Cross-sectional studies

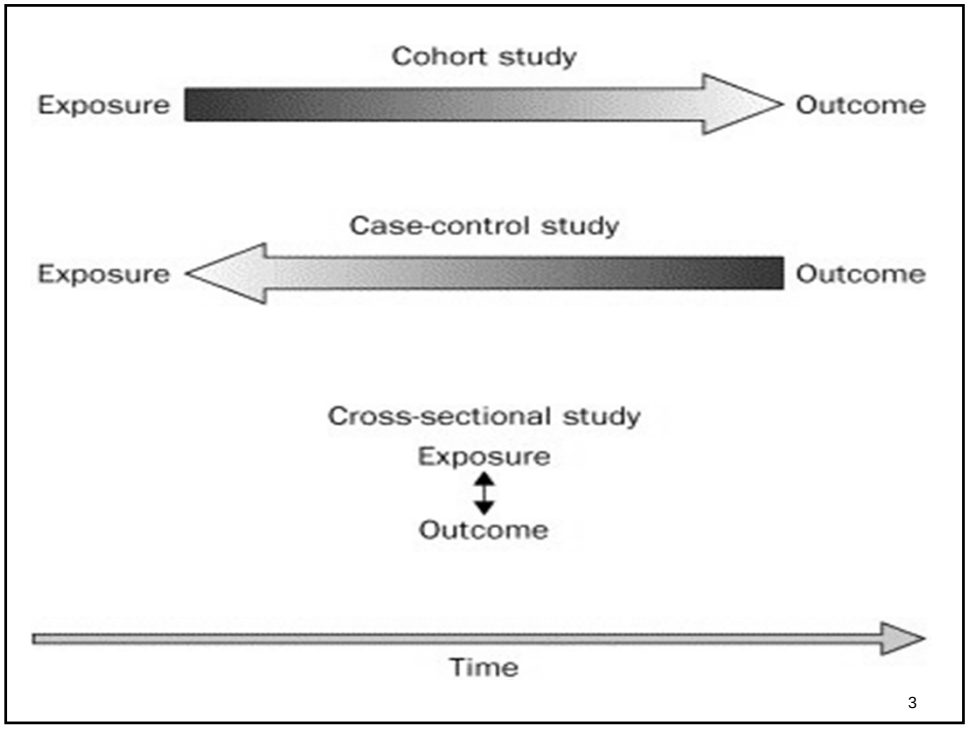
Atiporn Ingsahtit, MD., Ph.D. (Clin. Epid.)
Section of Clinical Epidemiology and Biostatistics
Faculty of Medicine Ramathibodi Hospital, Mahidol University

1

Concepts to take home

- Principle & types of cross-sectional study designs
- Advantages & disadvantages
- Prevalence, prevalence ratio, prevalence odds ratio
- Bias in cross-sectional studies
- Usefulness of cross-sectional studies

2



Principle of cross-sectional studies

- Conducted at a single point in time or over a short period of time (snapshot of population)
- Exposure status and disease status are measured at one point in time or over a period.
- Can be either descriptive or analytic, depend on design
  - Prevalence studies (descriptive cross-sectional study)
  - Comparison of prevalence among exposed and non-exposure (analytic cross-sectional study)

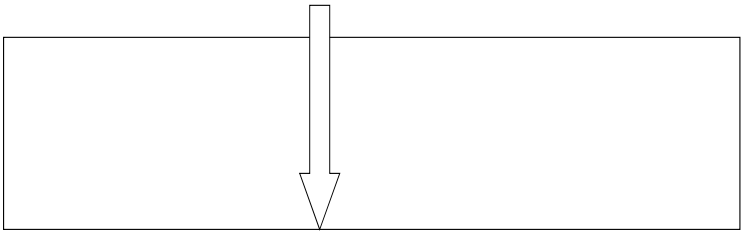
4

Analytic Cross-sectional Study

*Comparative groups

*One measurement, no follow up

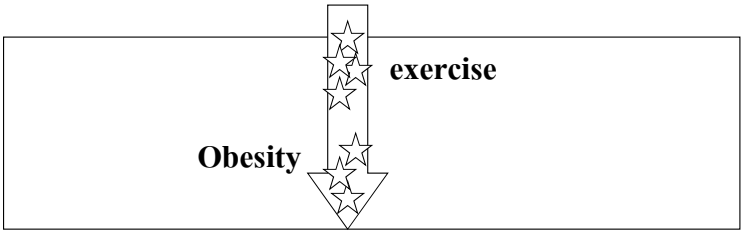
*Association ?



snapshot of population

5

Analytic Cross-sectional Study



	O+	O-
ex+	50	100
ex-	20	80

Relative prevalence O+ = (50/150)/(20/100)= 1.67

Association, no sequence

6

Types of cross-sectional studies

- **Descriptive cross-sectional study**
- **Analytic cross-sectional study**
- **Repeated cross-sectional study**

7

Cross-sectional studies

- **Descriptive**
  - Collected number of cases and number of total population.
  - Can assess only prevalence of disease or other health events, also called *"prevalence study".*

- **Analytic**
  - Expose and disease status are assessed. simultaneously
  - Can determine association between exposure and disease.

8

Descriptive cross-sectional study

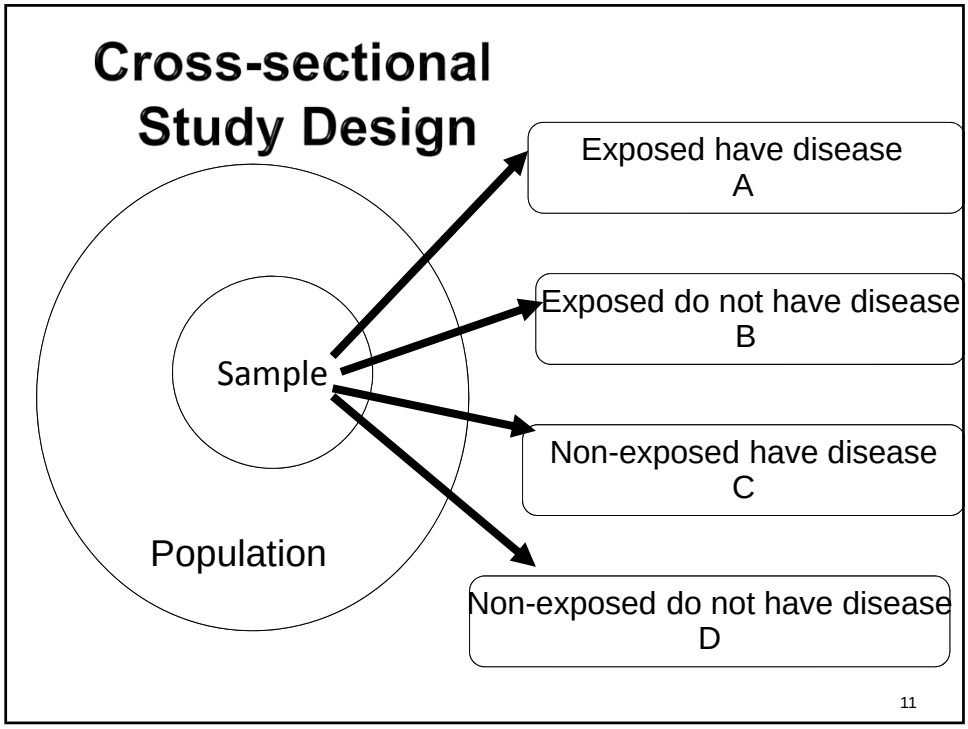
- Measures prevalence of disease at a single point in time or over a short period of time. Two types:
  - Point prevalence: *Do you currently use a NSAIDS ?*
  - Period prevalence: *Have you used a NSIADS in the past 6 months?*

9

Analytic cross-sectional study

- Measure association between expose and outcome.
- Expose and outcome are assessed simultaneously.
- Measure of association;
  - Prevalence ratio
  - Prevalence odds ratio

10



2 x 2 tables

Risk Factor	Disease: Yes	Disease: No	
Yes	A	B	A+B
No	C	D	C+D
	A+C	B+D	

12

Measure of prevalence

		Disease: Yes	Disease: No
Risk Factor	Yes	A	B
	No	C	D

$$\text{prevalence} = \frac{A+C}{A+B+C+D}$$

$$\text{Prevalence of disease among exposure} = \frac{A}{A+B}$$

$$\text{Prevalence of disease among non-exposure} = \frac{C}{C+D}$$

13

Measure of association

		Disease: Yes	Disease: No
Risk Factor	Yes	A	B
	No	C	D

1. Prevalence ratio

$$= \frac{\text{Prevalence of disease among exposure}}{\text{Prevalence of disease among non-exposure}}$$

$$= \frac{\frac{A}{A+B}}{\frac{C}{C+D}}$$

14

Measure of association

2. Prevalence odds ratio

Risk Factor	Disease: Yes	Disease: No
Yes	A	B
No	C	D

- Odds of exposure among cases

$$= \frac{\text{exposed cases}}{\text{all cases}} \Big/ \frac{\text{unexposed cases}}{\text{all cases}}$$

$$= \frac{A}{A+C} \Big/ \frac{C}{A+C} = \frac{A}{C}$$

- Odds of exposure among non-cases

$$= \frac{\text{exposed non-cases}}{\text{all non-cases}} \Big/ \frac{\text{unexposed non-case}}{\text{all non-cases}}$$

$$= \frac{B}{B+D} \Big/ \frac{D}{B+D} = \frac{B}{D}$$

Prevalence odds ratio (OR) $= \dfrac{\text{Odds of exposure among cases}}{\text{Odds of exposure among non-cases}}$

$= AD / BC$

15

Example: Medical exam & X-rays to diagnose osteoarthritis of the knee

Obesity	Osteoarthritis: yes	Osteoarthritis: no	
yes	80	20	100
no	40	60	100

16

Prevalence ratio

prevalence of osteoarthritis: 120/200 = 0.6

Prevalence of osteoarthritis among obese subjects: 80/100 = 0.8

Prevalence of osteoarthritis among non-obese subjects: 40/100 = 0.4

Prevalence ratio = 0.8/0.4 = 2.0

Interpretration: the proportion of people with OA is 2-fold greater if a person is obesity

17

Prevalence odds ratio

Prevalence odds ratio

$$= \frac{80 \times 60}{20 \times 40} = 6.0$$

Interpretation:

The odds that OA patients would be obesity appear to be about 6 times the odds that non-OA patients would be obesity.

The estimated OA diagnosis among the obese subjects is 6.0 times greater than that among the non-obese.

18

Repeated cross-sectional study

- Exposure and disease are determined at baseline and reassessed throughout a period of follow-up.
- Distinction between repeated cross-sectional study & longitudinal , prospective cohort

19

Repeated cross-sectional data

AGE (yr)	1985	1990	1995	2000	2005
40	A	B	C	D	E
35	B	C	D	E	F
30	C	D	E	F	G
25	D	E	F	G	H
20	E	F	G	H	I

Year

20

Longitudinal or cohort data

AGE (yr)					
40	A	B	C	D	E
35	B	C	D	E	F
30	C	D	E	F	G
25	D	E	F	G	H
20	E	F	G	H	I
	1985	**1990**	**1995**	**2000**	**2005**
			Year		

1

Advantages of cross-sectional studies

- Good for describing the magnitude and distribution of health problems.
- Generalizable results if population based sample
- Quick, conducted over short period of time, easy, inexpensive.
- Can study multiple exposures and disease outcomes simultaneously.

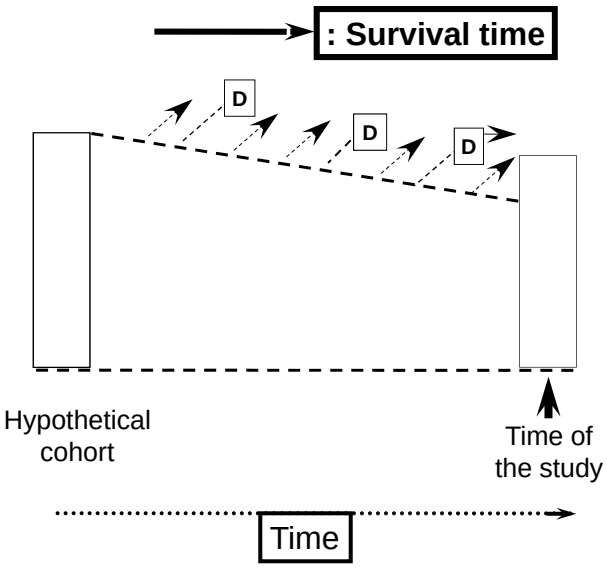
22

Disadvantages of cross-sectional studies

- Cannot establish sequence of events
  - Not for causation or prognosis
- Impractical for rare diseases if pop based sample (eg, gastric CA 1/10,000).
- Possible bias since only survivors are available for study

23

Cross–sectional study design



24

Bias in Cross-Sectional Studies

1. **Selection bias**
   - **Sampling bias: representativeness**
   - **Prevalence-incidence bias (Neyman bias)**
   - **Response and non-response bias**
2. **Measurement bias**
   - **Misclassified (misdiagnosed, undiagnosed)**
     - **Recall bias**
     - **Lead-time bias**
     - **Length biased sampling**
3. **Confounding**

25

Sampling in Epidemiology

- Definitions
  - **Sampling unit** – the basic unit around which a sampling procedure is planned
    - Person
    - Group – household, school, district, etc.
    - Component – eye, physiological response
  - **Sampling frame** – list of all of the sampling units in a population
  - **Sample** – collection of sampling units from the eligible population

26

Sampling in Epidemiology

- **Probability Sample**
  - Simple random sample
  - Stratified random sample
  - Cluster sample
  - Multistage sample
  - Systematic sample
- **Non-probability Sample**
  - Convenience sample
  - Consecutive sample
  - Quota sample
  - Volunteer sample

PROBABILITY SAMPLE

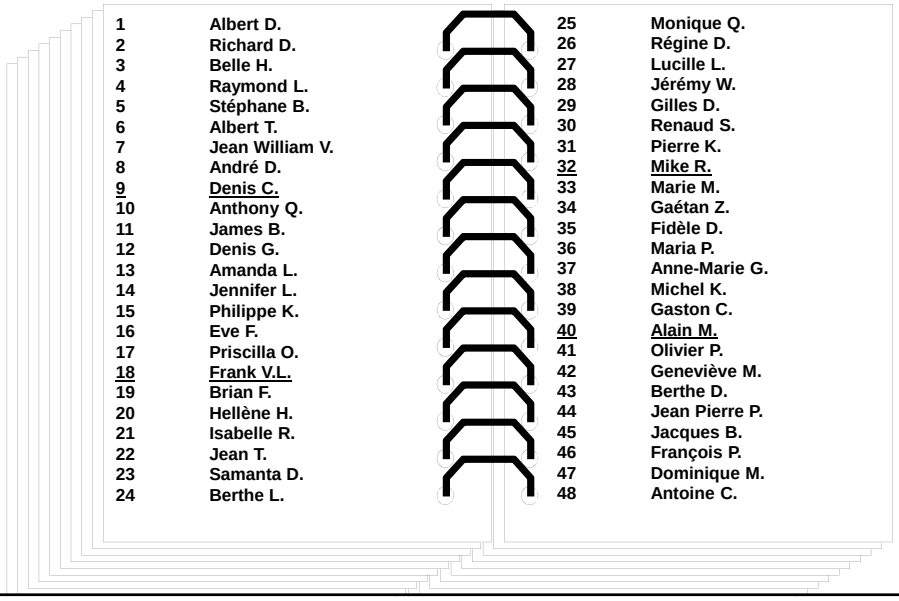
28

Sampling in Epidemiology

- Simple random sampling
  - Each sampling unit has an equal chance of being included in the is sample
  - In epidemiology, sampling generally done without replacement as this approach allows for a wider coverage of sampling units, and as a result smaller standard errors

Example of simple, random sampling

Numbers are selected at random

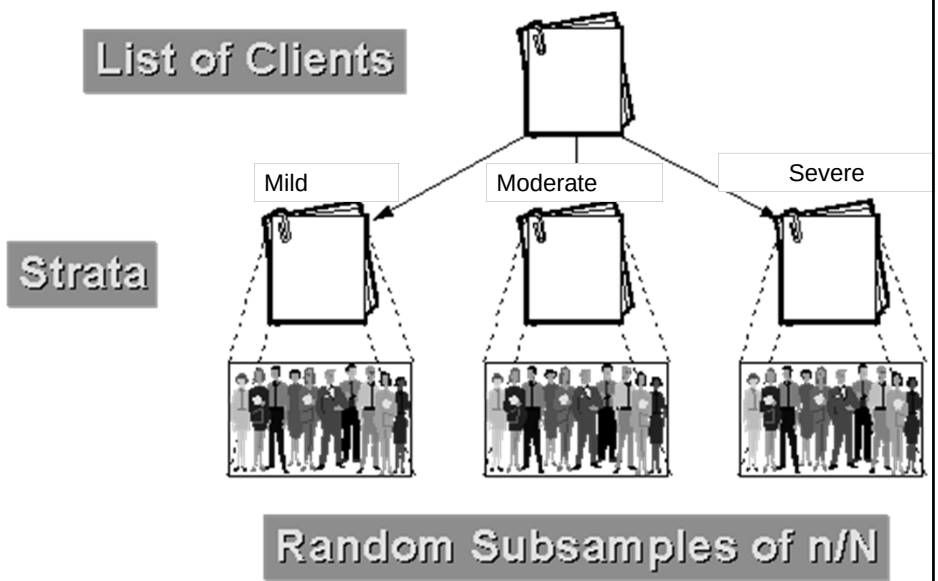


1	Albert D.	25	Monique Q.
2	Richard D.	26	Régine D.
3	Belle H.	27	Lucille L.
4	Raymond L.	28	Jérémy W.
5	Stéphane B.	29	Gilles D.
6	Albert T.	30	Renaud S.
7	Jean William V.	31	Pierre K.
8	André D.	32	Mike R.
9	Denis C.	33	Marie M.
10	Anthony Q.	34	Gaétan Z.
11	James B.	35	Fidèle D.
12	Denis G.	36	Maria P.
13	Amanda L.	37	Anne-Marie G.
14	Jennifer L.	38	Michel K.
15	Philippe K.	39	Gaston C.
16	Eve F.	40	Alain M.
17	Priscilla O.	41	Olivier P.
18	Frank V.L.	42	Geneviève M.
19	Brian F.	43	Berthe D.
20	Hellène H.	44	Jean Pierre P.
21	Isabelle R.	45	Jacques B.
22	Jean T.	46	François P.
23	Samanta D.	47	Dominique M.
24	Berthe L.	48	Antoine C.

Sampling in Epidemiology

- Stratified random sample
  - The sampling frame comprises groups, or strata, with certain characteristics
  - A sample of units are selected from each group or stratum

Stratified Random selection for drug trail in hypertension



Sampling in Epidemiology

- Cluster sampling
  - Clusters of sampling units are first selected randomly
  - Individual sampling units are then selected from within each cluster

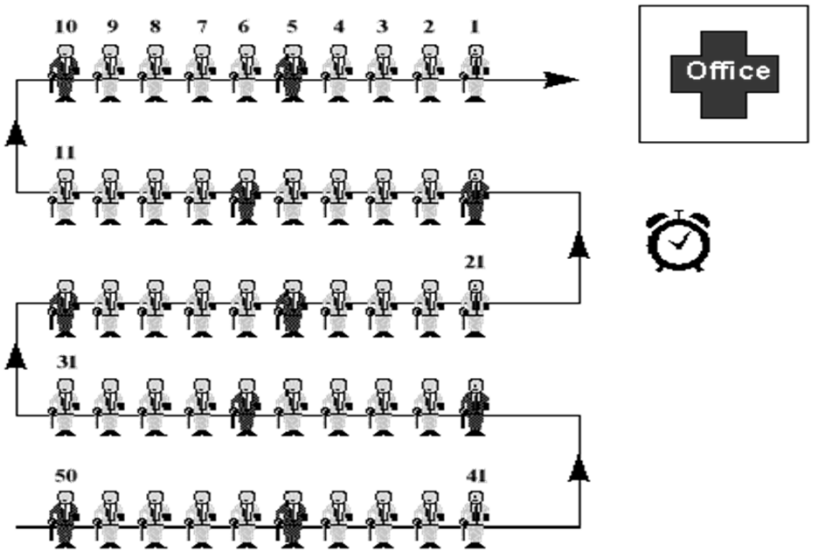
Sampling in Epidemiology

- Multistage sampling
- Similar to cluster sampling except that there are two sampling events, instead of one
  - Primary units are randomly selected
  - Individual units within primary units randomly selected for measurement

Sampling in Epidemiology

- Systematic sampling
  - The sampling units are spaced regularly throughout the sampling frame, e.g., every 3rd unit would be selected
  - May be used as either probability sample or not
    - Not a probability sample unless the starting point is randomly selected
    - Non-random sample if the starting point is determined by some other mechanism than chance

Systematic sampling



NON-PROBABILITY SAMPLE

37

Sampling in Epidemiology

- **Convenience sample**
  - **Case series of patients with a particular condition at a certain hospital**
  - **“Normal” graduate students walking down the hall are asked to donate blood for a study**
  - **Children with febrile seizures reporting to an emergency room**

Investigator decides who is enrolled in a study

Sampling in Epidemiology

- **Consecutive sample**
  - **A case series of <u>consecutive</u> patients with a condition of interest**
  - **Consecutive series means ALL patients with the condition within hospital or clinic, not just the patients the investigators happen to know about**
- **Advantages**
  - **Removes investigator from deciding who enters a study**
  - **Requires protocol with definitions of condition of interest**
  - **Straightforward way to enroll subjects**
- **Disadvantage**
  - **Non-random**

Sampling in Epidemiology

- **Quota sampling**:
  selecting fixed numbers of units in each of a number of categories.

QUOTA SAMPLING

- **Researcher uses some knowledge of the population to build some representativeness into the sampling plan**
- divides population into different strata and samples from each of them
- USUALLY BETTER THAN JUST CONVENIENCE

40

Prevalence-incidence bias (Neyman bias)

- It arises when **a gap in time** occurs between **exposure** and **selection of study subjects.**

41

Neyman bias example

- The study of myocardial infarction and snow shovelling (the exposure of interest) would miss individuals who died in their driveways and thus never reached a hospital.
- This eventuality might greatly lower the association of infarction associated with this strenuous activity.

42

Prevalence-incidence bias (Neyman bias)

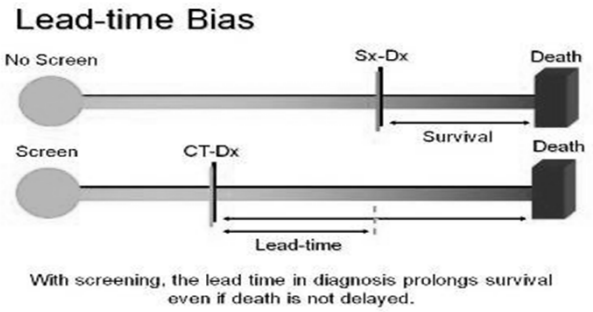
Framingham study

	Incidence			Prevalence		
	Developed CHD by exam 6	Did not develop CHD by exam 6	Total	CHD present at exam 6	No CHD present at exam 6	Total
High serum cholesterol	85	462	547	38	34	72
Low serum cholesterol	116	1511	1627	113	117	230
	201	1973	2174	151	151	302
ORs	**2.40**			**1.16**		

Friedman et al. Amer J Epid 1966;83:366

43

Lead-time bias



With screening, the lead time in diagnosis prolongs survival even if death is not delayed.

- Lung cancer-specific survival is measured from the time of diagnosis (Dx) of lung cancer to the time of death.
- If a lung cancer is screen-detected before symptoms (Sx), then the lead time in diagnosis equals the length of time between screening detection and when the first signs/symptoms would have appeared.
- Even if early treatment had no benefit, the survival of screened persons would be longer simply by the addition of the lead time.

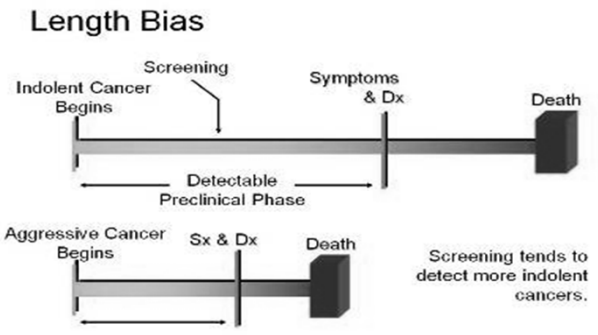
44

Length biased sampling

- **Length biased sampling**: diseases that have long duration will over-represent the magnitude of illness while short duration will under-represent illness

45

Length bias



- The cancers that grow slowly are easier to detect because they have a longer pre-symptomatic period of time when they are detectable.
- Thus, the screening test detects more slowly growing cancers.

46

Usefulness of cross-sectional study design

- Diagnostic test
- Prevalence study
  - Describe distribution of variables
  - Health care services
- Examine associations among variables
  - Hypothesis generating for causal links
- Prediction score

47

Accuracy of a Test Result

		Disease		
		Yes	No	
Test	Positive	a True positive	b False positive	
	Negative	c False negative	d True negative	
				a+b+c+d

Sensitivity = true positive rate = a / a + c
Specificity = true negative rate = d / b + d

Accuracy of a Test Result

Test	Disease	No disease	Total	EST	CAD	No	Total
+	a	b	a+b	+	80	10	90
-	c	d	c+d	-	20	90	110
	a+c	b+d	n		100	100	200

Term	General	Example	Definition
Sensitivity	a/(a+c)	80/100 (80%)	Proportion of those with the condition who have a positive test
Specificity	B/(b+d)	90/100 (90%)	Proportion of those without the condition who have a negative test
Accuracy	a+d/n	170/200 (85%)	Proportion of accurate diagnostic test
Positive predictive value	a/(a+b)	80/90 (90%)	Proportion of those with a positive test who have the condition
Negative predictive value	d/(c+d)	90/110 (82%)	Proportion of those with a negative test who do not have the condition

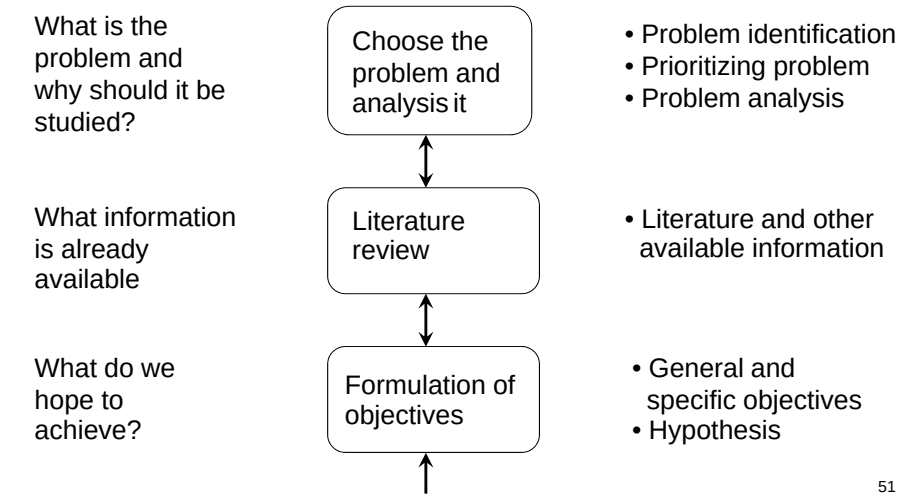
49

Accuracy of a Test Result

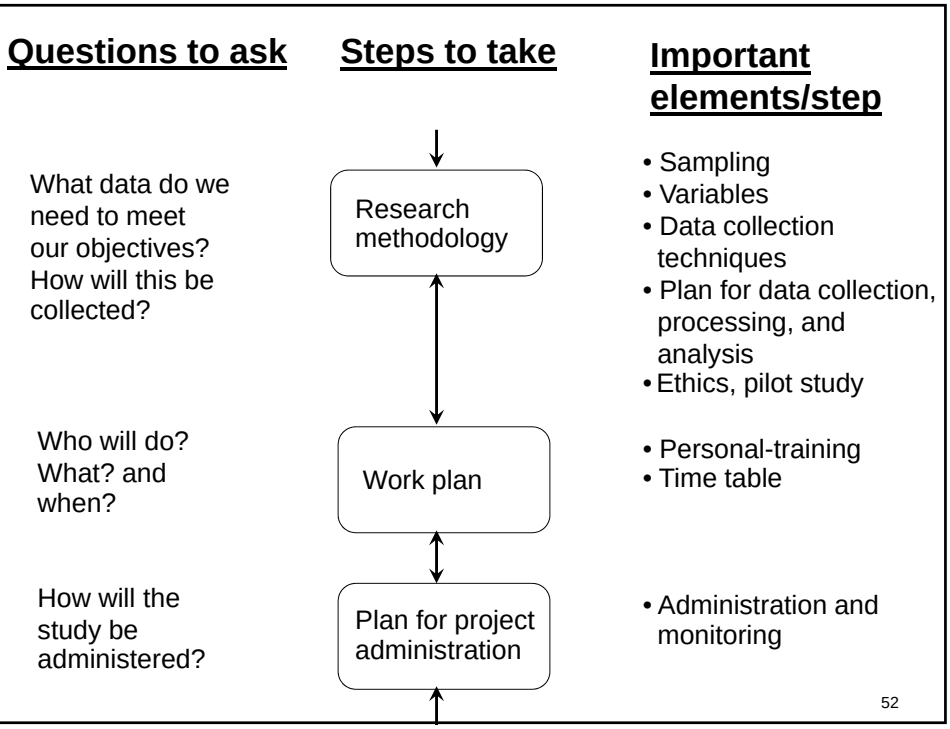
- Sensitivity: Is the test detecting true cases of disease?
  - (Ideal is 100%: 100% of cases are detected)
- Specificity: Is the test excluding those without disease?
  - (Ideal is 100%: 100% of non-cases are negative)

50

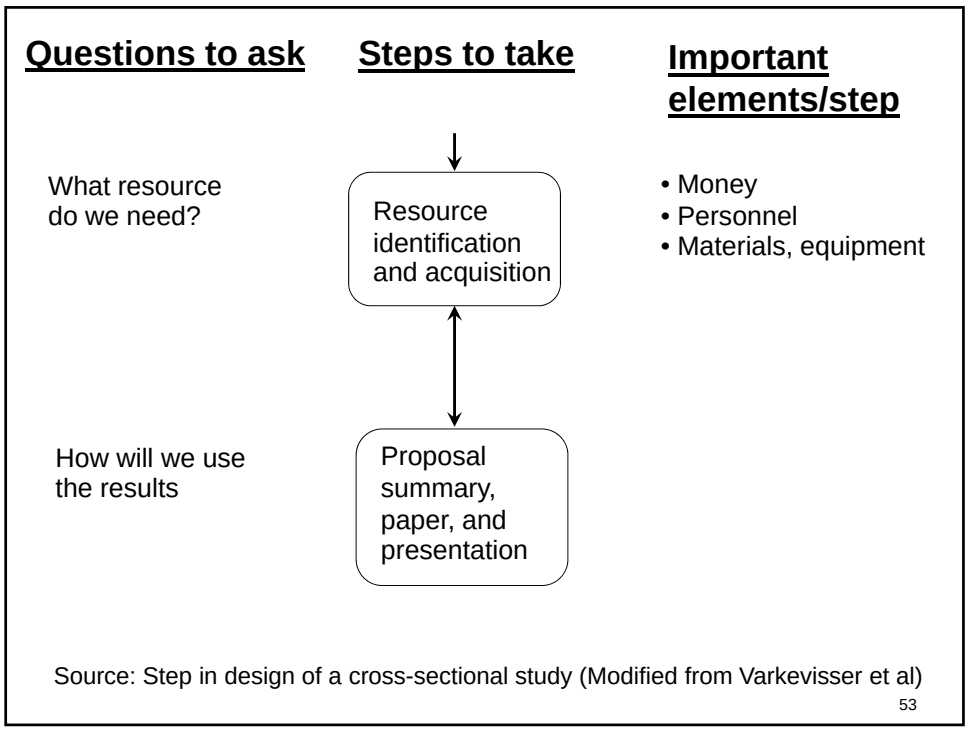
Steps of conducting cross-sectional study

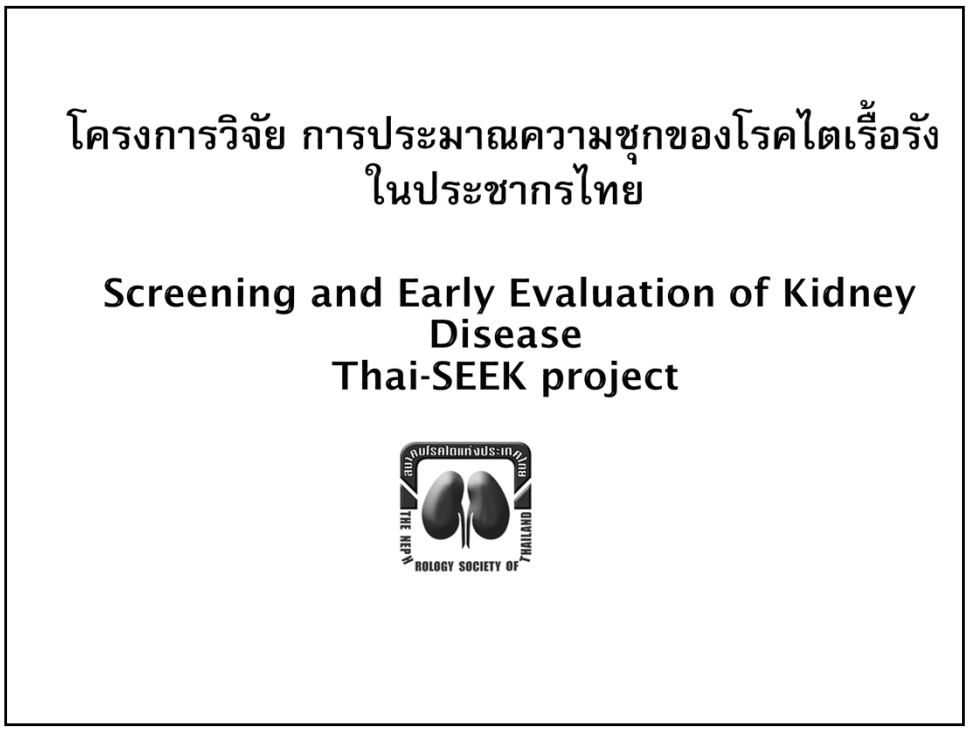


51



52


Questions to ask
Steps to take
Important elements/step
What resource do we need?
Resource identification and acquisition
• Money
• Personnel
• Materials, equipment
How will we use the results
Proposal summary, paper, and presentation
Source: Step in design of a cross-sectional study (Modified from Varkevisser et al)
53


โครงการวิจัย การประมาณความชุกของโรคไตเรื้อรัง ในประชากรไทย
Screening and Early Evaluation of Kidney Disease
Thai-SEEK project

Nephrol Dial Transplant (2010) 25: 1567–1575
doi: 10.1093/ndt/gfp669
Advance Access publication 27 December 2009

Prevalence and risk factors of chronic kidney disease in the Thai adult population: Thai SEEK study

Atiporn Ingsathit[1], Ammarin Thakkinstian[1], Amnart Chaiprasert[2], Pornpen Sangthawan[3], Pongsathorn Gojaseni[4], Kriwiporn Kiattisunthorn[5], Leena Ongaiyooth[5], Somlak Vanavanan[6], Dhavee Sirivongs[7], Prapaipim Thirakhupt[8], Bharati Mittal[9], Ajay K. Singh[9] and the Thai-SEEK Group

[1]Section for Clinical Epidemiology and Biostatistics, Research Center, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand, [2]Division of Nephrology, Department of Medicine, Phramongkutklao Hospital, Bangkok, Thailand, [3]Division of Nephrology, Department of Medicine, Faculty of Medicine, Prince of Songkla University, Songkla, Thailand, [4]Division of Nephrology, Department of Medicine, Bhumibol Adulyadej Hospital, Bangkok, Thailand, [5]Division of Nephrology, Department of Medicine, Faculty of Medicine, Siriraj Medical School and Hospital, Mahidol University, Bangkok, Thailand, [6]Division of Clinical Chemistry, Department of Pathology, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand, [7]Division of Nephrology, Department of Medicine, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand, [8]Division of Nephrology, Department of Pediatrics, Phramongkutklao Hospital, Bangkok, Thailand and [9]Renal Division, Brigham and Women's Hospital and Harvard Medical School, Boston, USA

Correspondence and offprint requests to: Atiporn Ingsathit; E-mail: teait@mahidol.ac.th

Primary objective

- To describe the ***distribution of CKD stages*** and ***severity***

Methodology

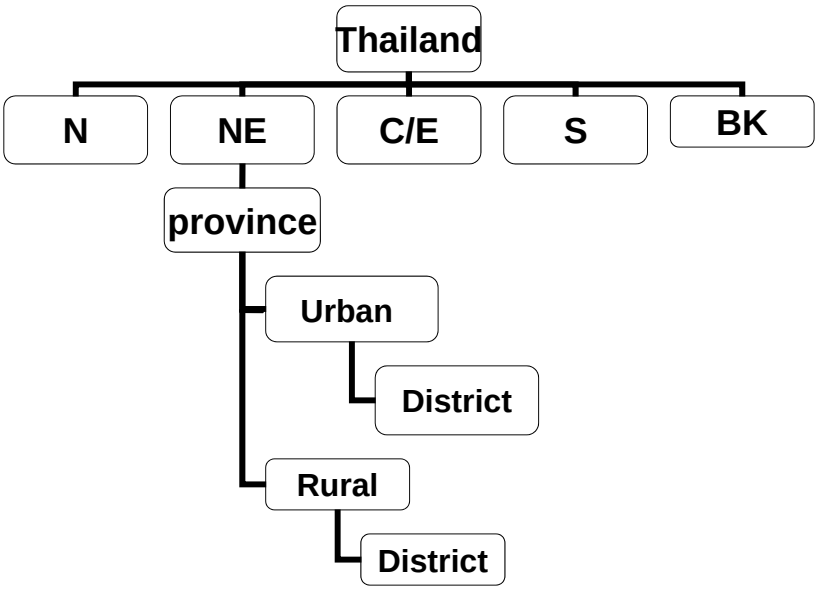
- **Study design**: Cross-sectional study
- **Study period:** August 2007 to January 2009

The study was approved by the IRB of the Faculty of Medicine at Ramathibodi Hospital, Mahidol University

Study subjects

- **Inclusion criteria**
  - Aged 18 or older
  - No menstruation period
  - No fever for at least a week before examination date
  - Willingness to participate and provide a signed consent form
- **Exclusion criteria**
  - Blood or urine specimens were not taken

Stratified-cluster random sampling

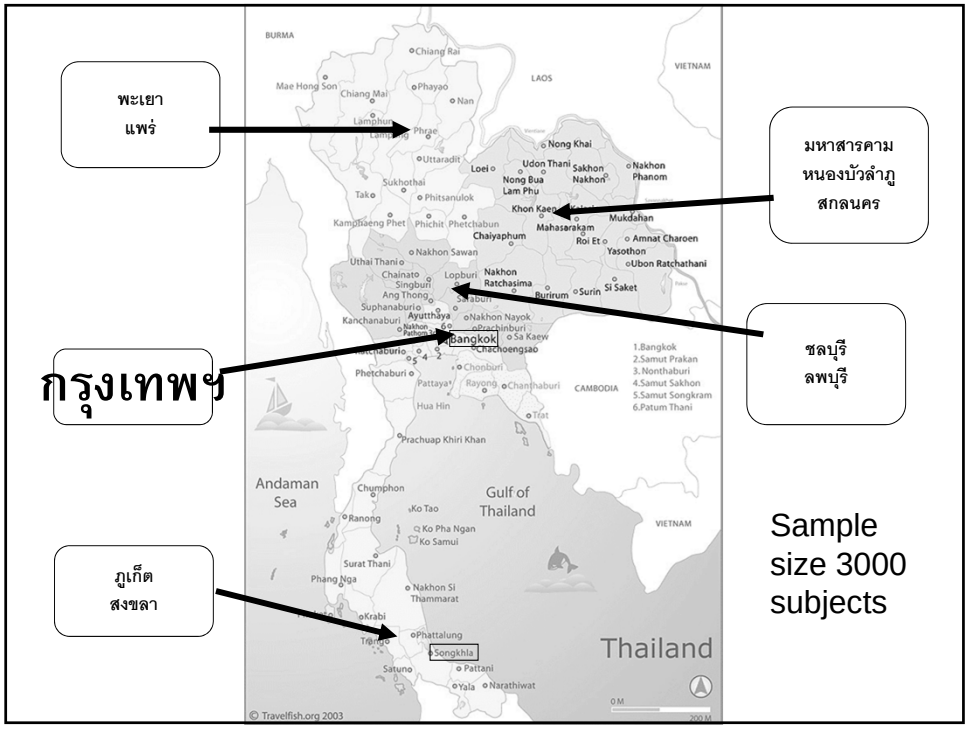


Sample size estimation

- Prevalence from previous studies 3%-13.7%
- Type I error = 0.05
- Design effect = 3
- Calculate 95% CI
  - Sample size 4,000 95%CI = 11.9-15.7
  - Sample size 3,000 95%CI = 11.7-16.0

Sample size estimation

- Prevalence from previous studies 3%-13.7%
- Type I error = 0.05
- Design effect = 3
- Calculate 95% CI
  - Sample size 4,000 95%CI = 11.9-15.7
  - Sample size 3,000 95%CI = 11.7-16.0



ภาค	ขนาดประชากร	ขนาดตัวอย่างต่อภาค	จำนวนจังหวัดตัวอย่าง	ขนาดตัวอย่างของจังหวัด	จังหวัดตัวอย่าง	อำเภอตัวอย่าง	ขนาดตัวอย่างของอำเภอ+10%*
กทม	5658953	272	1	272	กรุงเทพมหานคร	พระนคร,วัฒนา	150
กลาง	15030613	722	2	361	ชลบุรี	พานทอง,สัตหีบ	199
				361	ลพบุรี	พัฒนานิคม,ท่าหลวง	199
เหนือ	11883517	571	2	286	พะเยา	เมือง,จุน	157
				286	แพร่	สูงเม่น,สอง	157
ตะวันออกเฉียงเหนือ	21328112	1025	3	342	มหาสารคาม	นาเชือก,วาปีปทุม	188
				342	หนองบัวลำภู	นาวัง, นากลาง	188
				342	สกลนคร	นิคมน้ำอูน, กุสุมาลย์	188
ใต้	8516860	409	2	205	ภูเก็ต	เมือง,ถลาง	113
				205	สงขลา	สิงหนคร,นาหม่อม	113
รวม	62418056	3000	10	-			

Measurement

- Serum creatinine: Standardized with IDMS method
- Urine albumin: Immunoturbidimetry
- Hematuria: Trained technician at site

Pre-camp training

Camp day

Station 1 Inform consent

Station 2 Registration

Station 3 Blood sample collection

Station 4 Urine sample collection

Station 5 Interview

Station 6 Physical examination

Station 7 Education

Material

Station 8 Check point for completeness

RESULTS

CKD prevalence in Thai population

Thai SEEK study

3,459 general population

Age 45.2 (0.8), Male 45.3%

CKD staging								Overall N=3459	
I		II		III		IV+			
No*	Prevalence* (95%CI)	No.	Prevalence (95%CI)	No.	Prevalence(95%CI)	No.	Prevalence (95%CI)	No.	Prevalence (95%CI)
134	3.3	207	5.6	248	7.5	37	1.1	626	17.5
	(2.5, 4.1)		(4.2, 7.0)		(6.2, 8.8)		(0.7, 01.5)		(14.6, 20.4)
8.9 (6.8, 11.0)				8.6 (7.0, 10.3)					

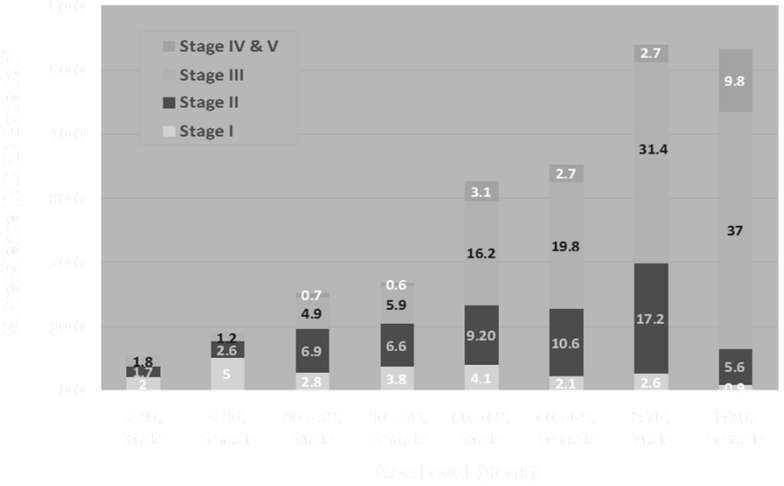
Atiporn Ingsathit , et al. Nephrol Dial Transplant. 2010 May;25(5):1567-75

Projection of expected numbers of adult population

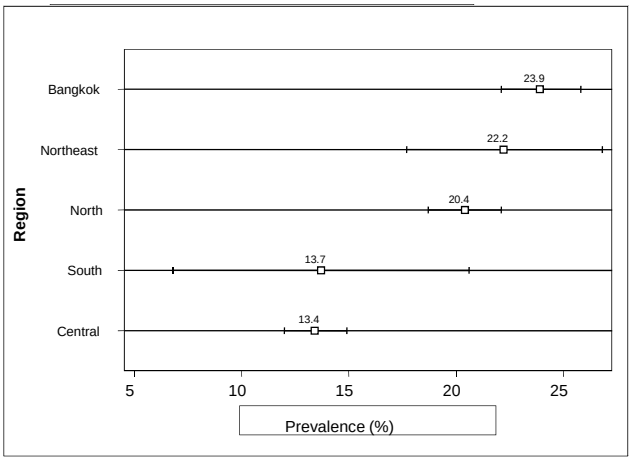
Thai SEEK study

Year	Adult Population	Expected CKD cases
2008	3.9 million	5.6 million
2009	3.9 million	5.7 million
2010	4.0 million	5.8 million
2011	4.8 million	7.0 million
2012	4.9 million	7.1 million
2013	5.0 million	7.2 million

Estimation of CKD prevalence according to age and gender



Estimation of CKD prevalence according to region



Risk factors associated with CKD

Factors	CKD group				Adjusted OR	
	Stage I-V		No CKD		OR (95% CI)	p-value
	number	%	number	%		
Age, year						
≥ 70	139	22.26	128	4.08	7.34 (4.18, 12.90)	<0.001
60 – 69	148	22.85	255	9.40	3.63 (2.26, 5.86)	0.001
40 – 59	237	39.19	1,227	43.85	1.71 (1.16, 2.52)	0.017
< 40	102	15.70	1,223	42.67	1	
History of kidney stone	74	11.30	95	3.72	2.72 (1.80, 4.12)	0.002
DM	183	28.48	251	8.40	2.72 (1.57, 4.73)	0.005
Hypertension	329	53.60	626	21.99	1.96 (1.44, 2.67)	0.002
Uric acid, mg/dl						
> 5.61	331	55.03	938	35.09	2.87 (1.77, 4.64)	0.002
4.40 – 5.61	166	26.58	960	33.49	1.50 (0.92, 2.46)	0.087
< 4.40	129	18.39	935	31.42	1	
Using traditional medicine	263	42.65	880	31.55	1.20 (1.02, 1.42)	0.035
Sex						
Female	356	57.77	1,534	53.86	1.70 (1.18, 2.43)	0.013
Male	270	42.23	1,299	46.14	1	

Cross-sectional Design

Rapid, Easy

Co-operative

Inexpensive

Prevalence study

First step of cohort

Cross-sectional association

Blinded: single

Causal relationship

Rare diseases

Not incidence

82

Summary

- Principle & types of cross-sectional study designs
- Advantages & disadvantages
- Prevalence, prevalence ratio, prevalence odds ratio
- Bias in cross-sectional studies
- Usefulness of cross-sectional studies

83