

정신과 연구에서 다양한 임상연구방법의 장단점

연세대학교 의과대학 정신과학교실,¹ 연세대학교 의과대학 의학행동과학연구소²

하 라 연^{1,2} · 조 현 상^{1,2}

Pros and Cons of Various Research Designs in Clinical Psychiatry

Ra Yeon Ha, MD,^{1,2} Hyun-Sang Cho, MD^{1,2}

¹Department of Psychiatry, Yonsei University College of Medicine, Seoul, Korea

²Institute of Behavioral Science in Medicine, Yonsei University College of Medicine, Seoul, Korea

An appropriate research design for hypotheses and purposes leads to a good quality of research results. In this review article, we summarized the types of research methods and described the characteristics of clinical trials. Research designs are categorized into observational studies and experimental ones, depending on data collecting methods. In an observational study, there are cross-sectional, cohort and case-control studies. Parallel groups design and crossover trial studies are representative designs in a randomized controlled trial study, a kind of experimental study. Clinical researchers should understand the characteristics of clinical research designs including advantages and disadvantages and choose the suitable design according to their study purposes and the nature of collected data or subjects.

Key Words Research design · Observational study · Experimental study · Clinical trials.

Received: July 9, 2012 / Revised: August 2, 2012 / Accepted: August 21, 2012

Address for correspondence: Hyun-Sang Cho, MD

Department of Psychiatry, Yonsei University College of Medicine, 50 Yonsei-ro, Seodaemun-gu, Seoul 120-752, Korea

Tel: +82-31-760-9403, Fax: +82-31-761-7582, E-mail: chs0225@yuhs.ac

서 론

일반적으로 연구를 수행하기 전에 중요한 과정이 연구계획의 수립이다. 특히 연구계획 수립은 연구를 위한 문제 제기 단계로서 가장 중요하며, 이 단계에서 쟁점이 명확히 제시되고 이 쟁점에 대해 연구과제가 정립되고 가설이 형성된다. 그 다음 단계로서 연구설계는 대상 및 표본 선정, 자료의 성격과 수집, 측정방법, 자료분석의 틀을 제시하게 된다. 이 연구설계는 다양한 연구방법을 토대로 한다.¹⁾

가장 간단한 연구방법 분류로는 자료 수집 혹은 실험 유무 방법에 따라 관찰연구와 실험연구가 있다. 관찰연구(observational study)는 자연발생적인 사건에 대한 관찰로 특정 개입(intervention)이 없다. 이 연구는 기술적인(descriptive) 연구뿐만 아니라 가설검증적인 연구에도 이용된다. 실험연구(experiment study)는 무작위 대조 설계로서 위험요인을 인위

적으로 투여(혹은 개입)하여 그 영향을 비교하는 방법이다. 관찰연구는 실험연구와 비교하여 비용투입이 적고 다양한 범주의 환자들을 포함시킬 수 있는 장점이 있다.²⁾ 또한 관찰연구는 위험요인과 예후지표를 파악하는데 일차적으로 적용되며, 무작위-대조 시험이 불가능하거나 그 자체가 비윤리적인 경우 사용되는 연구설계이다.³⁾ 또한 시간 혹은 시점에 따른 연구설계 방법으로 횡단면(cross-sectional) 연구와 추적(longitudinal) 연구가 있고, 연구진행 방향에 따라 후향적, 현황적, 전향적 연구로 나눌 수 있는데, 횡단면 연구는 현황적 연구이고 코호트 연구나 임상시험의 경우 전향적 연구가 대부분이다.

본 종설에서는 일반적 연구방법과 관련된 디자인을 간단히 설명하고, 정신과 영역에서의 임상연구 디자인의 장단점에 대해 기술하고자 한다.

본 론

관찰연구 : 역학적 연구 디자인

횡단면 연구(Cross-sectional study)

위험인자에 대한 노출과 질병발생이 거의 동시에 이뤄지는 경우에 적용되는 연구설계라고 할 수 있다. 일반적으로 특정 시점에서 모집단에서 추출된 표본집단으로 구성되는데, 짧은 기간 내에 필요한 결과를 얻을 수 있고 동시에 많은 종류의 질병과 위험요인의 관련성을 이해할 수 있어 비용이나 시간 등과 같은 현실요인에 제한을 덜 받는 장점이 있기 때문에 실제 임상연구에서 많이 적용된다. 그러나 위험요인과 질병 사이의 시간의 선행관계에 파악이 어려우므로 인과관계의 정확한 구별이 어렵고 유병률이 낮은 질병은 수행이 어려울 수도 있다. 또한 표본집단이 대표성을 갖기 위해서는 대상자 수가 충분히 커야 하고, 여러 질병 위험요인 중 특정 요인만을 찾기가 어려운 경우가 있다. 이 횡단면 연구의 예로서 Woo 등⁴⁾은 직장인에서 수면 행태 및 특성에 따른 생산성 손실을 파악하기 위해 일정 기간 동안 특정 지역의 직장인 700여 명을 대상으로 수면 및 업무 성과 관련 설문을 시행하여 수면의 질이 좋은 집단과 나쁜 집단의 생산성 변인들을 비교 분석하였다.

코호트 연구(Cohort study)

특정 질환이 발생하기 전에 연구 대상집단인 코호트에 대해 위험하다고 의심되는 요인에 대한 노출 여부에 대해 미리 조사해 놓고, 장기적으로 이 집단을 관찰한 후, 발생하는 질환과 의심되는 요인과의 상관성을 조사하는 설계방법이다. 질병 발생과 요인과의 상관성은 보통 비교위험도(relative risk)로 제시된다.

이 코호트 연구는 위험요인에 대한 노출에서부터 이후 질병의 진행 과정이 관찰 가능할 뿐만 아니라, 연구하고 싶은 요인을 연구자 의도에 따라 포함시킬 수 있다. 또한 원인과 결과의 전후 결과 해석이 가능해져서 발생률과 비교위험도 측정이 가능해진다. 단점으로는 장기간이 소요되는데, 추적 탈락자가 많아지면 연구 자체가 실패할 가능성이 높아지고 발생이 아주 드문 희귀 질환의 경우에는 이 연구설계를 사용하기 곤란하다.

Fisher 등⁵⁾은 아동기 집단 따돌림이 청소년기 자해 가능성을 높이는지 그리고 어느 집단 따돌림 아동이 자해 위험이 높은지 조사하기 위해 영국에서 1994년에서 1995년 사이에 태어난 2200여 명의 아동을 5세, 7세, 10세, 12세에 평가하여 12세 생일 직전 6개월동안 자해 행위의 비교위험도를 계산하였는데, 이 연구설계가 코호트 연구의 예라고 할 수 있다.

환자-대조군 연구(Case-control study)

현재 특정질병을 갖고 있는 환자군과 그 질병에 이환되지 않은 대조군을 설정하여 이 질병의 원인이라고 의심되는 요인에 대해 과거에 노출되었던 비율을 비교하여 특정 질환과 위험요인의 관련성을 연구하는 설계이다.

코호트 연구에 비해 필요한 연구 대상 수가 비교적 적고 투입되는 시간과 경비도 적게 들어 상대적으로 연구를 단기간 내에 수행할 수 있다. 또한 기존 자료를 활용할 수 있으며, 희귀 질환이나 긴 잠복기를 가진 질환도 연구가 가능하다. 그러나 수집할 수 있는 위험요인의 수가 제한되고 노출 요인에 대한 정보 수집도 개인의 기억이나 회상 능력에 의존할 수 밖에 없어 왜곡이나 비뚤림(bias)의 가능성이 있어 신뢰도에 문제가 있을 수 있다. 또한 시간적인 선행관계를 파악하기 어려워 인과관계의 설정이 곤란하다. 적합한 대조군 설정이 쉽지 않고 선정되더라도 오류의 개입 소지가 높으며, 특정 요인에 대한 위험도를 추정하는 것이 어렵다.

Zhang 등⁶⁾은 조현병 환자들이 일반 인구에 비해 흡연율이 높는데 이 같은 현상이 인지기능 저하를 완화시키기 위한 자가 치료의 한 형태인지 알아보기 위해 580명의 남자 중국인 조현병 환자와 나이, 교육 수준을 고려한 175명의 남자 정상 대조군을 대상으로 실험을 진행하였다. 이 연구설계가 환자-대조군 연구의 예이다.

상기의 관찰연구는 특정 개입이 없는 자연발생적인 측면이 강조되어 있다. 그러나 혼란변수(confounding factor)라는 효과로 연관성이 허위로 도출되거나 실제 연관성이 관찰되지 못할 수도 있다. 즉 특정 상태의 결과와 잠재적인 위험요인 사이에 찾을 수 있는 관련성에 대한 정확한 인과관계 결론을 내리기 힘들어진다.

실험연구(Experimental study)

실험연구는 특정요인을 인위적으로 조작, 통제하고 개입시켜 그 요인에 대한 영향을 알아보는 방법이다. 실험연구로서 연구 디자인과 방법론은 특정 약물이나 치료법의 개발과 발전을 위한 연구에 결정적으로 중요하다고 할 수 있다. 특히 새로운 치료법이나 개입의 효과를 평가하는 것을 목적으로 하는 연구를 시험(trial)이라고 하는데, 연구 대상이 의학적 조건에 있는 사람이라면 임상시험이라고 한다. 임상연구와 관련하여 다양한 이슈들이 있는데, 대조연구의 윤리적 문제, 자연연구 및 관찰연구의 적용 가능성, 효능 및 효과 연구의 장단점 등이 있다.

치료개입 효과를 조사하는 연구 디자인으로 2개의 일반적인 방식이 있는데 관찰연구와 무작위 대조시험이다. 관찰연구로는 사례연구, 사례 시리즈, 개방연구 등이 있다. 그러나 일반적으로 임상시험(clinical trial)이라 함은 무작위 대조시험(rando-

mized controlled trial) 혹은 최소한 대조시험을 의미하는 경우가 많다.

임상시험에서의 대표적 연구설계

어느 한 치료법이 다른 치료법보다 더 나은지 알기 위해서는 목적에 따른 각기 다른 연구 디자인이 필요하다. 여기서는 실제 많이 사용되는 임상시험 디자인을 기술할 것이다.

병행집단(Parallel groups) 연구

이 연구설계 방법은 임상시험에서 가장 많이 적용된다. 가장 단순하게는 대상 환자들을 두 치료집단 가운데 어느 한 집단에 무작위(randomization)로 배정하는 것이다. 즉, 무작위로 형성된 집단이 나란히 연구시험을 진행하는데 이 병행 설계시험의 목적은 각 참여자가 각각 배정된 치료를 배타적으로 받아 그 효과나 결과를 비교하는 것이다. 병행집단 설계는 연구 수행이 가장 용이하고 분석과 해석도 수월하여 정신과를 포함한 모든 의학영역에서 가장 간단하고 제일 흔히 적용되는 설계이다. 그러나 다른 연구방법과 비교하여 상대적으로 많은 참여자를 필요로 하여 시간과 비용이 더 증가될 수 있다. 병행집단연구는 시험군과 대조군 두 집단으로 나누는 것이 가장 기본 형태라고 할 수 있다.

실제 연구 예로 Ahn 등⁷⁾은 약물선택 왜곡을 피하기 위해 조현병 피험자들을 리스페리돈 군 혹은 지프라시돈 군으로 무작위 배정한 후 지프라시돈의 체중 증가와 대사성 부작용에 대한 효과를 연구하였다.

교차시험(Crossover trial)

교차시험에서는 일정 간격의 특정한 순서로 2개 이상의 연구 처치 혹은 치료를 받게 된다. 예를 들어 두 시기(two-period) 교차시험 설계에서는 각 참여자가 A 혹은 B 각각 두 치료를 받게 되는데 그 순서가 AB가 될지 BA가 될지는 무작위로 결정된다. 따라서 이 임상시험은 두 군으로 구성되어 AB 또는 BA 순서로 받는다. 교차시험은 피험자 내(within subjects) 비교로 이뤄지지만 병행 설계는 참가자 간(between subjects) 비교가 이뤄지게 된다. 이 설계는 한 치료만으로는 질병이 치료되거나 질병 양상이 달라지기 어렵고 주기적인 변화가 없는, 안정적이고 만성적인 질환에서 단기간 치료요법에 대한 반응을 보는 연구에 적합하다. 또한 교차시험에서는 각 피험자가 동시에 대조군으로 활용되기 때문에 병행설계에 필요한 피험자보다 적은 수도도 통계적으로나 임상적으로 타당한 결과를 얻을 수 있다. 단점으로는 두 치료의 간섭에 의한 이월 혹은 잔류 효과(carryover effect)의 가능성인데 이를 최소화하기 위해 두 치료 사이에 일정한 약효세척 기간을 항상 둔다. 그러나

이 기간이 효과적이지는 확실치 않고 이 간격을 설정하기도 쉽지 않다.⁸⁾

무작위 교차설계는 가장 타당성이 높은 실험설계라고 할 수 있다. 이 설계를 이용한 예를 들면, Heresco-Levy 등⁹⁾은 글라이신 혹은 위약을 2주의 약효세척 기간을 두고 6주동안 투여하여 글라이신의 효과를 관찰하였다. 또 Zarate 등¹⁰⁾은 치료 저항성 주요우울 환자에서 무작위 이중맹검 교차 설계로 케타민 혹은 위약을 일주일 간격으로 투여한 연구를 발표하였다.

시험연구의 종류 : Efficacy vs. Effectiveness trials

무작위 대조시험(randomized controlled trials, 이하 RCT)은 근거중심의학(evidence-based medicine)의 근간으로서 임상 결정이나 임상 실체에 중요한 영향을 미치고 있다. 임상시험은 사례보고나 비대조 사례연구들보다 치료효과를 더 엄밀하고 정확하게 평가할 수 있게 한다. 그러나 RCT와 같은 임상시험은 실행가능성이나 일반화 측면에서 중요한 제한이 있을 뿐만 아니라 위약 사용과 같은 윤리적 이슈들이 제기되고 있다. RCT의 강점과 약점에 대해서는 Table 1에 기술하였다.¹¹⁾ 이 내용들은 연구의 특정 목적에 따라 고려되고 수정되어야 할 것이다. 가장 중요한 결정 요소는 환자나 치료의 현실적인 실제 특성을 어느 정도로 제한시키고 표준화하느냐 하는 것이다. 예를 들어 한 효능(efficacy) 연구에서는 모집된 200명의 환자 중에 시험 거절, 이전 치료에서 불응 과거력, 발병 시기, 나이 등을 이유로 최종적으로는 10~20% 정도인 27명만이 피험자가 될 정도로 명확하게 정의된 협의의 환자들만이 포함되었으며, 일상적인 임상실제에서는 사용되지 않을 광범위한 평가도구들이 이용되었다.¹²⁾ 반면 효과(effectiveness) 연구는 폭넓은 환자군에 적용함으로써 연구 결과들을 실제 환자들에게로 일반화가 손쉬우며, 추적중단, 입원, 복약충실도, 삶의 질과 같은 좀 더 객관적이고 실질적이면서 쉽게 적용 가능한 영역을 측정하는 결과 도구들을 이용하여 평가한다. 조현병 환자에서 체중 감량 프로그램의 효과를 조사하는 한 효과 연구는 146명의 환자를 대상으로 하여 이 중 과체중인 109명의 환자에게 16주간 평균 6.7회기의 치료 프로그램을 적용하며 시험 시작시, 1년 후, 그리고 각 치료 회기시에 체중과 체질량지수(Body Mass Index)를 측정하여 그 결과를 정상 대조군과 비교하는 방법으로 시험하였다.¹³⁾ 효능과 효과 연구 각각의 특성에 대해서는 Table 2에 나타내었다.¹⁴⁾

결론

본 종설은 일반적 연구 방법들과 정신과 영역에서의 임상연구 방법의 장단점에 대해 개괄하였다. 일반적 연구 방법은 자료

수집 방법에 따라 관찰연구와 실험연구로 분류되고, 관찰연구에는 횡단면 연구, 코호트 연구, 환자-대조군 연구 등이 있고 실험연구는 특히 임상시험에서 많이 쓰이는 디자인인 무작

위 대조시험인 병행집단 연구와 교차시험이 대표적이다. 따라서 임상 연구자는 다양한 관찰연구와 실험연구, 효능 혹은 효과 연구의 특성과 장단점을 파악하여 연구 목적과 자신이 처

Table 1. Strengths and weaknesses in randomized controlled trials

Selected trial characteristics	Strengths	Weaknesses
Restricted inclusion criteria	High internal validity	Limited external validity/generalizability
	Specific signal detection capacity in carefully selected target population	Difficulty assessing optimal dosing in unrestricted, more heterogeneous or seriously ill sample
Randomization	Usable for regulatory and registration purposes	Decreased knowledge about response and side effect patterns in patients with comorbid psychiatric and/or medical conditions
	Controlling for measured and, especially, unmeasured group differences	Slow enrolment
Small-to-medium sized samples	Challenges associated with placebo controls, maintaining blind in the face of specific adverse effects	Selection bias towards a less generalizable sample (less severely ill, more chronically ill patients ; patients with prior stabilization or treatment phase)
	More homogeneous and carefully characterized samples	Reduced generalizability
Specialized settings	Greater potential for careful selection and diagnostic/assessment	Need for multiple research-oriented sites
	More control over study procedures	Low signal detection for rare outcomes
Frequent visits	Greater comfort using placebo controls	Reduced generalizability
	Study conduct by well-trained personnel with allocated research time	Fewer potential sites
Comprehensive assessments	Controlled treatment and assessment	Greater likelihood of professional patients
	Better assurance of patient safety	Lower enrolment rates
Cost	More systematic quantitative assessments for therapeutic and adverse effects	Potentially less access to patients of interests (eg, acute exacerbations, drug-naïve)
	Increased opportunity to facilitate/monitor adherence	Reduced generalizability
Comprehensive assessments	Greater ability to use informative laboratory tests	Reduced enrolment
	Better assessment of use of ancillary services or use of other medications	Increased burden and dropouts
Comprehensive assessments	Specific assessment of measurable outcomes (including safety and tolerability) using validated and reliable scales administered by well-trained personnel	Frequent quantitative assessments not done in clinical practice
	Usable for regulatory and registration purposes	Potential influence on specific, investigated effect by increased contact
Comprehensive assessments	Usable for potential marketing	Primary/secondary outcomes rarely assessed in clinical practice
		Patient/caregiver rated outcomes, quality of life and functional capacity rarely assessed in clinical practice
Comprehensive assessments		Use of quantitative measures unlikely in clinical practice and clinicians not trained in their use
		Need for careful training and ongoing supervision of raters
Comprehensive assessments		Increased burden and dropouts
		Increased per-patient costs

Table 2. Differential characteristics of clinical trials of “efficacy” vs. “effectiveness”

Clinical trials of “efficacy”	Clinical trials of “effectiveness”
Highly restricted inclusion criteria to reduce confounding biases	More relaxed exclusion criteria, permitting wider range of : Patients (eg. comorbidity not excluded)
Randomization and blinding, also to reduce bias	Treatment settings and interventions (including adjunctive treatments)
Treatment driven exclusively by study protocol	Emphasis on clinical need to determine treatment doses, etc
Patients remain only in the treatment group originally assigned	Levels and/or type of psychopathology
Fewer treatment adjustments are allowed	Forms of outcome criteria, such as :
Strict limitations on adjunctive treatment	Time to discontinuation
Measures taken to insure all members of treatment group receive same intervention (s)	Quality of life
	Preference of self-rating instruments or global ratings
Use of well-validated outcome assessment	
Advantages :	Advantages :
Higher internal validity for clinical effects	Higher external validity
Higher internal validity for adverse effects, tolerability	Arguably greater applicability to “real-world” practice settings
Contextual and human factors controlled for	Capacity to inform policy process
Considered “best quality” clinical evidence for informing treatment decisions	Longer duration can be easier achieved
	Can enrol large number of patients more easily
Disadvantages :	Disadvantages :
Stringent inclusion criteria limit external validity	Internal validity limited
Outcome measures may not reflect crucial advantages and limitations of Interventions being studied	Cannot be used to examine effective dose ranges
Outcome measures may not address issues most important to patients and families	Cannot make as meaningful clinical comparisons between agents
Often short in duration	

해 있는 현실과 자료 특성에 맞춘 연구 디자인이 필요하리라 생각된다.

중심 단어 : 연구디자인 · 관찰연구 · 실험연구 · 임상시험.

Acknowledgments

본 내용은 2011년 7월1일 개최된 생물정신의학회 제3차 연수강좌에서 발표되었음.

Conflicts of interest

The authors have no financial conflicts of interest.

REFERENCES

- 1) Song K, Han M, Cheong MY, Lim KS, Kim DK. Clinical research design and biostatistical methods. Korean J Urol 2005;46:835-841.
- 2) Feinstein AR. Epidemiologic analyses of causation: the unlearned scientific lessons of randomized trials. J Clin Epidemiol 1989;42:481-489; discussion 499-502.
- 3) Naylor CD, Guyatt GH. Users’ guides to the medical literature. X. How to use an article reporting variations in the outcomes of health services. The Evidence-Based Medicine Working Group. JAMA 1996;275:554-558.
- 4) Woo JM, Hyun SY, Lee SH, Kang SG, Lee JS, Kim L, et al. Productivity time lost by sleep disturbance among workers in Korea. J Korean Neuropsychiatr Assoc 2011;50:62-68.
- 5) Fisher HL, Moffitt TE, Houts RM, Belsky DW, Arseneault L, Caspi A. Bullying victimisation and risk of self harm in early adolescence:

- longitudinal cohort study. BMJ 2012;344:e2683.
- 6) Zhang XY, Chen da C, Xiu MH, Haile CN, Sun H, Lu L, et al. Cigarette smoking and cognitive function in Chinese male schizophrenia: a case-control study. PLoS One 2012;7:e36563.
- 7) Ahn YM, Lee KY, Kim KS, Kim DK, Kim YH, Kim CY, et al. Body weight gain and metabolic changes in patients with schizophrenia during 12-week randomized treatment of ziprasidone and risperidone. Korean J Psychopharmacol 2007;18:92-102.
- 8) Senn S. Cross-over Trials in Clinical Research. West Sussex: Wiley; 2002.
- 9) Heresco-Levy U, Javitt DC, Ermilov M, Mordel C, Silipo G, Lichtenstein M. Efficacy of high-dose glycine in the treatment of enduring negative symptoms of schizophrenia. Arch Gen Psychiatry 1999;56: 29-36.
- 10) Zarate CA Jr, Singh JB, Carlson PJ, Brutsche NE, Ameli R, Luckenbaugh DA, et al. A randomized trial of an N-methyl-D-aspartate antagonist in treatment-resistant major depression. Arch Gen Psychiatry 2006;63:856-864.
- 11) Correll CU, Kishimoto T, Kane JM. Randomized controlled trials in schizophrenia: opportunities, limitations, and trial design alternatives. Dialogues Clin Neurosci 2011;13:155-172.
- 12) Hofer A, Hummer M, Huber R, Kurz M, Walch T, Fleischhacker WW. Selection bias in clinical trials with antipsychotics. J Clin Psychopharmacol 2000;20:699-702.
- 13) Niv N, Cohen AN, Hamilton A, Reist C, Young AS. Effectiveness of a Psychosocial Weight Management Program for Individuals with Schizophrenia. J Behav Health Serv Res 2012. [Epub ahead of print]
- 14) Möller HJ. Effectiveness studies: advantages and disadvantages. Dialogues Clin Neurosci 2011;13:199-207.