

낙상 위험 선별검사 Timed Up and Go test의 예측 타당도 메타분석

A Meta-analysis of the Timed Up and Go test for Predicting Falls

■ 박성희¹, 이연석²

Seong-Hi Park¹, On-Seok Lee²

■ 순천향대학교 간호학과¹, 순천향대학교 의료IT공학과²

Department of Nursing, Soonchunhyang University¹, Department of Medical IT Engineering,
Soonchunhyang University²

■ 교신저자 : 박 성 희

주소 : 충남 아산시 신창면 순천향로 22, 순천향대학교 간호학과

전화 : 041-530-4854

팩스번호 : 041-570-2498

전자우편주소 : shpark2015@sch.ac.kr

Correspondence : Seong-Hi Park

Address : 22 Soonchunhyang-ro, Sinchang-myeon, Asan-si, Chungcheongnam-do, 31538 Korea.

Tel : +82-41-530-4854

FAX : +82-41-570-2498

E-mail : shpark2015@sch.ac.kr

Funding : None

Conflict of Interest : None

Received : Nov. 12. 2016

Revised : Dec. 16. 2016

Accepted : Dec. 17. 2016

Abstract

Purpose: Globally, falls are a major public health problem. The study aimed to evaluate the predictive validity of the Timed Up and Go test (TUGT) as a screening tool for fall risk.

Methods: An electronic search was performed Medline, EMBASE, CINAHL, Cochran Library, KoreaMed and the National Digital Science Library and other databases, using the following keywords: 'fall', 'fall risk assessment', 'fall screening', 'mobility scale', and 'risk assessment tool'. The QUADAS- II was applied to assess the internal validity of the diagnostic studies. Thirteen studies were analyzed using meta-analysis with MetaDisc 1.4.

Results: The selected 13 studies reporting predictive validity of TUGT of fall risks were meta-analyzed with a sample size of 1004 with high methodological quality. Overall predictive validity of TGUT was as follows. The pooled sensitivity 0.72 (95% confidence interval [CI]: 0.67-0.77), pooled specificity 0.58 (95% CI: 0.54-0.63) and sROC AUC was 0.75 respectively. Heterogeneity among studies was a moderate level in sensitivity.

Conclusion: The TGUT's predictive validity for fall risk is at a moderate level. Although there is a limit to interpret the results for heterogeneity between the literature, TGUT is an appropriate tool to apply to all patients at a potential risk of accidental fall in a hospital or long-term care facility.

Key words

Accidental falls, Timed up and go test, Sensitivity and specificity, Meta-analysis

I. 서론

낙상 예방은 오늘날 보건의료가 직면한 가장 중요한 도전과제 중 하나이다. 세계보건기구(World Health Organization)에 의하면, 낙상은 교통사고 다음으로 발생하는 비의도적 손상의 두 번째 주요 원인이고, 전 세계적으로 매년 42만 4천명이 낙상으로 사망하고 있으며, 특히 60세 이상 노인에서 그 빈도가 가장 높다고 보고하고 있다[1].

이에 영국과 미국노인병학회[2]와 National Institute for Health and Care Excellence (NICE)[3] 등에서는 낙상예방을 위한 진료지침을 개발하여 보급하고 있다. 이는 우리나라에서도 예외는 아니어서 대한내과학회를 주축으로 표준진료지침위원회가 구성되어 2015년 국내 낙상예방 진료지침을 개발하였다[4]. 본 진료지침에 따르면, 낙상 위험군 선별을 위해 낙상병력 청취와 보행과 균형을 평가하는 낙상 선별검사를 실시하도록 권장하고 있다. 이에 권고된 낙상의 위험도 선별검사는 Timed Up and Go test (TUGT)[5], 버그균형척도(Berg Balance Scale)[6], Performance-Oriented Mobility Assessment (POMA)[7]이며, 이는 전문가 의견 등 임상경험에 근거한 근거수준 E등급과 좋은 효과가 나쁜 효과보다 확실히 더 크고, 권고대로 했을 때 대부분 효과가 있는 권고수준 1의 강한 권고로 제시되었다[4].

낙상 예방을 위한 선별검사의 중요성에 대해서는 국내외 의료기관 인증기준[8,9]에도 포함되어 상급 종합병원에서 요양병원에 이르기까지 이를 준수하는 의료 질 향상 활동을 수행하고 있으며, 2016년 7월부터 시행되고 있는 환자안전법에서도 낙상과 같은 위해사건에 관한 환자안전사고보고시스템 구축과 환자안전지표 측정이 강화되고 있어 의료기관에서도 낙상과 같은 위해사건에 대한 현황을 모니터링하고 이를 예방하기 위한 근거중심의 실무 적

용을 더욱 증진해야 할 것으로 생각된다.

그러나 국내 낙상예방 진료지침에 제시된 TUGT나 버그균형척도는 주로 재활의학분야에서 사용되는 도구들이어서[10-12] 상대적으로 덜 알려져 있으며, 병원간호사회의 낙상예방간호 실무지침서에서는 보바스기념병원과 Morse 낙상위험사정도구를 제시하고 있다[13]. TUGT는 기본적인 동적 균형 및 이동 능력을 짧은 시간 내에 쉽게 측정할 수 있는 검사로 팔걸이가 있는 의자에 앉아 신호와 함께 일어서서 3m 전방 목표물 주위를 될 수 있는 한 빨리 돌아와 다시 의자에 앉는 시간을 측정한다. 기존에 사용하던 신발이나 보행 보조 기구를 사용하게 하거나 보조자의 도움을 받을 수는 없다. 검사 완료까지 10초 이내가 걸리면 정상이며, 노쇠한 노인의 경우 11-20초가 정상치의 상한선이다. 20초 이상 걸린다면 외부 활동 시 도움이 필요하다는 뜻이며, 추가적인 검사나 조치가 필요하다. 객관적인 측정을 통해 비교적 간단하게 낙상위험을 사정할 수 있는 도구이다[14,15].

따라서 본 연구에서는 환자안전에 대한 중요성이 더욱 강조되는 현 시점에서 급성기 병원 뿐 아니라 요양병원과 요양시설에서 그 활용이 보다 요구되는 TUGT에 초점을 맞추어 본 도구가 환자의 낙상위험을 선별하는데 추천할 수 있는 적절한 도구인지를 확인하고, 그 효용성에 대한 과학적 근거를 제시해 보고자 시도되었다.

II. 연구 방법

본 연구는 코크란 연합의 진단법 정확도에 대한 체계적 문헌고찰 핸드북[16] 및 PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 그룹이 제시한 체계적 문헌고찰 보고지침[17]에 따라 수행되었다.

1. 문헌검색전략

전자 데이터베이스를 이용한 검색은 2014년 8월 30일에 시행되었으며, Ovid-Medline, Embase, Cochrane Library와 CINAHL Complete, KoreaMed, 국가과학기술정보센터(NDSL), 한국교육학술정보원(KERIS) 검색 엔진이 이용되었다. 또한, 낙상 관련분야의 학술지가 빠짐없이 포함될 수 있도록 대한간호학회지를 비롯하여 8개 학회지를 2014년 12월 14일에 검색하였다. 이후 2016년 6월 2014년 8월 이후 출간된 문헌에 대해 재검색되었다.

사용된 검색어는 TUGT로 국한하기 보다는 'fall risk assessment', 'fall screening', 'mobility scale', 'risk assessment tool'로 다소 넓게 검색되었으며, 각 단어가 사용되는 형태를 검토하여 절단검색(truncation), 만능문자(wild cards) 및 인접연산자(positional operators)를 활용하였다. 진단법 평가 연구에 대한 검색 필터(search filter)는 SIGN (Scottish Intercollegiate Guidelines Network)에서 제시한 검색전략을 활용하였다.

2. 문헌 선택 및 자료추출

문헌선택 기준은 TUGT를 사용하여 전향적으로 낙상발생 예측의 정확성을 평가한 연구, TUGT를 이용하여 진양성, 위양성, 진음성, 위음성 결과를 보고한 연구, 동료심사(peer-reviewed)를 거쳐 출판된 문헌으로 국한하였다. 원저가 아니거나 낙상사정도구가 주 증제가 아닌 연구, 진단법 평가 연구가 아니거나 연구결과에서 민감도, 특이도만 제시하고 진양성, 위양성, 진음성, 위음성 값을 통해 2×2 이원분류표를 작성하는데 충분한 진단정확도 자료를 제공하지 않은 연구는 배제하였다.

검색된 문헌에서 일차적으로 중복된 문헌을 제거하

고 문헌의 제목 및 초록을 검토하였으며, 정확한 판단이 곤란한 경우는 원문을 찾아 선택 및 배제기준을 적용하여 문헌을 선별하였다. 자료추출은 우선 근거표(evidence table) 기본 서식을 작성하여 시범적으로 서식의 적절성을 검토한 후 사용되었다. 선택된 문헌의 연구유형, 연구장소, 연구기관 수, 연구대상의 성별 분포, 평균 연령, 낙상발생 추적관찰기간, 낙상 판단기준, 경계점수(cut-off point) 및 낙상위험사정도구의 진단결과를 기술하였다. 모든 과정은 2명의 저자가 각자 독립적으로 검토한 후 합의를 통해 결정하였으며, 의견의 불일치가 있는 경우는 제3자 개입의 원칙을 정하였으나 연구자간 이견없이 진행되었다.

3. 문헌의 질 평가

문헌의 질은 Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) 도구[18]를 활용하여 독립적으로 평가하였다. QUADAS-2는 진단법 평가 연구에 대한 질 평가 도구로 '편중의 위험(risk of bias)' 과 '적용(applicability)' 을 평가하게 되며, 평가항목은 대상자 선택(patient selection), 시험 진단법(index test), 참고표준검사(reference standard)와 과정 및 적시성(flow and timing)의 4가지 영역(domain)으로 구성되어 있다. 각 영역은 평가항목에 따라 예, 아니오, 불확실로 평가하고 최종적으로 편중의 위험과 적용성의 평가결과는 위험이 높음, 낮음, 불확실로 기술된다. 문헌의 질은 2명의 연구자가 독립적으로 평가한 후 명확하지 않는 부분은 논의를 통해 정리하였다.

4. 통계적 분석

MetaDiSc 1.4를 이용하여 메타분석하였다[19]. 진단법 메타분석에서 통합 추정치를 분석하고자 하는 경우, 연구간 이질성을 반영하기 위해 랜덤효과모형(random

effect model) 사용을 권고하므로[20], 통계적 모델에 대한 일반적 원칙은 이를 토대로 통합 민감도와 특이도, 양성 및 음성우도비와 진단교차비로 분석하였다. 또한, sROC 곡선 통계량은 곡선아래면적(Area under the curve [AUC])과 index Q^* 값을 통해 검사 정확도를 기술하였다. AUC의 수치는 $AUC=0.5$ 인 경우 비정보적인 검사, $0.5 < AUC \leq 0.7$ 은 정확성이 떨어짐, $0.7 < AUC \leq 0.9$ 는 중등도, $0.9 < AUC < 1$ 는 매우 정확함 및 $AUC=1$ 은 완벽한 검사로 판단하였고[19], ROC 곡선에서 민감도와 특이도의 동등점을 반영하는 index Q^* 는 100% 정확할 때 '1'을 기준으로 판단하였다[21]. 문헌들 간 이질성(heterogeneity)의 존재여부는 유의수준 5% 미만으로 하여 카이제곱 검정(χ^2 -test)과 Higgins의 I^2 동질성 검사 (I^2 test)로 평가하였고 I^2 의 판단기준은 $I^2 \leq 25\%$ 이면 이질성이 낮은 것으로, $25\% < I^2 \leq 75\%$ 는 중간정도의 이

질성, $I^2 > 75\%$ 이상은 이질성이 높다고 해석하였다[22].

III. 연구결과

1. 문헌선택 결과

전자 데이터베이스를 통해 총 2,571개 문헌이 검색되었다. 중복된 문헌을 250개를 제외한 2,321개 문헌을 토대로 우선 제목과 초록을 검토하였고 선택 및 배제기준에 부합된 연구인지 확인하였다. 제목과 초록만으로 정확히 판단하기 어려운 경우는 원문을 찾아 문헌을 선별하였다. 최종적으로 2,308개(99.4%) 문헌이 배제되고, 13개 문헌[23-35]이 최종 선택되었다. 자세한 문헌 선택 과정과 결과는 흐름도로 제시하였다 (Figure 1).

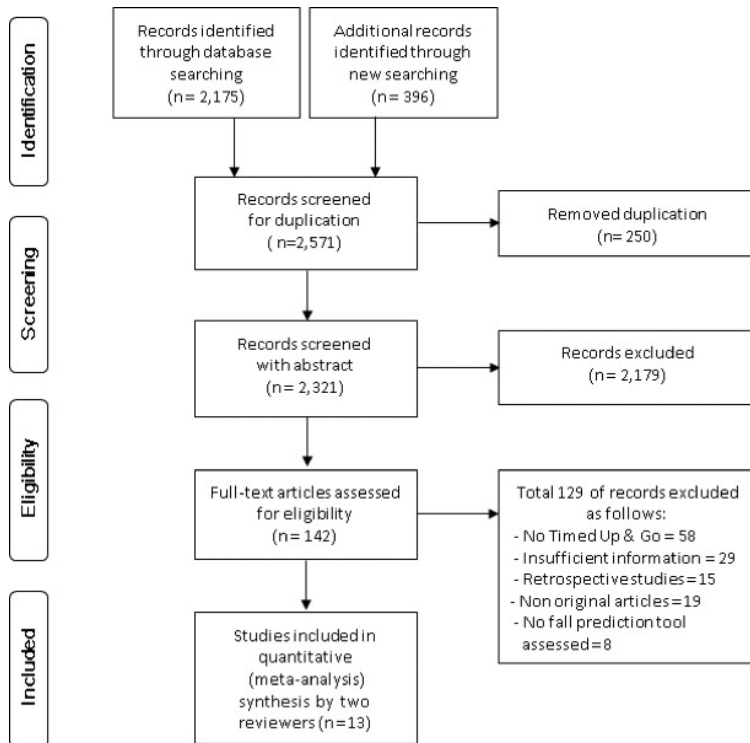


Figure 1. Flow diagram of article selection

2. 문헌의 질 평가 결과

선택된 13편 연구에 대한 평가결과, 전체적으로 각 영역에 대해 편중의 위험이 '높음'으로 판단된 문헌은 없었다. Jernigan 등[25]의 연구는 편의표본추출(convenience sampling) 되었고, 2개 문헌[23,35]을 제외

한 대부분의 문헌에서 대상자 선택과정이 명확히 기술되지 않았으나 모두 전향적인 연구이므로 연구결과를 해석하는데 영향을 주지 않았고, 모든 문헌에서 부적절한 배제가 없어 환자 선택 영역에서 편중의 위험은 없다고 해석하였다. 따라서 선택된 문헌은 모두 질 평가 영역을 충족하는 고품질의 문헌이었다(Table 1).

Table 1. Quality assessment results of the selected studies by QUADAS-2

Study	Risk of bias				Application concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Martinez et al.[23]	☺	☺	☺	☺	☺	☺	☺
Alexandre et al.[24]	☺	☺	☺	☺	☺	☺	☺
Jernigan et al.[25]	☺	☺	☺	☺	☺	☺	☺
Möller et al [26]	☺	☺	☺	☺	☺	☺	☺
Bhatt et al. [27]	☺	☺	☺	☺	☺	☺	☺
Marschollek et al [28]	☺	☺	☺	☺	☺	☺	☺
Kerr et al. [29]	☺	☺	☺	☺	☺	☺	☺
Wrisley & Kumar [30]	☺	☺	☺	☺	☺	☺	☺
Nilsagård et al [31]	☺	☺	☺	☺	☺	☺	☺
Schwesig et al. [32]	☺	☺	☺	☺	☺	☺	☺
Dite et al [33].	☺	☺	☺	☺	☺	☺	☺
Dibble et al [34].	☺	☺	☺	☺	☺	☺	☺
Whitney et al. [35]	☺	☺	☺	☺	☺	☺	☺

☺ Low risk; ☹ High risk; ? Unclear risk.

3. 선택된 문헌의 일반적 특성

낙상위험사정도구인 TUGT의 예측 타당도를 보고한 문헌은 총 13편으로 총 대상자는 1,004명이었다. 선택된 문헌들은 미국이 4편[25,27,30,34]으로 가장 많았고 독일[28,32], 브라질[23,24], 스웨덴[26,31]과 호주[29,33]가 각각 2편이었고, 영국이 1편[35]이었다. 병원 입원환자에게 적용된 연구는 2편[23,28]이었고, 외래에 내원하거나 연구에 자발적으로 참여한 지역사회 거주 주민들이었다. 연구대상자들은 파킨슨병[29,34], 다발성 경화증[31], 당뇨병으로 인한 말초신경증 등의 신경계 질환[25]이 있거나 지역사회 거주하는 60세 이상 된 노

인들이었다. TUGT가 사용된 연령층은 다양하였다. 물론 대부분의 연구는 노인이 주된 대상이었으나 만성질환이 있는 경우는 평균 연령이 다소 낮은 경향을 보였다. 낙상은 모든 연구에서 낙상자(fallers)를 대상으로 하였고, 낙상을 판별하는 참조표준기준은 1회가 11편으로 대부분이었으나 2회 이상 반복적인 낙상으로 정의한 연구도 3편[25,33,34] 있었다. 낙상의 고위험을 결정하는 TUGT의 경계점수는 가장 적게는 7.6초[27], 가장 많게는 30.1초[32]로 다양하여 일관성을 찾기 어려웠으나 대부분 10-15초 범위였다. 낙상발생의 추적관찰기간은 대부분 6개월 정도였고, 12개월 이상 장기간 추적된 경우도 3편[25,27,34] 있었다(Table 2).

Table 2. Characteristics of Selected Studies

Year of publication	Authors	Location	Participants			2×2 Table				Value (95% Confidence interval)							
			Subjects	Total (n)	M:F (n)	Mean age (years)	Aged criteria	Falls criteria	Cut off (sec)	Follow up (mos)	TP	FP	TN	SN	SP	DOR	
2015	Martinez et al.[23]	Brazil	Elderly inpatients	68	40:28	70.4±7.7	≥60	≥1	10.85	UD	14	5	7	42	0.67 (0.43-0.85)	0.89 (0.77-0.96)	16.80 (4.59-61.48)
2012	Alexandre et al.[24]	Brazil	Community-dwelling older adults	63	30:33	66.7±5.6	≥60	≥1	12.47	6	14	14	5	27	0.74 (0.49-0.91)	0.66 (0.49-0.80)	5.40 (1.61-18.07)
2012	Jernigan et al.[25]	USA	Diabetic peripheral neuropathy	36	21:15	57.3±6.1	40-65	≥2	10.7	12	9	3	1	23	0.90 (0.55-1.00)	0.88 (0.70-0.98)	69.00 (6.32-753.62)
2012	Möller et al.[26]	Sweden	Community-dwelling older adults	153	51:102	81.5±6.3	≥65	≥1	12.0	6	11	37	4	17	0.73 (0.45-0.92)	0.31 (0.20-0.46)	1.26 (0.35-4.55)
2011	Bhatt et al.[27]	USA	Community-dwelling older adults	119	44:75	71.0±6.0	≥65	≥1	7.6	12	33	23	26	33	0.56 (0.42-0.69)	0.59 (0.45-0.72)	1.82 (0.87-3.82)
2011	Marschollek et al.[28]	Germany	Geriatric inpatients	50	13:37	81.3	-	≥1	20.0	UD	17	21	2	6	0.89 (0.67-0.99)	0.22 (0.09-0.42)	2.43 (0.43-13.61)
2010	Kerr et al.[29]	Australia	People with Parkinson disease in community	101	68:33	66.4±8.2	-	≥1	10.1	6	33	20	15	33	0.69 (0.54-0.81)	0.62 (0.48-0.75)	3.63 (1.59-8.29)
2010	Wrisley & Kumar [30]	USA	Community-dwelling older adults	35	17:18	72.9±7.8	≥60	≥1	11.0	6	5	4	1	25	0.83 (0.36-1.00)	0.86 (0.68-0.96)	31.25 (2.86-341.88)
2009	Nilsagård et al.[31]	Sweden	People with multiple sclerosis in community	76	19:57	50.0	≥18	≥1	13.6	6	32	18	16	10	0.67 (0.52-0.80)	0.36 (0.19-0.56)	1.11 (0.42-2.96)
2009	Schwesig et al.[32]	Germany	Nursing home residents	108	16:92	81.9±9.5	≥51	≥1	30.1	6	15	26	10	57	0.60 (0.39-0.79)	0.69 (0.58-0.78)	3.29 (1.30-8.29)
2007	Dite et al.[33]	Australia	Rehabilitation hospital and general community	40	30:10	59.9-65.2	≥18	≥2	19.0	6	11	7	2	20	0.85 (0.55-0.98)	0.74 (0.54-0.89)	15.71 (2.77-89.10)
2006	Dibble et al.[34]	USA	People with Parkinson disease in community	45	-	69.9±11.3	-	≥2	7.95	12	23	14	2	6	0.92 (0.74-0.99)	0.30 (0.12-0.54)	4.93 (0.87-27.88)
2005	Whitney et al.[35]	UK	Consecutive patients at the fall clinic	110	27:83	79.3±7.2	≥63	≥1	15.0	2	39	38	9	24	0.81 (0.67-0.91)	0.39 (0.27-0.52)	2.74 (1.13-6.64)

TP= True positive; FP= False positive; FN= True negative; SN= Sensitivity; SP= Specificity; DOR= Diagnosis odds ratio; UD= Until discharge.

4. TUGT의 예측 타당도

TUGT의 예측 타당도 평가결과는 선택된 문헌을 대상으로 통합 민감도, 특이도, 양성 및 음성우도비

와 진단교차비, sROC 통계량을 종합적으로 기술하였다. 또한, 문헌들간 이질성의 원인을 조사하기 위해 연구대상의 특성 및 낙상기준에 따라 세부 분석하였다 (Table 3).

Table 3. Summary results of meta-analysis

Tools	No. of studies (patients)	Results of pooled diagnostic test accuracy (95% Confidence interval)										Results of sROC curve			
		Sensitivity				Specificity				Positive likelihood ratio	Negative likelihood ratio	Diagnostic odds ratio	AUC (SE)	Q* (SE)	
		Value	I ₂ (%)	X ²	p	Value	I ₂ (%)	X ²	p						
Total	13 (1004)	0.72 (0.67-0.77)	52.6	25.29	.014	0.58 (0.54-0.63)	88.5	104.78	<.001	1.78 (1.39-2.28)	0.52 (0.41-0.67)	4.02 (2.37-6.83)	0.75 (0.04)	0.69 (0.03)	
Sub-group analysis															
Subjects	Elderly (≥60)	6 (548)	0.69 (0.61-0.76)	45.4	9.16	.103	0.58 (0.52-0.64)	91.6	59.32	<.001	1.95 (1.27-2.97)	0.53 (0.38-0.75)	4.04 (1.78-9.17)	0.75 (0.05)	0.69 (0.04)
	with disease	4 (258)	0.74 (0.66-0.81)	66.1	8.85	.031	0.57 (0.48-0.65)	87.5	24.00	<.001	1.69 (1.03-2.76)	0.48 (0.24-0.96)	4.33 (1.23-15.28)	0.80 (0.15)	0.74 (0.13)
Falls	>1	10 (883)	0.69 (0.64-0.74)	38.8	14.70	.099	0.57 (0.53-0.62)	89.2	83.56	<.001	1.62 (1.26-2.09)	0.58 (0.47-0.71)	3.13 (1.89-5.19)	0.71 (0.04)	0.66 (0.03)
Criteria	>2	3 (121)	0.90 (0.77-0.97)	0.0	0.48	.789	0.67 (0.55-0.78)	89.2	18.53	.0001	2.96 (0.99-8.89)	0.20 (0.08-0.47)	14.60 (3.66-58.20)	0.93 (0.04)	0.86 (0.05)

ROC curve= Receiver operating characteristic curve; AUC=Area under the curve; SE=Standard error.

선택된 문헌 총 13편(1,004명)을 토대로 TUGT의 예측 타당도에 대한 메타분석 결과, 통합 민감도는 0.72 (95% CI: 0.67-0.77), 문헌들간 이질성은 52.6% (X²=25.29, p=.014)로 중간 정도였고, 통합 특이도는 0.58 (95% CI: 0.54-0.63), 문헌들간 이질성은 88.5% (X²=104.78, p<.001)로 높은 수준을 보였다. 통합 양성 우도비는 1.78 (95% CI: 1.39-2.28), 통합 음성 우도비는 0.52 (95% CI: 0.41-0.67), 통합 진단교차비는 4.02 (95% CI: 2.37-6.83)이었고, sROC AUC는 0.75 (SE=0.04), Q*값은 0.69 (SE=0.03)이었다(Figure 2).

연구대상의 특성에 따른 세부분석은 60세 이상 노인과 신경계 질환이 있는 환자군으로 구분하였다. 60세 이상 노인들만을 대상[23,24,26,27,30,35]으로 한 6편(548명)의 문헌을 분석한 결과, 통합 민감도는 0.69 (95% CI: 0.61-0.76), 문헌들간 이질성은 45.4% (X²=9.16, p=.103)로 중간 정도였고, 통합 특이도는

0.58 (95% CI: 0.52-0.64), 문헌들간 이질성은 91.6% (X²=59.32, p<.001)로 높은 수준이었다. sROC AUC는 0.75 (SE=0.05), Q*값은 0.69 (SE=0.04)이었다. 파킨슨 병 등의 신경계 질환이 있는 환자를 대상으로 한 4편의 연구[25,29,31,34](258명)에서 통합 민감도는 0.74 (95% CI: 0.66-0.81), 문헌들간 이질성은 66.1% (X²=8.85, p=.031), 통합 특이도는 0.57 (95% CI: 0.48-0.65), 문헌들간 이질성은 87.5% (X²=24.00, p<.001), sROC AUC는 0.80 (SE=0.15), Q*값은 0.74 (SE=0.13)이었다.

낙상판별 기준에 따라 살펴보면, 1회 이상을 낙상기준으로 본 문헌은 10편(883명)이었다. 메타분석 결과, 통합 민감도 0.69 (95% CI: 0.64-0.74), 문헌들간 이질성은 38.8% (X²=14.70, p=.099)였다. 통합 특이도는 0.57 (95% CI: 0.53-0.62), 문헌들간 이질성은 89.2% (X²=83.56, p<.001), sROC AUC 0.71 (SE=0.04), Q*값은 0.66 (SE=0.03)이었다. 2회 이상을 낙상기준으로

본 문헌 3편[25,33,34](121명)의 메타분석 결과, 통합 민감도 0.90 (95% CI: 0.77-0.97), 문헌들간 이질성은 0.0% ($X^2=0.48, p=.789$)로 전혀 없었고, 통합 특이도

0.67 (95% CI 0.55-0.78), 문헌들간 이질성은 89.2% ($X^2=18.53, p=.0001$), sROC AUC 0.93 (SE=0.04), Q^* 값 0.86 (SE=0.05)이었다.

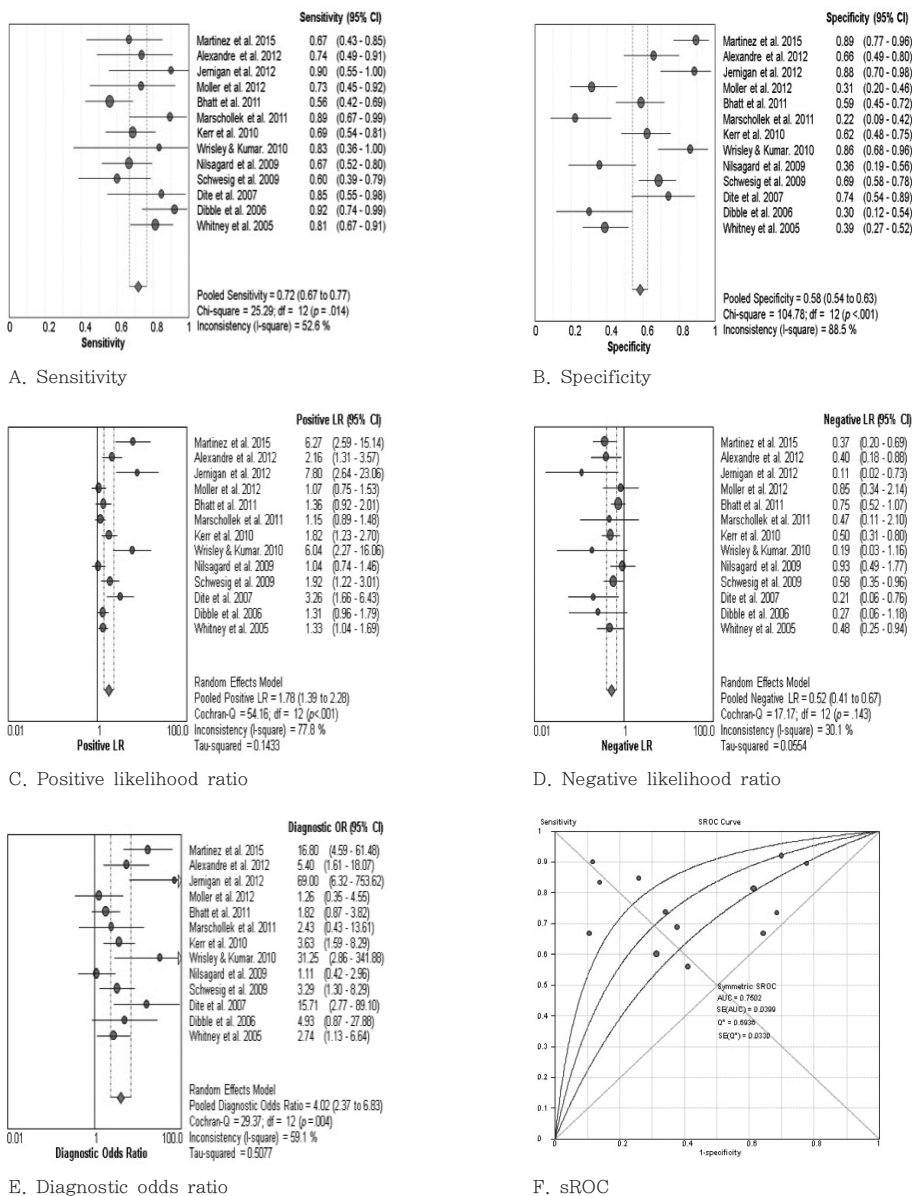


Figure 2. Diagnosis test accuracy of Timed Up and Go Test

IV. 고찰

예측 타당도는 검사나 평가도구에서 얻은 점수와 미래의 어떤 결과와의 관계로 추정되는 준거 타당도이다. 낙상위험사정도구는 낙상발생을 확진하는 진단검사가 아니라 낙상의 위험을 예측하는 선별검사이다. 이상적으로 선별검사는 민감도, 특이도, 양성 및 음성예측도가 모두 높을수록 좋으나 이는 현실적으로는 불가능하므로 이들 중 어떤 지표에 중점을 두게 된다[36]. 낙상이 발생할 가능성이 있는 환자를 고위험군에 포함하여 낙상발생을 최대한 억제하고자 한다면 민감도가 높아야 하고, 어느 정도의 낙상발생 환자가 고위험군에서 누락되더라도 낙상예방관리에 투입되는 노동력의 효율성을 높이고자 한다면 특이도가 높아야 한다[37]. 낙상은 연령에 상관없이 누구에게나 발생될 수 있지만 낙상 위험을 선별해야 하는 주요 대상은 노인이다. 노인에게 다발하는 낙상은 생명을 위협하며 사망을 초래한다. 따라서 주로 노인을 대상으로 적용되는 낙상 위험사정도구는 민감도가 높을수록 보다 적절할 것으로 판단된다. 이에 본 연구에서는 영국과 미국[2] 및 우리나라의 낙상예방 진료지침[4]에서 그 사용을 권고하고 있는 TUGT를 보다 효율적으로 활용하기 위해 총 13편(총 1,004명)의 진단법 평가연구들의 메타분석을 통해 정량적인 분석결과(quantitative summary)를 제공하고자 시도되었다.

선택된 총 13편의 문헌에서 보고된 TUGT의 민감도는 1편[27]을 제외하고는 모두 0.6 이상의 결과를 나타냈다. 통합 민감도는 0.72였고, 문헌들간의 이질성은 52.6%로 중간 수준이었다. TUGT의 특이도는 가장 낮게는 0.22부터 0.89로 큰 폭의 차이를 나타냈고, 0.5 미만을 보고한 문헌도 5편 [26,28,31,34,35]이나 되었다. 통합 특이도는 0.6 수준이었고 문헌들간 이질성도 80% 이상으로 높았다.

다만, 민감도의 관점에서만 보면 TUGT 도구의 낙상위험 발생에 대한 예측 타당도는 $0.7 < AUC \leq 0.9$ 사이의 값을 가지므로 중등도의 검사정확도를 갖는다고 해석할 수 있다. 이 결과는 NICE 진료지침[3]에서도 이미 지적된 바와 같이 낙상위험을 예측하는 도구들이 갖는 문제점이라고 할 수 있다.

실무에서의 TUGT 활용도를 살펴보면, 급성기 병원이나 요양시설 뿐 아니라 지역사회에 거주하는 노인들에게서 모두 특정 환경에 구분 없이 사용되고 있었다. 또한 노인 뿐 아니라 뇌졸중이나 파킨슨병과 같은 신경계 관련 질환에서도 낙상 예방을 위해 그 사용 범위가 넓음을 알 수 있었고, 예측 타당도의 차이는 크지 않아 균형능력의 상실이 예상되는 모든 환자에서 구별없이 사용될 수 있는 범용적 선별도구로 보인다. 특히 의사, 초시계와 같은 간단한 도구만으로도 쉽게 측정할 수 있고, 본 도구의 관찰자간 신뢰도가 높은 것으로 알려져 있어[5] 장점을 가진 편리한 도구로 생각되었다.

연구대상자의 특성에 따라 세부분석한 결과에서 60세 이상 노인에 비해 신경계 질환자에서 통합 민감도가 0.74로 다소 높았고, 문헌들간의 이질성의 수준은 큰 변화를 보이진 않았다. 따라서 TUGT는 균형능력에 이상이 있는 노인이나 환자들의 낙상발생 위험을 안정적으로 예측하는 선별도구로 해석할 수 있었다. 낙상을 판별하는 참조표준기준을 1회 낙상(통합 민감도 0.69)으로 했을 때보다 2회 이상 반복적인 낙상(통합 민감도 0.90)을 기준으로 할 때 민감도가 더 높았고 문헌들간 이질성도 0.0% ($p=.789$)로 전혀 없었다. 이는 낙상 고위험군과 저위험군을 구분할 때 노인이나 신경계 질환을 가진 환자의 경우 일반 성인에 비해 낙상 잠재력이 높은 집단이며, 1회 낙상은 압도적인 외적 요소에 의해 발생할 수 있고[38], 반복적인 낙상은 isolated fall 보다 intrinsic predisposition, 즉 의도하지 않은 자

세의 변화로 발생할 수 있으므로[39] 낙상의 위험성을 보다 정확히 예측할 수 있다고 판단된다. 따라서 2회 이상의 낙상을 참조표준기준으로 판별함이 합리적일 것으로 보이며, 본 연구를 통해 증명되었다고 볼 수 있다.

본 연구결과를 종합해 보면, 현재 사용하고 있는 연구대상의 특성 및 참조표준기준에 따라 민감도에 차이를 보이긴 하지만, 특정 환경에 구분없이 쉽게 측정할 수 있고 문헌들 간 이질성의 차이가 크지 않으며, 안정된 예측 타당도 값을 가지므로 낙상 위험을 선별하는 도구로 실무에서 사용하기에는 적절할 것으로 해석된다.

V. 결론

병원 및 일상 환경에서 가장 빈번한 낙상사고를 예방하는 차원에서 신뢰성 있는 낙상위험사정도구를 선택하고 활용하는 것은 매우 중요하다. 본 연구에서는 전향적 연구설계로 진행된 총 13편의 진단법 평가문헌을 토대로 한 메타분석을 통해 TUGT는 노인이나 뇌졸중 및 파킨슨병과 같은 균형능력에 문제가 있는 환자에서 안정적으로 낙상위험을 예측하는 선별도구로 증명되었다. 따라서 동 결과를 토대로 보다 실무에서 낙상예방 진료지침에서 권고에 따라 TUGT가 활발히 사용되고, 더 효과적인 낙상관리를 위해 현재 검증 없이 쓰이고 있는 다양한 도구들에 대한 과학적 근거를 찾는 꾸준한 시도가 있기를 기대해 본다.

VI. 참고문헌

1. World Health Organization. Fact sheet 344: Falls [Internet]. Geneva: World Health Organization; 2016 [cited 2016 October 24]. Available from <http://www.who.int/media-centre/factsheets/fs344/en/>.
2. Panel on Prevention of Falls in Older Persons, American Geriatrics Society and British Geriatrics Society. Summary of the Updated American Geriatrics Society/British Geriatrics Society Clinical Practice Guideline for Prevention of Falls in Older Persons. *Journal of the American Geriatrics Society*. 2011;59(1):148-157.
3. National Institututr for Health and Care Excellence. Falls: Assessment and prevention of falls in older people. NICE clinical guidelines 161. Manchester, UK: National Institute for Health and Care Excellence; 2013.
4. Kim KI, Jung HK, Kim CO, Kim SK, Cho HH, Kim DY, et al. Evidence-based Guideline for Fall Prevention in Korea. *Korean Journal of Medicine*. 2015; 89(6): 752-780.
5. Podsiadlo D, Richardson S. The timed ‘Up & Go’: A test of basic functional mobility for frail elderly persons. *Journal of the American Geriatrics Society*. 1991;39:142-148.
6. Berg K, Wood-Dauphinee SL, Williams JI, Gayton D. Measuring balance in the elderly: preliminary development of an instrument. *Physiotherapy Canada*. 1989;41:304-311.
7. Tinetti ME, Speechley M, Ginter SF. Risk factors for falls among elderly persons living in the community. *New England Journal of Medicine*. 1988;319(26):1701-1707.
8. Ministry of Health & Welfare, Korea Institute for Healthcare Accreditation [KOHA].

낙상 위험 선별검사 Timed Up and Go test의 예측 타당도 메타분석

- Guideline for healthcare accreditation [Internet]. Seoul: KOIHA. 2014 [cited 2015 January 10]. Available from: <https://www.koiha.or.kr/home/data/data/doView.act>
9. The Joint Commission. 2015 Comprehensive accreditation manual for hospitals. Oak Brook, IL: Joint Commission Resources; 2014.
 10. Kim MC, Ahn CS, Kim YS. The Effect of Exercise Program for Falls Prevention on Balance and Quality of Life in the Elderly Women. *Journal of Korean Society of Physical Medicine*. 2010;5(2):245-254.
 11. Lee KS. Effects of Balance Control on Physical Performance of Elderly Women. *Neurotherapy*. 2012;16(1):37-43.
 12. Lee HS, Choi JH. Correlation Between BBS, FRT, STI, TUG, MBI, and Falling in Stroke Patients. *Journal Korean Society of Physical Therapy*. 2008;20(4):1-8.
 13. Hospital Nurses Association. Fall prevention nursing practice guideline and summary. [Internet]. Seoul: Hospital Nurses Association; 2016 [cited 2016 October 24]. Available from http://khna.or.kr/bbs/linkfile/resource/khna_Fcare.pdf.
 14. Cho BL, Scarpace D, Alexander NB. Tests of Stepping as indicators of mobility, balance, and fall risk in balance impaired older adults. *Journal of the American Geriatrics Society*. 2004;52(7):1168-1173.
 15. Bohannon, RW. Reference values for the timed up and go test: A descriptive meta-analysis. *Journal of Geriatric Physical Therapy*. 2006;29(2):64-68.
 16. Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Analysing and presenting results. In: Deeks JJ, Bossuyt PM, Gatsonis C, editors. *Cochrane handbook for systematic reviews of diagnostic test accuracy version 1.0* [Internet]. The Cochrane Collaboration; 2010 [cited May 30, 2013]. Available from <http://srdta.cochrane.org/>.
 17. Moher D, Liberati A, Tetzlaff J, Altman DG, Antes G, Atkins D, et al. Preferred reporting items for systematic reviews and meta-Analyses: The PRISMA statement. *Annals of Internal Medicine*. 2009;151(4):264-269.
 18. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*. 2011;155(8):529-536.
 19. Zamora J, Abraira V, Muriel A, Khan KS, Coomarasamy A. Meta-DiSc: A software for meta-analysis of test accuracy data. *BMC Medical Research Methodology* 2006;6:31.
 20. Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Preventive Veterinary Medicine*. 2000;45(1-2):23-41.
 21. Walter SD. Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data. *Statistics in Medicine*. 2002;21(9):1237-1256.
 22. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Statistics in*

- Medicine. 2002;21(11):1539-1558.
23. Martinez BP, Gomes IB, Oliveira CS, et al. Accuracy of the Timed Up and Go test for predicting sarcopenia in elderly hospitalized patients. *Clinics*. 2015;70(5):369-372.
 24. Alexandre TS, Meira DM, Rico NC, Mizuta SK, Alexandre TS, Meira DM et al. Accuracy of Timed Up and Go Test for screening risk of falls among community-dwelling elderly. *Rev Bras Fisioterapia* 2012; 16(5):381-388.
 25. Jernigan SD, Pohl PS, Mahnken JD, Kluding PM, Diagnostic accuracy of fall risk assessment tools in people with diabetic peripheral neuropathy. *Physical Therapy*. 2012; 92(11):1461-1470.
 26. Möller UO, Kristensson J, Midlöv P, Ek-dahl C, Jakobsson U. Predictive validity and cut-off cores in four diagnostic tests for falls - A study in frail older people at home. *Physical and Occupational Therapy in Geriatrics*. 2012;30(3):189-201.
 27. Bhatt T, Espy D, Yang F, Pai YC. Dynamic gait stability, clinical correlates, and prognosis of falls among community-dwelling older adults. *Archives of Physical Medicine and Rehabilitation*. 2011;92(5), 799-805.
 28. Marschollek M, Rehwald A, Wolf KH, Gietzelt M, Nemitz G, ZuSchwabedissen HM, et al. Sensors vs. experts - A performance comparison of sensor-based fall risk assessment vs. conventional assessment in a sample of geriatric patients. *BMC Medical Informatics & Decision Making*. 2011;11:48.
 29. Kerr GK, Worringham CJ, Cole MH, Lach-erez PF, Wood JM, Silburn PA. Predictors of future falls in Parkinson disease. *Neurology*. 2010;75(2):116-124.
 30. Wrisley DM, Kumar NA. Functional gait assessment: concurrent, discriminative, and predictive validity in community-dwelling older adults. *Physical Therapy*. 2010;90(5):761-773.
 31. Nilsagård Y1, Lundholm C, Denison E, Gunnarsson LG. Predicting accidental falls in people with multiple sclerosis: A longitudinal study. *Clinical Rehabilitation*. 2009;23(3):259-269.
 32. Schwesig R, Kluttig A, Kriebel K, Becker S, Leuchte S. Prospective comparison of assessments to evaluate fall risk in a nursing home population. *Zeitschrift für Gerontologie and Geriatrie*. 2009; 42(6):473-478.
 33. Dite W, Connor HJ, Curtis HC. Clinical identification of multiple fall risk early after unilateral transtibial amputation. *Archives of Physical Medicine & Rehabilitation*. 2007;88(1):109-114.
 34. Dibble LE, Lange M. Predicting falls in individuals with Parkinson disease: a reconsideration of clinical balance measures. *Journal of Neurologic Physical Therapy*. 2006;30(2):60-67.
 35. Whitney JC, Lord SR, Close JC. Streamlining assessment and intervention in a falls clinic using the Timed Up and Go Test and Physiological Profile Assessments. *Age and Ageing*. 2005;34(6):567-571.
 36. Oh DY, Kim JH, Lee PK, Ahn ST, Lee JW.

Prevention of pressure ulcer using the pressure ulcer risk assessment based on Braden Scale. *Journal of the Korean Society of Plastic and Reconstructive Surgeons*. 2007;34(4):456-469.

37. Lee Y, Jeong I, Jun S. A comparative study on the validity among pressure ulcer risk assessment scales. *Journal of Korean Academy of Nursing*. 2003;33(2):162-169.
38. Thomas JI, Lane JV. A pilot study to explore the predictive validity of 4 measures of falls risk in frail elderly patients. *Archives of Physical Medicine and Rehabilitation*. 2005;86(8):1636-1640.
39. Faulkner KA, Redfern MS, Cauley JA, Landsittel DP, Studenski SA, Rosano C, et al. Multitasking: association between poorer performance and a history of recurrent falls. *Journal of the American Geriatrics Society*. 2007;55(4):570-576.