

근거 중심 치의학

2. 근거 수준과 비판적 평가

한림대학교 의과대학 강동성심병원 구강악안면외과학교실
 연구전임강사 홍 순 민

ABSTRACT

Evidence-Based Dentistry
 2. Evidence Level and Critical Appraisal

Department of Oral and Maxillofacial Surgery, KangDong Sacred Heart Hospital,
 Collage of Medicine, Hallym University
 Soon-Min Hong, D.D.S., M.S.D.

In this second part of series report on evidence-based dentistry, the level of evidence and the methods for critical appraisal are discussed. The epidemiologic studies, namely clinical studies have some bias per se, and the degree of bias is somewhat predetermined by the study design. The level of evidence can be defined as the reliability of the clinical study and it is dependent on the degree of bias. Thus, it is important to determine the type of the study and to understand its structure before critically appraising them. Systematic reviews of primary studies, randomized controlled trials, nonrandomized clinical trials, cohort studies, case-control studies, and case reports / series constitute the clinical studies and the level of evidence follows the order of the studies listed above. Critical appraisal is the most important procedure in evidence-based dentistry. It is done to determine the credibility of research papers and their usefulness in the clinician's own practice. Critical appraisal is consisted of multiple questions that are helpful for evaluating validity and usefulness of the studies concerning therapy, diagnosis, prognosis, and causation.

Key words : Dentistry, Evidence-Based Medicine, Epidemiologic Studies

서론

문헌 검색을 통해 찾아낸 논문은 비판적으로 평가해야 한다. 학계에서 인정 받는 공신력 있는 잡지에 실린 논문이라고 해서 반드시 모두 옳은 것은 아니다. 심지어는 의학 연구 논문의 99%에 오류가 개입되어 있다는 주장도 있었다.¹ 하지만 의학 논문의 99%에 오류가 있다고 하여 단지 남은 1%에서만 근거를 얻을 수는 없을 것이다. 오히려 비판적 평가는 연구 결과물들이 임상적 의사 결정 과정에서 이용하기에 적당한 정도로 오류가 적은가, 또는 확실하진 않지만 타당한가라는 질문에 대한 답을 얻는 과정이다. 우리가 환자를 진료하며 내리게 되는 결정 과정에는 많은 요소가 관여하며, 연구 증거물들은 이중 하나에 불과한 것이다.²

비판적 평가를 하기에 앞서 중요하게 생각해야 할 점은, 검색해낸 임상 연구의 종류를 먼저 파악해야 한다는 점이다. 이는 두 가지 이유 때문에 중요한데, 첫째는 임상 연구는 그 종류에 따라 각각 다른 근거 수준 (evidence level)을 갖게 되기 때문이다. 근거 수준이란 논문의 신뢰도를 의미하는 것으로, 연구자나 환자의 편향된 선입견에 따라 잘못된 결론을 유도할 가능성이 낮아질수록 근거 수준은 더 높아지게 된다.³ 둘째로는 비판적 평가의 방법은 개별 연구의 종류에 따라 다르게 적용하게 되기 때문이다. 임상적 문제는 치료, 진단, 예후, 원인 등 네 가지 종류로 나눌 수 있는데, 각 문제에 따라

적용 가능한 임상 연구의 종류가 정해지며 (표1) 이에 따라 비판적 평가를 각각 시행하게 된다.

따라서 제2부에서는 임상 연구의 종류와 특성에 따른 근거의 수준을 먼저 간략히 설명한 후에, 비판적 평가의 전반적인 과정과 각 연구 논문의 종류에 따른 개별적 평가 사항에 대해 설명할 것이다. 또한 임상 치과의사가 가장 많이 접할 치료 연구와 예후 연구에 대한 비판적 평가의 예를 제시하고자 한다. 비판적 평가에 대해 좀 더 자세한 정보를 원한다면 JAMA에 연재되었던 User's Guides to the Medical Literature 시리즈를 참고하는 것이 좋다.⁴ 또한 임상에서 적용하기 편한 쉬운 형태의 비판적 평가에 대한 자료를 원한다면 캐나다 Toronto 대학에서 운영하는 Centre for Evidence-Based Medicine 홈페이지⁵를 참고하는 것이 도움이 될 것이다.

임상 연구의 분류

임상 연구는 몇 가지 방법으로 구분이 가능한데, 크게 실험자가 대상인 환자에게 의도된 처치를 가했는가의 유무에 따라 관찰 연구 (observational study)와 실험 연구 (experimental study)로, 실험을 시작한 시점에 결과 측정에 사용할 사건이 발생하였는가의 유무에 따라 전향적 연구 (prospective study)와 후향적 연구 (retrospective study)로, 그리고 시간의 경과를 고려했는가에 따라 단면 연구

표 1. 연구의 목적과 이에 적합한 연구의 방법³

연구 주제	연구 목적	연구 방법
치료 (Therapy)	약물이나 수술의 효과를 검증함	가장 선호하는 방법은 무작위 대조 연구이다.
진단 (Prognosis)	새로운 진단 방법이 유효하고 신뢰성 있는지를 검증함	가장 선호하는 방법은 단면 연구이다.
예후 (Prognosis)	질병의 초기 단계에 있는 환자에서 향후 어떠한 일이 발생할 것인지를 밝힘	선호하는 방법은 코호트 연구나 환자-대조군 연구이다.
원인 (Causation)	질병의 원인이 될 것 같은 요소가 질병의 발생과 관계되는지 밝힘	선호하는 방법은 코호트 연구나 환자-대조군 연구이지만 발병률이 낮은 경우에는 증례 보고도 중요한 정보를 제공할 수 있다.

(cross-sectional study)와 추적 연구 (longitudinal study)로 분류할 수 있다⁶. 세부적으로는 무작위 대조 연구 (randomized Controlled Trial: RCT), 비무작위 대조연구 (nonrandomized controlled trial), 코호트 연구 (cohort study), 환자-대조군 연구 (case-control study), 단면 연구 (cross-sectional survey), 그리고 증례보고 및 증례나열 (case report and case series)로 구분된다^{3,7,8}.

과거에는 이들 임상 연구를 보고하는 방식이 표준화되지 않아서 여러 가지 혼란스러운 점이 있었으나, 이러한 혼란을 줄여주고 임상 연구의 질을 높이기 위한 목적으로 점차 각 임상 연구를 보고하는 방식들이 전문가들에 의해 표준화되고 있다. 현재 무작위 대조 연구의 보고 형식 (CONSORT)⁹, 비무작위 대조 연구의 보고 형식 (TREND)¹⁰, 진단 연구의 보고 형식 (STARD)¹¹, 무작위 대조 연구에 대한 메타 분석의 보고 형식 (QUOROM)¹², 그리고 관찰연구에 대한 메타 분석의 보고 형식 (MOOSE)¹³이 제시된 바 있다.

근거 수준

근거 수준 (evidence level)은 연구 논문이 얼마나 신뢰할 수 있는가를 나타내는 지표이다¹³. 연구가 목적인 바를 옳게 얻는데 지장을 주는 것을 연구 오차라고 하며, 근거 수준이 높은 연구는 이러한 오차가 적어서 진리에 더 가까운 결론을 얻을 수 있는 연구이다. 근거 수준을 나타내기 위해 치우침 (bias)이란 개념을 사용하기도 하는데, 치우침이란 연구의 결과나 결론을 진실에서 편향시킴으로써 처치의 효과를 과장되게 부풀리거나 줄일 수 있는 모든 요소와 과정을 의미한다. 따라서 근거 수준이 높은 연구일수록 치우침의 정도는 낮아진다.

근거 수준은 무작위 대조 연구에 대한 체계적 고찰과 메타 분석이 가장 높으며, 그 아래로 무작위 대조 연구, 비무작위 대조 연구, 코호트 연구, 환자-대조군 연구, 증례 보고, 권위자의 개인적 의견의

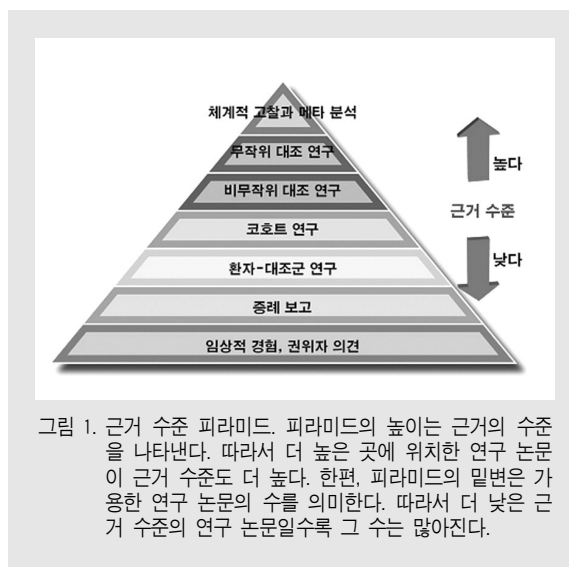


그림 1. 근거 수준 피라미드. 피라미드의 높이는 근거의 수준을 나타낸다. 따라서 더 높은 곳에 위치한 연구 논문의 근거 수준도 더 높다. 한편, 피라미드의 밑변은 가용한 연구 논문의 수를 의미한다. 따라서 더 낮은 근거 수준의 연구 논문일수록 그 수는 많아진다.

순서를 따른다 (그림 1)¹⁴⁻¹⁶. 이때 한가지 주의할 점은, 근거 수준이 더 높은 연구가 근거 수준이 낮은 연구보다 항상 우월하지는 않다는 점이다. 좋은 연구 방법을 적용하였더라도 실제 연구 과정에서 개입되는 오차를 잘 제어해 주지 않는다면 근거의 질은 떨어질 수밖에 없는 것이다.

비판적 평가의 전반적인 과정

개별 문헌에 대해 비판적 평가를 가할 때에는 근거 수준이 높은 문헌부터 시작하는 것이 좋다. 높은 근거 수준의 논문이 비판적 평가 결과 적절한 방법으로 합당한 결론을 내렸으며, 이것이 임상치과의사 본인의 개별적 임상 상황에 부합한다고 생각된다면, 낮은 근거 수준의 문헌을 읽어야만 하는 수고를 덜 수 있는 것이다. 논문 전문에 대한 평가는 대략 세 가지로 나뉘어진다^{17,18}.

- 연구의 결과는 유효한가? (Are the results of the study valid?)

이는 연구의 내부 유효성 (internal validity)을 평가하는 것으로, 연구의 목적 및 종류에 적절한 엄정한 실험 방법을 적용함으로써 치우침의 발생 가능성을 최소화시켰는지를 평가하는 것이다.

• 결과는 어떠하였는가? (What are the results?)

유효한 방법을 통해 얻어진 결과를 분석하는 것이다. 결과 분석을 위해 기존 논문에서는 주로 p값 등의 정성적인 (qualitative) 통계적 수치를 중시하였지만, 현재는 신뢰 구간 (confidence interval) 등 정량적인 (quantitative) 분석을 중요시한다.

• 이 결과는 내가 환자를 진료하는데 도움이 되는가? (Will the results help locally?)

이는 연구의 외부 유효성 (external validity)과 결과의 가치를 평가하는 것이다. 임상 연구는 제한된 조건의 참여자를 대상으로 제한된 조건 하에서 시행하는 경우가 많다. 이러한 요소들이 개별 치과 의사의 환자 및 진료 조건에 부합하는가를 평가하는 것이 외부 유효성 평가이다.

체계적 문헌 고찰에 대한 비판적 평가

체계적 문헌 고찰 (systematic review)이란 체계적이고, 투명하며, 재현 가능한 방법으로 특정 주제의 일차 연구를 검색, 선택, 평가, 그리고 결합한 논문을 의미한다. 일차 연구의 결과들을 수리적이고 통계적으로 결합한 체계적 문헌 고찰의 한 가지 종류를 메타 분석 (meta-analysis)이라고 한다¹⁹. 체계적 문헌 고찰은 기존의 권위자에 의한 문헌 고찰과는 다른 개념으로, 일차 문헌에 대해 객관적이고 표준화된 평가를 내리는 것이다. 체계적 문헌 고찰에 관한 비판적 평가는 간략히 다음 과정을 따른다²⁰.

(1) 도입부에서 적절하며 명확한 질문을 제기하였는가?

명확하게 잘 정의된 질문이 없다면 문헌 고찰이 그 목적에 얼마나 일치하는지, 그리고 독자가 답하고자 하는 질문에 얼마나 부합되는지 평가하기가 어렵다. 질문은 연구한 인구 집단, 시행한 처치, 고려한 측정 결과의 관점에 기초하여 제기해야 한다.

(2) 일차 연구의 포함 기준 및 제외 기준은 적절하며 명확히 기술되어 있는가?

고찰에 사용할 일차 연구의 인구 집단, 처치, 결과 측정 방법, 그리고 연구 종류에 기초하여 그 포함 및 제외 기준을 잘 정의하고 명확히 진술한다. 이는 실제로 일차 연구를 검색하기 이전에 정해져야 하는데, 왜냐하면 특정 의견의 논문만 선택하는 선택 치우침 (selection bias)을 피할 수 있고 연구의 재현 가능성을 높여주며 적당한 연구가 포함되었는지 여부를 독자가 판단할 수 있도록 해주기 때문이다²¹.

(3) 중요한 데이터베이스를 모두 포괄하여 철저한 문헌 검색이 이루어졌는가?

이는 선택 치우침과 출판 치우침 (publication bias)을 최소화해 주기 위한 전략이다. 일반적으로 중요한 치의학 잡지에서는 그 연구 결과가 명확한 논문을 많이 출판하고 결과가 명확하지 않은 것들 (즉, 통계학적 차이가 나지 않는 결과를 보이는 논문)은 잘 게재하지 않는 경향이 있는데, 이를 출판 치우침이라고 한다. 따라서 MEDLINE뿐만 아니라 적어도 EMBASE나 Cochrane library 등의 광범위한 데이터베이스를 참고해야만 한다.²¹ 또한 부가적으로 주요 잡지에서 검색 조건에 부합하는 논문을 일일이 찾아보고 (hand searching) 일차적으로 검색된 논문의 참고 문헌까지도 검색해야 한다²⁰. 더불어, 데이터베이스와 키워드를 포함한 검색 전략을 명확히 기술해야 한다²².

(4) 검색한 일차 연구의 신뢰성을 객관적인 방법으로 평가하였는가?

검색해낸 일차 문헌들은 모두 동일한 근거 수준을 가지고 있지는 않을 것이며 실험 방법상의 문제로 오류를 일으켰을 수도 있다. 따라서 적절한 평가 과정을 거치지 않은 채 개별 연구들을 단순히 결합한다면 그 결과는 신뢰성이 없을 것이다. 이를 위해 포함된 각 연구들에 대한 비평을 시행하고 이를 표로 순서대로 나열하는 것이 좋다. 방법은 재현 가능하며 치우침이 없어야 한다.

(5) 개별 일차 연구의 환자 선정, 실험 방법, 결과 측정의 방법이 유사하며 서로 비슷한 결

론을 내렸는가?

체계적 고찰의 궁극적인 목적은 동일한 주제의 연구 논문 결과를 산술적으로, 그리고 통계학적으로 통합하여 개별 연구에 비해 더 신뢰성 있는 결과를 제시하기 위함이다. 따라서 통합해야 할 일차 연구들은 서로 비슷한 환자에 대해 비슷한 방법을 적용하고 비슷한 결과를 측정했는 것들이 좋다. 이러한 항목 중 어느 하나라도 다른 연구들과 현저한 차이를 보이는 연구는 제외하거나 결과를 통합하는 과정에서 이에 대해 고려해 주어야 한다.

또한 비슷한 환자 집단에 비슷한 처치를 가하고 비슷한 항목에 대한 변화를 측정하였다면 그 결과도 비슷해야 하는 것은 당연하다고 할 것이다. 이를 위해 개별 연구의 결과가 비슷한지 평가해야 한다.²³ 결과에 동질성이 없다면 그 이유를 밝혀야 하며²⁴ 특별히 유별나게 큰 차이를 보이는 연구를 최종적인 분석에 포함시킬지 여부도 결정해야 한다.

(6) 연구의 관찰 결과들은 적절히 결합하였는가?

체계적 문헌 고찰에서는 개별 연구의 자료 하나 하나를 산술적, 통계적으로 결합하려는 노력을 한다. 이 방법의 장점은 통계적 분석의 대상 수를 늘려줌으로써 통계학적 유의성을 발견하기 쉽도록 해준다는 것이며²⁵ 이것이 체계적 문헌 고찰이 기존의 문헌 고찰에 비해 우월한 점이라고 할 수 있다.

(7) 이 연구의 결과는 내 환자를 진료하는데 있어 도움이 되는가?

이는 연구의 외부 유효성을 평가하고 연구 결과의 중요도를 파악하는 과정이다. 개별 연구에서 대상으로 한 환자의 특성과 진료 조건이 임상 치과의사 본인의 환자 및 진료 여건에 부합하는지를 확인한다. 이후 체계적 문헌 고찰에서 내린 결론이 환자에게 현저한 도움을 줄 수 있다면 이를 임상에 적용한다.

치료 연구에 대한 비판적 평가

치료가 주제인 연구 중 가장 근거 수준이 높은

것은 무작위 대조 연구이다. 비무작위 대조 연구를 평가할 때에는, 환자를 배정하는 방법에 의해 치우침이 개재될 가능성은 없었는지 평가하는 것이 가장 중요하다²⁶. 무작위 대조 연구의 비판적 평가는 다음 순서를 따른다²⁷.

(1) 도입부에서 적절하며 명확한 질문을 제기하였는가?

질문은 연구한 인구 집단, 시행한 처치, 고려한 측정 결과의 관점에 기초하여 구체적이고 명확하게 제기한다.

(2) 환자의 군은 무작위로 배정되었으며 논문에 그 구체적 방법을 기술하였는가?

무작위 배정은 참가자의 무작위적 분배와 배정은폐 (allocation concealment)의 두 단계로 이루어진다. 이는 알려지지 않은 환자의 잠재적 기저특성이 각 군에 비슷하게 배분되도록 해주기 때문에 최종 결과가 가해준 처치에 의해서만 발생되었다는 확신을 가질 수 있도록 해준다²⁸. 실제로 그릇된 방법으로 무작위 배정을 하면 치료 효과가 41% 과장되게 나타날 수 있으며, 논문에서 무작위 배정을 하였다고 진술하였지만 그 구체적인 방법에 대해 기술하지 않았을 때에는 30% 과장되게 나타났다는 보고가 있었다²⁸.

(3) 연구의 시작 시점에서 각 군에 분배된 환자들의 기저특성은 비슷하였는가?

무작위 배정이 완벽하게 이루어졌더라도 특히 참여한 환자의 수가 적을 때에는 각 군에 배정된 환자의 기저특성 사이에 차이가 있을 수 있다. 논문에서는 표를 통해 각 군에 배정된 환자의 연령, 인종, 성별, 병력 등의 특성에 차이가 없음을 보여야 하며, 차이가 클 때에는 결과 분석 시에 이를 고려해야만 한다.

(4) 환자, 처치 제공자 (의사), 그리고 결과 분석자에게 환자가 배당된 소속군을 비밀로 유지하였는가? (맹검이 성공적이었는가?)

맹검 (blindness)이란, 참가자, 처치 제공자, 그리고 때로는 결과 분석자 등 연구에 관련된 이들이

개별 환자에게 배당된 치료를 모르게 하는 것이다. 일반적으로 환자는 새로운 치료를 받거나 위약(placebo)이 아닌 치료를 받으면 치료 효과를 과장되게 보고하는 경향이 있으며 결과를 측정하는 연구가나 결과 분석자도 동일한 경향을 보인다. 하지만, 배당 은폐와는 다르게, 맹검은 항상 시행 가능하지는 않으며 특히 치과 처치를 가할 때에는 환자나 처치 제공자에게 맹검 상태를 유지하는 것은 불가능하거나 어렵다. 이러한 경우에도 결과 측정만은 맹검 하에서 시행할 수 있다²⁹.

(5) 처음 연구에 참여했던 모든 환자에 대해 결과를 측정하였는가? 처치 의도 분석을 시행하였는가? 충분히 오랜 기간 동안 추적 관찰하였는가?

추적 관찰 중 소실된 환자를 보고하고 그 이유를 설명하는 것은 중요하다. 사실은 추적 관찰 중 소실된 환자들이 더 유용한 정보를 제공하는 경우도 많다. (합병증 발생이나 처치에 의해 질환이 소실되어 소실되는 경우가 많다.) 이러한 정보를 모두 기술하였다고 하더라도, 추적 관찰된 참여자의 비율이 80% 미만이라면 그 연구는 받아들일만하지 못하다⁷. 또한 모든 환자는 처음 배정된 군에 포함시켜 분석해야 한다 (이를 처치 의도 분석: intention to treat analysis이라고 한다). 예컨대 연구 시작 시점에서 A군에 배정 받았던 환자가 중간에 B군에 할당된 치료를 받았다면 이 환자는 A군에 포함시켜 분석해야 하는 것이다. 이는 무작위 배정의 원칙을 지키기 위한 것으로, 치료의 차이에 의한 결과의 차이보다는 무작위 배정에 의해 균등하게 분배된 환자의 기저특성이 더 중요하기 때문이다. 마지막으로, 관심 결과가 발생하기에 충분한 기간 동안 추적 관찰을 시행해야 처치의 효과를 제대로 파악할 수 있다.

(6) 임상적으로 중요한 결과를 측정하였는가?

실험 논문에서 측정하는 결과에는 대리 결과(surrogate end point)와 최종 결과(true end point)가 있다. 대리 결과는 생리적, 해부학적 지표

로, 생체 내의 생물학적 활성을 측정한 결과를 의미하며, 최종 결과는 환자에게 실제적인 것들이며 환자가 어떻게 느끼는가, 기능하는가, 또는 생존하는가를 직접 결정해줄 수 있는 측정치이다. 예컨대 치아 우식증과 관련된 연구에서 치면 미생물 밀도를 측정한다면 이는 대리 결과를 측정한 것이며, 치료 받아야 할 우식증의 개수나 범위를 측정하였다면 이는 최종 결과를 측정한 것이다³⁰. 대리 결과를 측정하면 빠른 시간에 손쉽게 연구를 시행할 수 있다는 장점이 있지만, 대리 결과가 반드시 최종 결과를 예견할 수 있는 것은 아니기 때문에 치료 논문에서는 가급적 최종 결과를 측정하는 것이 좋다.

(7) 이 연구의 결과는 내 환자를 진료하는데 있어 도움이 되는가?

연구의 외부 유효성을 평가하고 연구 결과가 통계학적으로뿐만 아니라 임상적으로도 중요한 결론을 도출하게 되었는가 평가한다.

(8) 치료 연구에 대한 비판적 평가의 예

본 문헌고찰의 1편에서 문헌 검색의 예를 제시한 바 있다. 이를 통해 1998년 Harradine 등이 제3대구치의 발치 여부가 하악 전치부 충생의 발생에 미치는 영향을 평가한 무작위 대조 연구³¹를 검색해 낼 수 있었다. (이 연구는 PubMed에서 전문을 무료로 다운로드 받을 수 있다.) 이를 비판적으로 평가하고 그 예를 표 2에 제시한다.

진단 연구에 대한 비판적 평가

진단 연구에서는 새로이 시도하는 진단법을 기존의 표준 진단법에 비교하는 것이 기본이다. 진단 연구에 대한 비판적 평가의 방법은 다음과 같다^{32,33}.

(1) 연구할 진단 방법의 특성을 명확히 기술하였는가? 실제 진료에서 재현할 수 있는가?

측정법의 특성을 기술한다. 즉, 새로운 진단 방법을 어떻게 실행하는지, 그리고 그 결과는 어떻게 해석해야 하는지 명확하게 기술해야만 한다. 뿐만

표 2. 치료 연구에 대한 비판적 평가의 예

<p>A. 연구의 결과는 유효한가? (Are the results of the study valid?)</p>
<p>1. 도입부에서 적절하며 명확한 질문을 제기하였는가? (그렇다.) Introduction의 말미와 Subjects and Methods의 서두에 명확히 기술되어 있다. 이 논문의 목적은 “제3대구치의 조기 발치 (early extraction)가 하악 절치의 만기 총생 (late lower incisor crowding)에 어떠한 영향을 끼치는가”를 전향적으로 평가하기 위함이었다.</p>
<p>2. 환자의 군은 무작위로 배정되었으며 논문에 그 구체적 방법을 기술하였는가? (그렇다.) 교정을 완료한 청소년기의 환자를 대상으로 하였으며, 이들의 발치 여부는 무작위적으로 생성한 숫자로 결정하였다고 기술되어 있다.</p>
<p>3. 연구의 시작 시점에서 각 군에 분배된 환자들의 기저특성은 비슷하였는가? (아마도 그럴 것이다.) 발치를 시행한 군과 시행하지 않은 군 사이의 기저특성은 비교하지 않았다. 연구 시작 시점에서 연령이나 성별 등의 기저특성이 각 군마다 어떻게 분배되었는지에 대한 언급이 없다. 그러나 엄격한 포함 기준 및 제외 기준을 만족시키는 환자만을 연구의 대상으로 하였기 때문에 기저특성에 차이가 있을 가능성은 적다.</p>
<p>4. 환자, 처치 제공자 (의사), 그리고 결과 분석자에게 환자가 배당된 소속군을 비밀로 유지하였는가? (그렇다.) 발치는 외과적 처치이기 때문에 처치 제공자 및 환자에게 맹검을 유지하는 것은 불가능했다. 하지만 결과 분석자에게 맹검이 유지되었다는 사실이 기술되어 있다.</p>
<p>5. 처음 연구에 참여했던 모든 환자에 대해 결과를 측정하였는가? 충분히 오랜 기간 동안 추적 관찰하였는가? (추적 관찰 기간은 적절하였지만 추적 관찰 소실율은 높았다. 그러나 높은 소실에 대해 어느 정도는 정당화가 가능하다.)</p> <ul style="list-style-type: none"> • 처음 연구에 참여하였던 총 164명의 환자 중 77명 만이 최종 결과 측정이 가능하였다 (47%). 이는 매우 낮은 추적 관찰율이지만, 추적 관찰 기간이 길었고 (평균 66개월), 추적 관찰이 가능했던 환자와 그렇지 못한 환자에서 연구 시작 시점에 총생의 정도가 차이가 없었으며, 이들 사이에 거주지나 치료에 대한 태도의 차이가 없었다는 설명이 있었으므로 환자 소실이 많았음에도 불구하고 신뢰성이 있는 것으로 판단된다. • 연구 시작 시점 (교정 치료 완료 시점)에서 환자의 평균 연령은 14세 10개월이었으며 추적 관찰 기간은 평균 66개월이었다. 이는 10대 중반에서 10대 후반, 또는 20대 초반까지의 기간을 관통하며, 따라서 제3대구치가 맹출하기 전후의 기간을 포함하기 때문에 추적 관찰 기간으로 적절하다고 판단된다.
<p>6. 임상적으로 중요한 결과를 측정하였는가? (그렇다.) 이 연구에서는 각 환자의 연구 시작 시점 및 연구 종료 시점에서 인상 채득한 모델 상에서 Little’s Irregularity Index (LII), Intercanine width (ICW), 그리고 Arch length (AL)를 측정하였고, 각 측정치에 대한 시작 시점과 종료 시점의 차이를 결과로 이용하였다. 이들 측정치는 환자의 전치부 총생의 정도를 직접적으로 측정할 수 있는 것으로 여겨 진다.</p>
<p>B. 결과는 어떠하였는가? (What are the results?)</p>
<p>치료의 효과는 어느 정도였는가? 연구 시작 시점과 연구 종료 시점에서 측정한 LII 및 ICW의 차이값은 95% 신뢰 구간에서 두 군간에 유의한 차이가 없었으며 그 수치도 1.02mm 및 0.18mm만 차이가 났기 때문에 임상적인 중요성은 없었다고 할 수 있다. 그러나 AL의 변화량은 두 군간에 통계학적으로 유의한 차이가 있었다. 하지만 그 수치는 0.77mm 정도밖에 되지 않았기 때문에 임상적으로 중요하다고 볼 수 없었다. 즉, 세 가지 결과 측정치 모두가 두 군간에 임상적으로 중요한 차이는 나타내지 않았다고 할 수 있다.</p>
<p>C. 이 결과는 내가 환자를 진료하는데 도움이 되는가? (Will the results help locally?)</p>
<p>이 연구의 가장 큰 결점은, 연구 시작 시기에서 두 군의 기저특성 간에 차이가 있었는지 평가하지 않았다는 점과 추적 관찰 소실율이 높았다는 점이다. 그러나 이 연구는 무작위 배정의 원칙을 잘 지켰으며 엄격한 포함 기준 및 제외 기준을 적용하여 참여자를 결정했기 때문에 두 군의 기저특성 간에 차이가 없었거나 있었다라도 크지 않았을 것으로 판단된다. 또한 높은 추적 관찰 소실에 대해서도 어느 정도 정당화가 가능하다. 이 연구는 하악에 능동적 교정 치료를 가하지 않은 환자만을 대상으로 하였기 때문에, 교정 치료를 마친 하악 치아의 회귀 현상이 이 연구의 결과값에 어떠한 영향을 끼쳤으리라고 생각하기는 힘들다. 게다가 연구에 포함된 환자들은 모두 제3대구치의 치축이 제2대구치쪽을 향하면서 맹출 공간이 부족하였기 때문에, 맹출하는 제3대구치가 전방 치열쪽에 압력을 가하여 전치부 총생을 유발하는가를 평가하기에 매우 좋았다. 또한 결과 측정 대상 환자들 모두가, 청소년기 초에서 시작하여 성년 초, 또는 청소년기 말까지 장기간 추적 관찰이 가능했었다. 따라서 교정치료를 끝마쳤으며 청소년기에 있는 내 환자에게 적용할 수 있다.</p>
<p>D. 임상 진료에의 적용</p>
<p>환자에게 매복되어 있으며 증상이 없는 제3대구치의 발치를 권유하지 않는다. 교정 치료 후 하악 전치부 총생의 재발 현상은, 매복된 제3대구치의 맹출과는 큰 관계가 없는 것으로 보인다. 물론 제3대구치를 발치하지 않는 경우, 하악 전치부 총생이 발생한다면 그 양이 약간 더 클 가능성이 있지만 임상적으로는 그다지 중요해 보이지 않는다. 또한 제3대구치 발치로 인해 초래될 수 있는 다양한 합병증, 추가적 치료 비용, 그리고 고통을 생각한다면 발치를 권유하는 것은 적절해 보이지 않는다.</p>

아니라 환자 준비, 주의 사항, 그리고 합병증 등에 대해서도 자세히 기술해야 한다.

(2) 임상에서 마주칠 수 있는 광범위한 특성의 환자를 포괄하여 측정하였는가?

진단 연구는 측정할 특성이 광범위하게 분포된 환자군이나 무작위적으로 선택된 환자들만을 대상으로 하는 것이 좋다. 저자들은 환자 선택과 연구 조건을 명시해야 한다. 이는 새로운 진단 방법이 다양한 특성에 대해 동일한 정도의 진단 능력을 지녔는지를 평가하기 위함이다.

(3) 실험에 참여한 모든 환자는 새로운 진단 방법과 참고 기준이 될 수 있는 표준 진단법을 모두 시행 받았는가?

연구 대상 측정법은 적절한 참고 기준이 될 수 있는 표준 진단법과 비교해야만 한다. 이를 통해 대상 환자군에서 양성 결과와 음성 결과를 알아내는데 있어 새로운 진단법이 얼마나 효율적인가를 알 수 있다. 참고 기준이란, 그 정확성에 대한 한계가 명확히 정의된 기존의 진단법이다. 참고 기준이 없을 때에는 민감성과 특이성이 알려진 기존의 다른 진단법을 사용해야 한다²⁹.

(4) 새로운 진단 방법과 표준 진단법은 맹검 하에서, 그리고 독립적으로 시행되었는가?

연구 진단과 표준 진단은 맹검 하에서 실시해야 한다. 서로 다른 측정자가 연구 진단과 표준 진단을 시행해야 하며, 치우침을 배제하기 위해 한가지 진단을 하는 연구자는 다른 진단 결과를 몰라야 한다. 더불어, 연구 진단 및 표준 진단을 행할 때에는 가급적 시간차가 없는 것이 좋다. 시간이 길어지면 환자의 상태에 변화가 생길 수 있기 때문이다.

(5) 이 연구의 결과는 내 환자를 진료하는데 있어 도움이 되는가?

연구의 외부 유효성을 평가한다. 또한 논문에 나와있는 특이도 (specificity), 민감도 (sensitivity), 양성 예측치 (positive predictive value), 음성 예측치 (negative predictive value)를 참고하여 새로운 진단 방법이 임상에 적용할만한 가치가 있는지 결

정한다.

예후 연구에 대한 비판적 평가

예후 연구에서는 특정 질환이나 상태로 인해 야기될 수 있는 결과와 이 결과의 빈도를 관찰한다³³. 예후 연구의 비판적 평가는 다음과 같다³⁴.

(1) 환자들은 잘 설명되었고, 대표적이며, 연구를 시작할 당시 질환의 진행 정도가 비슷하였는가?

관심 상태 (질병)에 대해 자세히 설명해야 하며, 환자가 질환을 가졌는지 여부를 명확히 기술해야 한다. 중증 질환을 가진 환자만을 대상으로 한다면 우리가 임상에서 마주칠 수 있는 환자들을 대표하는 것이 아니며, 초기 질환에서 중증 질환으로 이환 되지 않는 환자들을 고려할 수 없기 때문에 가급적 초기 질환 상태의 환자들만을 대상으로 하는 것이 좋다.

(2) 추적 관찰은 충분히 오래, 그리고 완전히 하였는가?

질환이나 상태에 노출된 후 관심 결과가 나타나기까지는 오랜 시간이 걸리는 경우가 많기 때문에 연구 대상자가 중간에 탈락하는 경우가 많다. 하지만 탈락자가 너무 많으면 연구 결과가 편향될 수 있기 때문에 적어도 시작 시점에서의 대상자 중 80% 이상이 특정한 종결점 (end point)이나 결과 (outcome)에 다다를 때까지 추적 관찰 되어야 연구 결과가 유효한 것으로 판단한다²⁹.

(3) 결과의 기준은 명확하였고 평가자는 이를 맹검 하에서 결정하였는가?

결과로 간주하는 상태의 기준에 대해서는 연구 내에서 명확히 기술하고 그 이론적 근거를 제시하는 것이 좋다. 만약 결과 측정이 기술되어 있지 않다면, 또는 연구가 대리 결과에 기초한 결론을 얻었다면 이 연구 결과는 받아들이기 힘들다. 또한 결과 평가자가 참여자의 병력이나 관심 질환 및 상태에 대한 노출 여부를 모른다면 결과를 측정하는

표 3. 예후 연구에 대한 비판적 평가의 예

A. 연구의 결과는 유효한가? (Are the results of the study valid?)

1. 환자들은 잘 설명되었고, 대표적이며, 연구를 시작할 당시 질환의 진행 정도가 비슷하였는가? (어느 정도는 그렇다.)

환자는 흡연 중인 환자군과 흡연의 경험이 전혀 없는 사람 / 흡연을 중단한 사람 등이 포함된 비흡연자군으로 분류하였다. 흡연 기간, 양, 그리고 흡연을 중단하였다면 중단 기간 및 중단 시까지의 흡연 기간을 고려하는 것이 좋으나 이에 대한 구분은 하지 않았다. 그러나 이러한 요소들을 모두 고려한다면 통계적 과정이나 결과 해석이 매우 어려워질 수 있으므로 어느 정도는 적절한 환자군 분류였다고 할 수 있다.

2. 추적 관찰은 충분히 오래, 그리고 완전히 하였는가? (추적 관찰 소실은 적었으나 기간은 약간 짧았다.)

일반적으로 implant의 장기간의 생존율을 파악하기 위해서는 5년 이상의 생존율을 보는 것이 좋으나 이 연구는 3년까지의 생존율을 산출할 수 있었다. 그러나 일반적으로 implant의 실패가 식립 후 1년 이내에 집중되기 때문에 장기적 생존에 대해 어느 정도의 판단은 가능할 것으로 생각된다.

이 연구는 후향적 코호트 연구였기 때문에 추적 관찰 소실율은 산출할 수 없다. 그러나 다른 연구의 일환으로 식립하여 그 자료를 기록하고 있었던 대략 총 2900개 이상의 implant 중 2887개의 implant에 대해 식립 후 3년까지의 완전한 결과를 얻어 이를 연구에 포함시킬 수 있었기 때문에 (90%) 선택 치우침 (selection bias)이 개입되었을 가능성은 낮았을 것으로 판단된다.

3. 결과의 기준은 명확하였고 평가자는 이를 맹검 하에서 결정하였는가? (결과 기준은 명확했으나 맹검 여부는 모른다.)

이 연구는 변연골 소실이나 치주 지수 등의 대리 결과가 아닌, implant의 생존 자체, 즉 최종 결과를 평가의 기준으로 삼았기 때문에 결과 기준은 적절했다고 할 수 있다. Implant의 실패에 대한 기준은, 그 원인에 상관없이 implant를 제거해야만 하는 경우로 정의하였다.

결과 평가자가 환자의 흡연 여부에 대해 맹검 상태를 유지하였는가에 대한 언급은 없었다. 그러나 implant의 성공/실패는 비교적 명확하게 측정 가능한 것이기 때문에 맹검 여부가 아주 중요하다고 보기 힘들 수도 있다.

4. 특별히 다른 예후를 보이는 환자들의 소그룹 (subgroup)이 있다면, 이에 영향을 미친 외부 인자를 찾아내어 결과 분석에 이를 고려하였는가? (외부 인자에 대해 매우 심도 깊게 고려하였다.)

이 연구는 implant 성공/실패에 영향을 끼칠 수 있는 환자의 기저특성들인 전신 건강 상태 (ASA physical status), 연령, 성별, 그리고 인종이 실제로 implant 실패에 영향을 끼쳤는지를 통계학적 방법을 동원하여 체계적으로 검색하였다. 또한 식립된 implant의 수산화인회석 coating여부와 술전 항생제 투여 여부가 implant의 실패에 영향을 끼쳤는지도 통계학적 방법으로 평가하였다. 따라서 외부 인자에 대한 고려는 매우 적절했다고 할 수 있다.

B. 결과는 어떠하였는가? (What are the results?)

3년까지의 실패율은 비흡연자군에 식립한 implant에서 6%, 흡연자군에 식립한 implant에서 8.9%였다. 즉, 흡연자에 implant를 식립하면 비흡연자에 비해 실패율이 1.5배 정도 상승하였다. 이는 교란 요소를 고려하더라도 통계학적으로 유의한 차이였다. 한편, 이 연구에서는 implant 식립을 2단계 술식으로 진행하였는데, 실패는 주로 2차 수술로 fixture를 노출시켜 치유 지대주를 연결 시킨 후 최종 보철물이 연결되기 전까지의 시기에 집중되었다. 전신 건강 상태, 연령, 성별, 또는 인종 등의 기저특성들은 흡연이 implant 실패를 야기하는데 별다른 영향을 미치지 못했다. 그러나 수산화인회석이 coating된 implant를 흡연자에게 식립하면 실패율이 11.2% 감소 (16%에서 4.8%로)하였으며 술전에 흡연자에게 항생제를 투여하면 implant의 실패율이 10.2% 감소 (14.9%에서 4.7%로)하였다. 이는 모두 통계학적으로나 임상적으로 유의한 차이였다. 신뢰 구간에 대한 언급은 없었다.

C. 이 결과는 내가 환자를 진료하는데 도움이 되는가? (Will the results help locally?)

연구된 환자는 내 환자와 비슷한가? 그 결과는 치료를 가하거나 피하는 데에, 또는 환자에 대한 권고에 도움이 되는가?

이 연구는 임상적으로 가정할 수 있는 여러 가지 요소를 모두 고려하였기 때문에 외부 유효성이 높다고 할 수 있다. 즉, 다양한 연령, 인종, 성별, 그리고 전신 건강 상태를 모두 포괄한 환자들을 포함하였으며 결과 분석 시 이를 고려하였기 때문에 이 결과를 내 환자에게도 충분히 적용 가능하다고 할 수 있다. 또한 흡연 여부는 implant 실패에 주요한 영향을 끼친다고 결론 내릴 수 있기 때문에 환자에 대한 권고에 충분한 도움이 된다. 다만 추적 관찰 기간이 3년으로 짧았기 때문에 좀 더 장기간의 연구에 대한 추가적인 고려가 필요한 것으로 판단된다.

D. 임상 진료에의 적용

흡연은, implant의 실패율을 식립 후 3년까지 약 1.5배 정도 높일 수 있기 때문에 금연할 것을 환자에게 권유한다. 또한 흡연자에게 implant를 식립할 경우에는 술전, 술후에 항생제를 주의 깊게 투여하고 수산화인회석이 코팅되었다거나, 또는 표면 처리된 implant를 식립할 것을 고려한다. 2단계 술식으로 implant를 식립할 경우 2차 수술 후에 실패하는 경우가 많음을 반드시 명심하고 있는 것이 좋다. 또한 1단계 술식을 시행하면 implant 구조물이 구강 내로 계속 노출되어 실패율이 높아질 수 있기 때문에, 흡연자에게는 1단계 술식을 가급적 시행하지 않도록 한다.

데 있어서 치우침의 정도는 줄어들 것이다.

(4) 특별히 다른 예후를 보이는 환자들의 소그룹 (subgroup)이 있다면, 이에 영향을 미친 외부 인자를 찾아내어 결과 분석에 이를 고려하였는가?

참여자의 연령이나 성별, 또는 기타 인자가 예후에 영향을 미칠 수도 있다. 노출과 결과에 모두 관련되어 있는 제3의 변수로 인해 노출과 결과의 관계에 발생하는 왜곡을 교란 (confounding)이라고 하는데, 관찰 연구에서는 무작위 배정을 할 수 없기 때문에 교란이 발생할 가능성이 높아서 근거의 수준이 떨어진다³⁵. 따라서 가급적 연구 시작 전에, 또는 연구 종료 후에라도 가능한 교란 인자를 찾아내어 나열하고 결과 분석 시에 이를 고려하는 것이 좋다.

(5) 연구된 환자는 내 환자와 비슷한가? 그 결과는 치료를 가하거나 피하는 데에, 또는 환자에 대한 권고에 도움이 되는가?

연구의 외부 유효성을 평가한다. 연구 결과로 추론할 수 있는 예후의 정도를 파악하여 치료의 예후와 비용 간의 관계, 그리고 환자의 선호도를 파악하여 진료에 적용한다.

(6) 예후 연구에 대한 비판적 평가의 예

앞선 논문에서 예후 연구에 대한 질문의 예로서, “흡연 여부가 implant의 장기 성공에 어떠한 영향을 미치는가?”를 들었다. 이에 대해 PubMed 검색을 시행하여 몇 개의 문헌을 찾아낼 수 있었다. 이 중 2000년 Lambert 등이 발표하였던 논문³⁶을 비판적으로 평가하였으며 그 예를 표3에 정리하였다. 이 연구는 다른 목적의 무작위 대조 연구를 위해 축적한 자료를 이용하여 흡연이 implant의 성공에 미치는 영향을 평가하였는데, 그 연구 형태는 후향적 코호트 연구로 분류할 수 있었다.

원인 연구에 대한 비판적 평가

원인 연구는 특정 처치나 요소에 대한 노출이 질

환이나 상태를 유발하는가를 평가하는 연구이다. 원인에 대한 연구도 예후에 대한 연구와 비슷하게 윤리적 문제로 인해 관찰 연구만이 시행 가능하다. 원인 연구에 대한 비판적 평가는 다음 사항을 고려한다³⁷.

(1) 실험군과 대조군으로 분류된 참여자들의 특징은 구체적으로 기술되어 있으며, 원인 요소에 대한 노출 여부를 제외하고는 비슷한가?

두 군 간에는 원인 요소에 대한 노출의 여부를 제외하고 가능한 어떠한 차이도 없는 것이 좋다. 두 군에서의 참여자 선정과 제외 기준은 동일해야 하며, 양 군의 참여율이 비슷하게 높아야 한다. 참여율은 참여자 선정 기준을 충족시키는 사람 중 실제 연구에 참여한 사람의 비율을 일컫는 것으로, 참여율이 낮거나 양 군간의 참여율에 차이가 크다면 연구 결과가 편향될 수 있기 때문이다²⁹.

(2) 노출 여부 결정과 결과 측정은 두 군간에 동일하게 시행하였는가? (결과 측정자는 노출 여부에 대해 맹검 상태를 유지하였는가?)

만약 결과 측정자가 참여자의 원인 요소에 대한 노출 여부를 안다면, 노출된 참여자에 대해 더 정밀하게 측정하는 경향이 있으며 이를 감시 치우침 (surveillance bias)이라고 한다. 또한 환자도 본인이 원인 요소에 노출되었다는 사실을 안다면 더 열심히 실험에 참여하고 노출을 치료하려는 경향이 강하다. 따라서 측정자나 환자 모두 원인 요소의 노출 여부에 대해서는 맹검 하에서 연구하는 것이 좋다.

(3) 추적 관찰은 충분히 오래, 그리고 완전히 하였는가?

예후에 관한 연구에서와 동일하게, 결과 요소가 발생할 때까지 충분히 오래 추적 관찰해야 한다. 또한 탈락율은 20% 미만인 것이 좋다.

(4) 원인 요소와 결과 요소 간에 인과 관계를 충분히 도출할 수 있는가?

- 원인에 대한 노출은 결과에 선행하는가?

이는 원인이라고 생각하는 요소와 결과라고 생각

하는 요소가 인과 관계에 있는지, 아니면 단순한 상관 관계에 있는지를 가려내는 중요한 요소이다. 코호트 연구에서는 원인 요소와 결과 요소의 전후 관계가 분명하기 때문에 이를 밝혀내는 것은 어렵지 않지만, 환자-대조군 연구에서는 결과를 아는 상태에서 원인을 추정하는 것이기 때문에 어려울 수 있다.

• 용량-반응 관계가 있는가?

원인이라고 생각되는 요소에 더 많이, 또는 더 오래 노출되었다면 결과의 발생 빈도가 높아지거나 그 정도가 심해져야 한다.

• 인과 관계는 생물학적 관점에서 타당한가?

다른 인과 관계에 대해 다양하게 추론해 보았는지, 우리의 상식적인 생물학적 지식에 견주어 인과 관계가 타당한지 평가해야 한다.

(5) 연구의 결과는 중요하며 내 진료에 도움이 되는가?

연구 결과 질환이나 특정한 병적 상태를 유발할 수 있는 원인 요소를 제거하거나 줄이면 환자에게 실제로 얼마나 도움이 될 것인지와 연구의 외부 유효성을 평가한다.

결 론

지금껏 많은 의학 문헌이 오류를 포함하고 있다는 점과, 개별 임상가들은 이러한 오류를 발견할 수 있는 능력이 부족하다는 점 때문에 근거 중심 치의학에서는 비판적 평가의 과정을 매우 중요시해 왔으며, 따라서 본 논문에서도 이에 대해 매우 많은 비중을 두어 설명하였다. 실제로 엄격한 비판적 평가를 위해서는 수십, 수백 가지 항목에 대해 여러 가지 복잡한 수식과 이론을 이용한 방법을 적용해야 하지만, 본 논문에서 설명한 것들은 이중 가장 중요하고 직관적으로 이해하기 쉬운 항목들로,

이들만을 이용하더라도 해당 문헌에 대한 대략적으로 중요한 비판적 평가는 가능할 것이다. 실제로 단지 세 가지 항목- 무작위, 맹검, 그리고 참여자의 중도 탈락 정도-만으로도 무작위 대조 연구의 질을 확연히 평가할 수 있다는 보고가 있었다³⁸. 일단 비판적 평가의 전반적인 과정에 익숙해지면, 근거 중심 치의학에 연관된 수식들과 통계적 방법을 익히는 것이 좋다. 이를 통해 연구 결과의 유효성을 더 명확히 파악할 수 있으며 따라서 더 명확한 근거 하에 임상 연구를 활용할 수 있는 능력이 생기게 된다. 이에 관해 좀 더 심도 깊은 지식을 원한다면 Canadian Medical Association Journal에 Basic Statistics for Clinician이란 제목으로 연재되었던 논문들을 참고할 것을 추천한다³⁹⁻⁴².

비판적 평가의 과정을 통해 연구의 질과 임상 상황에의 적용 가능성을 평가한 후에는, 임상 치과의사 본인의 경험 하에서 이를 환자의 선호도 및 임상적 환경과 통합하여 실제 임상 진료에 적용한다. 또한 이렇게 연구 결과를 임상 진료에 적용한 후에는 그것이 성공적이었는지, 또는 개선할 점은 없었는지를 평가하여 차후의 진료에 참고하도록 한다. 이는 근거 중심 치의학의 마지막 두 단계인 적용하기와 평가하기에 해당하는 과정들로, 이들 과정은 엄격한 과학적 문제가 아닌, 개별 치과의사의 가치 판단과 관계된 문제이다. 이를 위해 임상 치과의사 개개인이 근거 중심 치의학을 실제로 여러 번 시행하며 경험을 쌓아갈 것을 추천하며 따라서 이들 두 과정에 관해서는 더 이상 논하지 않기로 한다. “근거 중심 치의학 (Evidence-Based Dentistry: EBD)”은 일부 치과의사들만을 위한 것이 아니라, “모든 치과의사의 치의학 (Everybody’s Dentistry: EBD)”이라는 점을 명심하고⁴³ 많은 치과의사들이 실제 임상 진료 과정에서 근거 중심 치의학의 원리를 사용할 것을 추천하는 바이다.

참 고 문 헌

1. Altman DG. The scandal of poor medical research. *BMJ* 1994;308:283-284.
2. Hill A, Spittlehouse C. What is critical appraisal? Hayward Medical Communications. Available at "www.evidence-based-medicine.co.uk/ebmfiles/WhatisCriticalAppraisal.pdf".
3. Greenhalgh T. How to read a paper: the basics of evidence-based medicine. London: BMJ Publishing, 2006
4. CRITICAL APPRAISAL; User Guides to the Medical Literature (JAMA). Accessible at "www.shef.ac.uk/scharr/ir/userg.html".
5. Centre for Evidence-Based Medicine; Critical appraisal worksheets. Available at "www.cebm.utoronto.ca/teach/materials/caworksheets.htm".
6. Levin KA. Study design I. *Evid Based Dent*. 2005;6:78-79.
7. Jadad A. Bias in RCT's: beyond the sequence generation. In: *Randomized Controlled Trials: A User's Guide*. London: BMJ Publishing; 1998. p. 28-45.
8. 안재억, 유근영, 이중환. 의학 보건학 통계분석. 서울: SPSS 아카데미, 2000.
9. Moher D, Schulz KF, Altman D; CONSORT Group. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA*. 2001;285:1987-91.
10. Des Jarlais DC, Lyles C, Crepaz N; TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *Am J Public Health*. 2004;94:361-6.
11. Bossuyt PM, Reitsma JB, Bruns DE, et al.; Standards for Reporting of Diagnostic Accuracy. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Standards for Reporting of Diagnostic Accuracy*. *Clin Chem*. 2003;49:1-6.
12. Moher D, Cook DJ, Eastwood S, et al. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Quality of Reporting of Meta-analyses*. *Lancet*. 1999;354:1896-900.
13. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA*. 2000;283:2008-12.
14. Woolf SH, Battista RN, Anderson GM, et al. Assessing the clinical effectiveness of preventive maneuvers: analytic principles and systematic methods in reviewing evidence and developing clinical practice recommendations. A report by the Canadian Task Force on the Periodic Health Examination. *J Clin Epidemiol* 1990;43:891-905.
15. Sackett D. Rules of evidence and clinical recommendations. *Can J Cardiol*. 1993;9:487-9.
16. Cook DJ, Guyatt GH, Laupacis A, Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest* 1992;102:305S-311S.
17. Sackett DL, Richardson WS, Rosenberg WMC, Haynes RB. *Evidence-based medicine: how to practice and teach EBM*. London: Churchill-Livingstone, 2000.
18. Attia J, Page J. A graphic framework for teaching critical appraisal of randomised controlled trials. *Evid Based Med* 2001;6:68-69
19. Oxman AD, Guyatt GH. The science of reviewing research. *Ann N Y Acad Sci* 1993;703:125-133
20. Oxman AD, Cook DJ, Guyatt. GH. User's guides to the medical literature. VI: how to use an overview. *JAMA* 1994;272:1367-1371.
21. Moher D, Cook DJ, Eastwood S, et al. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Quality of Reporting of Meta-analyses*. *Lancet*. 1999;354:1896-1900.
22. Sutherland SE. Evidence-based dentistry: Part V. Critical appraisal of the dental literature: papers about therapy. *J Can Dent Assoc*. 2001;67:442-445.

참 고 문 헌

23. Thompson SG. Why and how sources of heterogeneity should be investigated. In: Egger M, Davey Smith G, Altman DG, eds. Systematic reviews in health care: meta-analysis in context. London: BMJ Publications, 2001
24. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000;283:2008-2012.
25. Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. Eval Health Prof 2002;25:76-97.
26. Victora CG, Habicht JP, Bryce J. Evidence-based public health: moving beyond randomized trials. Am J Public Health. 2004;94:400-405.
27. Guyatt GH, Sackett DL, Cook DJ. User's guides to the medical literature. II: how to use an article about therapy or prevention. JAMA 1993; 270: 2598-2601 and 271: 59-63.
28. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodologic quality associated with estimates of treatment effects in controlled trials. JAMA. 1995;273:408-12.
29. SIGN 50: A guideline developers' handbook. Available at "www.sign.ac.uk/guidelines/fulltext/50/index.html".
30. Richards D. Outcomes, what outcomes? Evid Based Dent. 2005;6:1.
31. Harradine NW, Pearson MH, Toth B. The effect of extraction of third molars on late lower incisor crowding: a randomized controlled trial. Br J Orthod. 1998;25:117-22.
32. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? JAMA. 1994;271:389-391.
33. Jaeschke R, Gordon H, Guyatt , Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. what are the results and will they help me in caring for my patients? JAMA. 1994;271:703-707.
34. Laupacis A, Wells G, Richardson S, Tugwell P. Users' guides to the medical literature. V. How to use an article about prognosis. JAMA. 1994;272:234-237.
35. Levin KA. Study design II. Issues of chance, bias, confounding and contamination. Evid Based Dent. 2005;6:102-103.
36. Lambert PM, Morris HF, Ochi S. The Influence of Smoking on 3-Year Clinical Success of Osseointegrated Dental Implants. Ann Periodontol. 2000;5:79-89
37. Levine M, Walter S, Lee H ,et al. Users' guides to the medical literature. IV. How to use an article about harm. JAMA. 1994;271:1615-1619.
38. Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials. 1996;17:11-12.
39. Guyatt G, Jaeschke R, Heddle N, et al. Basic statistics for clinicians: 1. Hypothesis testing. CMAJ. 1995;152:27-32.
40. Guyatt G, Jaeschke R, Heddle N, et al. Basic statistics for clinicians: 2. Interpreting study results: confidence intervals. CMAJ. 1995;152:169-173.
41. Jaeschke R, Guyatt G, Shannon H, et al. Basic statistics for clinicians: 3. Assessing the effects of treatment: measures of association. CMAJ. 1995;152:351-357.
42. Guyatt G, Walter S, Shannon H, et al. Basic statistics for clinicians: 4. Correlation and regression. CMAJ. 1995;152:497-504.
43. Richards D. EBD--everybody's dentistry. Evid Based Dent. 2005;6:57.