

자료의 해석 및 논문작성

아주의과대학 예방의학교실 이 순 영

Skepticism is the chastity of the intellect

Don't give it away to the first attractive hypothesis that comes along.

M. B. Gregg

I. 자료의 해석

A. 자료의 질적 관리(Quality Control)

1. 연구설계의 중요 포인트

- 연구의 가설을 명확히 결정하였는가?
관련된 변수들을 구체화하였는가?(hypothesis)
- 혼란변수 등을 규명하였는가?(counfounding variables, effect modifiers)
- 연구집단을 특성을 정의하였는가?(external validity)
- 관련성을 밝히기 위한 통계력을 결정하였는가?(statistical power)
- 연구과정을 표준화 하였는가?(manual)
- 연구자들의 훈련과 자격(staffing)
- 질적 관리(quality control)

2. 질적 관리

cf) Quality assurance vs Quality control

- ① Quality assurance : 자료수집전 질적인 자료수집을 위한 과정, 활동(프로토 콜, 지침서)
- ② Quality control : 자료를 수집하고 분석하는 과정에서 질적인 자료수집을 위한 과정, 활동

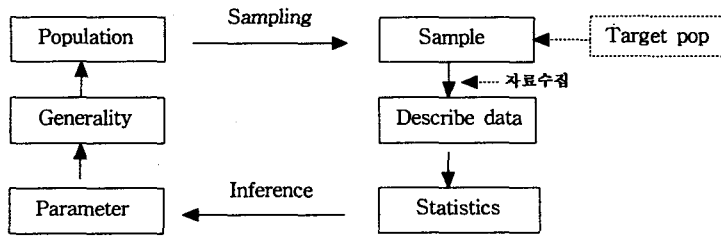
가. 종류

- 1) 관찰모니터링
- 2) 양적 평가 : blind fashion
 - 무작위 반복 측정(intra-observer, inter-observer)
 - quality control pools
 - 관찰자의 측정오차 모니터링

나. 타당도와 신뢰도 분석

- 1) 타당도: 민감도와 특이도로 추정(test vs gold rule, inter-method reliability)
 - 내적타당도 : target population에 대한 추론
 - 외적타당도 : 일반인구에 대한 추론

cf) study pop - actual pop - target pop - external pop



2) 신뢰도

- 지표 : percent agreement/kappa/correlation coefficient/reliability coefficient/coefficient of variation/percentage difference

cf) Regression toward the mean

: 같은 표본을 반복측정할 때 첫번째 측정이 반복측정과 무관하더라도 음적 관계가 있는 것처럼 보임

B. 통계적 관련성과 인과관계

: 역학적 연구의 기초 - 어떤 현상은 우연히 무작위로 발생하는 것이 아님

1. 통계적 연관성 평가

구 분	지 표	정 의	특 성
연관성 측정 (measure of association)	OR, RR	관계의 방향성과 강도의 양적 표현	표본의 수와는 무관함 관계의 신뢰성을 반영하지는 않음
유의성 검증 (test of significance)	CI p-value	관찰된 관계와 영가설과의 차이의 우연성 정도	표본의 수 및 연관성의 크기와 연관됨

2. 통계적 유의성 검증결과의 해석

가. 통계적으로 의미가 없을 때

“통계적으로 유의하지 않다(no significant)”고 반드시 “관련성이 없다(no association)”는 것을 의미하지는 않음

→ 1) 실제로 조사인구집단내 관계성이 미존재

2) 조사집단의 수가 실제의 관계를 찾아낼 만큼 크지 못함(표본수가 너무 적음)

나. 통계적으로 유의할 때

“그 자체가 원인-결과의 관계를 말하지는 않는다”

cf) p value와 신뢰구간

① p값 : 연구자가 용납할 수 있는 1종 오류(α)

② 신뢰구간 : 연관성의 통계적 분산(statistical variability of association)

95%의 신뢰구간의 의미 : - 대상인구 집단에서 반복표본을 추출할 때 참값을 얻을 확률

- 그 구간내 참값을 포함할 확률이 95%, 즉 참값이 그 범위에 포함될 확률이 95%

$$- CI = 100 * (1 - \alpha)$$

· 신뢰구간 해석 : 예) RR = 4.0 CI = 3.0 - 5.3

· 구간의 크기는 정확도(precision)을 의미함

- 구간의 넓을 때 : 분산이 크고 정확도가 떨어짐

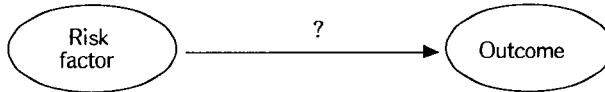
- 구간이 좁을 때 : 분산이 작고 정확도가 큼

· 표본수와의 관계

$$\ln(\theta) \text{의 CI} = \ln(OR) \pm Z_{1-\alpha/2} \times \sigma_{\ln(OR)}$$

$$\sigma_{\ln(OR)} = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}$$

3. 요인과 결과의 관련성의 가능성



1) Chance(Random error)

2) Bias(Systematic error)

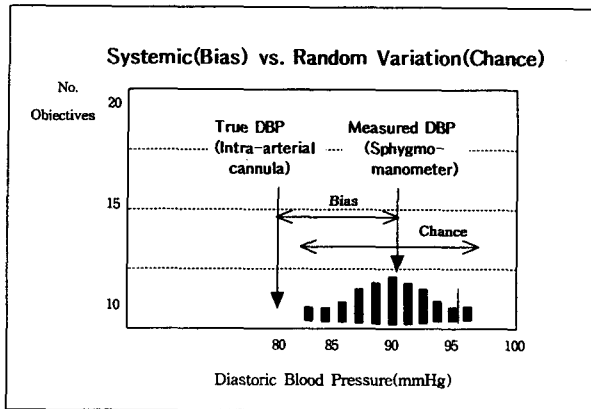
3) Confounding(By third factor)/ Effect modifier

4) Causal association

4. Bias(오류)

가. 정의

a process at any stage of inference tending to produce results that depart systematically from the true values



<systemic(bias) vs random variation(chance)>

나. 원인(source)

- 연구전 : 문헌고찰, 연구설계
- 연구진행과정 : 표본추출
- 연구진행후 : 분석, 논문작성

다. 선택오류

① Admission rate(Berksonian) bias

Disease of bone	Respiratory ds (yes)	Respiratory ds (no)
Exposed	↑↑a	b
Unexposed	c	d

$OR = (a/c)/(b/d) = ad/bc$

↑↑ ↑↑

→ nested / population based case-control study

② Detection bias

	Tumor Registry		D&C Registry	
	Case	Conrtol	Case	Conrtol
Estrogen (+)	↑ 45	7	↑ 59	↑ 42
Estrogen (-)	72	110	89	106

$OR = 9.8$

$OR = 1.7$

↑

↔

(compensatory bias, nondifferential)

→ matching on diagnostic procedure

③ Non-respondent bias(volunteer bias)

	Case	Control
Exposed	45	45
Unexposed	45	↓ 20

Response 90% 65%

$OR = (45/45)/(45/20) = 0.44$

↓

↓

라. 정보오류(information bias)

: 질병, 노출, 혼란변수 등에 대한 불완전한 정보에 의한 것으로 조사자, 조사대상자, 측정도구가 원인임

① Diagnostic suspicion bias

: 노출에 대한 사전정보가 노출의 결과를 찾는데 영향을 줌

② Exposure suspicion bias(interviewer bias)

	Case	Control
Exposed	↓a	b
Unexposed	↓c	d

$$OR = (a/c)/(b/d) = ad/bc$$

↑↑ ↑↓

③ recall bias

④ family information bias

5. 혼란변수(Confounding)

가. 정의(Criteria)

- ① 어떤 요인이 노출과 질병 둘 다 연관되어 있음(즉 위험요인과 어떤요인이 서로 상관관계가 있으면서 어떤 요인 자체도 결과에 영향을 미친다)
- ② 어떤 요인의 각층(strata)의 RO(specific RO)가 crude RO와 다름(Simpson's paradox).

예) 커피섭취와 혈압과의 관계

비만도	혈압	자주마심(%)	자주마시지 않음(%)
	고혈압	22(31)	64(49)
	정상	48(69)	66(51)
비만	고혈압	17(85)	63(79)
	정상	3(15)	17(21)
정상	고혈압	5(10)	1(2)
	정상	45(90)	49(98)
계		70(100)	130(100)

나. 종류

종류	crude RR/RO	adj. RR/RO
positive	3.5	2.1
negative	1.5	3.2

* Confounding is not an "all or none" phenomenon.

다. 교정방법

- 1) 자료수집전 : 무작위 선출/ 개인 짝짓기/ 빈도 짝짓기
- 2) 자료수집후 : 직접/ 간접/ Mantel-Haenszel/ Regression technique

예) ① A형간염과 우유 및 도너스 섭취

우유	환자	대조군	계	
먹음	37	21	58	Odd ratio = 3.9 Yates-corrected $\chi^2 = 9.24$ P value = 0.0002
먹지않음	13	29	42	
계	50	50	100	

도너스	환자	대조군	계	
먹음	40	20	60	Odd ratio = 6.0 Yates-corrected $\chi^2 = 15.04$ P value = 0.0001
먹지않음	10	30	40	
계	50	50	100	

② A형간염과 우유 및 도너스 섭취(우유섭취로 층화)

		우유 마심		우유 안마심	
		환자군	대조군	환자군	대조군
도너스	먹음	36	18	4	2
	안먹음	1	3	9	27
		Odds ratio = 6.0		Odds ratio = 6.0	

③ A형간염과 우유 및 도너스 섭취(도너스로 층화)

		도너스 먹음		도너스 안먹음	
		환자군	대조군	환자군	대조군
우유	마심	36	18	1	3
	안마심	4	2	9	27
		Odds ratio = 1.0		Odds ratio = 1.0	

④ A형간염과 우유 및 도너스 섭취, 2×4표

도너스	우유	환자군	대조군	ODDS RATIO
Yes	Yes	36	18	6.0
No	Yes	1	3	1.0
Yes	No	4	2	6.0
No	No	9	27	1.0(Ref)

6. Effect modifier

① 소집단간 노출과 결과사이의 상관정도가 다름.

(예: 연령이 effect modifier로 작용하는 경우 - 홍역예방주사의 효과,

Tetracycline이 치아에 미치는 영향)

- ② 혼란변수는 자료분석상의 artifact로 흔하게 발생되는 경우에 반하여, Effect modifier는 생물학적 현상을 나타내는 것으로 흔치않은 결과임

7. 인과관계 평가기준

가. 주요 기준

- 1) 통계적 연관성의 강도(Strength of association)
- 2) 시간적 속발성(Temporal sequence)
- 3) 용량-반응관계(Dose-response)
- 4) 결과의 일관성(Consistency)
- 5) 생물학적 가능성(Biologic plausibility)

나. 보조 기준

- 1) 특이성(Specificity)
- 2) 유추(Analogy)

II. 논문작성법

1. 논문구성

- 1) 논문요약 국문초록
영문초록(English abstract) : 참고문헌 뒤
- 2) 차례 본문차례 1. 가. (1) (가) ① ㉞
- 3) 서론
- 4) 연구목적
- 5) 연구방법
- 6) 연구결과
- 7) 고찰
- 8) 결론
- 9) 참고문헌
- 10) 부록(필요시)

2. 각 부분별 기술지침

- 1) 초록
 - (1) 연구의 기본목표와 범위
 - (2) 사용, 적용한 연구재료와 방법

- (3) 연구결과의 요약
- (4) 근간이 되는 연구결론
- (5) 중심 또는 색인 단어(key word) 선정 : 논문내용의 중심이 되는 단어를 6개 이내로 선정하여 초록 끝에 제시(국문,영문 각각).

2) 차례

논문 본문목차와 표 및 그림의 제목들을 각각 순서대로 나열

3) 서론

서론에서는 다음과 같은 내용이나 사항이 객관적, 논리적으로 언급, 설명

- (1) 연구과제의 의의와 중요성
- (2) 관련 지견의 분석(문헌고찰)
- (3) 구체적 연구주제
- (4) 가설 등의 명시

4) 연구목적

가) 궁극적 연구목적

최종적으로 도달하고자 하는 목적으로 대개 개념적 또는 추상적인 내용

나) 실행목표(또는 논리적 귀결, 명제)

본 연구를 통하여 구체적으로 관찰 또는 분석해 보고자 하는 실행적 연구목표로서 실측가능한 상태로 유도

5) 연구재료 및 방법

가) 연구의 궁극적 대상인 이론적 모집단의 특성을 규정

나) 연구대상에의 포함 또는 제외하는 기준

다) 개개의 대상 선정방법을 명시

라) 표본추출에 의하여 재료를 선정한 경우는 추출방법과 표본규모 결정방법을 제시

마) 대상의 일반적 특성을 제시

바) 연구설계, 수행과정 등을 쉽게 알아 볼 수 있는 적절한 도식(diagram) 즉 연구설계를 제시

6) 연구결과

가) 연구결론의 도출이나 유도에 직접적으로 관련되는 성적만을 제시함

나) 부정적 연구성적이라 할지라도 연구결론에 직접적인 연관이 있다고 판단되면 제시

다) 표(table) 작성

- (1) 표는 보충설명이나 본문의 내용을 읽지 않고도 표자체로서 제시하고자 하는 내용이 전달될수 있도록 작성
- (2) 맨 윗줄은 2줄로 함
- (3) 종선(세로줄) 사용은 되도록 피함
- (4) 표에 포함되게 되는 구성항목(위 예시에서 age group, control group, patient group) 중 그

내용이 비교적 동질적(homogenous)이고 합계가 의미를 지닌다고 판단되는 항목을 좌단에 배치(위의 예시에서 control 과 patient group의 합계의 표시가 특별한 의미를 주지 못함).

(5) 개개 구성항목의 범주(categories)를 구분할 때는 되도록 표준분류법에 따르도록 함 표준분류법이 없는 경우는 되도록 다른 연구자의 연구내용과 상호비교가 될 수 있도록 고려하여 분류함(예: age group의 경우 5-9, 10-14, 또는 10-19등으로 구분함 6-10, 11-15 또는 11-20등으로 구분하지 않음).

(6) 표 각주를 사용하는 경우는 개개에 대하여 "Note"로 표시하고 해설

(7) 표 작성에서 저자가 임의로 사용한 약자 또는 표기는 각주로 설명하며 표작성에 사용한 기호(예:-,+,등)에 대하여도 명확한 설명을 부기

cf) 연구가설과 백분율의 표현방법

백분율의 표현은 연구가설을 탐색적으로 조사하는데 중요

백분율←분모의 개념 : 표본선정시 고정된 집단

표. 첫출산시 연령에 따른 유방암의 발생수

첫출산 연령	<20	20-24	25-29	30-34	>34	계
환자군	320	1206	1011	463	220	3220
대조군	1422	4432	2893	1092	406	10245
계	1742	5638	3904	1555	626	13465

1) 행(row)백분율

첫출산 연령	<20	20-24	25-29	30-34	>34	계
환자군	9.9	37.5	31.4	14.4	6.8	3220
대조군	13.9	43.3	28.2	10.7	4.0	10245
계	12.9	41.9	29.0	11.5	4.6	13465

2) 열(column)백분율

첫출산 연령	<20	20-24	25-29	30-34	>34	계
환자군	18.4	21.4	25.9	29.8	35.1	3220
대조군	81.6	78.6	74.1	70.2	64.9	10245
계	100.0	100.0	100.0	100.0	100.0	13465

3) 전체백분율

첫출산 연령	<20	20-24	25-29	30-34	>34	계
환자군	2.4	9.0	7.5	3.4	1.6	3220
대조군	10.6	32.9	21.5	8.1	3.0	10245
계	13.0	41.9	29.0	11.5	4.6	13465

마) 그림작성

바) 중복제시 지양(그림, 표)

7) 고찰

고찰은 앞서 기술한 내용 즉 연구방법, 연구결과에 기술한 내용 각각에 대하여 순서대로 검토 기술함

가) 연구방법의 타당성과 신뢰성의 검토

(1) 연구자료 선택방법의 검토

(2) 현실적으로 감수할 수 밖에 없는 특정 비뚤림이 있을 경우는 그 불가피성을 명시

(3) 관찰항목 채택과정 검토

(4) 자료수집 방법의 검토

(5) 계량화 방법의 검토

나) 연구방법과 연구성적과의 연결

연구방법들에 대한 고찰내용을 연구성적의 해석이나 추론에 연관시켜 해석상의 제한점, 또는 본 연구가 지닌 강점 등에 관하여 기술

다) 연구성적의 고찰

(1) 연구목적과의 연관

(2) 명확한 논거 제시

(3) 상충적 성적의 고찰

라) 건의사항 제시

연구성적중 해석이나 추론할 수 없는 부분을 명시하며 추후 연구되어야 할 과제에 관한 의견을 제시

8) 결론

연구목적에 연관시켜 일반적이고 포괄적인 내용의 연구결론을 기술하는 부분으로 구체적인 숫자나 지엽적 사항의 나열은 가능한 한 배제하고 본 연구의 목적달성여부와 기대효과 및 향후 계속 추구되어야 할 연구과제에 대하여 제안함

9) 참고문헌

10) 부록

참고문헌

1. Flecher RH, Flecher SW, Wagner EH. Clinical Epidemiology. 2nd Ed. Baltimore, Williams & Wilkins, 1988
2. Gregg MB. Field Epidemiology. Oxford University Press, 1996
3. Gordis. Epidemiology. 1996
4. Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic Research. New York, Van Nostrand Reinhold Company, 1982
5. Rothman KJ. Modern Epidemiology. Boston/Tronto, Little, Brown and Company, 1986
6. Schlesselman, J J. Case control studies : Design, conduct, analysis. Oxford University Press, New York, 1982
7. 연세대학교 의과대학 예방의학교실. SAS를 이용한 의학통계 자료분석, 1995
8. 신영수, 안윤옥. 의학연구방법론. 서울대학교 출판부, 1997