자료 분석 및 해석

계명의과대학 이 충 원

분석을 계획하는 전략, 분석 시행 방법, 결과를 해석하는 지침에 대해서 기술한다.

분석전 계획 (PREANALYSIS PLANNING)

1. 무엇을 분석할 것인가? (What to Analyze)

성공적으로 분석을 하려면 먼저 연구를 시작하기 전에 분석 전략 (analytic strategy)을 잘 짜야한다. 적절한 가설을 설정하고 이 가설을 평가하기 위해서 관련 자료를 적절하게 수집, 기록, 관리, 분석 및 평가하는 것이 필요하다. 그래서 자료를 어떻게 무엇을 분석할지를 면담지를 디자인하기 전에 결정하여야 한다.

그림 8-1에 분석의 단계를 예시해 두었다.

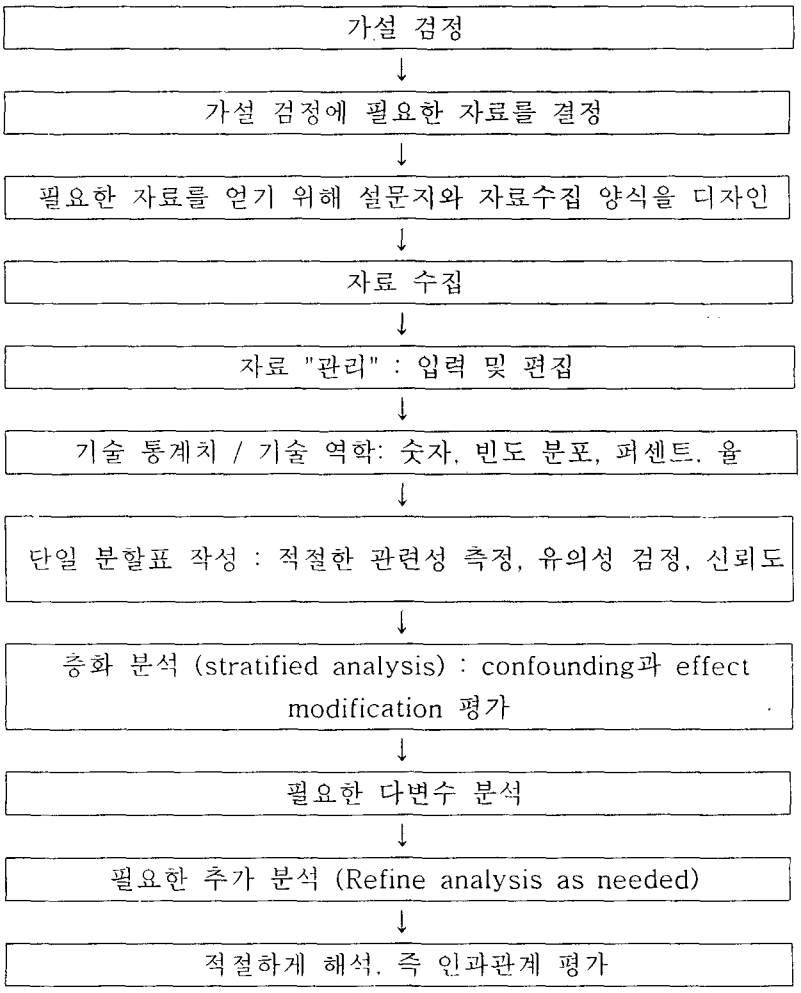

그림 8-1. 분석 단계

일단 검정해야할 가설이 설정이 되면 가설 검정을 위해 어떤 자료를 모아야 할지를 결정하여

야 한다. 중요할 수 있는 (potentially important) 위험인자에 대한 항목이 설문지에 없으면 유행에서의 그 역할을 파악할 수가 없다.

이외에 식별 정보, 임상 정보, 기술 요인에 대한 문항도 있어야 한다. 임상 정보는 한 환자가 case definition을 충족하는지를 결정할 수 있을 정도로 충분히 모아야 한다. 질병의 심각도, 입원여부, 후유증과 같은 임상적인 정보 역시 쓸모가 있다. 기술 요인은 시간, 장소, 사람과 관련된 요인은 대상자들의 특성을 파악하고, 그룹간의 비교도 (comparability)를 평가하기 위해서 수집하여야 하며 인관 관계에 대한 가설을 설정하는데 도움이 된다.

2. 자료 편집 (Data Editing)

보통 분석 역학 자료는 paper questionnaire로 자료를 수집한 후 컴퓨터에 입력한다. 그러나 요즘은 paper 과정을 거치지 않고 바로 컴퓨터로 입력하는 경우가 많다. 자료를 분격적으로 분석하기 전에 자료 관리 (data management)를 잘 해 두어야 한다.

- 중복 자료없이 정확한 수의 records를 입력
- 각 data field에 대해 질관리 점검 (quality-control checks)을 실시

2가지 종류의 질관리 점검이 있는데

- 범위 점검 (range check)

각 변수에 대해 자료수집 범위를 벗어나는 것 ("out of range")을 찾아낸다. 예) sex, weight

- 논리 점검 (logic check)

서로 논리적으로 연관되어 있는 내용의 질문들의 응답을 점검하여 논리적으로 어긋나는 것을 찾아낸다. 예를 들어 "gender"는 "male"로 되어 있으나, hysterectomy는 "yes"로 되어 있는 record. 날짜 대조 가능 (입원일이 질병 발생일보다 앞서는 경우; nosocomial infection은 예외)

자료 관리 시 주의해야할 2가지 추가 사항

- 모든 것, 특히 결정을 내린 사항을 문서화해둔다 (document). 질관리 시 일관성없는 것을 일관성있게 바꾼 내용을 반드시 기록해 둔다.
- 에러는 바로 수정한다 ("Never let an error age"). 특히 field 조사시에 바쁘고, 스트레스를 받아서 소홀히 하기 쉬운 사항이다.

3. 분석 전략 개발 (Developing the Analysis Strategy)

분석에 들어가기 전에 먼저 분석 전략을 세워야 한다 (표 8-1). 이는 term paper를 써기 전에 전체적인 윤곽을 짜는 것과 같다. 논리적인 순서에 따라 분석의 핵심 요소를 설계하는 것인데, 실제 자료분석 시에 하나의 지침이 된다.

표 8-1. 역학 분석 전략의 순서 (Sequence of an Epidemiologic Analysis Strategy)

==========

1. 자료 수집 방법을 설정하고 이에 따라 분석할 계획을 세운다.
2. subject matter, 생물학적으로 그를 듯한 가설, 연구가 진행되는 방식에 비추어 보아 가장 중요하다고 생각되는 변수를 찾아내어 목록을 만든다.
   - 문제의 폭로
   - 문제의 결과
   - 혼란변수
   - subgroup 분석을 위한 변수

3. 자료를 잘 알기 위해서 2단계에서 찾아낸 변수에 대해, 빈도 분포 분석과 기술 통계분석을 할 계획을 세운다.
4. 연구 집단의 특징을 알아보기 위해 임상적인 특징과 기술역학의 표를 만든다. (table shells을 조사에 앞서 그려 두어야 한다)
5. 폭로-질병 관계를 평가하기 위해서 연구 디자인, 기존의 지식, 가설 등에 기초를 두고 2X2 분할표를 만든다. (table shells을 조사에 앞서 그려 두어야 한다)
6. 자료에서 흥미로운 결과가 있으면 추가로 2X2 분할표를 그린다.
7. 디자인, 기존의 지식, 가설, 자료의 흥미있는 결과 등에 기초를 두고, 3X3 분할표를 그려서 분석을 다듬고 (즉, 양-반응, 민감도 분석 등) subgroup 분석을 실시한다.

---

1) 첫 번째 단계는 자료가 수집된 방법을 고려하는 것이다. 예를 들어
- cohort study 라면 폭로그룹을 중심으로 생각하고 율을 계산한다.
- case-control study 라면 cases와 controls을 중심으로 생각하고, match 되었다면 matched analysis를 실시한다.
- survey 자료 라면 samplig scheme을 검토하여 분석에 디자인 영향을 고려해야 할 때가 있다.

2) table shells을 초안 잡는 것이다. 보통 "가표" ("dummy table")이라고 불리는 것인데, 빈도 분포 또는 2X2 분할표 같은 것인데, 실제 data가 빠져 있다는 것을 제외하고는 완벽한 표의 형태를 취한다. 실제 분석을 진행하면서 빈 표에 자료를 채워 넣는다.

table shells은 간단한 표 (즉, 기술 역학)에서 복잡한 표 (즉, 분석 역학)로 논리적인 순서에 따라 진행되어야 한다. 이때 무슨 통계치 (즉, 비차비)를 계산하며 어떤 통계적인 검정 (즉, chi square)을 할지를 표시해 두어야 한다. 일반적인 표의 순서는 다음과 같다.

표 1 : 임상적인 특징 (즉, 증상 및 징후, 실험실 확진한 %, 입원한 %, 사망한 % 등)
표 2 : 기술 통계
- 시간 : 선 그래프 (추세) 또는 유행곡선
- 장소 : 거주지 또는 발생지, spot 또는 shaded map
- 사람 : 연구 대상자의 특성 (연령, 인종, 성별 등)

분석 역학을 위해
표 3 : 일차적인 연관성 표 (즉, 결과 상태에 따른 위험 요인)
표 4 : 표 3을 층화하여 서로 독립적인 effects로 분리하여 혼란 영향 (confounding)과 효과 조정자 (effect modification)의 현상이 있는지를 평가. 혼란 변수는 통제해 주어야 하고 효과 조정자는 찾아내어야 한다.
표 5 : 표 3의 분석을 좀 더 다듬는다 (즉, 양-반응관계, latency, 좀 더 민감하고 특이한 환례 정의 등)
표 6 : 특정한 subgroup 분석

다음의 A-I 까지의 table shells은 Kawasaki syndrome을 case-control study 하기 전에 디자인한 표의 순서이다. 이 증후군에 대한 명확한 진단법이 없으므로 환자는 fever에 추가해서, 가표 A에 제시된 5개의 임상적인 특징 중 최소한 4가지를 가져야 한다. 이 연구에서 검정할 3가지 가설은 antecedent viral illness, 최근의 carpet shampoo 폭로, 가계 수입 증가 등이었다.

Table Shell A. Diagnostic Criteria for Kawasaki Syndrome Cases with Onset October–December

CRITERION	NUMBER	PERCENT
1. Fever ≥ 5 days	—	(%)
2. Bilateral conjunctival injection		(%)
3. Oral changes	—	(%)
Injected lips		(%)
Injected pharynx	—	(%)
Dry, fissured lips	—	(%)
Strawberry tongue	—	(%)
4. Peripheral extremity changes	—	(%)
Edema	—	(%)
Erythema	—	(%)
Periungual desquamation	—	(%)
5. Rash	—	(%)
6. Cervical lymphadenopathy ≥ 1.5 cm	—	(%)

Table Shell B. Days of Hospitalization, Kawasaki Syndrome Cases with Onset October–December

DAYS OF HOSPITALIZATION	FREQUENCY
0	
1	
2	—
3	—
4	—
5	
6	—
7	—
8	—
9	
and so on to maximum	
Unknown	
Range	—
Mean	—
Median	

Table Shell C. Frequency Distribution of Serious Complications among Kawasaki Syndrome Cases with Onset October–December

CRITERION	NUMBER	PERCENT
Arthritis	—	(%)
Coronary artery aneurysm	—	(%)
Other complications (list)	—	(%)
Death	—	(%)

Table Shell D. Demographic Characteristics of Kawasaki Syndrome Cases with Onset October–December

DEMOGRAPHIC CHARACTERISTIC		NUMBER	PERCENT
Age	< 1 yr	—	(%)
	1 yr	—	(%)
	2 yr	—	(%)
	3 yr	—	(%)
	4 yr	—	(%)
	5 yr	—	(%)
	≥ 6 yr	—	(%)
Gender	Male	—	(%)
	Female	—	(%)
Race	White	—	(%)
	Black		(%)
	Asian		(%)
	Other	—	(%)

Table Shell E. Frequency Distribution by County of Residence, Kawasaki Syndrome Cases, October–December

COUNTY	NUMBER	PERCENT	POPULATION	ATTACK RATE
County A	—	(%)	—	—
County B	—	(%)	—	—
County C	—	(%)	—	—
County D	—	(%)	—	—
County E	—	(%)	—	—
County F	—	(%)	—	—

Table Shell F. Frequency Distribution by Household Income, Kawasaki Syndrome Cases, October–December

ANNUAL HOUSEHOLD INCOME[a]	NUMBER	PERCENT
< $15,000	—	(%)
$15,000–$29,999	—	(%)
$30,000–$44,999	—	(%)
≥ $45,000	—	(%)

[a]May need to revise categories of household income to portray range.

Table Shell G. Kawasaki Syndrome and Antecedent Illness, Case Control Study

		CASES	CONTROLS	TOTAL	
ANTECEDENT ILLNESS	YES	—	—	—	Odds ratio = ___ 95% CI = (,)
	NO	—	—	—	x^2 = _____, *P* value = _____
	TOTAL	—	—	—	

Table Shell H. Kawasaki Syndrome and Carpet Shampoo, Case Control Study

		CASES	CONTROLS	TOTAL	
CARPET SHAMPOO	YES	—	—	—	Odds ratio = ___ 95% CI = (,)
	NO	—	—	—	x^2 = _____, *P* value = _____
	TOTAL	—	—	—	

Table Shell I. Kawasaki Syndrome and Carpet Shampoo, Case Control Study

		CASES	CONTROLS	TOTAL	
HOUSEHOLD INCOME (IN THOUSANDS OF DOLLARS)	<15	—	—	—	x^2 = _____, *P* value = _____
	15–30	—	—	—	
	30–45	—	—	—	
	45+	—	—	—	
	TOTAL	—	—	—	

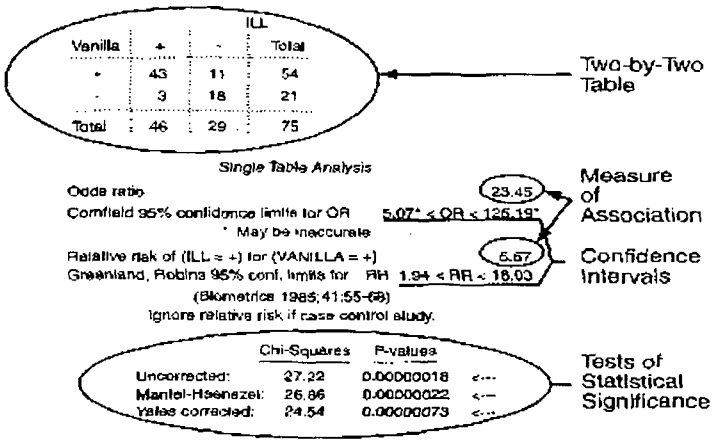

Figure 8-2. Typical *Epi Info* output from the analysis module, using the "tables" command. *Source:* CDC (1994).

그림 8-2는 Epi Info의 Analysis module을 실행시 나타나는 화면을 옮긴 것이다. "table" 명령어를 사용한다. 결과로 제시된 것에서 4가지 요소를 알아야 한다.

- 2X2 분할표
- 연관성 측정치 (measures of association)
- 통계적인 유의성 검정 (tests of statistical significance)
- 신뢰구간 (confidence intervals)

4. 2X2 분할표

대부분의 역학 조사에서 폭로와 health event가 이분 변수 (binary variables) (즉, 예, 아니오)로 측정된다. 일반적인 예가 표 8-2에 제시.

Table 8-2. Data layout and notation for standard two-by-two table.

	ILL	WELL	TOTAL	ATTACK RATE
EXPOSED	a	b	h_I	a/h_I
UNEXPOSED	c	d	h_{II}	c/h_{II}
TOTAL	v_1	v_0	t	v_1/t

South Carolina에서 발생한 유행의 예 (Salmonella gastroenteritis)에 대한 표가 표 8-3에 제시.

Table 8-3. Turkey Consumption and Gastrointestinal Illness, *Salmonella* Outbreak, South Carolina, 1990

		ILL	WELL	TOTAL	ATTACK RATE
ATE TURKEY?	YES	115	89	204	56.4%
	NO	5	36	41	12.2%
	TOTAL	120	125	245	49.0%

Source: Luby et al. (1993).[2]

연관성 측정 (MEASURES OF ASSOCIATION)

연관성의 측정치는 폭로와 health problem 간의 통계적인 연관성의 강도 또는 정도를 정량화한다. 연관성의 측정치를 때로 효과의 측정치 (measures of effect) 라고 도 하는데, 만약에 폭로가 질병과 인관 관계가 있다면 그 측정치가 폭로가 질병의 발생율에 미치는 영향을 정량화하기 때문이다.

1. Relative Risk (RR, Risk Ratio)

RR = 폭로 그룹의 위험 / 비폭로 그룹의 위험 = $(a/h_1) / (c/h_0)$

비폭로 (배경, 예상되는) 그룹에 비교해서 폭로된 그룹에서 나타나는 excess risk를 반영하며 이 excess를 비로 표현한다. 급성 유행에서 risk는 attack rate로 표현한다. 표 8-3에서 turkey를 먹으면 질병에 걸릴 상대적인 위험이 4.6이라는 의미이다. 즉 turkey를 먹지 않은 사람에 비해 먹은 사람은 질병에 걸릴 위험이 4.6배 더 높다는 것이다. 1보다 더 크면 폭로 그룹에서 발생율이 더 높다는 것이고 반면에 1보다 더 작으면 폭로 그룹에서 발생율이 더 낮다는 의미이다 (vaccination study 일 경우).

2. Odds Ratio (Cross-Product Ratio, Relative Odds)

대부분의 case-control study에서 폭로된 그룹과 비폭로 그룹의 참된 크기를 알지 못하므로 attack rate 또는 risk를 계산할 수 있는 분모를 알 수 없다. 그러나 odds ratio로 대신할 수 있다.

OR = ad / bc

community hospital에서 group A Streptococcus (GAS) surgical wound infection 유행에서 17개월 동안에 10 cases가 발생하였다. 조사자는 확률수 표 (난수표, table of random numbers)를 사용하여 그 기간 동안에 시행된 2,600 건의 surgical procedures 중에서 controls을 추출하였다. 보통 GAS surgical infections의 clusters는 수술실에 근무하는 직원 중에 있는 GAS carrier에 기인하는 경우가 많아서, 각 staff members를 폭로로 2X2 분할표를 작성하였다.

표 8-4에 nurse A에 대한 결과가 제시되어 있다. OR가 (8X49) / (2X5) = 39.2 이었는데, 이는 cases에서 controls에 비해서 nurse A에 폭로될 odds가 39배 더 높다는 것을 나타낸다. 다시 말해서 GAS surgical infection을 발생할 odds가 nurse A에 폭로된 사람에서 폭로되지 않은 사람에 비해 39배 더 높다는 의미로 볼 수 있다. rare disease (5% 이하) 일 경우 OR는 RR과 비슷해 진다 (이 경우 10 / 2,600 = 0.38%).

Table 8-4. Surgical Wound Infection and Exposure to Nurse A, Hospital M, Michigan, 1980

		CASE	CONTROL	TOTAL
EXPOSED TO NURSE A?	YES	8	5	13
	NO	2	49	51
	TOTAL	10	54	64

Source: Berkelman et al. (1982).[3]

OR는 역학적인 연구에서 연관의 정도를 측정하는데 여러모로 쓸모 있는 측정치이다.

3. Prevalence Ratio & Prevalence Odds Ratio

cross-sectional study (횡단면적 연구)에서 인구집단에서 health condition의 유병율을 측정한다. 유병율은 발생율 (rsik)과 질병의 기간 (duration of illness)에 따라 달라지므로 prevalent cases에 기초한 연관성 측정치는 폭로가 결과에 미치는 영향뿐만 아니라 질병의 기간 또는 생존 (duration & survival)에 미치는 영향까지도 반영한다.

Table 8-5. Data from a Hypothetical Cross-Sectional Survey

		HAVE CONDITION?			
		YES	NO	TOTAL	PREVALENCE
EXPOSED?	YES	20	80	100	0.20
	NO	20	380	400	0.05
	TOTAL	40	460	500	

prevalence ratio는 0.20 / 0.05 = 4.0 이고, prevalence OR는 (20X380) / (80X20) = 4.75 이다. prevalence 이 낮을 경우, 이 두 값은 비슷해 진다.

공중보건학적 영향의 측정 (MEASURES OF PUBLIC HEALTH IMPACT)

폭로-질병 연관을 공중보건학적인 관점에서 조망한 것이다. 특정한 인구모집단에서 어떤 폭로가 질병 빈도에 기여한 정도를 반영한다. 질병 위험의 증가와 관련된 폭로에서 (즉 흡연과 폐암) attributable risk percent는 만약에 그 폭로가 제거될 경우에 (또는 절대로 존재하지 않았다면) 예상되는 질병 load의 감소를 나타낸다. population attributable risk percent는 어떤 인구모집단에서 질병 중 폭로에 기인하는 분율을 나타낸다. 질병 감소와 관련된 폭로 (즉 vaccination)에서 prevented fraction은 질병 load에서 현재의 폭로수준에 기인해서 감소하는 정도를 나타낸다.

1. Attributable Risk Percent

(Attributable Fraction [또는 Proportion] among the Exposed, Etiologic Fraction)

폭로 그룹의 cases 중에서 폭로에 기인하는 분율이다. 이 측정치는 비폭로 그룹의 risk 수준이 폭로 그룹에도 역시 적용되어 excess risk 만이 폭로에 기인한다고 가정한다. 식은 다음과 같다.

Attributable risk percent = (폭로 그룹의 risk - 비폭로 그룹의 risk) / 폭로 그룹의 risk
= (RR - 1) / RR

fraction 으로 표시할 수도 있고 100을 곱하여 퍼센트로 표시할 수도 있다. 표 8-3의 자료를 이용해서 계산을 하면 (0.564 - 0.122) / 0.564 = 78.4% 인데, 이는 turkey를 먹은 사람들 중에 발생한 gastroenteritis의 3/4 이상이 turkey 소비에 기인한다고 할 수 있다. 나머지 21.6%는 그 인구모집단에서 평시에 일어나는 baseline 발생에 기인한다.

case-control study에서 OR가 RR과 비슷하다고 할 수 있다면, attributable risk percent는

Attributable risk percent = (OR - 1) / OR

2. Population Attributable Risk Percent (Population Attributable Fraction)

이는 cases 중 전체 인구집단 (폭로 그룹과 비폭로 그룹 모두 다)에서 폭로에 기인하는 분율이다.

Population attributable risk percent = (전체 risk - 비폭로 그룹 risk) / 전체 risk

= P(RR - 1) / [P(RR - 1) + 1]

(P는 폭로된 분율 = h1 / t)

turkey consumption data에 적용시켜 보면 (0.490 - 0.122) / 0.490 = 75.1% 이다. 대부분의 cases가 폭로된 경우, attributable risk percent와 population attributable risk percent는 비슷할 것이다. 여러 가지 원인 (즉 많은 만성 질환)과 흔치 않은 폭로를 가진 질병에 대해서 population attributable risk percent는 attributable risk percent에 비해 훨씬 더 적을 것이다.

population attributable risk percent는 population-based case-control study에서 OR과 controls 중 폭로된 분율인 P (즉 P = b / v0 ; controls이 전체 population을 대표한다고 가정)를 사용하여 계산할 수 있다.

3. Prevented Fraction in the Exposed Group (Vaccine efficacy)

risk ratio가 1 이하일 경우 폭로가 없었다면 발생하였을 새로운 cases의 분율이다. 다름 말로 prevented fraction은 vaccination과 같은 beneficial exposure에 의해 예방된 cases의 분율이다.

Prevented fraction among the exposed

= (비폭로 그룹의 risk - 폭로 그룹의 risk) / 비폭로 그룹의 risk

= 1 - RR

표 8-5의 자료는 1970년에 Texas-Arkansas border를 따라 measle outbreak가 발생하였을 때의 자료이다. 이 때 measle vaccination를 받았던 아이들에서도 발병하였으므로, vaccine의 effectiveness에 대한 논란이 있었다. 표 8-6의 자료를 보면

Table 8-6. Vaccination Status and Occurrence of Measles, Texarkana, 1970

	MEASLES	NO MEASLES	TOTAL	RISK PER 1,000	RELATIVE RISK
VACCINATED	27	6,323	6,350	4.2	0.04
NOT VACCINATED	512	4,323	4,835	105.9	(REFERENCE)
TOTAL	539	10,646	11,185	48.2	

Vaccine efficacy = (105.9 – 4.2) / 105.9 = 0.96

Source: Landrigan (1972).[4]

vaccine을 받은 어린이에서 risk는 vaccine 받지 않은 어린이의 risk의 4% 정도 되었다. vaccine efficacy는 96%로 계산되었는데, 이는 vaccination이 vaccine 맞은 어린이가 vaccine을 맞지 않았더라면 발생하였을 measle의 96%를 예방하였음을 나타낸다.

"attributable"과 "prevented" 라는 용어는 단순한 통계적인 연관성 이상인 인과관계 (cause-and-effect relationship)을 의미하므로 루틴으로 제시하지 말아야 하며, 인과관계를 잘 고려해 보고 인정이 될 때 비로소 제시하여야 한다.

통계적인 유의성 검정 (TESTS OF STATISTICAL SIGNIFICANCE)

통계적인 유의성 검정은 폭로가 실제로 질병과 연관이 되어있지 않았다면, 관찰된 결과가 우연만으로 (by chance alone) 발생할 가능성을 측정하는데 사용한다.

통계적인 검정에서 연구 표본 (study population)을 큰 "소스 인구모집단" ("source population")로부터 추출된 표본이라고 가정한다. 그리고 소스 인구모집단에서 질병의 발생율은 폭로된 그룹과 비폭로된 그룹 둘 다에서 동일하다고 가정한다. 다시 말해 소스 인구모집단에서 폭로가 질병과 연관이 없다고 가정하는 것이다. 이러한 가정을 영가설 (null hypothesis) 라고 한다.

가설 검정의 순서는

1. 영가설 설정
2. 연구 가설에 적절한 통계치를 계산
3. 통계치를 사건이 일어날 확률로 표시: P값 (이 확률은 영가설이 실제로 사실이라면, 관찰할 수 있는 것만큼 또는 관찰할 수 있는 것보다 더 강한 연관을 관찰할 확률이다). 다른 말로 P값이 작으면, 연구한 표본에서는 영가설이 받아들이기 어렵다는 것을 나타낸다.
4. 결론 : 만약에 P값이 미리 설정해둔 정해진 값보다 더 작으면 (보통 0.05), 영가설을 기각하고 (discard, reject) 대안가설을 받아들인다. 즉 이 표본에서 관찰된 연관이 "통계적으로 유의성" ("statistically significant") 이 있다고 할 수 있다.

영가설에 대한 결정을 내릴 때, 2가지 형태의 오차에 주의하여야 한다.

1. 제1형 오차 (type I error, α error) : 영가설이 맞는데, 기각할 확률.
2. 제2형 오차 (type II error, β error) : 영가설이 틀리는데, 기각 못할 확률.

연구를 시작하기 전에 영가설과 대안가설을 설정해두어야 한다. 연관의 방향성에 대해서 잘 모를 경우에는 폭로가 질병과 연관이 없다고 영가설 (즉 OR = 1)을 설정한다. (대안가설은 폭로가 질병과 연관이 있다는 것이다. 즉 OR ≠ 1) "non-directional hypothesis" → "two-tailed" test

"directional" hypothesis는 연관관계가 더 잘 알려져 있어서 방향설정을 할 수 있을 때 사용한다.

예) 흡연과 폐암간의 관련성 또는 MMR과 measles간의 관련성 연구 → 영가설: OR ≤ 1, 대안가설: OR > 1. 이 경우 "one-tailed" test 시행.

일반적으로 단측 검정일 경우 양측 검정에 비해 cutoff가 2배이다. 즉 양측 검정은 0.05 이나 단측 검정은 0.10 이다. 보통 field 역학에서는 양측 검정을 실시한다.

참고) "Fishing expeditions" ("Data dredging")

$1 - (0.95)^n$ (n=1, 2, 3, 4 ... ; 통계적 검정 시행 횟수)

n=1, 0.05; n=2, 0.0975; n=3, 0.143; n=4, 0.185 등이 된다.

다음은 field 역학에서 흔히 사용하는 2가지의 통계기법을 소개한다.

1. Fisher Exact Test

2X2 분할표에 대한 통계적인 검정의 "gold standard" 이며 test of choice 이며, 표의 cell 내의 표본 수가 적을 때 사용한다.

계산 방법은 부록 8-1에 제시되어 있다.

일반적으로 2X2 분할표의 cell 내의 기대치 (expected value)가 하나라도 5이하인 경우에 사용한다. 그러나 표본 수가 증가하면 계산하기가 힘들다. 이 때는 chi-square test가 Fisher exact test와 비슷해지므로, chi-square test를 사용한다.

2. Chi-Square test

대상자수가 최소한 30명이상이고 2X2 분할표의 cell 내의 기대치가 최소한 5 이상인 경우에 chi-squre test는 Fisher exact test와 비슷해진다.

Epi Info에서는 3가지 계산식을 제공한다.

$$\text{Pearson uncorrected } \chi^2 = \frac{t\,(ad-bc)^2}{(v_1)\,(v_0)\,(h_1)\,(h_0)}$$

$$\text{Yates corrected } \chi^2 = \frac{t\left(|ad-bc| - \left(\frac{t}{2}\right)\right)^2}{(v_1)\,(v_0)\,(h_1)\,(h_0)}$$

$$\text{Mantel-Haenzel } \chi^2 = \frac{(t-1)\,(ad-bc)^2}{(v_1)\,(v_0)\,(h_1)\,(h_0)}$$

Pearson chi-square 식이 가장 큰 값을 나타내며, Fisher exact test 값보다 더 크다. 그래서 Paerson 값이 제1종 오차에 빠질 위험이 더 크다. 반면에 Yates corrected chi-square는 가장 작은 값을 나타내어 제1종 오차의 위험이 가장 적은 반면에 제2종 오차의 위험은 더 커진다. Mantel-Haenzel 은 중간이며 주로 층화 분석 (stratified analysis)에서 사용한다.

표 8-7에 예가 제시되어 있다.

Table 8-7. Macaroni Consumption and Gastroenteritis, *Salmonella* Outbreak, South Carolina, 1990

	ILL	WELL	TOTAL	RISK	
EXPOSED	76	63	139	54.7%	Relative risk = 1.3
UNEXPOSED	44	62	106	41.5%	Odds ratio = 1.7
TOTAL	120	125	245		

$$\text{Uncorrected } \chi^2 = \frac{(245)\,(76\times 62 \quad 63\times 44)^2}{(120)\,(125)\,(139)\,(106)} = 4.17$$

$$\text{Mantel-Haenszel } \chi^2 = \frac{(245-1)\,(76\times 62 - 63\times 44)^2}{(120)\,(125)\,(139)\,(106)} = 4.16$$

$$\text{Corrected } \chi^2 = \frac{(245)\left(|76\times 62 - 63\times 44| - \frac{(245)}{2}\right)^2}{(120)\,(125)\,(139)\,(106)} = 3.66$$

The corresponding two-tailed *P* values are as follows:

Uncorrected $\chi^2 = 4.17$, *P* value = 0.041

Mantel-Haenszel $\chi^2 = 4.16$, *P* value = 0.042

Corrected $\chi^2 = 3.66$, *P* value = 0.056

Fisher exact *P* value (two-tail) = 0.053

Source: Luby et al. (1993).[2]

3. 어느 검정법을 사용하는가?

Fisher exact test는 2X2 분할표의 cell 내의 기대치가 5 이하인 경우에 사용한다. 2X2 분할표에서 모든 기대치가 5 이상이면, chi-square tests 중 하나를 선택하면 된다.

4. 연관성 측정 대 유의성 검정 (Measure of Association vs Test of Significance)

relative risk 또는 odds ratio 과 같은 연관성 측정치는 폭로와 질병간의 연관성의 강도와 방향성을 반영해준다. 이러한 측정치는 일반적으로 연구의 표본 크기에 독립적이므로, 소스 인구모집단에서 연관성의 강도를 가장 잘 추정해주는 것 ("best guess")으로 간주할 수 있다. 그러나 측정치는 측정치의 신뢰도를 나타내 주지는 않는다.

반면에 유의성 검정은 관찰된 결과가 우연에 의할 가능성을 나타내 준다. chi-square test 통계치가 연관의 크기와 연구의 표본 크기에 영향을 받을 수 있으나, 이들 두 가지 영향을 구분해 주지는 않는다. 그래서 이들 두 가지는 서로 보완적인 정보를 제공한다.

5. 통계학적인 검정 결과 해석 (Interpreting Statistical Test Results)

"유의하지 않다"는 것은 "연관성이 없음"을 항상 의미하지는 않는다. 연관성의 측정은 연관성의 방향과 강도를 나타낸다.

통계적인 유의성은 그 자체가 원인-효과 관계 (cause-effect relationship)를 나타내는 것은 아니다. 관찰된 연관성은 인과관계를 나타낼수도 있으나 chance, selection bias, information bias, confounding 그리고 연구 디자인, 실행, 분석에서의 오차 소스에 기인할 수도 있다. 통계적인 검정은 관찰된 연관성을 설명하는데 우연의 역할만을 측정해주며, 통계학적으로 유의하다는 것은 연관성이 단지 우연에 의할 가능성이 희박하다는 것을 나타낸다. 최종적인 역학적인 판단은 인과관계를 판단하는 기준 (다른 연구와의 일관성, 폭로와 질병간의 시간 순서, 생물학적인 가능성 등) 을 동원하여 신중하게 내려야 한다.

통계적인 유의성이 있는 것이 공중보건학적인 중요성을 항상 가진다고 볼 수 없다. 표본 크기가 클 경우 공중보건학적인 (임상적인) 중요성이 없는 약한 연관성이라고 할지라도 통계적으로 유의할 수 있다.

연관성 측정의 신뢰구간
(CONFIDENCE INTERVALS FOR MEASURES OF ASSOCIATION)

연관성의 통계학적인 변이 (statistical variability of the association)를 나타내 주는 또 다른 지수 (index)는 신뢰구간 (confidence interval) 이다. 소스 인구모집단을 반복하여 표본 추출을 하였을 때, 95% (100번 중 5번에서) 에서 진실된 연관성을 포함할 구간이다. 하나의 연구에서 얻은 신뢰구간은 진실된 값을 포함할 95%의 확률을 가지는 범위의 값이라고 해석할 수 있다.

1. 신뢰구간 해석

어떤 표본을 대상으로 흡연과 질병간의 RR가 4.0이고, 95% 신뢰구간이 3.0-5.3 이었을 경우의 의미는 한번의 연구에서 소스 인구모집단의 RR을 4.0으로 추정하였으며 동일한 연구를 반복해서 할 때 95%의 경우에서 RR 값이 3.0-5.3에 포함된다는 것이다.

신뢰구간의 넓이는 어떤 한 연구가 연관성 (RR)을 찾아낼 수 있는 정밀도 (precision)를 반영한다. 신뢰구간이 넓으면 변이의 정도가 크거나 또는 정밀하지 못함을 반영하며, 좁은 경우는 반

대로 변이가 거의 없고 정밀도가 높음을 나타낸다. 보통 연구에서 대상자의 수가 클수록 정밀도는 더 커지고 신뢰구간은 더 좁아진다.

신뢰구간을 영가설을 검정하는데 사용할 수 있다. 영가설에서 RR (또는 OR)이 1.0이라고 설정되므로, 1.0을 포함하는 신뢰구간은 영가설과 일관성이 있다. 이것은 영가설이 기각될 수 없는 경우에 해당된다. 반면에 1.0을 포함하지 않는 신뢰구간은 영가설이 기각되어야 되어야 한다는 것을 나타내므로 통계적인 유의성이 있다고 판단할 수 있다.

요약 폭로표 (SUMMARY EXPOSURE TABLES)

문제의 폭로를 표 8-8에서 처럼 하나의 표에 요약하는 것이 도움이 될 수 있다.

Table 8–8. Food-Specific Attack Rates for Persons Who Ate Sunday Lunch, *Salmonella* Outbreak, South Carolina, 1990[a]

	ATE			DID NOT EAT					
FOOD	# CASES	TOTAL	AR %	# CASES	TOTAL	AR %	RR	(95% CI)	P VALUE
Turkey	115	204	56	5	41	12	4.6	(2.0, 10.6)	<0.001
Ham	65	121	54	54	122	44	1.2	(0.9, 1.6)	0.178
Dressing	99	186	53	21	59	36	1.5	(1.0, 2.2)	0.027
Gravy	85	159	53	35	85	41	1.3	(1.0, 1.7)	0.090
Macaroni	76	139	55	44	106	42	1.3	(1.0, 1.7)	0.056
Beans	96	183	52	23	61	38	1.4	(1.0, 2.0)	0.065
Corn	80	153	52	40	92	43	1.2	(0.9, 1.6)	0.229
Rolls	78	158	49	41	84	49	1.0	(0.8, 1.3)	0.958
Butter	47	88	53	73	157	46	1.2	(0.9, 1.5)	0.365
Tea	102	203	50	18	42	43	1.2	(0.8, 1.7)	0.482
Coffee	9	28	32	111	217	51	0.6	(0.4, 1.1)	0.090
Cranberries	42	74	57	78	171	46	1.2	(1.0, 1.6)	0.144

[a]AR indicates attack rate; RR, relative risk, and CI, confidence interval.
Source: Luby et al. (1993).[2]

표 8-8에서 유행을 야기시킨 원인 요인을 찾아내기 위해서는 다음의 2가지 특징을 가진 식품이 있는지 찾아보아야 한다.

(1) 증가된 RR, OR 또는 chi-square 값 (작은 P 값) : 폭로된 그룹과 비폭로된 그룹간의 attack rate에 상당한 차이가 남을 시사한다.

(2) 대부분의 환례가 폭로되어, 폭로가 모든 환례는 아니지만 대부분의 예를 설명할 수 있어야 한다.

표 8-8에서 turkey가 가장 높은 RR을 나타내었으며 120 례의 환자 중 115명을 설명해 준다.

층화 분석 (STRATIFIED ANALYSIS)

2가지 서로 다른 폭로가 질병과 연관이 되어있는 경우, 이들 둘을 동시에 분석하는 방법은? 하나의 특정한 폭로와 하나의 특정한 질병간의 연관성을 파악하고자 할 때, 제3의 요인이 이들간의 연관성을 복잡하게 할 수 있다. 2가지 형태의 complications이 있는데, 혼란 영향 (confounding)과 효과 조정현상 (effect modification) 이다. 층화 분석 (stratified analysis)는 제3의 요인의 범주 (categories) 내에서 폭로-질병간의 상관성을 검사하는 방법으로 이러한 문제점을 다루는 방법 중 하나이다.

표 8-9에 제시된 자료는 가상적인 자료로서, 중학교 학생에서 hepatitis A 유행을 조사한 것이다. 조사자는 환례 50명, 대조례 50명에 대해 food consumption questionnaire를 실시하였다. 그 중 milk, donuts와 hepatitis A와의 상관성을 제시한 자료이다.

Table 8-9. Hepatitis A and Consumption of Milk and Donuts

MILK	CASES	CONTROLS	TOTAL	
EXPOSED	37	21	58	Odds ratio = 3.9
UNEXPOSED	13	29	42	Yates-corrected χ^2 = 9.24
TOTAL	50	50	100	P value = 0.0002
DONUTS	**CASES**	**CONTROLS**	**TOTAL**	
EXPOSED	40	20	60	Odds ratio = 6.0
UNEXPOSED	10	30	40	Yates-corrected χ^2 = 15.04
	50	50	100	P value = 0.0001

우유와 도너츠에서 OR이 증가되어 있었다. 도너츠는 보통 우유와 같이 먹으므로, 많은 사람들이 이들 2가지 식품에 동시에 폭로되어 있었다. 이를 해결하기 위해서

(1) 층화 분석을 실시한다. 폭로 변수와 층화 변수로 될 식품을 결정한다. 도너츠의 OR이 더 크므로 폭로 변수, 우유를 층화 변수로 정하고 분석실시 한 표가 8-10이다. 우유 소비에 관계없이 도너츠와의 상관성이 6.0이었다. 반대로 도너츠를 층화 변수, 우유를 폭로 변수로 하여 분석 실시하니 표 8-11에서 처럼 우유와 질병과 상관성이 없었다 (OR=1). 결과적으로 우유가 질병과 연관된 것으로 결론을 내릴 수 있다.

Table 8-10. Hepatitis A and Donut Consumption, Stratified by Milk

		DRANK MILK				DID NOT DRINK MILK	
		CASES	CONTROLS			CASES	CONTROLS
ATE DONUT?	YES	36	18	ATE DONUT?	YES	4	2
	NO	1	3		NO	9	27
		Odds ratio = 6.0				Odds ratio = 6.0	

Table 8-11. Hepatitis A and Milk Consumption, Stratified by Donuts

		ATE DONUT				DID NOT EAT DONUT	
		CASES	CONTROLS			CASES	CONTROLS
DRANK MILK?	YES	36	18	DRANK MILK?	YES	1	3
	NO	4	2		NO	9	27
		Odds ratio = 1.0				Odds ratio = 1.0	

(2) 다른 방법으로 표 8-12에 제시된 것처럼 2X4 분할표가 있다.

Table 8-12. Data Layout for Two-by-Four Table, Analyzing Two Exposures at Once

EXP 1	EXP 2	ILL	WELL	TOTAL	RISK	RISK RATIO	ODDS RATIO
Yes	Yes	a_{YY}	b_{YY}	h_{YY}	a_{YY}/h_{YY}	$Risk_{YY}/Risk_{NN}$	$a_{YY}d/b_{YY}c$
No	Yes	a_{NY}	b_{NY}	h_{NY}	a_{NY}/h_{NY}	$Risk_{NY}/Risk_{NN}$	$a_{NY}d/b_{NY}c$
Yes	No	a_{YN}	b_{YN}	h_{YN}	a_{YN}/h_{YN}	$Risk_{YN}/Risk_{NN}$	$a_{YN}d/b_{YN}c$
No	No	c	d	h_{NN}	c/h_{NN}	1.0 (Ref)	1.0 (Ref)

폭로 1은 "EXP1", 폭로 2는 "EXP2"로 제시되어 있다. EXP1과 EXP2 모두에 폭로되지 않은 범주를 기준으로 하여 (마지막 row) OR를 계산한다. 이 표에서 separate 및 joint effects를 평가할 수 있다.

row2와 row3을 참고하면, 우유가 아닌 도너츠가 원인 요인이라는 사실을 알 수 있다.

Table 8–13. Hepatitis A and Consumption of Milk and Donuts, in Two-by-Four Table Layout

DONUT	MILK	CASE	CONTROL	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)

1. 혼란 작용 (Confounding)

혼란 작용은 폭로-질병 관계가 제3의 요인 ("혼란 요인", "confounder")에 의해서 왜곡되는 것 (distortion) 이다. 혼란 요인이 가져야 할 조건은 다음과 같이 2가지이다.

(1) 폭로와는 독립적으로 결과 (질병)과 연관되어 있다 - 즉 비폭로 그룹에서도 역시 연관 (다른 말로 독립적인 "risk factor" 라야 한다)
(2) 폭로와 연관은 되어 있으나 폭로의 결과 (consequence of exposure)는 아니다.

폭로의 영향과 혼란 요인의 영향을 분리하기 위해서는 혼란 요인을 층화하여야 한다.
예) Alaska (younger)와 Arizona 주 (older)의 사망률 비교. 1988년에
Arizona 주의 crude mortality rate (CMR)가 7.9 / 1,000 명 (연령교정 7.5)
Alaska 주의 CMR 3.9 / 1,000 명 (연령교정 8.4)
이 경우 연령이 혼란 요인으로 작용할 수 있으므로, 연령 교정 사망률을 비교하여야 한다.
- 거주 지역에 상관없이 연령이 증가하면 사망률이 증가한다.
- 연령은 거주지역과 상관관계 있다.

simple 2X2 분할표 분석에서 2개 또는 그 이상의 위험 요인이 나타나면, 각각을 서로 층화하여 분석하여야 한다.

혼란 작용을 찾아내기 위해서
(1) 먼저 층에 따른 연관성 측정치 중 가장 작은 것과 가장 큰 것을 찾아서, crude value와 비교한다. 만약에 crude value가 가장 작은 것과 가장 큰 것의 범위 내에 들지 않으면, 혼란 작용이 존재한다. 그러나 혼란 작용이 분명하지 않을 경우가 있는데, 그 다음 단계로서
(2) 조정된 요약 연관성 측정치 (summary "adjusted" measure of association)를 계산한다. 보통 Mantel and Haenzel 방법으로 계산한다. 조정된 값과 crude 값을 비교해서 "상당히 차이" ("appreciably different")가 나는지 알아 본다. 그러나 어느 정도가 "상당한 차이"를 보이는지를 평가하는 hard-and-fast rule 또는 통계적으로 검정하는 방법이 없다. crude 값과 adjusted 값이 별다른 차이가 없을 경우, crude 값을 사용하는데 설명하기가 더 쉽기 때문이다. 두 값이 상당히 다를 경우 (10%? 또는 20%?)에는 조정된 값을 사용한다.

2. 효과 조정현상 (Effect Modification)

층화분석의 또 다른 사용처가 바로 효과 조정자를 평가하는 것이다. 효과 조정자는 간단하게 이야기하면 폭로와 결과간의 연관성의 정도가 인구모집단의 subgroups 간에 서로 다른 현상이다. 예를 들어 measles vaccine (폭로)이 15개월 후에 맞으면 질병 (결과)을 예방하는데 아주 효과적이나, 15개월 전에 맞으면 효과가 떨어진다. 이 때 층화 변수가 vaccination 맞는 연령이며 15개월을 기준으로 연관성이 달라진다.

다른 예로 tetracycline (폭로)은 어린이에서는 tooth mottling (결과)를 유발시키지만, 성인에서는 유발시키지 않는다. 이 때 연령 (어린이, 어른)이 층화변수익, 이에따라 효과가 달라진다.

결과적으로 혼란작용은 통제를 해주어야 하는 현상이나, 효과조정 현상은 찾아내어 밝혀주어야 하는 생물학적인 현상이다. 혼란작용은 자료의 artifact이므로 아주 흔하지만, 효과조정 현상은 생물학적인 현상이므로 혼란작용보다 훨씬 흔하지 않다.

효과조정 현상은 2가지의 과정을 거쳐서 찾아낸다.

(1) 연관성의 범위가 공중보건학적인 또는 과학적인 중요성을 가질 정도로 넓은가? (field 역학의 신조는 "차이는 차이가 나기위해서는 영향을 미쳐야 한다"

(2) 연관성의 범위가 정상적인 표본추출 변이 (normal sampling variation)를 나타낼 가능성이 있는가? 질적으로 ("결과를 눈으로 살피는 것") 또는 양적으로 평가를 하여야 한다 (지수회귀분석 같은 다변수분석 또는 heterogeneity의 통계적인 검정으로 시행).

추가 분석 (ADDITIONAL ANALYSES)

언급해야할 2가지 분야가 더 있는데, 바로 양-반응 작용과 모델링이다. 간단하게 개념만 설명한다.

1. 양-반응 (Dose Response)

역학에서 양-반응은 폭로의 양이 많아짐에 따라 질병의 위험이 같이 증가하는 것을 말한다. 폭로의 양은 폭로의 강도 (즉 L-tryptophan의 양 (milligrams), 일일 소비 담배개피 수 등) 또는 폭로의 기간 (즉 폭로된 개월 수 또는 연수)으로 평가한다.

만약에 폭로와 건강 문제간의 연관성이 확립이 되면, 다음 단계로 양-반응 관계를 찾아보게 된다. 양-반응관계의 입증은 인과관계 추론의 중요한 하나의 기준이 된다. 혼란 변수를 통제해 주면서 양-반응 관계를 검정해주는 통계적인 기법이 있다.

첫 단계는 자료를 양-반응관계를 평가하기 편리하게 정리하는 것이다. 하나의 방법이 2XH 분할표를 만드는데, H가 폭로의 범주 또는 용량이 된다.

표 8-14에 제시된 것처럼 가장 낮은 용량 (또는 비폭로 그룹)을 기준으로 각각의 용량에 대해서 OR를 계산하고 신뢰구간을 구한다.

Table 8-14. Data Layout and Notation for Dose-Response Table

	H.I.	WELL		*Odds ratio*
Dose 5	a_5	b_5	h_5	a_5d/b_5c
Dose 4	a_4	b_4	h_4	a_4d/b_4c
Dose 3	a_3	b_3	h_3	a_3d/b_3c
Dose 2	a_2	b_2	h_2	a_2d/b_2c
Dose 1	a_1	b_1	h_1	a_1d/b_1c
Dose 0	c	d	h_0	1.0 (reference)
	v_1	v_0	t	

이러한 표에 제시된 자료만 보아도 양-반응관계가 있는지 감을 잡을 수 있다. 만약에 OR이 monotonic 하게 증가하거나 또는 감소를 하면, 통계적으로 유의성 있는 관계가 있을 가능성이 높다. Mantel extension test로 양-반응 관계를 평가해 볼 수 있다.

2. 모델링 (Modeling)

단일변수 분석 또는 층화분석만으로 연구목적을 달성할 수 없을 때 모델링을 사용할 수 있다. 모델링은 자료를 특정한 통계적인 식 (statistical equations)에 fit 하는 기법이다. 한 방법이 회귀모델 (regression model) 인데 결과를 폭로 변수, 혼란변수, 상호작용 항 (효과조정자) 의 함수로 표현한다. 자료의 형태에 따라 회귀 모델 방법이 달라진다. 예를 들어 이항 결과변수 (ill/well, case/control, alive/dead 등) 일 경우 역학에서는 지수회귀분석법 (logistic regression) 을 많이 사용한다.

지수회귀분석에서 이항 결과 변수는 일련의 독립변수의 함수로 모델링된다. 독립변수는 관심이 되는 폭로이고 혼란변수와 기타의 복잡한 상호작용 항이 포함될 수 있다. β coefficients로 평가를 하는데,

(1) 모델에 결과 변수와 관심의 폭로 변수만 포함되면, e^{β} 는 OR과 같다.

(2) 다른 항이 더 포함되면, e^{β}는 식에 포함된 모든 다른 항에 대해 조정된 OR이 된다.

지수회귀분석으로 양-반응관계, 효과조정 현상 및 기타 복잡한 관계를 평가하는데 사용할 수 있다. case-control 디자인에서 match를 하였을 경우 conditional logistic regression 기법이 적절하다.

역학에서 사용하는 다른 모델에는 생명표 분석에 사용하는 Cox proportional hazards model, risk ratio analysis에 사용하는 binomial regression, rare-event data에 사용하는 Poisson regression 방법 등이 있다.

아무리 복잡한 분석 기법이라고 할지라도 잘못 수집된 자료를 보완해 주지는 못한다는 사실을 명심하여야 한다. 분석 기법은 단지 도구에 지나지 않는다.

CASE-CONTROL STUDY에서 MATCHING

연구 디자인이 다를 경우 분석 기법 역시 이에따라 실시하여야 한다. 짝짓기 (matching) 는 지금까지 언급한 분석기법과는 다른 방법을 필요로 한다.

짝짓기는 일반적으로 case-control study 디자인에서 하나 또는 그 이상의 특징 (관심의 폭로 이외의 변수) 에 있어 controls을 의도적으로 cases와 비슷하게 선택하는 것이다. 그 목적은 비교성을 높이고자 하는 것이다. 가장 적절한 짝짓기 변수는 보고자 하는 폭로-질병 관계에서 혼란변수로 작용할 수 있는 변수이다. 연령, 성별 또는 지역적인 지역과 같은 요인에 대해서 짝짓기를 하므로서 이들 변수의 cases와 controls간에 분포가 같아질 것이다.

2가지 형태의 짝짓기 방법이 있다.

(1) **pair matching**

특정한 case에 따라 control을 선택하는 방법이다. 각 case가 짝짓기 변수의 관점에서 독특할 경우에 적절한데, 예로서 50명의 cases가 지역적으로 아주 넓게 분포되어 있을 때, 각 case를 친구 또는 이웃 control로 짝 지울 수 있다.

(2) **category (또는 frequency) matching**

controls을 일종의 층화 표본추출 (stratified sampling) 하는 것인데, 짝짓기 변수의 각 범주 내에서 cases 수에 비례하도록 controls을 추출한다. 예를 들어 70명의 남자와 30명의 여자가 cases인 연구에서 100명의 controls이 필요하다면, 모든 건강한 남자의 pool에서 확률적으로 남자 70명을 추출하고 건강한 여자의 pool에서 같은 방법으로 여자 30명을 추출하는 것이다. pairs는 독특하지 (unique) 않으며 이러한 방법으로 짝짓기 한 자료는 층화분석으로 분석하여야 한다.

matching의 장점은 다음과 같다.

- neighborhood, friendship, 또는 sibship 과 같은 변수에 대해 짝짓기를 하면 여러 가지 통제하기가 곤란한 사회적인 요인에 의한 혼란작용을 통제할 수 있다.
- 비용과 시간에 있어 효율적이며 controls 등록이 원활해 진다. 예를 들어 각 case 를 면담할 때 친구 controls을 찾기 쉬우며 general population에서 얻은 controls 보다 협조를 더 쉽게 얻을 수 있을 것이다.
- 통계적인 효율 (statistical efficiency) 이 증가되어 신뢰구간이 좁아진다.

반면에 단점은 다음과 같다.

- 일단 짝짓기 한 변수는 질병과의 관련성을 알아 볼 수가 없다. 예를 들어 연령과 성별에 따라 짝짓기를 해 버리면 질병과 이들 변수와의 관련성을 평가해 볼 수 없다.
- 짝짓기 변수가 실제적으로 짝짓기가 까다로운 변수일 경우, 적절한 controls을 찾아내는데 비용과 시간이 비효율적일 수 있다. 짝짓기 변수의 수가 많아질수록 적절한 controls을 찾기 어려워진다.
- 혼란 변수가 아닌 변수에 대해서 짝짓기를 하거나 또는 적절한 controls을 발견할 수 없어 case를 버려야 할 경우에는 통계적인 효율을 감소시켜서 신뢰구간이 넓어진다.
- 분석이 복잡해 진다.

요약하면, 짝짓기는 다음과 같은 사실을 먼저 알고 시행할 경우에 바람직하고 도움이 된다.

(1) 짝짓기 변수와 질병과의 관련성을 평가할 필요가 없을 때
(2) 짝짓기 요인이 질병의 위험 요인이어서 혼란 작용을 가질 경우
(3) 짝짓기는 편리하거나 또는 최소한 추가적인 비용의 가치가 있다.

위의 사실이 확실하지 않을 경우, 짝짓기를 하지 말거나 또는 폭로 및 비폭로 그룹간에 분포가 아주 다르고 평가할 필요가 없는 강력한 위험 요인에 대해서만 짝짓기를 실시한다.

1. Matched Pairs

matched pair 분석을 위한 표는 앞에서 언급한 2X2 분할표와 보기는 비슷하나 실제로 아주 다르다. matched-pair 2X2 분할표에서 각 cell은 row와 column 기준을 만족시키는 matched pairs의 수이다.

표 8-15에서 E+는 "폭로", E-는 "비폭로" 를 나타낸다. cell f는 폭로된 case와 비폭로 control로 이루어진 짝의 수이다. cells e와 h는 case와 control이 동일한 폭로 범주에 속하므로 "concordant pairs" 라고 부르고 cells f와 g는 "discordant pairs" 라고 부른다.

Table 8–15. Data Layout and Notation for Matched-Pair Two-by-Two Table

		CONTROLS E+	CONTROLS E-	Total
CASES	E+	e	f	e + f
	E–	g	h	g + h
	TOTAL	e + g	f + h	e + f + g + h

Table 8–16. Continual Tampon Use during Index Menstrual Period in Case-Control Pairs, Toxic Shock Syndrome Study, 1980

		CONTROLS YES	CONTROLS NO	TOTAL
CASES	YES	33	9	42
	NO	1	1	2
	TOTAL	34	10	44 pairs

Odds ratio = 9 / 1 = 9.0
McNemar uncorrected chi-square test = $(9 - 1)^2 / (9 + 1) = 6.40$ ($P = 0.01$)
McNemar corrected chi-square test = $(|9 - 1| - 1)^2 / (9 + 1) = 4.90$ ($P = 0.03$)

Source: Shands et al. (1980).[7]

matched-pair 분석에서 discordant pairs 만이 정보를 제공해 줄 수 있다. 이 때 OR 계산식은

$$OR = f / g$$

유의성 검정은 McNemar chi-square test로 시행한다.

$$\text{Uncorrected McNemar test} = \frac{(f - g)^2}{(f + g)}$$

$$\text{Corrected McNemar test} = \frac{(|f - g| - 1)^2}{(f + g)}$$

표 8-16에 실제 자료가 예시되어 있다.

2. Matched Triples

하나의 case에 2개의 controls이 짝짓기된 예가 표 8-17에 제시되어 있다. 각 cell은 f_{ij}로 명명되어 있는데, i는 triplet에서 폭로된 cases의 수 (case가 폭로되면 1, case가 비폭로면 0), j는 폭로된 controls의 수이다. 그래서 f_{02}는 case는 비폭로이나 controls은 둘 다 폭로된 triplets의 수이다.

Table 8–17. Data Layout and Notation for a Matched Case-Control Study with Two Controls per Case

		PERCENT EXPOSED CONTROLS		
		2 of 2	1 of 2	0 of 2
CASES	E+	f_{12}	f_{11}	f_{10}
	E–	f_{02}	f_{01}	f_{00}

Table 8–18. Kawasaki Syndrome and Annual Household Income > $40,000, Washington State, 1986

		NUMBER OF EXPOSED CONTROLS		
		2 of 2	1 of 2	0 of 2
CASES	E+	0	1	7
	E–	0	4	4

Odds ratio = (2 × 7 + 1) / (2 × 0 + 4) = 3.8

Source: Dicker (1986).[4]

OR 의 일반적인 계산식은

OR = (폭로된 cases와 짝 지워진 비폭로된 controls 수) / (비폭로된 cases와 짝 지워진 폭로된 controls 수)

matched triplets에서 위의 일반식은 다음과 같이 변한다.

$$\text{Odds ratio} = \frac{2f_{10} + f_{11}}{2f_{02} + f_{01}}$$

표 8-17에 Washington State에서 Kawasaki syndrome을 case-control study한 예가 제시되어 있는데 16명의 case-subjects 각각에 대해 2명의 연령 및 이웃 (neighborhood)에 대해 짝지워 controls을 선택하였다. carpet cleaning과 연관없는 것으로 나타났으나, 높은 가계수입과는 연관이 있었다.

3. Larger Matched Sets & Variable Matching

이 경우에는 좀 더 표가 복잡해지나, 분석의 원칙은 동일하다. 분석은 Epi Info로 쉽게 수행할 수 있다.

짝짓기 디자인은 짝짓기 분석을 필요로 하는가?

일반적으로는 필요로 한다. pair-matched design에서 만약에 pairs가 unique 하면 (즉 형제, 친구 등), pair-matched 분석이 필요하다. 반면에 pairs가 성별 또는 인종처럼 unique 하지 않은 특징에 기초를 두고서 선택하였다면 층화 분석이 더 적당하다. 빈도 짝짓기 연구에서는 층화 분석이 필요하다.

실제적으로 적절한 matched 분석을 실시한 다음 "짝짓기를 풀고" ("break the match") unmatched analysis를 시행하는 역학자도 있다. 그래서 결과가 비슷하면, 자료를 짝짓기하지 않은 형태로 제시하기도 한다. 대부분의 경우 unmatched OR은 matched OR에비해 1.0에 더 가까워질 것이다 ("bias toward the null"). 이 보다는 덜 자주 "broken" 또는 unmatched OR는 null로 부터 더 멀어질 것이다 ("further from the null"). 혼란 작용과 연관된 이러한 차이는 별것

아닐 수도 있고 상당할 수도 있다. unmatched data에서 얻은 chi-square test 결과는 특히 잘못될 가능성이 높은데, 일반적으로 matched data를 McNemar test를 한 결과보다 더 크다. matched 분석 또는 unmatched 분석을 사용할 건지에 대해 결정을 내리는 것은 crude 또는 adjusted 결과를 제시할 것인지를 판단하는 것과 같다.

FIELD DATA 해석

증가된 RR 또는 OR 또는 통계적으로 유의한 chi-square test는 폭로가 질병의 진실된 원인이라는 것을 의미하는가? 확실하게 그렇지 않다. 연구 디자인, 실행, 분석 등의 오차로 인해서 연관성이 없는데 있는 것처럼 나타날 가능성이 충분히 있다. chance selection bias, information bias, confounding, investigator error 등으로 인해 잘못된 결과가 도출될 수 있다.

- chance : 아주 작은 P값이 영가설이 결과에 대한 설명이 아닐 가능성이 높다는 것을 나타내나, 우연이 완전하게 배제되는 것은 아니다. 항상 제1종, 제2종 오차가 있을 수 있다. 연구를 시행하여 P=0.01라는 값을 얻었다는 것은 "영가설이 진실이고 우연에 의한 결과" 인 100개의 표본에서 하나의 다른 표본을 연구한 것일 수 있다. 그리고 통계적인 검정은 우연의 역할 (role of chance) 만을 평가해준다는 점에 주의하여야 한다.

- selection bias : 연구 그룹 또는 연구 참가자를 등록하는데 있어, 결과적으로 폭로가 질병 위험에 미치는 영향을 잘못 추정하도록 하는 systematic error이다. 즉 연구에 참여하는 대상자에 기인하는 문제이다. 디자인 단계와 실행 단계 모두에서 일어날 수 있다. case-control study의 디자인에서 case 정의가 너무 loose 하거나, controls에서 무 증상 cases를 넣는 경우, 부적절한 control 그룹을 사용한 경우 등이다. 실행 단계에서 특정한 폭로와 질병 특성을 가진 적격한 (eligible) 대상자가 연구에 참가하지 않기로 하거나 또는 찾을 수 없을 경우이다. 예를 들어 문제의 폭로가 된 환자가 연구의 가설을 알고는 다른 환자에 비해 연구에 더 적극적으로 참여하는 경우에 2X2 분할표에서 cell a가 cell c에 비해 인위적으로 커져서 OR가 커지는 결과를 초래하게 된다.

- information bias : 연구 참가자에 대한 폭로 및 결과 자료를 수집하는데 systematic error가 개재되어 폭로가 질병 위험에 미치는 영향을 잘못 추정하게 되는 것이다. 잘못된 wording (설문지의 질문을 잘못 이해하는 것), 잘못된 recall (일주일 전 화요일에 점심으로 먹은 음식은?), 일관성없는 면담 방법, 그리고 만약에 대상자가 알면서 일부로 잘못된 정보를 제공할 경우에 나타날 수 있다.

- confounding : 혼란 작용을 일으킬 수 있는 변수의 리스트를 작성하고, 하나씩 혼란 작용을 일으킬 가능성이 있는지 점검해 본 후, 가능성 있는 혼란 변수를 필요에 따라 통제해 주어야 한다.
- investigator error : 연구자가 계산기를 잘못 두드리거나, 통계프로그램에서 값들을 잘못 recode 하거나, 엉뚱한 식 (또는 통계기법)을 사용하여 계산하여 잘못된 결과를 도출하는 것이다. 흔히 있을 수 있는 error이므로 연구자가 주의해서 점검하거나 또는 제3자가 재현하여 검정하는 과정을 거치는 것이 좋다.

이러한 5가지의 문제가 없다는 것을 확인한 후에도 연관성이 인과성을 갖기 위해서는 다음과

같은 인과관계 평가를 위한 기준을 적용시켜 보고, 조심스럽게 결론을 내려야 한다.

1. 연관의 강도 (Strength of the association)

일반적으로 연관성이 클수록 그 관련성이 진실일 가능성이 더 크다는 것이다. 즉 RR이 9.0일 경우 1.5일 경우 보다 인과관계를 가질 가능성이 더 크다는 것이다. 그러나 RR이 1.5일 경우라도 연관성이 진실일 가능성을 배제할 수 없다. selection bias, information bias, confounding 등으로 인해서 잘못된 결과를 도출할 가능성이 상대적으로 RR이 1.5일 경우에 비해서 9.0일 경우에 더 적어진다는 것이다.

2. 생물학적인 타당성 (Biological plausibility)

나타난 결과가 생물학적으로 맞는 결과인가? 병태 생리학 (pathophysiology), 알려진 vehicles, 질병의 자연사, animal models, 또는 기타 다른 생물학적인 요인과 일관성이 있는가? 그러나 가끔 전혀 새로운 사실이 밝혀질 경우가 있다.

3. 다른 연구와의 일관성 (Consistency with other studies)

기존의 다른 연구와 일관성을 가지는가? 특히 다른 대상자를 사용하여 다른 방법으로 시행한 연구와 일관성을 가지는가가 중요하다.

4. 폭로가 있은 후 질병이 발생 (Exposure precedes disease)

후향적인 연구가 가지는 가장 큰 단점으로 꼽히는 기준이다.

5. 양-반응 효과 (Dose-response effect)

양-반응 관계는 인과관계 정립을 위한 필요조건은 아니다. threshold effect를 나타내지 않는 인과관계가 있을 수 있다. 양-반응관계가 나타난다고 해도 confounding의 가능성을 배제해 주지는 않는다.

대부분의 field 역학에서 위의 5가지 조건을 모두 만족시키는 것은 없다. 시급한 공중보건이 달린 문제일 경우, 모든 조건을 만족시키지 않는다고 개입을 하지 않는 변명이 될 수 없다. George Comstock의 말처럼 "역학적인 추론 (reasoning)의 기교는 불완전한 자료로부터 분별있는 결론 (sensible conclusions)을 내리는 것이다". field 역학은 대중들의 건강을 증진하고 보호한다는 목적을 가지고 과학, 인과관계 추론, 건전한 실제적인 상식에 기초를 두고서 공중보건학적인 행위를 취하기 위한 도구이다.

부록 8-1. FISHER EXACT TEST

$$\Pr(a) = \frac{(v_1)!\,(v_0)!\,(h_1)!\,(h_0)!}{t!a!b!c!d!}$$

where k! ("k factorial") = 1 × 2 x . . . × k,
(e.g., 5! = 1 × 2 × 3 × 4 × 5 = 120)

Example

	H.I.	WELL	TOTAL	
EXPOSED	4	17	21	Odds ratio = undefined (cannot divide by zero)
UNEXPOSED	0	19	19	
TOTAL	4	36	40	

Based on the margins of this two-by-two table, cell a can take on values from 0 to 4, but none more extreme than 4. The probabilities for each value from 0 to 4 are

a	*b*	*c*	*d*	*Probability*	
4	17	0	19	21!36!21!19! / 40!4!17!0!19!	= 0.07
3	18	1	18	21!36!21!19! / 40!3!18!1!18!	= 0.28
2	19	2	17	21!36!21!19! / 40!2!19!2!17!	= 0.39
1	20	3	16	21!36!21!19! / 40!1!20!3!16!	= 0.22
0	21	4	15	21!36!21!19! / 40!0!21!4!15!	= 0.04.

Since there are no possible values of cell a more extreme than 4, the one-tailed *P* value is simply 0.07. The two-tailed *P* value is 0.07 + 0.04 = 0.11. Given a cutoff of 0.05, we could not reject the null hypothesis.

부록 8-2. Chi-square Table

Chi-Square Table

DEGREE OF FREEDOM	PROBABILITY						
	0.50	0.20	0.10	0.05	0.02	0.01	0.001
1	0.455	1.642	2.706	3.841	5.412	6.635	10.827
2	1.386	3.219	4.605	5.991	7.824	9.210	13.815
3	2.366	4.642	6.251	7.815	9.837	11.345	16.268
4	3.357	5.989	7.779	9.488	11.668	13.277	18.465
5	4.351	7.289	9.236	11.070	13.388	15.086	20.517
10	9.342	13.442	15.987	18.307	21.161	23.209	29.588
15	14.339	19.311	22.307	24.996	28.259	30.578	37.697
20	19.337	25.038	28.412	31.410	35.020	37.566	43.315
25	24.337	30.675	34.382	37.652	41.566	44.314	52.620
30	29.336	36.250	40.256	43.773	47.962	50.892	59.703

Note: The Pearson chi-square test and the Yates corrected chi-square test from a two-by-two table have one degree of freedom. The Mantel-Haenszel chi-square also has one degree of freedom, whether from a single two-by-two table or from stratified analysis.