

— 원저 —

임상기반 간호학 실험논문의 질 평가

정재심¹ · 임경춘² · 권정순³ · 곽찬영⁴ · 김경희⁵ · 김성재⁶ · 이경숙⁷ · 박광옥⁸ · 김주현⁹ · 김금순¹⁰ · 최명애¹¹

¹울산대학교 임상전문간호학; ²성신여자대학교 간호대학; ³서울아산병원, 간호팀장; ⁴한림대학교 간호대학; ⁵중앙대학교 간호학과; ⁶서울대학교 간호대학; ⁷강릉원주대학교 간호학과; ⁸순천대학교 간호학과; ⁹강원대학교 간호학과; ¹⁰서울대학교 간호대학; ¹¹서울대학교 간호대학

Assessing Quality and Risk of Bias in Experimental Studies on Clinical Nursing

Jeong, Jae Sim¹ · Lim, Kyung Choon² · Kwon, Jeong Soon³ · Kwak, Chan Yeong⁴ · Kim, Kyung Hee⁵ · Kim, Sung Jae⁶
Lee, Kyung-Sook⁷ · Park, Kwang Ok⁸ · Kim, Joo Hun⁹ · Kim, Keum Soon¹⁰ · Choe, Myoung Ae¹¹

¹Professor, Department of Clinical Nursing, University of Ulsan Graduate School of Industry; ²Professor, College of Nursing, Sungshin University;

³Team Leader, Nursing Department, Asan Medical Center; ⁴Professor, College of Nursing, Hanllm University;

⁵Professor, Department of Nursing, Chung-Ang University; ⁶Professor, College of Nursing, Seoul National University;

⁷Professor, Department of Nursing, Gangneung Wonju National University; ⁸Associate Professor, Suncheon University;

⁹Emeritus Professor, Department of Nursing, Kangwon National University; ¹⁰Emeritus Professor, College of Nursing, Seoul National University;

¹¹Emeritus Professor, College of Nursing, Seoul National University

Purpose: This study was conducted in order to explore quality assessment tools applicable to research in the nursing area and to suggest the results of applying such tools in assessing the quality of experimental studies in nursing. **Methods:** We selected three experimental and 119 quasi-experimental studies with control groups in the Journal of Korean Clinical Nursing Research. We applied Cochran collaboration's tool for assessing the quality of randomized controlled studies. Also, items developed by researchers were applied for assessing the quality of experimental studies in nursing. **Results:** Three studies (2.5%) used an adequate randomized sequence generation and two of them concealed the allocation sequence. The 63.1% of the studies applied blinding and 51% of the studies did not have missing values, or had insignificant missing values. In addition, 85.2% were not likely to report selectively, and 36.1% were assessed to have raised the risk of bias. **Conclusion:** Many studies had too a high risk of bias to be used as evidence-based data in the designs and methods of experimental studies. We found that it is necessary to inquire into research designs, methods, and quality assessment tools desirable for evidence-based practice in nursing.

Key words: Nursing, Experimental study, Quality assessment, Risk of bias, Evidence-based practice

서론

최근 의학을 비롯하여 보건 의료계열에서 근거기반실무의 중요성이 강조되고 있다. 근거기반실무를 위해서는 근거기반실무지침이 개발되어야 하고 근거기반실무지침의 개발단계에서 가장 핵심적인 과정 중의 하나가 연구의 질 평가이다.¹⁾ 연구의 질 평가는 연구 결과가 얼마나 비뚤림(risk of bias)없이 수행되었는지를 평가하여 근거로 적용이 가능한지를 분석하는데 사용된다.²⁻⁴⁾

간호학에서는 연구설계를 질적연구, 양적연구로 구분한 뒤, 양적 연구를 다시 실험연구와 비실험연구로 나누고 다시 각 범주별로 구체적인 연구설계 방법을 제시한다. 그러나 의학과 보건학 연구에서는 크게 분석적 연구(analytic study)와 비분석적연구(nonanalytic study)로 분류한 뒤 분석적 연구를 다시 중재(실험)연구와 관찰연구로 구분한다. 실험연구에는 무작위대조군연구(randomized controlled trial), 준무작위연구(quasi-randomized trial), 비무작위연구(non-randomized trial)가 있고, 관찰연구에는 대조전후연구(con-

주요어: 간호, 실험연구, 질 평가, 비뚤림, 근거기반실무

*본 연구는 2009년 병원간호사회에서 연구비를 지원받아 진행하였음.

Address reprint requests to : Lim, Kyung Choon

School of Nursing, College of Nursing, Sungshin University, 147, Mia-dong, gangbuk-Gu, 01133 Seoul, Korea

Tel: +82-2-920-7729 Fax: +82-2-920-2092 E-mail: kclim@sungshin.ac.kr

투고일: 2018년 11월 30일 심사회의일: 2018년 12월 04일 게재확정일: 2018년 12월 26일

trolled before and after study), 코호트연구(cohort study), 환자-대조군 연구(case-control study), 전후연구(before and after study), 단면연구(cross-sectional study), 증례연구(case-series) 등이 있다.²⁾ 여러 가지 다양한 연구설계에서 실험연구는 조작화, 무작위화, 무작위배정 등을 거치므로 중재효과를 평가하는데 있어 관찰연구나 조사연구에 비하여 비뚤림의 가능성이 적다. 특히 무작위 대조군을 적용하는 무작위대조군연구가 가장 근거등급이 높은 것으로 알려져 있다.

연구의 질 평가를 하기 위하여 연구 설계에 따라 각기 다른 질 평가 도구들이 개발되어 있다.⁵⁻⁶⁾ 무작위대조군연구의 경우 Jadad 척도(Jadad quality assessment scale)나 코크란연합 질 평가도구를 많이 적용한다.⁷⁾ 비무작위연구를 위한 질 평가 도구로는 Downs and Black 도구,⁸⁾ 환자-대조군 연구를 위한 Newcastle-Ottawa 척도,⁷⁾ MI-NORS 도구(Methodological Index for Non-Randomized Studies)⁹⁾ 등이 사용된다.

간호학 분야에는 무작위배정연구가 드물고, 연구의 주제나 연구를 위한 임상 여건상 유사실험설계가 많이 적용되므로 비무작위배정연구를 위한 질 평가 도구가 주로 적용되어야 한다. 그러나 현재 개발된 질 평가 도구들은 모두 의학과 보건학 연구 설계를 기반으로 하여 개발된 도구들이다. 또한 간호학 연구에서 사용하는 연구 설계의 분류와 명칭이 의학이나 보건학 연구와 달라서 이들 도구를 선택하는데 어려움이 있을 수 있다. 국내 간호학에서는 해마다 수십 종의 학회지를 통하여 수 백 편이상의 논문이 발간되고 있다. 선행연구에서 주제나 대상자, 그리고 연구설계에 중점을 두고 연구 동향이나 실태를 분석한 연구들은 여러 편 발표되었다. 일부 선행연구¹⁰⁾에서 연구의 질 평가를 동향 분석과 함께 적용하기는 하였으나 질 평가가 주 목적은 아니기 때문에 표준화된 질 평가 도구를 적용하지는 않았다. 따라서 이 연구는 간호학 분야의 논문에 활용가 능한 질 평가 도구를 탐색하고, 실제로 간호학 실험논문의 질 평가에 적용한 결과를 제시하기 위해 시도되었다.

연구 방법

1. 분석대상 연구논문

병원간호사회의 “임상간호연구” 학회지는 임상에서 수행된 실험 연구 논문이 가장 많이 발표되므로 근거기반간호 실무와 가장 밀접한 관련이 있을 것으로 판단되어 이 학회지에 발표된 실험연구논문을 대상으로 질 평가를 실시하였다. “임상간호연구” 1995년 1권 1호(창간호)부터 2008년 14권 2호까지 13년간 발간되어 총 26권에 수록된 모든 논문을 대상으로 총 341개의 논문 중 순수실험연구와 대조군이 있는 유사실험연구를 선택하였다. 대조군이 없는 단일군 원시

실험연구(pre-experimental design)는 결과 해석에 비뚤림이 많고 인과관계 확인이나 일반화에 제한점이 많으므로¹¹⁾ 분석대상에서 제외하였다.

2. 자료수집도구

1) 질 평가 도구의 검토와 개발

무작위대조군연구를 위한 질 평가 도구와 비무작위연구를 위한 질 평가 도구를 수집하여 내용과 항목을 비교분석하였다. 본 연구에서 분석 대상으로 한 간호학 실험연구는 거의 대부분이 비무작위연구이지만 일부 연구(순수실험연구와 일부 유사실험연구)는 무작위배정을 시도하였으므로 무작위대조군연구가 포함되어 있었다. 원칙적으로 무작위대조군연구와 비무작위대조군연구를 구분하여 각각의 질 평가도구를 적용해야 한다. 그러나 간호학 분야에서 비뚤림 위험을 감소시키는 연구를 장려함으로써 근거 등급이 높은 연구결과를 많이 산출하기 위해 무작위대조군연구의 수를 증가시켜야 한다는 취지에서 본 연구에서는 대조군이 있는 실험연구를 모두 무작위대조군연구의 질 평가 도구를 이용하여 분석하였다.

무작위대조군연구를 위한 질 평가 도구 중 한국어로 번역이 된 것을 검색하였고, Kim 등²⁾이 코크란연합의 무작위대조군연구 질 평가 도구를 번역하여 제시한 한국판 도구를 선정하였다. 그러나 6개의 도메인으로 구성된 코크란연합의 질 평가 도구로는 간호학 연구에서 필요한 질 평가의 다른 측면을 분석하기 어렵다고 판단하여 본 연구팀에서 몇 가지의 간호학실험연구 질 평가 문항을 개발하여 추가하였다. 코크란 연합의 질 평가 도구와 저자 등이 개발한 간호학실험연구 질 평가 문항의 세부내용은 다음과 같다.

(1) 코크란연합의 무작위대조군연구 질 평가 도구

(이하 “무작위대조군연구 질 평가 도구”)

Kim 등²⁾이 제시한 도구는 1) 무작위 배정 여부, 2) 배정 순서의 은폐 여부, 3) 참여자, 연구자, 결과평가자에 대한 눈가림 여부 4) 결과 자료의 충분한 제공 여부 5) 선택적 보고 여부 6) 비뚤림 위험성 여부 총 6가지의 항목으로 구성되었다. 각 항목은 “예”, “아니오”, “불확실”의 세 가지 범주로 평가하도록 되어 있으며, 각 항목의 응답 시 참고가 가능하도록 사례들이 제시되어 있다. 본 연구에서는 평가 결과를 양적으로 산정하기 위하여 “예”는 1점, “아니오”와 “불확실”은 0점을 배정하였고, 각 항목에 대한 사례도 각각 번호를 배정하여 해당 사례를 내용별로 분석하였다.

(2) 간호학실험연구 질 평가 문항(이하 “실험연구 질 평가 문항”)

자체 개발된 문항은 간호학 실험연구에서 비폴립 위험을 감소시키기 위해 고려해야 할 항목이지만 Kim 등²⁾의 질 평가 도구에서 평가하지 않는 항목과 간호학연구의 특성상 자료수집 방법에 많이 적용되는 사회심리측정도구에 대한 평가 내용 등이 포함되었다. 실험연구 질 평가 문항 개발에는 모든 연구원이 참여하여 초안을 작성한 후 2회에 걸쳐 동일한 논문 3개를 각각 평가한 후 문항 내용을 수정보완하였다. 최종적으로 개발된 문항은 총 14개로 구성되었으며 1) 눈가림의 적용과 정도, 2) 실험군과 대조군의 동질성 평가 여부와 내용, 3) 실험처치의 안전성 4) 기술, 대조군에 대한 처치, 5) 연구 대상자 수 산정 근거와 방법, 6) 교차실험설계에서의 이월효과 처리, 7) 사회심리측정도구의 신뢰도와 타당도 검증, 8) 생리적 측정도구의 정확도 검증, 9) 종속변수의 종류와 측정방법 및 결과, 10) 기타 질 평가 관련 문제점 기술의 항목으로 구성되었다.

3. 자료수집

서울대학교 간호대학 연구대상자보호심사위원회의 연구계획서 심의를 받은 뒤 자료수집을 시작하였다. 연구원들이 질 평가를 시행할 때 겪는 여러 문제점들을 논의하기 위해 전문가를 초빙하여 “연구의 질 평가 방법 및 질 평가도구 개발”에 관한 교육 목적의 세미나를 개최하였다. 연구원들은 사전에 1개의 논문에 대하여 잠정적으로 선정된 코크란연합의 질 평가 도구를 이용하여 분석해 본 후 평가결과 및 제반 문제점에 대하여 질문과 응답의 형식으로 진행되었다. 연구자들간 신뢰도를 높이기 위해 세미나 후 몇 차례의 논의를 거쳐 평가방법에 대한 기본적인 내용을 공유한 후 전체적인 연구의 질 평가는 분석대상 논문 한 편당 연구원 2명이 한 조가 되어 코크란연합의 무작위대조군 질 평가 도구와 본 연구팀이 개발한 간호학실험연구 질 평가 문항을 이용하여 각각 독립적으로 평가하였다. 일차 자료수집 후 두 명의 결과를 비교하여 불일치하거나 불분명한 항목에 대하여는 전체 연구팀 회의에서 최종적으로 검토하여 결정하였다.

질 평가 도구와 항목에 따른 분석 이후에 간호학 실험연구의 비폴립 정도를 평가하기 위한 이차분석을 일부 시도하였다. 간호학 연구에서는 결과 측정을 위한 자료수집도구로 생리학적 측정도구와 사회심리측정도구 등의 다양한 측면의 도구가 동일한 결과 변수에 중복 적용되는 경우가 많다. 사회심리측정도구가 자가보고인 경우 결과 측정에서 대상자 비폴립이 발생할 가능성이 높다. 따라서 대상자나 측정자의 영향을 받지 않아 측정 비폴립이 적은 것으로 알려진 생리학적 측정도구와 사회심리측정도구에서의 결과 유의성을 비교하여 간접적으로 대상자 비폴립 존재 유무를 확인하였다.

4. 자료분석

두 명의 연구원이 최종 합의하여 한 개의 질 평가 점수를 산정한 뒤 각 논문단위로 엑셀에 입력하여 SPSS/WIN 22.0 프로그램을 이용하여 분석하였다. 코크란연합의 질 평가 점수는 각 항목별로 평가하여 빈도와 백분율로 비교하였다. 사회심리측정도구와 생리적 측정도구의 결과 유의성 차이는 논문의 결과변수별로 “통계적으로 유의한 차이가 있음(different)”, “부분적으로 차이가 있음(partially different)”, “차이가 없음(not different)”으로 구분하였다. 부분적으로 차이가 있음은 측정도구에 하위범주가 있고 하위 범주별로 비교하도록 되어 있는 경우(예를 들면 SF-36과 같은 삶의 질 도구)에 하위 범주 일부에서만 차이가 있는 경우가 해당된다. 하위범주가 없는 측정도구인 경우에는 전체 값의 결과를 그대로 비교하였다. 통계는 chi-square test를 적용하였고, 유의수준 5%하에서 양측검정하였다.

연구 결과

1. 연구대상 실험논문 설계빈도

순수실험연구는 11편(8.9%)이었고 이중 전후대조군 설계가 6건(4.9%)으로 가장 많았다. 대조군이 있는 유사실험연구는 111편(90.9%)이었으며, 유사실험연구 중에서는 비동등성대조군 전후시차설계(nonequivalent control group non-synchronized design) 50건(41.0%)으로 가장 많았고 다음이 비동등성대조군 전후설계(nonequivalent control group pretest-posttest design) 34건(27.9%), 비동등성대조군 사후설계(nonequivalent control group posttest only design) 17건(13.9%)이었다(Table 1).

2. 무작위대조군연구 질 평가도구 적용

무작위대조군연구 질 평가 도구를 이용하여 순수실험연구와 대조군이 있는 유사실험연구(이하 실험연구) 122편을 분석한 결과는 Table 2와 Figure 1과 같다.

무작위 배정 순서 생성이 적절하게 된 경우는 3건(2.5%)에 불과하였으며, 무작위 방법을 시행하지 않은 경우가 66건(54.9%)으로 가장 많았고, 무작위라고 하였지만 연구자가 임의로 배정한 경우가 27건(21.6%)으로 두 번째로 많았다. 그 이외에 부적절한 방법 사용 8건, 일정한 규칙을 이용한 경우가 6건으로 많았다.

배정순서 은폐에 대하여는 무작위 배정이 실시된 총 3건의 연구와 불확실한 연구 6편을 합하여 9편의 연구를 평가하였다. 이중 배정순서가 은폐된 경우는 2건이었고, 다른 2건은 부적절한 은폐방법

을 사용하였으며, 평가하기 어려운 경우가 5건이었다.

눈가림이 시행되었다고 기술되어 있거나, 연구 설계의 정확상 눈가림이 시행된 것으로 판단되는 경우(예를 들면 비동등성 대조군 전후 시차 설계)를 모두 합하여 59건(48.3%)의 연구에서 눈가림이 시행되었다. 결과평가자에 대한 눈가림이 시행된 경우는 10건(8.2%)이었고, 의료장비에 의하여 측정되어 눈가림이 결과에 영향을 미치지 않는 경우가 8건(6.6%)이었다.

분석을 위한 결과자료에서 결측치가 없이 분석된 경우가 65건(53.3%)으로 가장 많았다. 결과자료가 부적절한 경우가 33건(27.0%)이었는데, 이중 상당수의 결측치가 존재하고 결과에 비뚤함을 초래할 수 있는 경우가 19건(15.6%), 무작위 배정된 중재를 받지 않은 경우가 많은데도 중재받은 대로만 분석을 수행한 경우가 14건(11.4%)으로 나타났다.

분석에 사용된 자료가 선택적으로 보고되었을 가능성이 없어서 연구 프로토콜대로 수행된 것으로 확인된 경우가 36건(29.5%), 연구 프로토콜은 없으나 정확상 사전에 계획된 대로 분석된 것으로 확신가능한 경우가 68건(55.7%)이었다.

앞의 5개 항목 이외에 다른 비뚤림 요인이 없었던 것으로 확인된 경우가 42건(34.4%)이었고, 연구설계 관련 비뚤림이 있었던 경우 22건(18.0%), 피험자의 기저군간에 심각한 불균형이 있었던 경우가 14건(11.5%)으로 나타났다. 그리고 기술이 명확하지 않아 확인하기 어려운 불확실한 경우가 44건(36.1%)이었다.

3. 무작위배정연구 질 평가 도구의 점수 분포

무작위배정연구 질 평가 도구의 1, 3~6번 문항(2번 문항은 1번에 종속적이어서 제외)에서 예인 경우를 1점으로 아니오나 불확실한 경우를 0점으로 처리하여 각 항목 점수를 더한 총점의 분포를 보면

3점이 가장 많아서 50건(41.0%)이었고, 다음이 2점 30건(24.6%), 1점 24건(19.7%)이었으며, 4점 이상인 경우도 14건(11.4%)으로 나타났다.

4. 간호학실험연구 질 평가 도구 적용

본 연구팀이 개발한 간호학 실험연구 질평가 도구를 적용하여 순수실험연구와 대조군이 있는 유사실험연구 총 122편을 분석한 결과는 Table 3과 같다.

눈가림에 대한 기술이 있었던 경우는 16건(13.1%)이었으나, 기술은 없지만 정확상 실시된 것으로 판단된 경우가 71건(58.2%)으로 나타났다. 눈가림의 방법은 연구대상자 눈가림이 가장 많아서 78건(89.7%)으로 대다수이었고, 삼중눈가림(연구대상자, 연구자, 자료수집자)은 시행된 연구가 없었으며, 이중눈가림이나 자료수집자 눈가림도 5건, 4건으로 적었다.

동질성 검증은 일반적 특성과 종속변수에 대하여 모두 검증을 실시한 경우가 76건(62.3%)이었으나, 한 가지만 실시한 경우도 36건(29.5%)였고, 실시하지 않은 경우도 10건(8.2%)이 있었다.

실험처치의 안전성에 대하여 기술한 논문은 3건에 불과하였고, 대부분의 연구에서 실험처치의 안전성에 대한 기술이 없었다.

대조군에 대한 처치에 대하여 기술하지 않은 경우가 35건(28.8%)이었고, 기술한 경우를 보면 다른 처치가 가장 많아서 60건(49.1%), 처치가 적용되지 않은 경우 25건(20.5%)이었다. Sham이나 placebo가 적용된 경우는 2건에 불과하였다.

교차실험 설계가 적용된 논문은 총 19건이었고, 교차실험에서 필수적으로 고려해야 할 이월효과 예방에 대한 내용이 기술된 논문은 7건(36.8%)이었다.

실험연구의 경우에는 연구결과의 타당도 검증을 위하여 연구대상자 수가 사전에 설계되어야 한다. 연구대상자 산정에 대한 기술이

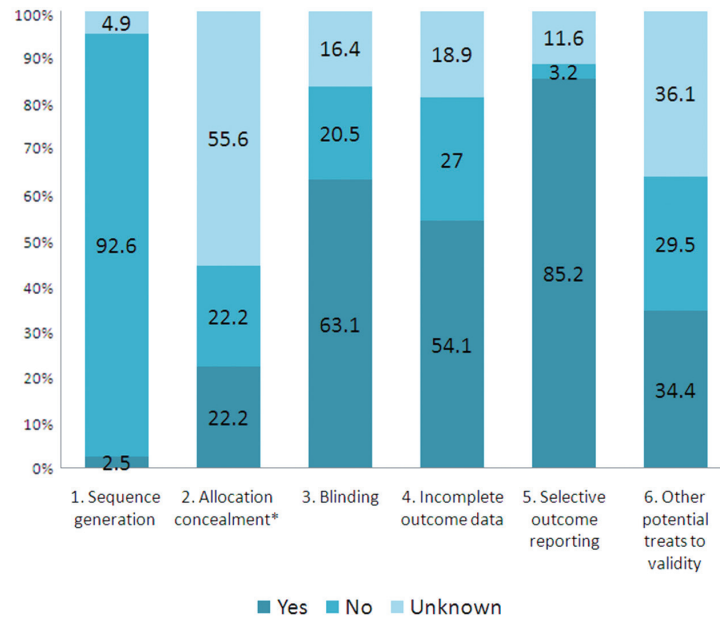
Table 1. Proportion of Experimental Study Designs

(N=122)

Design	Categories	n (%)
Experimental design	Pretest-post test design	6 (4.9)
	Posttest only control group design	1 (0.8)
	Cross-over design	2 (1.6)
	Randomized controlled trial	1 (0.8)
	Time series design	1 (0.8)
	Subtotal	11 (8.9)
Quasi-experimental design	Nonequivalent control group pretest-posttest design	34 (27.9)
	Nonequivalent control group posttest only design	17 (13.9)
	Nonequivalent control group non-synchronized design	50 (41.0)
	Simulated control group pretest-posttest design	2 (1.6)
	Reversed treatment nonequivalent control design with pretest and posttest	1 (0.8)
	One group cross-over design	3 (2.5)
	Nonequivalent control group non-synchronized posttest only design	4 (3.2)
	Subtotal	107 (87.7)
Total		122 (100.0)

Table 2. Evaluation of Quality and Risk of Bias in Experimental Studies based on Clinical Nursing according to the Cochrane Collaboration's Assessment Tool (N=122)

Categories		Subcategories	n (%)
1. Sequence generation	Yes	Referring to a random number table or using a computer random number generator	1 (0.8)
		Using coin tossing, shuffling cards or envelopes (Picked card should be put inside again), throwing dice, etc. However, clearly stated that 'random' is implemented in allocation sequence.	2 (1.7)
		Subtotal	3 (2.5)
	No	Do not perform random assignment	66 (54.9)
		Use of inadequate method	8 (6.4)
		Sequence generated by the odd or even number of patient ID or hospital numbers	6 (4.8)
		Sequence generated by some rule based on date of birth, day of admission	2 (1.7)
		Allocation by judgement of the clinician	1 (0.8)
		Allocation by the person responsible for determining the eligibility	27 (21.6)
		Allocation by preference of the participant	1 (0.8)
		Allocation by the results of a laboratory test or series of tests	1 (0.8)
		Allocation by the results of a laboratory test or order of medication preparation	1 (0.8)
		Subtotal	113 (92.6)
	Unclear	Insufficient information about the sequence generation process to permit judgment of 'Yes' or 'No'.	6 (4.9)
2. Allocation concealment (n=9)	Yes	Sequentially numbered, opaque, sealed envelopes	2 (22.2)
	No	Do not use allocation concealment	1 (11.1)
		Using an open random allocation schedule (e.g. a list of random numbers)	1 (11.1)
	Unclear	Insufficient information to permit judgement of 'Yes' or 'No'.	5 (55.6)
3. Blinding of participants, research personnel, and outcome assessors	Yes	Blinding of participants and key study personnel ensured, and unlikely that the blinding could have been broken	59 (48.3)
		Either participants or some key study personnel were not blinded, but outcome assessment was blinded and the non-blinding of others unlikely to introduce bias.	10 (8.2)
		No blinding, but the review authors judge that the outcome and the outcome measurement are not likely to be influenced by lack of blinding	8 (6.6)
		Subtotal	77 (63.1)
	No	No blinding or incomplete blinding, and the outcome or outcome measurement is likely to be influenced by lack of blinding	23 (18.9)
		Blinding of key study participants and personnel attempted, but likely that the blinding could have been broken,	2 (1.6)
		Subtotal	25 (20.5)
	Unclear	Insufficient information to permit judgement of 'Yes' or 'No'	20 (16.4)
4. Incomplete outcome data	Yes	No missing outcome data	65 (53.3)
		For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes not enough to have a clinically relevant impact on observed effect size	1 (0.8)
		Subtotal	66 (54.1)
	No	Reason for missing outcome data likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across intervention groups	19 (15.6)
		'As-treated' analysis done with substantial departure of the intervention received from that assigned at randomization	14 (11.4)
		Subtotal	33 (27.0)
	Unclear	Insufficient reporting of attrition/exclusions to permit judgement of 'Yes' or 'No'	23 (18.9)
5. Selective outcome reporting	Yes	The study protocol is available and all of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way	36 (29.5)
		The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified	68 (55.7)
		Subtotal	104 (85.2)
	No	Not all of the study's pre-specified primary outcomes have been reported	2 (1.6)
		One or more primary outcomes is reported using measurements, analysis methods or subsets of the data (e.g. subscales) that were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected adverse effect)	2 (1.6)
		Subtotal	4 (3.2)
	Unclear	Insufficient information to permit judgement of 'Yes' or 'No'. It is likely that the majority of studies will fall into this category	14 (11.6)
6. Other potential threat to validity	Yes	The study appears to be free of other sources of bias.	42 (34.4)
	No	There is at least one important risk of bias	22 (18.0)
		Had extreme baseline imbalance	14 (11.5)
		Subtotal	36 (29.5)
	Unclear	There may be a risk of bias, but there is either: insufficient information to assess whether an important risk of bias exists; or insufficient rationale or evidence that an identified problem will introduce bias	44 (36.1)



* Only "yes" and "unknown" responses to "1. sequence generation" item (n= 9) were analysed.

Figure 1. Summary of the assessment result of quality and risk of bias in experimental nursing studies according to the Cochrane collaboration's assessment tool (N=122).

Table 3. Evaluation of Risk of Bias in Experimental Studies based on Clinical Nursing according to the Quality and Risk of Bias Assessment Items Developed by Authors (N=122)

Variables	Categories	n (%)
Blind	Applied	16 (13.1)
	Not described, but assumed by situation	71 (58.2)
	Not applied	35 (28.7)
Method of blind (n=87)	Single (subject) blinded	78 (89.7)
	Double (subject, researcher) blinded	5 (5.7)
	Accessors (analyser) blinded	4 (4.6)
Homogeneity test	General characteristics only	32 (26.2)
	Dependent variables only	4 (3.3)
	Both	76 (62.3)
	Neither	10 (8.2)
Safety of experimental treatment	Described	3 (2.5)
	Not described	119 (97.5)
Treatment for the control group	Sham or placebo	2 (1.6)
	Different treatment	60 (49.1)
	No treatment	25 (20.5)
	No description	35 (28.8)
Effect size determination for sample size calculation	Based on previous studies	1 (0.8)
	Calculated arbitrarily by researcher	10 (8.2)
	No description	111 (91.0)
Description of carry-over effect prevention method in crossover design (n=19)	Yes	7 (36.8)
	No	12 (63.2)
Translation, back translation, and preliminary survey of unstandardized foreign language tools (n=24)	Yes	1 (4.0)
	No	23 (96.0)
Expert panel verification for developed, modified, or complemented tools (n=53)	Yes	29 (54.7)
	No	24 (45.3)
Reliability test of questionnaire (n=71)	Yes	57 (80.3)
	No	14 (19.7)
Calibration of biological measurement instrument (n=45)	Yes	12 (26.7)
	No	30 (66.7)
	Partially performed	3 (6.6)

있는 경우는 11건(9.0%)에 불과하였고, 연구대상자 산정에 대한 근거가 기술되어 있는 11건의 논문 중 효과크기 산정에 대한 근거가 기술된 경우는 4건에 불과하였다. 효과 크기를 선행연구나 pilot test 등을 통하여 적절하게 산정하여 제시한 경우는 1건에 불과하였고, 연구자가 임의로 산정한 경우가 10편으로 대다수였다.

설문지(사회심리측정 도구 포함)를 사용한 연구는 59%이었다. 설문지를 사용한 72개의 연구에서 총 122개의 설문지가 사용되었고, 설문지의 원저는 표준화된 한국판도구가 59.8%로 가장 많았고, 표준화되지 않은 도구도 19.7%가 사용되었다. 표준화되지 않은 외국도구 사용 시 번역, 역번역, 예비조사 시행여부를 확인하였을 때 23건(96.0%)의 설문지는 이러한 절차에 대한 기술이 없어 절차없이 사용한 것으로 드러났다.

설문지를 사용하면서 연구자가 개발하거나 수정보완하여 사용한 설문지 도구 53개 중 전문가의 타당성 검증은 29건(54.7%)의 설문지에서만 수행되었으나, 자신의 연구에서 신뢰도를 검증한 경우는 57건(80.3%)이었다.

생리적 측정도구를 이용하여 자료수집을 실시한 경우 45가지의 생리적 측정도구가 사용되었고, 이중 도구의 정확도 검증(calibration)이 시행되었다고 기술된 연구는 12건(26.7%)이었다.

5. 자료수집 방법에 따른 실험처치의 효과분석

자료수집 방법에 따라 실험처치의 효과 차이를 분석한 결과 사회심리적 측정도구를 이용한 설문지법(사회심리적 측정도구)은 차이가 있는 경우가 69.5%인데 비하여 생리적 측정법은 43.8%로 낮았다. 시각적 상사 척도(Visual Analogue Scale, VAS)나 설문지는 각각 68.6%, 66.7%로 사회심리적 측정도구와 유사하였다. 차이가 없는 경우도 사회심리적 측정도구, VAS, 설문지가 각각 13.3%, 20.0%, 18.2%로 유사하게 낮은 빈도를 나타내었다. 그러나 비교적 더 객관적인 것으로 알려진 생리적 변수와 관찰법은 42.7%와 35.1%로 실험처치에 차이가 없다는 결과가 더 많았다.

이를 다시 재분류하여 사회심리적 측정도구/VAS/설문지와 생리적 측정법/관찰법의 두 가지 범주로 나누어 분석한 결과 통계적으

로 매우 유의한 차이($p < .001$)를 나타내었다(Table 4).

논 의

근거기반간호에서 가장 근거 등급이 높은 연구설계는 다기관 무작위대조군연구(multicenter randomized controlled trial)이다.¹²⁾ 무작위배정은 실험연구에서 대상자에게 동등한 선택기회를 부여한 상태에서 처치조건에 할당하는 것으로 순수실험연구에서 필수요소이다.¹³⁾ 대상연구들의 분석결과 유사실험연구가 대부분을 차지하였는데, 유사실험설계의 경우 사전-사후 검증을 거쳤다고 하더라도 연구자가 파악하기 어려운 혼동변수(confounding variable)의 존재 가능성이 있으므로 선택 비뮴림(selection bias)이 존재할 여지가 크다.^{11,12)} 무작위배정을 하였다고 기술이 되어 있으나, 부적절한 경우가 많았는데, 환자를 등록번호나 입원 순서에 따라 배정한다든지 연구자 임의로 배정하는 것 등이 이에 해당된다. 흔히 연구자들이 무작위 배정과 무작위 추출을 기술하면서 혼동하는 경우가 많은데,¹⁴⁾ 무작위로 배정하였다는 것이 무작위 추출을 의미할 수도 있다고 생각되었으나, 확인하기는 어려웠다. 간호학연구의 특성상 무작위배정이 어려운 경우도 있으나, 실제로 무작위배정이 불가능한지에 대하여 연구 설계 단계에서 충분히 검토할 필요가 있는 것으로 생각된다.

무작위배정이 된 논문과 불확실한 논문 9편을 대상으로 배정순서는 은폐여부를 확인한 결과 무작위 배정순서에 의해 일련번호가 기록되어 있는 동일한 모양의 포장을 사용하였다는 연구는 2편에 불과하였다. 무작위배정시에는 연구자나 대상자가 배정되는 시점까지 배정순서를 은폐하도록 권장한다.¹²⁾ 이는 연구자 등이 사전에 배정순서를 알 경우 선택 비뮴림이 발생할 수 있는 것을 사전에 방지하기 위한 목적이며 밀봉한 봉투를 사용하는 등의 방법을 권장한다.¹²⁾ 예를 들면, 난수표로 선정된 무작위배정 순서가 기재된 표를 이용하여 배정하는 것도 연구자가 사전에 순서를 알 수 있으므로 은폐가 이루어지지 않은 것으로 간주한다.

눈가림(blinding)은 대상자가 자신이 연구대상자로 선정된 것을 알게 되는 경우에 평소와 다르게 행동하는 것이나 연구자가 누가

Table 4. Difference of Treatment Effect by the Type of Outcome Measurement Methods

(N=122)

Type of outcome measurement instruments*	Different n(%)	Partially different** n(%)	No difference n(%)	χ^2	p
Psychosocial/visual analogue scale/ self-administered questionnaire	119/173 (68.8)	27/173 (15.6)	27/173 (15.6)	326.00	< .001
Biological/observational	85/166 (51.2)	16/166 (9.6)	65/166 (39.2)		

*Other measurement(i.e., interview, etc) were excluded; **When the partial results of instrument were significantly different(in case of instrument with subcategories).

실험군이고 대조군인지 알게 될 경우 미치는 비뚤림 등을 방지하기 위하여 적용된다.¹²⁾ 눈가림은 무작위배정과 마찬가지로 비뚤림을 제거하기 위한 중요한 방법 중의 하나이다.¹²⁾ 특히 효과를 측정하는 결과 변수가 대상자나 측정자의 주관적인 인지상태에 따라 좌우되는 경우에는 눈가림을 하지 않고 자료가 수집되었을 때 연구의 내·외적 타당도에 대한 문제가 제기될 수 있다.¹¹⁾ 내적타당도를 위협하는 후광효과(Halo effect)는 대상자가 실험군에 속한 것을 연구자가 알고 있어서 연구 대상자의 반응을 정확히 사정하지 않고 추측하여 잘못 평가하는 것이다.¹¹⁾ 외적타당도를 위협하는 Hawthorne 효과는 대상자가 자신이 연구대상으로 선정된 사실을 알게 될 때 보통 때와 달리 반응하는데서 초래된 결과이다.¹¹⁾ 이와 같은 연구의 내·외적 타당도를 위협하는 요인들을 통제하기 위해서 연구자는 연구설계 과정에서 단일눈가림(대상자), 이중눈가림(대상자, 연구자), 삼중눈가림(대상자, 연구자, 측정자)방법 중에서 가장 적절한 방법을 선택하여 연구진행 과정에 적용하여야 한다.¹²⁾

간호학에서 많이 적용하는 비동등성 대조군 전후 시차설계의 경우 시간차이로 인하여 대상자 눈가림이 이루어진 것으로 판단할 수 있는 경우가 89.2%로 비교적 많았다. 그러나 연구자가 중재를 시행하고 결과변수도 동시에 측정하는 경우가 많아서 연구자 눈가림은 거의 이루어지지 않고 있었다. 이 경우 자료수집자를 실험군과 대조군이 누구인지 모르도록 한 제 3의 인물로 지정하는 자료수집자 눈가림을 고려해 볼 수 있는데,⁷⁾ 이렇게 한 연구는 4편(4.6%)으로 적었다(결과로 제시하지는 않았음).

결과를 분석할 때 실험연구의 경우 “배정된대로 분석(Intention To Treat analysis, ITT analysis)인지 “연구계획서대로 분석 또는 치료 받은대로 분석(Per Protocol analysis, PP analysis)인지를 구분한다.¹²⁾ 우선되는 것은 ITT 분석으로 이는 선정된 모든 연구대상자를 중도 탈락에 상관없이 중재가 일정 부분 적용되었으면 분석대상에 포함하는 것이다. PP 분석은 연구를 프로토콜대로 모두 완료한 대상자를 대상으로 분석하는 것을 말한다.¹²⁾ 근거기반연구에서 ITT 분석을 더 중요시하는 것은 중도탈락자가 연구 중재와 관련된 요인으로 탈락하였을 가능성을 배제하기 어렵고, 따라서 이 탈락자가 연구 결과에 미치는 영향의 정도를 파악하기 어렵기 때문이다. 그러나 본 연구결과에서 보면 많은 연구에서 초기에 연구대상자로 선택한 뒤 연구 수행 중 중도탈락자가 있으면, 분석에서는 이들을 모두 제외하고 결과를 제시하였다. 또한 연구대상자 선정과정에 대한 구체적인 기술이 없는 경우가 많아서 연구대상자 선정과 배제기준을 명확하게 파악하기 어려웠다.

무작위배정이 이루어지지 않은 연구에서는 선택비뚤림의 가능성이 크게 작용하므로 실험군과 대조군의 사전 동질성 검증은 필수적이고 매우 중요하다.¹²⁾ 일반적(인구사회학적) 특성과 임상적 특

성, 그리고 결과변수에 대한 사전 동질성 검증이 필요하나, 두 가지를 모두 시행한 연구는 62.3%이었고, 나머지는 한 가지만 실시하거나 실시하지 않은 경우도 있었다. 동질성 검증을 하였음에도 차이 검증에 대한 통계분석없이 단순히 빈도와 백분율만 제시하는 경우도 있었다.

대상자의 선정기준과 배제기준을 설정할 때 체계적인 문헌고찰 및 전문가의 의견이 반영되지 않아서 연구목적에 적합하지 않은 대상자가 선정되는 경우가 있었다. 또한 표본의 크기를 산출하는 방식에 문제가 있었다. 많은 연구에서 산출근거가 모호하거나 기술되어있지 않았다. 특히 효과크기의 경우에는 선행연구를 참고하거나 선행연구 부재 시 pilot test 등을 실시하여 그 결과를 토대로 산정하는 것이 권장되는 방법이나,¹⁵⁾ 연구자가 임의로 산정하는 경우가 대부분이었다.

측정도구의 신뢰도와 타당도는 연구 결과의 측정에 매우 중요하므로 타당도가 높은 자료수집도구의 선정이 필수적이고, 자신의 연구에서는 연구대상자가 도구개발 당시의 연구대상자와 차이가 있을 수밖에 없으므로 신뢰도를 다시 한 번 제시할 것을 권장한다.¹¹⁾ 본 연구 결과를 보면 기존의 측정도구를 수정보완하여 사용하는 경우가 많았고, 이 경우 타당도에 차이가 발생할 수 있으나, 소수의 전문가에 의한 내용타당성 검토 결과만을 제시하는 경우가 많았다. 또한 도구의 신뢰도를 제시하지 않거나, 도구의 원저자를 제시하지 않고 이를 사용한 후속연구자만을 기술하여 원저자가 누구인지 파악하기 어려운 경우도 많았다.

대상자나 측정자의 영향을 받지 않아 측정 비뚤림이 적은 것으로 알려진¹¹⁾ 생리학적 측정도구와 사회심리측정도구 등에서의 결과 유의성을 비교하여 간접적으로 대상자 비뚤림 존재 유무를 확인한 결과, 예상대로 생리학적 측정이나 관찰법에서 중재효과에 차이가 없는 경우가 사회심리학적 측정도구나 자가보고 설문지에서 보다 통계적으로 유의하게 높았다. 그러나 본 연구에서는 결과 측정 방법별로 전체적인 유의성만 분석한 것이어서, 무작위배정이나 눈가림 미실시에 의한 결과인지 확인하기는 어려웠다. 후속연구를 통하여 동일한 논문 내에서 한 결과변수를 다양한 방법으로 측정 한 경우로 제한하여 재분석할 필요가 있다. 이를 통하여 간호학 실험연구에서의 결과 측정 방법에 대한 지침을 일부라도 제시할 수 있을 것으로 생각한다.

질 평가 항목을 체크하면서 가장 아쉬웠던 것은 연구 방법에 대한 기술에 있어 부족한 사항이 많았던 점으로 기술만 정확하게 하였어도 질 평가에서 높은 점수가 나왔을 것으로 추정되는 부분들이 있었다. 예를 들어, 마사지를 중재로 적용하면서 같은 공간에서 같은 시간대에 동일한 중재자가 시행했는지에 대한 기술이 없으면, 중재자와 측정자가 구체적으로 기술되지 않았기 때문에 눈가림 여

부를 판단하기가 어려웠고, 측정자가 여러 명일 때 측정자간 신뢰도도 알 수 없었다. 대조군에 대한 처치의 내용에 대하여도 기술되어 있지 않거나, 무작위배정을 하였다고 했지만 구체적인 방법의 기술이 없는 경우들도 있었다. 이러한 경우 모두 불확실로 간주되므로 실제로 제대로 시행되었다 하더라도 질 평가에서 좋은 점수를 얻지 못하게 된다. 연구자가 논문을 투고할 때 체크리스트를 통해 설계방법에 따른 내용타당성을 확인하는 것도 논문의 질 평가를 상승시킬 수 있는 방안일 것이다.

결론 및 제언

이 연구는 그 동안 연구 현황이나 내용 분석에 집중되었던 간호학연구 분석을 근거기반실무를 위한 질 평가로 확대하였다는 데에 의의가 있다고 생각한다. 무작위배정연구 질 평가 도구를 적용하여서 부족한 점이 많이 도출된 것은 앞으로 간호학실험연구의 질 향상을 위한 제언으로 간주하는 것이 적절할 것이다. 추후 표준화된 비무작위배정연구 질 평가 도구를 이용하여 재분석하여 보면 간호학 실험연구에 대한 새로운 결과를 도출할 수도 있을 것이다.

본 연구를 수행하면서 연구 설계의 분류와 명칭이 의학이나 보건학 연구와 달라서 질 평가 도구 선정과 분석에 어려움이 있었다. 근거기반실무는 임상에서의 적용과 평가를 위주로 수행되므로 간호학 연구 설계도 근거기반실무를 고려한 방식으로 개선보완이 필요하다. 또한 여러 간호학술지 중 “임상간호연구”의 논문만을 대상으로 하였으므로 결과를 전체 간호학 실험연구에 확대하여 적용하기는 어렵다고 생각한다. 앞으로 간호학 연구의 질 평가에 대한 적절한 도구를 연구 설계별로 재탐색하거나 수정보완하여 더 많은 수의 연구를 대상으로 분석해보는 것이 필요하다.

REFERENCES

1. Sackett DL, Straus SE, Richardson WS, Rogenberg WM, Haynes RB. Evidence based medicine: How to practice and teach EBM. 2nd ed. London: Churchill Livingstone; 2000.
2. Kim SY, Shin SS, Han SK, Cho HM, Jung HJ. Study of developing quality assessment tool in experimental studies based on clinical research. Chuncheon: Health Insurance Review & Assessment Service. Hallym University Industry Academic Cooperation Foundation; 2009.
3. Fineout-Overholt E, Mazurek B, Stillwell SB, Williamson KM. Critical appraisal of the evidence: Part II. *Am J Nurs*. 2010;110(9):41-8. doi: 10.1097/01.NAJ.0000390503.92404.e6
4. Lim SM, Shin ES, Lee SH, Seo KH, Jung YM, Jang JE. Tools for assessing quality and risk of bias by levels of evidence. *J Korean Med Assess*. 2011;54:419-29. doi: 10.5124/jkma.2011.54.4.419
5. Croew M, Sheppard L. A review of critical appraisal tools show they lack rigor: alternative tool structure is proposed. *J Clin Epidemiol*. 2011;64:79-89. doi: 10.1016/j.jclinepi.2010.02.008
6. Jutel A. Beyond Evidence-Based Nursing: tool for practice. *J Nurs Man*. 2008;16:417-21. doi: 10.1111/j.1365-2834.2008.00861.x
7. Kim HJ, Kim SY. Quality assessment of non-randomized studies in the Korean Journal of Family Medicine. *J Korean Acad Fam Med*. 2009;30: 129-37.
8. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of healthcare interventions. *J Epidemiol Community Health*. 1998;52:377-84.
9. Slim K, Nini E, Forestier D, Kwiatkowski F, Panis Y, Chipponi J. Methodological index for non-randomized studies (MINORS): development and validation of a new instrument. *ANZ J Surg*. 2003;73:712-6.
10. Shin HS, Hyun MS, Ku MO, Cho MO, Kim SY, Jeong JS et al. Analysis of research papers published in the Journal of the Korean Academy of Nursing-focused on research trends, intervention studies, and level of evidence in the research. *J Korean Acad Nurs*. 2010;40:139-49. doi: 10.4040/jkan.2010.40.1.139
11. Lee EO, Lim NY, Park HA, Lee IS, Kim JI, Pai JI et al. Nursing research and statistical analysis. Seoul: Soomoonsa; 2009.
12. Park BJ. Evidence-based healthcare. Seoul: Korea Medical Book Publisher; 2009.
13. Viera AJ, Bangdiwala SM. Eliminating bias in randomized controlled trials: importance of allocation concealment and masking. *Fam Med*. 2007;39(2):132-7. doi: 10.1111.1111.417
14. Choe MA, Jeong JS, Lim KC, Kim JH, Kim KS, Kwon JS et al. Research trend of studies published in Journal of Korean Clinical Nursing Research, 1995-2008. *J Korean Clin Nurs Res*. 2011;16:95-105.
15. Browner WS, Newman TB, Hulley SB. Getting ready to estimate sample size: hypotheses and underlying principles. In: Hulley SB, Cummings SR, Browner WS, Grady DGG, Newman TB, editors. Designing clinical research. PA: Lippincott-Williams and Wilkins; 2007. p. 56.