REVIEW ARTICLE

1

투고일 : 2014. 5. 13 심사일 : 2014. 5. 16 게재확정일 : 2014. 7. 2

치의학 분야에서의 메타 분석

전남대학교 치의학전문대학원 교정학교실, 치의학 연구소
임 회 정

ABSTRACT

Meta-analysis in dental research

Department of Orthodontics, Chonnam National University School of Dentistry, Dental Science Research Institute
Hoi-Jeong Lim, Ph.D.

Recently, importance of meta-analysis is increasing in the field of dentistry, since it is not easy to settle controversies arising from conflicting studies. Meta-analysis is the statistical method of combining results from two or more individual studies that have been done on the same topic. Merits of meta-analysis include an increase in power, an improvement in precision, and the ability to address solution not provided by individual studies. However, it might mislead researchers when variation across studies and publication bias are not carefully taken into consideration. The purpose of this study is to help understand meta-analysis by making use of individual results in dental research paper.

Key words : Meta-analysis; systematic review; dental research

Corresponding Author
Hoi-Jeong Lim, Ph.D.
Department of Orthodontics, Chonnam National University School of Dentistry
Dental Science Research Institute
33 Yongbong-ro, Buk-gu, Gwangju 500-757, Korea
Tel : +82-62-530-5830 Fax : +82-62-530-5659, E-mail : hjlim@jnu.ac.kr

이 논문은 2010년도 전남대학교 학술연구비 지원 (#2010-2709) 에 의하여 연구되었음.

Ⅰ. 서론

1976년 Glass에 의해 메타분석이라는 용어가 처음 사용되었고, 1980년대 중반 이후부터 의학 분야에서, 2000년도 이후부터는 치의학 분야에서 본격적으로 사용되기 시작하여 현재는 메타분석 연구가 기하급수적으로 증가하고 있는 실정이다(Fig. 1). 메타분석은 각 연구들의 상반된 결론이나 논쟁이 야기되었을

때 혹은 짧은 기간에 적은 예산으로 문제의 해결을 요구할 때 효과적으로 사용되고 있다. 요즘 치의학 분야에서도 비슷한 주제를 가진 연구가 계속 증가하여 기존의 문헌 연구로는 경향을 파악하기 힘들고 관련 정보가 많아져 메타 분석의 필요성이 증가하고 있다.

메타분석은 과거 수십 편의 동일하거나 유사한 주제로 실시된 연구에서 개별연구에 나타난 실험결과를 공통된 효과크기로 전환하여 실험결과를 객관적이고 계량적으로 일반화시키는 분석이다. 메타분석의 장점으로는 개별 연구들을 합쳐 보다 큰 표본크기로 가설을 검정하기 때문에 통계적 검정력을 높일 수 있고, 여러 개별 연구들을 종합하여 효과크기의 평균치를 계산하는 것이므로 개별 연구의 효과 크기보다 더 정확한 효과크기를 추정할 수 있다. 또한 서로 상충하는 연구 결과가 있을 때 그 원인이 무엇인지 알아 볼 수 있다는 점 등을 들 수 있다. 메타분석의 단점으로는 사과와 오렌지를 섞는 것처럼 서로 비교할 수 없는 다른 성질의 연구결과들을 종합하려는 데에 문제가 있을 수 있으나 극단적인 자료는 제외하는 것으로 이를 해결할 수 있고, 질이 높은 연구와 낮은 연구를 구별하지 않고 종합하려는 데에 문제가 있을 수 있으나 개별 연구의 질적 수준에 따라 적절한 비중을 두고 종합하는 것으로 해결할 수 있다. 또한 출판된 연구만을 연구 대상으로 하기 때문에 (publication bias) 대표성에 문제가 있을 수 있으나 출판되지 않은 연구들도 연구 대상으로 포함시켜 이를 해결할 수 있다[1].

이 연구의 목적은 메타분석을 수행하기 위한 절차를 자세히 다루고 치의학 연구에서 메타분석이 어떻게 사용되었는지 알아보고자 하는 것이다.

Ⅱ. 체계적 문헌 고찰의 절차

1. 연구의 가설 설정

영가설이나 대안가설을 설정한다.

치의학 논문을 예로 들어 보자. 체계적 문헌 고찰을 한 연구[2]의 목적은 각 출판된 연구로부터 치아의 총생과 치아 우식증과의 관련성을 알아보고자 하는 것이었다. 체계적 문헌 고찰과 더불어 메타분석을 한 연구의 예로는 첫째[3], 이미 출판되어있는 통제된 혹은 통제되지 않은 전향적 임상시험들로부터 메타분석을 통해 미니스크류 식립 실패에 영향을 미치는 위험요인이 무엇인지 알아보는 것이었고, 둘째[4], 부착(bonding) 실패율과 부착하는데 걸리는 시간에 대한 SEP (Self-etch primers)와 AE (Acid-etch technique)방

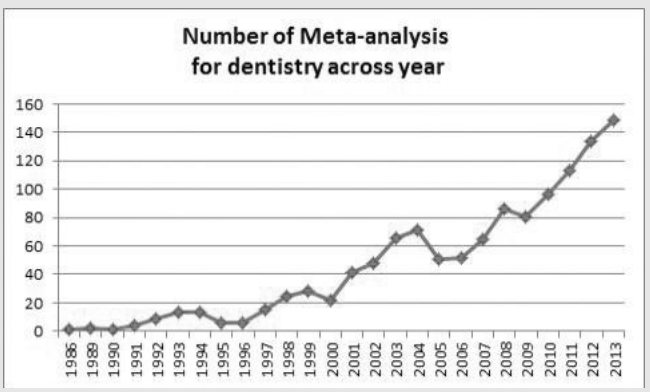

Fig. 1. Number of meta-analysis for dentistry across the year

법에 대한 비교를 메타분석 하고자 하는 것이었다.

2. 연구의 포함기준 및 제외기준 설정 후 문헌 탐색 및 논문 선택

논문 선택 시 선정된 주제와 관련된 발표 혹은 미발표 연구물들을 빠짐없이 수집함으로써 보다 객관적인 결론을 내릴 수 있는 근거를 제공해야하며 연구물의 대표적인 자료수집의 출처[1]로는 국내 및 국외의 석박사 학위논문들, 학회에서 발표된 논문들, 참고문헌, 서적, 연구물들의 목록들, 데이터베이스 활용, 현재 출판 예정 중에 있는 논문들이나 미발표 논문 등을 들 수 있다. 미발표 논문까지도 포함해야 하는 이유는 학술지에 게재된 논문들은 유의한 결과가 입증된 논문들을 선호하는 속성이 있기 때문에 효과 크기가 과대평가된 경우를 배제하기 어렵기 때문이다. 그러나 방대한 자료를 모두 수집하기란 경제적인 면에서나 시간과 노력의 측면에서도 쉬운 일이 아니다. 이런 경우 무작위로 연구논문들의 제한된 수를 표집하는 방법, 연도순에 따라 일정 기간으로 연구의 범위를 제한하거나 연구주제를 구체적으로 잡아 선행연구 결과들의 수집을 제한하는 등 분석의 적정 규모를 선정할 수 있다.

연구 자료들의 소재를 파악하는 방법(Cooper(1982))[5]으로는 첫째, 상향(ancestry) 추적 접근방법인 한 연구 발표 논문에서 인용한 참고문헌들을 근거로 하여 관련 주제의 연구들을 추적하는 방법이 있고, 둘째, 하향(descendency) 추적 연구방법인 Pubmed나 Medline 등의 컴퓨터 데이터 베이스를 이용한 정보검색방법으로 우리가 원하는 결과들을 빠르게 검색해 낼 수가 있어 메타분석 분야의 연구가 활발히 전개되고 있는 실정이다.

Fig. 2b의 치의학 저널의 플로우챠트[2]를 Fig. 2a을 참고하여 살펴보면, 데이터베이스 검색을 통해 확인된 문헌 수는 6911개였고, 다른 정보원에서 확인된 추가 문헌 수는 3개였다. 중복 제거 후 남은 문헌 수는 3820개였고, 그 중 배제된 문헌 수는 3801개였다. 3727개의 문헌은 관련성이 없었고, 56개는 영어로 쓰이지 않았고, 15개는 유병률(prevalence)만을 계산한 문헌이었고, 3개는 literature reviews, editorials, 혹은 expert opinion에 속하는 문헌이었다. 따라서 19개의 원문을 얻을 수 있었고, 그 중 11개가 배제되었다. 1개의 원문은 비교군이 없었고, 8개

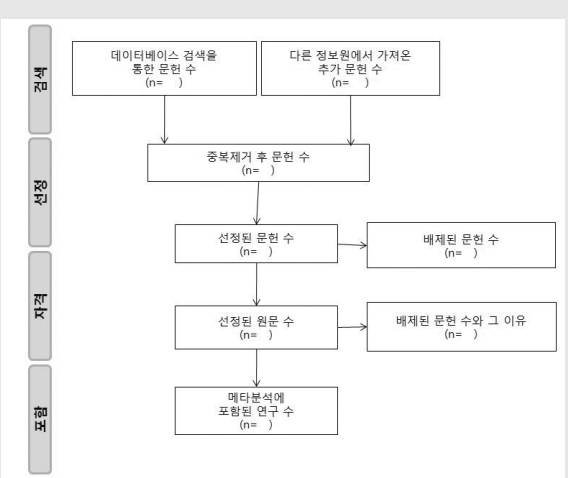

Fig. 2a. PRISMA Flow Chart

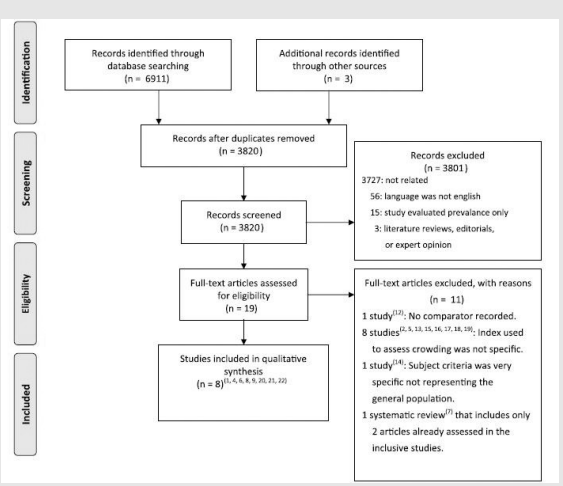

Fig. 2b. PRISMA Flow Chart[2]

는 crowding을 평가하는 index가 분명하지 않았고, 1개는 일반적인 모집단이 아닌 특정한 집단을 연구대상으로 삼았고, 나머지 1개는 체계적 문헌고찰을 한 연구였다. 결과적으로 메타분석에 포함된 연구 수는 8개였다.

3. 연구의 특성 추출 및 연구의 자료 코딩

1) 연구 특성의 추출

일반적으로 연구 논문의 발표일자, 연구대상의 연령계층, 남녀간의 표집 비율, 표본 크기, 실험 대상의 탈락, 연구 대상의 무선 혹은 비무선 표집, 연구 논문의 발표 형태(학술지, 단행본, 학위논문 등) 등을 추출할 수 있다. 아래의 Table 1[2]을 살펴보면 각각의 연구에서 연구 대상자의 나이, 성별, 인종, 표본크기, 구강상태 등을 추출한 것을 알 수 있다.

2) 메타분석용 코딩 매뉴얼과 코딩표의 개발

intra-rater error와 inter-rater error를 피하기 위해서는 코딩 매뉴얼의 작성이 무엇보다 중요하다. 코딩 매뉴얼이 구체적이고 명료하게 만들어지면 코딩의 오류를 방지할 수 있게 되고 결국 메타분석 연구의 신뢰도를 높일 수 있다. 코딩 작업은 한 사람 이상의 평가자에 의해 분류 및 코딩 작업이 이루어지고 평가자 상호간의 agreement를 계산하여 보고 해야 한다. 아래의 코딩 매뉴얼은 Lipsey et al. (2000)

Table III. Subject criteria in the 8 studies

	Subject criteria				
Study	*Age*	*Sex*	*Race*	*Sample size*	*Baseline oral health status*
Hixon et al[4]	Severe crowding: not reported Excellent occlusion: mean age, 18.3 years	Severe crowding: not reported Excellent occlusion: 61% males	White	Severe crowding: 20 Excellent occlusion: 106	Not reported
Roder and Arend[22]	14-16 years	Girls	Not reported	Subgroups: not reported	Not reported
Katz[8]	<25 years	Not reported	White	Subjects: 160 Subgroups: not reported	Adequate condition of occlusion and teeth present
Addy et al[1]	11.5-12.5 years	Role of sex is insignificant	Not reported	2656 pairs of contralateral crowded and uncrowded teeth	Not reported
Helm and Petersen[20]	33-39 years (mean age, 35.5 years)	Not reported	Not reported	Maxillary incisor crowding: 33 Mandibular incisor crowding: 39 Maxillary crowding: 41 Mandibular crowding: 51 Control group: 27	Not reported
Stahl and Grabowski[21]	Children with primary dentitions (mean age, 4.5 years) and mixed dentitions (mean age, 8.9 years)	Girls: 4306 Boys: 4558 Subgroups: not reported	Not reported	Subgroups: not reported	Not reported
Staufer and Landmesser[6]	Age groups, 18-34 years (n = 63) and ≥35 years (n = 62)	Women: 63 Men: 62 Subgroups: not reported	Not reported	18-34 years (n = 63) ≥35 years (n = 62)	Occlusal stability for a physiologically supported jaw and good state of care of mandibular canines and incisors
Alsoliman[9]	9-13 years	Subgroups: not reported	Not reported	Not reported	Not reported

Table1. General characteristics of included studies[2]

의 것을 번역한 것으로 아래 항목들을 살펴보고 본인의 코딩 매뉴얼을 개발하면 된다. 아래 항목들에는 논문 관련 정보, 표본의 일반 특성, 연구 디자인 특성, 종속변수 특성, 효과크기, 평균과 표준편차나 빈도 혹은 비율, 유의성 검정 결과, 계산된 효과크기 등을 포함하고 있고 연구 특징에 따라 코딩 매뉴얼을 바꿀 수 있다(Fig. 3).

논문 제목 : _____

_____

_____

_____ 1. 연구 ID 번호
2. 저자명 _____
3. 저널명 혹은 책제목과 페이지넘버 _____
_____ 4. 출판 연도
5. 국가 _____
_____ 6. 출판형태
(1) 책 (2) 학위 논문 (3) 저널에 출판된 논문 (4) book chapter (5)기타

표본의 일반 특성
_____ 7. 평균 나이
_____ 8. 인종
(1) >60% 백인 (2) >60% 흑인 (3) >60% 히스패닉
(4) >60% 기타소수인종 (5) <60% 인종이 섞임
(6) 인종이 섞임, 비율 측정 불가 (9) 모름
_____ 9. 성별
(1) <5% 남성 (2) 5%~50% 남성 (3) 50% 남성 (4) 50%~95% 남성
(5) >95% 남성 (9) 모름

연구 디자인 특성
_____ 10. 연구 디자인 종류
(1) case report (2) case-control study
(3) prospective study (4) Randomized Controlled Trial(RCT)
_____ 11. 조건에 따른 할당 방법:
(1) 매칭, 계층별, 블록화에 의한 무작위(random) 할당
(2) 단순 할당
(3) 비임의(nonrandom) 할당 후 사후 매칭
(4) 비임의 할당
(5) 기타 _____
(9) 언급 안함
_____ 12. 검정하기 전에 그룹들의 특성이 같은지 검정했는가?
(1) 네 (2) 아니오
_____ 13. 검정했다면 검정하기 전 검정 결과는?
(1) 차이가 없었다. (2) 중요한 차이가 있었다.
_____ 14. 전체 표본 크기
_____ 15. 환자 그룹(Case group) 표본 크기

_____ 16. 대조군 그룹(Control group) 표본 크기

종속 변수 특징
_____ 17. 효과크기(effect size) 종류
(1) 사전 비교 (2) 사후 비교 (3) 추적 관찰(follow-up) 비교
_____ 18. 실험 기간
19. 결과 변수(outcome measure) _____

효과 크기
_____ 20. 효과크기의 종류
(1) 평균과 표준편차 (2) t값 혹은 f값 (3) x^2값(df=1)
(4) 이분형(dichotomous) 일 때 빈도 혹은 비율
(5) 다분형(polychotomous) 일 때 빈도 혹은 비율
(6) 기타 _____
_____ 21. 효과크기가 발견된 페이지 넘버
_____ 22. 이 연구 자료는 어떤 그룹에서 더 성공적으로 나타났는가
(1) 치료그룹 (2) 대조군(control group) (3) 차이 없음 (9) 언급 안함

평균과 표준편차
_____ 23a. 치료군 평균
_____ 23b. 대조군 평균
_____ 24a. 치료군 표준편차
_____ 24b. 대조군 표준편차

빈도 혹은 비율
_____ 25a. 성공한 결과를 가진 치료군의 표본크기(N)
_____ 25b. 성공한 결과를 가진 대조군의 표본크기(N)
_____ 26a. 성공한 결과를 가진 치료군의 비율
_____ 26b. 성공한 결과를 가진 대조군의 비율

유의성 검정
_____ 27a. t값
_____ 27b. f값
_____ 27c. x^2값(df=1)

계산된 효과크기
_____ 28. 효과크기
_____ 29. 효과크기 계산 시 신뢰도
(1) high (2) moderate (3) some (4) slight (5) no estimation

Fig. 3. Manual Coding

출처 : Mark W. Lipsey & David B. Wilson(2000), pp. 231-232.[6]

4. 연구의 질 평가 (Quality assessment of included studies)

포함된 개별 연구들의 질 평가는 체계적 문헌 고찰에서 중요한 부분이다. 연구의 질 평가는 각 연구의 상황에 따라 개발되었으므로 각각의 상황에 따라 다른

평가가 이루어지고 연구의 종류, 대조군 여부, 결과변수 선택의 적절성, 표본크기 적절성, 측정 오차 측정 여부, 통계방법의 적절성 등의 질문으로 이루어진다. 예를 들어 메타분석을 한 치의학 논문[7]의 연구의 질 평가에 대해 살펴보면 아래의 항목에 대해 점수를 부여한 후 총점수를 구하여 연구의 질을 평가하였다.

1) Selection process의 서술 : 서술하지 않음(0), 부분적으로 서술(1), 자세히 서술(2)
2) 전향적인가 후향적인가 : 후향적 연구(0) 혹은 전향적 연구(2)
3) Consecutive cases : 불연속(unconsecutive)(0), 연속(consecutive)(1)
4) 표본크기(N) : N〈20(0) 혹은 N≥20(1)
5) 결과변수(outcome measure)의 선택 : 부적절(0), 부분적으로 적절(1), 적절(2)
6) Measurement error의 적절성 : 측정되지 않음(0), 부분적으로 적절(1), 적절(2)
7) 통계방법의 적절성 : 부적절(0), 적절(1)

total score : 7미만(Low), 7~8(Medium), 9~10(Medium-high), 11(High)

8개의 연구가 7점 미만(Low)을 받았고, 4개의 연구가 7~8점(Medium)을 받았다. 결론적으로 질이 좋은 연구들(High)이 아니기 때문에, 이 결과물들을 조심스럽게 해석해야 하며 이 결과물들로부터 내릴 결론은 확실하다고 할 수 없다.

또 다른 치의학 논문의 Table 2[2]의 연구의 질 평가에 대해 살펴보면 위의 평가 요인과는 다소 다르다는 것을 알 수 있다.

Table I. Quality assessment of included studies

Study	Study type[a]	Blinding[b]	Adequate reporting[c]	Comparator[d]	Validity, reliability of caries recording method[e]	Validity, reliability of crowding recording method[f]	Error of measurement[g]	Confounding factors[h]	Subgrouping[i]	Coding[j]	Score/grade
Hixon et al[4]	1	-	2	3	3	1	2	3 race, missing teeth, sex; NS	1 crowding	-	16/moderate
Roder and Arend[22]	1	-	2	2	1	2	-	2 missing teeth, sex; NS		-	10/moderate
Katz[8]	1	-	2	1	3	2	-	4 race, sex, missing teeth, baseline oral status	1 crowding	-	14/moderate
Addy et al[1]	1	-	2	2	1	1	-	4 race, sex, missing teeth, baseline oral status; NS		-	11/moderate
Helm and Petersen[20]	1	2	3	2	1	1	-	3 age, sex, missing teeth		-	13/moderate
Stahl and Grabowski[21]	1	-	2	2	1	1	-	1 missing teeth		-	8/low
Staufer and Landmesser[6]	1	-	3	3	1	2	1 crowding	3 age, missing teeth, baseline oral status	2	-	16/moderate
Alsoliman[3]	1	-	2	2	1	1	1 caries	2 missing teeth, sex; NS	-	1	11/moderate

NS, Not significant.
[a]Longitudinal study: 2 points; cross-sectional study: 1 point (maximum 2 points); [b]Blinding: 2 points; [c]Adequate reporting on subject criteria, subject distribution, and methodology: 1 point for each factor (maximum 3 points); [d]Control group of normal occlusion or minimal crowding (up to 2 mm): 2 points; control of different grades of crowding: 1 point (maximum 3 points); [e]Radiographic assessment: 2 points; visual examination: 1 point (maximum 2 points); [f]Quantitative index: 2 points; visual: 1 point (maximum 2 points); [g]Error of measurement for caries detection and measurement of crowding: 1 point for each factor (maximum 2 points); [h]Effect of confounding factors considered (age, race, sex, missing teeth, and baseline oral status): 1 point for each factor. When there was no sex-related difference between caries incidence, 1 point was given because no correction was needed. When confounding factors had no effect (NS): 1 point was given for each factor (maximum 5 points); [i]Subgrouping subjects to compare the effect of age and severity of crowding: 1 point for each factor (maximum 2 points); [j]Coding of subjects and variables: 1 point.

Table2. Quality assessment of included studies[2]

1) 연구의 종류 : 횡단연구(cross-sectional study)(1), 종단연구(longitudinal study)(2)
2) 눈가림(Blinding) 여부 : 눈가림이 아니면(0), 눈가림이면(2)
3) 연구대상자 선정 기준에 대한 적절한 보고 : 연구대상자 분포와 방법론을 서술하면 각각의 요인에 대해(1) (최대 3점)
4) 대조군 : 총생의 다른 등급을 가진 대조군이면(1), 정상 교합 혹은 2mm까지의 총생을 가진 대조군이 있으면(2) (최대 3점)
5) 충치 우식증을 보고하는 방법의 타당성 및 신뢰도 : 눈으로 평가(1), 방사선 평가(2)
6) 총생을 보고하는 방법의 타당성 및 신뢰도 : 눈으로 평가(1), 양적인 지수 (Quantitative index)(2) (최대 2점)
7) 측정 오차 (measurement error) : 충치 발견과 총생의 측정 오차 각 요인마다(1) (최대 2점)
8) 고려되어지는 교란변수들의 효과(age, race, sex, missing teeth, and baseline oral status) : 각각(1)
9) subgrouping : 나이효과와 총생의 정도를 비교하기 위한 연구대상자의 소그룹화, 각 요인마다(1) (최대 2점)
10) 코딩 : 연구대상자와 변수들에 코딩이 되어져 있지 않으면(0), 있으면(1) total score : 1~8(Low), 9~16(Moderate), 17~24(High)

1개의 연구가 1~8점(Low)을 받았고, 7개의 연구가 9~16점(Moderate)을 받았다. 역시 마찬가지로, 질이 좋은 연구들(High)이 아니기 때문에, 이 결과물들도 조심스럽게 해석해야 한다.

Ⅲ. 메타분석 수행의 절차

체계적 문헌고찰을 수행 후 코딩된 자료를 가지고 메타분석 수행에 따른 세부적인 절차는 다음과 같다.

1. 메타분석이 요구하고 있는 기본 전제를 충족하고 있는지를 아래와 같이 검토한다.

통합하고자 하는 주제의 선행 연구물들의 수가 충분해야 하며 메타분석에 사용될 선행연구의 연구 설계방식은 통제집단과 실험집단이 존재하는 실험연구여야 한다. 또한 각 집단의 평균과 표준편차, 표본크기 및 유의도 수준이 밝혀진 연구여야 하고 위의 정보가 제시되지 않더라도 T검정, F검정, 상관계수(r) 등에 의한 통계값이 제시된 연구여야 한다[8].

2. 결과변수의 자료는 이분형 자료, 연속형 자료, 순서형 자료 및 생존형 자료 등으로 구분될 수 있으며 이들과 관련된 요약통계량 유형을 검토하여 사전에 어떤 요약 통계량, 즉 효과크기로 결과들을 산출할지 정한다. 효과크기에는 표준화된 평균차(standardized mean difference: SMD), 오즈비(odds ratio: OR), 위험비(risk ratio: RR), 상관계수(r), 비율, p값 등이 있을 수 있다. 여기서는 연속형 자료와 이분형 자료에 대해서만 비교, 검토 할 예정인데 연속형 결과변수는 평균차이(mean difference)로 평가되는 반면 이분형 변수의 결과 변수는 비(ratio)로 표현된다.

치의학 예제[8]를 살펴보면 미니스크류 임플란트(MI) 그룹과 기존에 사용되었던 고정원 그룹(Conventional Anchorage group : CA)의 비교에서 각각의 고정원 손실 (anchorage loss)의 평균과 표준 편차, 표본크기를 추출하여 아래 수식에 있는 표준화된 평균 차이(SMD)를 효과크기(Effect Size: ES)로 계산 하였다.

$$ES_{sm} = \frac{\overline{X_e} - \overline{X_c}}{S_{pooled}}$$

$$\text{where } S_{pooled} = \sqrt{\frac{(n_1-1)S_1^2 + (n_1-1)S_2^2}{n_1+n_2-2}}$$

다른 치의학 예제[4]도 살펴보면 교정용 본딩을 위한 self-etch primer(SEP) 그룹과 기존에 사용되었던 acid-etch(AE) 방법을 사용한 그룹의 접착 실패율을 비교하고자 SEP로 bonding한 브라켓 빈도수와 접착에 실패한 빈도수(bond failure), AE 방법으로 bonding한 브라켓 빈도수와 접착에 실패한 빈도수를 각각의 논문에서 추출하여 아래 수식에 있는 승산비(OR)의 효과크기로 계산하였다.

$$ES_{OR} = \frac{ad}{bc}$$

효과크기의 개념은 각기 다른 연구 결과들을 수량적으로 통합하고자 할 때 하나의 공통 척도로 나타내어 의미있게 비교할 수 있도록 한 것이다.

3. 각 논문의 효과크기들의 동질성 검정을 시행한다.

이질성(heterogeneity)이란 체계적 문헌고찰에서 얻어진 개별연구들 사이에 효과크기가 통계적으로 서로 다른 것을 말한다. 이질성이 나타나는 원인 중에는 부적절하게 효과크기를 선택했거나 한 두 개의 극단적인 연구가 존재했을 때가 있을 수 있다. 자료형태에 따른 효과크기를 결정하여 분석을 수행한 후 산출된 결과에 대한 통계적 이질성을 검토해야 한다. 대표적인 방법으로는 forest plot에 포함된 개별 연구들의 신뢰구간들의 겹침 정도에 따라 이질성을 시각적으로 평가하는 방법 (거의 겹치지 않는다는 것은 이질성이 존재한다는 것) 과 통계적 검정으로 Q-검정을 사용하는 방법이 있는데 이에 대한 귀무가설은 개별연구 간에 모든 효과크기가 같다는 것이고 대립가설은 적어도 하나의 효과크기는 다르다는 것이다. 포함되는 개별 연구의 수가 적어 통계적 검정력이 낮은 점을 감안하여 통계적 유의수준을 통상적으로 사용되는 0.05 대신 0.1을 사용하고 이질성에 대한 통계량으로 I^2 통계량 (Higgins 등, 2003)이 사용되고, 이는 효과 추정치에서의 변동에 대한 퍼센트로 표현되는 것으로 나타나며 보통 0~25%까지는 낮은 이질성, 25~75%까지는 중간 정도의 이질성, 75%이상은 상당한 이질성이

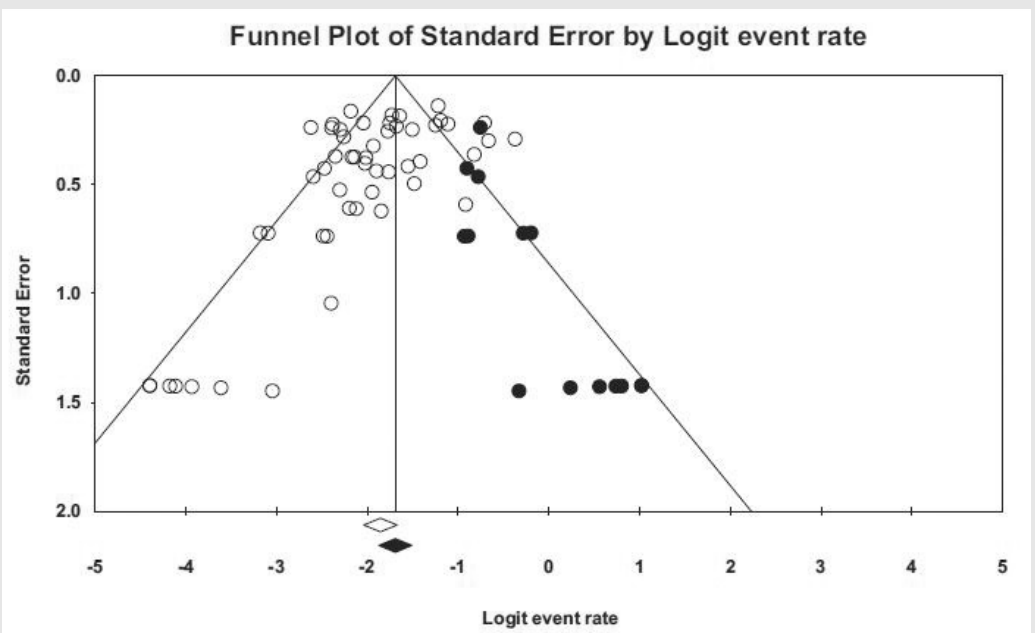

Fig. 4. Funnel plot of the 52 original studies included in the analysis(white dots) and the missing studies imputed by the trim-and-fill procedure (black dots).[3]

있는 것으로 해석된다[9].

이질성 문제가 크지 않을 때 메타분석을 수행할 수 있고, 이질성이 확인될 경우 원인을 파악해야 한다. 이질성의 원인을 탐색하는 방법으로는 하위군 분석, 메타-회귀분석, 민감도 분석 등이 있다. 하위군 분석의 경우 사전에 효과 추정치에 영향을 미칠만한 요인들을 선정한 후 탐색해야 연구 수행에서의 비뚤림 위험을 최소화할 수 있다. 이질성의 원인을 설명하지 못한 경우 일반적으로 메타분석은 시행되지 않으며 질적으로 검토하게 된다.

4. 메타분석에서는 효과 추정치 산출과 함께 funnel plot을 통한 출판 비뚤림(publication bias)의 영향을 확인할 수 있으며 결과에 미치는 다양한 영향을 탐색하면서 연구결과의 신뢰성을 검토하기 위하여 민감도 분석을 수행한다. 일반적으로 질평가 결과를 접목하여 연구의 질이 연구결과에 미치는 영향이 어떠한지 등을 검토한다.

Fig, 4에 의하면 원 연구에서 출판편의에 대한 funnel plot을 그렸을 때 하얀 점들이 비대칭으로 구성되어 출판편의가 존재하였다. 이 경우 출판편의를 제거하기 위해 trim-and-fill procedure를 통해서 대칭이 되도록 까만 점을 채워 넣어 다시 분석을 할 수 있다. 즉, trim-and-fill procedure는 출판 편의에 의해 누락되었을 것으로 예상되는 연구들을 추가하여 좌우대칭으로 만들어 준 후에 effect size(보정된)를 다시 계산해 주는 것을 말한다. 원래의 효과크기와 보정된 효과크기를 비교하여 그 차이가 유의하지 않을 정도로 작다면 원래의 효과크기가 옳다는 것을 좀 더 확신할 수 있다.

5. 메타분석 실시 및 결과 제시

치의학 예제[10]에서 나온 메타분석 결과를 살펴보면 두 그룹간의 고정원 손실 평균차이가 -2.4mm(95% CI=(-2.9, -1.8), p=0.00) 였다. 이렇게 마이너스 부호가 나온 것은 CA집단(Conventional Anchorage Group)의 평균이 MI집단(Miniscrew Implants Group)의 평균보다 높기 때문이다. 즉,

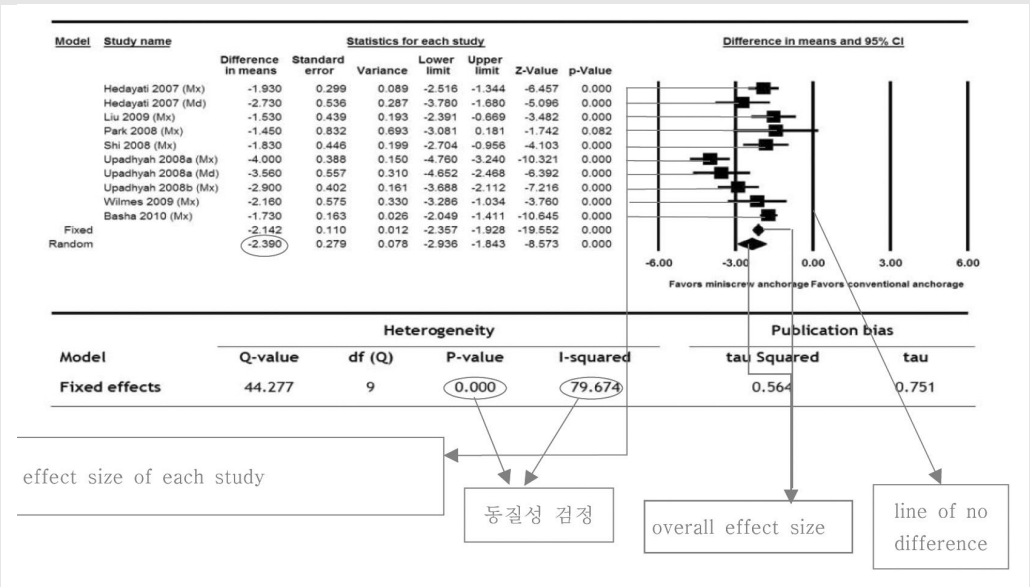

Fig. 5. Forest plot for the mean difference of the anchorage loss between the MI and conventional anchorage groups, including the number of source studies, the effect sizes with the 95% confidence intervals, the assessment of heterogeneity, and the statistical significance.[10]

MI집단의 고정원 손실이 CA집단보다 낮았다는 것을 의미한다.

이 그림은 forest plot 이라고 불리며 각각의 논문에 따른 사각형은 relative weight 혹은 sample size에 따라 크기가 결정되며 사각형 옆의 가로줄은 95% 신뢰구간(confidence interval)을 말한다. 그림에 나타난 마름모의 중심은 각각의 효과크기를 결합한 종합적인 효과크기(overall effect size) 를 말하며 마름모의 가로줄은 종합적인 효과크기의 95% 신뢰구간을 나타낸다. 0이 포함되는 것은 유의한 차이가 없음을 말하는데 여기서 0이 포함되지 않았으므로 두 그룹 간에 고정원 손실의 평균에 있어 유의한 차이가 있음을 의미한다(Fig. 5).

연속적인 결과변수에서는 차이없음을 나타낼 때는 0을 사용하고 이분형 결과변수에서는 효과가 없음을 나타내는 값으로 1을 사용한다.

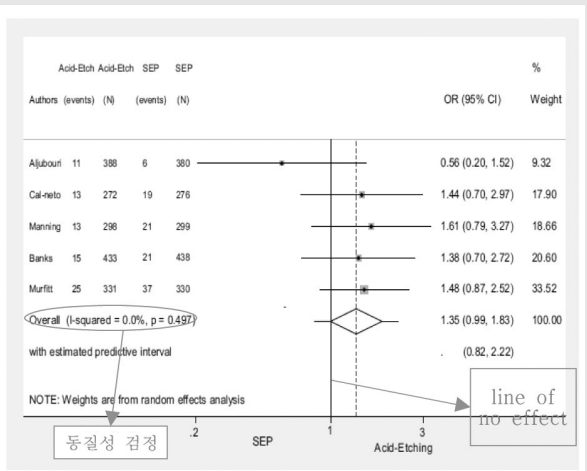

Fig. 6. Random-effects meta-analysis of bracket failure with SEP and AE[4]
N = total number in each group
Events = number of participants with the outcome in each group
Weight = influence of studies on overall meta-analysis

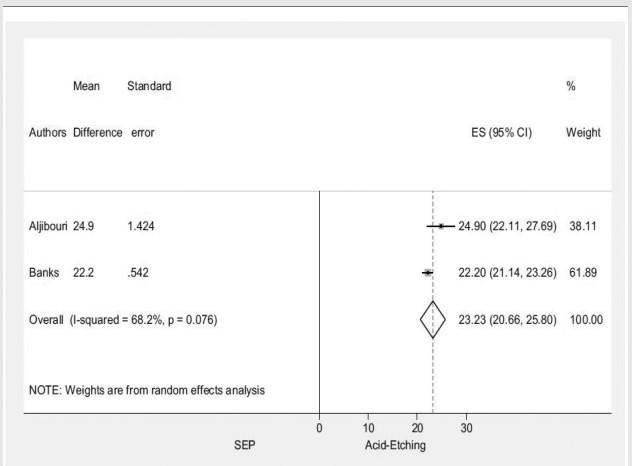

Fig. 7. Random-effects meta-analysis of required time to bond with SEP and AE.[4]

이 예제에서 동질성 검정 결과를 살펴보면 Q 테스트에 해당하는 P값이 0.000으로 매우 유의하다는 것을 알 수 있었다. 또한 I^2통계량이 79.67%로 50%가 넘어가 이질성이 존재한다는 것을 알게 되었다.

치의학 예제[4]에서 나온 메타분석 결과를 살펴보면 Overall effect size = 1.35(95% CI : 0.99-1.83) 으로 두 군 간의 실패율에 유의한 차이를 찾을 수 없었지만 SEP군과 AE군의 본딩에 필요한 시간을 비교한 메타분석 결과는 AE 그룹에서 23배 더 SEP 그룹에 비해 유의하게 많은 시간이 걸렸다는 것을 알 수 있었다. 그러나 Fig. 7에서는 두 개의 연구만으로 메타분석을 시행했기 때문에 SEP그룹에서 시간이 더 절약된다는 충분한 증거가 되지는 못했다.

동질성 검정 결과를 살펴보면 I^2값이 Fig. 6에서는 I^2값이 0%로 이질성이 없는 것으로 나타났고 Fig. 7에서는 I^2값이 68.2%로 이질성이 존재함을 알 수 있었다.

6. 민감도 분석 (Sensitivity Analysis) 실시

민감도 분석이란 메타분석에서 얻어진 연구결과가 맞는지 다시 한 번 확인하는 분석방법이다. 포함된 개별연구 중 한 편의 개별연구를 제거한 뒤 메타분석 연구 결과가 바뀌지 않는다면 그 연구 결과는 매우 안정적(robust)이라고 할 수 있지만, 연구 결과의 전반적인 방향이 바뀐다면 그 한 편의 연구에 의해 다수 논문의 전반적인 경향이 드러나지 않는 것이므로 결과의 안정성이 부족한 것으로 해석할 수 있다. 이 경우 해당 논문에 눈에 띄지 않는 이질성이 존재하지 않는지 검토할 필요가 있다. 아래 Figure 8의 그래프는 Fig. 6의 결과를 민감도 분석한 것이다. 즉 첫 번째 라인은 5개의 논문 중 첫 번째 Aljubouri 등의 논문을 제외한 4개의 논문을 가지고 메타분석을 한 결과이고 두 번째 라인은 Murfitt 등의 논문을 제외한 4개의 논문을 가지고 메타분석을 한 결과이다. 세 번째, 네 번째, 그리고 다섯 번째 라인 전부 마찬가지 방법으로 y축의 저자 논문을 제외한 메타분석 결과를 그린 것이다. 여기선 첫 번째 라인만 원래의 결과와 다르게 AE에서 유의하게 더 낮은 실패율을 보고한다고 하였다. 그 이

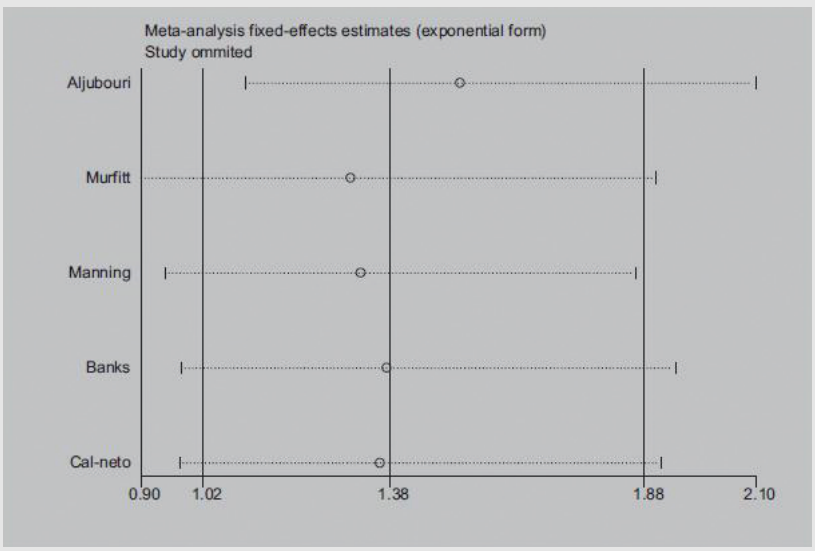

Fig. 8. Meta-analysis to investigate the influence of individual studies on the overall meta-analysis estimate.[4]

유는 Aljubouri 등의 논문에서 SEP에서 더 낮은 실패율을 보고하였기 때문으로 사료된다[4].

Ⅳ. 결론

메타분석 연구는 수년간에 걸쳐 축적된 연구들을 평가하고 분석하는 연구로 현 시점에서 절실히 필요한 연구 기법이고 가치가 높은 연구방법이라고 할 수 있겠다. 치의학 각 분야에서 이 절차들의 설명이 메타분석 수행을 하려는 연구자들에게 작으나마 보탬이 되기를 바란다.

참 고 문 헌

1. 오성삼. 메타분석의 이론과 실제. 1판 4쇄. 건국대학교 출판부. 2011.
2. Hafez HS, Shaarawy SM, Al-Sakiti AA, Mostafa YA. Dental crowding as a caries risk factor: A systematic review. Am J Orthod Dentofacial Orthop 2012;142(4):443-50
3. Papageorgiou SN, Zogakis IP, Papadopoulos MA. Failure rates and associated risk factors of orthodontic miniscrew implaEnts: A meta-analysis. Am J Orthod Dentofacial Orthop 2012;142(5):577-95
4. Fleming PS, Johal A, Pandis N. Self-etch primers and conventional acid-etch technique for orthodontic bonding: a systematic review and meta-analysis. Am J Orthod Dentofacial Orthop 2012;142(1):83-94.
5. Cooper H. Scientific guidelines for conducting integrative research review. Review of Educational Research 1982;52(2):291-302
6. Lipsey MW, Wilson DB. Practical Meta-Analysis (Applied social Research Methods) 264 pages SAGE Publications, Inc. 2000.
7. Fudale P, Antoszewska J. Are orthodontic distalizers reinforced with the temporary skeletal anchorage devices effective? Am J Orthod Dentofacial Orthop 2011;139(6):722-9
8. 박동아. 알기 쉬운 EBM ? 메타분석. 2014. 한국보건의료원. Available from http://hineca.kr/43.
9. Higgins JPT, Green S, Collaboration C. Cochrane Handbook for Systematic Reviews of Interventions. version 5.1.0 [Updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org
10. Papadopoulos MA, Papageorgiou SN, Zogakis IP. Clinical Effectiveness of Orthodontic Miniscrew Implants: a Meta-analysis. J Dent Res 2011;90(8):969-976.