예방의학회지 제32권 제3호(1999년 9월)
Korean J Prev Med 1999;32(3):361-373

Receiver Operating Characteristic 분석법을 이용한 업무관련성 근골격계질환 설문지 개발

김재용, 권호장[1], 주영수, 조수헌, 강대희, 성주헌,
최성우, 최재욱[2], 김재영[2], 김돈규[3]

서울대학교 의과대학 예방의학교실 및 의학연구원 환경의학연구소, 단국대학교 의과대학 예방의학교실[1],
고려대학교 의과대학 예방의학교실[2], 서귀포시 보건소[3]

Development of Work-related Musculoskeletal Disorder Questionnaire Using Receiver Operating Characteristic Analysis

Jaiyong Kim, Ho-Jang Kwon[1], Yeongsu Ju, Soo-Hun Cho, Daehee Kang,
Joohon Sung, Seong-Woo Choi, Jae-Wook Choi[2], Jae-Young Kim[2], Don-Gyu Kim[3]

Department of Preventive Medicine, Seoul National University College of Medicine & Institite of Environmental Medicine, SNUMRC
Department of Preventive Medicine, Dankuk University College of Medicine[1]
Department of Preventive Medicine, Korea University College of Medicine[2]
Seo-Gui-Po City Health Center[3]

Objectives : Receive Operating Characteristic(ROC) curve with the area under the ROC curve(AUC) is one of the most popular indicator to evaluate the criterion validity of the measurement tool. This study was conducted to develop a standardized questionnaire to discriminate workers at high-risk of work-related musculoskeletal disorders using ROC analysis.

Methods : The diagnostic results determined by rehabilitation medicine specialists in 370 persons(89 shipyard CAD workers, 113 telephone directory assistant operators, 79 women with occupation, and 89 housewives) were compared with participant' s own replies to 'the questionnair on the worker' s subjective physical symptoms' (Kwon, 1996). The AUC' s from four models with different methods in item selection and weighting were compared with each other. These 4 models were applied to 225 persons, working in an assembly line of motor vehicle, for the purpose of AUC reliability test.

Results : In a weighted model with 11 items, the AUC was 0.8155 in the primary study population, and 0.8026 in the secondary study population(p=0.3780). It was superior in the aspects of discriminability, reliability and convenience. A new questionnaire of musculoskeletal disorder could be constructed by this model.

Conclusion : A more valid questionnaire with a small number of items and the quantitative weight scores useful for the relative comparisons are the main results of this study. While the absolute reference value applicable to the wide range of populations was not estimated, the basic intent of this study, developing a surveillance tool through quantitative validation of the measures, would serve for the systematic disease prevention activities.

Korean J Prev Med 1999;32(3):361-373

Key words : Receiver operating characteristic, Questionnaire, Surveillance, Musculoskeletal disorder.

서 론

질병의 규모와 추이는 병인파악을 위한 역학적 추론에서 보건의료 문제의 우선순위 선정에 이르는 다방면에 요구되는 기초적인 정보다. 하지만, 질병통계에 관한 한 우리 나라는 외국에 비해 크게 뒤떨어져 있는 실정이다. 일부 악성종양과 만성질환에 대한 연구에서 유병률이나 발생률이 제시된 바 있으나 국가차원의 질병통계 자료로 간주하기에는 대표성과 타당성 측면에서 한계를 지니고 있고, 사망원인 자료나 의료보험자료와 같은 이차 자료는 자료원의 성격상 시의적절한 질병관리 및 예방활동에 활용하기 어렵다는 제한점을 지니고 있다. 소모적 사후대응과 합리적인 평가작업의 부재는 기초자료 부실의 결과라 할 수 있을 것이며, 이러한 악순환을 극복하기 위한 방안으로 체계적인 질병감시체계 구축에 대한 논의가 개진된 바도 있다(조수헌 등, 1997). 하지만, 법정전염병 신고체계에서 보이듯이 원론적인 개념을 현실에 적용하는 단계에서는 유명

접수 : 1999년 3월 31일, 채택 : 1999년 5월 24일.
본 연구는 보건복지부 1997년도 보건의료기술 연구개발사업 지원(과제번호:HMP-97-M-4-0048)에 의한 연구결과의 일부임.
교신저자 : 김재용 서울대학교 의과대학 예방의학교실 (전화 : 02-740-8332, 팩스 : 02-747-4830)

무실한 자발적 신고체계, 선택가능한 유인대책의 부족, 적절한 권한과 기능의 결핍 등 적지 않은 어려움에 처하게 되며, 개별 연구자가 한시적으로 수행한 각종 연구 및 조사 결과들을 능동적 질병감시에 활용하기에는 대표성, 지속성, 비용 등의 측면에서 한계가 존재한다.

Malison(1992)은 "너무 많은 질병에 대해 너무 많은 정보를 얻으려고 하는" 저개발국 질병관리체계의 난맥상을 지적하면서 주요 질환에 대해 반드시 필요한 정보를 우선 확보하는 접근방식과 실용적인 감시기준의 확립을 출발점으로 제시한 바 있다. 이러한 측면에서 간편 · 저렴하면서도 타당도가 검증된 표준화된 선별도구를 확보하는 것은 한계상황을 극복하고 기존 논의를 현실화시키는 계기가 될 수 있을 것이다.

측정도구의 타당도 검증과 관련하여 Philosophov 등(1997)은 환자의 실제 건강수준과 의학적 측정도구를 이용한 측정결과를 각기 요인 및 요인추정치(factor estimate)로 구별하여 표현한 바 있다. 이는 모든 측정값이 오차가 내재된 일종의 추정치라는 전제하에 역설적으로 타당도 검증의 중요성을 강조하기 위한 것이었다. 이를 위한 타당도 검증기법들 중에서 receiver operating characteristic(이하 ROC) 분석은 가장 대표적인 방법이다. 용어상의 개념은 Van Bemmel 등(1997)이 제시한 바 와 같이, 발신기(sender)가 보내는 신호(*s*)와 전달과정에서 개입되는 잡음(*n*)의 혼합물(*m*)이 수신기(receiver)에 도착할 때, 수신기가 신호를 선별적으로 포착하는 성능을 비교 · 분석하기 위한 기법이다(그림 1).

ROC곡선은 환자군 및 대조군에서 관찰된 모든 측정값 각각을 구분점으로 삼았을 때 계산된 민감도와 특이도를 이용하여 y축은 '민감도(sensitivity, true-positive fraction, TPF)', x축은 '1 - 특이도(1 - specficity, false-positive fraction, FPF)'로 이루어진 좌표 위에 점을 찍고, 그 점들을 연결한 하나의 곡선으로 표현된다. 또한, ROC곡선의 곡선하면적(area under the ROC curve, 이하 AUC)은 환자군과 정상인군에서 무작위로 한 명씩 추출하여 측정값을 비교하였을 때 환자가 정상인보다 더 비정상적인 측정값을 나타낼 확률(probability of correctly ranking a normal-abnormal pair)을 의미하는 기준타당도(criterion validity) 지표다. 공학분야에서 개발된 ROC 분석법이 의학분야에 본격적으로 활용된 것은 1960년대 후반으로, 정신과 및 정신생리학 분야의 실험적 연구와 방사선과에서 이용되기 시작했고, 의과학의 발전과 함께 다양한 임상 검사법이 개발되면서 동일 질환에 대한 서로 다른 검사법들의 성능을 정량적으로 비교 · 평가하는 표준적인 방법론으로 활용되고 있다(Morasso 등, 1996). 특히, ROC 분석법은 기준선택에 관련된 다양한 문제점들(유병률, 오분류의 영향, 정상 및 비정상군에서의 측정치 분포 등)을 동시에 고려할 수 있고, 시각적으로 결과를 제시함으로서 해석이 용이하며, 가능한 모든 측정범위를 포괄하여 측정도구의 기준타당도를 제시하는 장점을 지니고 있다(Zweig 등, 1993). 최근에는 설문지의 개발 및 표준화 작업에도 이 기법이 활용되어 각 나라 고유의 버전(version)을 만들거나, 보다 변별력이 우수한 설문도구를 개발하는데 이용되고 있다(Fombonne, 1991; Garyfallos 등, 1991; Harper 등, 1995).

설문지는 역학조사에서 가장 흔히 이용되는 측정도구로서 비교적 다수의 인구를 대상으로 간편하게 적용할 수 있으며 때로는 기존의 임상검사로는 확인할 수 없는 중요한 정보를 제공한다. 외국의 경우, 알콜 중독이나 우울증 등 일부 정신질환과 천식, 협심증, 요통 등에 대해 타당도 검증을 거친 질병특이적 설문도구가 개발되어 있으며, 임상치료의 효과나 삶의 질 등과 같이 참값을 규정하기 힘든 추상적 사안에 대한 설문도구는 일반적인 임상검사법보다 더욱 엄격한 예비검사와 타당도 검증을 요구받는다(Hartzema 등, 1998). 이에 비해, 기존 국내연구는 외국의 설문지를 번역하여 통계적 관련성을 기술하는 정도일뿐, 질병감시를 겨냥한 본격적인 표준설문도구의 개발시도는 상대적으로 부족한 것으로 판단된다.

이에, 본 연구는 상지(上肢)의 업무관련성 근골격계 질환(work-related musculoskeletal disorder, WRMSD)에 대한 설문지를 예로 삼아, 고위험군을 조기에 발견하여 시의적절한 중재조치를 취하는데 이용할만한 질병감시용 도구(surveillance tool)를 개발하는 것을 목적으로 하였다. 업무관련성 상지 근골격계 질환은 최근 우리 나라에서 발생이 증가하는 것으로 알려진 소위 신종 직업병의 하나로서 그 발병요인으로는 업무의 특성, 작업조건 및 환경, 심리적 요인, 인구사회학적 요인(연령, 성, 작업경력 등) 등이 알려져 있다. 이에 관한 국내 연구는 주로 전화교환원(박정일 등, 1989; 최재욱 등 1996)과 VDT 작업자(신병순 등, 1996; 임상혁 등 1997, 김수근 등, 1998)에 대한 것이며 최근에는 조립작업자, 포장작업자, 음악연주자 등 연구대상 직업군이 확대되고 있다. 개별 연구들은 일반적으로 설문지와 임상적 진단을 병행하여 증상호소율이나 유병률 등을 기술하고 일부 측정변수들과의 관련성을 제시하고 있지만, 서로 다른 기준에 따라 서로 다른 소집단 사이의 측정결과들을 상대적으로 비교하는데 치중되어 있어서 일반화에 어려움이 따를 것으로 예상되고 실제로 유사 또는 동일 업종 근로자에 대해서도 연구마다 측정결과가 크게 다르게 나타나고 있다. 또한, 근골격

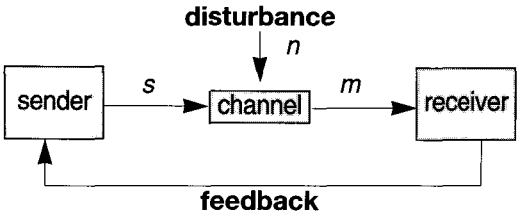

Figure 1. Sender, receiver, and transmission channel. (Reprinted from Van Bemmel et al.1997)

계 질환은 불확실한 병인, 다양하고 주관적인 증상호소, 불완전한 진단기준 등으로 인한 문제점을 안고 있으며, 의사의 문진이나 체계적인 이학적 검사를 통하여 근골격계 질환을 발견하는 일은 상당한 시간과 숙련을 요하는 일로서 현행의 근로자 건강진단 체계에서는 효과적인 선별이 어렵다. 하지만, 이 질환은 완전한 기능상실에 앞서 비교적 긴 기간 동안 통증 등 주관적인 증상을 호소하며, 일부 환자는 대증적인 자가치료의 경험을 거치게 되므로 적절한 시기에 특정 위험군 또는 개인에 대한 선별이 가능하다면 효과적인 예방 조치를 취할 여지가 많은 질환에 해당된다.

본 연구의 구체적인 목적은 문항수가 적으면서도 변별력이 우수한 자기기입식 설문지를 개발하는 것이다. 이를 위해 전문의의 진단기준 등과 같은 의학적 개념보다는 환자가 겪는 기능적 장애나 경험 등에 중점을 맞추고자 하였으며, 민감도를 중시해야 할 선별도구로서의 요구조건을 고려하되 적정 수준 이상의 특이도를 유지함으로써 역학적 또는 임상적 도구로서의 활용가능성도 타진하고자 하였다. 아울러, 이러한 연구성과를 기반으로 향후 질병특이적 설문도구, 질병감시기준 등을 개발하는데 기여하고자 한다.

대상 및 방법

1. 연구대상

1) 일차 연구대상군

1997년에 국내 모 조선업체에서 선박설계 작업자 89명에 대해 권호장 등(1996)이 개발하였던 '근로자의 신체 증상에 관한 설문지'를 일부 수정하여 설문조사와 검진을 실시하였다. 또한, 동일한 설문지를 활용하고 임상적 진단 결과가 뒷받침된 기존의 연구 자료들을 수집하였다.

1995년에 국내 모 통신회사의 전화번호 안내원 113명을 대상으로 수행한 연구(권호장 등, 1996)와 1997년에 선박설계 작업자를 대상으로 수행한 연구는 동일한 재활의학 전문의가 진단을 내리고 동일한 설문지를 사용한 것이다. 또한, 이들 연구를 보완하고, 직업군간 근골격계 질환의 유병률을 비교하기 위해 1998년에 직업여성 79명과 가정주부 89명을 대상으로 수행한 연구(김돈규 등, 1998)는 재활의학 전문의가 진단을 내리고 역시, 동일한 설문지를 사용하였다. 결과적으로, 최종분석에 포함된 일차 연구대상군은 총 370명(남자 89명, 여자 281명)이다. 이들 연구대상의 특성을 집단별로 기술하면 다음과 같다.

(1) 1997년에 국내 모 조선업체의 선박설계 부서에서 컴퓨터를 이용한 설계(computer aided design, CAD) 작업을 수행하는 근로자 전수(787명)를 대상으로 간이 점검표와 근로자 건강진단을 통해 근골격계 질환 유증상자 380명을 선별하였다. 이 중에서 통증이 1주 이상 지속되고 통증빈도가 1개월에 1회 이상이면서 진찰과정에서 이상소견이 의심되었던 근로자 65명에 대해 1개월 후에 위의 '근로자의 신체 증상에 관한 설문지'와 더불어 재활의학 전문의가 정밀검진(근전도 포함)을 실시하여 근막통증후군(myofascial pain syndrome, MPS)으로 확진된 환자군 30명을 선별하였다. 또한, 이상증상은 호소하고 있으나 근골격계 장애를 확진받지 못한 근로자중 설문조사에 성실히 응답한 32명과 근골격계 이상증상이 전혀 없는 근로자들 중에서 무작위로 선정된 27명 등 총 89명을 확보하였다. 사전에 성별에 제한을 두지는 않았으나, 사업장과 업무의 특성상 이들은 모두 남자였다

(2) 1995년에 충북지역의 3개 전화국에서 전화번호 안내원으로 1년 이상 근무한 125명의 여성 근로자를 대상으로 '경견완장애 특별검진'을 실시하였으며, 검진에 참여한 113명은 동일한 설문지를 기입하였고 재활의학 전문의가 진단을 내렸다. 평균 연령은 38세이며, 고졸자 85명, 대졸자 8명, 기타 18명이었다. 평균 근무기간은 17.4년이다.

(3) 직업여성군 79명은 동일 설문지를 이용하여 설문조사를 실시하였으며, 재활의학과 전문의가 동시에 임상적 진단을 실시하였다. 평균연령은 35세이며, 고졸자 27명, 대졸자 51명 및 기타 1명이었다. 구체적인 직업별로는 간호사 33명, 생활설계사 15명, 공무원 15명, 일반사무원 15명, 유치원 보모 1명이었다. 또한, 가정주부 89명은 동일 설문지를 이용하여 국내 모 대학에서 개최된 학습프로그램에 참여한 주부들을 대상으로 설문조사를 실시한 후, 재활의학 전문의가 1주 후에 임상적 진단을 실시하였다.

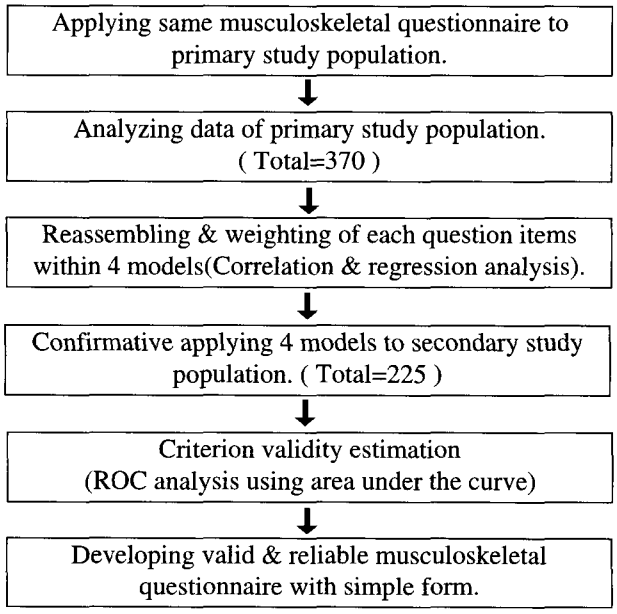

Figure 2. Schematic flow chart of the questionnair reassembling process.

2) 이차 연구대상군

1998년에 국내 모 자동차 조립공정 근로자 250명을 대상으로 실시된 근로자 특수건강진단에서 동일한 설문문항을 이용하여 근골격계 질환에 대한 선별검사를 수행하였다. 이들 연구대상에 대한 의학적 진단은 산업의학 전문의 1인이 내린 것이다. 본 연구에서 진단의사들은 사전에 Hales 등(1994)이 제시한 업무관련성 근골격계 질환의 정의를 숙지하고 있었으며, 설문지에 제시된 증상기준도 이와 동일한 것이었다. 또한, 이차 연구대상군에 대한 조사과정에서 요추부의 질환들은 별도로 구분되어 진단되었으며, 일차대상군과 동일한 기준에 따른 상지 근골격계 질환에 국한하여 '누적외상성 장애' 또는 '근골격계 장애'라는 진단명이 혼용되었음을 추후 확인하였다. 요통 요주의 및 추적관찰 필요, 추간판탈출증 등으로 진단된 경우를 제외한 이차 연구대상자수는 225명이었다.

2. 연구방법

1) 설문 조사

본 연구에서 사용한 '근로자의 신체 증상에 관한 설문지'는 미국 산업안전보건연구원(NIOSH, 1993)이 개발한 표준화된 설문지 문항들과 Karasek(1979)이 제안한 '업무긴장 모형(Job-strain model)'에서 사용된 문항들을 번역하여 사용한 것이다. 주요 인구학적 변수(연령, 근무기간, 교육정도, 가사부담 등)와 상지부위 근골격계 증상(증상 부위, 증상 지속기간, 빈도, 외상여부, 증상 정도), 정신사회적 스트레스에 대한 인지도(업무 재량도 6개 문항, 업무 요구도 5개 문항, 사회적 지지도 5개 문항) 등을 묻는 항목들이 포함되어 있다(표 1).

Table 1. *Item groups included in the 'Questionnaire on the worker's subjective physical symptoms'* (Kwon, 1996)

Content of question items groups	No. of items
Personal information(name, gender, address, education, marriage etc.)	10
Neck symptoms	13
Shoulder symptoms	13
Elbow symptoms	13
Hand and wrist symptoms	13
Ocular Symptoms	7
Past medical history and life style	13
Household affairs and hobbies	4
Conditions of past and current job	7
Self-consciousness on current job stress	16
Total	109

2) 대상질환 및 진단

본 연구는 Hales 등(1994)이 제시한 작업관련성 근골격계 질환의 정의를 기준으로 하였는데, 이는 반복되는 외부의 작업과 관련된 스트레스에 의해 점진적으로 신체의 일부가 손상을 받는 근골격계 질환을 의미하는 것으로서 적어도 1주일 이상, 또는 과거 1년간 적어도 한 달에 한 번 이상 지속되는, 상지관절 부위의 하나 이상의 증상들(통증, 뻣뻣함, 화끈거림, 찌릿찌릿함, 무감각 등)이 존재하는 경우를 지칭하는 것이다. 일차 연구대상군에서 재활의학과 전문의가 진단을 내린 대부분의 근골격계 질환은 근막통증후군이었으며, 소수에서 결절종이나 관절염이 근막통증후군과 병행하여 있었다. 이에 본 연구에서 일차대상군의 근골격계 질환은 전체 환례에 모두 포함되어 있는 근막통증후군으로 한정하여 분석하였다. 또한, 이차 연구대상군에서는 앞서 연구대상군에 대한 기술에서 언급한 바와 같이 누적외상성 장애 또는 근골격계 장애로 진단된 경우로 한정하여 분석하였다.

문진 및 이학적 검사상 국소적 동통이 현존하고 촉진시 근육내에 결절이 만져지며 결절내에 압통점이 존재하는 경우를 근막통증후군 '가능(possible)군'으로, 위의 기준에다가 압통점으로부터 전이통이 존재하거나 주사나 횡단면 파악(transverse snapping palpation)시 국소 연축반응이 보이거나 주사나 연장술에 의해 동통이 치유되는 경우를 근막통증후군 '확정(definite)군'으로 정의하였다. 본 연구에서는 이 최종적으로 '확정'된 군을 근막통증후군 환자군으로 정하였다.

3) 분석방법

(1) 근막통증후군 여부에 따른 교란요인들의 분포 파악

근골격계 질환의 발생에는 직업적 요인, 인적 특성, 심리적 요인 등이 복합적으로 작용하는 것으로 알려져 있다. 본 연구에서는 근막통증후군 여부에 따른 성별, 교육수준, 결혼 여부, 가사 노동, 노약자 보살핌, 연령, 작업 기간과 같은 관련 변수들의 분포를 파악하였다. 이는 이들 요인들이 교란요인으로 작용할 가능성을 검토하기 위한 것이었다.

(2) 설문문항의 선별 및 가중치 산정

설문문항의 선별 및 가중치 산정법이 서로 다른 4 가지 방법을 사용하였으며, 일차 연구대상군에서 수행한 선별방법 및 가중치 부여방법들 가운데 이차 연구대상군에서도 재현도가 높은 방법을 선정하는 것이 주안점이었다. 즉, 대상인구집단이 서로 다르더라도 통계적으로 유의한 차이가 없는 결과를 보여주는 방법을 찾고자 하는 것이다.

설문문항 선별에 따른 차이를 확인하기 위해 본 연구에서는 전체 52개 설문문항을 이용하는 방법(모형 1), 근막통증후근과의 관련성이 높은 문항군들 중에서 응답수준의 내적일치도(Cronbach's alpha 계수 이용)가 높은 설문항목들을 이용하는 방법(모형 2), 점근카이자승분포(asymptotic chisquaredistribution)를 따르는 score statistic을 이용하여 유의한 변수들을 자동적으로 선정하는 SAS® 6.12의 분석옵션을 이용하는 방법(모형 3), 신체부위별로 지난 1년간의 자각증상 여부를 묻고, 자각증상이 가장 심한 부위를 기준으로 증상의 빈도 및 중증도에 관련된 설문문항을 공통으로 사용하는 방법(모형

4) 등을 적용하였다.

선별된 각 문항에 대해 근막통증후군과의 단변수 로짓분석을 수행하여 산출된 각 문항별 모수추정치(parameter estimate) 즉, 로그변환된 교차비(odd ratio)를 문항별 가중치로 이용하였으며 각 문항의 응답수준을 지시변수(dummy variable)로 변환하여 산출된 모수추정치를 응답수준별 가중치로 사용하였다. 설문 응답수준에 따른 최종 가중치는 문항별 가중치와 응답수준별 가중치를 합산한 것이다. 신체부위별 문항과 자각증상 문항을 구별한 모형 4에서는 자각증상이 전혀 없는 경우에는 문항별 가중치만을 합산하였다. 이와 같이 단변수 분석의 모수추정치를 가중치로 이용한 경우로는 Harper 등(1995)을 들 수 있으며, Philosophov 등(1997)은 모수적 ROC분석을 통한 모의실험에서 변수가 많을수록, 비독립적으로 처리할수록 AUC가 증가함을 보여준 바 있다. 또한, 로그변환된 교차비가 설문문항의 가중치이므로 가중치의 단순 합산을 통해 교차비의 곱(product)을 가중치로 산정하게 하는 편의성을 제공한다. 모형별 문항선별 및 가중치 산출방법은 표 2에 제시한 바와 같다.

Table 2. Item selection criteria and total weight scores in each model

	Item selection criteria	Total weight score
Model 1	All the question items asking the subjective symptoms(n=52).	$\sum_{j=1}^{n}(R_j)$
Model 2	Question items with high internal consistency which is included in the question groups having statistically significant association with myofascial pain syndrome(n=10).	$\sum_{j=1}^{n}(\beta_{qj}+\beta_{sj})$
Model 3	Question items showing high score statistics derived from automatic variable selection process in logistic regression(n=10).	$\sum_{j=1}^{n}(\beta_{qj}+\beta_{sj})$
Model 4	Question items asking the symptom regions(m=4) and symptom severities(n=7).	$\sum_{i=1}^{m}\beta_{ri}+\sum_{j=1}^{n}(\beta_{qj}+\beta_{sj})$

n = number of the selected question items.
m = number of the symptom regions checked in the questionnair.
Rj = response score of jth question by the ordinal scale(1,2,3,... etc.).
βri = parameter estimate of ith region showing subjective symptom (i=1 for neck, 2 for shoulder, 3 for elbow, 4 for hand & wrist).
βqj = parameter estimate of jth question item for the subjective symptoms.
βsj = parameter estimate of each dummy variable by response level of subjective symptom severity in jth question item.

(3) ROC 분석

민감도와 특이도가 유병률과 무관하므로 ROC곡선은 조사대상 인구집단의 유병률 및 판별역치 효과(decision threshold effect)에 영향을 받지 않고 해당 측정도구의 질병 발견능력(disease detectability)을 시각적으로 비교하는데 유용하다. 반면, ROC 곡선에는 직접 구분점(또는 역치)이나 표본 수가 제시되지 않고, 표본 수가 감소하거나 측정치가 범주형 자료인 경우에는 불규칙한 직선형태를 띈다. 그리고, 계산상의 AUC가 동일하더라도 실제 ROC 곡선은 전혀 다른 형태를 가질 수 있으므로 ROC 분석의 결과를 이용할 때에는 전체 곡선을 제시하는 것이 권장되고 있다.

서로 다른 검사방법들의 판별력을 비교하는데 ROC 곡선을 이용하는 가장 초보적인 방법은 어느 곡선이 좌상측(左上側) 즉, (0,1)지점에 가깝게 그려졌는지 눈으로 확인하는 것이다. 하지만, 1980년 경부터 보다 엄밀한 통계적 방법론을 이용하는 방법들이 계속 개발되어 왔으며, ROC 분석을 통해 얻을 수 있는 단일한 판별력 지표로 가장 일반적으로 이용되어온 것이 AUC다. AUC의 계산과 비교방법은 환자군 및 정상군에서 관찰한 측정값의 분포에 대해 모수적 가정을 적용하는 방법과 비모수적 가정을 적용하는 방법으로 구분할 수 있다. Bamber(1975), Hanley 등(1982)이 소개한 비모수적 방법에 따르면 이표본통계량의 변형인 Mann-Whitney U-통계량과 AUC의 관계는 다음 공식 1과 같다.

$$AUC = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(\Psi_{ij} + \frac{1}{2}I_{ij})}{m \times n} \quad (1)$$

$\Psi_{ij} = 1$, when $Yj > Xi$
$\Psi_{ij} = 0$, when $Yj < Xi$
$I_{ij} = 1$, when $Yj = Xi$
$I_{ij} = 0$, when $Yj \neq Xi$

즉, 표본 X와 Y의 혼합표본 (Xi, Yj)에서 관계식 $\{Yj>Xi\}$를 만족하는 쌍의 개수(U-statistic)를 조합가능한 모든 쌍의 개수(m×n)로 나눈 것이 AUC이다.

DeLong(1983) 등을 거치면서 둘 이상의 AUC를 비교하기 위한 엄밀한 비모수적 방법론이 개발되었다. Zweig 등(1993)은 비모수적 방법의 장점을 지적한 바 있고, Centor(1991)는 모의실험을 통하여 측정값의 분포에 모수적 가정을 적용하기 어려운 경우에는 비모수적 방법이 적합하다는 결론을 제시한 바 있다. 본 연구에서는 비모수적 통계프로그램인 AccuROC® for Windows 95를 사용하였다.

또한, 공식 2에서 보듯이 ROC곡선의 기울기는 우도비(likelihood ratio, LR)를 의미하지만, 해석에는 주의를 요한다.

$$\frac{True\,(+)\,fraction}{False\,(+)\,fraction} = \frac{P(T(+)|D(+))}{P(T(+)|D(-))} = \mathrm{LR} \quad (2)$$

Bernard(1998)는 ROC곡선의 특정지점(x)에서 접선의 기울기와 원점(o)으로부터 그어진 직선의 기울기를 구분하여 각각 tangent(x)와 slope(o-x)로 표현하였다. Tangent(x)는 해당지점 x에서의 우도비의 순간적 변동량을 의미하며, slope(o-x)는 x이상을 양성, x미만을 음성으로 구분하는 2×2 표에서 산출되는 우도비를 의미한다. 접선의 기울기가 정보의 손실이 적고 측정도구의 진단적 가치를 가장

정확히 나타내는 것으로써, 일반적으로 접선의 기울기 1을 기준으로 구분점을 제시한다.

구분점 결정에 참고할 지표로는 효율(efficiency, $\frac{1}{2}$(*sensitivity+specificity*))을 이용하는 방법이 가장 기초적인데, 접선의 기울기가 1인 지점이 곧 효율이 가장 큰 지점에 해당된다. 하지만, 구분점은 측정도구 자체의 특성에만 의존하는 것이 아니며 임상적 의사결정(clinical decision making)이나 비용-편익분석(cost-benefit analysis)에서 그 이외의 다른 조건들이 영향을 미친다(Patton, 1978; Radack, 1986). Fombonne(1991)는 유병률 감소, 환례 정의(case definition)의 개선, 중재조치 비용의 증가 등을 구분점을 높여야 할 이유로 제한 바 있다. 이 경우는 민감도의 감소 및 특이도의 증가 즉, ROC곡선 상에서 기울기가 보다 가파른 지점으로 구분점을 옮기는 것이 된다. 또한, Swet(1996)는 비용-편익분석과 유병률에 따라 ROC곡선상의 최적 기울기를 결정하는 방법을 공식 3과 같이 제시하였다.

$$S_{opt} = \frac{1-p}{p} \times \frac{B_{True}(-) - C_{False}(+)}{B_{True}(+) - C_{False}(-)} \quad (3)$$

Sopt : Optimal slope.
p : Prevalence or prior probability.
$B_{True}(+)$, $B_{True}(-)$: Benefits of correct decisions.
$C_{False}(+)$, $C_{False}(-)$: Costs of incorrect decisions.

일단, 구분점이 정해진 후에는 원점으로부터 그어진 직선의 기울기와 질병의 유병률(prevalence, prior probability)을 알고 있는 경우, 베이스의 법칙(Bayes' rule)에 따라 다음 공식 4와 같이 양성예측도(positive predictive value, posterior probability)를 계산할 수 있다. 하지만, 이러한 적용은 분석표본의 유병률이 표적인구(target population)의 유병률과 같다는 가정을 만족할 경우에만 가능한 것이다(Linnet, 1998).

$$P = \frac{p \times LR}{p\,LR + (1-p)} \quad (4)$$

p=prevalence or prior probability.
P=positive predictive value or posterior probability.
LR=likelihood ratio.

본 연구에서는 업무관련성 상지 근골격계 질환에 관한 비용-편익분석이나 일반인구집단의 유병률은 조사하지 않았으므로 공식 3과 4를 이용한 결과해석은 어렵다. 하지만, 최적 기울기의 선정에 따른 구분점 변동의 양상을 확인하기 위해 기울기가 각각 0.5, 1.0, 2.0인 경우에 해당되는 구분점을 산출하였고 이를 유병률 및 설문지 합산총점의 분포, 효율 등과 비교하여 일정한 구분점을 제시하고자 하였다.

결 과

1. 일차 연구대상군의 일반적 특성

일차 연구대상군 370명을 직업군별로 분석한 결과, 연령은 가정 주부와 전화번호 안내원이 다른 군들에 비해 유의하게 높았다. 교육수준, 결혼 여부, 근무기간 등은 선박설계 작업자와 직업여성이 비슷하였고, 가정주부와 전화번호 안내원이 서로 또한 비슷하였다. 가사노동이나 노약자를 돌보는 부분에서는 직업여성과 전화번호 안내원이 서로 비슷한 양상을 보였다(표 3).

신체 부위별로 "최근 1년간 통증, 쑤심, 뻣뻣함, 저림, 감각 마비, 화끈거림 중 하나라도 느꼈던 증상이 있습니까?"라는 질문에 대한 설문문항에 "예"라고 응답한 경우를 기준으로 증상호소율을 비교한 결과, 유의한 차이가 있었다. 또한, 각 직업군별로 증상호소율과 근막통증후군 진단률 사이에 일관된 관련성이 있는지 확인하기 위하여 Breslow and Day's test for homogeneity를 시행하였다. 그 결과, 목(p=0.134), 어깨(p=0.708), 팔꿈치(p=0.692), 손 및 손목(p=0.141) 부위의

Table 3. Demographic and work-related characteristics by occupational groups

	Group (Persons, n=370)	Shipyard CAD workers (89)	Telephone directory operators (113)	Women with occupation† (79)	Housewives (89)
	Gender	Male	Female	Female	Female
		Mean (S.D.)	Mean (S.D.)	Mean (S.D.)	Mean (S.D.)
	Age*	32.3 (5.7)	38.5 (4.4)	35.4 (4.0)	39.9 (6.2)
	Work year*	9.0 (5.7)	18.1 (4.9)	9.8 (5.8)	15.7 (6.9)
		No. (%)	No. (%)	No. (%)	No. (%)
Education†	College	54 (60.7)	8 (7.1)	51 (64.6)	29 (32.6)
	High school	35 (39.3)	86 (76.1)	27 (34.2)	48 (53.9)
	Below middle school	-	17 (15.0)	-	3 (3.4)
	Missings	-	2 (1.8)	1 (1.3)	9 (10.1)
Marriage†	Married	65 (73.0)	107 (94.7)	60 (76.0)	83 (93.3)
	Not married	24 (27.0)	5 (4.4)	19 (24.1)	1 (1.1)
Household†	Labor	26 (29.2)	57 (50.4)	38 (48.1)	79 (88.8)
	Care	19 (21.4)	15 (13.3)	15 (19.0)	9 (10.1)
Symptom† Prevalence	Neck	59 (66.3)	100 (88.5)	46 (58.2)	49 (55.1)
	Shoulder	43 (53.2)	90 (79.7)	42 (53.2)	53 (59.6)
	Elbow	10 (11.2)	48 (42.5)	10 (12.7)	18 (20.2)
	Hand	38 (42.7)	84 (74.3)	24 (30.4)	38 (42.7)
	None of above	12 (13.5)	4 (3.5)	18 (22.8)	19 (21.4)
	Myofascial Pain Synd.(+)	30 (33.7)§	53 (46.9)	19 (24.1)	21 (23.6)

* : p <0.01, by ANOVA (between 4 groups)
† : p <0.01, by chi-square test (between 4 groups)
‡ : This group consists of nurses(34), insurance counselors(15), public officials(14), and clerks(16).
§ : In CAD workers, it's not a prevalence, but an arbitrary proportion. The minimal prevalence could be 3.8 %(30cases / 787subjects).
Labor : household loboring, alone / Care : taking care of children < 2yr. or elderly, disabled.

Table 4. Distribution of risk factors in MPS(+) group vs. MPS(-) group

Risk factors	Persons, incorporated	MPS(+) group (n=123)			MPS(-) group (n=247)		
		No.	Mean	(S.D.)	No.	Mean	(S.D.)
Age(yr.)	369	122	36.3	(5.8)	247	36.8	(5.9)
Work year	348	117	13.8	(7.1)	231	13.3	(6.9)
			No.	(%)*		No.	(%)†
Gender	Male	89	30	(24.4)		59	(23.9)
	Female	281	93	(75.6)		188	(76.1)
Education	College	142	44	(35.8)		98	(39.7)
	High school	196	64	(52.0)		132	(53.4)
	Below middle school	20	13	-		7	-
	Missings	12	2	-		10	-
Marrage	Married	15	105	(85.4)		210	(81.0)
	Not married	49	15	(12.3)		34	(13.8)
Household	Labor Yes.	200	62	(50.4)		138	(55.9)
	No.	149	51	(41.5)		98	(39.7)
	Care Yes.	58	16	(13.0)		42	(17.0)
	No.	84	19	(15.4)		65	(26.3)

* : The denominator is the number of persons in the MPS(+) group(n=123).
† : The denominator is the number of persons in the MPS(-) group(n=247).
Labor : household loboring, alone / Care : taking care of children < 2yr. or elderly, disabled.

Table 5. Symptom prevalence in MPS(+) group vs. MPS(-) group

Regions	Persons incorporated	MPS(+) group (n=123)		MPS(-) group (n=247)	
		No.	(%)*	No.	(%)†
Neck†	254	113	(91.8)	141	(57.9)
Shoulder†	228	106	(86.2)	122	(49.4)
Elbow†	86	42	(34.2)	44	(17.8)
Hand†	184	84	(68.3)	100	(40.5)

* : The denominator is the number of persons in the MPS(+) group(n=123).
†: The denominator is the number of persons in the MPS(-) group(n=247).
†: p <0.01, by chi-square test(between 2 groups)

Table 6. Selected questions and weights of each question in model 2 and model 3

Region	Questions in item	Weights by response level 1*	2	3	4	5	Items in Model 2	Items in Model 3
Neck	Frequency of symptom	6	21	23	32	35		●
	Severity of symptom	7	24	30	37	41	●	●
	Onset(before vs. after current job)	10	26				●	
	Relation to current job history	9	22	34			●	
Shoulder	Symptom since last week	16	25				●	●
	Self treatment since last year	10	21	27			●	●
	Frequency of symptom	5	15	22	20	31		●
	Severity of symptom	7	19	26	35	50	●	●
	Onset(before & after current job)	9	24				●	
	Relation to current job	8	19	29			●	●
	Symptom since last week	13	35				●	●
	Self treatment since last year	10	25	23			●	●

* : Weights in primay response level = weights in each item + 0

증상호소율과 근막통증후군 진단률의 관련성은 직업군이 달라도 유의한 차이를 보이지 않았다.

여기에서 선박설계 작업자는 전체 부서원 787명 중에서 2차에 걸친 검진을 통해 확진된 환례 30명과 임의선정된 대조군 59명으로 구성되었으므로 유병률을 정의할 수 없다. 다만, 유병률이 최소한 3.8%(30명/ 787명)보다는 클 것으로 추정할 수 있다.

근막통증후군 여부에 따른 성별, 연령, 교육수준, 결혼 여부, 가사 노동, 노약자 보살핌, 연령, 작업 기간과 같은 위험(혹은, 교란)요인들의 분포에는 유의한 차이가 관찰되지 않았다(표 4). (단, 교육수준의 경우, 중학교 이하인 20명과 미확인 12명은 비교하지 않았다.) 반면, "최근 1년간 통증, 쑤심, 뻣뻣함, 저림, 감각 마비, 화끈거림 중 하나라도 느꼈던 증상이 있었습니까?"라는 질문에 대한 응답결과는 근막통증후군 여부에 따라 유의한 차이가 관찰되었다(표 5)($p<0.01$).

근막통증후군과 신체 각 부위의 증상호소 사이의 관련성을 확인하기 위하여 로짓분석을 시행한 결과, 목(O.R.=5.1, C.I.=2.4~10.4), 어깨(O.R.=3.2, C.I.=1.7~5.7), 손 및 손목(O.R.=1.8, C.I.=1.1~3.0) 부위의 증상호소가 근막통증후군과 유의수준 5%에서 관련성이 있었으나, 팔꿈치 부위의 증상호소와는 유의한 관련성이 확인되지 않았다(O.R.=1.2, C.I.=0.7~2.1). 때문에, 기계적 변수선정 방법에 따른 문제점을 피하고자 모형 3에서는 팔꿈치에 관한 문항들을 제외하고 분석을 시작했다.

2. 문항선별 및 가중치 산정

모형 1은 전체 52문항에 대해 응답수준에 따른 순위(1,2,3.. 등)를 합산한 것이다. 모형 2를 적용하였을 때, 목과 어깨 부위에서 동일한 유형의 5개 문항(총 10문항)들을 선별할 수 있었다. 모형 3도 한 가지 유형의 문항만이 다를 뿐이었고 전체적으로 유사한 결과가 나왔다(표 6). 문항별 가중치와 응답수준별 가중치를 합산한 최종 가중치를 표 6에 제시하였는데, 응답수준

Table 7. Selected questions and weights of each question in model 4

Questions in item	Weights by response level				
	1	2	3	4	5
Neck Symptom	0	21			
Shoulder symptom	0	18			
Elbow symptom	0	9			
Hand and wrist symptom	0	12			
Duration	5	16	18	19	28
Frequency	7	21	24	31	39
Severity	9	25	38	44	40
Onset(before vs. after current job)	13	34			
Relation to current job history	13	19	34		
Symptom since last week	27	62			
Self treatment since last year	9	19	25		

Table 8. Demographic and work-related characteristics in secondary study population

Group (Persons, n=225)		Automobile assembling worker (n=225)	MSD(+)* group (n=29)	MSD(-) group (n=196)
		Mean (S.D.)	Mean (S.D.)	Mean (S.D.)
	Age†	32.6 (5.0)	33.2 (5.1)	32.5 (5.0)
	Work year†	8.7 (4.1)	8.8 (3.9)	8.7 (4.2)
		No. (%)§	No. (%)§	No. (%)§
Education	College	3 (1.3)	0 (0.0)	3 (1.53)
	High school‡	208 (92.4)	27 (93.1)	181 (92.4)
Below middle school‡		12 (5.3)	2 (6.9)	10 (5.1)
	Missings	2 (0.9)	0 (0.0)	2 (1.0)
Marrage	Married‡	165 (73.3)	23 (79.3)	142 (72.5)
	Not married‡	54 (24.0)	5 (17.2)	49 (25.0)
Symptom Prevalence	Neck	139 (61.8)	22 (75.9)	117 (59.7)
	Shoulder	139 (61.8)	25 (86.2)	114 (58.2)
	Elbow	64 (28.4)	14 (48.3)	50 (25.5)
	Hand	103 (45.8)	21 (72.4)	82 (41.8)
	None of above	32 (14.2)	0 (0.0)	32 (16.3)
	Musculoskeletal disorder(+)	29 (12.9)		

* : MSD means musculoskeletal disease.
† : p >0.05, by ANOVA (between MSD(+) group & MSD(-) group).
‡ : p >0.05, by Chi-square test (between MSD(+) group & MSD(-) group).
§ : In the percentage columns, the denominators are 225, 29, and 196, according to the order.

별 가중치의 기준값은 0이므로 첫 번째 응답수준의 최종가중치는 문항별가중치와 동일한 것이다. 신체부위별 4개 문항과 자각증상에 관한 7개 문항(총 11개 문항)으로 구성된 모형 4에서는 자각증상이 존재하는 신체부위가 전혀 없는 경우에는 기본값으로 문항별 가중치만을 부여하였다.

일차 연구대상군에서 나타난 점수를 직업군별로 비교한 결과, 전반적으로 전화교환원에서 점수가 높은 것을 알 수 있었다(그림 3). 직업군별 점수의 평균에 유의한 차이가 있는지 분산분석을 통해 검정한 결과, 모형 1, 2, 3 에서는 전화교환원과 나머지 직업군 사이에서만 유의한 차이가 나타났다(p=0.001). 한편, 모형 4를 이용한 합산총점의 직업군별 평균을 Duncan의 다중범위검정법을 이용하여 비교하였을 때, 전화교환원, 조선업 CAD 근로자, 직장여성 및 주부 등의 3군 사이에서 평균의 유의한 차이를 확인할 수 있었다. 각 모형들을 일차 연구대상군 총 370명에 적용하여 산출된 개인별 합산총점들의 분포를 살펴보면 모형 2, 3, 4가 근골격계 질환 여부에 따른 점수분포의 차이를 크게 함을 알 수 있었다(그림 4).

3. 이차 연구대상군의 일반적 특성

이차연구대상군은 평균연령(32.6세), 평균근무년수(8.7년), 결혼상태, 증상호소율 등에서 일차연구대상군과 통계적으로 유의한 차이가 없었다. 하지만, 교육수준은 92.4%가 고졸로서 일차연구대상군에 비해 분명한 차이가 나타났으며, 비교적 교육수준이 균일한 집단임을 알 수 있다(표 8).

이차 연구대상군에서도 상지 근골격계 질환 여부에 따른 연령, 교육수준, 결혼 여부 등의 분포에는 유의한 차이를 관찰할 수 없었다. 단, 가사노동, 노약자 보살핌과 같은 변수는 직업군의 특성을 고려하여 설문지에서 제외하였기에 확인할 수 없었다.

4. ROC 분석 결과

일차 연구대상군에서 적합된 4개 모형들의 문항선별 및 가중치 산정결과를 이차 연구 대상군에 적용한 결과, AUC는 전반적으로 감소하였다. 각 연구대상군 내에서는 모형별 AUC 사이에 통계적으로 유의한 차이를 확인할 수 없었다. 모형 1(p=0.1106)과 모형 4(p=0.3780)에서는 통계적으로 유의한 차이가 없는 AUC가 나타났다. 반면, 일차 연구대상군에서 높은 AUC를 나타내던 모형 2(p=0.0383)와 모형 3(p=0.0267)은 이차 연구대상군에서의 AUC와는 통계적으로 유의한 차이가 나타났다(표 7, 그림 4).

모형 2와 3의 AUC가 낮아진 원인을 파악하고자 개별문항들의 변별력을 확인하였다. 일차 연구대상군과 이차 연구대상군에서 가장 심한 증상을 중심으로 구성한 모형 4의 문항들과 목, 어깨, 팔꿈치, 손 및 손목 부위 문항들의 AUC 변동량의

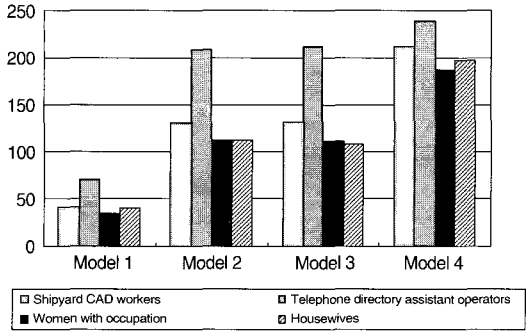

Figure 3. Mean of scores according to each model in four occupational groups.

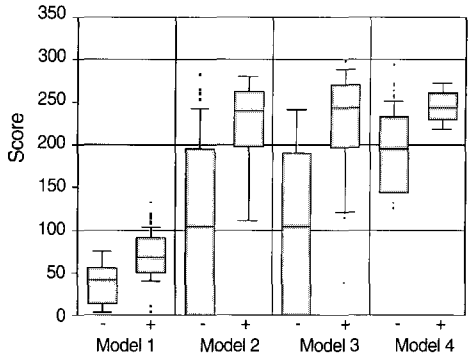

Figure 4. Box plots of score distribution with 90% confidence interval according to the diagnosis in 4 models. (-) : noraml / (+) : myofascial pain syndrome

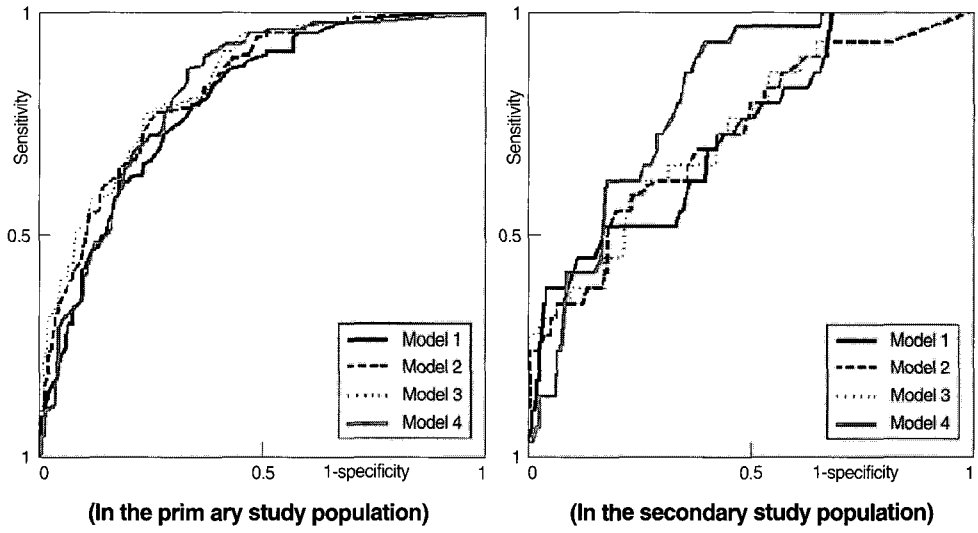

Figure 5. ROC plots in the primary and secondary study population by model 1 to 4. The area under the ROC curve by model 4 was filled with gray color.

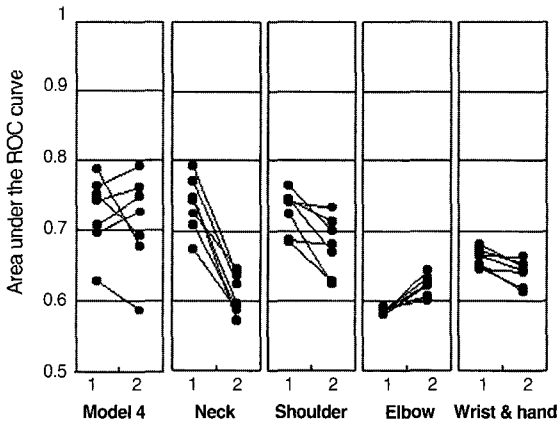

Figure 6. The changing patten of item AUC' s in primary and secondary study populations. (1=primary syudy population / 2=secondary study population)

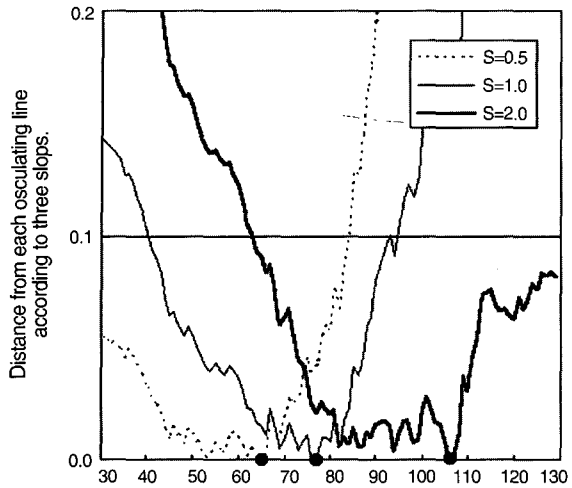

Figure 7. Distance from each osculating line according to three slopes by the order of cut-off points.

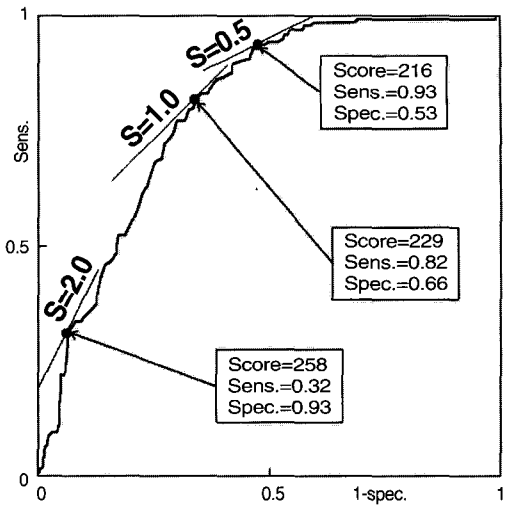

Figure 8. Designated cut-off point scpres according to each slope.(Sens : Sensitivity / Spec : Specificity)

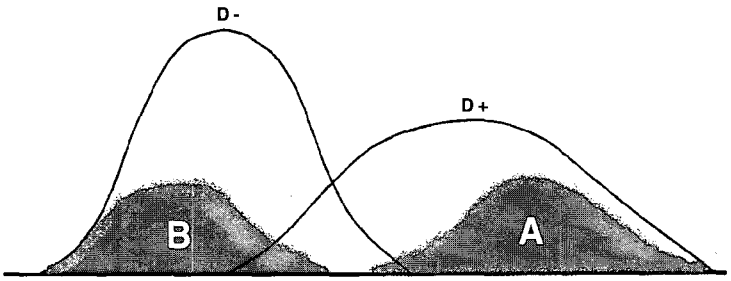

Figure 9. Selection of cases & controls for ROC analysis.

Table 9. The AUCs in the primary and secondary study populations by model 1 to 4

	AUC in the primary population. (95% C.I.)	AUC in the secondary population. (95% C.I.)	One-tailed p-value
Model 1	0.7942 (0.7480~0.8404)	0.7269 (0.6295~0.8242)	0.1106
Model 2	0.8245 (0.7815~0.8674)	0.7236 (0.6204~0.8268)	0.0383
Model 3	0.8344 (0.7927~0.8761)	0.7248 (0.6217~0.8278)	0.0267
Model 4	0.8155 (0.7714~0.8595)	0.8026 (0.7343~0.8709)	0.3780

평균치를 비교한 결과, 통계적으로 유의한 차이가 나타났으며(p=0.0003) Duncan의 다중범위검정법을 이용하여 목부위 문항들에서 나타난 AUC 변동량들의 평균이 다른 문항군들의 AUC 변동량의 평균과 유의한 차이를 보이는 것을 확인할 수 있었다. 또한, 통계적으로 유의한 평균의 차이를 보이지는 않았으나 어깨부위 모든 문항들도 AUC가 감소하는 경향을 볼 수 있었다(그림 6).

모형 4를 이용하여 전화안내원(113명), 직업여성(79명), 주부(89명), 자동차 조립공정 근로자(225명)등 총 506명(양성진단률=0.241)을 대상으로 최적기울기에 따른 구분점의 변동양상을 확인하였다. 단, 조선업체 근로자의 경우는 유병률을 정의할 수 없으므로 제외하였다. 이러한 모의실험 결과, 기울기가 0.5, 1.0, 2.0인 경우에 각각 65번째, 77번째, 106번째 구분점이 선정되었다(그림 7). 각 구분점의 합산총점은 각각 216점, 229점, 258점에 해당되었다(그림 8).

총 506명 가운데 122명(0.241)이 상지근골격계 질환으로 진단받았으며 정상인의 상위 사분위수는 234점, 환자군의 하위 4분위수는 238점이었다. 한편, 일차 연구대상군에서 해당 점수가 233점 및 229점, 이차 연구대상군에서는 244점 및 225점이었으므로, 최적기울기를 1.0로 상정한 구분점 229점이 가장 근접한 값으로 판단되며 효율은 0.7406이었다. 표적인구의 유병률을 0.241로 가정한다면 구분점 229점에서의 우도비(0.82÷(1-0.66)=2.4)를 이용하여 계산한 양성예측도(post-test probability)는 0.432다. 또한, 229점 이상을 역치로 가정하면 교차비는 8.9(5.3~14.8)였다.

고 찰

측정방법의 타당도를 평가하고자 하는 모든 접근방법에서 기본적으로 요구되는 것은 실제 건강 또는 질병 상태가 판정된 충분한 수의 표본 즉, 충분한 수의 참값이다. 본 연구에서 이용한 참값은 전문의의

진단이고 비교하고자 하는 측정도구는 설문문항이다. 하지만, 실험적 연구나 모의실험과 달리 실제 임상에서 다수의 참값을 확보한다는 것은 대단히 어려운 일이다. Metz(1978)는 ROC 분석 이외의 다른 통계적 방법론에서도 이러한 문제점들이 마찬가지로 존재한다는 점을 지적하면서, "물론, 참값의 정의는 궁극적으로 철학적 문제다. 그럼에도 불구하고 평가분석을 위해서는 진단적 참값의 기능적 기준(operational standards)이 반드시 확립되어야 한다"고 주장한 바 있다. 즉, 참값의 기준이 완벽한 것이길 기대할 수는 없어도, 최소한 평가하고자 하는 측정방법보다는 믿을 만한 것이어야 한다.

본 연구는 동일한 전문의가 모든 진단을 내리거나 또는 진단의 신뢰도 검정을 거친 전문의 집단을 통해 참값을 얻은 것은 아니므로, 전문의 사이의 진단기준 차이로 인한 변이가 존재하리라 예견된다. 하지만, 일차 연구대상군에서 직업군별로 증상호소율과 상지 근골격계 질환 진단률이 통계적으로 유의한 차이를 보이지 않았으며, 본 연구에서 진단을 내린 전문의들은 사전에 Hales 등(1994)이 제시한 업무관련성 근골격계 질환의 정의를 숙지하고 설문지에 표명된 업무관련성 근골격계 질환의 증상과 빈도를 문진과정에서 다시 한번 확인하는 절차를 거쳤다. 따라서, 본 연구에서 전문의간의 진단기준에 차이가 존재하지만 그 변이정도가 크지 않을 것이며, 진단결과의 타당도가 설문지보다 낮지는 않을 것이라는 최소한의 추정은 가능하다고 판단된다.

ROC 분석에서 중요한 또 하나의 문제는 환자군 및 대조군의 선정에 관한 것이다. Metz(1989), MacMahon(1988), Kelsey(1989)등은 적절한 환자군의 선정이 ROC 분석에 미치는 영향을 이론적 · 실험적으로 제시한 바 있으며, Gur(1990) 등은 적절한 대조군의 선정 역시 중요한 영향을 미친다는 점을 보여 주었다. 그림 9는 이러한 영향을 간략히 설명하기 위한 것이다. 대조군(D-)과 환자군(D+)를 이용한 ROC 곡선의 AUC는 0.5와 1.0 사이의 값을 취할 것이다. 그러나 만약 측정치가 높은 환자군의 일부(A)만이 선택되어 대조군(D-)과 비교되거나, 측정치가 낮은 정상인군의 일부(B)만이 선택되어 환자군(D+)과 비교된다면 이 측정도구의 변별력이 높아지는 방향으로 비뚤림이 일어난다. 즉, AUC가 더 커진다. 게다가 A와 B를 비교한다면 완벽한 변별력 즉, AUC=1이 된다. 따라서, ROC 분석을 수행하기 위한 표본선정 시에는 가능한 모든 질병변이(disease variant)가 포함되도록 하는 것이 특히 중요하다.

본 연구는 가능한 다양한 직종, 연령, 성별을 포괄하고자 하였다. 조선소 CAD 작업자에서 근막통증후군으로 확진된 환자군, 이상증상은 호소하고 있으나 근골격계 장애를 확진받지 못한 근로자군, 근골격계 이상증상이 전혀 없는 근로자군을 포함시키고, 전화번호 안내원 조사연구를 보완하고 직업군간 근골격계 장애의 유병률을 비교하기 위하기 위해 과도한 수작업을 하지 않거나 컴퓨터를 많이 이용하지 않는 직업 여성들과 주부를 임의로 선정하여 수행한 연구결과 등을 포함시킨 것은 ROC분석결과의 신뢰도를 높이는데 기여할 것으로 사료된다.

연구과정에서 검토한 4가지 모형 중에서 모형 2와 3은 두 연구대상군에서 산출된 AUC간에 통계적으로 유의한 차이를 보였다. 그 원인은 연구결과의 그림 6에서 보이듯이 모형 2, 3을 구성하고 있는 목부위 및 어깨부위 문항들의 AUC가 이차 연구대상군에서 낮아진 것에 반해, 신체부위에 대한 문항과 증상에 대한 문항을 계통적으로 배치한 모형 4의 문항들은 통계적으로 유의한 변동이 없었다. 즉, 자동차 조립공에서는 팔꿈치 부위 설문문항의 변별력이 통계적으로 유의하지는 않지만 상대적으로 높아진 것으로 판단되며, 신체부위에 상관없이 증상설문을 공통으로 사용하는 것이 보다 포괄적이고 유연한(flexible) 측정을 가능케 함으로서 재현도를 높인 것으로 추정된다. 또한, 모형 4가 모형 1에 비해 문항 수가 적음에도 불구하고 높은 AUC를 보인 것은 문항의 응답수준별 가중치를 부여하였기 때문인데, 가중치를 부여하지 않은 경우에 모형 4는 일차 연구대상군에서는 0.717, 이차 연구대상군에서는 0.605라는 저조한 AUC가 나타났다. 즉, 신체부위에 대한 문항들과 증상에 대한 문항들을 계통적으로 배치하는 방법과 문항별 가중치를 부여한 것이 모형 4의 재현도 및 변별력을 높이는데 기여한 것으로 판단된다.

특히, 문항별 가중치는 변별력을 높인다는 측면뿐만 아니라 설문조사의 결과를 정량적 지표로 산출함으로써 해석과 활용이 간단한 질병감시도구의 요건을 충족시킨다는 중요한 의의가 있다. 예를 들어, 현행 근로자 건강진단에는 이미 자각증상을 묻는 설문지가 도입되어 있으나 그로부터 얻어진 자료는 방치되어 있는 실정이다. 활용방안이나 기준이 전혀 없다는 점은 이러한 문제발생의 주요원인 중에 하나라고 판단된다. 본 연구에서는 임의로 선정한 세 가지의 최적기울기와 유병률에 대한 가정, 설문지 합산총점의 분포, 효율 등을 고려하여 구분점, 양성예측도 등을 추정하고자 시도하였는데, 이는 근로자 건강진단 등의 실제 현장에서의 설문지를 활용할 때 일정 정도나마 참고할 기준을 제시할 필요성이 있다고 판단되었기 때문이다. 향후, 상지 근골계 질환의 직업군병 유병률 및 그 변동양상, 조기진단 및 질병감시로 인한 손익 등에 대한 파악이 시도된다면 보다 의미있는 구분점 선정이 가능할 것이다.

또한, 다양한 직업군 및 조건하에서 측정된 결과를 확보한다면 보다 객관적인 지표를 개발할 수 있을 것이지만, 설문조사 고유의 특성을 고려할 때 일정한 한계가 예상된다. 즉, 주관적 증상호소가 주를 이루는 근골격계 질환은 직장의 업무환경을 포함한 사회심리적 조건들이 발병과 관련이 있고 인구집단별 증상호소 양상도 다를 것으로 예상되므로 본 연구에서 도출된 설문지의 합산총점의 분포를 근거로 일반인구집단에서 사용할만한 참고치를 제시하기는 어렵다고 판단된다. 하지만, 본 연구는 개별 환자에 대한 의학적 확진도구를 지향하기 보다는 인구집단에서 확보된 관찰값과 통계학적 확률론에 의거하여 질병감시용 도구 또는 진단보조 도구

로 활용할만한 설문도구를 개발하고자 하였으며, 앞에 언급한 질병 외적 요인들에 일정한 동질성을 가정할 수 있는 경우에는 측정값의 상대적 비교가 가능할 것으로 판단된다. 예를 들어, 일정한 관찰기간 내에 특정업종의 고위험 공정 근로자군을 판별하기 위해 부서별 평균값을 비교하거나 인구집단별 증상호소 특성을 관찰하는 등의 후속작업이 필요할 것이다.

본 연구진은 새롭게 설계된 설문지를 1998말에서 1999년초 사이에 수행된 국내 모 자동차 조립공장 및 조선업체 근로자의 건강진단에 실제 적용하는 과정에서 일부 문항을 개선할 수 있었으며, 향후 본 설문지의 실용성에 대한 후속연구를 수행할 예정이다. 이와 같은 표준화 작업을 통해 확보된 측정도구는 일차 검진 및 역학적 예비조사의 효율과 정확성을 높이고, 고위험 집단을 조기에 파악하거나 발생규모를 유추하는 등의 거시적 접근에 요구되는 기본 정보를 제공함으로써 보건계획 및 정책수립, 집행, 평가 등에 보다 일관된 근거를 제시하는데 기여할 수 있을 것이다.

요 약

ROC곡선의 AUC는 측정도구의 기준 타당도를 나타내는 가장 일반화된 지표다. 본 연구는 ROC분석법을 이용하여 현행의 근로자건강진단에서 업무관련성 근골격계 질환의 고위험군을 변별하는 표준 설문지를 개발하고자 하였다.

컴퓨터를 이용하는 선박 설계업 종사자 89명, 전화번호 안내원 113명, 일반 직업여성 79명, 주부 89명 등 총 370명의 일차 연구대상군에 대한 재활의학과 전문의의 최종 진단결과를 기준으로 1996년에 개발된 '근로자의 신체 증상에 관한 설문지'의 응답결과를 비교하였다. 근골격계 질환과의 관련성이 높은 문항조합을 선정하고 문항별 가중치를 산출하기 위해 로짓회귀분석, 상관분석 등을 실시하였으며, 문항조합 및 가중치 산출방법이 서로 다른 4가지 설문모형에 따른 AUC를 비교하였다. 또한, 국내 모 자동차조립공장 근로자 225명의 설문결과와 산업의학 전문의의 진단결과 자료를 이용하여 4가지 설문모형의 AUC 재현도를 확인하였다. 분석 결과, 통계적으로 유의한 차이는 없었으나 문항수를 줄여도 문항별 · 응답수준별 가중치를 부여하면 AUC가 일관되게 증가함을 확인하였다. 증상문항 4개와 신체부위문항 7개를 통합한 11개 문항에 가중치를 부여하는 모형이 변별력, 재현도, 편의성 측면에서 우수한 것으로 나타나, 이를 기준으로 새로운 업무관련성 근골격계 질환 설문지를 설계할 수 있었다.

문항수가 적으면서도 타당도는 높은설문지를 개발하고, 상대적인 비교평가에 쓰일 수 있는 정량적 가중치를 제시한 것이 본 연구의 주요성과라 할 수 있다. 본 연구는 전문의 사이의 진단기준 차이를 고려하지 못한 점, 다양한 인구집단에 적용할만한 절대적인 참고치를 제시하지 못한 점 등에서 한계가 있다. 그러나, "측정도구의 정량적 타당도 검증을 통한 질병감시용 도구 개발"이라는 본 연구의 기본 취지 및 접근방법은 향후 조직적인 질병예방활동에 활용될 여지가 있을 것이다.

참고문헌

권호장, 하미나, 윤덕로, 조수헌, 강대희, 주영수, 백도명, 백남종. VDT작업자에서 업무로 인한 정신사회적 스트레스에 대한 인지가 근골격계장애에 미치는 영향. 대한산업의학회지 1996; 8(3): 570-577.

김돈규, 조수헌, 한태륜, 권호장, 하미나, 백남종. VDT 업무가 근골격계 장해에 미치는 영향. 대한산업의학회지 1998; 10(4) : 524-533

김수근, 정해관. VDT작업자에서 어깨 근육장애와 혈청 CPK의 관련성. 대한산업의학회지 1998; 10(2):172-179.

박정근. 일부 영상단말기(VDT) 작업자의 작업자세에 관한 조사연구. 서울대학교 보건대학원 보건학 석사학위논문. 1990.

박정일, 조경환, 이승한. 여성 국제 전화 교환원들에 있어서의 경견완 장애 Ⅰ. 자각적 증상. 대한산업의학회지 1989; 1(2): 141-150.

신병순, 박정일. VDT작업자의 견관절 근육의 압통역치. 대한산업의학회지 1996; 8(1): 15-26.

조수헌, 강대희, 김재용. 직업병 및 산업재해의 감시체계(2). 산업보건 1997; 6: 37-49

임상혁, 이윤근, 조정진, 손정일, 송재철. 은행 창구 작업자(VDT 작업자)의 경견완장애 자각증상 호소율과 관련 요인에 관한 연구. 대한산업의학회지 1997; 9(1): 85-98.

최재욱, 염용태, 송동빈, 박종태, 장성훈, 최정애. 반복 작업 근로자들에서의 경견완장애에 관한 연구. 대한산업의학회지 1996; 8(2): 301-319.

Bamber D. The area above the ordinal dominance graph and the area below the receiver operating graph. *J Math Psychol* 1975; 12: 387-415.

Bernard CK, Choi. Sloped of a receiver operating characteristic curve and likelihood ratiofor a daignostic test. *Am J Epidemiol* 1998; 148(11): 1127-1132.

DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biomatrics* 1988 ; 44: 837-845.

Gur D, King JL, Rockette HE, Britton CA, Thaete FL, Hoy RJ. Practical issues of experimental ROC analysis-selection of controls. Statistics in Radiology 1990 ; 25: 583-586.

Fombonne E. The use of questionnaires in child psychiatry research: Measuring their performance and choosing an optimal cut-off. *J Child Psychol Psychiat* 1991; 32(4): 677-693.

Garyfallos G, karastergiou A, Adamopoulou A, Moutzoukis C, Alagiozidou E, Mala D, Garugallos A. Greek version of the general health questionnaire: accuracy of translation and validity. *Acta Psychiatr Scand* 1991; 84: 371-378.

Hales TR, Sauter SL, Peterson MR, Fine LJ, Anderson VP, Schlhifer LR, Ochs TT, Bernard BP. Musculoskeletal disorders among visual display terminal users in a telecomunications company. *Ergonomics* 1994 ; 37(10) : 1603-1621

Halperin W, Baker EL Jr. Public Health Surveillance. N.Y.; Van Nostrand Reinhold. 1992.

Hanley JA, McNeil BJ. A method of comparing the area under the receiver operating characteristic curves derived from the same cases. *Radiology* 1983; 148: 839-843.

Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic curve. *Radiology* 1982; 143: 29-36.

Harper AC et al. Development and validation of the Curtin Back Screening Questionnaire (CBSQ): a discriminative disability measure. *Pain* 1995; 60: 73-81.

Hartzema AG, Porta P, Tilson HH. Pharmacoepidemiology: An introduction, 3rd ed. Cincinnati; Harvey Whitney Books

Co., 1998.

Hosmer DW, Lemeshow S. Applied logistic regression. N.Y.; John Wiley & Sons, Inc. 1989.

Karasek R. Job demands, job decision latitude and mental strain : Implication for job redesign. *Am Sci Q* 1979 ; 24 : 285-308

Linnet K. A review on the methodology for assessing diagnostic tests. *Clin Chem* 1988; 34(7): 1379-1386.

Kelsey CA, Mettler FA Jr. ROC analysis can reveal best diagnostic method. *Diagn Imag* 1989; 11: 155-161.

MacMahon H, Metz CE, Doi K, Kim T, Giger ML, Chan HP. Digital chest radiography: Effect on diagnostic accuracy of hard copy, conventional video, and gray scale video display formats. *Radiology* 1988; 168: 669-673.

Metz CE. Basic principles of ROC analysis. *Semin Nucl Med* 1978 Oct; 8(4): 283-298.

Metz CE. Some practical issues of experimental design and data analysis in radiological ROC studies. Invest Radiol 1989; 24: 234-245.

Malison MD. Surveillance in developing countries. In : Halpein W & Baker EL(ed). Public Health Surveillance. New York, Van Nostrand Reinhold ; 1992; p56-61.

Morasso G et al. Assessing psychological distress in cancer patients: validation of a self-administered questionnaire. *Oncology* 1996; 53: 295-302.

Patton DD. Introduction to clinical decision making. *Semin Nucl Med* 1978 ; 8(4): 273-282.

Philosophov LV, Ryabinina LG. Medical diagnostic decision rules based on mutually dependent diagnostic factors. *Comput Biol Med* 1997; 27(4): 329-347.

Radack KL, Rouan G, Hedges J. The likelihood ratio-an improved measure for reporting and evaluating daignostic test results. *Arch Pathol Lab Med 1986* ; 110: 689-693.

Swets JA. Signal detection theory and ROC analysis in psychology and diagnostics-collected papers. Bolt Beranek and Newman Inc. and Harvard Medical School. A Volume in the Scientific Psychology Series, New Jersey, 1996.

Van Bemmel JH, Musen MA, eds. Handbook of medical informatics. Houten/Diegem; Springer, 1997.

Zweig MH , Campbell G. Receiver-operating charac-teristic(ROC) plots: a fundimental evaluation tool in clinical medicine. *Clin Chem* 1993; 39(4): 561-577.