

## 직장인들의 건강수준 평가를 위한 측정도구의 신뢰도와 타당도 분석

고상백 · 장세진 · 강명근 · 차봉석 · 박종구

연세대학교 원주의과대학 예방의학교실 및 직업의학연구소

= Abstract =

### Reliability and Validity on Measurement Instrument for Health Status Assessment in Occupational Workers

Sang Baek Koh, Sei Jin Chang, Myung Guen Kang, Bong Suk Cha, Jong Ku Park

*Department of Preventive Medicine and Institute of Occupational Medicine,*

*Yonsei University Wonju College of Medicine*

In order to test scaling assumption, and to assess the validity, reliability, and acceptability of the Short Form 36(SF-36) health survey questionnaire, we conducted a survey. Samples were 296 workers who had been employed in small sized companies. All scale passed for item internal consistency(100% success rate) and item discriminant validity(100% success rate). Reliability coefficients were ranged from the lowest 0.51 to the highest of 0.85. For 87.5% of the total workers, inconsistent responses were not observed. Only 3.0% of the total workers failed two or more checks. Factor analysis was performed using principal axis factor method and quartimax rotation.

In this survey, the SF-36 retained available psychometric properties even when used in a generally healthy worker group. But further study with some consideration to develop health status measurement is expected; first, the definition of health status should be rationalized. Second, the measurement of outcome is an important consideration in evaluations of quality of care. But ambiguities hinder understanding of this important topic. Third, internal consistency should be interpreted with caution as an indication reliability because it ignores potentially important sources of variation that can occur over time.

---

Key words : health status, measurement instrument, scaling assumption, reliability, validity

## I. 서 론

오늘날 의료의 목표는 질병의 치료에 그치는 것이 아니라 더 나은 삶을 제공하고, 신체적 기능 및 정신적 안녕상태를 유지하는 것이다. 이는 건강관련 패러다임의 변화를 반영하는 개념이다. 구체적으로 병태생리학적 이상에 근거한 질병중심의 사고로 부터 신체적 기능, 안녕수준, 삶의 질 등에 근거한 건강중심의 사고로의 이동이다. 과거에는 비용과 의료의 질을 분리해서 사고하였다면, 이제는 제한된 비용으로 최선의 결과를 얻고자 하는데 주안점을 두고 있다(Greenfield와 Nelson, 1992). 이에 최근 수년 동안 보건의료 분야에서는 의학적 개입(*medical intervention*)의 효율 및 효과를 측정하고, 의료의 질을 평가하며, 인구집단의 의료수요를 예측하고, 임상사들의 임상적 결정에 도움을 줄뿐만 아니라, 건강수준의 차이점에 대한 원인과 결과를 이해하기 위해 건강을 확인하고 평가할 수 있는 도구를 개발하기 위해 많은 노력을 기울이고 있다(Ware, 1981; Wasson, 1992). 따라서 건강수준에 대한 평가 및 측정은 보건의료 분야에서 최근 가장 활발히 연구되는 분야 중의 하나이다.

그러나 많은 연구자들과 임상사들은 환자의 건강상태를 파악하기 위해 전통적인 임상진단 도구를 이용하고 있으며, 이 측정법을 통해 생물학적 기능과 병태생리학적 기전에 관련된 구체적인 현상을 평가하고자 한다. 이러한 전통적 도구는 의료분야에서 오랜기간 사용되어 왔고, 이미 알고 있는 임상결과와 참고치를 근거로 해서 해석되어지며, 어떤 면에서는 이 객관적 측정법은 환자 진단에 정확성을 기하고 질병치료에 현실적으로 많은 기여를 하였다. 그러나 그 이용범위가 제한적이며 건강수준을 평가하는데 이용하기 어렵고, 신체적 활동, 감정적 안정, 사회적 생활, 전반적 안녕상태와 질병 및 치료에 대한 환자의 경험을 평가하는 데는 한계가 있다. 이에따라 이를 보완하고 인구집단을 대상으로 건강수준을 평가하거나, 환자의 치료결과를 지속적으로 평가하기 위해 건강수준 측정을 위한 새로운 도구를 개발할 필요성이 증가하고 있다.

건강수준을 측정하려는 첫 시도는 Sheldon 등(1935)의 Disability Index, Moskowitz 등(1957)의 PULSES profile, Steinbrocker 등(1949)의 American Rheumatism Association Classification 등이 50년대 이전 부터 시작되었으며, 이때의 측정도구 대부분은 병원에서 임상사들이 신체기능 및 특정질병에 국한된 건강수준을 파악할 목적으로 개발하기 시작하였다. 그 이후 정신건강을 평가하기 위해 Beck (1961), Zung (1965), Hamilton (1967) 등의 우울척도를 비롯하여 현재에도 많이 이용되는 Ware 등(1979)의 Mental Health Inventory, Goldberg 등(1979)의 General Health Questionnaire 등에 이르기 까지 다양한 측정도구가 개발되어 왔다. 이와같은 측정도구들은 과거 흔히 사용되어 왔던 사망률, 이환율(합병증, 전염성 질환 등), 병원이용률, 환자만족도 등과 같은 비교적 쉽게 정의된 요소보다 의료 결과의 측면에서 건강수준을 더 잘 반영하는 특징을 지니고 있으나(Geigle와 Jonnes, 1990), 일차원적인(*unidimensional*) 특성을 지니고 있어 총괄적인 건강수준을 반영하기에 미흡한 면이 있다.

다른 한편으로 임상에서 뿐만 아니라 보건의료정책 수립 및 보건의료 서비스의 공급과 이용이 건강수준에 어떤 영향을 미쳤는지 알아보기 위해 Fanshel과 Bush(1970)의 Quality of Well-Being Scale, Bergner (1976)의 Sickness Impact Profile(SIP), RAND의 General Health Measure(1979)와 Health Insurance Experimental(1980) 등 다차원적인(*multidimensional*) 특성을 지니는 척도가 등장하기 시작하였다. 그러나 이 도구들은 설문항목이 너무 많아 조사 및 분석시 실제적이지 못한 단점이 있다(Jenkinson 등, 1993; Ware, 1993). 예를들어 Health Insurance Experimental의 경우 108개의 문항으로 구성되어 있어 응답하는데 평균 45분이 소요되며, 136 문항으로 구성된 Sickness Impact Profile의 경우에도 조사표를 완성하는데 평균 30분이 걸린다(Stewart 등, 1988). 결국 건강수준을 측정하기 위한 지속적인 연구와 여러 분야의 경험을 토대로 실용적인 측면을 보강하여 최근에는 Nottingham

Health Profile(Hunt 등, 1981), COOP Charts(Nelson 등, 1990), Medical Outcome Study Short Form 36 (Ware와 Sherbourne, 1992) 등의 간단한 측정도구가 개발되었다.

이러한 배경하에 건강수준평가기표는 해결하고자 하는 주요 문제점을 평가하는 지표로서 신뢰도와 타당도가 높아야 한다. 특히 건강수준 측정시 임상적으로 중요한 변화를 확인하고자 할 때 이와 관련성이 높아야 하며, 의료제공자와 환자 모두에게 받아들여져야 한다.

이 중 SF-36(Short Form-36)은 최근에 고안되어 수 정보완된 것으로 8개 항목 36개 문항으로 이루어져 있으며, 임상 연구 및 보건 정책 평가, 일반인을 대상으로 한 조사에 사용될 수 있으며, 신체적, 정신적 상태를 비롯한 건강에 영향을 끼치는 전반적인 면을 다루고 있다. 여러 연구를 통해 타당도와 신뢰도가 검증되었으며 다양한 임상 결과를 추출해 낼 수 있고, 대규모 집단을 대상으로 한 연구도 가능성이 증명되었으며(Brazier 등, 1992 ; Ware와 Sherbourne, 1992; Garrat 등, 1993; Ware, 1993; Jenkinson 등, 1993; Perneger 등, 1995) 비교적 동질성을 갖는 집단 구성원의 건강 수준 측정에도 적절하다(Ziebland, 1995).

우리나라의 경우, 일반인들의 건강을 평가하기 위해 성인병 검진 및 근로자 건강검진 등 집단검진을 시행하고 있으며, 건강자, 요주의자, 환자 등 세가지로 구분하여 질병의 조기발견, 조기치료 및 건강증진을 위해 많은 노력을 기울이고 있다. 그러나 집단을 대상으로 빠른 시일내에 저렴한 비용으로, 많은 종류의 검사를 실시하다보면 수검자의 의견을 반영하기 어렵고, 그 판정결과의 신빙성이라든가, 검사성적이 함축하는 의미가 제한을 받는 등 여러문제가 야기될 수 있다. 또한 이 검사는 신체적 측면만을 강조하고 있으며, 그 중에서도 예방가능하거나 질병관리가 가능한 질병중심으로 검사하고 있어 정신건강 및 사회적 기능에 관련된 측정과 평가가 간과되고 있어 총괄적 건강수준 측정에 미흡한 측면이 있다. 이에 검진시 현재 시행되고 있는 검사항목과 이 연구에서 검토하고자 하는 건

건강수준평가기표를 병행해서 사용한다면 보다 효과적일 것이라 생각된다.

따라서 이 연구는 의료제공자 뿐만 아니라 환자 또는 일반 대중의 관점에서 건강을 확인하고 평가할 수 있는 도구로 알려진 SF-36의 방법론적 타당성을 검증해 보고, 우리나라 일반 근로자들의 신체적 활동, 감정적 안정, 사회적 생활, 전반적인 안녕상태에 대한 그들의 경험을 평가함으로써 보건관리의 기초자료를 마련하고자 하는데 그 목적이 있다.

## II. 연구 방법

### 1. 연구대상 및 자료수집 방법

강원도 W병원 건강관리과에서 보건관리대행을 시행하고 있는 50인 이상 100인 미만의 영세소규모사업장을 대상으로 1996년 8월 1일 부터 1996년 11월 30일 사이에 구조화된 설문지를 배부하여 자기기입에 의한 조사방법으로 자료를 수집하였다. 조사대상자는 24개의 영세 소규모 사업장 총 1608명 중 W시의 한 공단에 위치하는 5개 사업장을 집락추출하였고, 이중 조사 시점 당시 자료수집이 가능했던 310명을 연구대상집단으로 하였다. 310명의 대상자 중 조사항목의 반 이상을 응답하지 않은 자 14명을 제외한 총 296명을 최종 분석 대상으로 하였다.

조사대상자의 일반적 특성을 보면 성별로는 남자는 206명(70.8%), 여자는 85명(29.2%)이었으며, 평균연령은 35.4세(표준편차 9.8)였으며, 34세 이하가 149명(50.3%)으로 가장 많았으며, 35-44세가 93명(31.4%), 45-54세가 31명(10.5%), 55세 이상이 23명(7.8%)이었다. 교육수준은 중졸 88명(29.7%), 고졸 127명(42.9%), 대졸 81명(27.4%)였다. 조사대상자의 직종은 생산직은 111명(37.5%), 사무직(관리직 포함)은 185명(62.5%)이었다.

### 2. 조사도구

조사도구는 구조화된 설문지(별첨)를 이용하였으

며, 크게 사회인구학적 특성, 건강수준에 대한 평가(SF-36)로 구성되었으며, 건강수준을 평가하기 위해 기능수준(functional status), 안녕수준(well-being), 전반적인 건강평가(overall evaluation of health) 등 3개 영역으로 구분하여 조사하였다. 설문지의 구성은 표1과 같다.

건강수준은 9개 하부영역의 총 36문항으로 구성되어 있다. 이중 신체적 기능(PF), 사회적 기능(SF), 신체적 역할제한(RP), 감정적 역할제한(RE), 정신건강(MH), 활력(VT), 통증(BP), 일반 건강(GH) 등의 8개 항목을 이용하여 건강수준의 8개 영역을 측정하고자 하였다. 나머지 다른 하나의 항목은 건강상태의 변화를 조사하였다.

### 3. 분석방법

#### 1) 척도화 가정(Scaling assumption)

건강수준을 평가하는 8개 항목은 Likert 척도로 각 문항을 합산하였다. 즉, 각 문항을 건강에 가장 나쁜 영향을 미치는 내용을 1점으로 하여 문항에 따라 최고 점은 2점에서 6점으로 하였다. 이렇게 점수화한 각 문항을 항목별로 합산하였으며, 합산한 점수는 100점으로 환산하였다. 그 결과 각 항목에서 최고 높은 점수를 받은 자는 100점이 된다(Ware, 1993). 따라서 관련된 문항으로 구성된 항목들은 다음과 같은 가정(scaling assumption)을 만족해야 한다(Ware, 1993;

Mchorney 등, 1994).

#### ① 등분산성

동일한 항목 안에 있는 문항에 응답한 각 개인별 분포 및 분산은 비슷하여야 한다. 그렇지 않으면 항목별로 합산할 때 항목 표준화 과정을 거쳐야 한다. 이 연구에서는 표준편차로 표현하였다.

#### ② 문항내적일치도(item internal consistency)

문항내적일치도는 Howard와 Forehand(1962)가 제안한 방법을 이용하여 Ware(1993)가 통계적 방법으로 제시하였는데, 건강수준을 평가하기 위하여 이미 개념화한 8개의 항목과 각 문항에 대해 상관관계(item-total correlation)를 구하여 평가하였다. 각 항목을 구성하는 문항의 수가 SF-36과 같이 비교적 적을 경우, 이 검정이 유용하게 사용된다. 각 항목과 문항이 겹쳐있는 부분(overlap)의 상관계수가 0.40 이상인 경우를 문항내적일치도가 높다고 평가하였다.

#### ③ 문항판별타당도(item discriminant validity)

문항판별타당도는 이미 개념화된 건강개념(설문지를 구성하는 8개 차원의 건강개념)을 뜻하는 항목과 관련된 문항이 그 항목내에서 통일성을 지닐 수 있는가에 초점을 맞추고 있다. 이는 SF-36의 각 문항이 다른 개념을 의미하는 항목과의 상관계수에 비하여, 같은 개념을 지니는 항목과의 상관계수가 더 높을 때 의미가 있다고 할 수 있다. 총 점정수에 대한 성공수의 백분율로 성공률을 구하여 판별타당도를 검정하였다.

Table 1. Dimension of the health status evaluation

Area		Dimension	No. of questions
General characteristics			14
Health status (SF-36)	Functional status	Physical functioning(PF)	10
		Social functioning(SF)	2
		Role limitation-Physical(RP)	4
		Role limitation-Emotion(RE)	3
	Well-being	Mental health(MH)	5
		Vitality(VT)	4
		Bodily pain(BP)	2
		General health(GH)	5
	Overall evaluation of health	Health change	1

## 2) 신뢰도(reliability)와 타당도(validity)

신뢰도란 측정하려고 하는 것을 얼마나 안정적으로 일관성 있게 측정 하였느냐의 문제로, 검사도구가 얼마나 오차없이 측정하였느냐 하는 문제이다. 즉, 신뢰도란 측정의 일관성을 말한다. 따라서 본 연구에서는 신뢰도를 추정하기 위해 내적 일치도를 나타내는 Cronbach's  $\alpha$  계수를 구하였으며, 응답일관지수(Response Consistency Index; RCI)를 구하였다. Cronbach's  $\alpha$  계수는 건강수준을 평가하기 위한 8개 항목별로 각각 나누어서 SPSS/PC를 이용하여 구하였다. 또한 응답일관지수는 Ware(1993)가 제시한 방법을 이용하였는데, 비슷한 개념을 내포하는 문항 15쌍을 조사하여 산출하였다.

또한 타당도를 알아보기 위하여 구성타당도(construct validity)의 접근방법의 하나인 인자분석을 시행하였다. 건강수준에 대한 인자분석은 탐색적 인자분석을 이용하였으며, 주축인자법과 직각회전(quartimax) 방식으로 고유값 1.0 이상을 기준으로 인자를 선택하여 타당도를 평가하였다.

## III. 결 과

### 1. 척도화 가정(scaling assumption)에 대한 검증

건강수준을 평가하기 앞서 평가도구를 구성하는 척도에 대한 기본 가정을 만족하는지를 알아보았다. 표2는 건강수준표를 구성하는 각 문항에 대한 평균과 표준편차를 보여주고 있다. 같은 항목에 속하는 문항들의 표준편차는 비슷한 분포를 보이고 있다.

문항내적일치도를 보기 위하여 8개의 항목과 각 문항에 대해 상관관계를 구하였다. 또한 표2에 근거하여 각 항목과 동일한 개념을 반영하는 문항이 겹쳐있는 부분의 상관계수의 범위를 표3에 요약, 제시하였다. 신체적 활동 제한, 통증, 감정제한, 정신건강에서 문항-항목 상관관계의 범위가 0.10이하였고, 사회기능 및 일반건강에서 상관계수의 범위가 가장 컸다. 내적일치도에 대한 검증은 모든 영역에서 0.40이상으로 성공률

은 100%였다.

문항판별타당도는 표2의 상관행렬표에서 동일한 행의 같은 개념을 지니는 항목의 상관계수와 다른 개념을 의미하는 항목의 상관계수를 비교하여 검증하였다. 예를 들어 신체적 기능을 평가하는 문항1에 경우 신체적 기능 항목의 상관계수 0.61이 다른 항목의 상관계수들보다 높기 때문에 항목판별타당도는 높다. 이와같이 모든 항목에 대해 항목판별타당도를 구하여 표3에 제시하였고, 그 성공률은 100%였다.

### 2. 신뢰도 및 타당도

내적일치도를 알아보기 위해 Cronbach's  $\alpha$  계수를 구하였다(표4). 신체적 기능 항목의 신뢰도 계수가 0.85로 가장 높았으며, 사회적 기능의 항목의 신뢰도 계수가 0.51로 가장 낮았다. 8개 항목에 대한 신뢰도 검증결과 활력과 사회적 기능 항목을 제외한 다른 항목에서 만족할 만한 결과를 얻었다. 또한 전체 대상자 296명 중에서 259명(87.5%)이 일관된 응답률을 보였고, 두 개 이상 일관되지 못한 응답을 한 사람은 9명으로 3.0%에 불과하였다(표5).

타당도를 평가하기 위해 건강수준에 대한 인자분석을 시행한 결과 고유값 1.0 이상인 인자는 8개였으며, 총 분산의 62.4%를 설명할 수 있었다. 요인1에는 신체적 제한 항목들이 포함되었고 그중 문항1만 요인8에 분류되었다. 요인2는 정신건강 문항1,2,4, 활력 문항3,4, 사회적기능 문항2가 분류되었으며, 요인3에는 역할제한 항목(신체적 역할제한 문항3 제외)이 선정되었으며, 요인4에는 일반건강(문항4 제외)이, 요인5에는 정신건강 문항3,4, 활력 문항 1,2가 포함되었다. 요인6에는 통증 및 사회적기능 문항1이 분류되었으며, 요인7에는 일반건강 문항4, 신체적 역할제한 문항3이 포함되었고, 요인8에는 신체적 기능 문항1이 선정되었다(표6).

### 3. 근로자들의 일반적 건강수준

연구대상자 전체에 대한 건강수준 척도를 표7에 제

**Table 2. Item Mean, Standard Deviation, and Correlations Coefficients<sup>a</sup> between SF-36 Items and Hypothesized Scales**

Item	Mean	S.D	PF	RP	BP	GH	VT	SF	RE	MH
PF1	2.14	.63	.61*	.30	.29	.43	.33	.26	.30	.33
PF2	2.63	.54	.78*	.27	.32	.34	.22	.24	.19	.24
PF3	2.78	.45	.71*	.30	.35	.33	.21	.25	.26	.23
PF4	2.57	.56	.70*	.28	.22	.41	.26	.24	.22	.26
PF5	2.83	.41	.64*	.21	.23	.33	.18	.16	.21	.17
PF6	2.65	.55	.69*	.26	.23	.22	.08	.19	.23	.17
PF7	2.73	.48	.66*	.19	.34	.29	.20	.23	.14	.23
PF8	2.88	.36	.66*	.24	.23	.12	.16	.24	.17	.25
PF9	2.93	.30	.67*	.22	.24	.14	.18	.23	.20	.28
PF10	2.94	.28	.60*	.25	.24	.23	.16	.19	.19	.21
RP1	1.79	.41	.36	.79*	.32	.21	.41	.34	.62	.26
RP2	1.79	.41	.20	.72*	.16	.18	.17	.22	.51	.17
RP3	1.48	.50	.28	.73*	.31	.32	.26	.39	.37	.24
RP4	1.81	.39	.30	.78*	.32	.18	.32	.33	.47	.27
BP1	4.68	1.12	.34	.31	.91*	.44	.32	.46	.32	.27
BP2	4.34	.91	.39	.35	.87*	.39	.34	.39	.23	.29
GH1	3.36	.90	.34	.21	.32	.70*	.28	.25	.24	.27
GH2	3.81	.89	.42	.25	.30	.60*	.30	.24	.21	.39
GH3	3.27	.99	.28	.21	.36	.76*	.32	.24	.29	.33
GH4	3.20	1.05	.21	.23	.22	.59*	.21	.30	.16	.25
GH5	3.25	1.03	.33	.16	.43	.81*	.43	.31	.25	.38
VT1	3.75	1.57	.10	.17	.23	.27	.74*	.22	.25	.38
VT2	3.40	1.58	.21	.26	.28	.37	.76*	.31	.34	.46
VT3	4.90	1.13	.26	.34	.26	.23	.60*	.37	.27	.45
VT4	4.29	1.35	.34	.36	.29	.39	.72*	.38	.42	.54
SF1	3.67	1.07	.19	.34	.42	.29	.32	.88*	.33	.34
SF2	4.48	.67	.37	.32	.30	.33	.37	.61*	.29	.47
RE1	1.78	.41	.27	.58	.29	.22	.45	.30	.76*	.30
RE2	1.75	.44	.23	.52	.21	.26	.22	.28	.83*	.27
RE3	1.65	.48	.27	.47	.25	.31	.42	.39	.82*	.41
MH1	4.36	1.26	.32	.32	.30	.31	.36	.40	.31	.65*
MH2	4.66	1.31	.24	.17	.20	.28	.26	.32	.21	.61*
MH3	3.06	1.67	.16	.16	.21	.36	.59	.29	.32	.68*
MH4	4.68	1.28	.26	.26	.16	.32	.43	.36	.34	.69*
MH5	3.19	1.59	.23	.16	.19	.28	.43	.30	.18	.69

<sup>a</sup>Item-Total correlation for overlap.

\* Correlation between an item and its hypothesized scale (overlap)

PF; Physical functioning SF; Social functioning RP; Role limitation-Physical

RE; Role limitation-Emotion MH; Mental health(MH) VT; Vitality

BP; Bodily pain GH; General health

**Table 3. Results of Item Internal Consistency and Discriminant Validity**

Scale	k <sup>a</sup>	Range of Correlation		Internal consistency <sup>d</sup>		Discriminant Validity <sup>e</sup>	
		Item Internal Consistency <sup>b</sup>	Item Discriminant validity <sup>c</sup>	Success/Total	Success Rate(%)	Success/Total	Success Rate(%)
PF	10	.60 - .78	.08 - .43	10/10	100	80/80	100
RP	4	.72 - .79	.16 - .62	4/4	100	32/32	100
BP	2	.87 - .91	.23 - .46	2/2	100	16/16	100
GH	5	.59 - .81	.16 - .43	5/5	100	40/40	100
VT	4	.60 - .76	.10 - .54	4/4	100	32/32	100
SF	2	.61 - .88	.19 - .47	2/2	100	16/16	100
RE	3	.76 - .83	.21 - .58	3/3	100	24/24	100
MH	5	.61 - .69	.16 - .59	5/5	100	40/40	100

<sup>a</sup>Number of item and number of item-internal consistency tests per scale.

<sup>b</sup>Correlation between items and hypothesized scale corrected for overlap.

<sup>c</sup>Correlation between items and other scale.

<sup>d</sup>Number  $\geq$  0.40.

<sup>e</sup>Number of items with correlations significantly higher/total number of correlations.

PF; Physical functioning SF; Social functioning RP; Role limitation-Physical

RE; Role limitation-Emotion MH; Mental health(MH) VT; Vitality

BP; Bodily pain GH; General health

**Table 4. Cronbach's  $\alpha$  coefficients for subscales of SF-36**

Scale	Cronbach's $\alpha$
PF	.85
RP	.74
BP	.75
GH	.72
VT	.65
SF	.51
RE	.75
MH	.70

PF; Physical functioning SF; Social functioning RP; Role limitation-Physical

RE; Role limitation-Emotion MH; Mental health(MH) VT; Vitality

BP; Bodily pain GH; General health

**Table 5. Frequency distribution for the Response Consistency Index(RCI)**

Inconsistent Responses	f	%	Cum %
0	259	87.5	87.5
1	28	9.5	97.0
2	6	2.0	99.0
3	2	0.7	99.7
4	0	0.0	99.7
5	1	0.3	100.0

**Table 6.** Fator analysis for the SF-36

Factor coefficients of individual questionnaire after rotation								
	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7	Factor8
PF2	.64	MH1 .63	RP1 .74	GH1 .69	MH3 .79	BP1 .69	GH4 .54	PF1 .51
PF3	.67	MH2 .61	RP2 .76	GH2 .42	MH5 .63	BP2 .71		
PF4	.49	MH4 .73	RP4 .62	GH3 .73			RP3 .43	
PF5	.59			GH5 .70	VT1 .73	SF1 .50		
PF6	.66	VT3 .72	RE1 .70		VT2 .75			
PF7	.63	VT4 .69	RE2 .76					
PF8	.82		RE3 .58					
PF9	.83	SF2 .63						
PF10	.74							
Eigenvalue	9.17	3.14	2.33	1.95	1.70	1.35	1.12	1.09

PF; Physical functioning SF; Social functioning RP; Role limitation-Physical  
 RE; Role limitation-Emotion MH; Mental health(MH) VT; Vitality  
 BP; Bodily pain GH; General health

**Table 7.** Normative data for the workers(Total sample)

	PF	RP	BP	GH	VT	SF	RE	MH
Percentage nonrespondent	7.8	3.0	1.0	2.0	1.7	4.4	4.4	3.7
Mean	85.7	72.2	68.2	58.1	61.8	80.0	73.3	60.0
Percentile: 25 <sup>th</sup>	80.0	50.0	60.0	45.0	50.0	62.5	33.3	48.0
50 <sup>th</sup>	90.0	100.0	70.0	60.0	60.0	75.0	100.0	60.0
75 <sup>th</sup>	95.0	100.0	80.0	70.0	80.0	87.5	100.0	76.0
S.D.	15.4	32.3	18.1	16.4	19.7	16.8	36.0	19.7
Range	10-100	0-90	0-100	10-100	0-100	37.5-100	0-100	12-100
% Ceiling	23.4	0.0	18.1	1.0	2.4	18.7	57.6	2.8
% Floor	0.0	9.4	0.3	0.0	0.7	0.0	12.4	0.0

PF; Physical functioning SF; Social functioning RP; Role limitation-Physical  
 RE; Role limitation-Emotion MH; Mental health(MH) VT; Vitality  
 BP; Bodily pain GH; General health

시하였다. 구체적으로 살펴볼 때, 신체적 기능 항목의 경우 평균점수는 85.7(범위 0-100점)이며, 신체적 기능이 완전한 상태에 있는 응답자가 23.4%였고, 최저 점수에 해당하는 응답자는 없었다. 또한 응답자의 75 percentile 값은 95점이었다. 신체적 역할제한 항목의 평균점수는 72.2(범위 0-90점)이며, 최저점수인 0점에 해당하는 응답자는 9.4%였으며, 응답자의 75 percentile 값은 95점이었다. 통증 항목의 평균점수는 68.2(범위 0-100점)이며, 통증이 없이 지내는 자는 18.1%

였고, 통증으로 지장이 많은 자는 0.3%였다. 응답자의 75 percentile 값은 80점 이었다. 일반 건강 항목의 평균점수는 58.1(범위 10-100점)이며, 응답자 중 최고 점수에 1.0%가 분포하고 있고, 응답자의 75 percentile 값은 70점 이었다. 활력 항목의 평균점수는 61.8(범위 0-100점)이며, 응답자 중 최고점수에 2.4%가 분포하고 있고, 최저점수인 0점에 해당하는 응답자는 0.7%였다. 응답자의 75 percentile 값은 80점이었다. 사회적 기능 항목의 평균점수는 80.0점(범위 37.5-100)



이고 응답자 중 최고점수에 18.7%가 분포하고 있으며, 응답자의 75 percentile값은 87.5점이었다. 감정적 역할제한의 평균점수는 73.3(범위 0-100점)이며, 응답자 중 최고점수에 57.6%가 분포하고 있고, 최저점수인 0점에 해당하는 응답자는 12.4%였다. 정신건강 항목의 평균점수는 60.0(범위 12-100점)이며, 응답자 중 최고점수는 2.8%였으며, 응답자의 75 percentile은 76 점이였다.

## V. 고 찰

지난 10여년 동안 세계의 여러 곳에서 경제, 정치 및 사회적 변화와 더불어 보건의료 서비스 분야에서도 근본적인 변화와 성장이 있어 왔다. 구체적으로 보건의료 재정조달과 서비스 체계를 담당하는 정부의 역할 및 보건의료 공급자와 구매자와의 관계가 바뀌고 있으며, 이것이 공급자의 행태와 효율 그리고 진료의 질에 영향을 미치게 되었으며, 그 결과 일반 주민의 건강수준 평가에 많은 관심과 시사점을 던져 주고 있다. 특히 여러 국가에서 의료재정의 부담자측이 의료제공자들에게 비용을 관리하도록 압력을 가함에 따라 의료 결과(medical outcome)에 대한 측정여부와 측정방법을 둘러싸고 논쟁이 가열되고 있다. 의료서비스의 어떤 특성이 결과를 개선시키는데에 대해서 어느 바가 거의 없어도 의료서비스로 인해 국민의 건강수준이 개선되었다고 하면 비용절감을 위한 정책은 성공적이라고 볼 수 있다. 이에 따라 의료결과에 대한 연구의 필요성이 제기되고 있다(Ware, 1993).

보건의료서비스의 이상적 결과는 정상으로 회복되거나, 그 나이와 의학적 조건(medical condition)에 걸맞은 삶의 질을 회복하는 것이다. 질병과 불구(disability)에 대한 건강면접조사가 실시되기 시작한 이래로, 사람들의 건강인지 방식을 신뢰성 있게 측정하는 방법을 찾기 위한 모색이 줄기차게 지속되어 왔다. 최근들어 건강에 대한 자기보고방식의 측정 결과를 표준화하고 이를 이용하여 결과를 모니터링하는데 괄목할 만한 발전이 있었다. 따라서 표준화된 자기보고 연

구가 환자의 관점에서 결과를 측정하는 가장 우수한 방법으로 대두하고 있는 시점(Geigle과 Jones, 1990; Ware, 1992)에서 이러한 방법에 대해 검토해 보는 것은 시의적절하다고 생각된다.

이 연구에서는 최근 고안되어 여러 분야에 널리 쓰이는 건강수준평가표(SF-36)를 우리나라 실정에 맞게 수정 보완하여 근로자들을 대상으로 적용하여 보았다. 설문지는 자기기입에 의한 조사방법으로 조사하였으며, 조사에 응한 응답자는 총 310명 이었으나, 조사항목의 반이상을 응답하지 않은 자 14명을 제외한 총 296명을 분석 대상으로 하였다. 각 항목에 대한 결측치는 신체적 기능 7.8%, 신체적 역할제한 3.0%, 통증 1.0%, 일반건강 2.0%, 활력 1.7%, 사회적 기능 4.4%, 감정적 역할제한 4.4%, 정신건강 3.7%로 전체자료의 완결성은 비교적 양호하였다.

건강수준 평가에 앞서 분석을 제대로 이해하고, 점수가 가지는 의미를 적절하게 해석하기 위해서 문항들을 항목별로 묶어서 점수화하는데 필요한 가정이 선행되어야 한다. 특히 건강수준 평가표가 단순하고 간단한 자기기입식 접근방법의 성격을 띠 경우, 많은 내용을 담고 있는 설문지에 비해 정확도와 포괄성이 상실될 위험성이 있으므로 척도화 가정은 매우 중요한 의미를 지닌다. 물론 설문지가 간단하고 단순하더라도 응답자의 부담감이 감소하기 때문에 정확도가 떨어지지 않는다는 것이 여러 연구를 통해 밝혀졌지만, 반드시 검토하고 넘어가야 할 과제이며 자료의 표준화를 위해서도 이 분석은 필수적이다. 척도화 가정은 문항내적일치도와 문항판별타당도를 통해 평가하였다. 문항내적일치도는 각 항목과 그 항목의 개념을 공유하는 문항간의 상관표에서 겹쳐 있는 부분의 상관계수를 구하여 각 문항이 그 항목에 속할 수 있는지를 보았다. 만약 상관계수가 0.40이하이면 그 문항은 그 항목에 구성이 되기 어렵다고 볼 수 있다. 이 연구에서는 모든 문항이 이를 만족하여 성공률은 100%였다.

문항판별타당도는 이미 개념화된 건강개념을 뜻하는 항목과 관련된 문항이 다른 항목과 구별되어 같은

의미의 항목내에서 통일성을 지닐 수 있는가에 초점을 맞추고 있다. 이는 표2의 상관행렬표에서 동일행의 같은 개념을 지니는 항목의 상관계수와 다른 개념을 의미하는 항목의 상관계수를 비교하여 검정하였다. 신체적 기능을 평가하는 문항1에 경우 신체적 기능 항목의 상관계수 0.61이 다른 항목의 상관계수들보다 높기 때문에 항목 판별 타당도는 높다고 할 수 있다. 본 연구에서 성공률은 100%로 적합한 결과를 보였다.

신뢰도는 측정도구를 어떤 대상에 여러번 적용할 경우 동일한 결과를 보일 것인가에 대한 추정으로 자료의 질을 나타내는 방법 중의 하나이다. 추정방법에 따라 신뢰도는 단순 적률 상관계수 추정공식에 의한 재검사 신뢰도 및 동형검사 신뢰도가 있으며, 한 검사를 구성하는 부분검사나 문항간의 일관성 정도를 보기 위한 내적 일관성 신뢰도(internal consistency reliability) 등이 있다(성태제, 1995). 이 연구에서는 내적 일관성 신뢰도를 선택하였으며, 모든 검사에 대한 반응의 일관성에 기초한 Cronbach's  $\alpha$ 계수를 구하였다. 이는 검사도구의 신뢰도를 과소 추정하는 단점이 있으나, 재검사 신뢰도와 동형검사 신뢰도를 산출해야 할 경우 동일 대상자에게 검사를 두 번 실시하는 번거로움과, 검사간격과 동형성 정도에 따라 신뢰도 계수가 변화하는 문제점을 피할 수 있고, 신뢰도 추정시 검사를 양분하지 않아도 되는 장점과 문항간의 일관성에 의하여 단일한 신뢰도 추정 결과를 얻는 장점이 있다. 이 연구의 건강수준 8개 영역에 걸쳐 나타난 Cronbach's  $\alpha$ 계수는 신체적 기능의 항목이 0.85로 가장 높았으며, 사회적 기능의 항목이 0.51로 가장 낮았다. 이 결과는 Nerenz 등(1992)의 신뢰도 계수 범위 (0.43-0.90), Brazier 등(1992)의 (0.60-0.81), Jenkinson 등(1993)의 (0.76-0.90), McHorney 등(1994) (0.63-0.94)의 신뢰도 계수 범위와 유사하다. 일반적으로 동일 대상에게 적용하여 개인간 측정값을 비교하기 위해서는 신뢰도 계수가 0.90 이상으로 높아야 한다. 하지만 건강수준을 이용한 일반 인구 집단 비교 등 집단간의 비교에는 신뢰도의 일반적 기준이 그리 높지 않아도

되며, 연구자에 따라 다르지만 대체로 신뢰도 계수가 0.50 혹은 0.70 이상인 경우 수용할 수 있다고 제안하고 있다(Helmstadter, 1964; Nunnally, 1978; Ware, 1993). 이 연구에서는 활력과 사회적 기능 항목을 제외한 다른 항목에서 0.70 이상의 결과를 얻었으며, 0.50을 기준으로 할 경우 모든 항목이 신뢰할 만한 측정도로 인정받을 수 있는 결과이며, 집단용 설문지로 수용할만 하다고 생각된다.

측정도구의 구성과 관련하여 신뢰도 계수에 관해 좀더 구체적으로 고찰해 보면, 신뢰도 계수는 측정도구 구성의 동질성 여부와 문항 수에 많은 영향을 받는다. 초기 연구의 경우 정신건강, 신체적 증상, 기능수준, 일반적 건강 인지도, 흡연 등을 하나의 영역으로 묶어 건강수준을 평가하였는데, 이질적 개념이 혼재됨으로 인해 신뢰도 계수는 낮은 결과를 보였고, 해석상의 어려움을 초래한 바 있다(Macmillan, 1957; Tyler와 Fiske, 1968). 다시 말해 동질성에 대하여 고려를 하지 않으면 문제가 발생하며, 신뢰계수는 낮은 경향을 보인다. 따라서 본 건강수준표는 건강수준을 평가하고, 구성의 동질성을 확보하기 위해 8개 영역으로 나누어 각 영역에 따라 같은 개념끼리 신뢰도를 산출하였다. 또한 신뢰도 계수는 항목내 문항수에 영향을 받는데, 문항수가 많으면 많을수록 신뢰도 계수는 높은 경향을 보인다. 본 연구에서도 가장 낮은 신뢰도 계수를 보이는 사회적 기능 항목은 그 문항 수가 두 문항에 불과하고, 가장 높은 신뢰도 계수를 보이는 신체적 기능 항목은 10문항으로 8개 영역 중 가장 많은 문항으로 구성되었다. 그러므로 신뢰도를 높이기 위해서는 문항수를 늘리면 된다. 그러나 실용적인 측면에서 응답자의 부담을 덜어주고, 널리 이용되기 위해서는 측정도구가 간편해야 한다. 따라서 두 가지 측면을 고려하여 축약형 설문지가 가지는 장점을 최대한 발휘하면서, 집단용 설문지로 수용가능한 신뢰도를 유지해야 할 것이다(Ware, 1993).

건강수준 평가표의 자료의 질을 평가하기 위한 또 다른 방법으로 각 개인의 응답을 분석하였는데, 36개 항목 중 비슷한 개념을 내포하는 문항 15쌍을 조사하

**Table 8.** Comparison of content of Medical Outcome Study and other widely used health surveys\*

concepts	Long-Form	SF-36	HIE	NHP	SIP	DUKE	COOP	MHIQ	QWA
Physical functioning	●	●	●	●	●	●	●	●	●
Social functioning	●	●	●	●	●	●	●	●	●
Role functioning	●	●	●	●	●	●	●	●	●
Psychologic distress	●	●	●	●	●	●	●	●	●
Psychologic well-being	●	●	●	●	●	●	●	●	●
Health perception	●	●	●	●	●	●	●	●	●
Pain	●	●	●	●	●	●	●	●	●
Energy/Fatigue	●	●	●	●	●	●	●	●	●
Report Health transition	●	●	●	●	●	●	●	●	●
Symptom/Problem	●	●	●	●	●	●	●	●	●
Sleep	●	●	●	●	●	●	●	●	●
Cognitive functioning	●	●	●	●	●	●	●	●	●
Sexual functioning	●	●	●	●	●	●	●	●	●
Health distress	●	●	●	●	●	●	●	●	●
Family functioning	●	●	●	●	●	●	●	●	●
Self esteem	●	●	●	●	●	●	●	●	●
Eating	●	●	●	●	●	●	●	●	●
Recreation/Hobbies	●	●	●	●	●	●	●	●	●
Communication	●	●	●	●	●	●	●	●	●
Quality of Life	●	●	●	●	●	●	●	●	●

Long Form: MOS 149-Item Functional Status and Well-Being Survey

SF-36: MOS 36-Item Short Form Health Survey, HIE: Health Insurance Experiment

NHP: Nottingham Health Profile, SIP: Sickness Impact Profile, DUKE: Duke Health Profile

COOP: Dartmouth COOP Function charts, MHIQ: McMaster Health Index Questionnaire,

QWB: Quality of Well-Being Scale

\*Source; cited from SF-36 Health Survey-Table 8.1 (Ware, 1993)

여 비밀관된 응답을 한 빈도를 계산하였다. 예를들어 '4.5리(1Km 정도 걷기)'에 긍정적인 응답을 한 사람이 '마을주변 산책'에 부정적으로 응답하였다면 일관성이 없다고 결론내릴 수 있다. 본 연구에서는 전체 대상자 중 87.5%가 일관된 응답율을 보였고, 이는 미국 일반 인구 집단을 대상으로 한 연구(Ware, 1993)의 90.3% 보다 다소 작은 수치이나, 두 개 이상 일관성이 없는 응답을 한 사람은 3.0% 정도로서 자료 분석에 무리가 없다고 판단된다.

타당도는 연구자가 측정하고자 하는 것을 어느정도 제대로 측정하였는가를 나타내주는 지표가 된다. 타당도의 문제가 제기되는 것은 조사설계와 측정의 단계에서이다. 즉 검사점수가 검사의 사용목적에 얼마나 부합하느냐의 문제이다. 결국 검사도구 목적의 적합성에 해당된다. 타당도는 많은 사람에 의해 혼미스

러울 정도로 다양한 타당도가 제안 되었으나, 일반적으로 내용타당도(content validity), 기준타당도(criterion validity), 구성타당도(construct validity)로 구분할 수 있다(Streiner와 Norman, 1989; Ware, 1993).

내용타당도는 본 연구결과에서 특별히 제시하지 않았으나, 검사의 문항, 질문, 목적이 측정을 위하여 규정된 영역이나 전체를 대표하는 정도를 나타내므로 중요하다고 할 수 있다. 따라서 내용타당성의 근거는 검사개발에서 중심관건이며, 내용선택, 문항유형 선택, 점수화 등의 의사결정을 하는데 중요한 역할을 한다(성태제, 1995). 건강수준 측정도구 역시 측정할 영역을 포괄하기 위하여 오래된 역사적 배경을 가지고 있으며, 이 토대위에 다양한 측정도구의 비교(표 8)와 함께 그 타당성을 주장하고 있다(Ware, 1993).

기준타당도는 검사도구를 통해 얻어진 측정하고자

하는 내용의 점수와 다른 어떤 기준과의 상관관계에 의하여 검사도구의 타당성을 검증한다. 여기에는 동일한 시점(공유타당도; concurrent validity)이나 현재 상태로 부터 일정한 간격을 두고(예측타당도; predictive validity) 수집된 자료 및 기준에 타당성을 보장받고 있는 검사와의 유사성 혹은 연관성에 의하여 타당성을 검증하는 방법을 포괄하고 있다. 예컨대 동일한 시점에서 임상적 기준을 가지고 평가하고자 할 때, 질병의 경중도(severity)라는 기준은 건강수준을 검증하는 적절한 접근 방법이 될 것이다. 이외에도 세부항목에 대한 타당성을 보기 위해 보건의료서비스 이용률을 조사하여 설문지의 일반건강 항목과의 연관성을 보는 것도 좋은 방법 중의 하나일 것이다.

마지막으로 구성타당도는 건강개념이 조작적 정의에 입각해서 측정도구를 구성하고 있는 하부개념들이 조작적 정의에서 규명한 건강수준을 제대로 측정하는가를 보는 것이다. 만약 연구 결과가 조작적으로 규정한 어떤 건강수준의 특성을 제대로 측정하지 못하거나 다른 특성을 측정한다면, 구성타당도가 결여되어 있는 것이다. 구성타당도를 검증하는 방법은 상관계수법, 실험설계법, 인자분석 등이 있다(성태제, 1995). 상관계수법은 건강수준을 평가하기 위하여 이미 개념화한 8개의 항목과 각 문항에 대해 상관계수를 구하여 검증할 수 있는데, 우리는 척도화 가정을 통해 검토를 하였으며, Pernegger 등(1995)은 이러한 검증 방법을 내적타당성(internal validity)이라는 표현으로 쓰기도 하였다. 구성타당도를 검증하기 위하여 가장 많이 쓰이는 검증 방법은 인자분석이다. 인자분석은 복잡하고 정의되지 않은 많은 변수들이 상호관계를 맺고 있어서 직접적으로 해석하기 어려운 여러 변수들 간의 구조적 연관관계를, 상관이 높은 변수들을 모아 요인으로 규명하고, 상대적으로 독립적이면서, 변수들의 저변구조(underlying structure)를 이해하기 위해 개념상의 의미를 부여할 수 있는 인자들을 상정하여 그 의미를 부여하는 통계적 방법이다(김기영과 전명식, 1990). 따라서 본 연구에서는 건강수준을 측정하고자 하는 36개의 문항과 공통인자(common factor)간에 내재하

는 인과관계를 탐구하고, 그 결과를 건강수준 평가표에서 설정하는 8개의 항목과 비교해 보고자 탐색적 인자분석을 시행하였다. 탐색적 인자분석을 주축인자법과 직교회전법을 이용하여 시행한 결과, 고유값 1.0 이상인 요인이 8개 였으며, 총 분산의 62.4%를 설명할 수 있었다. 요인1에는 신체적 제한 항목들이 포함(문항1 제외)되었고, 요인2는 정신건강 문항1,2,4, 활력 문항3,4, 사회적기능 문항2가 분류되었으며, 요인3에는 역할제한 항목(신체적 역할제한 문항3 제외)이 선정되었으며, 요인4에는 일반건강(문항4 제외)이, 요인5에는 정신건강 문항3,4, 활력 문항 1,2가 포함되었다. 요인6에는 통증 및 사회적기능 문항1이 분류되었으며, 요인7에는 일반건강 문항4, 신체적 역할제한 문항3이 포함되었고, 요인8에는 신체적 기능 문항1이 선정되었다. 이는 SF-36의 건강수준 8개 영역이 다소 혼재되어 있는 소견이며, 이후 연구 수행시 설문지 구성에 세심한 고려와 재조정이 필요함을 시사하는 결과라 하겠다. Garratt 등(1993)이 시행한 인자분석도 SF-36과 차이가 있으며, Garratt 등(1993)의 결과와 이 연구와 비교해 볼 때, 그가 보고한 요인1의 신체적 기능항목과 요인 2의 정신건강과 활력은 이 연구와 거의 일치된 결과를 보이며, 요인 3의 신체적 역할 제한, 사회적 기능, 통증이 한 공통인자로 묶여 있는 결과와는 다소 차이를 보인다. 이를 종합하면, 인자 분석을 통한 타당도의 검증은 대상집단이나 표본수가 다른 조사마다 인자구조가 달라질 수 있다는 것을 시사하며, 앞에서 제기한 다각적 접근을 통해 설문지로서의 근거는 마련되었다고 생각될 수 있으나 좀더 설문지 구성에 대한 연구가 지속되어야 할 필요성은 있다. 또한 Gronlund와 Linn(1990)이 타당도를 이해하는데 주의해야 할 점으로 제시한 “타당도는 연구대상자 집단에 사용된 측정도구나 검사에 의해 얻어진 검사결과의 해석에 대한 적합성이지 검사자체와 관련된 것이 아니며, 검사의 타당도라 표현하지만 정확히는 검사결과로부터 만들어진 해석에 대한 타당성”이라는 지적을 음미해 볼 필요가 있다.

많은 연구결과와 보고를 보면 건강수준을 해석하고

자 할 때 두가지 접근 방식으로 평가하고 있다. 즉, 기준(norms)에 입각한 접근방법과 준거(criteria)에 입각한 접근방법이다. 기준에 입각한 접근은 일반인구 집단의 점수 분포에 따라 개인의 점수와 집단의 평균을 해석하는 것이다. 이 방법은 단일한 기준이 없을 때 특히 유용하다. 연령이 다른 남성과 여성은 건강 수준에 대해 달리 보고하기 때문에 이 그룹과 서로 다른 직업군에는 각각 다른 기준이 사용될 수 있다(Jenkinson 등, 1993; Garratt 등, 1993; Ware, 1993; Perneger 등, 1995). 반면 준거에 입각한 접근은 의학적 조건, 질병의 경중도, 보건의료서비스의 이용 등에 관한 기준 변수(criterion variables)를 포함시켜 그러한 연구를 해석하는 데 도움을 준다(Stewart 등, 1989; Mchorney 등, 1993; Ware, 1993). 이 연구에서는 기준에 입각한 접근 방법으로 근로자들의 건강수준을 평가하였다. 연구대상자 전체의 건강수준은 표 7에 제시된 바와 같으며, 다른 나라의 보고와 비교해 보면 영국(Jenkinson 등, 1993; Garratt 등, 1993), 미국(Ware, 1993), 프랑스(Perneger 등, 1995)의 일반 인구 대상자의 건강수준 보고보다 다소 낮은 편이나 전반적으로 보아 이와 유사한 경향을 보인다.

끝으로 이 연구를 수행하면서 몇 가지 고려할 사항이 있다. 첫째, 건강수준의 정의가 합리적인가 하는 문제제기이다. 이 연구에서는 내용타당도에서 다양한 측정도구의 비교와 함께 그 타당성을 거론한 바 있다. 그러나 Dickinson 등(1993)이 지적하듯이 건강에 대한 정의와 그에 필요한 유형학(typology)과 관련하여 여기에는 여러 가지 곤란함과 불일치가 있을 수 있다는 점을 전제하여야 한다. 특히 평가의 근본 원칙이 적절한 도구를 선택하는 것이라면, 건강수준의 정의가 합리적으로 구성되어야 한다. 그런데 하나의 도구가 개선된 산전관리, 자발적 건강증진, 운동 프로그램, 고혈압의 약물치료 등을 평가하는데 적합하다고 가정할 수 있는가 하는 점이다. 우리는 Hunt와 McKenna(1993)의 SF-36은 '건강수준'을 측정하는 것이 아닌 '건강인지'의 측정수단이라고 비판한 점을 고려할 필요가 있다. 둘째, 결과라는 용어의 모호함이다. 결과의

측정은 진료의 질을 평가하는 데 있어 중요한 고려사항이다. 하지만 명확하지 않다면 이 주제를 이해하는데 걸림돌이 될 것이다. 가장 모호한 것은 무엇이 결과를 구성하는가에 대한 것이다. Ware(1993)는 이 용어를 보건의료의 효과를 언급하는데 사용하였다. 그러나 결과에 영향을 미치는 더 중요한 요인은 질병의 자연 경과이다(Christie와 Watt, 1993). Ware는 암묵적으로 치료결과와 환자의 의학적 상태 사이의 관련성을 인정한다. 진료의 질에 대한 가장 적절한 측정법은 어떤 특정 시점에서 질병의 자연 경과와 치료 후 질병의 경과 간의 차, 즉 순치료효과(net treatment effect)이다. 따라서 질병의 자연 경과가 제대로 규명되어 있지 않다면 결과 중에서 보건의료가 기여하는 부분 역시 알 수 없으므로 대상자 선정시 유의해야 한다는 점이다. 셋째, 이 연구에서는 신뢰도를 평가함에 있어 내적일치도를 구하였는데, 만약 검사-재검사 신뢰도를 구할 경우 두 신뢰도 계수간에 차이가 있을 수 있다는 점이다. 자기기입식 검사를 이용할 때, 흔히 내적일치도를 이용하지만, 더 확실한 것은 여러 단점과 번거로움에도 불구하고 단기간 내에 동일한 집단을 대상으로 동일한 검사를 최소한 두 번 반복하여 검사-재검사 신뢰도를 보는 것이다. 예를 들어 Garratt 등(1993)은 SF-36을 결과측정의 평가도구로 사용하였는데, 이때 개입에 따른 변화의 맥락에서 치료 전후의 점수를 측정하는 것이라 한다면, SF-36은 시간 변수를 고려한 신뢰도를 중요하게 고려하여야 한다.

그런데 건강상의 특별한 변화가 없는 데도 점수가 시간의 경과에 따라 큰 변이를 보인다면 치료로 인한 변화를 구명하는 것은 더욱 어렵고 측정된 치료효과의 정확도는 감소되며 이에 따라 내적 일치도가 높다 하더라도 그 근거가 미약할 수 있다(Sheldon, 1993). 따라서 보고되는 신뢰도 평가방법들이 대부분 내적일치도의 측정에 근거해 있다해도, Streiner와 Norman(1989)이 지적한 것처럼 내적일치도는 시간에 따른 변이의 주요한 발생원인을 간과할 소지가 있기 때문에 신뢰도의 지표로 사용할 때는 해석에 유의할 필요가 있다. 또한 도구의 신뢰도는 임상적이고 조직적 맥락

의 범위 안에서 적절하게 실험되어야 한다. 이러한 결과가 이용가능해야 비로소 결과 측정 수단으로서 도구의 통계학적 속성이 밝혀지는 것이다. 그전까지 신뢰도는 여전히 불확실하다고 할 수 있다. 그러나 위와 같은 점들은 지속적인 연구를 통해 해결할 수 있으며, 많은 논쟁점과 비판은 이 분야에 새로운 토양이 되리라 생각된다.

## V. 참고문헌

- 김기영, 전명식. SAS 인자분석. 서울, 자유아카데미, 1990
- 성태제. 타당도와 신뢰도. 서울, 양서원, 1995
- Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An Inventory for measuring depression. Arch Gen Psychiatry 1961; 4: 561
- Bergner M, Bobbitt Ra, Carter WB. The sickness impact profile: conceptual formulation and methodology for the development of a health status measure. Int J Health Serv 1976; 6: 393
- Brazier JE, Harper R, Jones NM, O'Cathain A, Tomas KJ, Usherwood T, Westlake L. Validation the SF-36 health survey questionnaire: new outcome measure for primary care. BMJ 1992; 305: 160-164
- Christie J, Watt A. The term outcome is ambiguous. BMJ 1993; 307: 126
- Dickinson E, Gompertz p, Harwood R, Ebrahim S. Definitions of health status should be rationalised. BMJ 1993; 307: 126
- Fanshel S, Bush JW. A health-status index and its application to health-services outcomes. Operations Research 1970; 18: 1021
- Garratt AM, Ruta DA, Abdalla MI, Buckingham JK, Russell IT. The SF 36 health survey questionnaire: an outcome measure suitable for routine use within the NHS? BMJ 1993; 306: 1440-1444
- Geigle R, Jones SB. Outcomes measurement: A report from the front. Inquiry 1990; 27: 7-13
- Goldberg DP, Hillier VF. A scaled version of the general health questionnaire. Psychol Med 1979; 9: 139
- Greenfield S, Nelson EC. Recent development and future issues in the use of health status assessment measures in clinical settings. Med Care 1992; 30 (5): MS23-41
- Gronlund NE, Linn RL. Measurement and evaluation in teaching(6th ed.) New York: Macmillan, 1990
- Hamilton M. Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol 1967; 6: 278
- Helmstadter GC. Principles of psychological measurement. New York, NY: Appleton-Century-Crofts, Inc. 1964
- Howard KI, Forehand GG. A method for correcting item-total correlations for the effect of relevant item inclusion. Educational and Psychological Measurement 1962; 22: 731-735
- Hunt SM, McKenna SP. Measuring patients' views of their health. SF 36 misses the mark. BMJ 1993; 307: 125
- Hunt SM, McKenna S, McEwen J, Williams J, Papp E. The Nottingham Health Profile: Subjective health status and medical consultations. Social Science in Medicine 1981; 15A: 221-229
- Jenkinson C, Coulter A, Wright L. Short form 36 (SF 36) health survey questionnaire: normative data for adults of working age. BMJ 1993; 306: 1437-1440
- Macmillan AM. The health opinion survey: Technique for estimating prevalence of psychoneurotic and related types of disorder in communities. Psychological Reports 1957; 3: 377-387
- McHorney C, Ware JE, Lu R, Sherbourne CD. The MOS 36-Item short form health survey (SF 36): III. Tests of data quality, scaling assumption, and reliability across diverse patient groups. Med Care 1994; 32(1): 40-66
- McHorney CA, Ware JE, Raczek AE. The MOS 36-item short form health survey (SF 36). II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Med Care 1993; 31: 247-263
- Moskowitz E, McCann CB. Classification of disability in the chronically ill and aging. J Chronic Dis 1957; 5: 342
- Nelson EC, Langdgraf JM, Hays RD, Kirk JW, Wasson JH, Keller A, Zubkoff M. The COOP func-

- tion chart: A system to measure patient function in physician' offices. In M Lipkin(ED. ), Functional status measurement in primary care. New-York: Springer-Verlag 1990
- Nerenz DR, Repasky DP, Whitehouse FW, Kahkonen DM. Ongoing assement of health status in patients with diabetes mellitus. *Med Care* 1992; 30 (Suppl): MS112-124
- Nunnally JC. Psychometric theory, 2nd edition. New York, NY: McGraw-Hill. 1978
- Perneger TV, Leplege A, Etter JF, Rougemont A. Validation of a French-Language version of the MOS 36-Item short form health survey (SF-36) in young healthy adults. *J cli Epidemiol* 1995; 48(8): 1051-1060
- Sheldon MP. A physical achievement record for use with crippled chidren. *Journal of Health and Physical Education* 1935; 6: 30
- Sheldon T. Reliability of SF 36 remains uncertain. *BMJ* 1993; 307: 125
- Steinbrocker O, Traeger CH, Baterman RC. Therapeutic criteria in rheumatoid arthritis. *JAMA* 1949; 140: 659
- Stewart AL, Greefield S, Hays RD, Rogers WH, Berry SD, McGlynn EA. Functional status and well-being of patients with chronic conditions: results from the medical outcomes study. *JAMA* 1989; 262: 914
- Stewart AL, Hays RD, Ware JE. The MOS short form general health survey. Relialbilty and validity in a patient population. *med care* 1988; 26(7): 724
- Streiner DL, Norman GR. Health measurement scale; A practical guide to their development and use. Oxford university press, 1989
- Tyler TA, Fiske DW. Homogeneity indices and text length. *Educational and Psychological Measurement* 1968; 28: 767
- Ware JE. Measuring patients' views: the Optimum outcome measure. *BMJ* 1993; 306: 1429
- Ware JE. SF-36 Health Survey. Manual & Interpretation Guide. Boston, MA. : The Health institute; 1993
- Ware JE, Brook RH, Davies AR, Lohr KN. Choosing measures of health status for individuals in general populations. *Am J Public Health* 1981; 71: 620
- Ware JE, Brool RH, Davies-Avery A, Williams KN, Stewart AL, Rogers WH, Donald CA, Johnston SA. Conceptualization and measurement of health for adults in the Health Insurance Study. Volum I: Model of Health and methodology. Santa Monica, CA: The RAND Corporation; 1980
- Ware JE, Johnston SA, Davies-Avery A, Brook RH. Conceptualization and measurement of health for adults in the Health Insurance Study. Volum III: Mental Health. Santa Monica, CA: The RAND Corporation; 1979
- Wasson J, Keller A, Rubenstein L, Hays R, Nelson E, Johnson D, The Dartmouth Primary Care COOP Project. Benifits and obstacles of health status assessment in ambulatory settings. *Med Care* 1992; 30(5): MS42
- Ware JE, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36). I. Conceptural framwork and item selection. *Med care* 1992; 30: 473
- Zung WWK. A self-rating depression scale. *Arch Gen Psychiatry* 1965; 12: 63

## 건강수준평가표의 설문 내용

1. (일반적 건강 문항1) 평소 당신의 건강 상태는 어떻다고 생각하십니까?
2. (건강 상태 변화) 작년과 비교하여, 현재의 당신의 건강은 어떤 상태라고 생각하십니까?
3. (신체적 기능) 다음은 평상시 당신의 활동에 관한 항목들입니다.  
이러한 활동을 할 때 지장이 있으십니까? 있다면 어느정도 입니까?  
문항1) 달리기, 무거운 물건 들어올리기, 격렬한 운동을 할때  
문항2) 탁자 옮기기, 집안 청소, 배드민턴 치기.  
문항3) 시장 보기.  
문항4) 한꺼번에 두세 계단씩 오르기.  
문항5) 한 계단씩 오르기.  
문항6) 구부리기, 무릎꿇기.  
문항7) 4~5리(1km정도) 걷기.  
문항8) 마을 주변 산책, 동네 한바퀴 걷기.  
문항9) 집 주변 걷기.  
문항10) 혼자서 목욕이나 옷 입기.
4. (신체적 역할제한) 지난 한달간 당신의 신체적인 어려움으로 일상생활이나 직장에서 다음과 같은 어려움이 있었습니까?  
문항1) 일하는 시간을 줄여야만 했다.  
문항2) 원하는 만큼 일을 하지 못했다.  
문항3) 어떤 일을 할 때 힘이 들 때가 있었다.  
문항4) 대부분의 일을 할 때 어려움이 있었다.
5. (감정적 역할제한) 지난 한달간 감정적인 어려움으로 일상생활이나 직장에서 다음과 같은 어려움이 있었습니까?  
문항1) 일하는 시간을 줄여야만 했다.  
문항2) 원하는 만큼 일을 하지 못했다.  
문항3) 보통 때만큼 일에 집중할 수 없었다.
6. (사회적 기능 문항1) 지난 한달간 신체상의 어려움이 감정상의 어려움으로 사회생활(가족간, 친구간, 이웃간) 에 어려움이 있었다면 어느정도 입니까?
7. (통증 문항1) 지난 한달간 당신의 신체적인 아픔은 어느정도 었습니까?
8. (통증 문항2) 신체적 아픔으로 인해 정상적인 생활에 지장이 있었습니까?
9. 다음의 질문들은 지난 한달간 여러분이 어떻게 느꼈고, 어떤 일이 있었는지에 관한 것입니다.  
활력항목 문항1) 나는 의욕이 넘쳤다.  
정신건강 문항1) 나는 신경질적이었다.  
정신건강 문항2) 나는 의기소침했었다.  
정신건강 문항3) 나는 안정되고 평온했다.  
활력항목 문항2) 나는 힘이 넘쳤다.  
정신건강 문항4) 나는 우울하고 상심했었다.  
활력항목 문항3) 나는 지쳤었다.  
정신건강 문항5) 나는 행복했었다.  
활력항목 문항4) 나는 피곤했었다.
10. (사회적 기능 문항2) 지난 한달간 몸이 불편하다든지 아니면 기분이 내키지 않아서 친척이나 친구 등을 방문하는 것과 같은 일들을 하는데 어려움이 있었다.
11. (일반건강) 다음의 질문에 대해 답해 주십시오.  
문항2) 나는 다른 사람들보다 더 쉽게 병에 걸리는 것 같다.  
문항3) 나는 누구보다도 건강하다.  
문항4) 나의 건강은 점점 나빠질 것이다.  
문항5) 나의 건강상태는 매우 좋다.