

근거에 기반한 문헌 평가

Evidence-based Appraisal of Clinical Studies

- 배희준 -

교신저자 배 희 준

주소: 경기도 성남시 분당구 구미동 300번지
분당서울대학교병원 신경과 (463-707)

■ braindoc@snu.ac.kr

I. 근거에 기반한 문헌 평가란?

근거중심의학의 아버지인 영국 옥스포드 대학의 Sackett교수는 그의 저서 'Evidence-based medicine: How to practice and teach EBM' 에서 근거중심 의학을 아래와 같이 정의하였다.

「근거중심의학이란 임상 의사의 전문적인 견해 (clinical expertise)와 환자의 가치(patient value)에 현재까지의 의학연구에서 얻을 수 있는 최선의 근거(best research evidence)를 통합(integration)시키는 것이다.」 (1)

필자는 근거중심 의학을 근거중심 의학답게 만드는 것은 '최선의 근거'라는 문구에 있다고 생각한다. 근거중심 의학은 여러 가지 자료원 통해 얻을 수 있는 근거들이 최선의 것인지를 판단하는 것에서 출발한다고도 할 수 있다. 하루에도 천 편이 넘는 의학논문이 Medline에 등재가 되고 이중 진료에 영향을 미칠 수 있는 임상시험만도 수십 편이 넘는 상황에서 의학적 결정의 준거가 되는 의학연구 문헌에 대한 평가 능력은 어찌 보면 보건의료인이 가져야 할 기본 소양일지도 모른다.

근거중심 의학이 무엇이고 근거에 기반한 문헌 평가가 근거중심 의학에서 어떠한 역할을 하는지 이해하는 데에는 증례를 통하는 것이 가장 효과적이고 근거중심 의학 관련 교육 책자도 모두 이런 형태를 취하고 있다. 필자도 같은 방식을 사용하려고 한다.

「45세 남자인 동료 의사 홍길동씨는 어머니가 뇌졸중으로 고생한 끝에 사망하여 평소에도 뇌졸중에 대한 걱정이나 두려움이 많았다. 담배는 하루에 한 갑 정도를 피고 있고 술은 일주일에 한 두 차례 가볍게 친구들과 어울려 한잔하는 정도이다. 일이년에 한번씩은 정기적으로 신체검사를 하며 이제까지 별다른 질환을 앓은 적은 없다. 키는 175cm이고 몸무게는 80kg로 과체중 상태이다 (Body Mass Index 26.1). 2년 만에 시행한 올해 신체검사에서 홍길동씨는 자신의 혈압이 높음 (150/90mmHg)을 알게 되었다. 홍길동씨는 아스피린이 심혈관계질환 예방효과가 있다는 사실을 기억하고 여러분에게 자신이 아스피린을 먹어야 하는 것은 아닌지 상의를 해왔다. 여러분은 아스피린 복용에 관하여 홍길동씨에게 어떠한 조언을 할 것인가? 이를 위해 어떠한 과정이 필요할 것인가?」

모두 느끼시겠지만 상당히 어려운 문제이다. EBM적인 접근법에 익숙하지 않은 의사라면 자신의 이제까지 경험과 공부한 것들, 그리고 조금 더 성의를 보인다면 MEDLINE을 검

색하여 관련 논문 초록을 몇 편 검토해 보고 아스피린 투여 여부를 결정할 것이다. 필자가 이야기 하고자 하는 것은 아스피린을 투여하거나 하지 않는 결정 자체의 잘잘못에 관한 것이 아니다. 이 주제에 관해 공부를 해보신 분이라면 아시겠지만 현재 어떤 쪽으로든 결정이 난 문제는 아니다. 강조하고 싶은 것은 근거중심의학적인 접근과 과거의 방법이 어떻게 다른냐는 것이다.

근거중심의학에 익숙한 의사라면 일단 상술한 흥길동씨의 치료방침을 결정하는데 있어 대답을 요하는 임상적 질문(answerable clinical question)이 무엇인지를 구체화할 것이다. 질문은 “뇌졸중의 일차예방에 아스피린이 유효한가?”로 구체화 해볼 수 있다. 다음 단계는 질문에 답하기 위해 필요한 근거를 찾는 것이다. 의학관련주제의 근거를 찾는데 가장 보편화된 방법이 MEDLINE검색을 하는 것이므로 (그러나 가장 보편적이기는 하지만 MEDLINE이 의학관련근거를 모두 포괄하지는 못한다는 사실을 알아야 한다) 이를 이용하기로 결정한다 (따라서 이 과정에서 문제의 특성에 따라 다른 데이터베이스를 이용할 필요는 없는지 검토가 필요하다). 단시간 내에 원하는 결과를 얻기 위해서는 적절한 검색어를 선정하는 과정도 중요하며 여기에도 몇 가지 원칙과 방법론이 필요하며 이 문제는 다른 필자들이 다룰 것이므로 생략하겠다. 상술한 문제에 대해서는 실제 “primary prevention”과 “aspirin” 및 “stroke”을 검색어로 선정하여 검색을 시행하였다. 2006년 11월10일 현재 모두 208개의 논문이 검색되었다. 다음 검색을 “randomized controlled trial”로 제한하였더니 모두 25개가 검색이 되었다. 제목과 초록을 검토하였을 때 검색의 목적에 가장 적합한 논문을 1989년 NEJM에 실린 “Final report on the aspirin component of the ongoing Physicians’ Health Study” 이었다.(2) 논문을 내용을 정리하면 아래와 같다.

- 모두 22,071명의 건강한 남자 의사를 대상으로 연구를 시행하였다.

- 11,037명이 아스피린 325mg을, 11,034명이 위약을 무작위 배정 받았고 평균 60.2개월 동안 이중맹검법에 의해 추적 관찰이 이루어 졌다.
- 뇌졸중, 심근경색, 혈관질환에 의한 사망의 발생을 관찰하여 primary end point로 삼았다.
- 심근경색의 경우 아스피린 투여군에서 44% 상대위험도 감소가 관찰되나 ($p < 0.00001$) 뇌졸중의 경우 통계적인 유의성은 없으나 오히려 아스피린 투여군에서 22% 뇌졸중의 위험도가 증가하였고 이는 주로 뇌출혈의 증가에 기인한 것이었다. 아스피린 투여군에서 위약군에 비해 혈관질환의 사망 역시 유의한 차이를 보이지 않았다 (상대 위험도, 0.96)

따라서 상술한 연구 결과에 의거하여 아스피린 복용은 흥길동씨에게 의미가 없다는 결론을 내릴 수도 있다. 그러나 위와 같은 결정을 단일한 하나의 연구에 근거하여 내리는 데에는 여러 가지 문제점이 발생한다. 임상에서 동일한 치료에 대해 두 가지 연구가 상이한 결과를 보이는 경우를 종종 경험하게 된다. 이처럼 이론의 여지가 있는 문제에 대해 하나의 연구 보고에 의존하여 결론을 내리는 것은 매우 위험한 일이고 이때 필요한 것이 메타분석이다. 메타분석은 개별적인 연구들에서 공통적으로 존재하는 하나의 모수가 있다고 가정하고 이 모수를 추정하여 병합하는 행위를 말하며 아래와 같은 장점을 지닌다.

- 메타분석에서 종합된 결과는 폭 넓고 강건한 증거가 된다.
- 메타분석에서 종합된 결과는 비교적 바이어스가 적은 표본으로부터 얻은 결과이다.
- 메타분석으로 여러 연구에서 나온 정보를 종합하게 되면 통계적 검정력이 증가한다.
- 메타분석의 종합방법은 여러 연구로부터 증거를 축적하는 표준적인 과정이다.

Primary prevention, aspirin, stroke을 검색어로 검색을 한 결과를 다시 “meta-analysis”로 제한하였을 때 모두 20개 논문이 검색이 되었다. 우리 문제에 가장 적합한 연구를 찾기 위해 초록을 검토하였고 이중 JAMA 2006년 1월호에 실린 “Aspirin for the primary prevention of cardiovascular events in women and men”이란 제목의 메타분석이 가장 적합한 한 것으로 결정하였고 원문을 검토하였다(3). 검토 결과 이제까지 연구결과로는 건강한 사람이나 혈관질환의 위험인자는 있으나 임상적으로 혈관질환의 증거는 없는 남자에게서 뇌졸중의 예방을 위해 아스피린이 유용하다는 근거는 없고 따라서 아스피린 복용을 권하지 않기로 결정을 내렸다.

그러나 이 결정의 근거에도 여러 가지 약점이 있다. 이제까지 뇌졸중의 일차예방에 대한 연구는 대부분 백인을 대상으로 한 것들이고 따라서 뇌졸중의 발생률이 상대적으로 높으면서 뇌졸중의 유형도 다소 다른 우리나라 사람에게 적용하는 것에는 무리가 있다. 이외에도 meta-analysis의 대상이 되었던 “randomized controlled trial”의 연구대상들 중 남자는 다수가 남자의사 대상의 두가지 연구에서 비롯되었다점과 최근 수행된 여자 대상의 연구에서는 아스피린이 뇌졸중 예방 효과가 있는 것으로 나온 점 등이 쉽게 드러나는 약점일 것이다.

따라서 의료진은 홍길동씨에게 아스피린의 처방을 권하지 않는다는 결정을 내리기 전에 이러한 문제에 대해서도 홍길동씨와 충분히 논의해야 할 것이다. 만일 뇌졸중의 전문가인 의료진 자신이 홍길동씨의 뇌졸중의 위험도가 높다고 판단하고 홍길동씨 자신 또한 여러 근거의 불충분함에도 불구하고 뇌졸중의 일차예방을 위해 아스피린의 복용을 강력히 원한다면 아스피린을 권한다는 결정 또한 타당할 수 있다고 저자는 생각한다.

상술한 과정을 통해 근거중심의학의 임상에 적용해 나아가는 과정과 근거중심의학의 핵심요소 중 하나가 근거의 타당성을 판단하는데 있는 것을 쉽게 이해할 수 있을 것이다.

II. 근거에 기반한 문헌 평가의 방법론: 치료에 관한 논문을 중심으로

의학논문 중 의료에 직접적인 영향을 크게 주는 것은 치료에 관한 논문이다. 그리고 같은 조건 하에 시행된 동일한 치료법에 관한 무작위 임상시험이 여러 개가 있을 경우 이들의 결과를 종합하여 검토한 논문이 단독 연구에 의한 결과 보다는 더 좋은 해답이 될 것이다. 이러한 이유로 문헌 검색은 체계적 검토 논문(systematic reviews)을 찾는 데에서 시작하는 것이 권고됨은 이미 이야기 하였다. 그러나 체계적 검토 논문도 각각의 임상시험을 개별적으로 평가하는 데에서 시작하며(이들 각각의 임상시험이 타당한 방법으로 시행되었음을 확인해야 한다) 현 시점에서 체계적인 검토 논문 보다는 개별적인 임상시험이 더욱 흔하기 때문에 본고도 개별적인 임상시험에 대한 평가 방법을 중심으로 진행하고자 한다.

치료에 관한 유용한 증거를 찾기 위해서는 어디에서부터 비판적 분석을 시작할지를 결정해야 한다. 타당성 분석에서 시작할 수도 있고 임상적 중요성 분석에서 시작할 수도 있다 (해당 연구가 타당하지 않다면 아무리 치료의 효과가 훌륭하다고 하더라도 무슨 의미가 있겠으며, 또한 연구 결과가 임상적으로 의미가 없다면 연구의 타당성에 누가 신경을 쓰겠는가?). 한 가지 질문에 답한 뒤 또 다른 질문에도 답해야 한다는 사실만 기억한다면 어디서 시작하는지는 중요하지 않다. 설명의 편의를 위하여 인터페론 치료가 요구되는 2차성 진행성 다발성 경화증 환자(secondary progressive multiple sclerosis)(4)를 예를 들어 보겠다. 임상적으로 제기된 질문은 다음과 같다: “이차성 진행성 다발성 경화증 환자에서 인터페론 투여가 진행의 가능성을 감소시킬 수 있을 것인가?”

이 임상시험의 결과는 타당한가?

1. 치료에 대한 환자 배정은 무작위적이며 무작위 배정표 (randomization list) 는 비밀이었는가?

타당성 분석의 첫 단계는 무작위 배정에 관한 것이다. 연구자가 환자를 각 치료군에 배정하는 데 있어 동전을 던지는 것과 비슷한 방법을 사용하였는지를 검토해야 한다. 무작위 배정을 중요하게 여겨야 하는 이유는 이것이 치료의 시작 시점에서 각 치료군 사이에 우리가 치료를 통하여 방지하고자 하는 사건의 위험성을 비슷하게 만드는 최선의 방법이기 때문이다. 무작위 배정은 각 군을 예후인자(질병의 정도나 다른 예후 결정인자들)에 대하여 균형을 맞추게 한다. 그리고 치료군 사이에 이러한 예후인자의 분포가 균일하지 않으면 치료의 효과가 과장, 혹은 축소되거나 심지어는 치료로 인해 병세가 악화되는 것처럼 보일 수도 있다. 무작위 배정은 대상이 되는 질환에 대해 충분히 알지 못하는 상태에서 치료군들 사이에 이러한 예후인자의 균형이 가능하게 할 수 있다.

또한 무작위 배정 과정이 환자를 임상시험에 모집한 임상 의사에게 비밀로 이루어 졌는지를 살펴 보아야 한다. 만약 그러하다면 해당 의사는 다음 모집 환자가 어떠한 치료를 받게 될지 모를 것이고 따라서 의식적이던 무의식적이던 비교하려는 치료군들 사이의 균형에 영향을 미칠 수 없을 것이다. 무작위 배정의 실패와 마찬가지로 배정 과정에 대한 비밀엄수의 실패는 치료효과를 과장하거나 축소하는 방향으로 연구결과를 왜곡할 수 있다.

일반적으로 무작위 배정을 했다는 것은 자랑할 만한 일이며 서 논문 제목이나 좀더 흔하게는 초록에 명기되기 때문에 이를 확인하는 것은 어렵지 않다. 대조적으로 무작위 배정표가 비밀이었던지 여부는 흔히 기술되는 내용은 아니지만 무작위 배정 과정이 전화를 통해서나 기타 환자가 임상시험에 모집되는 위치에서 떨어져 있는 곳에서 수행된다면 어느 정도

비밀성을 확신할 수 있다. 전술한 다발성 경화증에서 인터페론의 효과에 관한 임상시험에서는 제목이 무작위 배정에 의한 것임을 알려주고 있으며 방법론을 훑어보기만 하여도 이 다기관임상연구의 무작위 배정이 중앙의 한 곳에서 수행되었음을 알 수 있다.

만약 연구가 무작위 배정에 의한 것이 아니라면 더 이상 읽지 말고 다른 연구를 찾아보기를 권하며, 무작위 배정에 의한 연구를 전혀 찾을 수 없을 경우에만 다시 살펴 볼 만하다.

2. 환자에 대한 추적관찰은 충분히 길고 완전한가?

일단 연구가 무작위적 방법으로 수행되었음이 확인되었다면 다음으로 필요한 것은 연구에 동원된 모든 환자들의 임상상이 충분히 연구결과에 반영되었는가를 살피는 것이다. 이상적으로는 추적관찰 기간 동안에 소실된 환자가 전혀 없는 것이며 이는 소실된 환자에게 나타난 질병결과가 연구의 결론에 영향을 미칠 수도 있기 때문이다. 예를 들어 실험적 치료의 부작용 때문에 환자의 일부가 중간에 소실되고 이것이 결과의 분석에 반영되지 않는다면 이는 치료 효과의 과장으로 이어질 수 있다.

어느 정도까지 중도탈락이 허용될 수 있는 것일까? 임상시험의 결론을 확신하기 위해서는 추적관찰 중 소실된 환자들을 모두 최악의 경우로 가정하고도 (예후가 좋을 것으로 추정되는 군에서 소실된 환자들은 모두 예후가 나쁜 쪽으로 가정하고 예후가 나쁠 것으로 추정되는 군에서 소실된 환자들은 모두 예후가 좋을 것으로 가정하자) 원래의 결론을 유지할 수 있어야 한다. 20% 이상의 추적관찰 소실이 있을 경우 전술한 최악의 가정 하에 이루어지는 분석을 무시히 통과할 수 있는 임상시험은 매우 드물다. “Evidence-based Medicine”이나 ACP Journal Club은 추적관찰율이 80% 미만의 임상시험은 출판하지 않고 있다. 인터페론 연구의 경우 추적관찰율은

90% 이상이었다.

또한 추적관찰 기간은 임상적으로 의미있는 질병결과를 관찰할 수 있을 만큼 충분히 길어야 한다. 예를 들어 인터페론의 효과를 평가하기 위한 연구가 환자를 한달 정도만 추적관찰한다면 유용한 결과를 얻을 가능성은 거의 없다. 목표질환의 성격을 고려한다면 추적관찰 기간은 적어도 수개월 이상이어야 하고 이상적으로는 수년은 되어야 한다. 인터페론 연구의 경우 각 환자들의 추적관찰 기간은 적어도 2년 이상이었다.

3. 환자는 모두 애초에 무작위 배정된 군에 따라 분석되었는가?

무작위 배정 후 일어나는 어떤 일이든지 대상환자에게 관찰하고자 하는 사건이 일어날 가능성에 영향을 미칠 수 있기 때문에 연구에 포함된 모든 환자들을 처음 배정된 데로 분석하는 것이 중요하다 (비록 일부 환자들이 중간에 투약을 중단하거나 고의적으로 배정된 것과 다른 치료를 받더라도 처음 배정된 군에 따라 분석하여야 한다). 이미 언급한 바와 같이 연구대상 약물을, 비록 그것이 위약일지라도, 복용한 환자와 그렇지 않은 환자는 서로 다른 질병결과를 보인다. 무작위 배정의 의미를 살리기 위해서는 배정된 치료를 실제 받았는지 여부와는 관계없이 애초에 배정된 데로 환자를 분류하여 분석하는 “intention-to-treat analysis”가 요구된다. 인터페론 연구의 경우 이러한 “intention-to-treat analysis”가 사용되었고 전술한 세가지 타당성 기준을 모두 통과하였다.

4. 이중 맹검법이 시행되었는가?

5. 각군은 시험의 대상이 되는 치료법 외에는 모든 측면에서 동일하게 취급되었는가?

중요성은 떨어지지만 연구의 타당성을 평가하는 데 있어 몇 가지 고려해야 할 점들이 있다. 치료 효과에 대한 예측으로 인

해 환자의 증상에 대한 호소나 의사의 증상에 대한 해석이 왜곡되는 것을 피하기 위해 이중 맹검법은 필수적이다. 또한 이는 연구대상이 되는 실험적 치료법 외의 다른 치료법의 분포가 대상 환자군 중 어느 한 군에 치우치는 것을 예방한다 (예를 들어 말기 암환자의 치료법에 관한 위약 대조군 연구에서 해당환자가 위약대조군에 속하는 것을 환자나 의사가 알게 된다면 환자에게 대체요법이 시행될 가능성이 증가할 것이고 이는 연구 결과에 심각한 영향을 미칠 수 있다)

환자나 의사에게 치료법에 맹검 상태를 유지하는 것이 힘들 경우 (수술에 관한 임상시험에서처럼), 해당 환자에게 배정된 치료법에 대해 모르는 제 3의 의사가 임상 기록을 (치료에 대한 언급을 제거한 상태에서) 평가하거나 질병결과에 대한 특수하고 객관적인 측정을 수행하게 하는 방법이 있다. 비록 인터페론 연구가 이중맹검법에 의한 위약-대조군 임상시험이지만 인터페론 주사가 국소적인 피부반응을 종종 일으키기 때문에 담당의사는 환자들 중 누가 인터페론 치료를 정말 받고 있는지에 관해 어느 정도 추정이 가능하다. 그러나 이 점에 대해서는 연구에서 질병결과에 대한 최종 판정은 환자를 직접 치료하지 않는 제 3의 의사에 의해 수행되었다는 점을 고려한다면 안심할 수 있다.

6. 각군은 임상시험의 시작 단계에서는 유사하였는가?

마지막으로 각 환자군들이 치료법을 제외하고는 예후에 영향을 줄 수 있는 모든 측면에서 유사한지를 살펴보아야 한다. 만약 그렇지 않다면 이들 잠재적 예후인자에 대한 보정이 이루어졌는지를 조사해야 한다. 인터페론 연구에서 위약군과 인터페론 투여군 사이에 예후인자에 대해 의미있는 차이는 없었다.

해당 연구의 타당성 있는 결과가 임상적으로 얼마나 중요한가?

임상시험의 결과가 임상적으로 얼마나 중요한지를 결정하기 위해서는 다음 두 가지 요소를 고려해야 한다. 첫째, 시험 결과에 대한 가장 적절한 임상적 표현 수단을 찾아야 한다; 둘째, 이 결과를 해당 질환에 대한 다른 치료법의 결과와 비교해야 한다.

1. 치료효과는 어느 정도 인가?

연구 결과를 묘사하는 방법에는 여러 가지가 있으며 이 중 중요한 것들을 표 1에 모아놓았다: 다시 한번 인터페론 연구를 이용하여 이를 설명해 보자. 표 1에서 실제 임상시험의 결과를 살펴보면 33개월째 장애의 진행은 대조군에 배정된 다발성 경화증 환자의 경우 50%에서 발생했고 (이를 대조군에서의 사건 발생률 (Control event rate: CER) 라고 부른다) 인터페론

투여군의 경우 39% (이를 실험군에서의 사건발생률 (Experimental event rate: EER) 이라고 부른다) 에서 발생하였다. 이 차이는 통계적으로 의미있는 것이었으나 이를 어떻게 임상적으로 유용한 형식으로 표현할 것인가가 문제이다. 최근까지 임상논문에서 주로 쓰인 방법은 상대위험감소율 (Relative risk reduction(RRR)) 로 대조군과 실험군에서의 사건 발생률 차이의 절대값을 대조군에서의 사건 발생율로 나눈 값이다. 인터페론 연구에서 RRR은 (50%-39%)/50%로 22%이며 인터페론 투여가 장애의 진행에 대한 상대적인 위험을 22%까지 줄인다고는 것을 의미한다. 비슷한 방법으로 실험적인 치료법이 긍정적 사건의 가능성을 증가시키는 것을 상대편익증가율 (relative benefit increase (RBI); 역시 |CER-EER|/CER로 계산된다)이라고 부른다. 마지막으로 치료가 부정적인 사건의 가능성을 증가시킬 경우, 같은 식을 사용하여 상대 위험 증가율 (relative risk increase)을 계산할 수 있다.

상대위험감소율의 단점 중 하나가 표1의 하단의 가상례에

표 1. 치료의 효과에 대한 임상적으로 유용한 측정 수단

	사건발생율= 33개월까지 장애의 진행		상대위험 감소율 (Relative risk reduction) (RRR= CER-EER /CER)	절대위험 감소율 (Absolute risk reduction) (ARR= CER-EER)	Number needed to treat (NNT = 1/ARR)
	대조군의 사건발생율 (위약군에서) (CER = Control event rate)	실험군의 사건발생율 (인터페론 투여군에서) (EER = Experimental event rate)			
실제 임상시험례 Lancet 1998; 352: 1491-7	50 %	39%	(50%-39%) / 50% = 22%	50%-39% = 11%	1/11%=9
효과가 미약한 가상례	0.00050%	0.00039%	(00050% - 0.00039%) /0.00050% =22%	0.00050% - 0.00039% = 0.00011%	1/0.00011% = 909090

드러나 있다: 즉 이는 치료 없이 사건이 발생할 가능성 (CER)을 반영하지 못하며 따라서 치료효과의 정도를 표현하지 못한다. 예를 들어 장애의 진행율이 대조군에서 0.00050% 정도로 매우 미약하고 실험치료군 역시 0.00039%로 보잘 것 없더라도 RRR는 22%로 상대적으로 매우 높을 수 있다. 이와는 대조적으로 절대 위험 감소율 (대조군에서의 사건발생율과 실험군에서의 사건 발생율 사이에 절대적인 차이를 말함)은 두 군 사이의 발생율의 차이를 구별하게 해주면서 동시에 대조군에서의 기본적인 사건 발생율을 반영할 수 있다. 인터페론 연구에서 절대 위험 감소율 (ARR)은 50%-39%=11%이다. 기본적으로 사건 발생율이 매우 낮은 가상례에서는 절대 위험 감소율 역시 매우 낮아 0.00011%에 지나지 않는다. 따라서 절대 위험 감소율은 상대 위험 감소율 보다 치료 효과를 측정하는데 있어 좀더 유용한 수단이다. 실험적 치료법이 긍정적인 사건의 발생율을 증가시킬 때 절대 편익 증가율 (absolute benefit increase: ABI)을 사용할 수 있으며 역시 두 사건 발생율의 산수적인 차이로부터 계산된다. 비슷하게 실험적 치료법이 부정적 사건의 발생율을 증가시킬 때 절대위험증가율 (absolute risk increase: ARI)이 계산될 수 있다.

그러나 절대위험감소율은 기억하기가 쉽지 않다 (특히 1.0 미만일 때 그렇다). 한편 한 명의 사건 발생을 막기 위해 임상 시험 기간 동안 치료법을 시행해야 하는 환자의 숫자를 NNT (number needed to treat) 라고 하며 이는 절대위험감소율의 역수로 계산하며 정수로 표현될 수 있다. 인터페론 연구에서 NNT는 9명이라는 것은 33개월 동안 9명의 환자를 치료해야 한 명의 장애의 진행을 막을 수 있다는 것을 의미한다. 표 5.3의 하단에 있는 가상례의 경우 한 명의 악화를 예방하기 위해 100만 명에 가까운 환자를 치료해야 함을 의미하며, NNT의 임상적 의미를 역설적으로 강조하고 있다.

비슷한 방법으로 한명의 환자에게 부작용이 발생하기 전까지 치료해야 할 환자의 수(NNH; number needed to harm)

를 이용하여 치료의 부작용에 대해 이야기 할 수도 있다. NNH는 1/절대위험증가율 (Absolute risk increase, ARI)로 계산할 수 있다. 인터페론 연구에서 2년의 연구기간 동안 감기 증상이나 주사를 맞은 자리의 피부괴사와 같은 부작용이 대조군의 경우는 37%에서 발생하였으나 인터페론 투여군에서는 64% 발생하였다. 절대위험증가율은 27%이고 따라서 33개월 동안의 NNH는 1/27%로 4명이다. 한명의 환자에게 33개월 사이에 부작용이 발생하는 데는 단지 4명의 환자에게 치료를 시행하는 것으로 충분하다는 결론을 얻을 수 있다. 따라서 NNT와 NNH는 하나의 바람직하지 않은 질병결과를 막거나 유발하기 위해 소요되는 노력의 양에 대한 명확한 계측 수단을 제공할 수 있으며, 노력성고비 (임상에서 개략적으로 사용할 수 있는 일종의 비용효과분석)를 표현하는 방법으로써의 매력적이다. 실제로 NNT (그리고 NNH) 는 매우 광범위하게 사용되고 있다.

NNT를 이해하려면 두 가지 측면에 관한 추가적인 고려가 필요하다. 첫째, NNT의 의미에는 추적관찰기간이 내포되어 있다는 점이다. 예를 하나 들어 보자. 1991년 발표된 경동맥내막절제술의 효과에 대한 논문(5)을 살펴보면, 증상을 일으킨 중증의 경동맥 협착증에서 한명의 환자에게서 생명을 위협할 수 있는 뇌졸중이나 이로 인한 사망을 막기 위해 시행해야 할 경동맥 내막절제술의 수, 즉 NNT는 10이며 이는 2년 동안 한 건이 질병결과, 즉 뇌졸중의 발병이나 사망을 막기 위해 10명에게 수술을 시행하여야 한다는 의미이다 (수술 자체는 한 시간 안에 끝난다). 이처럼 NNT에 시간의 개념이 내포되어 있는 이 이유로 서로 다른 추적관찰 기간을 지닌 NNT를 비교하고자 할 때는 추적관찰기간의 차이에 대한 가정을 하고 이를 보정할 필요가 있고 이를 식으로 표현하면 아래와 같다.

$$NNT_{\text{가상적인}} = NNT_{\text{실제 관찰된}} \times (\text{실제 관찰기간} / \text{가상적인 관찰기간})$$

2. 치료효과에 대한 추정은 얼마나 정밀한가?

NNT와 관련하여 중요한 두 번째 사항은 다른 임상적인 측정에서와 마찬가지로 추정에 관한 문제이다. NNT는 참값에 관한 추정이고 이를 위해 실제 NNT값이 포함될 것이라고 신뢰성을 가지고 추정할 수 있는 구간을 규정하여야 한다 (95% 신뢰구간이 NNT값의 95%가 속하는 범위를 구체화한 것이다). 이를 구하는 공식은 ARR의 95% 신뢰구간을 표준오차를 이용하여 구한 다음 여기에 역수를 취하면 된다. 아래는 이 과정을 수식으로 표현한 것이다.

$$SE \text{ (standard error)} = \sqrt{P_1(1-P_1)/N_1 + P_2(1-P_2)/N_2}$$

$$95\% \text{ Confidence Interval of ARR} = ARR \pm 1.96 * SE$$

$$NNT = 1/ARR$$

해당 연구가 타당성이 있으며 임상적인 중요성도 충분하다면 이를 환자에게 적용할 수 있는가?

이제까지 탐색하여 비평한 근거가 타당성이 있고 임상적으로도 중요하다면 다음 단계는 이를 실제 환자에게 적용하는 것이 가능한지를 살펴보는 것이다. 따라서 이는 근거중심의학의 임상적 적용과정이라고 볼 수도 있다. 본고의 범위를 벗어나므로 간단하게 필수 요소만을 요약해 보겠다.

1. 해당 연구 결과를 적용하지 못할 정도로 실제 환자와 연구대상 환자 사이에 차이가 존재하는가?

먼저 해당 연구 결과를 적용하지 못할 정도의 차이가 연구대상 환자군과 실제 환자 사이에 존재하는지를 고려해야 한다. 해당 연구의 대상 선정 기준에 실제 환자가 부합되는지를 살펴보고 조금이라도 어긋날 때 해당 연구 결과를 적용하지 않는

것이 한가지 방책일 것이다. 그러나 이는 분별있는 방법은 아니다. 좀 더 합리적인 접근방법은 실제 환자와 해당 연구의 환자군 사이의 사회인구학적 특징이나 병태생물학적인 차이가 연구결과를 적용할 수 없을 정도로 유의한지를 알아 보는 것이다. 유의한 차이가 있을 때에만 해당 연구 결과를 무시하고 다른 근거를 찾기 위한 작업을 재개한다. 이처럼 실제 환자와 연구대상 환자군 사이에 연구결과를 적용할 수 없을 정도의 차이에 있을 때에만 해당 연구 결과를 무시하는 임상적인 접근법을 택함에 따라 엄격한 접근을 했을 때 흔히 일어났던 특정한 연구의 결과를 무시하는 일이 적어질 수 있다.

2. 지금 상황에서 시행이 가능한 치료법인가?

다음으로 해당 치료법이 현재 상황에서 가능한 것인지를 생각해보아야 한다. 여기에는 환자나 현재의 보건의료제도가 해당 치료법의 시행과 감독, 그리고 추적 관찰에 들어가는 비용을 감당할 수 있는지에 대한 고려가 포함된다.

3. 해당 치료법 적용 시 실제 환자에서 기대되는 잠재적 편익과 손실은 무엇인가?

검토한 연구 결과가 실제 환자에게 적용 가능하고 해당 치료법이 현실적으로 시행 가능하다고 판단된다면 다음은 특정 환자에서 해당 치료법을 시행했을 때의 득과 실을 따지는 것이다. 이때 전술한 NNT와 NNH가 활용될 수 있다

4. 치료법 자체와 질병 치료 결과가 실제 환자에게 가지는 가치와 기대는 무엇인가?

특정환자에 있어 치료에 의한 이득과 위험을 개별화하는 데 성공했을지라도 여전히 여기에는 환자 자신의 가치관과 기호에 대한 고려가 부족하다. 그렇다면 이러한 것들을 어떻게 치

료에 대한 권고에 통합시킬 수 있을 것인가를 고민해야 할 것이다. 이를 위해 근거중심의학에서 권장하는 방법 중 하나가 LHH (likelihood of being helped vs. harmed; 득실의 가능성)를 이용하는 것이다.

III. 결어

이상으로 두서 없이 근거중심의학 방법론의 임상적용을 특히 문헌 평가란 관점에서 살펴보았다. 필자의 조그만 노력이 이 글을 읽으시는 여러 선생님들께 근거중심의학이라는 흐름을 경험할 수 있는 기회가 되었기를 기대한다.

참고문헌

1. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. Evidence-based medicine: How to practice and teach EBM, 2nd ed. London; Churchill Livingstone, 2000.
2. Steering Committee of the Physicians' Health Study Research Group. Final report on the aspirin component of the ongoing Physicians' Health Study. *N Engl J M* 1989; 321(3):129-135.
3. Berger JS, Roncaglioni MC, Avanzini F, Pangrazzi I, Tognoni G, Brown DL. Aspirin for the primary prevention of cardiovascular events in women and men: a sex-specific meta-analysis of randomized controlled trials. *JAMA* 2006; 295(3):306-313.
4. Placebo-controlled multicentre randomised trial of interferon beta-1b in treatment of secondary progressive multiple sclerosis. European Study Group on interferon beta-1b in secondary progressive MS. *Lancet* 1998; 352(9139):1491-1497.
5. Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade carotid stenosis. North American Symptomatic Carotid Endarterectomy Trial Collaborators. *N Engl J Med* 1991; 325(7):445-453.