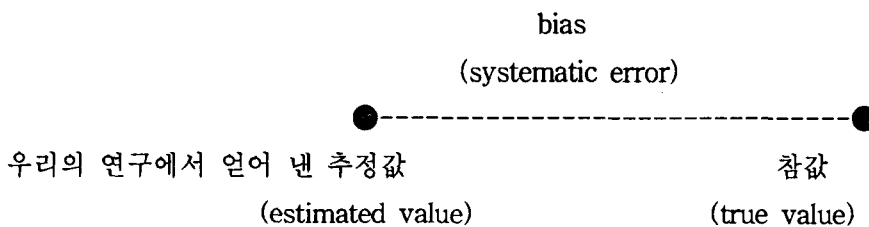


역학연구에서의 BIAS

이 원 철
가톨릭 의과대학

1. BIAS 의 정의

Bias란 진리인 참값(true value)과 우리가 연구하여 추정해 낸 값(estimated value)의 차이를 말한다 (그림 1).



(그림 1) Bias

즉 우리가 추정해 낸 값이 참값과 동일하다면 그 연구에는 bias가 전혀 없는 셈이고, 우리가 추정해 낸 값이 참값과 크게 다르게 되면 그 연구는 크나큰 bias를 지닌 셈이다. Bias가 큰 연구는 그 결과를 믿을 수 없게 되는 것이고 그 연구결과는 논문에 게

재될 가치가 없어지게 되며 반대로 bias가 없다면 그 연구는 매우 훌륭하게 수행된 것을 의미한다.

Bias는 다른 표현으로 하면 해당연구가 처음부터 끝까지 오류없이 행하여졌는지를 점검하는 것으로 볼 수 있다. 왜냐하면 연구 도중도중에 어떠한 오류가 생긴다면 그 오류로 인하여 우리의 추정값은 참값으로 부터 점차 멀어지게 되겠기 때문이다.

2. BIAS 의 종류

그렇다면 어떠한 오류들 때문에 bias가 발생하는가?

Bias 는 크게 아래 3가지로 구분하여 설명되고 있는데 선택 bias(selection bias), 정보 bias(information bias), 교란 bias(confounding bias) 이다.

2-1. 선택 Bias (Selection Bias)

선택 bias는 우리가 조사하고자 하는 전체 대상 인구중에서 조사를 위하여 뽑힌 (sampling 된) 조사대상자가 된 집단의 특성과, 조사에서 제외된 집단의 특성이 다를 때 구조적으로 다를 때 나타나는 오차(error)로 설명할 수도 있다. 쉽게 이야기 하자면 우리가 대상으로 하고 있는 집단은 평균연령이 45세인데, 조사를 위하여 우리 손안에 들어있는 집단의 평균연령이 35세이고 조사에서 제외된 집단의 평균연령이 55세 이라고 하면, 어떤 목적의 조사이던지 간에 우리 손안의 대상자들만으로 전체집단의 참값을 추정하겠다는 것은 매우 힘든 일이 된다. 다른 예를 들면 우리가 당뇨병을 대상으로 연구하고자 할 경우에, 만약 병원에 입원해 있는 당뇨병 환자만을 대상으로 하여 연구를 수행한다고 하면, 당뇨병을 일으키자 마자 사망하여 병원에 들리지도 못한 사람은 연구대상에 포함되지 못하고, 질병이 경하여서 병원에 입원할 필요가 없던 사람도 연구대상에서 포함되지 못하며 병원에 올만한 경제적인 여유가 없기 때문에 집에서 앓고 있는 사람도 대상에 포함되지 못한다. 즉 우리가 대상으로 하고 있는 병원에 입원한 환자들은 전체의 그 질병(당뇨병)을 대표할 수 있지 못하게 된다.

앞에서 정의한 대로 표현하자면 병원에 입원한 환자가 나타내는 값과 그 외의 그 질

환을 앓고 있는 집단이 나타내는 값은 다를 수 있다. 이러한 것을 선택 bias라고 하는데 bias 때문에 우리는 우리가 조사한 결과를 바탕으로 일반화(generalization)를 할 수 없게 되는 것이다.

Vinyl chloride Monomer 폭로와 태아사망간의 관련성에 관한 연구를 예로 들어 선택 bias 를 설명하면 다음과 같다.

여기서는 설명을 단순화하기 위하여 cross-sectional 한 design 으로 자료를 구하였다고 가정하자.

한 공장내에 10,000명의 근로자들이 근무하고 있다고 할 때 VCM 폭로와 태아사망에 관하여 2 x 2 표로 작성하면 다음 4가지로 구성된다.

- 즉, VCM에 폭로되고 있고 태아사망을 경험한 군 - A
- VCM에 폭로되고 있으나 태아사망은 경험한 적이 없는 군 - B
- VCM에 폭로된 바는 없으나 태아사망을 경험한 군 - C
- VCM에 폭로된 바도 없고 태아사망도 경험한 적이 없는 군 - D

		VCM 폭로 여부		
		폭로	비폭로	
태아사망	유	A 4,000	C 2,000	
	무	B 1,000	D 3,000	
				10,000

(표 1)

우리가 모집단 전체에 대하여 조사하였다면 표1과 같은 결과가 나왔을 것이다. (실제 VCM과 태아사망을 조사하게 되면 결과가 이렇다는 것은 아니고 이러한 분포를 보이는

것으로 가정하자) 그러면 VCM 과 태아사망의 관련성의 정도를 나타내는 지수인 odds ratio는 AD/BC 로서 $4,000 \times 3,000 / 1,000 \times 2,000 = 6$ 을 나타내게 된다.

그런데 전체대상인 10,000명을 조사하기가 불가능하였기 때문에 이중에서 1,000명을 sampling 하여 조사하였다고 하자. 그러면 1,000명을 sampling 한 후 위와 같은 4가지 가능성대로 구분하였다고 하면 당연히 표 2와 같이 분류되어 이에 대한 odds ratio 역시 $400 \times 300 / 100 \times 200 = 6$ 을 나타내어 참값 (모집단에서의 값) 과 추정값(sample한 집단에서의 값)이 동일하며 bias가 없는 결과를 나타내어야 한다.

		VCM 폭로 여부		
		폭로	비폭로	
태아 사망	유	A 400	C 200	
	무	B 100	D 300	
				1,000

(표 2)

		VCM 폭로 여부		
		폭로	비폭로	
태아 사망	유	A 200	C 200	
	무	B 300	D 300	
				1,000

(표 3)

그러나 sampling 된 1,000명을 4가지 가능성대로 구분하였더니 표 2 와는 다르게 표 3처럼 나타나 odds ratio = $200 \times 300 / 300 \times 200 = 1$ 을 나타내었다면 이는 참값이 6임에도 불구하고 추정치는 1을 나타내어 bias가 생긴 것이다. 이 결과대로 그대로 결론을 지으면 (추정된 odds ratio가 1이기 때문에) 'VCM과 태아사망은 관련성이 없다'라고 결론을 내리게 되겠으나 이는 틀린 결론이 된다 (왜냐하면 이 예에서의 실제 odds ratio는 6이기 때문이다).

이러한 bias가 일어난 이유는 무엇인가 하면 표1에서 표 2로 표본추출되는 경우 4가지 cell이 모두 동일한 확률로 추출된 반면 (이렇게 sampling 되어야 sampling이 잘 된 것이다). 표1에서 표3으로 표본추출된 경우에는 각 cell 의 samplinge 될 확률이 다르게 나타나 (즉 A cell 에서는 4,000명분에 200으로 2/40의 확율이고, B에서는 1,000명의 300으로 3/10, C에서는 200/2,000으로 1/10, D에서는 300/3,000으로 1/10을 나타내

어) 일부에서는 과도히 sampling 되고 일부에서는 너무 적게 sampling 되는 바람에 엉뚱한 결과를 초래하게 된 것이다. 이러한 현상을 selection bias라고 부른다.

2-2. 정보 bias

정보 bias 는 비교하고자 하는 두군에서 각각 그 폭로(exposure)나 결과(outcome)를 측정할 때, 측정상의 문제로 인하여 양군에서 측정의 정확도(accuracy)의 차이가 생기는 것을 말한다. 이 정보 bias는 관찰 bias 또는 분류 bias 라고도 불린다. 위에서 예를 든 VCM과 태아사망에 관한 연구를 수행한다고 하면 연구방법에 따라서 정보 bias가 발생할 수 있다. 즉 태아사망을 의사가 진단한 것에 의하여 결정하였다고 하면 문제가 없겠으나 이 연구에서 태아사망의 결정방법을 해당 산모가 '태아사망을 경험하였다, 경험하지 않았다'에 관하여 답변한 것을 근거로 하여 태아사망을 측정한다면 분류 오류가 발생할 가능성은 크다.

이러한 분류오류의 또다른 예로 기억 bias(recall bias)를 들 수 있다.

예를 들어 이를 설명하면 다음과 같다. 선천성 기형을 연구하기 위한 환자-대조군 연구에서는 기형인 아기를 출산한 산모가 환자군이 되고, 정상아를 출산한 산모가 대조군이 될 것이다.

그런데 이 상황에서 정상아의 산모보다 기형아의 산모가 예전 기억을 더 소상하게 기억하여 낸다고 하면 위에 설명한 차별 분류오류(이러한 상황에서는 이를 recall bias라고 부른다)가 발생할 수 있다.

즉, 기형아를 낳았다고 하는 그 사실이, 그 대상 산모로 하여금 임신당시의 모든 일들을 소상하게 되돌아 보게끔 할 것이고 따라서 정상아의 산모보다 약을 복용한 사실이라던가, 충격을 받았던가, 전염성질환을 경험했다던가 하는 것의 기억들을 더 정확하게 기억해 낼 가능성이 크게 될 것이다. 따라서 이를 토대로 분석하여 보면, 실제로 생물학적인 관련성 때문에 유의한 결과가 나타나는 것이 아니고 더 정확하게 대답했다는 이유 때문에 통계적으로 유의한 결과가 나타날 가능성이 있다는 말이다.

위에 열거한 예는 모두 차별 분류오류(differential misclassification)에 대한 것이다. 즉 비교하고자 하는 두 군에서의 분류오류의 정도가 다르기 때문에 차별(differential)이라는 표현을 사용하였다. 이러한 차별 분류오류는 우리가 분석하여 낸 추정값을 참값보다 크게 추정하게 할 수도 있고, 참값보다 적게 추정하게 할 수도 있다. 위에 든

예는 참값보다 크게 추정한 경우들이 되겠으나 이와 반대되는 예를 찾는것은 그리 어렵지 않다.

지금까지 설명한 차별 분류오류와는 다른 종류의 무차별 분류오류 (nondifferential misclassification)가 있는데 이는 비교하고자 하는 두군에서 동일한 비율로 분류오류가 나타난 경우를 말한다. 예를 들면 다음과 같다.

음주군에서의 후두암(laryngeal cancer)과 비음주군에서의 후두암의 발생율을 비교하려고 한다. 1,000명의 음주자와 1,000명의 비음주자를 대상으로 하였을때 실제로는 음주자 1,000명중 50명이 발생하고, 비음주자 1,000명중 10명이 발생한다고 하면 비교위험도 (의 참값)는 5이다.

그런데 음주자 1,000명중 500명만이 음주한다고 대답하였기 때문에 나머지 500명이 비음주자로 구분되었고, 비음주자군 1,000명에서도 500명만이 비음주자로 대답하고 나머지 500명이 음주자로 잘못 대답하였다고 하면 실제 상황과는 다르게 엉뚱한 분류가 생기게 된다.

이 때문에 음주군에서는 (음주군 500명중의 25명과 비음주군의 5명이 발생하여) 30명이 발생하였고, 비음주군으로 구분된 1,000명에서는 (음주군의 25명과 비음주군의 5명이 합하여) 30명이 발생하였다. 이러한 상황에서(잘못 추정된) 비교위험도는 1이 되고 만다.

위의 예는 비록 자연스럽게 못한 가상적인 예이긴 하지만 실제로는 5의 비교위험도를 지니는 관련성이, 정보 bias로 인하여 비교위험도가 1로 추정되는 오류를 범하게 된다는 것을 잘 설명하여 주고 있다.

여기서 강조하여 설명하고자 하는 점은, 차별 분류오류는 bias의 방향을 예측하기가 어려우나 무차별 분류오류는 bias의 방향이 항상 일정하다는 점이다. 즉 자료에서 분석하여 낸 추정치가 4이라고 하면 차별 분류오류가 있는 경우에는 참값이 이보다 클 수도 있고(예를 들어 5 또는 6) 작을 수도 있으나(예를 들어 2 또는 3), 무차별 분류오류는 항상 이보다 큰 곳에 위치하게 되므로(예를 들면 참값이 2 또는 3이 아니고 5 또는 6에 있음을 의미한다). 만약 추정치 4가 통계적으로 유의하다면 참값 역시 유의한 결과를 보일 것으로 생각할 수 있다. 이를 다르게 표현하면 무차별 분류오류에서는 추정값이 참값보다 항상 관련성이 없는 방향으로 bias가 일어난다 (toward the null bias)고 표현할 수 있다 (이에 관하여 한가지 추가로 더 설명하자면 odds ratio가 0.3으로 추정되었다면 참값은 0.2 또는 0.1 에 위치한다는 의미이다).

2-3. 교란 BIAS (Confounding Bias)

세번째로 구분되는 bias는 교란 bias(confounding bias)이다. 교란 bias를 한마디로 설명하자면 효과가 섞이는 것(mixed up)을 의미한다.

여기서는 이러한 교란 bias 를 설명할 때 가장 흔히 쓰이는 예로서 음주와 폐암의 관련성에 대한 연구를 그 예로 들고자 한다. 음주가 폐암을 일으키는지를 알아보려고 연구하였더니 연관이 있는 것으로 나타나 ‘음주와 폐암은 관련성이 있다’하고 결론을 지어 보고하였다고 하자. 그러나 이것은 참값(true value)이 아니었다. 참값은 음주와 폐암과는 관련이 없는 것이 참값이었다. 그렇다면 왜 이러한 bias가 생겼을까? (우리가 추정하여 낸 값은 관련이 있다는 것이고 참값은 관련이 없다는 것이니 여기에 bias가 존재한다).

그 이유는 우리가 보고자 하였던 음주와 폐암과의 관련성 연구에 흡연의 효과가 섞여 나타났기 때문이다. 흡연은 폐암의 원인이고 흡연하는 사람은 주로 음주를 하기 때문에 이러한 결과가 나타난 것이다. 그렇다면 이러한 종류의 bias를 없애는 방법은 섞여진 효과를 제거하고 순수한 효과만을 추정하고자 하는 것이 된다. 즉 흡연의 효과를 제거한 후의 음주가 폐암에 미치는 영향을 조사하여야 한다는 말이다.

이러한 설명을(가상적인) 표로 나타낸 것이 표 4이다.

표 4. 음주와 폐암과의 관련성을 평가하는데 있어서 흡연이 confounderr(교란인자)로 작용하는 것을 도식화한 예

* 대상자 전원의 경우

음 주	폐 암 발 생			전 체	폐암발생율
	발생함	발생하지 않음			
음 주 군	60	940		1,000	60/1,000
비음주군	40	960		1,000	40/1,000
전 체	100	1,900		2,000	50/1,000

$$\text{비교위험도(rate ratio)} = 60/40 = 1.5$$

* 비흡연군만을 따로 구분하여 분석한 경우

폐 암 발 생				폐암발생율
음 주	발생함	발생하지 않음	전 체	
음 주 군	4	496	400	10/1,000
비음주군	6	594	600	10/1,000
전 체	10	990	1,000	10/1,000

비교위험도(rate ratio) = 10/10 = 1.0

* 흡연군만을 따로 구분하여 분석한 경우

폐 암 발 생				폐암발생율
음 주	발생함	발생하지 않음	전 체	
음 주 군	56	544	600	93/1,000
비음주군	34	366	400	85/1,000
전 체	90	910	1,000	90/1,000

비교위험도(rate ratio) = 93/85 = 1.1

표에서 보는 바와 같이 모든 대상자를 조사하였을 때 음주군에게서의 폐암발생율(60/1,000)은 비음주군에서의 폐암발생율(40/1,000)보다 50%가 더 높게 나타났다. 그러나 대상자들(2,000명)을 흡연군(1,000명)과 비흡연군(1,000명)으로 구분한 후 각 군에서의 음주군과 비음주군에서의 폐암발생율을 보면 실제로 거의 차이가 없는 것을 볼 수 있다. (비흡연군에서는 음주군의 폐암발생율이 10/1,000이고 비음주군이 10/1,000이어 동일하며, 흡연군에서는 음주군의 폐암발생율이 93/1,000이고 비음주군의 폐암발생율이 85/1,000이어서 거의 비슷하다)

이러한 결과는 비음주군의 흡연율(40/1,000)보다 음주군의 흡연율(60/1,000)이 더 높았기 때문이며 동시에 폐암이 비흡연군에서 보다 흡연군에서 더 많이 발생(90/1,000 대 10/1,000)하였기 때문이다.

이러한 교란 bias를 일으키는 인자를 교란인자,(confounding factor, confounder)라고

부르는데, 어떠한 인자가 교란인자로 작용하기 위하여서는 결과인자 (여기서는 폐암)와도 관련성이 있어야 하고 동시에 폭로인자(여기서는 음주)와도 관련성이 있어야 한다.

2-4. 교란 bias를 제거하는 일반적인 방법

1) 교란 bias의 통제

그러면 어떠한 방법으로 교란 bias를 제거할 수 있을 것인가?

(1) 첫번째 단계는 연구설계(design) 과정에서 교란 bias가 끼어들지 않도록 교란요인을 제거하는 개념이다. 여기에는 3가지 구체적인 방법들이 있다.

① 첫번째 방법은 교란변수의 값을 제한(restriction)하는 방식이다. 예로 ‘음주가 폐암에 미치는 영향을 보고자 하었는데 흡연의 효과가 섞이는 것 때문에 참값을 측정하지 못하였다고 하면 흡연이 교란요인(confounder)로 작용하고 있는 셈이다.

그러면 만약 비흡연군만을 따로 떼어서 음주와 폐암과의 관련성을 보았다고 하면 흡연이 교란요인으로 끼어드는 것을 방지할 수 있게 된다. 이와 같이 대상자들을 교란요인에 따라 구분하여 이 중의 한 소집단(subgroup)만을 대상으로 분석하던가 또는 각 소집단을 개별적으로 분석하여 결과를 내면 그 요인에 의한 교란bias는 일어나지 않게 된다(물론 이러한 경우 전체 대상군에 있어서의 효과를 측정하지는 못하게 되나, 여하간 교란요인의 효과는 통제된 것이다).

② 두번째는 짝짓기(matching)에 의한 방식이다. 위의 예를 계속 인용한다면 음주군과 비음주군을 각각 1,000명씩을 대상으로 이들에게서 폐암발생이 같은지 다른지를 보고자 하는것이 우리의 연구 주제라고 한다면 첫번째 음주자를 대상자로 선정했을때 그 사람이 하루 1갑 담배를 피운다면 비음주자 중에서 담배를 하루 1갑 피우는 사람을 선정하여 음주자의 짝(pair)으로 정하고, 두번째 음주자를 보았더니 그 사람의 흡연량이 하루 10개피라고 하면 비음주자 중에서 흡연량이 하루 10개피인 사람을 정하여 그 짝으로 정한다. 이러한 방식대로 1,000번을 반복하면 두군(음주군과 비음주군)은 흡연량이 동일하게 구성되어 있기 때문에 흡연량이 음주와 폐암발생간의 효과추정에 교란요

인으로 작용할 수가 없게 되는 것이다. 그러나 짝짓기(matching)를 하려면 일반적인 많은 경비가 필요하고 실제로 우리가 노리는 효과보다 단점이 더 많기 때문에 어떤 경우이고 무조건 추천할 만한 방법은 아니다.

③ 세번째 방식은 무작위추출(randomization)방식이다. 이 방식은 동물실험과 같은 상황에서나 가능한 방법이지 산업보건역학 연구에서 일반적으로 가능한 방법은 아니다.

예를 들어서 2,000명의 대상자를 두군으로 나눈다고 하자. 1번부터 2,000번이 쓰인 종이를 보자기에 넣고 하나씩 꺼내어 처음 선정된 1,000명이 한군이고 나머지가 다른 한군으로 배정(random allocation)되면 이는 무작위 추출이다. 여기서 가장 중요한 점은 무작위 추출에 의하여 배정된 두군(각 1,000명씩)은 모든 면에서 동일하다는 점이다. 키를 재어도 두군은 같고 몸무게를 재어도 두군은 같다. 심지어 우리가 측정하지 못하는 모든 면에서 두군은 같다고 하여도 과언이 아니다(물론 실제 측정하면 조금씩 다를 수 있다. 그러나 이것은 random error이다).

무작위추출에 의하여 배정된 이러한 상황에서는 교란인자(confounder)가 있을수가 없고 교란 bias도 일어날 수가 없다는 점이다.

그러나 앞서서도 언급하였듯이 산업보건역학 분야에서는 이와 같은 실험적인 조치를 취한다는 것이 불가능하다. 단지 우리가 코호트 연구나 환자 대조군 연구에서 두 군을 비교하고자 할때, 연구 시작 단계에서 가능하면 두 군에서 우리가 관찰하고자 하는 요인 이외의 모든 변수의 값이 동일하게 되도록 노력하여야 한다는 말이다. 위에 언급한 짝짓기(matching)은 그 한 예로 볼 수 있다.

(2) 교란 bias를 제거하는 두번째 단계는 자료분석(data analysis) 단계에서 교란 bias를 제거하는 방식이다.

일반적으로 이 부분을 ① 표준화 (standardization)와 ② 다변량분석 (multivariate analysis)으로 구분하여 설명하고 있다.