Cross-sectional Study Design

İlker Kayı, MD, DPH

Koç University School of Medicine

10th RMHS, 2019

Outline

- Design of x-sectional studies
- Measures obtained in x-sectional studies
- When to use
- Sample selection
- Sample size calculation

Source: https://www.youtube.com/watch?v=h8vaYG5z5lw

CASE-CONTROL STUDY

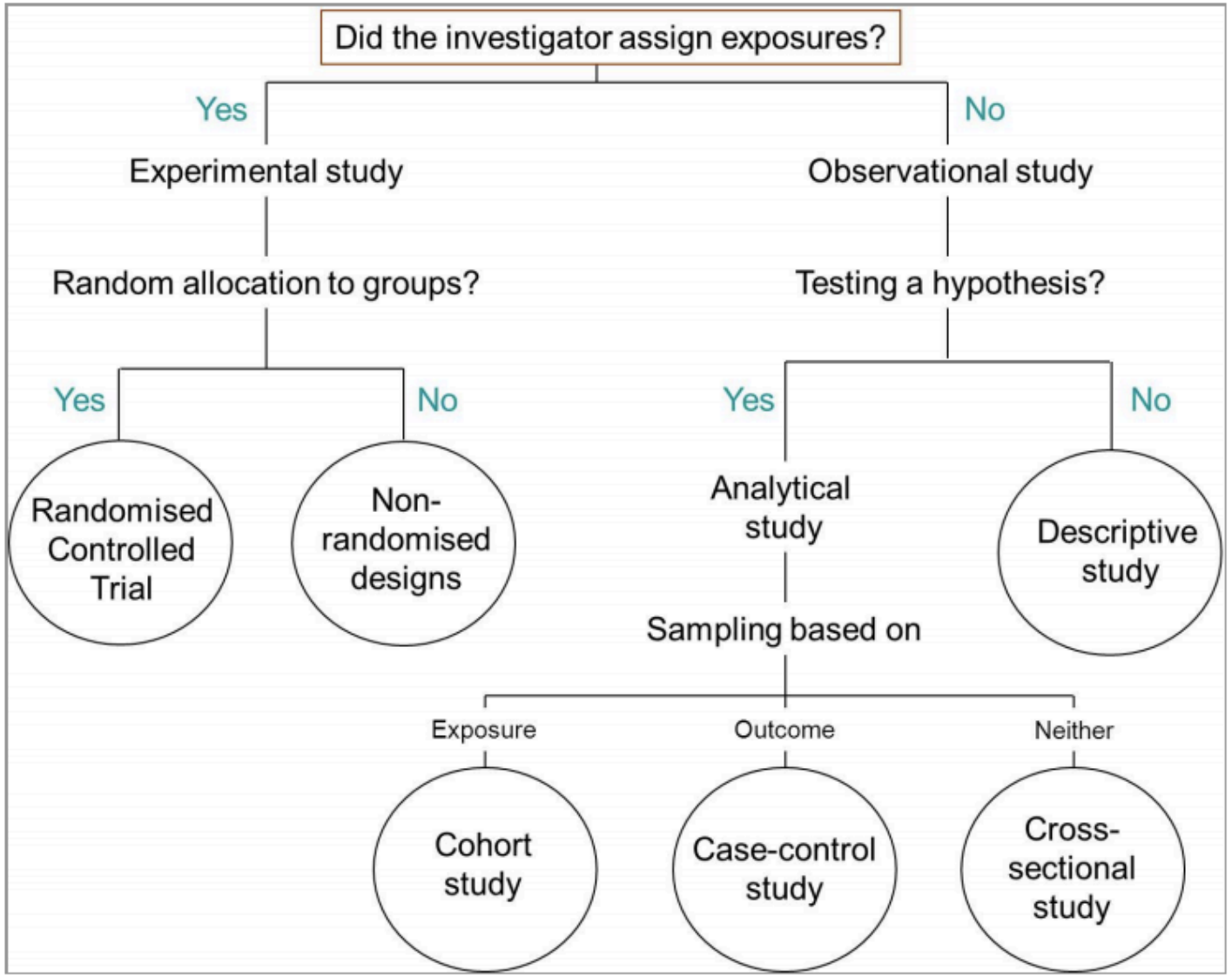
Exposure

X-SECTIONAL STUDY

Outcome

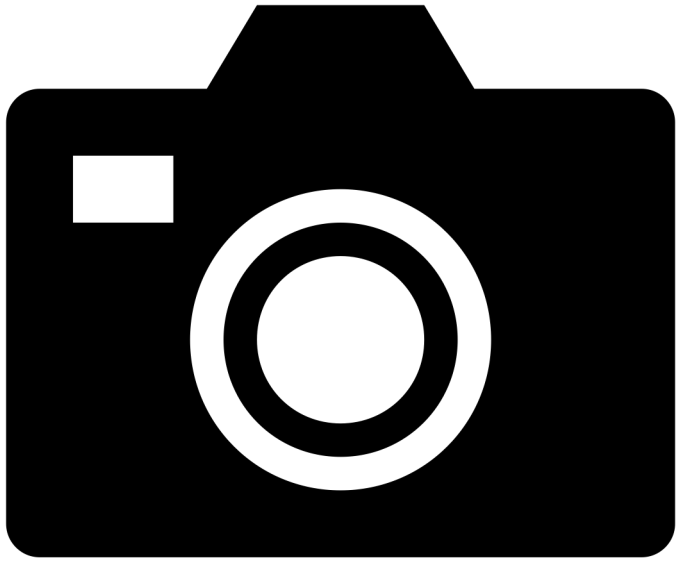
COHORT STUDY

What kind of study?



Definition

- Cross-sectional studies examine the relationship between diseases (or other health-related characteristics) and other variables of interest as they exist in a defined population at a particular point in time.
  - “Snapshot” of a society
  - Prevalence studies
  - Disease frequency studies



X-Sectional Studies

- The ***study domain*** of a survey is a particular segment of a demographic population, or, within a geographical area, a collection of institutions or other functional units in health care.
    - The population surveyed can be very large, as in national surveys, or relatively small, as when a single school or village is surveyed.
- The ***occurrence relation*** usually involves many different outcomes of interest.
    - The interest may be whether a factor has an association with the outcome and/or whether determinant-outcome relations differ according to third variables (such as sociodemographic characters etc.)

X-Sectional studies

- Data collection tool: Surveys, health records (databases), clinical observations
- Direction: None
  - People are studied at a “point” in time, without follow-up.
- Outcome measure: Prevalence
- A cross-sectional study can be combined with a follow-up study to create a cohort or a case-control study.
- Repeated cross-sectional studies can provide data for the change in a selected measure in population.

Exposure

X-SECTIONAL STUDY

Outcome

GATS 2 | Global Adult Tobacco Survey

FACT SHEET | INDIA 2016-17

GATS Objectives

The Global Adult Tobacco Survey (GATS) is a global standard for systematically monitoring adult tobacco use (smoking and smokeless) and tracking key tobacco control indicators.
GATS is a nationally representative survey, using a consistent and standard protocol across countries including India. GATS enhances countries' capacity to design, implement and evaluate tobacco control programs. It will also assist countries to fulfill their obligations under the World Health Organization (WHO) Framework Convention on Tobacco Control (FCTC) to generate comparable data within and across countries. WHO has developed MPOWER, a package of selected demand reduction measures contained in the WHO FCTC that includes:

mpower

- Monitor tobacco use & prevention policies
- Protect people from tobacco smoke
- Offer help to quit tobacco use
- Warn about the dangers of tobacco
- Enforce bans on tobacco advertising, promotion, & sponsorship
- Raise taxes on tobacco

GATS Methodology

GATS uses a global standardized methodology. It includes information on respondents' background characteristics, tobacco use (smoking and smokeless), cessation, secondhand smoke, economics, media, and knowledge, attitudes and perceptions towards tobacco use. GATS is a household survey of persons 15 years of age or older conducted in all 30 states of India and two Union Territories. The first round of GATS was conducted between June 2009 and January 2010. The second round of GATS was conducted between August 2016 to February 2017 by Tata Institute of Social Sciences, Mumbai for the Ministry of Health & Family Welfare, Government of India.
A multi-stage sample design was used for both rounds of GATS. From each of the sampled household, one household member 15 years of age or older was randomly selected for individual interview. In the first round 69,296 individual interviews were completed with an overall response rate of 91.8%. In the second round, a total of 74,037 individual interviews were completed with an overall response rate of 92.9%.

GATS 2 Highlights

TOBACCO USE

- 19.0% of men, 2.0% of women and 10.7% (99.5 million) of all adults currently smoke tobacco.
- 29.6% of men, 12.8% of women and 21.4% (199.4 million) of all adults currently use smokeless tobacco.
- 42.4% of men, 14.2% of women and 28.6% (266.8 million) of all adults currently use tobacco (smoked and/or smokeless tobacco).

CESSATION

- 55.4% of current smokers are planning or thinking of quitting smoking and 49.6% of current smokeless tobacco users are planning or thinking of quitting smokeless tobacco use.
- 48.8% of current smokers were advised by health care provider to quit smoking and 31.7% of current smokeless tobacco users were advised by health care provider to quit use of smokeless tobacco.

SECONDHAND SMOKE

- 38.7% of adults were exposed to second hand smoke at home.
- 30.2% of adults who work indoors are exposed to second-hand smoke at their workplace.
- 7.4% of adults were exposed to second hand smoke at restaurants.

MEDIA

- 19.2% of adults noticed smoking tobacco advertisement and 18.3% of adults noticed smokeless tobacco advertisement.
- 68.0% of adults noticed anti-smoking tobacco information on television or radio and 59.3% of adults noticed anti-smokeless tobacco information on television or radio.

KNOWLEDGE, ATTITUDES & PERCEPTIONS

- 92.4% of adults believed that smoking causes serious illness and 95.6% of adults believed that use of smokeless tobacco causes serious illness.

Ministry of Health & Family Welfare, Government of India

World Health Organization

CDC

TISS

Use of x-sectional studies

- Determining the prevalence or frequency of a health problem
  - Risk factors
  - Priorities
- Determinants of health and disease
  - Associations – "Odds Ratio"
  - Correlations
- Evaluation of interventions or policy implementations
  - Healthcare
  - Compliance with treatment
  - Satisfaction
  - Impact of institutional policies
- Surveillance

Turkish National Health Survey 2016

Bireylerin tütün mamulü kullanma durumunun cinsiyet ve yaş grubuna göre dağılımı, 2010, 2012, 2014, 2016
The percentage of individuals' status of smoking tobacco products by sex and age group, 2010, 2012, 2014, 2016

[15+ yaş - age] **(%)**

	2010			2012			2014			2016		
	Toplam Total	**Erkek** Male	**Kadın** Female	**Toplam** Total	**Erkek** Male	**Kadın** Female	**Toplam** Total	**Erkek** Male	**Kadın** Female	**Toplam** Total	**Erkek** Male	**Kadın** Female
Her gün kullanan Daily smoker	**25.4**	**39.0**	**12.3**	**23.2**	**35.9**	**10.8**	**27.3**	**41.8**	**13.1**	**26.5**	**40.1**	**13.3**
15-24	**16.4**	27.1	6.1	**14.3**	24.1	4.6	**18.5**	31.4	5.7	**18.1**	28.2	7.8
25-34	**32.7**	48.2	17.0	**30.5**	45.9	14.9	**35.1**	51.2	18.8	**33.2**	49.6	16.6
35-44	**34.5**	49.2	19.5	**30.9**	44.4	17.3	**34.9**	49.9	19.7	**35.2**	50.6	19.6
45-54	**28.8**	43.7	13.8	**27.7**	42.0	13.4	**32.7**	48.7	16.5	**31.6**	45.3	17.7
55-64	**20.4**	32.7	8.8	**17.4**	27.9	7.4	**24.0**	38.2	10.2	**22.8**	35.0	10.9
65-74	**11.2**	20.6	4.2	**10.1**	17.8	3.8	**12.1**	22.4	3.4	**13.5**	24.2	4.4
75+	**7.3**	15.1	0.9	**5.6**	12.6	0.8	**5.0**	8.9	2.4	**4.8**	10.7	1.0

A Study Example

Sleep Med. 2018 May 24;48:140-147. doi: 10.1016/j.sleep.2018.04.013. [Epub ahead of print]

Physical activity and sleep problems in 38 low- and middle-income countries.

Vancampfort D[1], Stubbs B[2], Smith L[3], Hallgren M[4], Firth J[5], Herring MP[6], Probst M[7], Koyanagi A[8].

Author information

Abstract

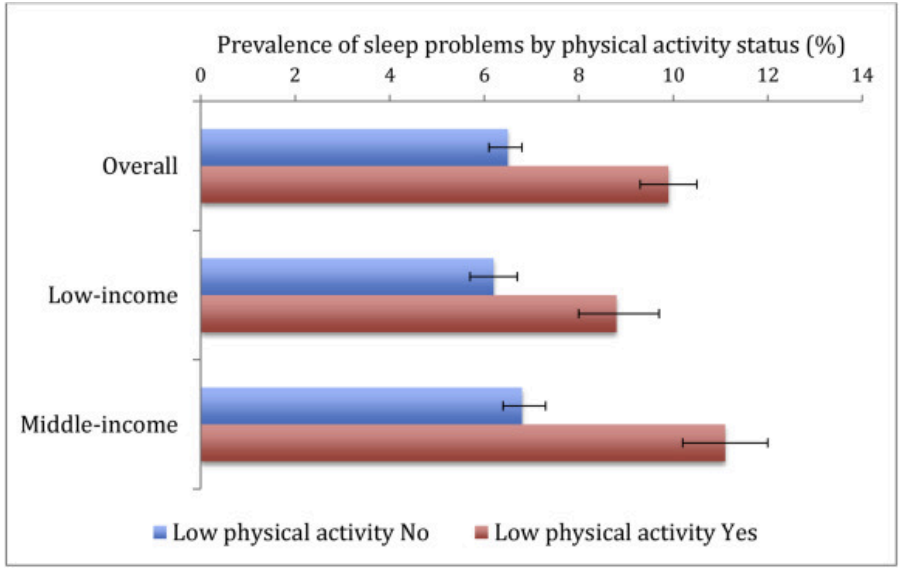
OBJECTIVE: Although physical activity (PA) is associated with a reduction of a wide range of sleep problems, it remains uncertain whether complying with the international guidelines of 150 min of moderate to vigorous PA per week can reduce sleep problems in adults. This research investigated the relationship between compliance with the PA recommendations of the World Health Organization and sleep problems in 38 low- and middle-income countries (LMICs).

METHODS: Cross-sectional, community-based data from the World Health Survey were analyzed. Adjusted logistic regression analyses were undertaken to explore the relationship between PA levels using the International Physical Activity Questionnaire and self-reported sleep problems (such as difficulties falling asleep, waking up frequently during the night or waking up too early in the morning) in the last 30 days.

RESULTS: Across 204,315 individuals (38.6 ± 16.1 years; 49.3% males), the overall prevalence (95% CI) of low PA and sleep problems were 29.9% (29.1-30.8%) and 7.5% (7.2-7.9%), respectively. After adjusting for socio-demographics, obesity, chronic physical conditions, depression, and anxiety; not complying with PA recommendations was associated with higher odds for sleep problems overall [odds ratio (OR) = 1.23, 95% CI = 1.10-1.38] as well as across the entire age range: 18-34 years (OR = 1.26; 95% CI = 1.02-1.57); 35-64 years (OR = 1.17; 95% CI = 1.01-1.35); and age ≥65 years (OR = 1.40; 95% CI = 1.11-1.76).

CONCLUSIONS: Not complying with international PA recommendations is associated with higher odds of sleep problems, independently of depression and anxiety in LMICs. Future longitudinal and interventional studies are warranted to assess whether increasing PA levels may improve sleep problems in this setting.

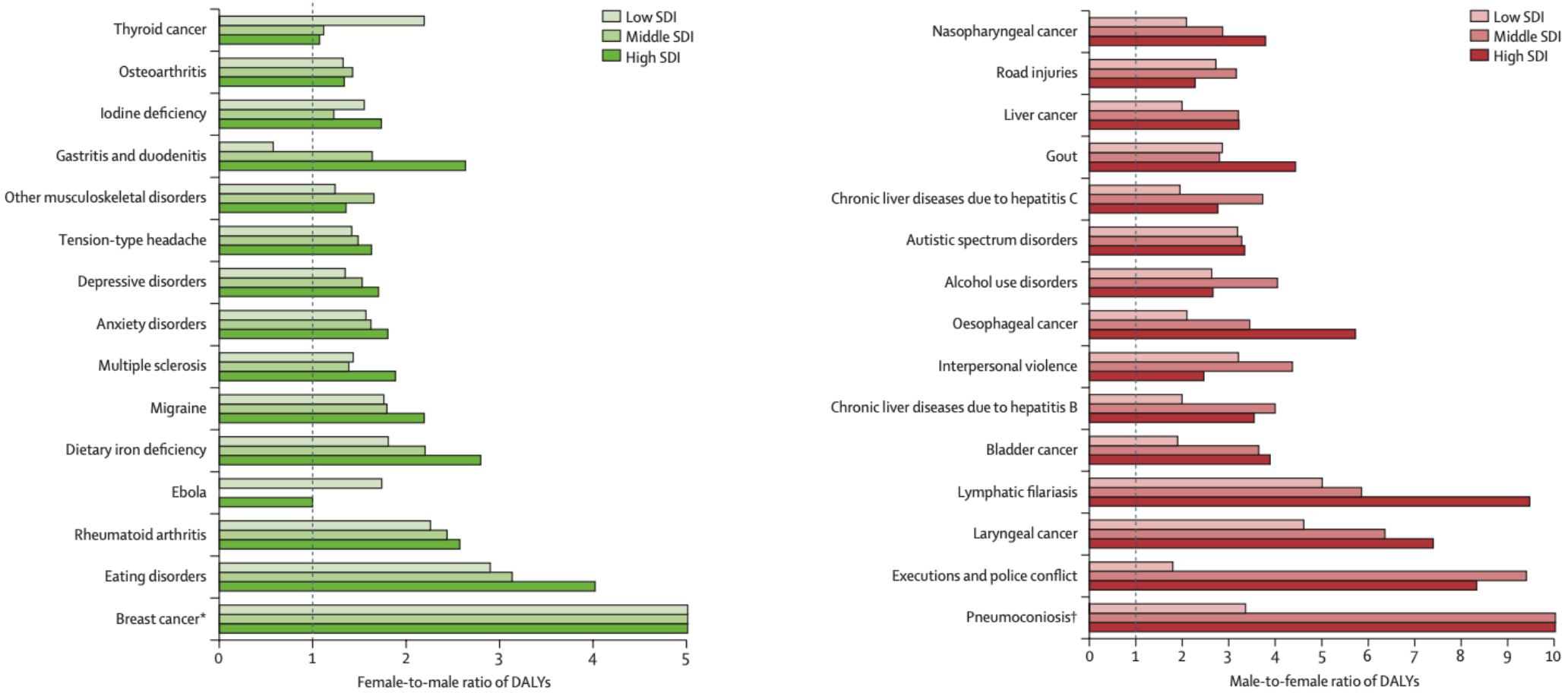
KEYWORDS: Exercise; Insomnia; Physical activity; Sleep



Gender norms and health: insights from global survey data

*Ann M Weber, Beniamino Cislaghi, Valerie Meausoone, Safa Abdalla, Iván Mejía-Guevara, Pooja Loftus, Emma Hallgren, Ilana Seff, Lindsay Stark, Cesar G Victora, Romina Buffarini, Aluísio J D Barros, Benjamin W Domingue, Devika Bhushan, Ribhav Gupta, Jason M Nagata, Holly B Shakya, Linda M Richter, Shane A Norris, Thoai D Ngo, Sophia Chae, Nicole Haberland, Katharine McCarthy, Mark R Cullen, Gary L Darmstadt on behalf of the Gender Equality, Norms, and Health Steering Committee**

Despite global commitments to achieving gender equality and improving health and wellbeing for all, quantitative data *Lancet* 2019; 393: 2455-68



Ratios of age-standardised DALYs by sex for different diseases grouped by SDI

Strengths

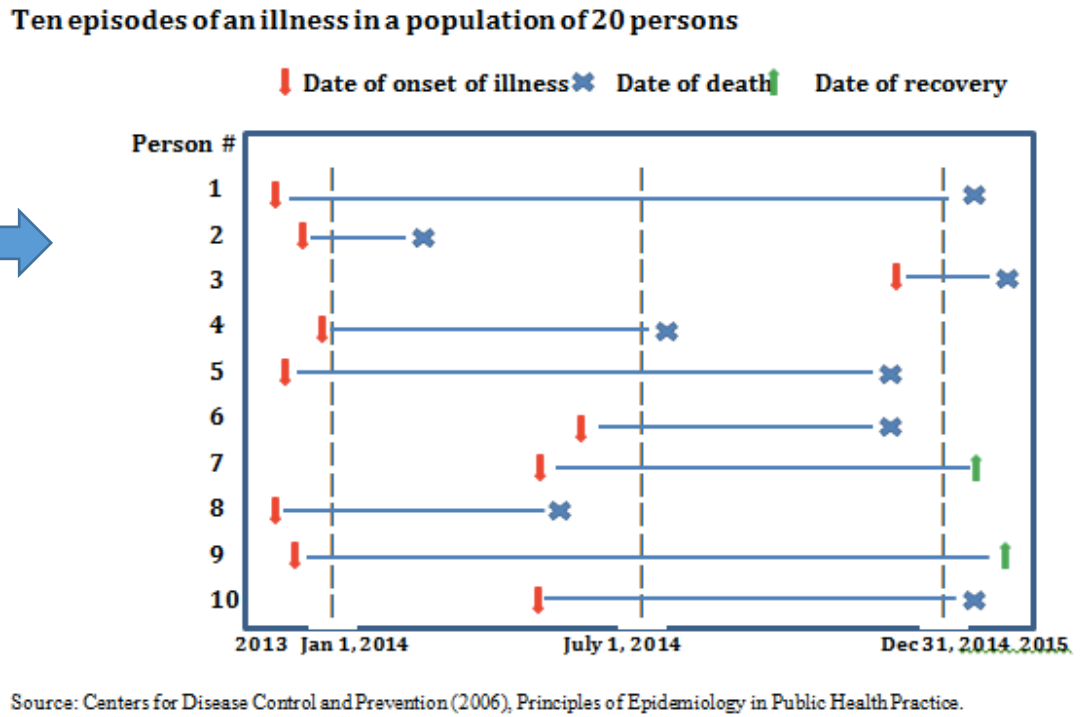
- Relatively quick and easy to conduct (no long periods of follow-up).
- Data on all variables is only collected once.
- Able to measure prevalence for all factors under investigation.
- Multiple outcomes and exposures can be studied.
- The prevalence of disease or other health related characteristics are important in public health for assessing the burden of disease in a specified population and in planning and allocating health resources.
- Good for descriptive analyses and for generating hypotheses.

Weaknesses

- Difficult to determine causality: whether the outcome followed exposure in time or exposure resulted from the outcome.
- Not suitable for studying rare diseases or diseases with a short duration.
- As cross-sectional studies measure prevalent rather than incident cases, the data will always reflect determinants of survival as well as etiology.
- Unable to measure incidence.
- Associations identified may be difficult to interpret.
- Susceptible to bias due to low response and misclassification due to recall bias.

Potential sources of bias

- Selection bias (Representativeness)
  - Sample size and selection
- Length time bias
  - Studies “survivors” and “stayers”
- Non-response



Source: Centers for Disease Control and Prevention (2006), Principles of Epidemiology in Public Health Practice.

- Select an optimum sample size
- Perform repeated contacts
- Select a random replacement for each subject
- Compare the demographic characteristics of the respondents with those of the non-respondents

Comparison of observational studies

	Cohort	Case-Control	X-Sectional
Design	Group assignment is based on exposure status of the participants. Then they are followed up for a certain time period to observe the expected outcome of interest.	Group assignment is based on outcome status of the participants. Then exposure is assessed in both groups.	Outcome and exposure are assessed together at the same time.
Analysis	Incidence Ratio and Rate Incidence Rate Ratio Advanced modeling methods (Survival Analysis, Cox Regression etc)	Odds Ratio Logistic Regression Models	Prevalance Odds Ratio Logistic Regression Models
Strengths	Temporality between exposure and outcome is well defined Allows studying multiple outcomes Efficient for rare exposures	Inexpensive Efficient for rare outcomes or outcomes with long latency Allows studying multiple exposures	Inexpensive and fast Useful before cohort studies Useful for public health monitoring
Limitations	Time-consuming and expensive Retrospective cohort studies assessment of exposure might be problematic	Prone to certain biases such as recall bias or selection bias Sometimes the temporality of the outcome might be an issue	Does not provide a causal relationship

Sampling and Calculating Sample Size

Faster keyboarding

I cdnuolt blveiee taht I cluod aulaclty uesdnatnrd waht I was rdanieg. The phaonmneal pweor of the hmuan mnid, aoccdrnig to a rscheearch at Cmabrigde Uinervtisy. It dn'seot mttaer in waht oredr the ltteers in a wrod are, the olny iprmoatnt tihng is taht the frist and lsat ltteer be in the rghit pclae. The rset can be a taotl mses and you can sitll raed it wouthit a porbelm.

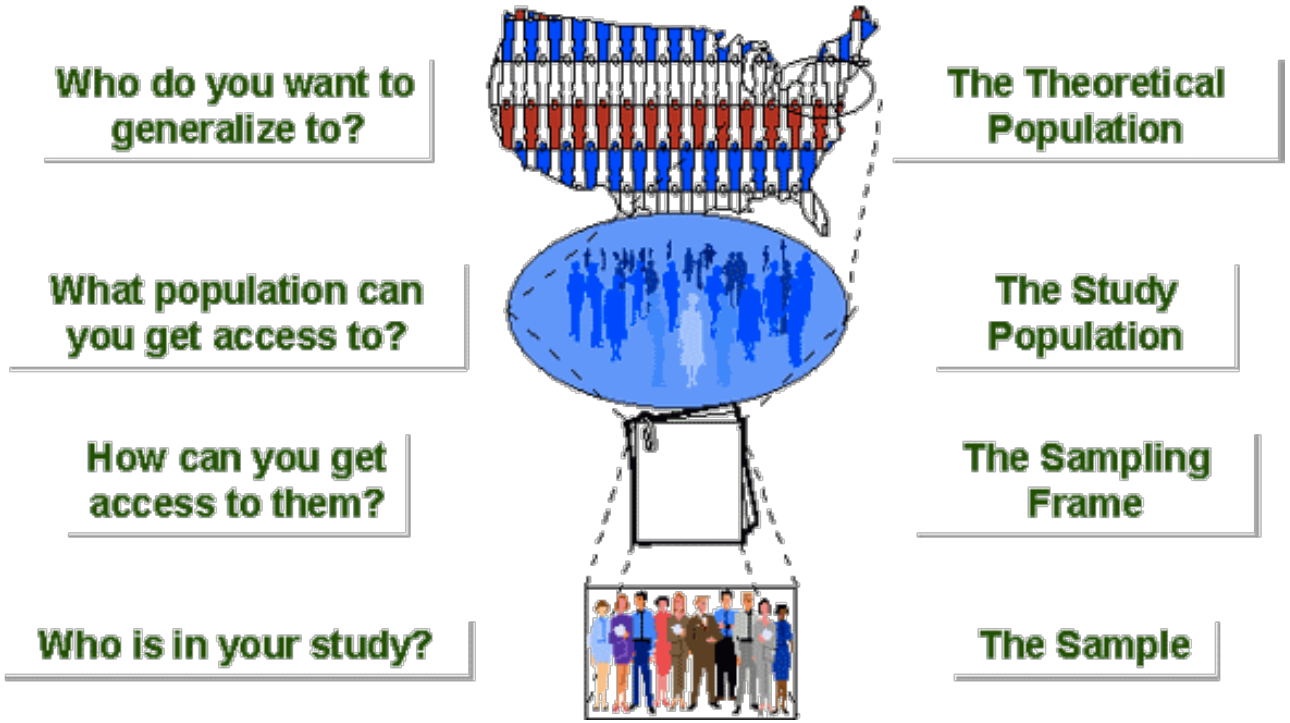
- Gary C. Ramseyer's First Internet Gallery of Statistics Jokes http://davidmlane.com/hyperstat/humorf.html (#162)

Sampling

- Sampling is the process of selecting units (e.g., people, organizations) from a population of interest so that by studying the sample we may fairly generalize our results back to the population from which they were chosen.

- In a broad epidemiological sense, the term ‘sampling’ refers to the process of facilitating access to a suitable selection of observation units or to the existing information about them.

- There are two general types ofvsampling methods:
  - Probabilistic (statistical)
  - Non-probabilistic (non-statistical) sampling.



Terminology

- **Population:** The group you wish to generalize in your study
- **Sampling frame:** a list of potentially eligible or accessible observation units
- **Sampling unit:** one of the units into which an aggregate is divided for the purpose of sampling, each unit being regarded as individual and indivisible when the selection is made (person, household, a product etc.)
- **Sampling fraction: t**he ratio of the sample size to the population size
- **Sample:** a subset of a frame where elements are selected based on a randomised process

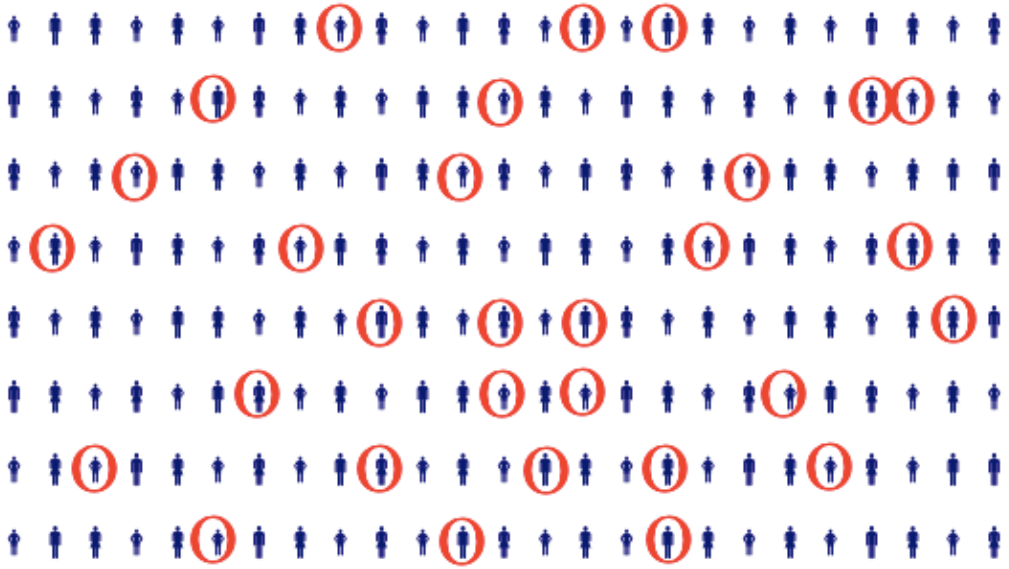
Types of sampling techniques

- Probabilistic
  - involves generating a list of potentially eligible observation units (sampling frame)
  - utilizes some form of random selection → there is a process or procedure that the different units in your population have equal probabilities of being chosen.
- Non-Probabilistic
  - No available sampling frame
  - There is no presumed equal chances to be chosen to the sample
  - Mainly there are two types
    - Convenience or Haphazard
    - Purposeful (Expert, Quota, Snowball, Heterogeneity)

Simple random sampling

- All subsets of the sampling frame are given an equal probability of being selected.

- Steps to follow:
  - assign a random unique number to each member of the population (lottery, table of random numbers, statistical software packages)
  - rank the sampling frame according to the randomly assigned numbers
  - select the first n of the ordered random numbers, where n is the required sample size



Systematic random sampling

- It is essential that the units in the population are randomly ordered.
  - Starting point bias

N = 100

want n = 20

N/n = 5

select a random number from 1-5: chose 4

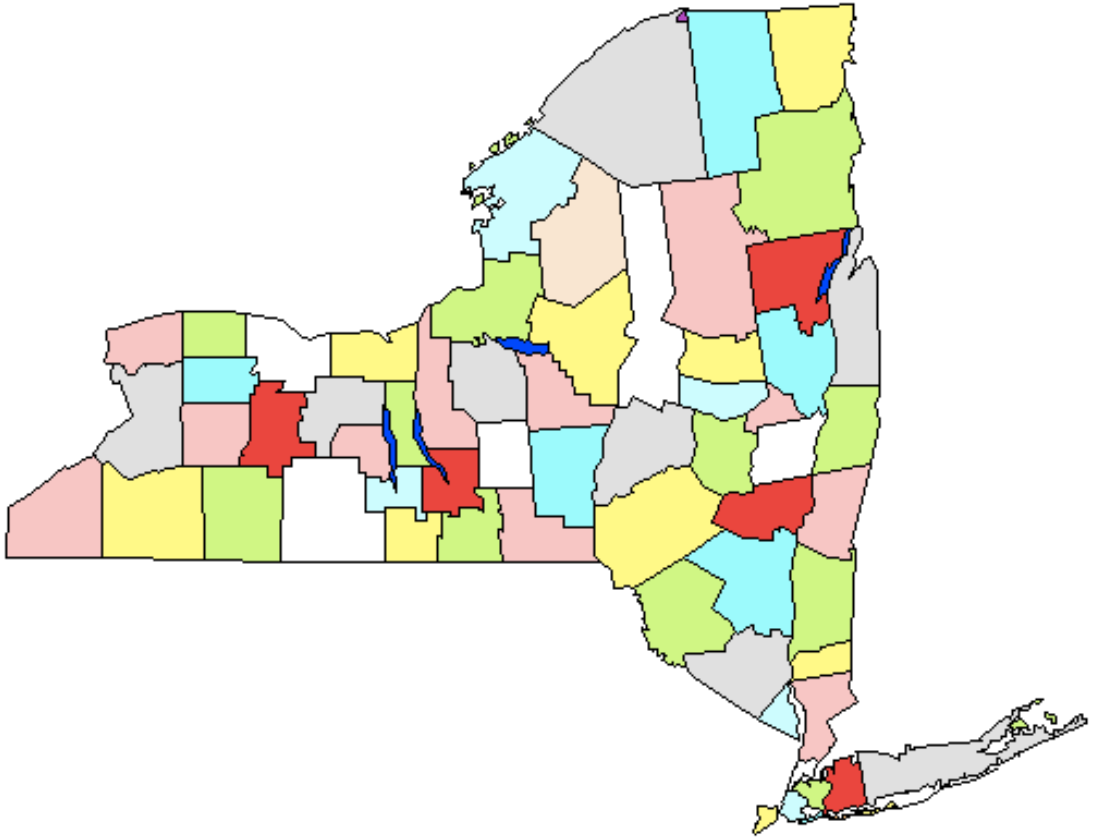
start with #4 and take every 5th unit

1	26	51	76
2	27	52	77
3	28	53	78
4	29	54	79
5	30	55	80
6	31	56	81
7	32	57	82
8	33	58	83
9	34	59	84
10	35	60	85
11	36	61	86
12	37	62	87
13	38	63	88
14	39	64	89
15	40	65	90
16	41	66	91
17	42	67	92
18	43	68	93
19	44	69	94
20	45	70	95
21	46	71	96
22	47	72	97
23	48	73	98
24	49	74	99
25	50	75	100

Cluster sampling

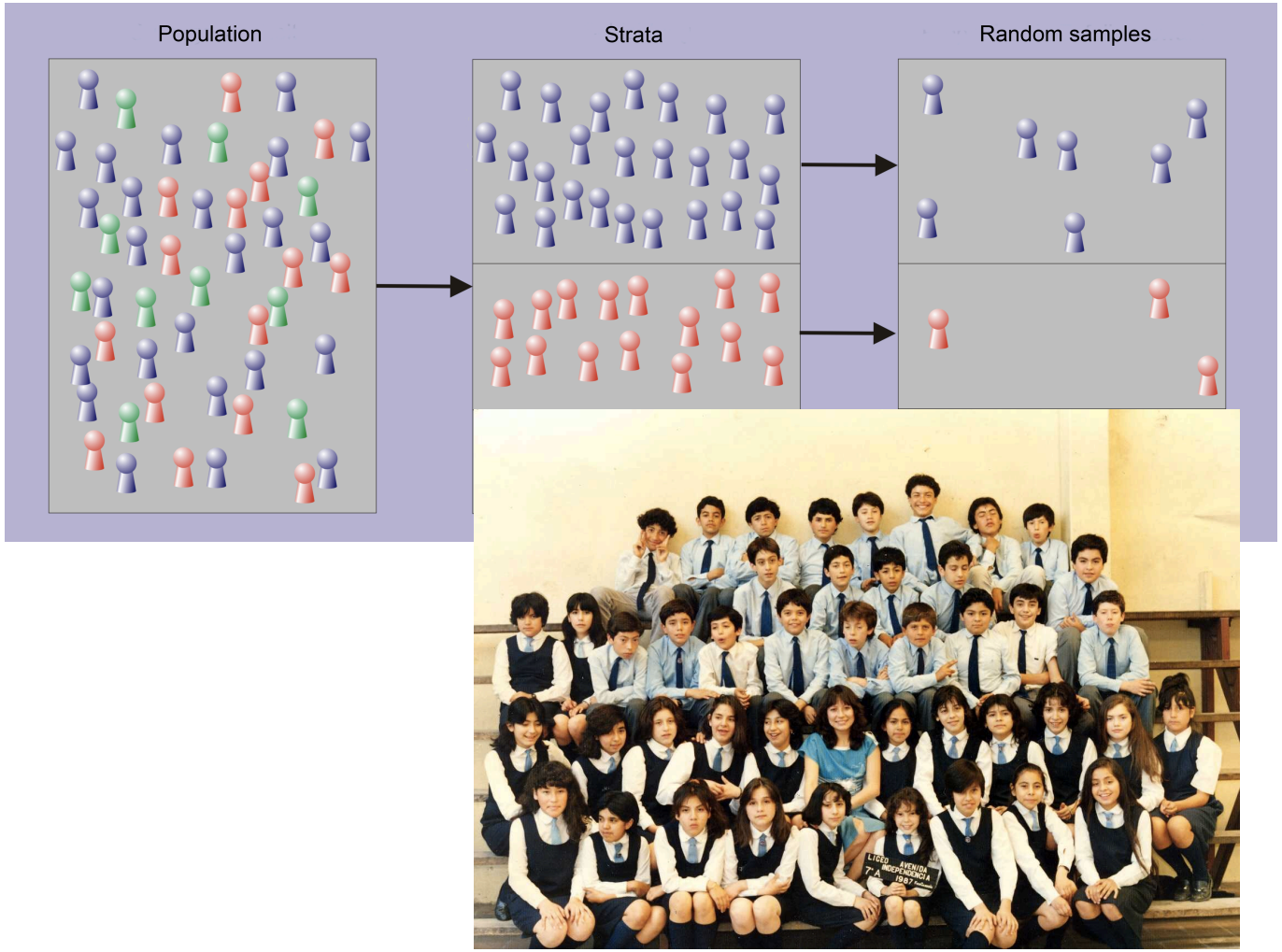
- Sampling unit is not individuals but clusters → village, neighborhood, school, factories etc.
- Design Effect (DEFF) is used to ensure precision

- Steps to follow
  - divide population into clusters (usually along geographic boundaries/cluster units such as schools)
  - randomly sample clusters
  - measure all units within sampled clusters



Stratified random sampling

- Stratified sampling divides the population into non-overlapping subgroups (strata) according to some important characteristic, such as sex, race, age category, or socioeconomic status, and selects a sample from each homogeneous subgroup.
- It assures that you will be able to represent not only the overall population, but also key subgroups of the population,



Type of statistical sampling	Sampling unit: example	Sampling frame: list, sampled units in bold
Random sampling	Individual school children in a region	1, **2**, 3, 4, **5, 6,** 7, **8**, 9, 10, 11, **12, 13, 14**, 15, … *(Individuals are randomly chosen from the list)*
Systematic sampling	Individual school children in a region	1, **2**, 3, **4**, 5, **6**, 7, **8**, 9, **10**, 11, **12**, 13, **14**, 15, … *(Every n*th *individual is chosen from list)*
Cluster sampling	Classes of school children in a region	1, 2, **3, 4**, 5, **6,** 7, 8, 9, **10, 11, 12,** 13, 14, **15**, … *(Classes are randomly chosen from the list; all pupils from the selected classes are invited)*

Illustration of random sampling, systematic sampling, and cluster sampling

Sample size calculation

- PS Power and Sample Size Calculator http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize
- G*Power: Statistical Power Analyses http://www.gpower.hhu.de/
- CDC Epi Info™ User Guide: https://www.cdc.gov/epiinfo/user-guide/statcalc/statcalcandopenepi.html
- CDC OpenEpi: http://openepi.com/Menu/OE_Menu.htm

Expand All | Collapse
Home
Info and Help
Language/Options/Settings
Calculator
Counts
Std.Mort.Ratio
Proportion
Two by Two Table
Dose-Response
R by C Table
Matched Case Control
Screening
Person Time
1 Rate
Compare 2 Rates
Continuous Variables
Mean CI
Median/%ile CI
t test
ANOVA
Sample Size
Proportion
Unmatched CC
Cohort/RCT
Mean Difference
Power
Random numbers
Searches
Google--Internet
PubMed--MEDLARS
Internet Links
Download OpenEpi
Development

OpenEpi Open Source Epidemiologic Statistics for Public Health

Now in English, French, Spanish, Italian, and Portuguese

Version 3.01 Updated 2013/04/06 ***Try it in a Smartphone browser!***

OpenEpi provides statistics for counts and measurements in descriptive and analytic studies, stratified analysis with exact confidence limits, matched pair and person-time analysis, sample size and power calculations, random numbers, sensitivity, specificity and other evaluation statistics, R x C tables, chi-square for dose-response, and links to other useful sites.

OpenEpi is free and **open source** software for epidemiologic statistics. It can be run from a web server or downloaded and run without a web connection. A server is not required. The programs are written in JavaScript and HTML, and should be compatible with recent Linux, Mac, and PC browsers, regardless of operating system. (If you are seeing this, your browser settings are allowing JavaScript.) The programs can be run in the browsers of many iPhone and Android cellphones

Test results are provided for each module so that you can judge reliability, although it is always a good idea to check important results with software from more than one source. Links to hundreds of Internet calculators are provided.

The programs have an open source license and can be downloaded, distributed, or translated. Some of the components from other sources have licensing statements in the source code files. Licenses referred to are available in full text at OpenSource.org/licenses. OpenEpi development was supported in part by a grant from the Bill and Melinda Gates Foundation to Emory University, Rollins School of Public Health.

A toolkit for creating new modules and for translation is included. Please let us know if you would like to collaborate in this way. Suggestions, comments, and expressions of interest in contributing to this effort should be sent by email to: andy.dean@gmail.com, cdckms@sph.emory.edu, and msoe@cdc.gov

Suggested citation: **Dean AG, Sullivan KM, Soe MM. OpenEpi: Open Source Epidemiologic Statistics for Public Health, Version. www.OpenEpi.com, updated 2013/04/06, accessed 2018/06/17.**

Example for a cross-sectional study

- A simple random or systematic sample is to be used to survey a urban city population of 280,000 concerning satisfaction with particular aspects of health care. The expected result is unknown, but guesstimated to be about 50% positive and 50% negative.

- What is the sample size needed?

- What would be the sample size if were to use cluster sampling with a design effect of 1.8?

OpenEpi page

Sample Size for % Frequency in a Population (Random Sample)		
Population size	280000	If large, leave as one million
Anticipated % frequency(p)	50	Between 0 & 99.99. If unknown, use 50%
Confidence limits as +/- percent of 100	5	Absolute precision %
Design effect (for complex sample surveys--DEFF)	1.0	1.0 for random sample

Sample Size for Frequency in a Population

Population size(for finite population correction factor or fpc)(*N*):	280000
Hypothesized % frequency of outcome factor in the population (*p*):	50%+/-5
Confidence limits as % of 100(absolute +/- %)(*d*):	5%
Design effect (for cluster surveys-*DEFF*):	1

Sample Size(*n*) for Various Confidence Levels

ConfidenceLevel(%)	Sample Size
95%	384
80%	165
90%	271
97%	471
99%	662
99.9%	1079
99.99%	1507

Equation

Sample size $n = [DEFF*Np(1-p)]/ [(d^2/Z^2_{1-\alpha/2}*(N-1)+p*(1-p)]$

Results from OpenEpi, Version 3, open source calculator--SSPropor
Print from the browser with ctrl-P
or select text to copy and paste to other programs.

Example for case-control studies

- 25 postoperative infections occurred in a hospital. How many controls should be chosen for each to conduct a case-control study in which Nurse A was present at 5 (20%) surgeries, but at only 50% of other surgeries?

OpenEpi page

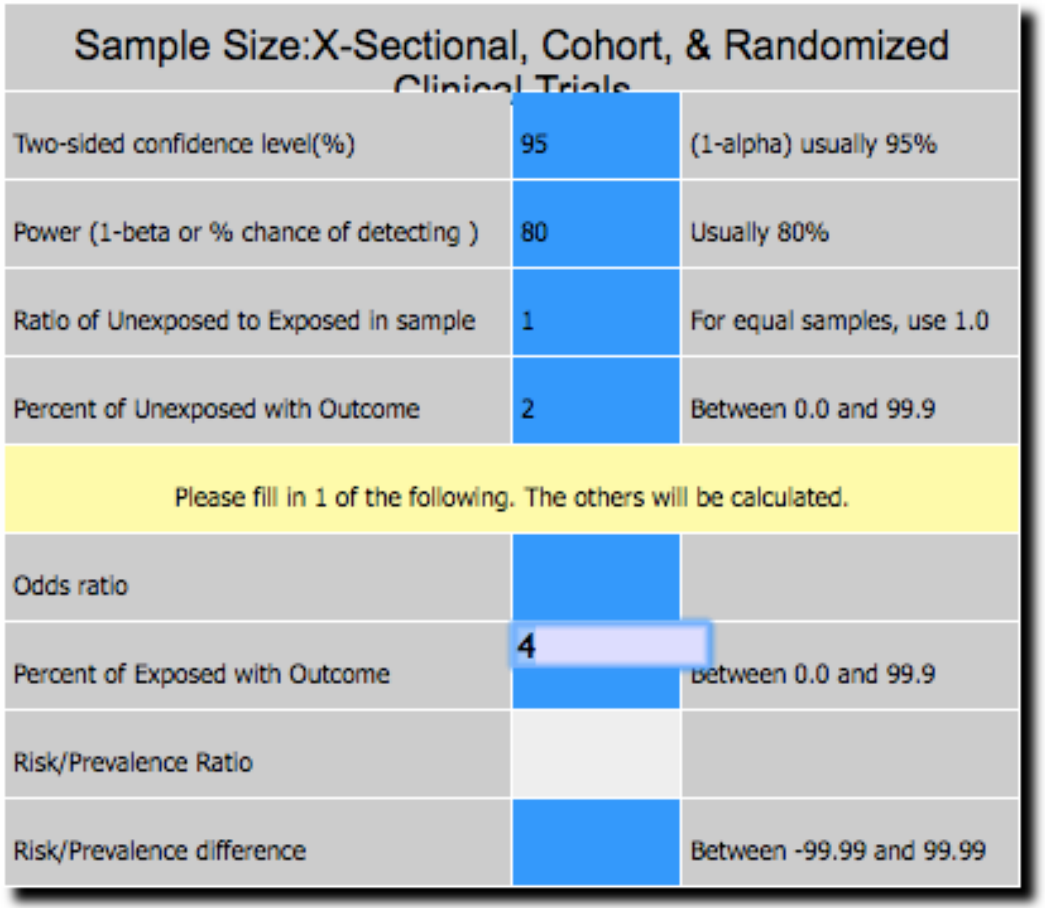
Sample Size for Unmatched Case Control Study		
Two-sided confidence level	95	(1-alpha) usually 95%
Power(% chance of detecting)	80	Usually 80%
Ratio of Controls to Cases	1	For equal samples, use 1.0
Percent of controls exposed	20	Between 0.0 and 99.99
Please fill in one of the following. The other will be calculated.		
Odds ratio		
Percent of cases with exposure	50	Between 0.0 and 99.99

Sample Size for Unmatched Case Control Study		
Two-sided confidence level	95	(1-alpha) usually 95%
Power(% chance of detecting)	80	Usually 80%
Ratio of Controls to Cases	4	For equal samples, use 1.0
Percent of controls exposed	20	Between 0.0 and 99.99
Please fill in one of the following. The other will be calculated.		
Odds ratio	4.00	
Percent of cases with exposure	50	Between 0.0 and 99.99

Example for a cohort study

- A phase I clinical trial is to be conducted to determine the frequency of adverse effects from a new medication.
- Two groups of equal size will be randomly allocated to receive the drug or a placebo.
- The effect being measured is expected to occur in 2% of the placebo group and 4% in the drug group.
- How large must the groups be to detect an adverse effect with 95% confidence and 80% power?

OpenEpi page



Sample Size:X-Sectional, Cohort, & Randomized Clinical Trials		
Two-sided confidence level(%)	95	(1-alpha) usually 95%
Power (1-beta or % chance of detecting)	80	Usually 80%
Ratio of Unexposed to Exposed in sample	1	For equal samples, use 1.0
Percent of Unexposed with Outcome	2	Between 0.0 and 99.9
Please fill in 1 of the following. The others will be calculated.		
Odds ratio		
Percent of Exposed with Outcome	4	Between 0.0 and 99.9
Risk/Prevalence Ratio		
Risk/Prevalence difference		Between -99.99 and 99.99

Sample Size:X-Sectional, Cohort, & Randomized Clinical Trials

Two-sided significance level(1-alpha):	95
Power(1-beta, % chance of detecting):	80
Ratio of sample size, Unexposed/Exposed:	1
Percent of Unexposed with Outcome:	2
Percent of Exposed with Outcome:	4
Odds Ratio:	2
Risk/Prevalence Ratio:	2
Risk/Prevalence difference:	2

	Kelsey	Fleiss	Fleiss with CC
Sample Size - Exposed	1146	1144	1242
Sample Size-Nonexposed	1146	1144	1242
Total sample size:	2292	2288	2484

References

Kelsey et al., Methods in Observational Epidemiology 2nd Edition, Table 12-15
Fleiss, Statistical Methods for Rates and Proportions, formulas 3.18 &3.19

CC = continuity correction
Results are rounded up to the nearest integer.
Print from the browser menu or select, copy, and paste to other programs.

G*Power

- http://www.gpower.hhu.de/
- G*Power is a tool to compute statistical power analyses for many different *t* tests, *F* tests, χ2 tests, *z*tests and some exact tests.
- G*Power can also be used to compute effect sizes and to display graphically the results of power analyses.

Effect Size (ES)

- ES is the *magnitude of the difference between groups*.
- The *absolute effect size* is the difference between the average, or mean, outcomes in two different intervention groups.
- Effect size helps readers understand the magnitude of differences found, whereas statistical significance examines whether the findings are likely to be due to chance.

Common effect size indices

Index	Description[b]	Effect Size	Comments
Between groups			
Cohen's d[a]	$d = M_1 - M_2 / s$ $M_1 - M_2$ is the difference between the group means (M); s is the standard deviation of either group	Small 0.2 Medium 0.5 Large 0.8 Very large 1.3	Can be used at planning stage to find the sample size required for sufficient power for your study
Odds ratio (OR)	$\frac{\text{Group 1 odds of outcome}}{\text{Group 2 odds of outcome}}$ If OR = 1, the odds of outcome are equally likely in both groups	Small 1.5 Medium 2 Large 3	For binary outcome variables Compares odds of outcome occurring from one intervention vs another
Relative risk or risk ratio (RR)	Ratio of probability of outcome in group 1 vs group 2; If RR = 1, the outcome is equally probable in both groups	Small 2 Medium 3 Large 4	Compares probabilities of outcome occurring from one intervention to another
Measures of association			
Pearson's r correlation	Range, −1 to 1	Small ±0.2 Medium ±0.5 Large ± 0.8	Measures the degree of linear relationship between two quantitative variables
r^2 coefficient of determination	Range, 0 to 1; Usually expressed as percent	Small 0.04 Medium 0.25 Large 0.64	Proportion of variance in one variable explained by the other

[a] Adapted from Ferguson et al.[9]
[b] Based on Soper.[7]

Example 1

- Your pilot study analyzed with a Student t-test reveals that group 1 (N=29) has a mean score of 30.1 (SD, 2.8) and that group 2 (N=30) has a mean score of 28.5 (SD, 3.5). The calculated P value= .06, and on the surface, the difference appears not significantly different.

- What would be the sample size?

Example 2

- A drug called X is being tested for its side effects on thrombocytes. Researchers plan to assess the thrombocyte levels of their patients before they administer the drug and do the same assessment after 3 months. However they need to know the sample size.
- In a previous study with a similar design the results of the participants were as follows:
  - Baseline assessment of Thrombocystes → Mean: 268.37 and SD: 60.28
  - Second assessment after 10 weeks → Mean: 264.62 and SD: 57.40
  - Correlation: 0.872

Thank you