

역학논문의 비판 (3)

카톨릭대학교 의과대학 예방의학교실

이 원 철, 박 정 일, 맹 광 호

지금까지 우리는 이 글들을 통하여 Vinyl chloride가 태아사망을 증가시키는 것을 예로 하여서 역학논문을 대할 때에 어떠한 관점에서 살펴보아야 할 것인가를 논하고 있다.

앞에서 이미 밝힌바와 같이 이러한 비판은 1)자료수집단계 2)자료분석단계 3)자료의 해석단계에 걸쳐서 시행하는 것이 바람직하다.

지난 두번의 글을 통하여 이미 자료수집단계와 자료분석단계에 대한 관점을 설명한 바 있다. 이번 글에서는 자료의 해석단계에 대하여 다루고자 한다.

- 저자 주 -

V. 역학연구에 대한 비판

3. 자료의 해석단계

1) 연구의 주요결과를 기술하여 보라.

이 연구에 있어서의 주요결과는 VCM에 폭로된 근로자들의 부인들에게서 나타난 태아사망률 (자연유산률)은 VCM에 폭로되지 않은 대조군 근로자들의 부인들이 경험한 태아사망률보다 대략 두배정도 높았으며, 또한 폭로전의 태아사망률과 비교하여도 2배정도 높다는 것이 이 연구의 주요한 결과이다.

이를 표를 인용하여 다시 표현하자면 지난 3월호에 실린 역학연구논문의 비판 (1)의 표1을 보면 폭로전 기간동안의 대조군과 VCM폭로군은 연령보정한 태아사망률이 6.9와 6.1을 보였으나, VCM에 폭로후에는 폭로군의 태아사망률이 15.8을 보여 대략 두배에 해당하는 태아사망률을 나타낸 것이다.

이렇듯 우선 그 해당연구의 주요결과가 무엇인지를 명확히 하는 것이 필요하다.

2) Bias가 있었다면 논문의 결과에 어떠한 영향을 주었겠는가?

이것은 해당논문에서 추정한 값이 과연 참값과 같겠느냐 혹은 크게 다르겠느냐 하는 문제이다. bias의 정의 그리고 그 종류들에 대하여는 지난 2월호에 설명한 바와 같다. 이를 기준으로 하여 위 연구에 대하여 '과연 어떠한 bias가 있었겠는가?'를 따

져보고자 한다. 그리고 그러한 bias가 논문의 결과에 어떠한 영향을 주었겠는가를 판단하여 보도록 하겠다.

(1) 선택 bias

선택 bias는 우리가 조사하고자 하는 전체 대상 인구중에서 조사를 위하여 뽑힌 (sampling된) 조사 대상자가 된 집단의 특성과, 조사에서 제외된 집단의 특성이 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 조사이던간에 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 (당뇨

병)을 대표할 수 있지 못하게 된다.

앞에서 정의한 대로 표현하자면 병원에 입원한 환자가 나타내는 값과 그 외의 그 질환을 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.

우리가 예로 들고 있는 VCM과 태아사망률에 관한 연구에서 보면 이 연구가 대상으로 하고 있는 VCM 폭로근로자들이 VCM에 폭로되고 있는 모든 근로자들을 대표할 수 있는가 하는 문제로 연결된다. 논문에서 저자들은 현재에 VCM 작업에 종사하고 있는 모든 근로자들이 대상이 되었다고 밝히고 있기 때문에 (물론 이 지역에 제한된 이야기이겠으나) 표본추출에 의한 문제는 없을 것으로 본다. 그러나 모든 대상자들중에서 과연 이 연구의 참여율이 어느 정도를 차지하느냐가 두번째 문제로 제기된다.

논문에서 저자들은 62% 내지 77%가 참여한 것으로 밝히고 있다. 참여율이 저조하면 저조할수록 그러한 제한된 대상으로 부터 나온 연구결과를 대상자전체에게 일반화 시키는 것이 문제가 될 것은 당연하다. 지금 이 연구에서는 이러한 참여율로 보아 선택 bias가 있어났다고 보기에 어렵다. 그러나 선택 bias가 전혀 없을 것이라고 판단하기도 힘든 상황이므로 한마디로 선택 bias의 유무를 결정하는 것은 조금 어려운 상태이다.

(2) 정보 bias

정보 bias는 비교하고자 하는 두군에서 각각 그 폭로 (exposure)나 결과 (outcome)를 측정할 때, 측정상의 문제로 인하여 양군에서 측정의 질 (정확도, accuracy)의 차이가 생기는 것을 말한다. 이 정보 bias는 관찰 bias라고도 불린다. 위의 VCM과 태아사망에 관한 연구에서 제일 문제가 되는 것은 이러한 정보 bias이다. 왜냐하면, 태아사망을 의사가 진단한 것에 의하여 결정하였다고 하면 문제가 없겠으나 이 연구에서 채택한 태아사망의 결정방법은 해당 산모가 '태아사망을 경험하였다, 경험하지 않았다'에 관하여 답변한 것을 근거로 하여 태아사망

을 결정하였기 때문이다.

한가지 예를 들도록 하자. 흡연군과 비흡연군에서 각각 폐질환의 일종인 폐기종 (emphysema)의 발생률을 비교하기 위하여 추적조사를 하기로 하였다. 폐기종은 병원에서 흉부 X-선 검사등의 진단 방법을 사용하지 않으면, 이 질환을 지니고 있으면서도 진단할 수 없는 그러한 질환이다. 만약에 흡연군에 속한 사람들이, 자신들의 건강에 더 관심이 많거나 (흡연의 유해성 때문에), 흡연으로 인한 장해 (예를 들어 기관지염)때문에 병원을 찾는 회수가 많으면, 이들에게서 나타나는 폐기종의 발생률은 병원을 찾는 회수가 빈번하다는 이유때문만으로 비흡연군에 비하여 더 높게 나타날 가능성이 있다. 두 군의 추적조사를 비교성이 있도록 (comparable) 신경쓰지 않는다면 정보 bias가 생길 수 있다. 즉 비흡연군에 비하여 흡연군에서 과다한 폐기종이 나타날 수 있는데 이는 흡연에 의한 생물학적 작용에 의한 것(참값)이 아니고 병원을 자주 방문하였다는 이유 때문에 이러한 결과(bias된 추정값)가 나타나는 것이다.

이는 차별 분류오류 (differential misclassification)에 속한다 [이에 상반되는 개념으로 무차별 분류오류 (nondifferential misclassification)가 있는데 이는 후에 설명하기도 하자].

이러한 차별 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 기억하여낸다고 하면 위에 설명한 차별 분류오류(이러한 상황에서는 이를 recall bias 라고 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. 따

라서 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 두가지 예는 모두 차별 분류오류(differential misclassification)에 대한 것이다. 이러한 차별 분류오류는 우리가 분석하여 낸 추정값을 참값보다 크게 추정하게 할 수도 있고, 참값보다 적게 추정하게 할 수도 있다. 위에 든 2가지 예는 모두 참값보다 크게 추정한 경우들이 되겠으나 이와 반대되는 예를 찾는 것은 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 발생률을 비교하려고 한다. 1,000명의 음주자와 1,000명의 비음주자를 대상으로 하였을 때 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 가상의 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 있다는 말이다). 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 0.1에 위치한다는 말이다).

무차별 분류오류(nondifferential misclassification)이 toward the null bias를 일으킨다는 점에 대하여는 J. Kelsey 등이 저술한 “Methods in Observation Epidemiology”의 p293에 자세히 설명되어 있으므로 이를 참조하기 바란다. 무차별 분류오류(nondifferential misclassification)는 random misclassification으로도 불리운다.

(3) 교란 bias(confounding bias)

교란 bias를 한마디로 설명하자면 효과가 섞이는 것(mixed up)이라는 설명은 이미 전번글에서 밝힌 바와 같다. Vinyl chloride와 태아사망간의 관련성에서 이를 살펴보자면 vinyl chloride의 폭로때문에 태아사망이 일어난다는 것을 밝히려 하고 하는데 다른 어떤 요인이 vinyl chloride의 폭로와도 관련이 있으면서 태아사망에 영향을 주기 때문에 그 요인의 효과가 섞여서 태아사망이 나타나면 이는 confounding bias가 일어난 셈이 되는 것이다.

이러한 상황에서의 초점은 우리는 다른 요인들이 태아사망에 미치는 영향을 다 제거한 후의, 순수하게 vinyl chloride가 태아사망에 미치는 영향만을 보고자 하는 것이 우리의 관심사이다.

이러한 교란 bias를 통제하는(또는 제거하는) 방

법에 대하여는 이미 전번 글에서 비교적 자세하게 다룬 바 있다. 여기서 예로 들고 있는 vinyl chloride와 태아사망의 연구에서는 연령(age)이 이미 보정되었고, 다른 특별한 요인이 교란요인으로 작용할 것이라고 판단할 근거가 없기 때문에 교란 bias가 영향을 줄 것이라고 비판하기는 어렵다.

3) 무차별 분류오류(random misclassification)는 결과에 어떠한 영향을 주었겠는가?

무차별 분류오류는 이미 설명한 바와 같다. 한가지 더 첨가하여 설명하고자 하는 것은 무차별 분류오류가 폭로(exposure)에서 생길 때와 질병(outcome)에서 생길 때 조금 달라질 수 있다는 것을 잠깐 비추려고 한다.

폭로(exposure)에서 분류오류가 생기는 경우는 모든 경우에 toward the null bias가 발생하게 되나 질병(outcome)에서 분류오류가 생기는 경우에는 연구방법에 따라서 (환자-대조군 연구이나 또는 관찰연구이나) 또는 보고자 하는 지수가 무엇이나에 따라서(비교위험도를 보고자 하느냐 또는 귀속위험도(attributable risk)를 보고자 하느냐에 따라서) bias가 없을 수도 있고 toward the null bias가 발생할 수도 있다.

이에 관하여는 KJ Rothman의 "Modern Epidemiology" p87을 참조하면 더 자세한 이해를 얻을 수 있을 것이다.

위 vinyl chloride와 태아사망에 관한 연구에서도 차별(differential) 분류오류가 아닌 무차별 분류오류(nondifferential, random misclassification) 이므로 결과의 해석에 큰 문제가 되지 못한다.

왜냐하면 이 연구에서는 비교위험도 2에 가까운 유의한 결과를 얻었으므로 무차별 분류오류가 있었다라든가 참값은 그 보다는 더 큰 위치(3 또는 4 등)에 있었기 때문이다.

이론적으로 한가지 문제가 생길 수 있는 경우는 추정값이 비교위험도 1-2를 나타내고 이것이 통계적으로 유의하지 못하였다고 했을 경우에 무차별 분류오류가 있었다고 하면 참값은 이보다 큰 2 또는 3에 위치할 가능성이 있고 이것이 통계적으로 유의할 수 있게 되므로 이러한 경우에는 결과의 해

석에 문제가 생기게 된다.

4) 이 결과는 어느 대상에게 일반화(generalization) 될 수 있겠는가?

일반화(generalization)은 매우 중요하고 현실적인 개념들 중의 하나이다. 우리가 쥐 10마리를 대상으로 연구하는 것은 모든 쥐에서 그렇게 나오리라고 가정하고 결과를 내려고 하는 것이지 해당 10마리의 결과를 알고 싶기 때문은 아닌 것이다.

어떠한 연구를 행한 후에 그 결과를 해당자들 모두에게 적용하려고 하는 일반화(generalization)가 가능하려면 우선 1) 연구대상수(sample size)가 커야 하고 2) bias가 없어야 하며 3) 연구방법이 원인 결과를 판정하는데에 있어서 우수한 방법이어야 할 것이다.

그러나 위의 vinyl chloride와 태아사망에 관한 연구는 1) 대상연구자 수가 적고 2) 관찰 bias(information bias)의 가능성이 크며 3) 단면연구(cross-sectional study)로 시행하였기 때문에 이 연구결과 하나만으로 전체 vinyl chloride 작업자들에게 이를 적용시키는 것은 시기상조라고 볼 수 있다.

이와 유사한 연구를 다른 대상자들에게도 몇차례 시행한 후에 일반화를 하는 것이 적합한 것이다.

일반화(generalization)의 개념에 대한 언급이 다소 소홀한 점이 있으나 이는 후에 더 자세히 다루도록 하기로 하겠다.

5) 자료의 해석에 조심스러웠는가?

연구결과를 저자들이 해석하는 데에 있어서 어떠한지를 비판하는 면이다. 극단적인 예로 표현하자면 기술적인(descriptive) 조사를 수행하여 놓고서 그 결과를 가지고 어떠한 관련성에 대한 결론을 강하게 표현하였다고 하면 이는 아마도 눈먼 사람이 코끼리 다리를 만져보고 코끼리 전체가 어떻게 하고 판단하는 것과 다를 바 없겠다.

위의 예로 든 연구에서는 위 결과를 가지고 저자들이 vinyl chloride의 유전적인 효과가 가능("possible")하다고 결론내렸기 때문에 자료의 해석에 조심스러웠다("conservative")고 할 수 있을 것이다. ♣