UCSF
University of California
San Francisco

Observational Study Designs III: Less Commonly Used Study Designs

Lydia B. Zablotska, MD, PhD
Professor, Department of Epidemiology and Biostatistics

Highlights from Week #4

- **Discussed design features of case-control studies**
  - Case-control studies as approximation of cohort studies
  - Selection of cases and controls (neighborhood, friends, RDD)
  - Measures of association estimated by c-c studies based on various control sampling methods
  - Matching
    - Case-control matching introduces **selection bias**
    - Matching factor as confounder, as independent risk factor associated only with E or only with D, as a mediator
    - Overmatching or undermatching (unnecessary matching on a variable that is not a confounder)
    - **Key advantage** is to control for confounding more efficiently!

Highlights from Week #4

- **Special types of case-control studies**
  - Nested case-control
  - Case-cohort
  - Two-stage case-control study
  - Counter-matched studies
- **Advantages and disadvantages of the special types of case-control studies**

3

UCSF

Learning Objectives

1. **Case-only design**
   1. Case-specular
   2. Case-crossover
2. **Study designs used in pharmacoepidemiology**
   1. Case-crossover (CXO)
   2. Elf-controlled case series (SCCS)
3. **Study designs used in genetic epidemiology**
   1. Family-based association studies
   2. Case-only Gene x Environment Interaction studies
4. **Cross-sectional study designs**
   1. Standard, case / control
   2. Repeated survey, survey follow-up, intervention follow-up
   3. Proportionate mortality
5. **Ecological Studies**
6. **Comparison of study designs**
7. **Levels of Evidence and their role in Evidence-Based Medicine**

4

UCSF

UCSF

1. Case-only design
Case-Specular Design

- **First developed for study of magnetic waves and cancer**
- **Premise:**
  - To discriminate between the **magnetic field hypothesis**, which postulates that the development of childhood cancer is affected by magnetic fields (wire codes used as proxies for magnetic fields) and the **neighborhood hypothesis**, which postulates that the development of childhood cancer is affected by some characteristics of the neighborhood other than magnetic fields (wire codes used as proxies for these unknown neighborhood characteristics).
- **Specular – relating to a mirror or reflection and describes virtual controls.**
- **Case-specular method overcomes confounding by neighborhood in studies were controls are not matched to cases on neighborhood.**
- **Case-specular method compares the wire code of cases with wire codes of specular residences, in which either the position of the residence or the position of the power line is switched around the center of the street.**

5

UCSF

Assumptions of the Case-Specular Method

1. residence-specular probability matrix is symmetrical.
   - The probability of encountering a residence with actual wire code X and with specular wire code Y is the same as the probability of encountering a residence with actual wire code Y and with specular wire code X. In other words
2. Residences on the same side of the street as power lines are not systematically different from residences on the opposite side.
   - E.g., socioeconomic conditions or other potential risk factors
3. Coding of case residences and their speculars is done in an unbiased (nondifferential) way.

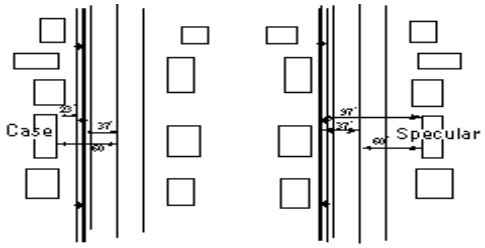
6

UCSF

UCSF

Example 1:
Hypothetical study of electromagnetic fields and childhood cancer

- Use physical properties to distinguish controls' environmental 'exposures'.
- Example: Study of EMF exposure and disease.
  - Measure case's home's distance to electrical wires.
  - Then 'flip' block and measure distance from specular home to electrical wires for 'control's' distance.



7 Zaffanella et al. 1998

UCSF

TABLE 4. Data for Hypothetical Example

Actual	Specular VHCC	OHCC	LCC
VHCC	4	64	8
OHCC	32	19	28
LCC	2	14	123

As a hypothetical example, assume that 294 case-specular pairs coded according to a three-level wire code classification (VHCC, OHCC, and LCC) are distributed as in Table 4. The logistic model in Equation 2 is used with the non-zero $x - z$ vectors, which correspond to the off-diagonal (discordant) pairs in the case-specular matrix.

Case, Specular

	VHCC, OHCC	VHCC, LCC	OHCC, VHCC	OHCC, LCC	LCC, VHCC	LCC, OHCC
$x - z$	(1, −1)	(1, 0)	(−1, 1)	(0, 1)	(−1, 0)	(0, −1)
Number	64	8	32	28	2	14

We calculated odds ratios and their 95% confidence intervals (CI) using a logistic regression program as described above. The estimated hazard ratio for VHCC (*vs* LCC) was 4.0 with 95% CI = 2.01–7.97, whereas the estimated ratio for OHCC (*vs* LCC) was 2.0 with 95% CI = 1.10–3.62.

linked with magnetic field exposure in the home. Five categories are commonly used: very high current configurations (VHCC), ordinary high current configurations (OHCC), ordinary low current configurations (OLCC), very low current configurations (VLCC), and underground wiring (UG). As described in Table 1, these categories are characterized by the type or class of power line, which relates to its potential to carry current, and by the distance to the residence. Both class and distance are estimated by visual inspection.

LCC – a three-level code that combines UG, VLCC, and OLCC

- Analyzed as a matched pair case-control data using conditional logistic regression

8 Zaffanella et al. 1998

UCSF

UCSF

Example 2:
Interphone study: Case-control study

Methods An interview-based case–control study with 2708 glioma and 2409 meningioma cases and matched controls was conducted in 13 countries using a common protocol.

Results A reduced odds ratio (OR) related to ever having been a regular mobile phone user was seen for glioma [OR 0.81; 95% confidence interval (CI) 0.70–0.94] and meningioma (OR 0.79; 95% CI 0.68–0.91), possibly reflecting participation bias or other methodological limitations. No elevated OR was observed ⩾10 years after first phone use (glioma: OR 0.98; 95% CI 0.76–1.26; meningioma: OR 0.83; 95% CI 0.61–1.14). ORs were <1.0 for all deciles of lifetime number of phone calls and nine deciles of cumulative call time. In the 10th decile of recalled cumulative call time, ⩾1640 h, the OR was 1.40 (95% CI 1.03–1.89) for glioma, and 1.15 (95% CI 0.81–1.62) for meningioma; but there are implausible values of reported use in this group. ORs for glioma tended to be greater in the temporal lobe than in other lobes of the brain, but the CIs around the lobe-specific estimates were wide. ORs for glioma tended to be greater in subjects who reported usual phone use on the same side of the head as their tumour than on the opposite side.

Conclusions Overall, no increase in risk of glioma or meningioma was observed with use of mobile phones. There were suggestions of an increased risk of glioma at the highest exposure levels, but biases and error prevent a causal interpretation. The possible effects of long-term heavy use of mobile phones require further investigation.

- There was some indication of risk of glioma in the top 10% of usage (OR=1.40), which corresponds to more than 100,000 minutes of cell phone use. If a person has used it for 10 years, that would correspond to 30 minutes a day, every day.

9 Cardis et al. for INTERPHONE Int J Epidemiol 2010

UCSF

One Conclusion Emerges From Interphone Study: Controversy Will Continue

By Judy Peres

The 13-nation Interphone study was supposed to establish, once and for all, whether a link between cell phone use and brain tumors exists. But after 10 years and more than $25 million, the only clear message to emerge was that more research was needed.

The study, which the International Agency for Research on Cancer sponsored and the *International Journal of Epidemiology* reported last month, spawned confusing press reports. The *Evening Times* (Glasgow) trumpeted, "Phones not linked to cancer," while the *Sunday Times* (London) declared, "Heavy mobile users risk cancer"—and other headline writers tried to take positions between the extremes.

If the lay press didn't know what to make of the results, the Interphone researchers also seemed to have trouble agreeing on what the study showed.

"We have not demonstrated an increased risk of brain cancer, but I think it's important to point out that we also have not demonstrated that there's no risk."

Maria Feychting, M.D., Ph.D., of the Karolinska Institute in Stockholm, an Interphone study group member, said in a press release, "The use of mobile phones for over 10 years shows no increased risk of brain tumors." But principal investigator Elisabeth Cardis, Ph.D., of the Centre for Research in Environmental Epidemiology in Barcelona, told *Microwave News*, "To me, there's certainly smoke here. Overall, my opinion is that the results show a real effect."

In an interview with JNCI 2 weeks later, Cardis was more careful: "We have not demonstrated an increased risk of brain cancer," she said. "But I think it's important to point out that. . . we also have not demonstrated that there's no risk."

Range of Interpretation

Cardis acknowledged "a range of interpretation among the investigators about whether there is an effect [of cell phone use on the risk of brain tumors] or whether the results are related to bias. . . . Some investigators think there is probably no risk, and the increased odds ratios are just artifacts. Others, on the contrary, think many of the odds ratios may be underestimated. All of these interpretations are legitimate."


Elisabeth Cardis, Ph.D.

Jack Siemiatycki, Ph.D., of the University of Montreal, who directed one of the Canadian Interphone centers, concluded, "For the great majority of users—about 90%—the message is rather encouraging. There was absolutely no evidence of any risk. There is some doubt in the top 10% of usage, which corresponds to more than 100,000 minutes of cell phone use. If a

10

UCSF

UCSF

Devilish Details

The Interphone study has neither reassured the concerned nor convinced skeptics of a link between cell phone use and risk of brain cancer. The two sides interpret the findings differently, pointing to details of study design, statistical analysis, and trends in cell phone usage. Among the arguments on each side:

Those who think the study supports a link argue that:

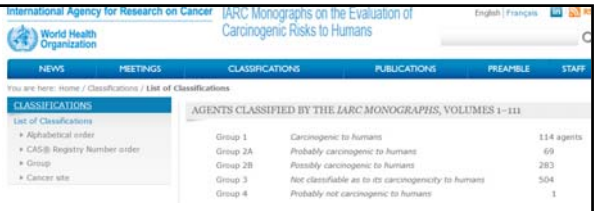
- The study found several odds ratios of less than 1 (meaning that cell phone users had fewer brain tumors than nonusers). Most analyses of the data concluded that assuming that cell phones are actually protective was implausible and, therefore, attributed the results to one or more sources of bias. That assumption implied that the few findings of excess risk could be underestimated. "The increased risk is all the more significant against the background that the study is affected by downward bias," said Rodolfo Saracci, M.D., of the National Research Council in Pisa, who cowrote a commentary on the Interphone report in the *International Journal of Epidemiology*.
- Several previous studies, especially by the Swedish research group led by Lennart Hardell, M.D., found that cell phone users did have a higher risk of malignant brain tumors. The risk was greatest for tumors on the same side of the head where the phone was typically held and was especially high for those who had started using their cell phones when they were younger than 20 years.
- Typical cell phone use today is much greater than 10 years ago, when the Interphone cases were recruited. "The average user today would fall into the high-use, high-risk group in about 13 years," said Berkeley's Joel Moskowitz, Ph.D.
- Prevalence of cell phone use has increased explosively, with many people using them instead of land lines. An estimated 5 billion cell phones are in use worldwide. So, if cell phone use were carcinogenic, it could reach pandemic proportions.
- The effect of a carcinogen can take 20–30 years of exposure to show up, making even some skeptics admit that more substantial evidence might yet emerge.
- The increased risk that the Interphone study found was greater for tumors on the same side of the head (although laterality was established by self-report, which allows for another source of bias).
- Perhaps most important, "You can't prove a negative," as Robert Tarone, Ph.D., of the International Epidemiology Institute put it.

Those who don't think the study supports a link argue that:

- The only statistically significantly increased risk estimate (odds ratio [OR] = 1.40) in the main analysis showed up in the top decile of cumulative cell phone use. But data of reported use in this group included some implausible values: 10 case patients (and no control subjects) said that they used their cell phones for 12 or more hours per day. The first nine deciles showed no upward trend. The odds ratio was significantly decreased (OR = 0.71) in the ninth decile. And, even in the top decile, the increased risk was not in the longest-use group but in the short-term users: those who had started using a cell phone 1–4 years before their recruitment. "This looks more like random variation than like a dose–response effect," said Tarone. "If you had a carcinogen, you'd expect to see increasing risk with increasing exposure."
- No known biological mechanism to explain any increased risk is evident.
- The subset analysis that showed the biggest odds ratios was arguably inappropriate. In this analysis, reported not in the main study but in an appendix, the researchers compared more-frequent cell phone users to minimal users (rather than to nonusers). Saracci believes that this analysis was "logically justified," of the type that occupational epidemiology often uses to eliminate a source of bias. But Berry was skeptical. "This is the kind of thing people do when they don't have a positive study—they look for things that are positive. But it raises the bar. If you now show an effect, the results have to be really compelling, and they're not."

JNCI | News 929

UCSF



International Agency for Research on Cancer

World Health Organization

PRESS RELEASE
N° 208

31 May 2011

IARC CLASSIFIES RADIOFREQUENCY ELECTROMAGNETIC FIELDS AS POSSIBLY CARCINOGENIC TO HUMANS

Lyon, France, May 31, 2011 -- The WHO/International Agency for Research on Cancer (IARC) has classified radiofrequency electromagnetic fields as possibly carcinogenic to humans (Group 2B), based on an increased risk for glioma, a malignant type of brain cancer[1], associated with wireless phone use.

UCSF

UCSF

Summary

Carcinogenicity of Cell Phone Radiation: 2B or not 2B…

Well

Cellphone Radiation May Cause Cancer, Advisory Panel Says

MICRO WAVE NEWS

A Report on Non-Ionizing Radiation

Home | News Center | About Us | EMF/EMR Meters | EMF/EMR Directory

IARC: Cell Phone Radiation Is a Possible Human Carcinogen

Small Group Will File Minority Opinion

Search

Return to Mai

UCSF

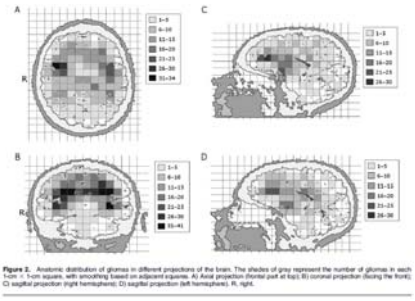
Example 2:
Interphone study: Case-only analysis

Table 1. Location of Gliomas in the Hemispheres of the Brain in Relation to the Self-Reported Side of Mobile Telephone Use Among Glioma Patients From 7 European Countries, 2000–2004[*][a]

Side of Glioma	Laterality of Phone Use		
	Right	Left	Total
Right	160	47	207
Left	123	86	209
Total	283	133	416

[*] *P* for heterogeneity < 0.001.

[a] Statistical significance for heterogeneity was evaluated on the basis of Fisher's exact test.



- Tumor laterality coincided with laterality of phone use.
- A statistically significant excess of gliomas on the self-reported side of mobile phone use (prone to recall bias)

UCSF

UCSF

Example 2: Interphone study: Case-only analysis

- **To avoid recall bias:**
  - Distance to the nearest source of exposure ("exposure line" from the ear canal to the corner of the mouth, 6.7 cm) on the same side of the head as glioma, irrespective of patient's reported typical side of phone use
- **Statistical analysis:**
  - Mann-Whitney test to compare distances of glioma from the exposure line, unconditional logistic regression

Table 4. Distance Between Glioma and Source of Mobile Telephone Exposure, in Relation to Exposure Variables, Among Glioma Patients From 7 European Countries, 2000–2004

	Mean Distance[a], cm	Distance From Exposure[a]				*P* for Heterogeneity
		≤5 cm		>5 cm		
		No.	%	No.	%	
Glioma cases	6.25	200	23	688	77	
Regularity of mobile phone use[b]						0.39
Regular use	6.29	107	22	388	78	
Never-regular use	6.19	91	24	287	76	
Cumulative call time, hours[c]						0.41
0.001–46	6.29	33	21	125	79	
47–339	6.27	38	25	114	75	
>339	6.36	30	19	129	81	
Duration of use, years[d]						0.82
1.5–4	6.31	65	21	239	79	
5–9	6.28	30	21	112	79	
≥10	6.38	10	24	32	76	
Laterality[e]						0.80
Ipsilateral	6.37	51	21	195	79	
Contralateral[f]	6.29	37	22	133	78	
Speculars	6.24	166	19	722	81	0.71
Ipsilateral speculars	6.26	47	19	199	81	
Contralateral speculars[f]	6.36	30	18	140	82	

[a] Distance between the midpoint of the glioma and the presumed source of exposure.
[b] Information was missing on 15 cases.
[c] Cumulative call time <0.001 hour, never-regular use, or missing information for 419 cases.
[d] Regular use <1.5 years, never-regular use, or missing information for 400 cases.
[e] Use on both sides, glioma located centrally, or missing information on both sides for 472 cases.
[f] Distance was calculated to the closest (ipsilateral) exposure line despite the knowledge of contralaterality.

15 Larjavaara et al. Am J Epidemiol 2011

UCSF

Example 2: Interphone study: Case-only analysis

Table 5. Odds Ratio for a Distance of ≤5 cm Between the Glioma Midpoint and the Typical Source of Mobile Telephone Exposure in All Regular Mobile Phone Users Compared With Never-Regular Users (Case-Case-Analysis) Among Glioma Patients From 7 European Countries, 2000–2004

Exposure Characteristic	Crude OR	95% CI	Adjusted OR[a]	95% CI
Frequency of use (regular use)	0.87	0.63, 1.20	0.80	0.56, 1.15
Cumulative call time, hours				
0.001–46	0.82	0.52, 1.29	0.82	0.51, 1.31
47–339	1.04	0.67, 1.60	0.97	0.60, 1.56
>339	0.72	0.46, 1.15	0.58	0.35, 0.96
Laterality of use				
Ipsilateral	0.82	0.56, 1.21	0.80	0.52, 1.22
Contralateral	0.87	0.56, 1.34	0.77	0.47, 1.24
Duration of use, years				
1.5–4	0.86	0.60, 1.23	0.85	0.57, 1.25
5–9	0.84	0.53, 1.35	0.71	0.43, 1.18
≥10	0.99	0.47, 2.08	0.85	0.39, 1.86

Abbreviations: CI, confidence interval; OR, odds ratio.
[a] Adjusted for age, education, sex, and country.

- Tumors were located closest to exposure among never-regular and contralateral users (non-significant)
- **No excess risk** in the highly exposed parts among regular users versus never-regular users were found
- All of the odds ratios for the higher categories of intensity or duration of mobile phone use were below unity in these analyses, indicating **no excess risk in the highly exposed parts among regular users versus never-regular users**, although all of the upper confidence limits were above 1.

16

UCSF

UCSF

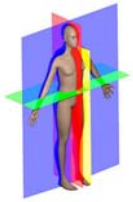
Example:
Interphone study:
Case-specular analysis

Original Contribution

Location of Gliomas in Relation to Mobile Telephone Use: A Case-Case and Case-Specular Analysis

Suvi Larjavaara*, Joachim Schüz, Anthony Swerdlow, Maria Feychting, Christoffer Johansen, Susanna Lagorio, Tore Tynes, Lars Klaeboe, Sven Reidar Tonjer, Maria Blettner, Gabriele Berg-Beckhoff, Brigitte Schlehofer, Minouk Schoemaker, Juliet Britton, Riitta Mäntylä, Stefan Lönn, Anders Ahlbom, Olof Flodmark, Anders Lilja, Stefano Martini, Emanuela Rastelli, Antonello Vidiri, Veikko Kähärä, Jani Raitanen, Sirpa Heinävaara, and Anssi Auvinen

- Side of use was ignored
- A hypothetical reference location was assigned for each glioma, and the distances from the actual and specular locations to the handset were compared
- Specular controls were mirror images on the opposite side of the same hemisphere in terms of sagittal and coronal axes



Anatomical planes in a human:
- median or sagittal plane
- a parasagittal plane
- frontal or coronal plane
- transverse or axial plane

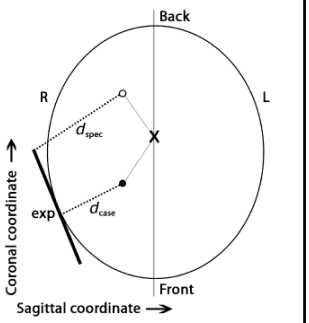


Figure 1. Schematic representation of assignment of the coronal coordinates for a specular analysis of mobile phone use and glioma risk. In the axial projection (*x*-axis, sagittal coordinate; *y*-axis, coronal coordinate), the midpoint for a case is indicated with a solid circle and the corresponding specular location is indicated with an open circle. The distance from the source of exposure (exp) is denoted by *d* separately for the case (d_{case}) and for the specular location (d_{spec}). Axial coordinates were obtained in a similar fashion using a coronal projection. R, right; L, left.

17 Larjavaara et al. 2011

UCSF

Example:
Interphone study:
Case-specular analysis

Table 6. Odds Ratio for the Distance Between the Glioma Midpoint and the Typical Position of the Mobile Telephone, Measured as Both a Categorical Variable (≤5 cm) and a Continuous Variable, in All Cases Compared With Speculars (Case-Specular Analysis) Among Glioma Patients From 7 European Countries, 2000–2004

Exposure Characteristic	Glioma Midpoint ≤5 cm (vs. >5 cm) From Mobile Phone		Increasing Distance From Exposure Point, per cm	
	OR	95% CI	OR	95% CI
Total (all cases vs. all speculars)	1.22	0.99, 1.51	1.00	0.95, 1.07
Never-regular users				
Regular use	1.19	0.89, 1.59	0.99	0.92, 1.08
Never-regular use	1.30	0.95, 1.80	1.01	0.92, 1.11
Cumulative call time, hours				
0.001–46	1.39	0.81, 2.38	1.00	0.87, 1.16
47–339	1.21	0.74, 1.97	0.99	0.86, 1.13
>339	1.00	0.59, 1.69	1.01	0.88, 1.16
Duration of use, years				
1.5–4	1.15	0.80, 1.66	0.98	0.89, 1.09
5–9	1.04	0.61, 1.76	1.02	0.89, 1.18
≥10	2.00	0.68, 5.85	1.08	0.82, 1.42

- The mean distances between exposure line and tumor location were similar for cases and speculars.
- Glioma cases were closer to the exposure line in long-term users, but the differences remained nonsignificant.

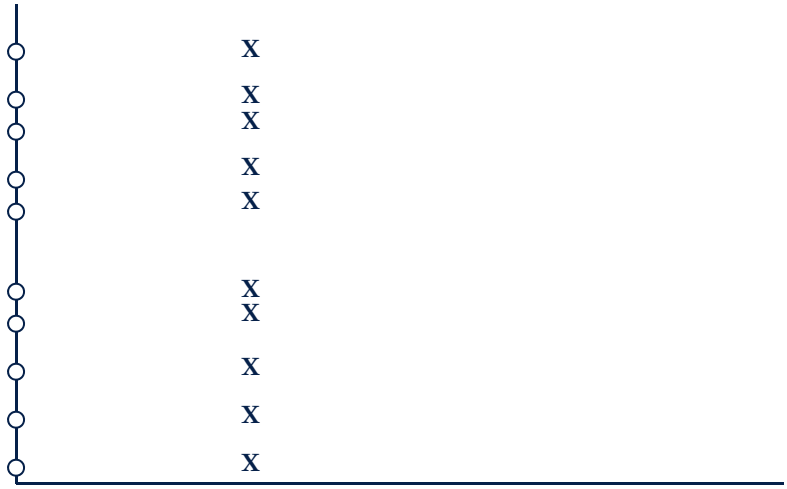
18 Larjavaara et al. Am J Epidemiol 2011

UCSF

UCSF

1. Case-only design

Case-Crossover Study



19

UCSF

1. Case-only design

Case-Crossover Study

- One or more time periods are selected as matched "control" periods for the case.
- Compare exposure status at the time of disease onset to the 'control' exposure status within the same individual.
- Depends on the assumption that neither exposure nor confounders are changing over time in a systematic way.
  - i.e. cyclic manner
- Exposure must vary over time within individuals
- Exposure must have a short duration, short induction period and a transient effect

20

UCSF

UCSF

Example:
Physical Exertion & MI

- A number of different exposure periods can be measured.
- One might also use a bidirectional approach to measuring exposures.
- The choice of control time periods is governed by the availability and representativeness of the exposure data for the control period

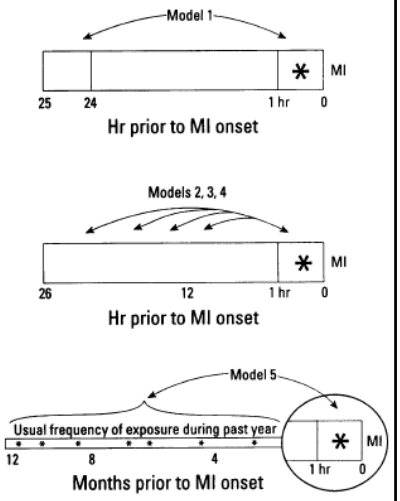


Figure 3. Examples of three different approaches for the choice of controls in a unidirectional case–crossover design. (*A*) Single control time intervals. (*B*) Multiple control time intervals. (*C*) Usual frequency over some predefined interval prior to the case interval. Data from

21 Tager, 2000

UCSF

Strengths and Weaknesses of Case-Crossover Studies?

Table 4. Case-only studies.

Type of study	Features of sampling of control exposure	Rationale	Potential threats to validity (sources of bias)
Case–crossover			
Unidirectional	Selected from relevant time period(s) prior to case exposure period	Controls for subject confounders that remain fixed over time (e.g., sex, ethnicity)	General and confounder-specific time trends in exposure Applies to both uni- and bidirectional sampling
Bidirectional	Selected from relevant time periods before and after case exposure period	Potential to control for temporal trends in exposure	Time trends in confounder–exposure risks Carryover effects of exposure Case selection bias Information bias
Case–specular (developed specifically to study effects of *magnetic* fields)	Defined by wire code of hypothetical residences located in a virtual situation in which either the position of the power line is switched around the center of the street of the case residence	Avoids bias in control selection in relation to important neighborhood characteristics	Failure of the assumption of symmetrical probably of a given wire code Bias in wire coding

22 Tager, 2000

UCSF

UCSF

2. Pharmacoepidemiology study designs

Case-crossover (CXO) and self-controlled case series (SCCS)

- CXO depend on on strong assumptions, being suitable for transient exposures with short term effects so that inter- mittency of drug use and the length of the exposure time window may have an impact on the estimates obtained
  - a conditional logistic regression model is employed with more than one control mo- ment
- SCCS follows the cohort design approach with an assumption that the occurrence of the event of interest does not influence the chance of subsequent exposure
-

UCSF

Example

- To examine the consistency of relative risk estimates of hip/femur fractures (HFF) associated with the use of benzodiazepines (BZD) across case-only designs in two databases, when a common protocol was applied.
- Exposure to BZD was divided into non-use, current, recent and past use.
- Results:
  - conditional logistic regression was used
  - For CXO, current use of BZD was associated with an increased risk of HFF.
  - In the SCCS, IRRs for current exposure differed in direction in the two databases.
  - Pre-exposure time risk window of 30 days (an incident HFF has a short-term impact on the likelihood of being prescribed a BZD).
  - Analyses removing this period produced similar results in both databases.

 Requena et al. 2016

UCSF

UCSF

3. Genetic Epidemiology Study Designs
Case-only design

- The strength of the association between genes and diseases so far have been weak, and a large proportion of familial clustering of disease remains unexplained
- A combination of genes, or a combination of genes with one or more environmental exposures, may account for some of this 'missing' heritability.
- Answer: harmonization of existing and planned large-scale cohort studies. Examples: ***InterLymph, Genome-Wide Association Study for Pancreatic Cancer (PanScan3), GWAS of Renal Cell Carcinoma (RCC), Lung Cancer Cohort Consortium (LC3), Breast and Prostate Cancer Cohort Consortium (BPC3)***
- *Advantages of case-only design for G-E interactions:*
  - Smaller sample sizes required to detect interaction between genes and environmental exposures
  - when it is difficult to recruit control subjects
  - when, in a multicenter study, controls are not comparable among centers because of different methods of recruitment and/or exposure assessment

3. Genetic Epidemiology Study Designs
Case-only design

- **Assumptions:**
  - Assume independence of G-E in source population
  - Assume genotypes are independent in source population
  - Can't look at main effects.
  - Can only look at multiplicative departures.
- **In practice:**
  - Mendelian randomization: alleles will segregate randomly in the population, regardless of environmental exposures
  - Check this in controls to see if the assumption holds in the source population

3. Genetic Epidemiology Study Designs

- Use theoretical considerations to construct a distribution of exposure in the source population.
- Use this distribution in place of an observed control series.
- Use laws of inheritance and other assumptions in study designs (e.g., gives a population of genotypes).
- Hardy-Weinberg Principle:
  - Genotypes will reflect allele frequency distributions in the general population.
- I.e., frequencies in a population remain constant—they are in equilibrium—from generation to generation unless specific disturbing influences are introduced.

3. Genetic Epidemiology Study Designs
Case-Only GxE Studies

- Search for gene-gene or gene-environment interactions.
- OR among cases only estimates departure from multiplicative joint effect of G and E.

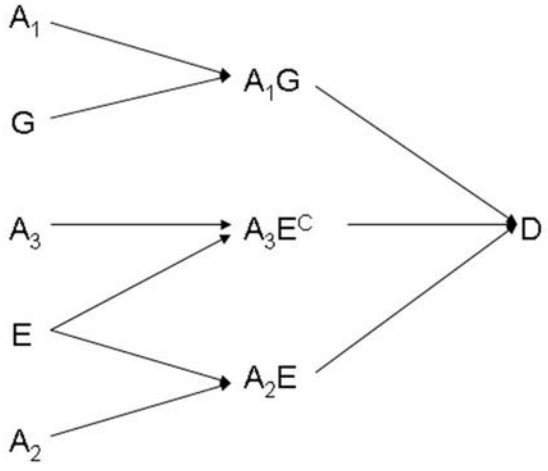
	E+		E-	
	G+	G-	G+	G-
Case	A_{11}	A_{10}	A_{01}	A_{00}
Control	B_{11}	B_{10}	B_{01}	B_{00}

Case-only GxE associations do not always imply mechanistic interaction

- VanderWeele et al. showed that association amongst cases can arise between the genetic and environmental factors when there is in fact no mechanistic gene-environment interaction.
- When it can be assumed the genetic and environmental factors themselves can never prevent the outcome (monotonicity assumption), a positive association amongst cases implies a mechanistic gene-environment interaction. Without this assumption that the effects of the two factors are never preventive, a multiplicative interaction greater than 2 is needed to conclude the presence of a mechanistic interaction.

29 VanderWeele et al. *Genet Epidemiol* 2010. UCSF

Multiplicative interaction in case-only study with no mechanistic interaction



30 VanderWeele et al. *Genet Epidemiol* 2010. UCSF

UCSF

Example

- Milne et al. 2006 studied interaction between maternal folate supplementation and child's methylenetetrahydrofolate reductase (MTHFR) gene polymorphisms among infants with acute lymphoblastic leukaemia (ALL).
- Both folate supplementation and the MTHFR polymorphisms have been associated with a protective effect for ALL.
- The case-only OR for MTHFR C677T genotype and folate supplementation was 1.25 (95% CI: 0.31–5.10). In addition to statistical instability due to small sample sizes, this OR could be explained by the possibility the effect of this gene is causative for some individuals and preventive for others.
- Even if the case-only estimate of 1.25 had been statistically significant, conclusions could not be drawn about mechanistic interaction unless the monotonicity assumption held and effects of MTHFR gene and folate supplementation were in the same direction for all individuals.

3. Genetic Epidemiology Study Designs

Comparison of Designs

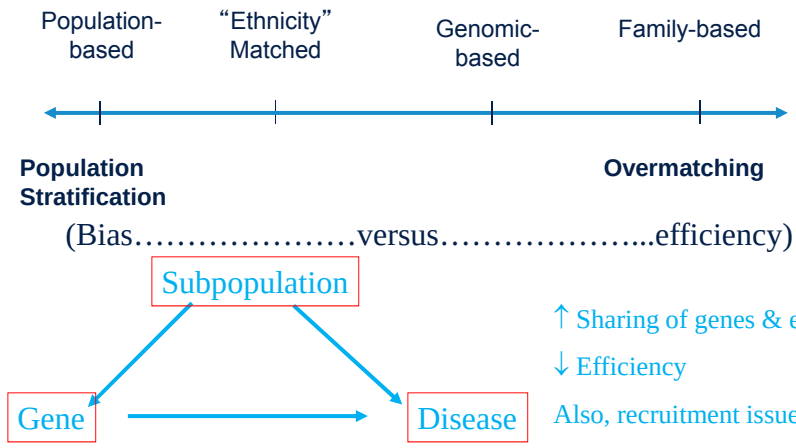
• Family-based designs can be less efficient than population-based designs.

	Rare Recessive High Risk	Common Low Risk	Rare Dominant High Risk
Population-based	100%	100%	100%
Case-sibling	69%	51%	50%
Case-cousin	97%	88%	88%
TDT	231%	102%	101%

• Further, family-based designs can require more recruitment efforts.

 Witte et al. 1999

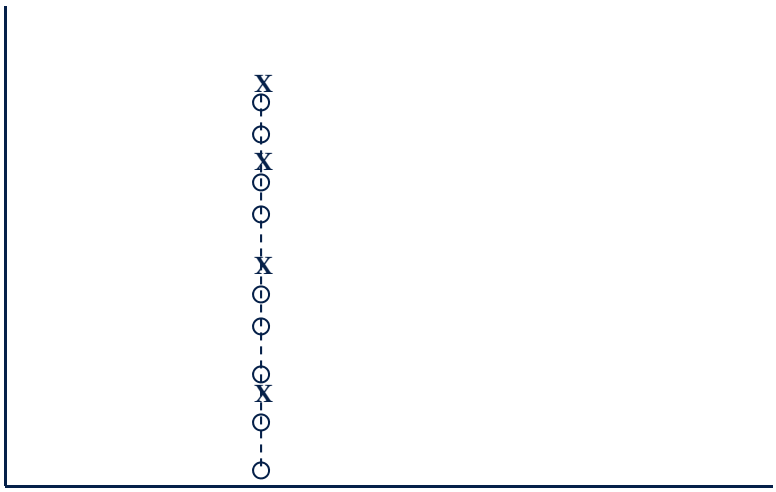
Continuum of Genetic Epi Study Designs



33

UCSF

4. Cross-sectional study design



34

UCSF

UCSF

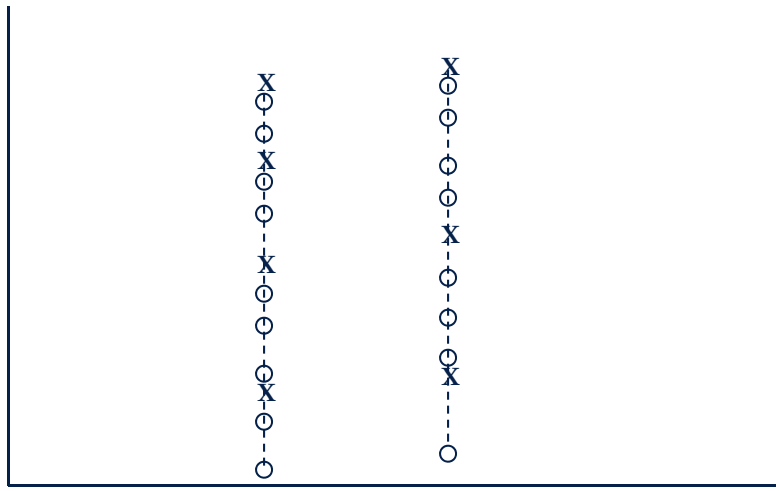
4. Cross-sectional studies

- Subjects all in population at the time of ascertainment or a representative sample.
- Often exposures that can not change:
  - E.g., blood type or other invariable personal characteristics.
- Cross-sectional analyses of baseline information in cohort studies provides possible exposure-disease associations that can later be confirmed.
- Cases in a cross-sectional study will over represent cases with a long duration of illness and under represent those with a short duration of illness.

4a. Case or control cross-sectional studies

- Use cases only to determine estimates of disease prevalence etc. among different groups
  - (e.g., defined by geographical region).
- Use controls to estimate exposure prevalence in population.

4b. Repeated survey



37

UCSF

4b. Repeated survey

- Combine >=2 cross-sectional studies of the same source population at different times.
- Individuals are not followed (even though population is).
- For testing etiologic hypotheses not much better than a basic cross-sectional study.
- Use this design to:
  - Study population trends
  - Evaluate effectiveness of population interventions initiated between surveys
- Impact of changes in exposures on disease rates.

38

UCSF

UCSF

4c. Survey follow-up

- Cross-sectional study followed by a cohort study of people still at risk for disease.
- Use this design to
  - Estimate prevalence and incidence rates in the same source population
  - Identify persons still at risk of developing the disease

39

UCSF

4d. Intervention follow-up

- Combines an intervention with a cohort study
  - Each has different follow-up period and outcome variable.
- 1st follow-up period short
  - used to assess the effect of an intervention / exposure on an outcome (not primary disease).
- 2nd follow-up period is generally longer
  - used to observe disease occurrence.
- Good for examining relationships between acute biological / behavioral responses & chronic health effects.

40

UCSF

UCSF

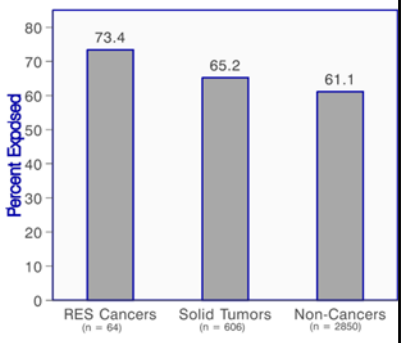
4e. Proportionate study

- Special type of case-control (or cross-sectional) study.
- A group of individuals with (or dying from) the index disease of interest is compared with a group of individuals with (or dying from) certain other diseases.

41

Example:
Proportionate mortality study

- Occupational exposure to low-level ionizing radiation on cancer.
- All certified deaths among employees of the Hanford nuclear power facility between 1944 and 1972 were classified by cause of death and exposure status
  - based on company records of radiation monitoring.
- Proportion of deaths that were exposed to ionizing radiation (>= 1 + badge exposure) among 3,520 male employees, by cause of death
  - RES Cancers include lymphomas, myelomas, and leukemias.



42 Mancuso et al. Health Physics 1977

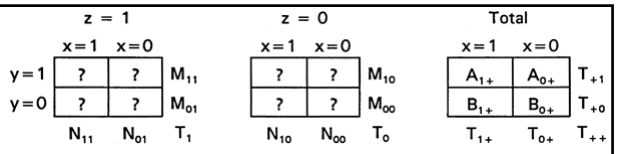
5. Ecological studies

Levels of Measurement

- Aggregate measures: summaries of observations from individuals in each group.
  - e.g., means
- Environmental measures: physical characteristics of the place in which members of each group live or work
  - e.g., air pollution level, hours of sunlight
- Global measures: attributes of groups, organizations, or places for which there is no distinct analogue at the individual level
  - e.g., population density, level of social disorganization, existence of a specific law, or type of health-care system

UCSF

5. Ecological



y – Disease; x – Exposure; z - Covariates

RGL fig. 25-1

Levels of Analysis

- The common level for which data on all variables are reduced and analyzed.
- In an individual analysis every subject has a value assigned to each variable (e.g., predictor, such as air pollution, may be ecologic) → we know all cells with "?"
- In a complete ecologic analysis, all variables are ecologic measures; We only know the marginal distributions of each variable → know only T
- In a partially ecologic analysis, additional information is available on certain joint distributions → we know M, N, or A/B
- Multilevel analysis - What type of analysis is this and what do we know?

UCSF

UCSF

5. Ecological study designs

Distinguished by:

- Exposure measurement
  - Exploratory
  - Etiologic
- Grouping
  - Place
  - Time
  - Mixture

5. Ecological study designs

Multiple-group design

Exploratory studies

- The rate of disease is compared among many groups during one period of time to search for spatial patterns.

Example: NCI cancer study (by county).

- Nearby regions tend to be more similar because unmeasured risk factors tend to cluster in space.
- Regions with the smallest number of cases show greater variability and tend to have the most extreme rates.

Etiologic studies

Example: Association between average exposure (census tracts) and childhood cancers around nuclear power plants

5. Ecological study designs

Multiple-group design

Exploratory studies

Example: Migrant study

Compare the rate of disease may between migrants / offspring and residents of the countries of immigration and emigration.

- e.g., Environmental or behavioral risk factors, Genetic risk factors

Cancer Site	Japanese			
	Japan	Not US Born	US Born	US Caucasians
Stomach (M)	100	72	38	17
Colorectal (F)	100	218	209	483
Breast	100	166	136	591

47 MacMahon B, Pugh TF. Epidemiology. 1970:178.

UCSF

5. Ecological study designs

Time-trend Design

- Follow one group or population over time to assess relationship between a change in exposure frequency and a change in disease frequency.
  - Example: NCI study of artificial sweetener consumption and bladder cancer.
- Potential Issues
  - Changes in disease classification over time
  - Changes in treatment (e.g, for mortality)
  - Latent periods between exposure to risk factor and disease

48

UCSF

UCSF

5. Ecological study designs

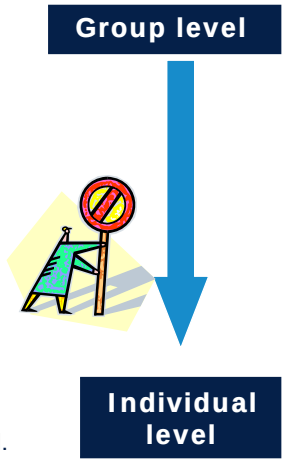
Mixed (time-trend) Design

- A mixture of the two previous designs.
- A number of groups or populations are followed over time to assess a possible association between a change in exposure frequency and a change in disease frequency.
- Example: Change in annual CVD mortality rate for males between 1948 and 1964 in 83 British towns by age and water level hardness.

UCSF

Ecological fallacy:

- Fallacy of drawing inferences regarding associations at the individual level based on the group-level data
  - **The group-level data:**
    - inverse linear relationship between alcohol consumption and CHD mortality
    - Those who consume large quantities of alcohol have the smallest mortality
  - **The individual-level data:**
    - relationship is J-shaped
    - non-drinkers and those who consume large quantities have higher mortality than those who consume small to moderate amounts of alcohol.



UCSF

UCSF

Ecological fallacy, *continued:*

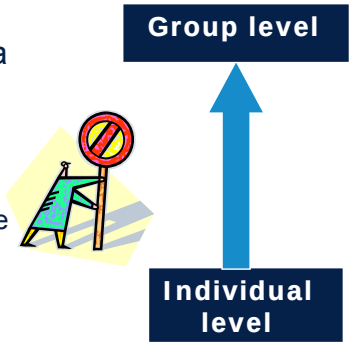
- This does not mean that every ecological study has *ecological fallacy*!
- The importance of the *ecological fallacy* may differ for different research questions
- Potential strategies to reduce *ecological fallacy:*
  - **Use smaller units to make groups more homogeneous**
  - **Supplement ecological variables with individual-level variables**

51

Atomistic fallacy:

- drawing inferences at a higher level from analyses performed at a lower level
- **Example:**

– in a case-control collect information on various possible exposures but ignore the geographic, spatial, and social context in which a person lives



52

Limitations of ecologic analysis

- ***Ecologic bias***
- **Inability to control for confounders**
  - Criteria for confounding are not the same as in individual-level studies
- **Within-group misclassification**
  - Non-differential misclassification of a dichotomous E within groups leads to bias AWAY from the null
- **Lack of adequate data**
- **Problems with temporality**
- **Collinearity**

UCSF

In-Class Exercise

OPEN ACCESS Freely available online

PLOS ONE

The Relationship of Sugar to Population-Level Diabetes Prevalence: An Econometric Analysis of Repeated Cross-Sectional Data

Sanjay Basu[1*], Paula Yoffe[2], Nancy Hills[3], Robert H. Lustig[4,5]

- **Q1: What types of measures were used?**
- **Q2: What type of design was used?**
- **Q3: What type of analysis was performed?**
- **Q4: Do you agree with this conclusion:**

"The observed relationship between dietary sugar exposure and diabetes in this statistical assessment was not mitigated by adjusting for confounders related to socioeconomics, aging, physical activity, or obesity. This suggests that sugar should be investigated for its role in diabetes pathogenesis apart from its contributions to obesity."

 Basu et al. 2013

UCSF

UCSF

In-Class Exercise Answer

OPEN ACCESS Freely available online PLOS ONE

The Relationship of Sugar to Population-Level Diabetes Prevalence: An Econometric Analysis of Repeated Cross-Sectional Data

Sanjay Basu[1*], Paula Yoffe[2], Nancy Hills[3], Robert H. Lustig[4,5]

- **Q1: What types of measures were used?**

Answer: Both E and D were ecological variables of aggregate type, e.g., market availability of different food items (sugars, fruits, etc.) in kcal/person/day in each country for each year

- **Q2: What type of design was used?**

Answer: **Multiple-group design, exploratory study**

- **Q3: What type of analysis was performed?**

Answer: Completely ecologic analysis

- **Q4: Do you agree with this conclusion:**

Answer: The authors did not explain the intended level of inference.

"The observed relationship between dietary sugar exposure and diabetes in this statistical assessment was *not mitigated by adjusting for confounders related to socioeconomics, aging, physical activity, or obesity*. This *suggests that sugar should be investigated* for its role in diabetes pathogenesis apart from its contributions to obesity."

55

UCSF

American Journal of Epidemiology
© The Author 2011. Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.
Vol. 174, No. 10
DOI: 10.1093/aje/kwr241
Advance Access publication:
September 22, 2011

Special Article

Brenden Dufault* and Neil Klar

The Quality of Modern Cross-Sectional Ecologic Studies: A Bibliometric Review

Recommendations for reporting the results of ecological studies

Appendix Table 1. Suggested Additions to the STROBE Document for Ecologic Reporting[a]

STROBE Section	Recommendation Addition(s)
Title and abstract	Indicate the study's design with a commonly used term in the title or abstract.
	Ecologic study—Use one of the following terms in your description: "ecological," "ecologic," or "aggregate."
Introduction	
Objectives	State specific objectives, including any prespecified hypotheses.
	Ecologic study—Specify the level(s) of inference to which your investigation is intended to apply.
Methods	
Study design	Present key elements of the study design early on in the paper.
	Ecologic study—Provide a rationale or justification for the design, taking your study objectives into account.
Participants	Ecologic study—Give the eligibility criteria for the ecologic units, as well as their sources and any sampling methods used. Provide a rationale for their selection. If the units were aggregated for the study, explain how this was done and the criteria by which they were constructed.
Discussion	
Limitations	Discuss limitations of the study, taking into account sources of potential bias or imprecision.
	Ecologic study—Provide a discussion that either informs the reader of the potential for cross-level bias or explains why it is not a possibility for your results and inferences.

Abbreviation: STROBE, Strengthening the Reporting of Observational Studies in Epidemiology.
[a] Patterned after the STROBE statement (17).

56

UCSF

UCSF

6. Criteria for comparing study designs

Three general criteria:

1. Relevance of the information to the investigator
   - Extent to which expected findings will satisfy the specific objectives of the study
   - Investigator's desire to estimate specific population parameters

UCSF

6. Criteria for comparing study designs

Three general criteria:

2. Quality or accuracy of the information expected in the data
   - Ability of the investigator to determine that the exposure preceded disease occurrence
   - Ability of the investigator to eliminate the possibility that the statistical findings were due to various methodological problems or sources of error

UCSF

UCSF

6. Criteria for comparing study designs

Three general criteria:

3. Cost of the information

– The ultimate worth of a study is the total value of all derived information--now and in the future--relative to the total (direct and indirect) costs of the study

UCSF

6. Comparisons between case-control and cohort study designs

Characteristics	Case-control	Cohort study
Select subjects based on	Disease status	Exposure status
Directionality	Always backward	Always forward
Timing	Contemporary	Retrospective/prospective
Exposure	Good for common exposures	Good for rare exposures
Examine >1 exposure	Yes	Generally yes (think Framingham), but remember problems with cohorts enriched for primary exposure of interest)
Disease frequency	Good for rare diseases	Good for common diseases
Study >1 outcome	No	Yes
Establish temporal order	Temporality generally not clear	Temporality generally clear
Incidence calculation	Can calculate OR, which approximate incidence risk/rate ratio in certain situations	Can calculate incidence risk/rate ratios depending on study design
Inherent study selection problem	Difficult to ascertain appropriate control group	Not applicable since start with a source population
Subject to biases	Susceptible to more biases, particularly recall bias	Less subject to biases, except to loss to follow-up (loss of subjects due to migration, lack of participation, withdrawal & death)
Cost-effectiveness	Cheaper and less time consuming	Expensive and time consuming

UCSF

UCSF

7. Levels of Evidence and their role in Evidence-Based Medicine

- A cornerstone of EBM is the hierarchical system of classifying evidence.
- This hierarchy is known as the levels of evidence.
- Health researchers are encouraged to find the highest level of evidence to answer etiological questions, clinicians – to answer clinical questions.

UCSF

No current consensus

Canadian Task Force on the Periodic Health Examination's Levels of Evidence*

Level	Type of evidence
I	At least 1 RCT with proper randomization
II.1	Well designed cohort or case-control study
II.2	Time series comparisons or dramatic results from uncontrolled studies
III	Expert opinions

Levels of Evidence for Prognostic Studies*

Level	Type of evidence
I	High quality prospective cohort study with adequate power or systematic review of these studies
II	Lesser quality prospective cohort, retrospective cohort study, untreated controls from an RCT, or systematic review of these studies
III	Case-control study or systematic review of these studies
IV	Case series
V	Expert opinion; case report or clinical example; or evidence based on physiology, bench research or "first principles"

Levels of Evidence from Sackett*

Level	Type of evidence
I	Large RCTs with clear cut results
II	Small RCTs with unclear results
III	Cohort and case-control studies
IV	Historical cohort or case-control studies
V	Case series, studies with no controls

Levels of Evidence for Therapeutic Studies*

Level	Type of evidence
1A	Systematic review (with homogeneity) of RCTs
1B	Individual RCT (with narrow confidence intervals)
1C	All or none study
2A	Systematic review (with homogeneity) of cohort studies
2B	Individual Cohort study (including low quality RCT, e.g. <80% follow-up)
2C	"Outcomes" research; Ecological studies
3A	Systematic review (with homogeneity) of case-control studies
3B	Individual Case-control study
4	Case series (and poor quality cohort and case-control study
5	Expert opinion without explicit critical appraisal or based on physiology bench research or "first principles"

 Burns et al. 2011

UCSF

UCSF

Summary

- Less commonly used study designs provide interesting opportunities for examining various causal relationships.
- Researchers need to be aware of multiple assumptions of such study designs; failure to verify that these assumptions are met could lead to spurious results.
- These study designs are particularly useful as 'extensions' of traditional study designs and provide efficient use of resources.

63

UCSF

Extra

Intellectually curious?

64

UCSF

UCSF

Which study design to choose?

- In theory, it's possible to use each design to test a hypothesis
  - Example: Suppose you want to study the relationship between dietary Vitamin A and lung cancer….

UCSF

Cohort study option

Subjects are chosen on the basis of exposure status and followed to assess the occurrence of disease

- High Vitamin A consumption ----------> lung cancer or not

- Low Vitamin A Consumption ----------> lung cancer or not

What are the advantages and disadvantages of this option?

UCSF

UCSF

Experimental study option

Special type of cohort study in which investigator assigns the exposure to individuals, preferably at random

Investigator assigns exposure to:

- High Vit A consumption ----------> lung cancer
  or not
- Low Vit A consumption ----------> lung cancer
  or not

What are the advantages and disadvantages of this option?

67

Case-control study option

Cases with the disease and controls who generally do not have the disease are chosen and past exposure to a factor is determined

- Prior Vitamin A consumption <---------- lung cancer cases
- Prior Vitamin A consumption <---------- controls

What are the advantages and disadvantages of this option?

68

In practice, choice of study design depends on:

- State of knowledge
- Frequency of exposure and disease
- Time, cost and other feasibility considerations
- Each study design has unique and complementary advantages and disadvantages
- Each design represents a different way of harvesting information
- Selection of one over another depends on the particular research question, concerns of about data quality and efficiency, and practical and ethical considerations

69

UCSF

Recommended references

1. H. Checkoway H, N. Pearce N, D. Kriebel. Selecting appropriate study designs to address specific research questions in occupational epidemiology. *Occup Environ Med* 2007 Sep;64(9):633-8.
2. J. H. Fowke. Issues in the design of molecular and genetic epidemiologic studies. *J Prev Med Public Health* 2009 Nov;42(6):343-8.
3. I. B. Tager. Current view of epidemiologic study designs for occupational and environmental lung diseases. *Environ Health Perspect* 2000 Aug;108 Suppl 4:615-23.

70

UCSF

UCSF