

종 설

임상진료지침의 개발

신 승 수

아주대학교 의과대학 호흡기내과학교실

Development of Clinical Practice Guidelines

Seungsoo Sheen, M.D.

Department of Pulmonary and Critical Care Medicine,
Ajou University School of Medicine, Suwon, Korea

서론

임상진료지침은 “특정 임상 환경에서 적절한 진료에 대한 의료진의 결정을 돕기 위해 체계적으로 개발된 지침”을 의미한다(1). 임상진료지침은 다양한 목적을 가지며, 그 중 대표적인 내용은 다음과 같다(2).

- 근거중심의 명확하고 접근가능한 표준을 제정
- 외래 및 입원 환자를 대상으로 보다 용이하고 객관적으로 임상적 의사결정을 내림
- 전문가의 업무수행을 평가할 수 있는 척도를 제공
- 업무의 분담을 제정(예: 일차 의료의 대 전문의)
- 최신의 가장 적절한 의료에 대한 환자 및 의료인 대

* 교신저자 : 신승수, 경기도 수원시 영통구 원천동 아주대학교 의과대학 호흡기내과학교실
Tel) 031-219-5122, E-mail) sssheen@ajou.ac.kr

상 교육

- 보건의료 서비스의 비용-효과를 개선
- 의료를 대상으로 하는 외부의 통제 도구로 이용

양질의 진료를 위해 임상진료지침의 도입과 적극적인 활용이 필요하다는 인식이 의료계 내외에서 증가하고 있으나, 이를 지나치게 남용하고 의존하는데 대한 우려의 목소리 또한 크다. 일선 의사들이 임상진료지침에 대해 가지는 우려의 핵심은 근거중심의학이라는 대명제를 앞세워 정치가 및 보건서비스의 관리자들이 환자의 치료라기보다는 질병의 치료를 강제하기 위해 이를 이용하는 것이 아닌가라는 점이다. “평균적으로” 치료 효과가 있다는 근거에 영합하여 환자에 대해 판단 내리려는 태도는 경계할 필요가 있다. 이 밖에도 임상진료지침이 가질 수 있는 단점들은 다음과 같다(2).

- 학문적으로 의심의 여지가 있고 문제 있는 의료를 “전문가의 의견”이라는 이름으로 공식화시킬 가능성이 있음
- 의료 내용의 다양성이 감소함으로써 최고의 치료보다는 “평균적인” 치료로 표준화될 가능성이 있음
- 개별 환자들이 분리되어 민감하게 다루어지지 못하게 하고 의학의 혁신을 저해함
- 이런 상황에서 유능한 의사라면 “이것은 시행하였을 것”이라는 법적 강요의 용도로 사용될 가능성이 있음
- 국가적 또는 범지역적 차원에서 개발된 임상진료지침은 특정 지역의 요구를 반영하지 못할 수 있고, 지역의 의사들을 종속화시킬 수 있음
- 전문 의료기관 환자를 대상으로 개발된 임상진료지침은 일차 진료환경 환자의 인구학적, 임상적, 치료적

인 면의 차이점을 반영하지 못할 수 있음

- 임상진료지침이 서로 다른 전문가집단 사이의 힘의 균형(예: 임상 대 의학연구자, 의료구매자 대 공급자)을 바람직하지 않은 방향으로 기울게 하는 결과를 빚을 수 있어, 임상진료지침의 개발이 정치적인 행위로 인식될 수 있음

모든 임상진료지침이 애초에 목적인 임상적인 개선을 달성하는 것은 아니며, 임상진료지침에 의한 결과의 편차는 매우 큰 편이다. 효과적인 임상진료지침이 되기 위해서는 실제로 임상진료지침의 개발전략(어디서, 어떻게 개발되었는가?), 보급전략(어떻게 임상외사의 관심을 끌 수 있는가?), 실천전략(어떻게 임상외사가 이 지침을 따라하게 하는가?)과 같은 세 가지 요소를 고려해야 한다. 이상 세 요소의 측면에서 본 효과적인 임상진료지침의 기준은 다음의 표 1과 같다(3). 이에 따르면, 가장 효과적인 임상진료지침이란 실제로 이를 이용할 지역의 사람들에 의해 개발되고, 특별한 교육 방법으로 소개되고, 의뢰와 동시에 특정 환자에 적합한 해답이 바로 제시될 수 있는 것이다.

다른 모든 논문들과 마찬가지로 임상진료지침 역시 표준화된 양식으로 제시된다면 평가가 용이하다. 그러나 아직까지 이런 형식을 모두 제대로 지키는 임상진료지침은 그리 많지 않다. 임상진료지침이 갖춰야 할 바람직한 형식의 목록은 다음과 같다(4).

- 목표(objective) - 보건의료상의 문제와 대상환자, 의료공급자, 의료환경 등을 포함하는 임상진료지침의 일차 목표
- 선택권(options) - 임상진료지침을 도출하는데 고려할 수 있는 임상적으로 선택가능한 진료내용들

표 1. 효과적인 임상진료지침의 기준(3)

효과	개발전략	보급전략	실천전략
높음	내부 개발	특별교육(예: 문제중심학습의 일부로)	의뢰 즉시 환자 지향적인 해답을 제시
평균 이상	내부, 외부 함께 개발	보수교육(예: 임상강좌)	환자 특이적인 피드백
평균 이하	외부, 지역적	대상자에게 우편 홍보	일반적인 피드백
낮음	외부, 국가적	논문 게재	일반적인 조언

- 결과(outcomes) - 대안적인 진료들의 비교에 이용할 수 있는, 의미있는 보건의료 및 경제학적인 결과들
- 근거(evidence) - 언제, 어떻게 근거가 수집, 선택, 완성되었는가
- 가치(values) - 선택가능한 임상진료에 따른 각각의 결과에 대해 어떻게 가치가 매겨졌으며, 이 작업에 누가 참여하였는가
- 편익(benefits), 위해(harms), 비용(costs) - 임상진료 지침을 실천함으로써 환자들에게서 기대할 수 있는 편익, 위해, 비용의 종류 및 규모
- 권고(recommendations) - 핵심 권고사항의 요약
- 타당화(validation) - 외부 평가, 다른 임상진료지침과의 비교, 지침이용의 임상적인 검정에 대한 보고
- 후원자(sponsors) - 임상진료지침의 개발, 재정지원, 확인과정에 참여한 사람들을 거명

임상진료지침의 개발 방법

임상진료지침의 개발을 위해서는 해당 분야에 대한 학문적, 임상적 지식 이외에도 다음과 같은 다양한 분야에 대한 지식과 기술이 필요하다(5).

- 문헌 검색
- 역학
- 통계학
- 보건의료 서비스 연구
- 임상적 경험
- 그룹관리
- 문헌 작성 및 교정

임상진료지침을 개발하기 위해 사용되는 방법은 개발의 주체가 되는 기관이 가진 고유한 철학과 접근법에 따라 차이가 있다. 이 글에서는 임상진료지침의 개발을 위한 방법을 과거 Woolf가 제안한 바 있는 구분법인 1) informal consensus development, 2) formal consensus development, 3) evidence based development, 4) explicit evidence based development, 이렇게 네 가지를

중심으로 설명하도록 한다(6).

1. Consensus Based Development Method

임상진료지침의 개발을 위해 전문가의 전반적이고 주관적인 합의(consensus)를 이용하는 방법을 의미하며, 가장 많이 사용하는 방법이다. 고전적인 합의모형은 전문가집단을 교류시켜 모으고 특정 상황의 진료에 대해 합의를 얻어낸다. 다른 엄격한 임상진료지침의 개발방법에 비해 빠르고 저렴하다는 장점이 있다. 또한 전문가의 의견이 임상 의료 내용의 주를 이루는 전통적인 치료와 수련 모습과 잘 들어맞는다. 하지만 이 방법에는 중요한 문제가 존재한다.

임상진료지침의 강점은 지침의 개발에 참여한 사람들의 바이어스효과를 얼마나 잘 차단하였는가와 관련된다. 이 방법으로는 참여자의 바이어스가 합의과정에서 제거될 수 없기 때문에 신뢰도가 떨어진다. 만일 임상진료지침을 합의과정에 다른 사람들을 참여시켰다면 틀림없이 근본부터 다른 임상진료지침이 나왔을 가능성이 크다. 전문가의 의견이란 상대하는 전문가가 누구냐에 따라 100% 달라질 수 있으며, 전혀 다른 답을 들을 수도 있는 것이다. 보통 다년간의 경험과 특정 분야의 많은 저술을 근거로 선정되는 전문가 참여자는 치료에 대해 상당한 바이어스를 가지고 있을 가능성이 크다. 결국 이들 참여자의 좋은 의도에도 불구하고 이들이 자신의 임상 및 연구 경험을 무시한 채 특정 처치의 편익과 위해, 비용에 대해 정확히 평가하기란 불가능하다. 또한 formal 및 informal consensus 방법 모두 근거와 명확히 연계된 권고를 필수적으로 요구하는 것은 아니기 때문에 권고 가운데 어떤 것이 강력한 경험적인 근거에 바탕을 둔 것이고 어떤 것이 단순히 전문가들의 의견에 바탕을 둔 것인지를 감별하기란 쉽지 않다.

1) Informal Consensus Development

수십 년간 임상진료지침의 개발자들은 거의 대부분 전문가의 의견을 바탕으로 만들었다. 임상진료지침은

전문학회, 정부기관, 전문가로 구성된 위원회에서 토론을 통해 합의에 이르렀으며 때로는 단 한 번의 모임만으로 권고안이 만들어지는 경우도 있었다. 참여자가 무엇을 권고할 것인지를 결정하기가 용이한 접근법을 informal consensus development(혹은 global subjective judgment)라고 부르는데, 이는 이 같은 과정을 거쳐서 내려진 권고가 잘 정의되어 있지 않은 경우가 흔하기 때문이다. 보통 지침에는 단지 권고만이 담겨 있을 뿐, 이 같은 권고가 나오기까지의 일련의 과정에 대해서는 별로 언급되어 있지 않다. 고찰면에 과학적인 근거가 실리기도 하나 바이어스의 개입 없이 검토가 이루어졌는지, 권고에 영향을 미친 것이 단지 근거만이인지를 독자들이 판단할 수 있도록 도와주는 방법론상의 정보는 부실하다.

하지만, informal consensus development는 여전히 임상진료지침의 가장 흔한 개발방법이다. 이 방법은 전문가들이 특정 주제에 대해 잘 알고 있는 경우 합의를 보다 신속히 모을 수 있고 근거와 근거 사이에 놓인 공백을 연계할 수 있는 연구경험이 있기 때문에 단순하면서도 유연성이 있다. 또한 복잡한 분석방법이 필요 없으며, 정형화된 분석 방법에 익숙하지 않은 패널 구성원에게도 수용이 용이하다는 장점이 있다.

이 방법의 문제점은 이 같은 방법에 의해 만들어진 지침의 질이 그리 좋지 못하다는데 있다(6). 첫째, 적정성을 규정하는 근거로 전문가의 의견을 적용하는 것이 근본적인 타당도가 의심된다. 즉, 특정 처치가 편익이 있을 것이라고 패널의 구성원들이 생각한다고 해서 실제로 그 같은 일이 사실이라는 것을 보장해주지는 못한다. 둘째, 어떻게 합의를 얻어내는가가 명확하게 규정되어 있지 않으므로 의문이 발생할 수 있다. 체계적인 방법 없이 모임에서 만들어지는 지침은 구성원의 역할, 주도적으로 거리낌 없이 말하는 개인, 조직의 정책 등에 따라 쉽게 영향 받는다. 셋째, 규정된 방법의 부재로 인해 권고가 과학적인 근거에 의한 것인지 패널의 바이어스가 근거를 잘못 해석한 부분은 없는지를 독자가 판단하기 어렵게 만든다. 결국, 합의의 과정이 단지 전문가

그룹의 바이어스를 공식화하고 조장하는 것으로 그칠 수 있고 설령 완전한 바이어스는 아니라 하더라도 임상진료지침이 의도하고 있는 이용자보다는 특정 전문가의 좁은 시각을 반영하는데 그칠 수 있다.

설령 패널이 바람직한 분석방법을 적용했다 하여도 그 방법이 제대로 기술되어 있지 않다면 독단적인 권고로 비칠 수 있다. 예를 들어, 민간 보험회사가 이용하는 지침의 분석방법은 매우 복잡한 편이나 이들은 자신의 영업상의 비밀을 보호하기 위해 구체적인 방법을 제시하지 않기 때문에 그 신뢰성이 의문이 생길 수밖에 없다. 마찬가지로 전문학회가 자신들의 수입증가에 도움이 되는 특정 처치를 권유하였을 때, 권유에 이르기까지의 방법과 원칙이 기술되어 있지 않다면 이해관계의 측면에서 의심을 품어보아야 한다.

2) Formal Consensus Development

1970년대부터 합의를 구하기 위한 보다 정형화된 방법이 미국에 소개되기 시작하였다. 1977년 미국의 National Institutes of Health(NIH) Consensus Development Program이 이 방법을 도입하였다(7). 바이어스의 발생가능성을 줄이기 위해 NIH 패널에는 방법론의 전문가, 임상 의사, 일반대중의 대표자 등이 망라되어 있으나 “특정 주제에 대해 강력한 옹호자로 알려진” 인물은 배제시킨다. 각 영역의 의견을 대표하는 전문가들이 공개된 자리에서 근거를 발표하고 토론하도록 초청받는다. 만 이틀 만에 걸친 모임을 통해 전문가 패널은 권고에 대한 합의를 도출한다. 지침은 총회와 공개 토론 후에 비공개회의를 통해 개발되며, 제3일째에 청중과 대중에게 발표된다. 합의에 이르기까지 패널리스트가 밤을 새워가며 작업하는 경우도 종종 있었다. 이 방법은 약 100여개 이상의 주제에 적용되었다. 비록 이 같은 방법이 informal 방법에 비해 구조화된 것이기는 하나 명확한 기준의 부재와 한번의 모임만으로 쫓기듯이 권고안을 만들어내야 하는 부분에 대해 비판이 일어났다.

1982년 이후 American Medical Association

Diagnostic and Therapeutic Technology Assessment 프로그램에서는 전문가 의견을 모으는 간단한 방법을 통해 적절성 평가를 수행하기 시작하였다(8). 이 방법에서는 1700명 이상의 연락가능한 전문가의 목록을 관리하고 있다. 특정 주제가 정해지면 관리중인 관련 전문가에 설문지를 발송한다. 설문지는 패널리스트에게 특정 처치의 안전성과 효과에 대해 “확립됨”, “기대해볼만함”, “실험적임”, “의심스러움”, “수용불가능” 가운데서 의견을 선택하도록 만들어져 있다. 결과는 다시 우편으로 수집되어 계산되고 JAMA에 실리게 된다. 이 결과물은 진정한 임상진료지침이라기 보다는 포괄적인 적절성 분류법이라고 말하는 것이 더 적절하다. 우편을 이용한 방법은 전문가 사이의 의견교환 없이 단순한 투표행위에 그치는 것이므로 진정한 의미의 합의라고 말하기는 어렵다.

1980년대 RAND 협회에서는 합의에 도달하기 위한 좀더 정형화된 방법을 도입하였다(9). 전문가 패널은 특정 처치에 대한 수백에서 수천에 이르는 잠재적인 적응증이 적힌 목록 외에 이의 과학적 근거를 살핀 배경논문을 제공받는다. 이어서 2단계 델파이 기법이 뒤따른다. 첫 패널 모임 이전에 패널 구성원은 각각의 적응증에 대한 특정 처치의 적절성을 질문 받는다. 적절성의 평가지표는 9점 척도로 되어 있는데, 1점은 “전혀 적절하지 않음”, 5는 “애매함”, 9는 “매우 적절함”의 의미를 가진다. 패널 모임에서 목록상의 각각의 적응증을 살피고, 그룹에 의해 매겨진 각 적응증별 점수의 분포를 검토하고, 자신의 판단과 전체적인 점수 사이의 차이를 비교하며, 불일치의 원인에 대해 토론한다. 그 다음 패널 구성원은 모임에서의 토론을 근거로 애초에 자신이 매겼던 점수를 수정한다. 적절성 점수의 마지막 목록은 각각의 적응증의 적절성에 대한 패널리스트의 합의의 정도를 반영한다. 이 방법은 근래 들어 많이 이용되기 시작하였으며, 이에 의해 만들어진 적절성 점수는 민간회사, 병원, HMO, 대규모 보험회사 등에서 이용되고 있다. 이 방법의 제한점은 이 것의 결과물인 적절성 점수 목록을 의사들이 실제 임상에 적용하기엔 어렵다는 점이다. 또한 의

견을 적절성을 규정하는 바탕으로 삼고 합의를 도출한다는 것 역시 근본적인 제한점으로 지적된다. 물론 패널 내에서 과학적인 근거가 언급되기는 하지만 지침상에는 근거의 질과 권고 사이가 명확하게 연계되어 있지 않다.

2. Evidence Based Method

1980년대 몇몇 미국 및 국제 기관에서 임상진료지침을 좀더 과학적인 근거와 연계시키려는 노력을 하기 시작하였다. Canadian Task Force on the Periodic Health Examination(CTFPHE), U.S. Preventive Services Task Force(USPSTF), American College of Physicians(ACP)의 Clinical Efficacy Assessment Program 등이 대담가능한 특정 질문을 정의하고, 근거의 선정기준을 제정하고, 적절한 근거를 검색하고 평가하기 위한 체계적 과정을 규정한 근거중심 접근법을 이용하여 일련의 임상진료지침을 만들어내기 시작하였다. 이 과정에서 패널은 모아진 근거의 강도가 명확히 반영된 권고를 제시한다. 이 같은 접근법은 임상진료지침의 신뢰도와 타당도를 개선함으로써 신뢰를 얻었으며, 뒤이어 다른 그룹에서도 많이 채택하게 되었다. 그렇지만 다른 그룹이 많이 채용함에 따라 근거중심 접근법이 여러 다른 중요한 결과를 계량화하기 보다는 특정 결과(예: 질병특이사항률)에만 집중되는 결과를 낳기도 하였다(10). 근거중심방법에서 한 단계 더 나아가 기반을 둔 explicit 개발법도 제안되었는데, 이는 특정한 한 가지 결과만이 아니라 가능한 모든 중요한 임상 결과에 대해 특정 처치의 효과를 확률적, 체계적으로 추정하는 방법이다(11,12).

1) Evidence Based Development

기존의 주관적인 합의 중심 지침에 대응하여 권고와 근거의 질을 연계시킨 방법을 채용하려는 노력의 일환으로 1980년 ACP의 Clinical Efficacy Assessment Project에서는 전문가 자문위원을 통해 근거를 평가하고 이를 바탕으로 지침을 개발하는 방법을 채택하였다.

Annals of Internal Medicine에 실리는 지침은 그 지침이 근거하고 있는 구체적인 과학적 근거가 담긴 배경논문이 함께 수록되었다. 이후 이 지침은 Blue Cross and Blue Shield Association의 의료비지출 정책에 반영되었으며, Joint American College of Cardiology/American Heart Association Task Force on Assessment of Cardiovascular Procedures 같은 다른 전문학회에서도 비슷한 접근법을 채택하기 시작하였다(6).

임상 예방 서비스에 관한 지침의 개발에 권고와 근거의 보다 명확한 연계를 도입한 사례로서, 1979년 CTFPHE는 선별검사 및 기타 예방 서비스에 대한 권고에서 근거의 질을 반영하기 위해 권고의 강도에 등급을 매기기 시작하였다. 근거의 강도는 연구설계의 종류에 따라 구분되었다. 이후 지침을 개발하는 다른 패널들도 이 같은 근거구분의 법칙을 채택하였다. USPSTF는 1984년 설립 당시부터 비슷한 근거법칙을 채택하기 시작하여 1989년의 첫 보고서에는 이를 확대 반영하였으며 이후의 논문에서는 임상적 효과의 평가를 하는데 있어 더욱 명확한 기준과 체계적 방법을 채택하였다.

비록 근거중심 접근법이 임상진료지침 개발의 과학적인 엄격성을 제고시킴으로써 신뢰를 얻었지만, 근거를 찾을 수 없어 적절한 권고를 제시하기 어려운 경우가 흔히 생긴다는 문제점이 발생하였다. 처치를 긍정도 부정도 하지 않는 중립적인 권고 또한 문제가 된다. 비판

가들은 근거중심 방법에 지나치게 경도되어 있으면 현재 임상연구를 통해 타당성이 증명된 일부 처치를 제외한 나머지 대부분의 의학적 처치는 배제되고 말 것이라고 주장한다. 따라서 지침 개발 패널의 많은 수가 근거중심 접근법과 전문가 의견의 수렴을 통합한 방법을 채용하고 있다. 대신 지침 내에서 어떤 권고가 과학적 근거를 바탕으로 한 것이고 어떤 권고가 전문가의 의견을 바탕으로 한 것인지를 명확히 기술할 것을 요구한다.

2) Explicit Evidence Based Development

이 방법에서는 개발자들이 대상 처치의 편익, 위해, 비용을 규정하고 각각의 결과의 발생가능성을 명확히 추정한다. 추정을 위해서는 가능한 한 과학적 근거와 정형화된 분석방법을 이용해야 하고, 부득이한 경우 전문가의 의견을 이용하기도 하지만 대신 이 같은 추정의 출처를 명확히 기술해야 한다. 이 같은 결과중심(outcome-based) 접근법에 그치지 않고, 환자의 선호도까지 융합시키는 선호도중심(preference-based) 접근법을 채택할 수도 있다. 마지막으로 이상의 각 가정에 따른 결과를 표 2와 같은 형태의 “대차대조표”에 실어 환자, 의사, 정책관련자들이 각 선택에 따른 잠재적인 편익, 위해, 비용을 살필 수 있게 해준다(12).

최근 들어 이 방법은 채택하는 기관과 학회가 늘어나고 있기는 하나, 아직까지는 비교적 새로운 방법이고 따

표 2. 70세 이상 남자를 대상으로 한 전립선특이항원(PSA) 선별검사의 편익과 위해의 대차대조표(선별검사 매 100,000명당)(10)

편익과 위해		명	증가한 인년	개선된 인년
추정 편익	사망 예방	449	3,019	829
	전이 예방	1,019		
추정 위해	선별검사 양성 또는 직장수지검사	34,000	-245	-3,229
	조직 생검	23,460		
	광범위 전립선절제술	3,360		
	-사망	22		
	-실금	301		
	-음위	1,569		
-실금+음위	467	-5,050		
계			2,774	-24,358

라서 지침의 개발에 이 방법을 적용한 경험은 그렇게 많지 않다(10). 비판자들은 시간과 여력이 없는 지침 개발자들이 이 같은 복잡한 분석을 수행하기란 거의 불가능하다고 말한다.

임상진료지침의 개발 과정

임상진료지침을 개발하는 과정에서는 많은 최근의 문헌의 검토가 이루어지게 되며, 주제의 선정부터 임상진료지침의 문장을 다듬기까지의 일련의 과정이 사전에 규정되어 있다. 그 대략적인 단계는 다음과 같다(5).

- 대상주제와 이용자 선정
- 패널의 구성 및 운영
- 근거의 파악 및 평가
- 근거를 임상진료지침으로 변화
- 임상진료지침의 검토와 갱신

1. 대상 주제와 이용자 선정

임상진료지침이 다루기 적절한 주제를 선정하기 위해서 사전에 고려해야 하는 몇 가지 선정 기준이 존재한다. 비록 대부분의 기관은 이 같은 기준의 일부를 준수하고는 있으나, 모든 기준을 철저히 지키는 곳은 거의 없는 형편이다. 예를 들어, 전문학회는 회원들의 특수한 관심사와 요구를 반영하기를 바라는 반면, 행정기관에서는 다른 정치적인 고려를 가지고 있기 마련이다. 다음은 임상진료지침 개발을 위한 대상 주제 선정시 고려해야 하는 기준들의 목록이다(2).

- 대상 주제의 유병률, 위험도, 비용이 큰가?
- 임상진료상의 편차가 크거나 설명되지 않는 부분이 있는가?
- 대상 주제가 환자진료상의 과정과 결과의 측면에서 중요한 것인가?
- 개선의 여지가 있는 것인가?
- 시간 및 금전적인 투자만큼의 가치가 있는 것인가?
- 연구진 모두가 관심을 가질만한 대상 주제인가?

- 의견의 일치가 있는가?
- 변화가 환자에게 편익을 안겨줄 것인가?
- 변화가 실천가능한 것인가?

이상의 평가를 통해 대상 주제가 정해지면 주제와 관련된 근거를 평가하기 이전에 대상 주제를 세부적으로 재정의해야 할 필요가 있다. 이런 과정이 생략된다면, 범위가 되는 임상적 상태나 질문이 너무 방대해질 수 있다. 예를 들어, 단순히 당뇨의 지침이라고 한다면 일차, 이차 및 삼차의료의 영역 모두를 포함할 수 있고, 선별 검사, 진단, 식이요법, 약물요법, 위험요인관리, 전문의 자문을 요하는 적응증 등 치료의 다원적 요소 모두를 포함할 수 있다. 지침 안에 이들 모두를 넣는 것이 가능할지는 모르나, 이를 위해 들여야 하는 시간과 비용상의 노력은 막대할 것이다. 따라서 개발자들은 지침 안에 담을 내용의 범위를 명확하게 해야 한다. 근거중심적인 개발방법도 많은 노력을 필요로 하며, 특히 이 같은 지침의 개발 경험이 부족한 연구자들은 이를 위해 필요한 자원을 과소추정하는 오류를 흔히 범한다.

대상 이용자와 임상진료지침이 의도한 임상환경을 규정하는 것 역시 중요한 부분이다. 특정 임상환경과 대상 이용자를 고려치 않은 임상진료지침이 효과적으로 이용되기를 기대하기란 어렵다. 실제로 일차진료를 주제로 하여 Centers for Disease Control and Prevention (CDC), American Cancer Society (ACS), NIH에서 개발한 임상진료지침 중 실천에 필요한 비용, 시간이나 기술, 경쟁관계에 있는 보다 흔한 문제에 대한 수요, 환자의 선호도와 같이 임상진료지침이 수용되고 실천될 것인지를 결정짓는 핵심 요소에 관심을 두지 않은 경우가 많이 있었다(10).

2. 패널의 구성 및 운영

물론 근거중심방법이 임상진료지침의 일관성과 신뢰성을 향상시킬 수는 있으나 임상진료지침의 패널 구성은 여전히 과정과 결과물에 큰 영향을 미치는 요소이다.

패널의 크기와 구성은 과학적, 현실적, 정치적 관심 사이에서 균형을 유지해야 한다. 임상진료지침의 패널은 보통 10명(최소 6명)에서 15명(최대 20명) 사이로 구성된다(5,10). 이 정도 규모 내에서 그 크기가 작을수록 의사결정이 용이하고 비용이 절감되나 다양한 의견교환은 덜 이루어지며, 클수록 효과적인 그룹 운영이 어렵다. 광범위한 임상범위를 다뤄야 하는 것과 같은 특수한 경우에는 전체를 대표하는 그룹과 기능적인 소그룹 운영의 필요성 사이에서 저울질 해보아야 한다. 특화된 전문가와 일반의사의 상대적 분포 또한 중요한 고려사항이다. 해당 분야의 전문가는 과학적 고찰과정에 대한 중요한 통찰력을 제공하고 임상진료지침 결과물의 신뢰성을 제고시키는 중요한 원천이다. 반면, 동시에 전문가는 자신이 개입하고 있는 특정 치료나 연구에 의해 바이어스를 가질 가능성이 크다. 같은 근거를 놓고도 특정 처치를 선호하는 입장으로 바이어스가 있는 단일한 전문가 그룹은 각 분야의 전문가로 이루어진(multidisciplinary) 그룹과는 다른 결론을 내릴 가능성이 있다. 예를 들어, 혈관외과 의사만으로 이루어진 그룹은 외과 의사와 내과 의사가 고르게 참여한 그룹에 비해 경동맥수술을 더 선호하는 경향을 보일 것이다. 이런 경향을 막기 위해 패널에는 방법론의 전문가, 일반의사, 의사가 아닌 의료전문가, 환자나 소비자 대표 등을 포함시킴으로써 균형을 유지하고자 한다. 이는 권유를 만들어 내는데 있어 근거의 유무에 대한 충분한 토론을 이끌어내기 위해 중요하다. 개인의 바이어스는 각 분야의 전문가들이 모임으로써 균형이 유지될 수 있으며, 이 같은 균형의 결과로 보다 타당한 임상진료지침이 도출될 수 있다. Agency for Health Care Policy and Research(AHCPR)의 패널에는 처음부터 의사가 아닌 사람이 참여하고 있으며, 근래 들어 USPSTF, CTFPHE, ACP 패널에도 방법론의 전문가인 일반 의사의 참여를 강조하고 있다(10). 또한 임상진료지침의 초고를 대상으로 다차원적인 동료간심사를 함으로써 전문가와 각 분야의 대표로 구성된 패널의 효과를 극대화하기 위해 노력한다. 그렇지만, 각 분야의 대표자들이 참여하는 방식은 모든 패널리스트들이 전

과정에 동등하게 참여하거나 합의의 도출에 이르는 것을 어렵게 만들기도 한다.

패널 내의 각 구성원은 그룹의 리더, 전문가 자원, 기술적인 지원, 행정적인 지원 등 각 분야마다 필요로 하는 역할을 맡는다. 그룹의 구성원은 자신이 몸담고 있는 특정 분야를 대표하여 참여하며 이용가능한 근거와 임상진료에서의 자신의 지식을 바탕으로 권고를 개발하는 역할을 담당한다. 그룹 리더는 그룹의 기능을 원활하게 하고(group process) 그룹이 목표를 달성하도록(group task) 독려하는 역할을 담당한다. 이 임무는 임상환경과 의학적 지식 모두에 익숙한(그렇다고 옹호자라는 의미는 아니다) 사람이 담당하는 것이 바람직하다. 토론을 독려하고 자신의 의견을 다른 그룹구성원들에게 강요하는 것이 아니라 진정한 합의를 이끌어내는 능력이 필요하다. 이를 위해서는 임상적인 기술뿐만 아니라 그룹을 상대하는 기술도 필요하다(5). 그룹모임에서는 비정형화된 방법보다는 정형화된 방법(예: 델파이 기법)을 시도하는 것이 더 나은 결과를 도출할 수 있을 것이다(13).

3. 근거의 파악 및 평가

근거를 파악하고 평가하는 최적의 방법은 체계적 종설(systematic review)이다. 체계적 종설의 목적은 이용가능한 근거를 모으고 고려대상 임상질문에 대한 각 근거의 잠재적인 적용가능성을 평가하며, 근거의 바이어스 가능성을 살피고, 관찰한 결과를 요약하는 것이다. 관심 임상질문을 파악하는 것은 적합한 근거의 기본적인 영역을 설정(연구설계, 출판년도 등)하는 것을 돕는다. 예를 들어, 질문이 처치의 효능에 대한 것이라면 검색 대상이 되는 연구는 주로 무작위임상시험이 될 것이며, 질문이 예후에 대한 것이라면 검색 대상 연구는 주로 코호트연구가 될 것이다. 근거를 수집하는 첫 번째 단계는 기존에 출판된 적당하고 최신의 체계적 종설이 있는가이다. 코크란 연구자 그룹과 접촉하여 해당 주제에 대한 검토가 진행 중인지 알아본다.

만일 최근의 체계적 종설을 발견할 수 없다면 메드라

인 검색부터 시작하도록 한다. 메드라인 검색에서는 연구의 종류에 따라 적절한 검색조건이 존재한다. 예를 들어, 치료효과에 대한 최고의 근거는 무작위임상시험이며, 위험요인에 대한 최고의 근거는 전향적 코호트연구이다. 코크란 라이브러리의 일부인 코크란 대조시험 등록부에는 데이터베이스와 수작업검색을 통해 찾은 218,000건이 넘는 임상시험이 수록되어 있는데, 이들도 검색해 보아야 한다. 논문의 참조문헌을 살펴보면 단순 컴퓨터 검색만으로는 찾지 못했던 다른 논문들을 추가로 찾아낼 수 있으며, 해당 분야의 전문가에게 찾아낸 문헌의 목록을 보여주어 누락된 중요한 문헌이 있는지를 확인한다. 특정 질병의 정보가 담긴 정부 문서와 보고서도 메드라인에는 누락되어 있기 때문에 검토자는 해당 질병의 연구나 관리프로그램을 담당하고 있는 정부부처와 접촉해볼 필요가 있다. 영어로 기술되지 않은 연구 검색이나 특정 분야에 특화된 데이터베이스 검색, 관련 학술지의 수작업 검색, 출판되지 않은 자료의 검색 등을 통해 미처 못 찾은 연구들을 추가로 찾을 수 있으나 이를 위해서는 적지 않은 노력과 비용이 소요된다. 여러 검색 전략의 비용-효과에 대한 고찰은 아직까지 충분하지 못하며, 검색범위를 정하기 위한 전략을 세우는데 있어 이용가능한 자원을 고려하여 시행하는 것이 가장 바람직하다(5).

연구들을 수집한 후에는 관심 임상질문의 대상으로 적절한지와 바이어스의 가능성 측면에서 근거를 평가한다. 초록을 살펴보는 것만으로도 연구를 좀더 세부적으로 평가할 가치가 있는지를 판단하게 도와준다. 막연한 기준보다는 구체적인 기준을 이용하는 것이 이 과정의 신뢰성을 향상시키는데 도움이 된다(14). 근거중심방법의 핵심이면서도 논쟁의 중심에 위치한 단계가 연구결과의 내적, 외적 타당도의 평가를 포함하는 개별 연구들의 질 평가과정이다. 내적 타당도는 연구설계와 이의 수행과 관련된 지표이다. 비록 무작위 대조 임상시험과 같은 특정 연구설계가 바이어스에 덜 취약하다고는 하나 충분치 않은 검증력, 적절하지 못한 설계나 분석, 불충

분한 추적관찰, 부적절한 결과변수의 선택, 효과적이지 못한 처치의 시행 등과 같이 개별 연구의 내적 타당도를 저해할 수 있는 다른 요인 또한 많이 존재한다. 교란변수, 선택 바이어스, 측정상의 오류는 관찰연구에서 중요하게 고려해야 하는 부분이다. 또한 중요한 것으로서 개별 연구의 결과가 일반 인구집단에도 적용가능한가의 여부 즉, 외적 타당도를 들 수 있다. 대표성을 갖추지 못한 연구대상 인구나 임상 환경, 실천불가능한 처치를 담고 있다면 아무리 잘 시행된 연구라 할지라도 외적 타당도는 낮을 수밖에 없다. 설령 연구가 좋은 효능(이상적인 환경에서 결과)을 보인다는 결과가 나온다고 해도 실제 환경에 적용해 보면 낮게 나오기 쉽다. 때로는 결정적인 근거가 부족한 중요 질문에 대해 전문가의 의견을 모아야 하는 경우가 있다. 이를 위해서는 비정형화된 방법과 정형화된 방법(예: 델파이 기법) 모두 이용가능하다. 그러나 이후 권고를 만들 때에는 의견에 바탕을 둔 권고와 근거에 바탕을 둔 권고를 명확히 구별하여 제시해야 한다.

수집된 연구 중 적절하다고 판단되는 연구를 대상으로 특정 처치의 근거를 요약한다. 이를 위해 대상 처치 및 경쟁관계에 있는 다른 처치를 편익, 위해, 비용 등의 측면에서 비교해 볼 필요가 있다. 가능하면 여러 연구의 결과를 요약하기 위해 메타분석을 이용해 볼 수도 있다. 이 같은 과정을 통해 편익(예: 이환율과 사망률의 감소, 삶의 질 증가)과 위해(예: 치명적 또는 비치명적인 부작용)의 요약 추정치를 대차대조표를 통해 비교할 수 있다. 특정 결과에 대해 환자의 선호도를 추정할 수 있다면, 결과를 질보정여명(QALY)같은 공통적인 지표로 변환하여 특정 전략의 전반적인 효과(편익과 위해의 차)를 계산할 수도 있다. 물론 치료가 사망률에 미치는 효과는 명확한 반면 부작용은 미미하다면(예: 울혈성 심부전에 대한 안지오텐신 전환효소 억제제), 세세한 비교표가 불필요할 수도 있다. 그렇지만 치료가 주로 증상을 개선시키는 역할에 그치고 오히려 적지 않은 위험부담을 감수해야 하는 것(예: 전립선비대증 수술)이라면 대차대조표

를 통한 비교가 필요하다. 비용-효과분석을 시도하는 경우도 있으나, 불행히도 이 같은 모형 수립의 복잡성이나 각각의 지표들을 추정하는데 필요한 다양한 가정들로 인해 도출된 결론에 대해 일반적인 임상사의가 가지는 신뢰감의 정도는 낮은 편이다. 특히 이 같은 결과가 실제 임상환경에서 우세하게 행해지는 행위에 반하는 근거를 제시하는 경우 더욱 심하다.

요약된 근거를 바이어스에 대한 취약성의 정도에 따라 범주화한다. 이는 “근거의 강도(strength of evidence)”를 임상진료지침의 독자에게 전달할 수 있는 가장 신속한 방법이다(15,16). 범주화를 위한 분류법은 여러 가지가 있으나 연구설계에 따른 분류법이 효과를 평가하기에 가장 적절한 방법이다. 다음의 표 3은 지침에 실리는 근거의 분류표 예이다(5). 지침의 개발자는 근거가 있는지를 살피기 위해 명확한 몇 개의 기준을 이용하여야 한다.

4. 근거를 임상진료지침으로 변화

현실의 의료현장에서 강력한 근거를 가진 채 이루어

지는 처치란 거의 없다고 보아도 틀리지 않다. 따라서 강력한 권고를 가진 지침이 활용될 수 있는 여지는 그리 넓지 않다. 예를 들어, 특정 질병에 대해 지침에서 강력하게 권고하는 치료라 할지라도 임상적, 보건학적, 정책적, 경제적 측면에서 재해석할 필요가 있다. 따라서 근거를 권고가 담긴 지침으로 변환하기 위한 첫 번째 단계는 수집, 요약된 근거를 어떻게 해석할 것인지에 대해 패널 간의 의견을 구하고 모으는 작업이며, 이를 위한 구체적인 규정을 사전에 마련해야 한다. 근거를 해석하는 과정에서 소규모 무작위 임상시험이나 대조 관찰연구 등에서 나온 근거를 어느 정도까지 인정해야 하는가와 특정 인구집단을 대상으로 삼은 연구의 결과를 지침의 대상이 되는 인구집단에 적용(외삽)하는데 문제는 없는가와 같은 일반화가능성의 문제를 평가하기 위해 의견이 필요하다. 단지 임상적인 판단과 경험에 근거하여 권고를 하는 것은 바이어스와 개인적인 취향에 취약할 우려가 있다. 따라서 전문가의 의견이 담당할 역할의 한계를 정한 다음에는 이를 어떻게 수집하고 평가할 것인가를 정해야 한다. 어떤 방법이 이를 달성하기 위한 최적의 방법 인지는 결론이 없으나 가능한 한 그 과정을 명확하게 규정하는 것이 필요하다. 과학적인 근거와 전문가의 의견

표 3. 근거범주의 제시 예(5)

근거의 범주	
Ia	무작위대조시험의 메타분석
Ib	단일 무작위대조시험
IIa	무작위화가 이루어지지 않은 대조시험
IIb	하나 이상의 다른 형태의 준실험연구
III	관찰연구(예: 비교연구, 상관성연구, 환자-대조군연구)
IV	전문가위원회의 보고서나 의견, 권위자의 의견이나 경험

표 4. 권고의 강도 제시 예(5)

권고의 강도	
A	I 등급 근거에 직접적으로 기반
B	II 등급 근거에 직접적으로 기반 또는 I 등급 근거로부터 유추된 권고
C	III 등급 근거에 직접적으로 기반 또는 I 또는 II 등급 근거로부터 유추된 권고
D	IV 등급 근거에 직접적으로 기반 또는 I, II, 또는 III 등급 근거로부터 유추된 권고

외에 임상진료지침이 감안해야 하는 요소로 보유자원과 실천가능성을 들 수 있다. 검사나 치료의 비용이 타당한지에 대한 판단은 비용-효과의 계산 결과, 관점(예: 의사, 지불자, 사회 등), 보건의료 시스템의 자원상의 제약(예: 제한이 많은 공공시스템, 민간보험시스템) 등에 의해 좌우된다. 실천가능성의 문제는 의료진이 권고를 실천하기 위해 필요로 하는 시간, 기술, 인력, 장비 등의 고려와 환자와 의료시스템이 이를 감당할 수 있는지의 문제를 고려해야 한다. 지침에 담긴 권고는 표 4의 예와 같이 등급의 형태로 제시하는 것이 일반적이다(5). 이 같은 정보는 지침을 개발한 패널이 이 지침의 결과에 대해 어느 정도의 확신을 가지고 있는가를 독자에게 잘 전달해 준다. 이 같은 “권고의 강도” 분류법은 단순한 것부터 복잡한 것까지 다양하게 존재하나 반드시 그 가운데 특정한 것이 더 우월하지는 않다. 또한, 강력한 근거가 반드시 강력한 권고로 이어지는 것은 아니다. 등급은 패널 내에서 근거의 강도에 대해 충분한 논의를 거친 후 구성원의 투표에 의해 이루어지는 경우가 흔하다.

5. 임상진료지침의 검토와 갱신

임상진료지침의 신뢰성을 유지하기 위한 마지막 요소로 전문가, 전문학회, 정부기관, 소비자단체의 대표자, 지침의 잠재이용자 등을 망라한 다양한 외부 검토자에 의한 동료간검토를 들 수 있다. 지침의 잠재이용자에는 임상 의사 외에도 보험자단체의 대표자가 포함된다는 사실이 중요한데, 이는 이들이 임상진료지침의 실천에 핵심적인 구성요소이기 때문이다. 비록 다양한 범위의 검토자들로부터 의견을 받는 것이 바람직하고 중요하지만 다양한 검토자의 시각과 동기는 지침이 겨냥하고 있는 대상자의 시각과 동기와는 다를 수 있음을 잊지 말아야 한다(10). 예를 들어, 해당 분야의 학문적인 전문가들은 과학적인 주제를 알리려는데 중요성을 둬으로써 일차의료의사들과는 다른 우선순위를 가지고 있을 가능성이 크다. 임상진료지침은 내용타당도, 명확도, 적용가능도의 측면에 대해 외부의 평가를 거쳐야 한다. 외부평가자

는 다음의 세 분야, 즉 임상적인 내용에 대해 충분한 경험이 있고 지침의 문헌검색이 충분하였는지 임상적인 감각이 충족되었는지를 판단할 수 있는 능력이 있는 사람; 체계적인 종설이나 지침개발의 전문가로서 지침의 개발방법을 평가할 수 있는 사람; 지침의 잠재적인 이용자로서 이 지침의 유용성을 판단할 수 있는 사람으로 구성되어야 한다(5).

임상진료지침의 발간 이후 새로운 근거가 출판될 때마다 지침을 갱신할 수도 있으나, 체계적인 고찰을 갱신하는 시기를 미리 정하여 지침을 갱신하는 것이 더 바람직하다.

참고문헌

1. Field MJ, Lohr KN. Clinical practice guidelines: direction of a new agency. *Washington DC: Institute of Medicine*, 1990.
2. 트리샤 그린할프, 신승수 역. 의학논문 제대로 읽기. 서울: 몸과마음, 2001:192-205.
3. Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice. A systematic review of rigorous evaluations. *Lancet* 1993;342:1317-1322.
4. Hayward RS, Wilson MC, Tunis SR, Bass EB, Rubin HR, Haynes RB. More informative abstracts of articles describing clinical practice guidelines. *Ann Intern Med* 1993;118:731-737.
5. Shekelle PG, Woolf SH, Eccles M, Grimshaw J. Developing Guidelines. *BMJ* 1999;318:593-596.
6. Woolf SH. Practice guidelines, a new reality in medicine, II: methods for developing guidelines. *Arch Intern Med*. 1992;152:946-952.
7. Institute of Medicine, Council on Health Care Technology. Consensus Development at the NIH: Improving the Program. *Washington DC: National Academic Press*; 1990.

8. American Medical Association. DATTA Procedures. Chicago, Ill: *American Medical Association*; 1989.
9. Park RE, Fink A, Brook RH, Chassin MR, Kahn KL, Merrick NJ, et al. Physician Ratings of appropriate indications for six medical and surgical procedures. *Am J Public Health*, 1986;76:766-772.
10. Berg AO, Atkins D, Tierney W. Clinical Practice Guidelines in Practice and Education. *J Gen Intern Med* 1997;12:S25-S33.
11. Eddy DM. Practice policies: guidelines for methods. *JAMA*. 1990;263:1839-1841.
12. Eddy DM. A Manual for Assessing Health Practices and Designing Practice Policies: The Explicit Approach. *Philadelphia, Pa: American College of Physicians*; 1992.
13. Shekelle PG, Schriger DL. Evaluating the use of the appropriateness method in the agency for health care policy and research clinical practice guideline development process. *Health Serv Res* 1996;31:453-468.
14. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408-412.
15. Colditz GA, Miller JN, Mosteller F. How study design affects outcomes in comparisons of therapy. I: medical. *Stat Med* 1989;8:441-454.
16. Miller JN, Colditz GA, Mosteller F. How study design affects outcomes in comparisons of therapy. II: surgical. *Stat Med* 1989;8:455-466.