생각하는 사회적 가치에 차이가 발생하기 때문에, 이러한 차이를 최소화하기 위해서는 측정 전에 건강상태 혹은 질병에 대한 사전교육이 필요하며, 이를 통해서 지식의 많고 적음으로 인해 발생하는 문제를 해결할 수 있다고 하였다 [24-26]. GBD 그룹과 DDW 그룹은 실제로 장애가중치 측정에 일반인들을 참여시킴으로써 이를 확인하였다. 다만 이처럼 일반인들이 참여하기 위해서는 질병에 대한 충분한 사전교육이 전제되어야 하며, 이는 질병에 대한 충분한 사전 이해 없이 일반인들을 측정에 참여시킬 수는 없다는 것을 의미한다. 본 연구는 의료인들만을 장애가중치 측정에 참여시켰기 때문에 일반인들에 의한 장애가중치 결과를 비교할 수 없었다. 그러나 추후 연구에서는 이에 대한 보완이 필요할 것으로 생각된다.

또 다른 측면은 측정에 참여한 사람들의 판단에 영향을 줄 수 있는 측정과정의 문제이다. 측정과정에서 발생할 수 있는 문제는 두 가지 유형으로 나뉘는데, 첫 번째는 측정자들 각자 판단의 한계(limitations in judgement)로 인해 발생하는 선호도의 불일치(inconsistencies)인데 framing effect라고도 한다. 두 번째는 상황특이적(situation-specific) 변수로 인해 발생하는 불일치의 유형이다. 우선 첫 번째 유형으로 인한 문제는 측정방법을 개

선함으로써 해결이 가능한데, 본 연구에서는 PTO1과 PTO2를 이용하여 framing effect를 최소화시켰다. 두 번째 유형은 건강상태(health status)의 정의에 대한 불일치로 발생하는데, 이 문제는 연구자가 각각의 건강상태에 대한 정의를 표준화함으로써 해결 가능하다. 네덜란드 그룹에서는 EuroQol 5D+C를 이용하여 각각의 질병을 정의함으로써 표준화를 시도하였는데, 본 연구에서는 질병에 대한 별도의 정의를 하지 않았다. 이유는 측정에 참여하는 구성원 모두가 의사였기 때문에 각 질병의 심각성이나 예후에 대해서 충분히 알고 있다는 것을 전제로 하였다. 그러나 후후 장애가중치 측정을 의료분야에 종사하지 않는 비전문가들까지 확대하기 위해서는 각 질병들에 대한 명확한 정의를 통해서 표준화가 되어야 할 것으로 생각된다.

한편 측정상황에서 발생하는 다중오차요인(multiple sources of error)을 동시에 분석하고, PTO를 이용한 측정과정과 결과를 일반화시키기 위하여 일반화가능도 이론을 이용하였다 [27,28]. 이 이론은 ANOVA의 체계를 이용하여, 오차요인들을 분산성분(variance component)으로 분해하여 분석하였다. 이 일반화가능도 이론을 이용한 분석을 위해서 Crick과 Brennan에 의해서 개발된 GENOVA를 이용하였다 [29].

G study의 결과는 기여도(contribution)의 비중이 높은 것이 측정의 일반화에 있어서 오차요인으로서 가장 많은 비중을 차지함을 의미한다. 이를 바탕으로 D study는 연구목적에 맞는 가장 최적의 측정조건을 제시하기 위한 과정이며, 이를 위해서는 일반화가능도 계수(G coefficient)를 산출해야 한다.

2. 연구 결과에 대한 고찰

Table 3에서 내삽을 통해 측정된 7개의 공통핵심질환들과 재측정을 통해서 나온 7개의 공통핵심질환들 각각의 rs는 다소 낮은 편이었는데, 이는 질환들의 장애가중치 측정에 있어 우선순위 설정에서는 별다른 차이가 없었으나, 각 질환별 장애

가중치 자체는 측정자들의 선호도에 따라 편차가 발생한 것으로 추정된다.

패널 참가자들의 구성에 있어서는 남녀 구성비를 동일하게 하지 않았다는 점이 한계라고 볼 수 있으나, 일반화가능도 이론을 통해서 분석한 결과(G study) 다중오차요인으로서 패널 구성원들보다는 지표질환의 수가 훨씬 기여도가 컸음을 고려할 때 남녀 구성비는 그다지 큰 영향을 끼치지 못했을 것으로 추정된다. 또한 오차요인으로서 PTO 측정의 횟수와 패널 구성원들의 수가 일반화가능도 계수에 큰 영향을 주지 못한 것은 측정 진행자(facilitator)에 의한 효과로 추정되는데, 이는 진행자가 PTO 측정시에 각 패널 구성원들 사이의 격차를 줄이려는 노력을 한 결과라고 볼 수 있을 것이다.

D study 결과를 통해서 이번 연구가 최적의 측정조건에서 이루어졌음을 알 수 있었을 뿐만 아니라 측정의 최적조건은 단순히 패널 구성원들의 수나 측정횟수를 늘리는 것만으로 측정의 신뢰도가 높아지는 것이 아님을 알 수 있었다. 이는 사전연구(pilot study)에서 얻어진 자료를 분석(G study)하여, D study로 확장시킨다면 실제 연구에 사용되는 측정 절차의 각 조건을 효율적으로 최소화하면서 어느 수준 이상의 신뢰도를 확보할 수 있음을 의미한다.

이와 같이 측정된 장애가중치 결과는 한국에서 전체 질병으로 인한 질병부담으로서 DALY를 측정하는데 중요한 필수적인 요소로 사용될 것이며, 결국 이러한 질병부담측정을 통해 앞으로 보건의료정책의 우선순위 결정에 실질적인 도움을 줄 수 있을 것이다. 그러나 이번에 측정된 장애가중치가 불변의 값은 아니며 시간이 지남에 따라 질병구조의 변천에 따라 장애가중치도 변하게 될 것이며, 따라서 이러한 장애가중치 측정은 주기적으로 이루어져야 할 것이다.

결론

본 연구는 현재 WHO를 필두로 세계 여러 나라에서 질병부담측정 연구의 일

환으로 진행하고 있는 장애가중치 측정 연구 중에서, PTO를 이용한 한국인 질병의 장애가중치 측정도구를 개발하고 이에 대한 신뢰도 평가를 하고자 하였다.

연구를 통해서 측정된 장애가중치는 패널 구성원 사이의 상관관계나 측정-재측정, 그리고 일반화가능도 연구를 통한 분석결과를 볼 때 신뢰할만한 결과임을 알 수 있었다. 이는 WHO의 GBD 그룹과 네덜란드 그룹에서 사용한 프로토콜을 바탕으로 개발한 한국형 장애가중치 측정 프로토콜이 우리나라에서 적용 가능함을 의미하며, 향후 장애가중치에 대한 국제비교연구에 많은 도움이 될 것으로 기대된다. 또한 최근 세계 각국에서 국민들의 상병수준을 단일한 지표로 결합하여 건강수준을 측정하기 위해 산출된 장애가중치에 대한 비교연구가 활발하게 이루어지고 있으며, 이러한 측면에서 한국형 장애가중치 측정도구의 개발이 절실하게 요구되고 있었다는 점에서 이번 장애가중치 측정도구의 개발을 통한 한국인 질병의 장애가중치 측정은 의의가 있다고 하겠다.

참고문헌

1. 한국보건의료관리연구원. 질병부담 추정에 관한 기초연구. 1998
2. 최용준. 건강생활년을 이용한 우리 나라 주요 암 질환의 질병부담 측정. 서울대학교 대학원 의학석사 학위논문. 2002
3. Nord E. The PTO approach to valuing health care programs. *Med Decis Making* 1995; 15: 201-208
4. Murray CIL, Lopez AD. The Global Burden of Disease: A comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020. WHO. 1996
5. Mathers CD, Vos T, Lopez AD, Salomon J, Ezzati M. National Burden of Disease Studies: A Practical Guide, Edition 2.0. WHO. 2001
6. Stouhard MEA. Disability weights for diseases: A modified protocol and results for a Western European region. *Eur J Public Health* 2000; 10: 24-30
7. Stouhard MEA, Essink-Bot ML, Bonsel GJ, Kramers PGN, van de Water HPA, Gunning-Schepers LJ, van de Maas PJ. Disability weights for diseases in the

- Netherlands. 1997
8. Yoon SJ, Kwon YD, Kim BY. Estimate the disability weight of major cancers in Korea using Delphi methods. *Korea J Prev Med* 2000; 33(4): 409-414 (Korean)
 9. 조희숙, 성주현, 최인정, 이태진, 김재용, 송윤미, 정은경, 신해림, 신영수. 한국형질병체계 분류에 의한 사망손실연수의 평가. 대한예방의학회 추계학술대회. 2001
 10. 한국질병부담추정 연구팀. 질병부담연구 및 보건의료체계성과에 대한 세미나. 2002년 3월
 11. 김성숙, 김양분. 일반화가능도 이론. *교육과학사*; 2001(7-59쪽)
 12. Streiner DL, Norman GR. Health Measurement Scale : A practical guide to their development and use, 2nd ed. Oxford: Oxford University Press; 2001. p. 104-143
 13. Froberg DG, Kane RL. Methodology for Measuring Health-State Preferences I : Measurement Strategies. *J Clin Epidemiol* 1989; 42(4): 345-354
 14. Froberg DG, Kane RL. Methodology for Measuring Health-State Preferences II : Scaling Methods. *J Clin Epidemiol* 1989; 42(5): 459-471
 15. Kaplan RM, Ernst JA. Do Category Rating Scales Produce Biased Preference Weights for a Health Index. *Med Care* 1983; 21: 193-207
 16. Torrance GW. Measurement of Health-State Utilities for Economic Appraisal: a Review. *J Health Econ* 1986; 5: 1-30
 17. Patrick DL, Bush JW, Chen MM. Methods for Measuring Levels of Well-Being for a Health Status Index. *Health Serv Res* 1973; 8(3): 228-245
 18. Essink-Bot ML, Pereira J, Packer C, Schwarzingler M, Burstrom K, the European Disability Weights Group. Cross-national comparability of burden of disease estimates: the European Disability Weights Project. *Bull World Health Organ* 2002; 80(8): 644-652
 19. Üstün TB, Rehm J, Chatterji S, Saxena S, Trotter R, Room R, Bickenbach J and the WHO/NIH Joint Project CAR Study Group. Multiple-informant ranking of the disabling effects of different health conditions in 14 countries. *Lancet* 1999 Jul 10; 354(9173): 111-115
 20. Anand S, Hanson K. Disability-adjusted life years: a critical review. *J Health Econ* 1997; 16: 685-702
 21. Arnesen T, Nord E. The value of DALY life: problems with ethics and validity of disability adjusted life years. *BMJ* 1999; 319: 1423-1425
 22. Mansley EC, Elbasha EH. Preferences and person trade-offs: forcing consistency or inconsistency in health-related quality of life measures?. *Health Econ* 2003; 12: 187-198
 23. European Disability Weights Study. Facilitator instructions and manual. 2000
 24. Froberg DG, Kane RL. Methodology for Measuring Health-State Preferences III : Population and Context Effects. *J Clin Epidemiol* 1989; 42(6): 585-592
 25. Froberg DG, Kane RL. Methodology for Measuring Health-State Preferences V : Progress and Research Agenda. *J Clin Epidemiol* 1989; 42(7): 675-685
 26. Ubel PA, Richardson J, Menzel P. Societal Value, the Person Trade-Off, and the Dilemma of Whose Values to Measure for Cost-Effectiveness Analysis. *Health Econ* 2000; 9: 127-136
 27. Cronbach LJ, Gleser GC, Nanda H, Rajartnam N. The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles. John Wiley & Sons Inc; 1972. p. 1-14
 28. Keppel G. Design and Analysis: A Researcher's Handbook, 3rd ed. Englewood Cliffs: Prentice-Hall Inc; 1991. p. 188-202
 29. Crick JE, Brennan RL. GENOVA: A Generalized Analysis of Variance System (FORTRAN IV computer program and manual). Dochester, Mass.: Computer Facilities, University of Massachusetts at Boston, 1982