

10 Measurement error

2005. 5. 18.
의료관리학교실
김부경

측정오류에 대하여 이 장에서 논의할 것들은 다음과 같다.

- ▷ 원인(Sources)
- ▷ 평가(Assessment)
- ▷ 결과(Consequence)
- ▷ 노출과 질환의 관계(Exposure-disease relationship)

1. Source of measurement error

- 노출상태나 정도에 대한 미측정은 거의 모든 연구에서 어느정도는 발생하는데 이는 노출 여부나 그 정도를 확정하는 것이 불가능하기 때문임.
- 질문이나 인터뷰는 잘못된 정보를 포함할 수 있는데, 대상자들이 그들의 노출에 대해 기억하지 못하거나 반대로 과장되게 표현할 수 있기 때문임. 혈압과 같이 질환의 특징이 한 개인에서 여러번 측정해야 할 경우의 문제는 대상자를 단회 조사하게 되는 거임. 그러한 경우 긴 기간 동안의 평균값을 참값으로 구함.
- 기록은 과거의 노출에 대한 내용을 포함함. (고용, 의료, 생체징후)
- health outcome의 측정 오류들의 예
 - 담석증 또는 자궁내막증의 유병률은 실제 질환을 가진 사람들보다 작게 측정될 수 있음. 이는 이 질환을 가진 많은 사람들이 무증상이므로 검사를 받지 않기 때문임.
 - 진단 기술이 비슷한 병태 생리적 과정에 있는 많은 사람들의 sub-group을 구별하지 못할 수도 있음. 관절염을 가진 모든 사람들 중 류마티스성 관절염과 같은 특별한 타입을 확정하지 못하는 경우가 그러한 예임.
 - 진단법이 특정 개인에게 적합한 방법이라 하더라도, 그것이 인구 전체에 적용 가능한 확실하지 않음. 예를들어 pneumococcal 백신의 효과를 보기 위해 medicaid 수혜자들을 컴퓨터로 추출하여 백신 접종자와 비접종자간의 폐렴 발생률을 비교함. 연구자들은 pneumococcal pneumonia에 대한 연구를 하고 있으나, 컴퓨터는 pneumonia 환자들을 추출함. pneumococcal pneumonia는 전형적 폐렴의 1/3밖에 되지 않으므로 이것은 연구의 잘못된 분류를 야기함.

2. Assessing measurement error

- 측정(measure)이란 광범위하게 어떤 방법으로든 연구 대상자의 어떤 특징에 대한 데이터를 포착하는 것임. 자료 수집 방법은 질문, 인터뷰, 진찰, 실험, 의료 기록으로부터의 추출, 행동의 직접 관찰, 또는 다른 접근들.
- 특징이나 자료 수집 방법과 관계없이 각각의 연구 대상자들의 특징에는 참값이 있음. 그러나 각 대상자의 참값은 종종 알려지지 않음.
- 측정 오류의 결과는 그 양을 측정하는 것이 필요함. 측정 오류의 양적 지표들은 연구 디자인 단계에서 같은 특징에 대한 다른 측정법들 사이에서 유용함.

- 이 부분에서는 측정 오류의 지표들을 소개함.
 - 신뢰도(Reliability): 좋은 측정법은 여러 번 반복했을 때 같은 값이 나와야 함.
 - 타당도(Validity): 좋은 측정법은 정확한 값을 산출하는 것.
- 신뢰도와 타당도는 어떤 측정 방법에도 광범위하게 적용됨. 그러나 그것들의 양을 측정하는 데 접근하는 것은 측정의 크기에 의존함. 역학자들이 가장 관심있는 것이 그것임.

1. Continuous: 측정은 분율을 포함한 분수 값을 취함. 예를들면 몸무게는 거의 연속적인 숫자들인데, 일정 단위로 나타난 것임. 실제적인 이유에서 지속적 측정은 정해진 숫자를 기록하게 되고, 이 선택은 중재적인 것이지 측정의 내부적 문제가 아님.

2. Categorical

- 순위척도(ordinal): 측정된 특징이 세 개 혹은 네 개의 불연속 값일 수 있음. 예를 들면 질병의 정도를 mild, moderate, severe로 나누는 것.
- 명목척도(nominal): 둘 혹은 그 이상의 값이 가능하나, 그들 중 특별한 순서는 없음. 예를 들면 결혼상태에 관한 것인데, 기혼, 미혼, 이혼, 별거와 같은 것. 종종 단 두가지 값만 가능한 경우가 있음. 남성 여성과 같은 성별.

- 범주화 측정(categorical measure)에서 측정 오류는 보통 misclassification임.
- 측정 단위 간격이 '세밀한(fine)'것에서 '넓은(coarse)'것이 있을 수 있다. fine 측정은 coarse 값 하나로 묶을 수 있음. 예를들면 나이와 같이 연속적인 ordinal한 값을 몇 개의 연령 그룹으로 나눌 수 있음.

가. Reliability

- 앞서 말했듯, 신뢰도는 여러번 반복했을 경우의 동의를 내포함. 신뢰도는 역시 재생산이란 말과 같음. 덧붙이자면 얼마나 반복했는가와 관계가 있음.
- Intra-observer 신뢰도는 한 사람이 여러번 측정 했을 때의 신뢰도를 말함. Inter-observer 신뢰도는 여러 사람이 같은 방법으로 측정했을 때의 신뢰도를 말함. Test-retest란 같은 대상사를 다른 시간에 측정했을 때의 동의를 말함.

(1) 일치 Concordance(percent agreement)

- 일치란 categorical measure임. 어떤 categorical measurement는 여러 대상자를 비슷한 환경에서 두 번 측정함. Concordance는 단순히 테스트받은 대상자 전원에 대하여 두 번에서 같은 값이 나온 비율임.
- Concordance의 중요한 한계: 우연에 의한 일치를 구할 수 없음.

Table 10-1. Data Layout for Reliability Assessment on a Binary Measure

MEASUREMENT NO. 1	Measurement No. 2		TOTAL
	+	-	
+	<i>a</i>	<i>b</i>	<i>a + b</i>
-	<i>c</i>	<i>d</i>	<i>c + d</i>
Total	<i>a + c</i>	<i>b + d</i>	<i>N</i>

Table 10-2. Expected Results If Two Observers Each Assign a Positive Result to a Random 10% of Study Subjects

OBSERVER NO. 1	Observer No. 2	
	+	-
+	1	9
-	9	81

○ 일치도 = $a+d/N$, 표 10-2의 일치도는 $90/100=90\%$ 임.

〈예 1〉

		의사 1	
		양성	음성
의사 2	양성	1	3
	음성	2	94

〈예 2〉

		의사 1	
		양성	음성
의사 2	양성	43	3
	음성	2	52

그림 10-7. 일치도가 95%인 예

○ 예 1, 예2 모두 일치도는 95%임.

(2) 카파지수 Kappa

○ 카파지수(κ)는 우연에 의한 일치를 보정한 것.
 카파수식 (관찰일치도-기대일치도)/(1-기대일치도)

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

P_o 는 관찰된 일치율이고, P_e 는 우연에 의해 기대되는 일치도임.

○ 표 10-1에서 카파 일치도와 기대일치도 수식을 구하면 다음과 같다.

$$P_o = \frac{a + d}{N}$$

$$P_e = \left[\frac{(a + b)(a + c)}{N} + \frac{(b + d)(c + d)}{N} \right] / N$$

		의사 1							
		양성	음성	계					
의사 2	양성	0.12	3.88	4	의사 2	양성	20.7	25.3	46
	음성	2.88	93.12	96		음성	24.3	29.7	54
		3	97	100			45	55	100

그림 10-6. 우연에 의한 기대 일치도 예

- 예1의 카파지수 = $(95 - 93.24) / (1 - 0.9324) = 0.26$
- 예2의 카파지수 = $(95 - 50.4) / (1 - 0.5040) = 0.9$
- 카파값은 보통 0에서 1까지 있음. $P_o < P_e$ 인 경우 카파값은 음수가 되고 이것은 우연에 의한 일치보다 더 상관성이 없음을 나타냄.

Table 10-3. Guidelines for Interpretation of Kappa

KAPPA	INTERPRETATION
>.80	Almost perfect
.61-.80	Substantial
.41-.60	Moderate
.21-.40	Fair
.00-.20	Slight
<.00	Poor

[Source: Landis and Koch (1977).]

- Example: Wright와 그의 동료들은 human papilloma virus(HPV)를 자궁경부암에 대한 screening test로 삼았다. 그들이 채취한 것과 임상 의사들이 채취한 것의 결과는 다음과 같다.

CLINICIAN-COLLECTED	Self-collected		TOTAL
	+	-	
+	170	132	302
-	128	985	1113
Total	298	1117	1415

관찰일치도는 $(170+985/1415)=0.816$

기대일치도는 $(302 \times 298/1415 + 1113 \times 1117/1415)/1415=0.666$

카파지수는 $(0.816-0.666)/(1-0.666)=0.45$

(3) Intraclass correlation coefficient

○ 연속 측정법에서 신뢰도는 변수분석에 기초한 ICC로 측정할 수 있음. ICC에는 variance model의 다른 분석에 따른 몇가지 종류가 있음. 이러한 모델들은 각각의 대상자에 대한 반복적 관찰이 상호교환 할 수 있는나에 따라 다름.

○ 가장 간단한 상황은 두 독립된 관찰자가 각각의 대상자를 관찰하고, 이 두 관찰자를 상호교환할 수 있다고 보는 것임. 염분 섭취 연구를 예로들면, 각각의 50명의 대상자에게서 소변 검체를 얻음. Urine의 sodium 분석의 신뢰도를 얻기 위하여 두 개의 샘플을 각 사람의 원래 검체에서 얻어, 총 100개의 결과를 분석함. 각 사람은 #1, #2 라벨된 샘플이 있고, 이것들은 상호교환 할 수 있음.

○ ICC를 계산하는 법. 위의 소변- 염분 관계에서 보듯, 이 100개의 변이들 사이에는 두가지 요소가 있음. 하나는 사람간 변이(between-person B)이고, 다른 하나는 사람내 변이(within-person W)임. 그러므로 변이의 합은 B + W이고, ICC의 식은 다음과 같음

$$ICC = B / B+W$$

다르게 말하면 ICC는 총변이의 합에 대한 사람간의 변이의 구성비임.

○ ICC가 0이라면 B=0라는 의미이고, 이것은 사람간의 변이가 없이 50명 전원이 같음을 의미함.

나. 타당도 (Validity)

○ 타당도란 측정 결과가 사실을 반영하느냐를 내포함. 측정의 타당도를 평가하기 위해서는 각 대상자들의 실제 상태를 알아야 함. 이것은 이미 확립되어 타당도가 인정된 다른 방법으로 측정해야 함. 그것을 gold standard라함.

○ 역학에서는 실제 관심 대상의 특정 사실은 종종 두가지인데, 질환 상태의 존재 또는 부재와 같은 것임. 예를들어 알츠하이머 질환의 gold standard는 뇌조직의 현미경적 소견의 존재 유무와 같은 것을 말함. 그러나 모든 사람에게 gold standard를 적용하는 것은 위험하고, 시간과 비용이 들기도 함. 이러한 것이 가능한 다른 측정법을 찾게 함.

(1) 민감도와 특이도(Sensitivity and specificity)

○ 대상자들에게 gold standard를 적용하여 질환이 있는지 없는지를 조사하면, 어떤 대상자는 질환이 있고, 어떤 사람들은 없을 것임. 이제는 새로운 평가 방법으로 질환의 유무를 평가함. Gold standard와 새로운 방법의 결과의 차이가 표 10-4에 요약되어 있음.

a = number of true positives
 b = number of false positives
 c = number of false negatives
 d = number of true negatives

Table 10-4. Data Layout for Assessing the Validity of a Binary Test

TEST RESULT	Condition Present?	
	+	-
+	a	b
-	c	d
Total	$a + c$	$b + d$

Sensitivity = $a / (a + c)$
 Specificity = $d / (b + d)$

- 타당도의 두가지 요소의 정의
 - 민감도(Sensitivity): 실제로 질환이 있을 경우 평가방법이 얼마나 그것을 찾을 것인가? 실제 case는 $a+c$ 이고 평가에서 양성으로 나온 것은 a 이므로 민감도는 $a/(a+c)$ 로 측정됨.
 - 특이도(Specificity): 실제로 질환이 없을 경우 평가가 얼마나 음성의 결과를 낼 것인가? 실제 case가 아닌 경우는 $b+d$ 이고, 평가에서 음성 결과는 d 이므로 특이도는 $d/(b+d)$ 로 측정됨.
- 민감도와 특이도의 범위는 0에서 1까지임.
- 민감도와 특이도가 함께 높은 것이 이상적이지만, 둘 다 고립시켜 생각할 수 없음. 실제로 모든 평가들은 불완전하므로 민감도와 특이도는 0과 1 사이에 존재함.

■ Example: Catanzaro와 그의 동료들은 Mycobacterium tuberculosis Direct test (E-MTD)라 불리는 활동성 결핵의 새로운 진단법을 증명하려 했다. 이 방법은 TB bacillus의 유전 물질을 발견하는 것이다. 7개의 의료기관에서 결핵으로 의심되는 338명의 환자들에게 이 검사가 시행되었다. 이 환자들은 임상적 표준지침에 따라 광범위한 검사를 받았다. 이들 중 72명이 활동성 결핵으로 결정되었고, E-MTD 검사에서는 60명이 양성으로 나왔으므로 민감도는 $60/72=0.833$ 이다. 266명의 환자가 결핵이 아닌 것으로 나타났고, E-MTD 검사에서는 259명이 음성이었으므로 특이도는 $259/266=0.974$ 이다.

- 민감도와 특이도는 여러 평가들의 수행능력을 비교하는데 유용함. 예를들어 결핵 검사에서 AFB는 민감도가 0.597이고, 특이도가 0.917로 둘 다 E-MTD보다 낮음. E-MTD test에서 AFB보다 더 정확한 결과를 얻을 수 있음.
- 민감도와 특이도는 둘 다 유병률에 의존하지 않음. 그러나 실제로는 민감도와 특이도는 대상인구에 따라 다양하게 나타남. Catanzard의 연구에서 임상에서 처음 TB로 의심된 환자 중 TB 가능성이 낮다고 의심한 환자의 민감도는 $5/12=0.42$ 였고, TB 가능성이 높다고 의심한 경우 $33/40=0.83$ 이었음. 이것은 아마도 TB균이 많아 더 잘 검출된 것이라 추측됨. 민감도와 특이도는 어떤 대상인구 집단에 어떤 방법으로 평가를 수행했느냐를 반영했을 경우 가장 잘 해석됨.

(2) 수용자 작용 특성 곡선 Receiver operating characteristic (ROC) curves

○ 역학에서 관심 대상이 되는 특징은 종종 두가지로 나타낼 경우가 많지만, 그것에 대한 평가나 측정은 순위적이거나 연속적인 결과를 포함할 경우가 많음. 이럴 경우 타당도는 어떻게 증명할 것인가?

■ Example: Buchsbaum은 외래 환자들중 알콜 남용이 있는 환자들을 찾기 위해 네가지 질문을 담은 CAGE라는 도구를 이용했다. CAGE의 질문은 “예”, “아니오”로만 대답할 수 있게 되어 있고, “예”라는 대답은 알콜에 대한 문제를 가지고 있음을 포함하고 있다. 예를들면 “당신은 술을 끊어야겠다고 느낀 적이 있습니까?”라는 질문이다. CAGE score는 “예”라고 대답한 횟수로 정의되고, 범위는 0에서 4까지 있다.

Table 10-5. CAGE Questionnaire Scores in Patients With and Without True Alcohol Abuse or Dependence

CAGE score	True Alcohol Abuse or Dependence	
	PRESENT	ABSENT
4	56	1
3	74	10
2	86	34
1	45	54
0	33	428
Total	294	527

[Source: Buchsbaum et al. (1991).]

표 10-5에서 보여주는 것은 알콜남용이나 중독이 있는 환자들이 그렇지 않은 사람들보다 일반적으로 CAGE 점수가 높다는 것이다. 그러나 CAGE 점수의 두가지 분포가 겹쳐져 있다.

- 이러한 평가의 performance를 측정하는 한가지 방법은 분류기준값(cutoff value)을 정하는 것임. 결과가 분류기준값을 넘을 경우 양성으로 판정하고, 분류기준값의 아래에 있을 경우 음성으로 판정함. 민감도와 특이도는 그 분류기준을 이용하여 계산할 수 있음.
- 표 10-6에서 1과 2 사이를 분류 기준값으로 삼았다면 2, 3, 4는 양성이고, 0 과 1은 음성일 것임. 이 때 CAGE의 민감도는 $(56+74+86)/294=0.735$ 이고, 특이도는 $(54+428)/527=0.915$ 임.
- 민감도와 특이도는 어떤 값을 분류기준값으로 삼았느냐에 따라 다름. 표 10-6에서 보듯 분류기준값이 달라짐에 따라 CAGE의 민감도와 특이도가 달라짐.

Table 10-6. Sensitivity and Specificity of CAGE Questionnaire Score at Six Possible Cutoff Values

CAGE SCORE	True Alcohol Abuse or Dependence		CUTOFF LOCATION	SENSITIVITY	SPECIFICITY
	PRESENT	ABSENT			
4	56	1	←	0.000	1.000
3	74	10	←	0.190	0.998
2	86	34	←	0.442	0.979
1	45	54	←	0.735	0.915
0	33	428	←	0.888	0.812
Total	294	527			

○ ROC 그래프는 가능한 모든 분류기준값에 대한 민감도와 특이도를 보여줌. 그림 10-2는 ROC곡선은 보여주는데 세로축은 민감도, 가로축은 1-특이도임.

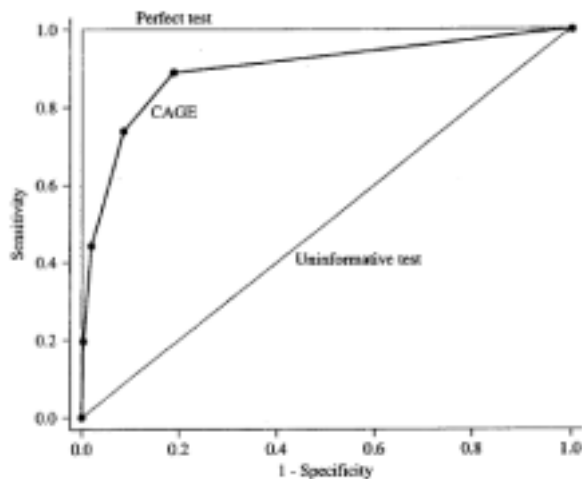


Figure 10-2. A Receiver Operating Characteristic (ROC) Plot.

- 가운데의 두꺼운 선은 CAGE 점수에서 가능한 모든 분류기준값에 대한 민감도와 특이도를 표시한 것임. 이것은 표10-6을 그림으로 나타낸 것임.
- "uninformative test"라고 적힌 선은 gold standard에서 양성(+)과 음성(-)을 구별할 수 없는 경우임. 이 선에서 모든 민감도는 특이도를 잃는 정도와 같이 커지므로 평가로서 의미가 없음.
- "perfect test"라고 적힌 선은 양성(+)과 음성(-)을 정확하게 구별함. 이 선은 민감도와 특이도 모두 1임을 의미함.
- 일반적으로 곡선이 왼쪽 모서리에 가까울수록 정확한 평가방법임을 의미함.
- AUC(area under the curve)는 평가의 정확도의 한가지 측정임. Uninformative test의 AUC=0.5이며, perfect test의 AUC=1.0 임. 그림10-2의 AUC=0.89임. AUC가 1에 가까울수록 좋은 평가방법임.

3. Consequences of measurement error

- 역학 연구에는 결과에 대한 측정오류의 영향이 있을 수 있음. 만약 질환의 유무에 의해 노출에 대한 확인에 영향이 있었다면 결과는 양성적으로나 음성적으로 바이어스가 있을 수 있음. 예를들면 아픈사람은 이전의 노출에 대해 더 잘 기억할 수 있음. 이럴 경우 노출과 질환 사이에 실제로는 그렇지 않은데 양의 관계가 있는 것으로 나타남.
- 또 하나의 예를 들면 제2형 당뇨 환자들의 몸무게의 경우, 실제로 당뇨가 발생한 기간동안의 몸무게가 아닐 수도 있음. 실제로 식사 조절을 통해 좀 더 낮게 나왔을 가능성이 있음. 2형 당뇨환자들의 몸무게를 비교하기 위해서는 아직 식사조절을 하지 않았거나 짧은 기간 밖에 되지 않은 사람들을 대상으로 해야함.
- 마찬가지로 질병 상태에 대한 다른 잘못된 분류들은 노출에 대하여 높거나 낮게 나올 수 있음. 발생률에 대한 잘못된 분류는 연관성의 크기를 크게 할 수 있음.
- 노출에 대한 측정의 비선택적 오류는 대상자의 노출 상태가 아픈사람과 그렇지 않은 사람간에 빈도나 정도에 있어서 비슷할 때 일어남.

■ Example: 노출에 대한 두가지에 대하여 완전히 분류되었다고 가정한다면 다음과 같은 표를 얻을 수 있다.

EXPOSURE	CASES	CONTROLS	ODDS RATIO
Yes	150	75	$\frac{150}{150} + \frac{75}{225} = 3.0$
No	150	225	
	300	300	

그러나 노출된 사람중 3분의 1이 잘못 분류된 것이라면 결과는 다음과 같다.

EXPOSURE	CASES	CONTROLS	ODDS RATIO
Yes	150 - 50*	75 - 25*	$\frac{100}{200} + \frac{50}{250} = 2.5$
No	150 + 50	225 + 25	
	300	300	

*150 × 1/3 = 50, 75 × 1/3 = 25

관찰된 odds ratio가 1에 가까울수록 의미가 없는 것이다.

실제로 노출되지 않은 사람의 20%가 잘못 분류되었다면 다음과 같은 결과를 얻는다.

EXPOSURE	CASES	CONTROLS	ODDS RATIO
Yes	150 - 50 + 30	75 - 25 + 45	$\frac{130}{170} \div \frac{95}{205} = 1.65$
No	150 + 50 - 30*	225 + 25 - 45	
	300	300	

*150 x 2 = 30, 225 x 2 = 45

결과는 무효에 더 가까워졌으나 연구의 결과는 더 영향력있게 되었다.

- 노출된 사람들 중에 노출되지 않았다고 분류되는 사람들이 있음. 이들 때문에 민감도는 100% 보다 작게 됨.
- 노출되지 않은 사람들 중에 노출되었다고 분류되는 사람들이 있음. 이들 때문에 특이도가 100%보다 작게 됨.

■ Example: 일년동안 식품 A에 대한 소비와 질병발생에 대하여 관찰하였다. 질병에 대한 진단에는 오류가 없다고 가정한다면 식품A를 소비한 사람들에서 10배나 더 질병이 발생했다. 다음 표와 같다.

The truth:

CONSUMPTION OF FOOD A	Illness			RELATIVE RISK
	YES	NO	TOTAL	
Yes	10	990	1000	10
No	10	9990	10,000	

만약 식품A를 구매한 사람들의 80%만이 이것을 먹었다면 민감도는 80%가 되고, 식품A를 먹은 사람들은 모두 제대로 분류되었다면 특이도는 100%가 된다. 이 결과를 나타내면 다음과 같다.

Exposure status ascertained with 80% sensitivity and 100% specificity:

CONSUMPTION OF FOOD A	Illness			RELATIVE RISK
	YES	NO	TOTAL	
Yes	10(.8) = 8	990(.8) = 792	800	8.5
No	10 + 2 = 12	9990 + 198 = 10,188	10,200	

반대로 A를 먹은 모든 사람은 잘 분류되었다면 민감도는 100%이고, A를 먹지 않은 사람의 20%가 잘못 분류된 것이라면 특이도가 80%이다. 이때의 결과는 다음과 같다.

Exposure status ascertained with 100% sensitivity and 80% specificity:

CONSUMPTION OF FOOD A	Illness		TOTAL	RELATIVE RISK
	YES	NO		
Yes	$10 + 2 = 12$	$990 + 1998 = 2988$	3000	4.0
No	$10(.8) = 8$	$9990(.8) = 7992$	8000	

이 경우에 민감도가 완전하지 못한 것보다 특이도가 완전하지 못할 때 대상자의 절대적인 숫자가 잘못 분류됨을 볼 수 있다.

- Example: 식품 A의 소비와 질병의 관계에서 이번에는 노출에 대한 것이 아닌 질병의 상태에 대한 잘못된 분류를 가정해보자. 다시 잘못 분류된 것이 하나도 없을 때를 보면 아까와 같다.

The truth:

CONSUMPTION OF FOOD A	Illness		TOTAL	RELATIVE RISK
	YES	NO		
Yes	10	990	1000	10
No	10	9990	10,000	

만약 아픈 사람의 10%가 다음과 같지 않다면 민감도는 90%일 것이고, 결과는 다음과 같다.

Illness status ascertained with 90% sensitivity and 100% specificity:

CONSUMPTION OF FOOD A	Illness		TOTAL	RELATIVE RISK
	YES	NO		
Yes	$10(0.9) = 9$	$990 + 1 = 991$	1000	10
No	$10(.9) = 9$	$9990 + 1 = 9991$	10,000	

반대로 100명 중 한명이 질환이 있음에도 없다고 믿은 경우를 가정해보자. 결과는 다음과 같다.

Illness status ascertained with 100% sensitivity and 99% specificity:

CONSUMPTION OF FOOD A	Illness			RELATIVE RISK
	YES	NO	TOTAL	
Yes	10 + 10 = 20	990(.99) = 980	1000	1.82
No	10 + 100 = 110	9990(.99) = 9890	10,000	

관찰된 상대 위험도는 특이도가 완전하지 못할 경우 10에서 1.82가 되어 무효한 값에 가깝게 감소함을 알 수 있다.

- 측정 오류를 최소화하는 것은 모든 역학 연구에서 주관적임.
- 노출이나 질병 모두에 대해 우리는 우리가 만든 tool을 믿으려함. 연구가 신뢰도와 타당도를 얻는 것은 어떤 틀을 어느 연구에 사용하느냐에 달려있음.
- 연구자는 자신이 원하는 것이 분류하는 것에 강하게 작용하지 않도록 조심해야 함.

■ Example: Anal intercourse가 anal cancer의 소인이 되는 것을 조사하기 위하여 Daling은 case-control study를 하였다. 1974년부터 1979년까지 서부 워싱턴주에서 anal cancer로 진단되는 사람들을 매독 history가 있는 사람들과 비교했다. 이 조사에서 매독의 신환들은 대부분 최근에 다른 남자와 sexual contact을 한 history가 있는 것으로 알려졌다. Anal cancer가 있는 47명 중 8명(17%)이 매독이 있었다. 반면 다른 암에서는 1~2%만이 그러했다.

Daling은 이 분포가 결혼 상태에도 적용될 것이라 생각했고, 결과는 그의 이론을 뒷받침해 주었다. Anal cancer를 가진 환자의 24%가 결혼을 하지 않았고, rectal이나 colon cancer(대조군)에서는 8%만이 결혼을 하지 않았다.

그러나 등록을 기초로 하는 이런 연구는 잘못 분류될 가능성이 많다. 결혼하지 않은 남자의 소수만이 anal intercourse를 할 것이므로 그것이 실제 노출은 아님. 그러나 강한 상관성을 보였기 때문에 빠르게 개인 면담 조사로 들어갔고, 결과는 더욱 확실해졌다. 이전에 anal intercourse를 한 적이 있는 사람이 anal cancer환자 중에는 25.9%, 다른 암에서는 단지 1.6%만이 있었다.