

# 국내 운동 관련 메타분석 논문의 질 평가

송영신<sup>1</sup> · 강문희<sup>2</sup> · 김선애<sup>3</sup> · 신인수<sup>4</sup>

<sup>1</sup>충남대학교 간호대학, <sup>2</sup>충남대학교 간호대학·간호과학연구소, <sup>3</sup>꽃동네대학교 간호학과, <sup>4</sup>진주대학교 사범대학 교육학과

## Review of Meta-analysis Research on Exercise in South Korea

Song, Youngshin<sup>1</sup> · Gang, Moonhee<sup>2</sup> · Kim, Sun-Ae<sup>3</sup> · Shin, In-Soo<sup>4</sup>

<sup>1</sup>College of Nursing, Chungnam National University, Daejeon

<sup>2</sup>College of Nursing · Research Institute of Nursing, Chungnam National University, Daejeon

<sup>3</sup>Department of Nursing, Kkottongnae University, Cheongju

<sup>4</sup>Department of Education, College of Education, Jeonju University, Jeonju, Korea

**Purpose:** The purpose of this study was to evaluate the quality of meta-analysis regarding exercise using Assessment of Multiple Systematic Reviews (AMSTAR) as well as to compare effect size according to outcomes. **Methods:** Electronic databases including the Korean Studies Information Service System (KISS), the National Assembly Library and the DBpia, HAKJISA and RISS4U for the dates 1990 to January 2014 were searched for 'meta-analysis' and 'exercise' in the fields of medical, nursing, physical therapy and physical exercise in Korea. AMSTAR was scored for quality assessment of the 33 articles included in the study. Data were analyzed using descriptive statistics, t-test, ANOVA and  $\chi^2$ -test. **Results:** The mean score for AMSTAR evaluations was 4.18 (SD= 1.78) and about 67% were classified at the low-quality level and 30% at the moderate-quality level. The scores of quality were statistically different by field of research, number of participants, number of databases, financial support and approval by IRB. The effect size that presented in individual studies were different by type of exercise in the applied intervention. **Conclusion:** This critical appraisal of meta-analysis published in various field that focused on exercise indicates that a guideline such as the PRISMA checklist should be strongly recommended for optimum reporting of meta-analysis across research fields.

**Key words:** Exercise, Meta-analysis

## 서론

### 1. 연구의 필요성

최근 의로지식과 기술이 빠르게 발전하고 대상자 중심의 근거기반증거가 강조되며 인터넷을 통한 실시간 정보교류가 가능해지면 서 메타분석 연구가 계속 증가하고 있다[1]. 2000년 이후 현재까지 10

년 동안 보건 및 의료 분야에서 발표된 연구 중 'meta-analysis'를 검색 용어로 하였을 때 국제적 데이터베이스인 PubMed에서 56,342편, 국내 데이터베이스인 RISS에서 학술지 게재 논문 2,435편, 학위논문 3,785편이 검색되었다. 이를 통해 국내외적으로 메타분석 연구가 활발히 진행되고 있음을 알 수 있다. 메타분석은 특정한 연구주제에 대한 두개 이상의 개별 연구들을 통계적으로 결합하여 평균 또는 공통 효과를 추정하는 양적 연구 방법이다[2]. 즉, 동일한 주제에서

주요어: 운동, 메타분석

Address reprint requests to : Gang, Moonhee

College of Nursing, Chungnam National University, 266 Moonhwa-ro, Jung-gu, Daejeon 301-747, Korea  
Tel: +82-42-580-8332 Fax: +82-42-580-8309 E-mail: mhgang@cnu.ac.kr

Received: March 24, 2014 Revised: April 16, 2014 Accepted: July 17, 2014

This is an Open Access article distributed under the terms of the Creative Commons Attribution NoDerivs License. (<http://creativecommons.org/licenses/by-nd/4.0>)  
If the original work is properly cited and retained without any modification or reproduction, it can be used and re-distributed in any format and medium.

도 연구 대상의 표집, 측정도구, 자료 분석 방법이나 연구 결과가 서로 다르므로 선행 연구들을 종합하여 또 다른 대안이나 결론을 도출하기 위한 연구통합법이라고 할 수 있다. 메타분석을 통한 연구통합의 결과는 매우 강력한 과학적 근거가 되고 있으며 다양한 분야에서 활용되고 있다. 특히, 메타분석은 치료 효과나 효과의 불확실성 정도에 대한 정량적 측정과 비교가 가능하므로 보건 및 의료분야에서 임상적 판단에 매우 유용하다. 메타분석은 간호사에게 임상적 의사결정과 간호수행에 필요한 과학적 근거를 제공해 주며, 이는 근거기반 간호에서 가장 높은 근거수준으로 분류되고 있다[2,3].

메타분석이 높은 근거수준으로 활용되기 위해서는 양적 증가뿐만 아니라 연구 방법이나 결과에 대한 신뢰성과 타당성이 우선 전제되어야 한다. 메타분석은 동질성과 타당성이 높거나 낮은 개별 연구들을 종합하므로 왜곡된 결론을 도출할 가능성이 있다. 메타분석에 대한 서로 다른 표본과 측정방법간의 비논리적 비교, 질 낮은 데이터의 활용문제, 연구물 선택의 편차문제 및 독립성 가정의 위반문제가 계속 비판되고 있다[4]. 따라서, 메타분석 결과를 적용하기 전에 메타분석 연구들에서 자료 수집 절차와 방법, 통계적 분석 방법 등이 적절하게 수행되었는지에 대한 연구 방법론적 검증은 필수적이다. 최근 메타분석에 대한 비판을 극복하고 신뢰할 수 있는 연구 결과를 확보하기 위한 전략으로 메타분석 연구의 질 평가가 제안되고 있다[5]. 외국의 경우 1990년대부터 메타분석 연구의 질 평가를 꾸준히 수행해 왔다[6-12]. 하지만 국내의 경우 거의 이루어지지 않았으며 선행 연구에 의하면 체계적인 연구의 질 평가 도구의 사용 없이 연구특성만을 서술하여 경영학 분야의 메타분석 연구의 현황을 보고한 연구[13], 연구자가 고안한 평가내용으로 국내 메타분석 연구의 방법론적 고찰을 수행한 연구[14], APA 평가도구를 사용하여 특수교육 분야의 메타분석 연구의 질을 평가한 연구[2], AMSTAR 평가도구를 사용하여 국내 간호학 분야 메타분석 연구의 질을 평가한 연구[3] 뿐이었다. 이와 같이 몇몇 학문분야별로 메타분석 연구의 질 평가가 시도는 되고 있으나 아직 초기 단계로 매우 부족한 실정이었다. 또한, 메타분석 연구의 질을 평가하는 도구나 평가하는 내용에도 차이가 있어 연구의 질 평가가 적절하게 수행되었는지를 판단하는데 어려움이 있었다. 따라서, 추후 신뢰도와 타당도가 검증된 평가도구를 일관되게 사용하여 다양한 주제 관련 메타분석 연구에 대한 체계적인 연구의 질 평가를 수행할 수 있는 연구가 요구되었다.

지금까지 과학적이고 체계적인 메타분석 연구의 질 평가를 위해 다양한 평가도구들이 개발되어 사용되어왔다. 의학, 약학, 치의학 및 간호학 분야에서는 1990년대 초부터 Postdam Consultation on Meta-analysis [15], QUOROM (Quality of Reporting of Meta-analyses) [5], MOOSE (Meta-analysis of Observation Studies in Epidemiology)

[16], PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analysis)[17], OQAQ (The Overview Quality Assessment Questionnaire)[18], AMSTAR (Assessment of Multiple Systematic Reviews) [19,20] 등을 주로 사용하여 메타분석 연구의 질 평가를 수행해 왔다[13-19]. 1990년 초에 개발된 Postdam Consultation on Meta-analysis [15]은 문항이 적고 일부 과학적 연구 방법론에만 초점을 두고 있으며 QUOROM (Quality of Reporting of Meta-analyses)[20]은 비교적 체계적으로 연구의 질을 평가하였으나 평가 항목수가 많고 실시하는데 시간이 오래 걸리는 단점이 있다. 이를 보완하여 관찰연구의 질을 평가하는 MOOSE (Meta-analysis of Observation Studies in Epidemiology)[16], 무작위 실험 연구의 질을 평가하는 PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analysis)[17]가 개발되었으나 대상 논문의 연구 설계에 따라 적용이 제한되며 27-35개 항목의 체크리스트 형식으로 구성되어 현재는 체계적 문헌고찰의 보고기준으로만 활용되고 있다.

이중 AMSTAR는 기존의 OQAQ (Overview Quality Assessment Questionnaire)와 Sack's checklist를 토대로 평가 내용과 방식을 보완하여 가장 최근까지 개발 및 수정되었으며 유일하게 신뢰도와 타당도가 검증된 도구이다[19,20]. AMSTAR는 문헌고찰의 계획 수립, 문헌 선택과 자료 추출 방법, 포괄적인 문헌검색, 출판상태의 포함 기준, 포함된 연구목록 제시, 연구특성 제시, 연구의 질 평가, 연구의 질을 결론 도출 시 활용, 개별 연구 결과의 결합방법의 적절성, 출판 편의 가능성 평가, 이해상충에 대한 기술을 포함하는 11개 문항으로 구성되었으며 기존 도구에 비해 메타분석 연구의 연구 방법론적 검증과 함께 연구의 질을 체계적으로 평가할 수 있는 도구이다 [19,20]. 또한, AMSTAR는 평가 문항이 간결하고 적용이 용이하며 0점 또는 1점으로 점수화가 가능하므로 메타분석 연구의 질 평가수준을 객관적으로 비교 및 분석할 수 있는 장점이 있다. AMSTAR는 최근 보건 및 의료 분야를 포함하여 학문분야에 제한 없이 전 세계적으로 널리 사용되고 있으며[7,9-11,19,20], 우리나라의 경우 2011년부터 한국보건연구원에서 AMSTAR를 승인을 받아 메타분석 연구의 질 평가 도구로 사용할 것을 권고하고 있다[21]. 또한, 다양한 AMSTAR를 사용한 메타분석 연구가 축적되면서 서로 다른 분야나 주제에 관한 메타분석 연구의 질 정도를 객관적으로 비교할 수 있으며, 이는 메타분석 연구의 발전과 연구 결과의 활용에 새로운 표준지침이 될 것으로 예상된다.

한편, 운동은 간호중재로 가장 많이 활용되고 있으며 간호학뿐만 아니라 의학, 체육학과 무용학, 교육학 등 다양한 분야에서 메타분석 연구의 주요 주제로 다루어지고 있다. '운동과 '메타분석'을 검색 용어로 하였을 때 국외의 경우 PubMed에서 2,000편 이상의 연구가 검색되었으며, 국내의 경우 한국교육학술정보원에서 제공하는

RISS에서는 150여 편의 학술지 게재 연구가 검색되어 학문분야별로 다양하게 연구되고 있음을 알 수 있다. 하지만 아직 국내에서 운동 관련 메타분석 질 평가 연구가 없어 국내에서 출판된 운동 관련 메타분석 연구의 자료 수집 절차, 자료 수집 방법, 관련된 변수의 수, 통계적 분석 방법 등에 대한 적절한 논의가 거의 이루어지지 못하고 있는 실정이다. 또한, 운동 중재의 임상적인 활용성 및 효과성을 판단하고 중재의 근거기반을 제공하기 위한 운동 관련 메타분석 연구의 결과에 관한 다양한 분석이 이루어지지 않았다.

따라서, 본 연구는 '운동과 메타분석'을 주요어로 하여 다양한 학문분야에서 현재시점까지 국내 학술지에 게재된 메타분석 연구를 대상으로 연구의 특성을 파악하고 신뢰도와 타당도가 검증된 평가 도구인 AMSTAR를 사용하여 연구의 질을 평가하고자 한다. 또한, 개별 메타분석 연구에서 제시한 효과크기를 토대로 가장 높은 수준의 효과크기를 보이는 운동중재가 무엇인지, 운동의 효과는 무엇인지, 운동에 참여한 대상자의 특성과 측정방법에 대해 비교함으로써 간호사에게 대상자 중심의 효과적인 운동간호중재를 위한 임상적 의사결정과 간호수행의 근거기반을 제공하고자 한다. 이는 추후 메타분석 연구의 질 향상 방안을 제시하며 더 나아가 근거기반에 기초한 운동 간호실무 향상에도 기여할 것이다.

2. 연구 목적

본 연구의 목적은 AMSTAR를 이용하여 국내 학술지에 게재된 운동 관련 메타분석 논문의 질을 평가하는 것이며 구체적인 목적은 다음과 같다.

첫째, 국내 학술지에 게재된 운동 관련 메타분석 연구의 특성을 파악한다.

둘째, 국내 학술지에 게재된 운동 관련 메타분석 연구의 질 평가를 실시한다.

셋째, 국내 학술지에 게재된 운동 관련 메타분석 연구의 연구특성에 따른 질 평가 점수를 파악한다.

넷째, 국내 학술지에 게재된 운동 관련 메타분석 연구의 연구특성에 따른 효과크기를 파악한다.

연구 방법

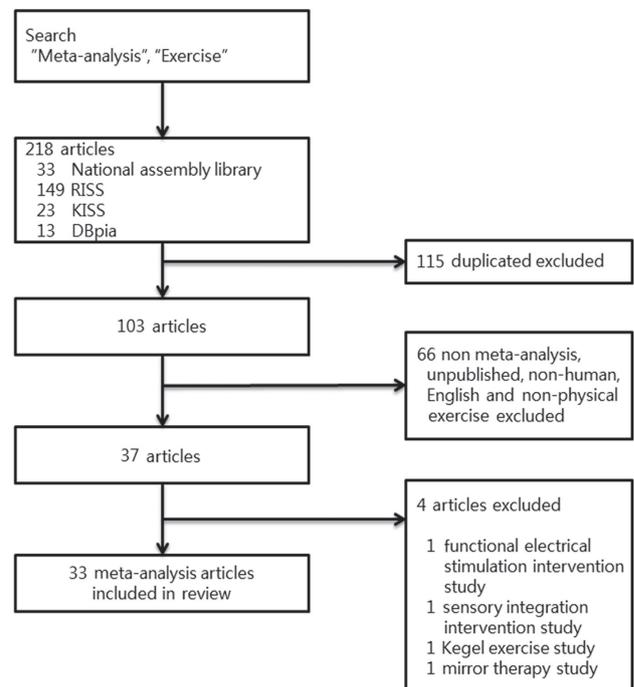
1. 연구 설계

본 연구는 국내 학술지에 게재된 운동 관련 메타분석 연구를 AMSTAR 보고기준에 따라 연구의 질을 평가하는 서술적 조사 연구이다.

2. 분석 대상 논문의 선정 및 자료 수집

국내 의학, 간호학, 체육학 등 다양한 학문분야의 학술지에 게재된 메타분석 연구를 검색하기 위해 한국학술정보원(Koreanstudies Information Service System [KISS]), DBpia, RISS4U, 학지사의 뉴논문, 국회도서관 데이터베이스를 이용하여 한국어로 '메타분석', '운동'을 주요어로 하여 2014년 1월까지 국내 학술지에 게재된 논문을 검색하였다. 본 연구를 위해 연구자는 체계적 문헌고찰, 메타분석 및 질 평가 관련 워크숍에 참여하여 메타분석 연구의 질 평가 수준을 높이기 위하여 노력하였다. 각자 독립적으로 자료검색과 자료추출에 참여하였으며 공동 연구자이며 메타분석 전공 교수 1명이 전반적인 자료 수집 및 선정 과정을 함께 검토하였다. 자료검색은 같은 주요어를 입력한 후 사전에 합의한 선정기준에 따라 자료를 추출하였으며 인터넷과 도서관 상호대차 등을 통해 자료를 검색 및 수집하였다. 분석 대상 논문의 선정기준은 첫째, 메타분석 연구, 둘째, 운동관련 효과크기를 분석한 메타분석 연구, 셋째, 한국어로 발표한 메타분석 연구이다.

검색전략에 의해 검색된 총 논문 수는 218편이며 이중 중복 115편 제외, 제목과 초록으로 부적합하여 제외된 일차연구 66편, 전문 검토 후 배제된 논문이 4편으로 최종 33편의 논문이 선택되었다 (Figure 1).



RISS=Research information sharing service; KISS=Korean studies information service system; DBpia=Databasepia.

Figure 1. Literature search strategy.

### 3. 연구 도구

#### 1) 메타분석 논문의 질 평가 도구

본 연구에서는 메타분석 연구의 질 평가를 위해 개발된 AMSTAR [19]를 국내에서 한국어로 번역하여 신뢰도와 타당도가 검증된 AMSTAR 한국어 버전[21]을 사용하였다. 이 도구는 사전에 문헌 고찰의 계획 수립 여부, 문헌선택과 자료 추출 방법, 포괄적인 문헌 검색, 출판상태의 포함 기준, 포함된 연구목록 제시, 연구특성 제시, 연구의 질 평가, 연구의 질을 결론 도출 시 활용, 개별 연구 결과의 결합 방법의 적절성 여부, 출판편의의 가능성 평가 여부, 이해상충에 대한 기술 여부를 평가하며 총 11문항의 체크리스트 형식으로 구성되어 있다. 각 항목은 '예', '아니오', '대답할 수 없음', '적용할 수 없음'으로 평가하며 '예'는 1점, '아니오', '대답할 수 없음', '적용할 수 없음'은 0점을 부여하였다. 총점 범위는 0-11점이며 선행 연구[3,17,22]에 근거하여 0-4점은 연구의 질이 낮은 수준, 5-8점은 중간 수준, 9-11점은 높은 수준으로 분류하였다.

본 연구자 3명은 연구자 간 측정의 일치도를 높이기 위해 연구수행 전에 5편의 논문을 대상으로 AMSTAR를 이용하여 평가하고 문항에 대한 이해와 판단기준을 합의하는 사전 훈련을 시행하였다. AMSTAR는 신뢰도와 타당도가 검증된 평가 도구로서 선행 연구 [19,20]에서 측정자 간 일치도는 Cohen's kappa = .84-.95, Intra class coefficients는 .84-.95였으며, 본 연구에서 측정자간 일치도는 Cohen's kappa = .74, Intra class coefficients = .96이었다.

#### 2) 대상 논문의 연구특성

본 연구에서 연구의 특성은 논문이 게재된 학문분야, 연구 대상, 메타분석에 활용한 검색 데이터베이스, 검색의 범위와 출판년도, 저자의 수, 메타분석에 이용된 논문의 수와 총 대상자 수, 연구비 지원여부, 분석에 이용된 통계프로그램, 효과크기로 사용된 지수의 유형, 분석모형 등을 조사하였다.

### 4. 대상자의 윤리적 고려

본 연구는 C대학교 간호대학 생명윤리심의위원회의 심의(IRB No. 2014-010)을 거쳐 수행되었다.

### 5. 자료 분석

#### 1) 코딩

최종 질 평가 대상 33편의 논문에 대한 코딩은 사전에 핵심질문과 부합되도록 자료추출양식을 작성하여 사용하였으며 연구자 3명

이 동시에 독립적으로 자료를 추출한 후 그 일치여부를 확인하였다. 의견 불일치 시에는 공동 연구자이며 메타분석의 전공교수의 자문을 구한 후 토론을 통해 투표로 합의된 의견을 코딩을 하였으며 이러한 과정을 거쳐 모든 문항에 대한 협의 및 검토가 이루어져 최종적으로 불일치한 문항은 없었다. 연구특성은 개별 논문의 내용을 토대로 하였으며 논문이 게재된 학문분야는 학술지명과 학회 홈페이지를 통해 확인하였다. 대상 논문의 질 평가는 포함 및 배제된 연구목록 제시 항목에 대해서 최종 연구목록만 보고한 경우, 포함된 연구의 특성 제시 항목에 대해서 목록만 보고한 경우, 개별 연구 결과의 결합방법의 적절성 항목에 대해서 동질성 검정결과가 없는 경우는 모두 '0점'으로 합의하여 처리하였다.

#### 2) 통계 분석

수집된 자료는 SPSS 21.0 프로그램을 이용하여 분석하였다. 대상 논문의 연구특성과 AMSTAR를 이용한 대상 논문의 질 평가 정도는 빈도, 백분율, 평균과 표준편차로 분석하였다. 대상 논문의 연구특성에 따른 질 평가 정도는 t-test, ANOVA를 이용하여 분석하였다. 대상 논문의 연구특성에 따른 효과크기 정도와 차이는 빈도, 백분율과  $\chi^2$ -test, Fisher's exact test를 이용하여 분석하였다.

## 연구 결과

### 1. 연구 특성

최종 대상 논문은 33편이었으며 연구특성 결과는 Table 1과 같다. 논문의 출판년도는 1995에서 2013년까지였고, 저자의 수는 최저 1명에서 최고 5명이었다. 논문이 게재된 학문분야별로는 체육학 16편, 간호학 7편, 의학 2편, 물리치료 3편, 기타 교육학 등이 5편이었다.

### 2. 대상 논문의 연구의 질 평가

대상 논문의 연구의 질 평가 결과는 Table 2와 같다. 질 평가 점수는 11점 만점에 평균 4.18( $\pm 1.78$ )점이었으며 0-4점의 낮은 수준은 66.7%, 5-8점의 중간 수준은 30.3%, 9-11점의 높은 수준은 3.0%로 나타났다. 11개의 하위영역별 평균점수를 점수범위 최소 0점-최대 1점 기준으로 제시하면, '사전에 문헌고찰의 계획 수립여부' 항목은 평균점수 1점으로 33편 모두 기준을 충족하였으며, '연구특성 제시' 0.70점, '개별 연구 결과의 결합 방법의 적절성 여부' 0.64점, '출판상태의 포함기준' 0.55점, '포괄적인 문헌 검색' 0.42점, '문헌선택과 자료 추출 방법', '출판편의의 가능성 평가 여부', '이해상충의 기술' 0.21점, '연구의 질 평가', '연구의 질을 결론 도출 시 활용' 0.12점의 순으로 낮

**Table 1.** Study Characteristics of Studies Included

(N=33)

ID	Year	Number of authors	Field of research	Participants	Number of studies included	Number of databases	Topic	Study design	Main outcomes of interest (effect size)
1	2013	1	PE	Elderly people, women	29	1	Muscle strength	RCT	Dance (ES=1.46), aerobic exercise (ES=1.07): large effect size
2	2013	1	PE	Elementary school students	20	2	Physical self-concept	RCT	Self-concept (ES=0.50)
3	2013	2	N	Elderly people	16	4	Endurance, flexibility	RCT/Quasi	Lower limb muscle endurance (ES=1.06), whole body endurance (ES=0.41), upper body flexibility (ES=0.29)
4	2013	2	N	People with falls	52	13	Fall	RCT	Physical factors: 3 months: balance (ES=0.54), flexibility (ES=0.40), muscle strength (ES=0.48), daily activity living change (ES=0.71) / 6 months: balance (ES=0.59), flexibility (ES=0.35), daily activity living change (1.30) / psychological factor: phobia of falls (3 month-ES=0.37; 6 month-ES=0.58)
5	2013	3	N	Obese children	61	3	Obesity	RCT/Quasi	Total effect size (ES=-0.23): exercise therapy (ES=-0.31), exercise & nutrition education (ES=-0.50)
6	2012	2	Others	Children	50	3	Effect of exercise	RCT	Physiologic function (ES=0.44), physical function (ES=0.47)
7	2012	1	PE	Obese children	42	4	Body fat	RCT	Aerobic exercise (ES=0.86), 5 times/week (ES=0.86), 30-60min./1 session (ES=0.71), duration of 9-12 weeks (ES=0.74)
8	2012	3	PE	Elementary school students	22	2	Effect of exercise	RCT	Effect of taekwondo (ES=0.73)
9	2012	2	PE	Elementary school students	15	1	Body fat	RCT	MES (Mean Effect Size): aerobic exercise, more than 40 min.
10	2012	1	PE	Children with a developmental disability	31	-	Physical activity	RCT/Quasi	Total effect size of physical activity (ES=1.11): cognitive (ES=1.51), psychomotor (ES=1.59), affective (ES=0.94)
11	2011	2	PT	Elderly people	17	2	Muscle strength	RCT/Quasi	Muscle strength (ES=0.82)
12	2010	3	Others	Adults, elderly people	82	1	Physical function	RCT	Adults (ES=0.43), elderly people (ES=0.52)
13	2010	2	N	Adults, patients with hypertension	15	4	Blood pressure	RCT	Systolic pressure (z=2.89, p=.004), diastolic pressure (z=2.78, p=.001)
14	2010	1	PT	Men	46	1	Growth hormone	Quasi	Total effect size: athletes (ES=2.62), nonathletes (ES=1.35)
15	2009	1	PT	Patients with ankylosing spondylitis	10	4	Physical function	RCT/Quasi	Exercise effect (ES=0.55)
16	2009	3	N	Middle-aged women	14	3	Bone mineral density	RCT	BMD (ES=0.32)
17	2009	1	PE	All ages	139	1	Physical function	RCT	Effect size in physiologic: complex exercise, adolescents (ES=0.93)
18	2009	1	PE	Elderly people	46	1	Physiologic, physical, psychological function	RCT	Physiologic function (ES=0.56), physical function (ES=0.60), psychological function (ES=1.14)
19	2009	1	PE	All ages	126	1	Physical function	RCT	Infants & children (ES=0.43), adolescents (ES=0.52), adults (ES=0.43), elderly people (ES=0.52)
20	2008	1	PE	Adults, elderly people	42	1	Psychological function	RCT	Adults (ES=0.67), elderly people (ES=0.47)

N=Nursing; M=Medical; PE=Physical exercise; PT=Physical therapy; Others=Education etc.; RCT=Randomized clinical trial; Quasi=Quasi experimental; BMD=Bone marrow density; HbA1C=Glycosylated hemoglobin; HDL=High density lipoprotein; LDL=Low density lipoprotein; NREM=Non rapid eye movement; RHR=Resting heart rate; VE=Ventilation; VO<sub>2</sub> Max=Maximal oxygen uptake.

Table 1. Study Characteristics of Studies Included (Continued)

(N=33)

ID	Year	Number of authors	Field of research	Participants	Number of studies included	Number of databases	Topic	Study design	Main outcomes of interest (effect size)
21	2007	1	PE	Elderly people	16	1	Body fat rate	RCT/Quasi	Body fat rate (ES=0.32)
22	2007	5	M	Obese population	30	1	Obesity	RCT	Body fat (ES=-0.88), TG (ES=0.80)
23	2007	1	M	Elderly people	63	7	Exercise participation & completion rate	Correlation	No significant correlation between exercise participation, exercise completion, & exercise intensity adherence (r=.002)
24	2006	2	PE	Children & adolescents (<19year)	68	3	Fitness	RCT/Quasi	Cardiovascular endurance (ES=0.84), balance (ES=0.88)
25	2006	1	PE	Elderly people	37	1	Exercise participation rate	RCT	Exercise participation: duration of exercise is shorter, group exercise (R <sup>2</sup> =0.71, p<.001)
26	2005	2	N	Adults, patients with type 2 diabetes	11	1	Metabolism of glucose and lipid, cardiopulmonary function	RCT	HbA1c (ES=0.55), fasting blood glucose (ES=0.38), VO <sub>2</sub> max (ES=0.50), cholesterol (ES=0.27), HDL (ES=0.11), LDL (ES=0.12)
27	2003	1	Others	Men (<29)	28	1	Male hormone	RCT/Quasi	Testosterone: athletes (ES=0.15), nonathletes (ES=0.27)
28	2001	1	Others	Adults (<29)	29	-	Growth hormone	Quasi	Total effect size: maximal exercise intensity-athletes (ES=1.85), nonathletes (ES=1.42)
29	1999	1	Others	All ages	38	1	Sleep	RCT	NREM 4 phase (ES=0.75), 3 phase (ES=0.34), total sleep hours (ES=0.31)
30	1999	2	PE	Men, athlete/nonathletes	17	1	Testosterone	Quasi	Athletes (ES=-0.90), nonathletes (ES=-1.17)
31	1998	2	N	Patients with arthritis	28	2	Pain, depression, activity disability	RCT/Quasi	Integrated program: pain (ES=0.25), depression (ES=0.11), activity (ES=0.16) Exercise program: pain (ES=0.30), activity (ES=0.60)
32	1998	3	PE	All ages	75	2	Physiologic factors	Quasi	Body fat (ES=-0.75), RHR (-0.85), VO <sub>2</sub> max (ES=0.67), VE (ES=1.07)
33	1995	1	PE	All ages	11	3	Cognitive strategy (internal & external attention)	RCT/Quasi	Total cognitive strategy (ES=0.34), external attention strategy (ES=0.49), internal attention strategy (ES=-0.08)

N=Nursing; M=Medical; PE=Physical exercise; PT=Physical therapy; Others=Education etc.; RCT=Randomized clinical trial; Quasi=Quasi experimental; BMD=Bone marrow density; HbA1C=Glycosylated hemoglobin; HDL=High density lipoprotein; LDL=Low density lipoprotein; NREM=Non rapid eye movement; RHR=Resting heart rate; VE=Ventilation; VO<sub>2</sub> Max=Maximal oxygen uptake.

Table 2. Scores of Assessment of Multiple Systematic Reviews

No.	Items	Yes	M ± SD
		n (%)	
1	Was an 'a priori' design provided?	33 (100.0)	1.00 ± 0.00
2	Was there duplicate study selection and data extraction?	7 (20.6)	0.21 ± 0.41
3	Was a comprehensive literature search performed?	14 (41.2)	0.42 ± 0.50
4	Was the status of publication (i.e. grey literature) used as an inclusion criterion?	18 (54.5)	0.55 ± 0.51
5	Was a list of studies (included and excluded) provided?	0 (0.0)	0.00 ± 0.00
6	Were the characteristics of the included studies provided?	23 (69.7)	0.70 ± 0.47
7	Was the scientific quality of the included studies assessed and documented?	4 (12.1)	0.12 ± 0.33
8	Was the scientific quality of the included studies used appropriately in formulating conclusions?	4 (12.1)	0.12 ± 0.33
9	Were the methods used to combine the findings of studies appropriate?	21 (63.6)	0.64 ± 0.49
10	Was the likelihood of publication bias assessed?	7 (21.2)	0.21 ± 0.42
11	Was any conflict of interest stated?	7 (21.2)	0.21 ± 0.42
Total			4.18 ± 1.78
Level of quality			
Low (0-4)		22 (66.7)	
Moderate (5-8)		10 (30.3)	
High (9-11)		1 (3.0)	

았으며, '포함된 연구목록 제시' 항목의 경우 점수가 평균 점수가 0 점으로 33편 모두 기준을 충족하지 못한 것으로 나타났다.

### 3. 대상 논문의 연구특성에 따른 질 평가

대상 논문 33편의 연구특성에 따른 질 평가 결과는 Table 3과 같

다. 논문이 게재된 학문분야별로 질 평가 점수는 유의한 차이가 있었으며 간호학분야에서 질 평가 점수가 가장 높았다( $F=5.16, p=.003$ ). 총 대상자수에 따라 차이가 있었으며 대상자 수가 1,000명 이상에서 유의하게 높게 나타났고( $F=8.75, p=.002$ ), 메타분석에 활용한 검색 데이터베이스의 수에 따라 차이가 있었으며 4개 이상에서 유의하게 높게 나타났다( $F=3.33, p=.033$ ). 논문이 게재된 학술지

**Table 3.** Comparison of AMSTAR Scores by Characteristics of the Meta-analyses Included

Characteristics	Categories	n (%)	M ± SD	t or F (p)	Scheffé test
Year of publication	1995-1999	5 (15.2)	3.00 ± 1.00	2.13 (.136)	
	2000-2009	14 (42.4)	4.00 ± 1.62		
	2010-2013	14 (42.4)	4.79 ± 1.97		
Field of research	Nursing <sup>a</sup>	7 (21.2)	5.86 ± 2.19	5.16 (.003)	b < a
	Medical	2 (6.1)	3.00 ± 1.41		
	Physical exercise <sup>b</sup>	16 (48.5)	3.50 ± 1.21		
	Physical therapy	3 (9.1)	6.00 ± 1.00		
	Others	5 (15.1)	3.40 ± 0.89		
Number of authors	1	17 (51.5)	3.71 ± 1.49	1.31 (.283)	
	2	10 (30.3)	4.60 ± 2.07		
	≥ 3	6 (18.2)	4.83 ± 1.94		
Number of studies included	≤ 20	11 (33.3)	1.74 ± 0.74	0.87 (.429)	
	21-49	13 (39.4)	3.77 ± 1.30		
	≥ 50	9 (27.3)	4.11 ± 2.37		
Number of total participants	≤ 499 <sup>a</sup>	6 (18.2)	3.00 ± 0.89	8.75 (.002)	a = b < c
	500-999 <sup>b</sup>	11 (33.3)	4.09 ± 1.45		
	≥ 1,000 <sup>c</sup>	6 (18.2)	6.33 ± 1.75		
	Unknown	10 (30.3)			
Number of databases	0	2 (6.1)	4.50 ± 0.71	3.33 (.033)	a < b
	1 <sup>a</sup>	15 (45.4)	3.27 ± 1.10		
	2-3	10 (30.3)	4.70 ± 1.64		
	≥ 4 <sup>b</sup>	6 (18.2)	5.50 ± 2.59		
Participants	Children/adolescents	7 (21.2)	4.43 ± 1.40	0.81 (.527)	
	Adults	9 (27.3)	3.67 ± 1.58		
	Elderly people	7 (21.2)	4.14 ± 2.12		
	Adults + elderly people	6 (18.2)	5.17 ± 2.32		
	All	4 (12.1)	3.50 ± 1.29		
Having disease of subjects	Yes	9 (27.3)	4.56 ± 1.51	0.73 (.468)	
	No	24 (72.7)	4.04 ± 1.88		
Language	Korean only	17 (51.5)	4.00 ± 1.27	0.89 (.455)	
	English only	9 (27.3)	4.22 ± 2.11		
	Korean + English	5 (15.1)	5.20 ± 2.78		
	Unknown	2 (6.1)	3.00 ± 0.00		
Journal type	SCI (E) or SSCI <sup>a</sup>	4 (12.1)	5.75 ± 2.75	3.35 (.032)	b < a
	KCI accredited	21 (63.6)	4.33 ± 1.53		
	KCI candidate	4 (12.1)	3.75 ± 1.25		
	Others <sup>b</sup>	4 (12.1)	2.25 ± 0.50		
Study design	RCT	18 (54.5)	4.22 ± 1.77	2.12 (.119)	
	Quasi	4 (12.1)	2.75 ± 0.96		
	RCT + quasi	10 (30.3)	4.90 ± 1.73		
	Correlational study	1 (3.0)	2.00 ± 0.00		
Financial support	Yes	9 (27.3)	5.33 ± 2.06	2.45 (.020)	
	No	24 (72.7)	3.75 ± 1.48		
Institutional review board	Yes	1 (3.0)	9.00 ± 0.00	3.10 (.004)	
	No	32 (97.0)	4.03 ± 1.56		
Participants flow diagram	Yes	5 (15.1)	5.40 ± 2.88	1.08 (.096)	
	No	28 (84.9)	3.96 ± 1.48		
Type of effect size	d	29 (87.9)	4.28 ± 1.77	0.37 (.692)	
	r	3 (9.1)	3.33 ± 2.30		
	odds ratio	1 (3.0)	4.00 ± 0.00		

SCI (E)=Science citation index expanded; SSCI=Social science citation index; KCI=Korean citation index; RCT=Randomized clinical trial; Quasi=Quasi experimental.

의 유형에 따라 차이가 있었으며 SCI (E)나 SSCI에서 유의하게 높게 나타났다( $F=3.35, p=.032$ ). 연구비 지원 여부에 따라 유의한 차이가 있었으며 연구비를 지원 받은 경우가 받지 않은 경우 보다 높게 나타났다( $t=2.45, p=.020$ ), IRB 승인 여부에 따라 유의한 차이가 있었으며 IRB 승인을 받은 경우가 받지 않은 경우에 비해 높게 나타났다( $t=3.10, p=.004$ ).

#### 4. 연구특성에 따른 효과크기

최종 선택된 총 33편의 논문 중 운동의 형태나 그 효과에 대한 효과크기가 없거나 상관관계수, odds ratio로 크기를 나타낸 6편 논문을 제외하고 27편의 논문에서 제시된 주요 변수에 따른 효과크기를 분석한 결과 Table 4와 같다. 각 논문에서 제시된 효과크기는 2-17개로 총 170개의 효과크기가 추출되었으며 이에 대한 주요 변수는 개별 연구자가 직접 제시한 기준과 특성을 반영하여 본 연구자가 운동, 대상자, 측정방법으로 범주화하였다. 운동은 운동의 형태와 효과로 구분하여 신체조성 및 체력은 신체적 효과로, 면역이나 호르몬 등은 생리적 효과로, 자기효능감, 우울 등은 심리적 효과로 세분화하였다. 대상자는 성별, 연령, 질환의 유무에 따라 구분하였고 측정방법은 객관적 측정과 자가 보고식으로 구분하여 효과크기를 비교하고 차이를 분석하였다. 추출된 효과크기는 0.8 이상은 'large',

0.2-0.8 미만은 'moderate', 0.2 미만은 'small'로 분류하였다[23].

연구 결과 운동중재에 따른 효과크기에서 유산소운동의 경우 67개로 가장 많았고 저항운동은 3개로 가장 적었으며 효과크기의 정도에 따른 세 집단 간에 통계적으로 유의한 차이가 있었다( $\chi^2=22.43, p=.002$ ). 한편, 효과크기는 운동의 효과, 대상자의 성별, 연령, 질환 유무, 측정방법에 따라 다양한 분포를 보였으나 통계적으로 유의한 차이는 없었다.

### 논 의

본 연구는 국내 학술지에 게재된 운동 관련 메타분석 논문의 연구특성과 효과크기의 차이를 파악하고 연구의 질 평가를 수행하여 메타분석 연구의 질 향상과 운동 관련 근거기반 간호실무 향상에 기여하고자 수행되었다. 주요 연구 결과를 토대로 논의하고자 한다.

본 연구에서 AMSTAR를 사용하여 총 33편의 메타분석 연구의 질을 평가한 결과 11점 만점에 평균 4.18점으로 낮게 나타났다. 대상 논문 전체의 질 평가 수준은 낮은 수준 66.7%, 중간수준 30.3%, 높은 수준 3.0%로 나타났다. 이러한 결과는 동일한 도구를 사용하여 여성의 유방촬영술에 관한 메타분석 연구의 질 평가 결과(3.24점)[24] 보다 높았으며 간호학 분야 메타분석 연구의 질 평가 결과(5.61점)[3], 아동폭력예방프로그램에 관한 메타분석 연구의 질 평가 결

Table 4. Classification of Effect Sizes by Study Characteristics

(N= 170)

Characteristics	Categories	Variables	Number of effect size	Large effect size ( $\geq 0.8$ )	Moderate effect size (0.2-0.7)	Small effect size ( $< 0.2$ )	$\chi^2 (p)$
			n (%)	n (%)	n (%)	n (%)	
Exercise	Type	Aerobic	67 (39.4)	0 (0.0)	48 (41.0)	19 (48.7)	22.43 (.002*)
		Anaerobic	16 (9.4)	1 (7.1)	9 (7.7)	6 (15.4)	
		Resistance	3 (1.8)	1 (7.1)	2 (1.7)	0 (0.0)	
		Combined	37 (21.8)	7 (50.0)	27 (23.1)	3 (7.7)	
		Non specific	47 (27.6)	5 (35.8)	31 (26.5)	11 (28.2)	
	Effect	Physical effect	94 (55.3)	3 (21.4)	69 (59.0)	22 (56.4)	7.35 (.118)
Physiologic effect		51 (30.0)	7 (50.0)	32 (27.3)	12 (30.8)		
Psychologic effect		25 (14.7)	4 (28.6)	16 (13.7)	5 (12.8)		
Participants	Gender	Male only	8 (4.7)	2 (14.3)	3 (2.6)	3 (7.7)	6.04 (.135*)
		Female only	5 (2.9)	0 (0.0)	3 (2.6)	2 (5.1)	
		Both	157 (92.4)	12 (85.7)	111 (94.8)	34 (87.2)	
	Age	Children/adolescents	57 (33.5)	2 (14.3)	45 (38.5)	10 (25.6)	13.61 (.067*)
		Adults	30 (17.6)	5 (35.7)	17 (14.5)	8 (20.5)	
		Elderly people	38 (22.4)	1 (7.1)	24 (20.5)	13 (33.3)	
		Adults + elderly people	9 (5.3)	0 (0.0)	8 (6.8)	1 (2.6)	
		All	36 (21.2)	6 (42.9)	23 (19.7)	7 (18.0)	
	Patient	Yes	47 (27.6)	6 (42.9)	33 (28.2)	8 (20.5)	2.63 (.269)
		No	123 (72.4)	8 (57.1)	84 (71.8)	31 (79.5)	
Measurement	Observation		118 (69.4)	31 (79.5)	77 (65.8)	10 (71.4)	2.61 (.272)
	Self-reporting		52 (30.6)	8 (20.5)	40 (34.2)	4 (28.6)	

\*Fisher's exact test.

과(7.02점)[22] 보다 낮았다. 근거기반 운동중재의 필요성에 대한 인식의 확대로 그 수요가 증가하고 있음을 고려할 때 추후 메타분석 연구의 질을 향상시키기 위한 효율적인 전략이 요구된다.

AMSTAR의 세부 항목별 점수를 살펴보면 '선택과 배제된 연구들의 목록 제시에서 대상 논문 전체가 기준을 충족하지 못 했다. 모든 논문들이 최종 분석에 사용된 논문의 목록만을 제시할 뿐 배제 논문의 목록은 제시하지 않았다. 이는 치의학 분야[11]와 간호학 분야 [3]에서 메타분석 연구의 질을 평가한 결과 66.5-88.1%에서 연구목록 제시 기준을 충족하지 못했음을 보고한 결과와 유사하였다. 이러한 결과는 학회나 기관의 논문의 페이지 수 제한 규정과 연구의 질 평가 도구로 사용된 PRISMA [17]와 AMSTAR [20]의 평가기준의 차이에서 기인한 결과로 해석된다. PRISMA의 경우 선정된 논문에 대한 자세한 기술을 요구하는 반면에 배제 논문은 배제 이유와 수만을 제시하도록 권장하고 있다[17]. 따라서, 메타분석 연구의 방법이나 분석 과정에 대해 좀 더 투명하게 평가하기 위해서는 메타분석 연구의 질 평가 도구와 보고 기준 간에 평가내용과 기준의 일치된 적용이 요구된다.

메타분석 수행 전 '대상 논문의 질 평가'는 연구 결과의 신뢰성과 밀접한 관련이 있다. 메타분석은 대상 논문의 연구 설계, 문헌의 질, 근거의 양과 일관성 및 직접성 등을 고려하여 논문의 질을 평가하고 이를 반영하여 종합적으로 결론을 제시할 때 신뢰성과 타당성이 높아지게 된다[21]. 하지만 본 연구 결과 대상 논문의 12.1%에서만 연구의 질 평가와 이를 반영한 결론의 기준을 충족한 것으로 나타났다. 이를 제외한 대부분의 메타분석에서 대상 논문의 질 평가 없이 연구 결과를 합성하고 의미를 해석하는 것으로 나타나 연구 결과에 대한 올바른 해석을 이끄는 데 다소 제한점이 있었다. 최근 질 평가 도구를 사용하여 논문을 평가하는 것이 일반화되고 있으므로 이를 확대하여 공식적으로 학회지의 심사 기준에 질 평가 기준을 도입한다면 메타분석 연구의 질 향상에 기여할 것으로 생각된다.

본 연구 결과 '두 명 이상의 독립적 자료추출' 평가 항목에서 대상 논문의 20.6%만이 기준을 충족한 것으로 나타났다. 이를 제외한 대부분의 논문에서는 자료추출 과정에 대한 언급이 없거나, 혹은 자료 추출 대신 추출된 자료의 코딩과정에 2명 이상이 참여하여 정확성과 신뢰도를 높이고자 하였다고만 서술하고 있었다. 자료코딩의 신뢰성 확보를 위해 충분하고 일관된 문헌 추출이 선행되어야 함에도 불구하고[20] 아직까지 이에 대한 연구자들의 인식수준이 낮음을 알 수 있다. 메타분석에서 검색전략을 어떻게 수립하고 수행하는가에 따라 민감도 및 특이도 뿐 아니라 출판편의의 문제가 개입될 수 있다. 따라서, 2명 이상의 연구자가 체계적인 프로토콜을 작성하여 일관되고 신뢰할 수 있는 자료를 추출할 수 있도록 노력해야 한

다[21]. 또한, 이러한 체계적이고 과학적인 자료추출의 중요성에 대한 연구자의 인식과 검색능력을 향상시킬 수 있는 교육과 지원이 요구된다.

AMSTAR 평가 문항 중 '이해의 상충의 기술'에 대해 대상 논문의 21.2%를 제외하고 대부분의 연구에서 이에 대한 언급이 없어 판단하기가 어려웠다. 국내 학술지의 경우 출판규정에 따라 차이는 있으나 이해상충에 대한 기술을 별도의 문서로 제출하도록 하고 있다. 또한, 외국과 달리 대부분 논문 양식에서 이해상충의 기술은 의무 사항이 아니어서 누락율이 높게 나타난 결과로 예상된다. 그러나 추후 메타분석 연구에 포함된 개별 연구들의 연구비 출처나 자원에 대한 명시는 물론 이해상충 여부까지 제시할 수 있도록 학회나 출판기관의 협조와 논의가 필요하다고 본다.

본 연구 결과 '출판 여부를 포함 기준 제시' 평가 항목에서 대상 논문의 54.5%가 기준을 충족한 것으로 나타났다. 미출판 석사와 박사 학위 논문만을 대상으로 분석하거나 반대로 회색논문을 전혀 포함하지 않은 논문도 있었다. 이는 출판편의에 대한 연구자의 이해부족으로 나타난 결과로 볼 수 있다. 출판편의는 연구의 질과도 밀접한 관련이 있으므로 신뢰성 높은 연구 결과를 확보하기 위해서는 연구자의 출판편의에 대한 이해를 높이고 석사와 박사학위 논문만을 대상으로 메타분석을 수행하는 것은 지양되어야 할 것이다.

본 연구에서 '연구의 결과를 종합하기 위한 적절한 방법 사용' 평가 항목에서 효과크기 계산 방법 및 공식은 제시하였으나 그에 따른 분산계산 공식을 누락한 연구가 많은 것으로 나타났다. 메타분석은 분산 계산 공식을 토대로 역변량 가중치를 구해서 연구 결과를 종합하게 된다. 분석에 포함된 개별 연구 결과들이 한 모집단에서 추출된 것인지, 다른 모집단에서 추출된 이질적인지는 동질성 평가에 따라 결정되며 이는 고정효과 모형 또는 랜덤효과 모형을 선택하는 근거가 된다[2]. 그러나 본 연구 결과 동질성 평가가 누락되거나 모델 선택의 통계적 근거나 기준을 제시하지 않은 연구가 많았다. 이는 메타분석 결과의 타당성을 저하시킬 수 있는 주요 요인이 되므로 반드시 개선이 필요하다.

본 연구에서 연구특성에 따른 메타분석 연구의 질 평가 점수의 차이를 분석한 결과 학문분야별로 수행된 메타분석 연구의 질은 통계적으로 유의한 차이가 있었다. 즉, 간호학과 물리치료학 분야에서 수행된 연구들의 AMSTAR 질 평가 점수가 높았으며 그 외 체육학과 의학 분야의 점수는 낮게 나타났다. 이와 같은 국내 학문간의 차이는 한국보건의료연구원에서 제시한 메타분석 보고 기준의 지침 개발 활동 등에 참여하여 학문 간의 특성을 반영하는 효율적인 보고지침을 공유함으로써 해소될 것으로 생각된다. 그 외 연구특성 중 포함된 연구들의 총 대상자 수가 많은 메타분석 연구와 검색에 사용된 데이터베이스 수가 4개 이상인 경우 AMSTAR 질 평가 점

수가 높아 선행 연구와도 같은 결과를 보였다[22]. 그러나 분석 단위 논문 편수에 따른 질 평가 점수에서는 통계적으로 유의한 차이가 없었다. 선행 연구[11]에 따르면 엄격한 선정 기준과 배제 기준을 적용 할수록 선정되는 논문 편수는 제한되지만 메타분석 연구의 질 평가 점수는 높게 나타났다. 따라서, 추후 연구자들은 메타분석 연구 수행 시 논문의 편수에 의미를 두기 보다는 정밀하게 계획된 엄격한 선정 및 배제 기준의 적용에 중점을 두어야 할 것이다. 본 연구에서는 2000년 이전과 2000년에서 2009년, 2010년 이후의 출판년도 간에 논문의 질 평가점수에서 통계적으로 유의한 차이는 없었다. 그러나 최근에 이를수록 논문의 질이 향상되어 2010년 이후 출판된 논문의 질이 유의하게 높다고 보고한 연구 결과[3]와 유사한 결과를 보였다.

메타분석 연구의 장점은 효과크기를 이용하여 개별 연구 결과를 통합하여 관심 변수의 효과에 대한 방향과 크기를 제시할 수 있다는 것이다[4]. 즉, 메타분석 연구의 최종 산물은 효과크기라고 할 수 있다. 따라서, 이러한 효과크기가 연구와 실무에 어떤 중요한 의미를 부여하는지, 효과크기를 어떻게 연구와 실무에 활용할 것인지에 대한 종합적인 정보 제공을 위한 고찰이 필요하다고 본다. 이에 본 연구에서는 운동 관련 메타분석 연구에서 제시한 효과크기를 기준[23]에 따라 분류하여 연구특성에 따라 비교하였다. 본 연구 결과 운동중재의 종류에 따라 효과크기는 중간 수준의 효과크기가 가장 많았으며 통계적으로 유의하게 다르게 분포하였다. 즉, 높은 효과크기에는 복합운동이, 낮은 효과크기에는 유산소운동이 많이 분포하여 연구자와 실무자들에게 운동중재의 방향을 제시할 수 있는 결과를 도출하였다. 또한, 선행 연구의 결과[25-28]에서 유산소 운동 등의 단일 운동중재보다 두 가지 이상의 복합운동중재에서 다양한 연령층의 신체적, 생리적 효과가 더 높게 나타난 것으로 보고하였는데 본 연구 결과를 통해 운동종류 별 효과크기의 양상을 구체적으로 확인할 수 있었다. 한편, 효과크기는 운동의 효과, 대상자의 성별, 연령, 질환 유무, 측정방법에 따라 다양한 분포를 보였으나 통계적으로 유의한 차이는 없었다. 대상 논문의 약 80%인 성인과 노인이 포함된 메타분석 연구의 효과크기가 높은 반면, 아동 대상에서의 효과크기는 중간과 낮은 수준으로 분포하였다. 이처럼 낮은 혹은 중간 수준의 효과크기 결과에 대해 아동의 성장과 성숙 정도가 개인별로 다르며 이러한 특성이 효과크기에 영향을 미친 것으로 보고된 바 있다[29]. 또한, 아동은 각 연령에 따라 신체적, 심리사회적 및 인지수준이 다르므로 자가보고식 도구의 한계를 극복할 수 있는 대안이 필요함을 논의하기도 하였다[29,30]. 즉, 대상자 자체의 특성에 따라 효과 측정에서 발생하는 일관성의 결여와 연구 설계가 효과크기에 영향을 준 것으로 보이며 도구의 이질성을 극복할 수 있는 도구개발 노력이 필요하다. 또한, 대상자 선정 시 대상자의

개인적 특성별로 분류될 수 있는 상호배타적인 기준의 제시 및 세분화된 연령 대상 운동중재의 디자인 적용이 낮은 효과크기를 높일 수 있는 방법으로 생각된다.

본 연구는 운동 관련 메타분석 논문의 연구특성 및 효과크기 분석과 연구의 질 평가를 처음 시도하였으며 간호학뿐 아니라 다양한 학문적으로 몇 가지 의의가 있었다고 본다. 첫째, 본 연구를 통해 다양한 학문 분야에서 운동 관련 메타분석 연구에 대한 관심과 수요가 증가하고 있음을 구체적으로 확인할 수 있었다. 따라서, 학문간 교류를 통해 메타분석 연구의 질 수준을 높이는 노력이 필요하며 운동프로그램의 체계적 고찰과 메타분석 결과에 대한 정보교환이 필요하다. 둘째, 현재 간호분야에서 수행되고 있는 운동 중재 연구의 경향을 확인할 수 있었으며, 일반인 뿐 아니라 급만성질환자 대상의 운동 중재 연구의 필요성을 확인하였다. 셋째, 본 연구에서는 개별 운동 관련 메타분석 연구의 결과를 바탕으로 연구특성과 효과크기를 다양한 측면에서 통합적으로 고찰해 봄으로써 추후 운동 관련 연구의 방향과 연구수행의 근거를 제공하였다. 넷째, 본 연구를 통해 운동중재의 종류에 따라 효과크기의 수준에 차이가 있음을 확인하였으므로 간호사는 운동간호중재를 적용할 때 임상적 의사결정과 간호수행의 과학적 근거로 적용해야 할 것이다. 다섯째, 체계적 고찰과 메타분석 연구 결과를 간호학뿐 아니라 여러 학문분야에서 의사결정에 활용할 수 있도록 연구 가이드라인에 맞는 간호교육이 필요하다. 이러한 과정을 통해 질 높은 체계적 고찰과 메타분석 연구가 활발해 질 것으로 예상된다. 한편, 간호학과 같은 의료계의 경우 학문적 범위를 고려하여 일반인뿐만 아니라 만성질환자나 급성질환자를 대상으로 하는 다양한 실험 연구 및 비실험 연구의 수행이 우선 요구된다고 본다. 수준 높은 메타분석 연구를 통해 이들 연구 결과를 통합하여 올바른 실무의 방향을 제시하는 것이 궁극적으로 근거기반 간호실무 향상에 기여할 것이다. 하지만 본 연구는 다양한 메타분석 연구 질 평가도구 중 AMSTAR를 사용하여 제시한 기준에 따라 메타분석 연구의 질을 평가하였고 국내에서 출판된 운동 관련 메타분석 논문들만을 포함하였으므로 해석과 전체 연구로 일반화하는 데는 제한이 있을 수 있다.

## 결 론

본 연구는 국내 학술지에 게재된 운동 관련 메타분석 33개 논문을 대상으로 연구특성을 파악하고 연구특성 별 효과크기를 비교하였으며, AMSTAR를 사용하여 연구의 질을 평가하였다.

본 연구 결과 운동종류에 따라 효과크기에 유의한 차이가 나타났으며, 연구의 질 평가점수는 11점 만점에 4.18점으로 33개 논문 중 66.7%가 낮은 수준의 질 평가점수 범주에 포함되었다. AMSTAR 평

가 점수를 연구특성에 따라 비교한 결과 간호학과 물리치료학 분야의 논문에서 유의하게 높았으며, 총 대상자가 1,000명이 넘는 경우, 데이터베이스를 4개 이상 사용한 경우, SSCI나 SCI (E)급에 게재된 경우, 연구비 지원이나 IRB 승인을 받은 경우에 연구의 질 평가 점수가 통계적으로 유의하게 높게 나타났다.

이와 같은 연구 결과를 토대로 현재까지의 운동 관련 메타분석 연구의 질 수준과 개선되어야 할 보고기준을 확인할 수 있었다. 또한, 다양한 학문분야별 운동 관련 메타분석 연구의 질과 효과크기의 수준을 확인함으로써 운동의 다 학문적인 특성과 임상 및 지역사회에서의 활용성 및 효과성을 파악할 수 있었으며, 이를 바탕으로 추후 메타분석 연구의 질 향상을 위한 교육적 전략과 근거 기반 간호실무 향상을 위해 활용될 수 있을 것으로 기대한다.

## REFERENCES

1. Yoo JY, Oh EG. Level of beliefs, knowledge and performance for evidence-based practice among nurses experienced in preceptor role. *Journal of Korean Academy of Nursing Administration*. 2012;18(2):202-212. <http://dx.doi.org/10.11111/jkana.2012.18.2.202>
2. Shin IS, Park EY. Review of the meta-analysis research in special education and related field. *Korean Journal of Physical and Multiple Disabilities*. 2011;54(4):157-176.
3. Kim JH, Kim AK. A quality assessment of meta-analyses of nursing in South Korea. *Journal of Korean Academy of Nursing*. 2013;43(6):736-745. <http://dx.doi.org/10.4040/jkan.2013.43.6.736>
4. Jang DH, Shin IS. Historical development of meta-analysis as an educational research methodology. *The Journal of Curriculum and Evaluation*. 2011;14(3):309-332.
5. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: The QUOROM statement. *Quality of reporting of meta-analyses*. *Lancet*. 1999;354(9193):1896-1900.
6. Dixon E, Hameed M, Sutherland F, Cook DJ, Doig C. Evaluating meta-analyses in the general surgical literature: A critical appraisal. *Annals of Surgery*. 2005;241(3):450-459.
7. MacDonald SL, Canfield SE, Fesperman SF, Dahm P. Assessment of the methodological quality of systematic reviews published in the urological literature from 1998 to 2008. *The Journal of Urology*. 2010;184(2):648-653. <http://dx.doi.org/10.1016/j.juro.2010.03.127>
8. Dijkman BG, Abouali JA, Kooistra BW, Conter HJ, Poolman RW, Kulkarni AV, et al. Twenty years of meta-analyses in orthopaedic surgery: Has quality kept up with quantity? *The Journal of Bone and Joint Surgery American Volume*. 2010;92(1):48-57. <http://dx.doi.org/10.2106/jbjs.i.00251>
9. De Vito C, Manzoli L, Marzuillo C, Anastasi D, Boccia A, Villari P. A systematic review evaluating the potential for bias and the methodological quality of meta-analyses in vaccinology. *Vaccine*. 2007;25(52):8794-8806. <http://dx.doi.org/10.1016/j.vaccine.2007.10.034>
10. Melchioris AC, Correr CJ, Venson R, Pontarolo R. An analysis of quality of systematic reviews on pharmacist health interventions. *International Journal of Clinical Pharmacy*. 2012;34(1):32-42. <http://dx.doi.org/10.1007/s11096-011-9592-0>
11. Papageorgiou SN, Papadopoulos MA, Athanasiou AE. Evaluation of methodology and quality characteristics of systematic reviews in orthodontics. *Orthodontics & Craniofacial Research*. 2011;14(3):116-137. <http://dx.doi.org/10.1111/j.1601-6343.2011.01522.x>
12. Suebnukarn S, Ngamboonsirisingh S, Rattanabanlang A. A systematic evaluation of the quality of meta-analyses in endodontics. *Journal of Endodontics*. 2010;36(4):602-608. <http://dx.doi.org/10.1016/j.joen.2009.12.019>
13. Kim YK. A review of the meta-analysis in major academic journal of business management in Korea. *Korea Journal of Business Administration*. 2010;23(4):1833-1858.
14. Suh MO. The review of meta-analysis and research methodology proposed. *Korean Journal of Educational Research*. 2011;49(2):1-23.
15. Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care from the potsdam consultation on meta-analysis. *Journal of Clinical Epidemiology*. 1995;48(1):167-171.
16. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: A proposal for reporting. Meta-analysis of observational studies in epidemiology (MOOSE) group. *JAMA: the Journal of the American Medical Association*. 2000;283(15):2008-2012.
17. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Journal of Clinical Epidemiology*. 2009;62(10):1006-1012. <http://dx.doi.org/10.1016/j.jclinepi.2009.06.005>
18. Oxman AD, Guyatt GH, Singer J, Goldsmith CH, Hutchison BG, Milner RA, et al. Agreement among reviewers of review articles. *Journal of Clinical Epidemiology*. 1991;44(1):91-98.
19. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: A measurement tool to assess the methodological quality of systematic reviews. *BMC Medical Research Methodology*. 2007;7:10. <http://dx.doi.org/10.1186/1471-2288-7-10>
20. Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *Journal of Clinical Epidemiology*. 2009;62(10):1013-1020. <http://dx.doi.org/10.1016/j.jclinepi.2008.10.009>
21. Kim SY, Park JE, Seo HJ, Lee YJ, Jang BH, Son HJ, et al. NECA's guidance for undertaking systematic reviews and meta-analyses for intervention. Seoul: National Evidence-based Healthcare Collaborating Agency; 2011.
22. Matjasko JL, Vivolo-Kantor AM, Massetti GM, Holland KM, Holt MK, Cruz JD. A systematic meta-review of evaluations of youth violence prevention programs: Common and divergent findings from 25 years of meta-analyses and systematic reviews. *Aggression and Violent Behavior*. 2012;17(6):540-552. <http://dx.doi.org/10.1016/j.avb.2012.06.006>
23. Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
24. Burda BU, Norris SL, Holmer HK, Ogden LA, Smith ME. Quality varies across clinical practice guidelines for mammography screening in

- women aged 40-49 years as assessed by AGREE and AMSTAR instruments. *Journal of Clinical Epidemiology*. 2011;64(9):968-976. <http://dx.doi.org/10.1016/j.jclinepi.2010.12.005>
25. Lee MH, Baek SS. A meta-analysis of optimal exercise time and frequency for fat loss in overweight elementary school students. *The Korean Journal of Elementary Physical Education*. 2012;18(2):199-210.
  26. Ho SS, Dhaliwal SS, Hills AP, Pal S. The effect of 12 weeks of aerobic, resistance or combination exercise training on cardiovascular risk factors in the overweight and obese in a randomized trial. *BMC Public Health*. 2012;12:704. <http://dx.doi.org/10.1186/1471-2458-12-704>
  27. Davidson LE, Hudson R, Kilpatrick K, Kuk JL, McMillan K, Janiszewski PM, et al. Effects of exercise modality on insulin resistance and functional limitation in older adults: A randomized controlled trial. *Archives of Internal Medicine*. 2009;169(2):122-131. <http://dx.doi.org/10.1001/archinternmed.2008.558>
  28. Chung CW, Lee S, Hwang SW, Park EH. Systematic review of exercise effects on health outcomes in women with breast cancer. *Asian Nursing Research*. 2013;7(3):149-159. <http://dx.doi.org/10.1016/j.anr.2013.07.005>
  29. Sung KS, Yoon YM, Kim EJ. Meta-analysis of the effects of obesity management program for children. *Child Health Nursing Research*. 2013;19(4):262-269. <http://dx.doi.org/10.4094/chnr.2013.19.4.262>
  30. McCartney K, Rosenthal R. Effect size, practical importance, and social policy for children. *Child Development*. 2000;71(1):173-180.