

Measures of Effect & Potential Impact

Madhukar Pai, MD, PhD

Assistant Professor

McGill University, Montreal, Canada

madhukar.pai@mcgill.ca



McGill

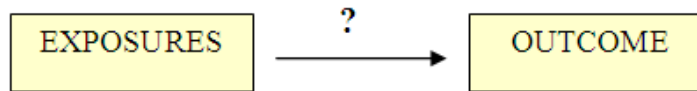
Big Picture



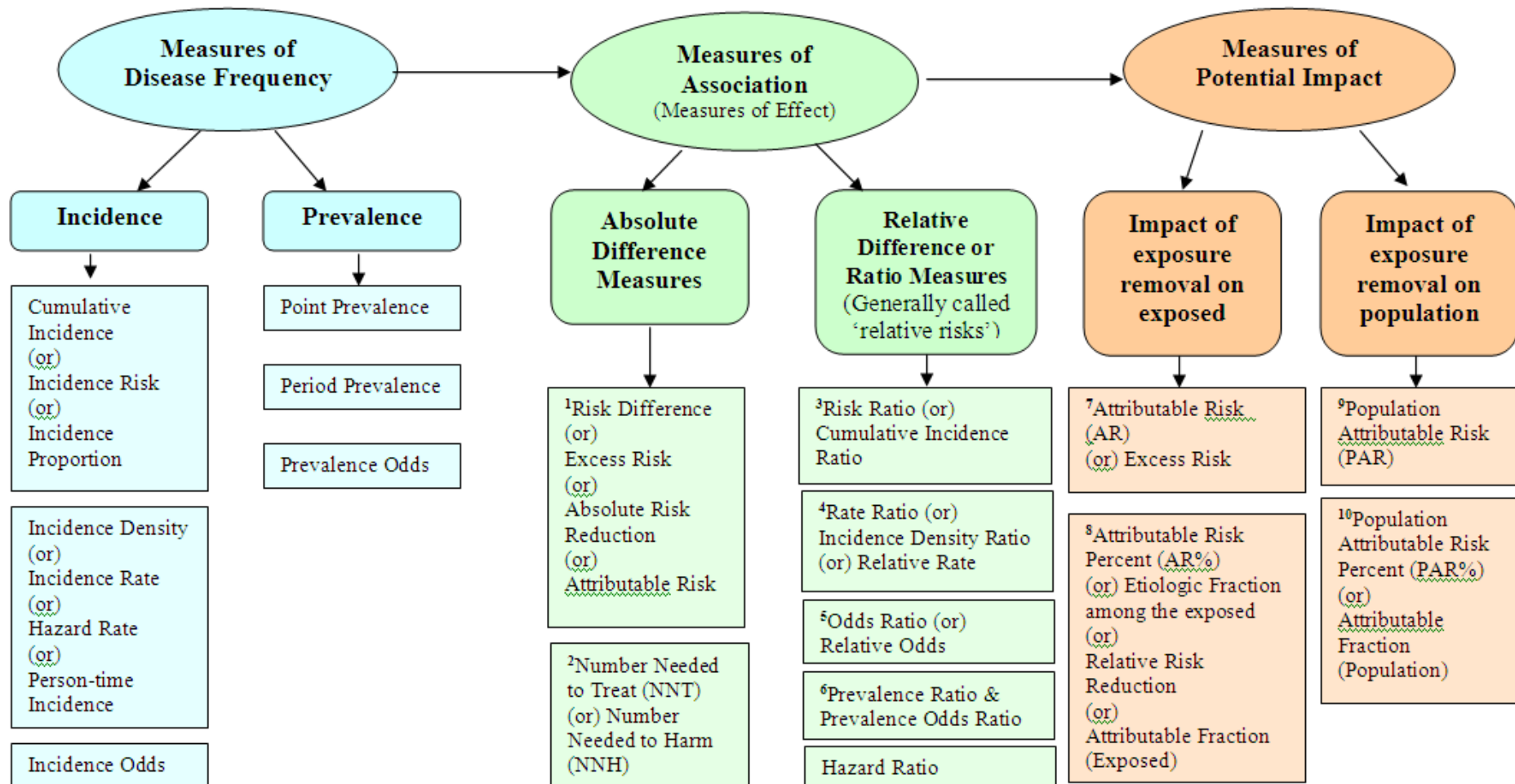
- First, we calculate measure of disease frequency in Group 1 vs. Group 2 (e.g. exposed vs. unexposed; treatment vs. placebo, etc.)
- Then we calculate measures of effect (which aims to quantify the strength of the association):
 - Either as a ratio or as a difference:
 - Ratio = (Measure of disease, group 1) / (Measure of disease, group 2)
 - Difference = (Measure of disease, group 1) - (Measure of disease, group 2)
- Lastly, we calculate measures of impact, to address the question: if we removed or reduced the exposure, then how much of the disease burden can we reduce?
 - Impact of exposure removal in the exposed group (AR, AR%)
 - Impact of exposure remove in the entire population (PAR, PAR%)

Note: there is some overlap between measures of effect and impact

AN OVERVIEW OF MEASUREMENTS IN EPIDEMIOLOGY [VER 3, 2007]



Epidemiology is about identifying associations between exposures and outcomes. To identify any association, exposures and outcomes must first be measured in a quantitative manner. Then rates of occurrence of events are computed. These measures are called “*measures of disease frequency*.” Once measured, the association between exposures and outcomes are then evaluated by calculating “*measures of association or effect*.” Finally, the impact of removal of an exposure on the outcome is evaluated by computing “*measures of potential impact*.” In general, measures of disease frequency are needed to generate measures of association, and both these are needed to get measures of impact. There is some overlap between these measures, and terminology is poorly standardized.



Measures of effect: standard 2 x 2 contingency epi table for count data (cohort or case-control or RCT or cross-sectional)

	Disease - yes	Disease - no	Column total (Margins)
Exposure - yes	a	b	a+b
Exposure - no	c	d	c+d
Row total (Margins)	a+c	b+d	a+b+c+d

Measures of effect: 2 x 2 contingency table for a cohort study or RCT with person-time data

	Disease - yes	Disease - no	Total
Exposure - yes	a	-	PY_e
Exposure - no	c	-	PY_0
Total	a+c	-	$PY_e + PY_0$

Person-time in the unexposed group = PY_0

Person-time in the exposed group = PY_e

STATA format for a 2x2 table [note that exposure and disease are transposed!]: be careful when using software – they may not use the conventional 2 x 2 format

	Exposure - yes	Exposure - no	Column total (Margins)
Disease- yes	a	b	a+b
Disease- no	c	d	c+d
Row total (Margins)	a+c	b+d	a+b+c+d

Some textbooks also use this format



MEASURES OF EFFECT

Ratio measures (loosely called “relative risks”)

- In a cohort study or RCT with count data:

Risk Ratio = CIR = I_e / I_0 [where I refers to cumulative incidence]

$$RR = CIR = \{a/(a+b)\} / \{c/(c+d)\}$$

- In a cohort study or RCT with person-time data:

Rate Ratio = IDR = I_e / I_0 [where I refers to incidence density]

- In a case control study (count data):

RR = OR = odds of exp given D / odds of exp given no D

$$RR = OR = ad / bc$$

- In a cross-sectional study (count data):

RR = Prevalence Ratio = $\text{prevalence}_e / \text{prevalence}_0$

$$RR = \text{Prevalence Odds Ratio} = ad / bc$$

Risk difference measures

- In a cohort study or RCT with count data:

Risk Difference = $RD = I_e - I_0$ [where I refers to cumulative incidence]

$$RD = \{a/(a+b)\} - \{c/(c+d)\}$$

RD is also called “Excess Risk” or “Attributable Risk”

- In a cohort study or RCT with person-time data:

Rate Difference = $RD = I_e - I_0$ [where I refers to incidence density]

Plug & chug epi software



OpenEpi Open Source Epidemiologic Statistics for Public Health
Version 2 Updated 2007-06-05

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

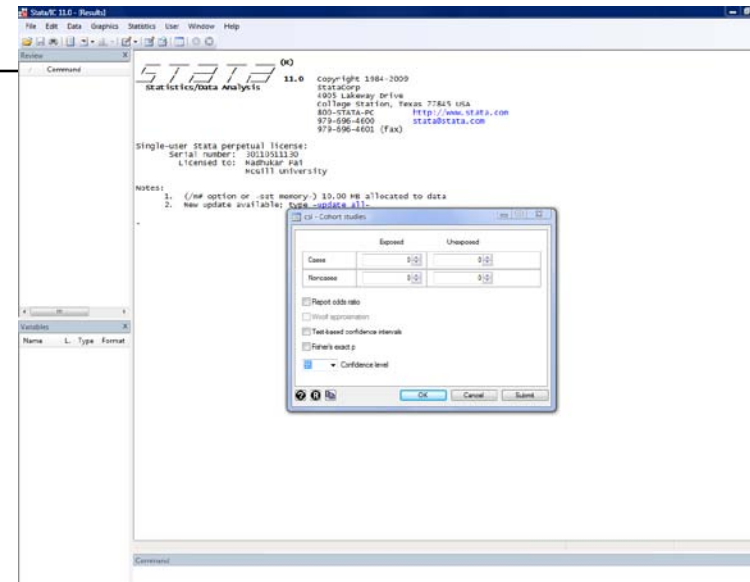
OpenEpi is free and open source software for epidemiologic statistics. It can be run from a web server or downloaded and run without a web connection. A server is not required. The programs are written in JavaScript and HTML, and should be compatible with recent Linux, Mac, and PC browsers, regardless of operating system. (If you are seeing this, your browser settings are allowing JavaScript.) A new tabbed interface avoids popup windows except for help files.

Text 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 msoc@sph.emory.edu

Suggested citation: Dean AG, Sullivan KM, Soe MM. OpenEpi: Open Source Epidemiologic Statistics for Public Health, updated (date). www.OpenEpi.com, accessed (date).



STATA 11.0

Copyright 1984-2009
STATACORP
4901 Lakeway Drive
College Station, Texas 77845 USA
805-576TA-PC http://www.stata.com
979-596-4400 statadata.com
979-596-4403 (Fax)

Single-user STATA perpetual license:
Serial number: J01195111.00
Licensed to: RABHAR, RAB
ncu111 university

Notes:
1. /nw option or -set memory- 10.00 MB allocated to data
2. New update available: [type update all](#)

ci - Cohort Studies

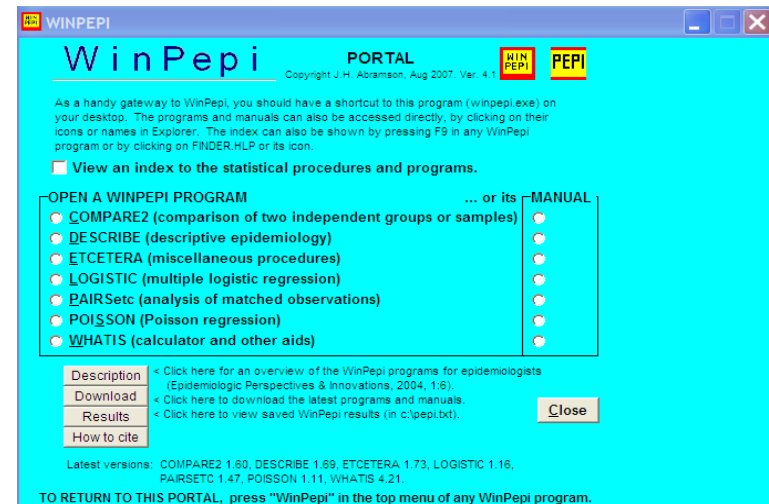
	Exposed	Unexposed
Case	1/20	1/20
Noncase	1/20	1/20

☐ Report odds ratio
☐ Wald approximation
☐ Test based confidence intervals
☐ Fisher's exact p
☒ Confidence limit

OK Cancel Submit

Be sure you can do all calculations by hand, before using these software!

In exams, you will be expected to do them by hand (using calculators)



WINPEPI PORTAL Copyright J. H. Abramson, Aug 2007, Ver. 4.1

As a handy gateway to WinPepi, you should have a shortcut to this program (winpepi.exe) on your desktop. The programs and manuals can also be accessed directly, by clicking on their icons or names in Explorer. The index can also be shown by pressing F9 in any WinPepi program or by clicking on FINDER.HLP or its icon.

View an index to the statistical procedures and programs.

OPEN A WINPEPI PROGRAM ... or its MANUAL

- COMPARE2 (comparison of two independent groups or samples)
- DESCRIBE (descriptive epidemiology)
- ETCETERA (miscellaneous procedures)
- LOGISTIC (multiple logistic regression)
- PAIRSETC (analysis of matched observations)
- POISSON (Poisson regression)
- WHATIS (calculator and other aids)

Description: Click here for an overview of the WinPepi programs for epidemiologists (Epidemiologic Perspectives & Innovations, 2004, 1:8).
Download: Click here to download the latest programs and manuals.
Results: Click here to view saved WinPepi results (in c:\pepi.txt).

Close

Latest versions: COMPARE2 1.60, DESCRIBE 1.69, ETCETERA 1.73, LOGISTIC 1.16, PAIRSETC 1.47, POISSON 1.11, WHATIS 4.21.

TO RETURN TO THIS PORTAL, press "WinPepi" in the top menu of any WinPepi program.

Example: Measures of effect in RCTs

Ice cream evoked headaches (ICE-H) study: randomised trial of accelerated versus cautious ice cream eating regimen

Maya Kaczorowski, Janusz Kaczorowski

Cold stimulus headache, also known as ice cream headache, is a common problem and is reported to occur in about a third of a randomly selected population.¹ It was further suggested that the ice cream headache could be induced only in hot weather.² A Medline search from 1966 to August 2002 with the MeSH terms and combination operators “ice cream,” “headache,” and “randomized controlled trial” to identify English language trials in this area produced no results.

In order to fill this important knowledge gap, we compared the effect of two ice cream eating regimens on the incidence of ice cream induced headaches in a prospective randomised manner. The study was carried out during the winter to test whether this phenomenon was restricted to hot weather only.

Characteristics of the 145 participants at enrolment in the ICE-H trial. Values are numbers (percentages) of participants unless stated otherwise

Characteristic	Cautious eating group (n=72)	Accelerated eating group (n=73)
Female	42 (58)	27 (37)
Mean (SD) age (years)	12.7 (0.8)	12.7 (0.8)
Lifetime prevalence of ice cream headaches	58 (81)	57 (78)
Regular headaches (weekly)	18 (25)	19 (26)

Participants who received red dot questionnaires were given 100 ml of ice cream but were instructed to eat it in < 5 seconds. The temperature of the ice cream was not formally regulated throughout the study. There were six eating sessions between December 2001 and

Dalewood Middle School, Hamilton, ON, Canada
Maya Kaczorowski
grade 8 student

Department of Family Medicine, McMaster University, Hamilton, ON, Canada

Janusz Kaczorowski
associate professor

Correspondence to: J Kaczorowski
kaczorow@mcmaster.ca

BMJ 2002;325:1445-6

20 (27%) of 73 students in the accelerated eating group (eat <5 seconds) reported ice cream evoked headache compared with 9 (13%) of 72 students in the cautious eating group (eat >30 seconds).

Measures of effect

	Headache - yes	Headache - no	Column total (Margins)
Accelerated eating	20	53	73
Cautious eating	9	63	72
Row total (Margins)	29	116	145

Cumulative incidence in accelerated group =

Cumulative incidence in cautious group =

Risk Ratio or Cumulative Incidence Ratio (CIR) =

Risk difference =

Odds ratio (OR) =

Measures of effect

	Headache - yes	Headache - no	Column total (Margins)
Accelerated eating	20	53	73
Cautious eating	9	63	72
Row total (Margins)	29	116	145

Cumulative incidence in accelerated group = $20/73 = 27\%$

Cumulative incidence in cautious group = $9/72 = 12.5\%$

Risk Ratio or Cumulative Incidence Ratio (CIR) = $27/12.5 = 2.2$ [95% CI 1.07 to 4.49]

Risk difference = $27\% - 12.5\% = 14.5\%$ [95% CI 0.7% to 29%]

Odds ratio (OR) = $(20*63) / (53*9) = 2.6$ [95% CI 1.04 to 7.13]

Example: measures of effect in cohort studies

Prospective cohort study of prone sleeping position and sudden infant death syndrome

TERENCE DWYER ANNE-LOUISE B. PONSONBY
NEVILLE M. NEWMAN LAURA E. GIBBONS

Studies of the link between prone sleeping position and sudden infant death syndrome have been criticised on grounds of recall bias and for not taking into account possible confounding effects. To avoid recall bias and to allow measurement of important biological factors a prospective cohort study of the cause of sudden infant death syndrome (SIDS) is being conducted. The infants included are those at high risk of the syndrome as assessed by a perinatal score. Of the 3110 members of the cohort born between January, 1988, and end of March, 1990, 23 infants later died of SIDS. Sleep position information was available for 15 of these. Matched analysis to control for the confounding effects of infant birthweight and maternal age indicated that prone sleeping position was associated with an increased risk of SIDS (OR 4.47 95% CI [1.30–15.43]). The findings are strengthened by the results of a concurrent retrospective case-control study of 42 SIDS cases in which the prone position was also associated with an increased risk of SIDS (unadjusted OR 3.45 [1.59–7.49]).

Lancet 1991; **337**: 1244–47.

Here we present data from a prospective study on the relation between SIDS and prone sleeping position. Several likely confounders are controlled for. In addition, the strength of association between SIDS and prone or non-prone position on the basis of prospectively collected data was compared with that found by retrospective data collected on the same infants.

Methods

Tasmania, the island state of Australia, has approximately 7000 livebirths per annum, and its rate of SIDS—3.5 per 1000 livebirths—is considerably higher than that in other Australian states.¹⁹ A prospective cohort study was started in January, 1988, to investigate the cause of SIDS in Tasmania. The six obstetric hospitals taking part cover approximately 93% of livebirths in the state. Infants born within these hospitals are assessed according to a local scoring system²⁰ to predict infants at high risk of SIDS. The composite score is based on maternal age, birthweight, season of birth, sex, duration of the second stage of labour, and infant feeding. Infants with a score over a cut-off point²⁰ are eligible for the study. The cut-off point identifies a group which represents approximately one fifth of livebirths in the state. Multiple births are also included in the study. Infants with severe neonatal disease or a major congenital anomaly and infants for adoption are excluded from the study.



Example: cohort study

	SIDS	No SIDS	Column total (Margins)
Prone	9	837	846
Non-prone	6	1755	1761
Row total (Margins)	15	2592	2607

Lancet 1991

Cumulative incidence among
exposed (prone) =

Cumulative incidence among
unexposed =

Cumulative incidence ratio (CIR)
=

Risk difference =

Odds ratio (OR) =

Example: cohort study

	SIDS	No SIDS	Column total (Margins)
Prone	9	837	846
Non-prone	6	1755	1761
Row total (Margins)	15	2592	2607

Lancet 1991

Cumulative incidence among exposed (prone) = 1.06% or 10.6 per 1000

Cumulative incidence among unexposed = 0.34% or 3.4 per 1000

Cumulative incidence ratio (CIR) = 3.12 [95% CI 1.11 to 8.74]

Risk difference = 0.7% or 7 per 1000

Odds ratio (OR) = 3.15 [95% CI 1.00 to 10.77]

Example: measures of effect in case-control studies

Prone sleep position and the sudden infant death syndrome in King County, Washington: A case-control study

James A. Taylor, MD, James W. Krieger, MD, MPH, Donald T. Reay, MD, Robert L. Davis, MD, MPH, Richard Harruff, MD, PhD, and Lois K. Cheney

From the Department of Pediatrics, University of Washington, the Seattle-King County Department of Public Health, and the King County Medical Examiner's Office, Seattle, Washington

Objective: To determine whether the prone sleep position was associated with an increased risk of the sudden infant death syndrome (SIDS).

Study design: Population-based case-control study.

Participants: Case subjects were infants who died of SIDS in King County, Washington. Control subjects were randomly selected infants born in King County. Up to four control subjects were matched on date of birth to each case subject.

Methods: During the study period, November 1992 through October 1994, sleep-position data were collected on infants who died of SIDS by the King County Medical Examiner's Office during their investigation of the deaths. Parents of infants chosen as control subjects were contacted by telephone, and sleep position information was obtained. Infants who usually slept on their abdomen were classified as sleeping prone; those who usually slept on the side or back were categorized as sleeping nonprone. The adjusted odds ratio for prone sleep position as a risk factor for SIDS was calculated with conditional logistic regression after control for race, birth weight, maternal age, maternal marital status, household income, and maternal cigarette smoking during pregnancy.

Results: Sleep position data were collected on 47 infants with SIDS (77% of eligible infants) and 142 matched control subjects; 57.4% of infants who died of SIDS usually slept prone versus 24.6% of control subjects ($p < 0.00001$). The unadjusted odds ratio for prone sleep position as a risk factor for SIDS was 4.69 (95% confidence interval: 2.17, 10.17). After control for potentially confounding variables, the adjusted odds ratio for prone sleep position was 3.12 (95% confidence interval: 1.08, 9.03).

Conclusion: Prone sleep position was significantly associated with an increased risk of SIDS among a group of American infants. (J Pediatr 1996;128:626-30)



Overall, 57.4% (27/47) of all infants who died of SIDS usually slept in the prone position as opposed to 24.6% (35/142) of control infants

Example: case-control study

	Case	Control	Column total
Prone	27	35	62
Non-prone	20	107	127
Row total	47	142	189

Exposure odds among cases =

Exposure odds among controls =

Disease odds among exposed =

Diseased odds among non-exp =

Exposure odds ratio =

Disease odds ratio =

J Pediatrics 1996

Can you compute any incidence measure in a case-control study?

Example: case-control study

	Case	Control	Column total
Prone	27	35	62
Non-prone	20	107	127
Row total	47	142	189

Exposure odds among cases =
 $27/20$

Exposure odds among controls =
 $35/107$

Disease odds among exposed =
 $27/35$

Diseased odds among non-exp =
 $20/107$

Exposure odds ratio = $(27/20) / (35/107) = 4.13$ [95% CI 1.95 to 8.76]

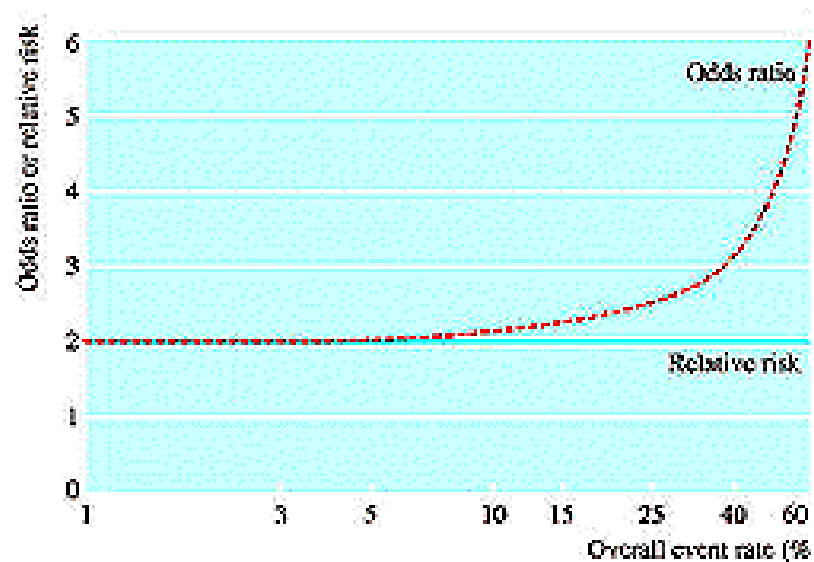
Disease odds ratio = $(27/35) / (20/107) = 4.13$ [95% CI 1.95 to 8.76]

J Pediatrics 1996

Later, we will see that, under specific circumstances, an OR from a case-control study will be a valid estimate of the IDR

OR vs RR: when do they diverge?

The odds will be close to the relative risk if the end point occurs relatively infrequently, say in less than 20%. If the outcome is more common then the odds ratio will considerably overestimate the relative risk



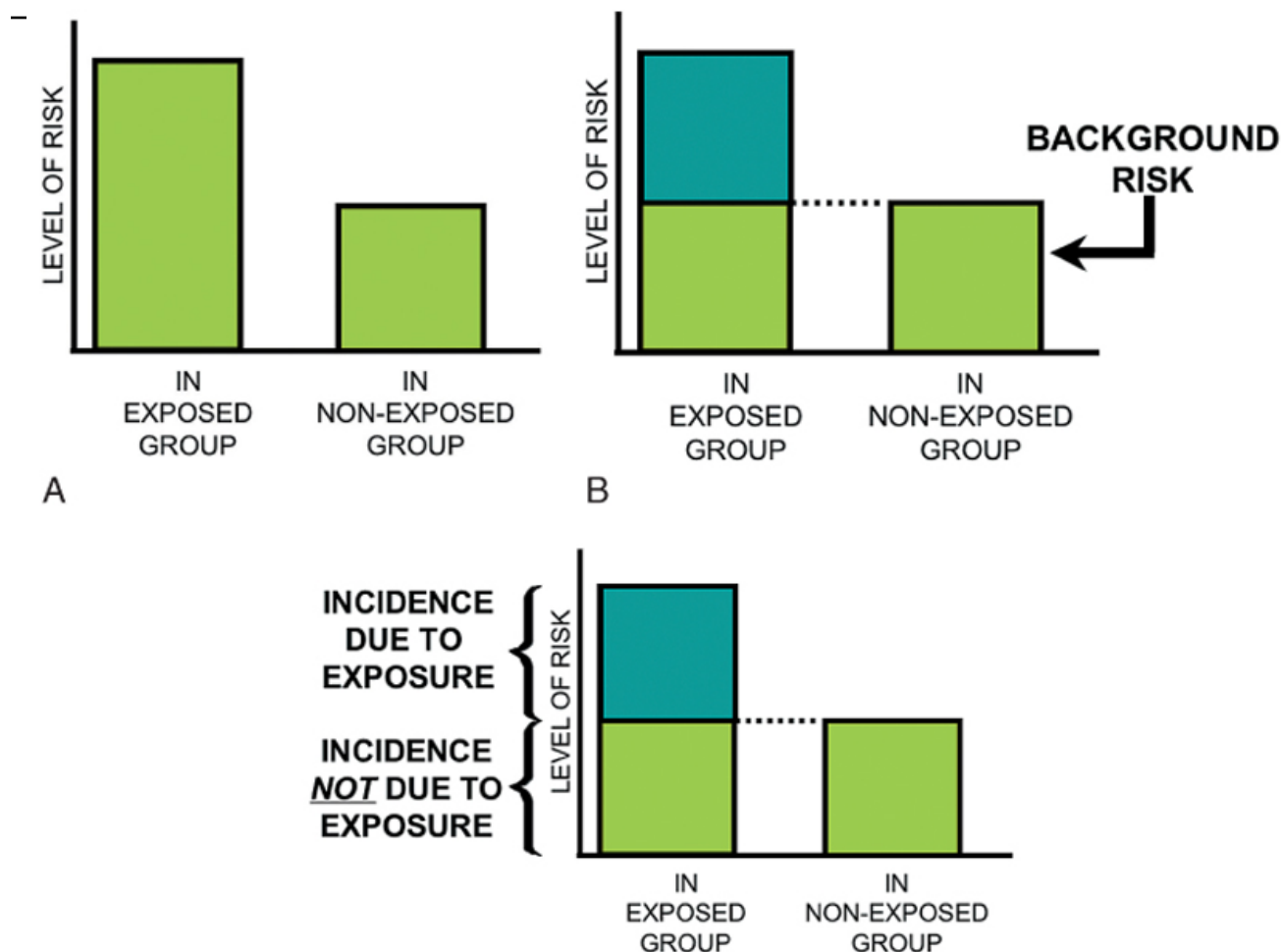


MEASURES OF POTENTIAL IMPACT

Measures of potential impact

- Impact of removing exposure in:
 - Exposed people (e.g. smokers)
 - All people (entire population – made up of both exposed and unexposed people)

The concept of background risk



C

Measures of potential impact

Measures of potential impact among the exposed:

- AR: Attributable risk = $(I_e - I_o)$
- AR%: Attributable risk % = $(I_e - I_o) / I_e = AR / I_e = (RR-1) / RR$

Measures of potential impact in the whole population:

- PAR: Population attributable risk = $I_t - I_o$
- PAR%: Population attributable risk % = PAR / I_t
 - Alternative formula: $\frac{P_{exp} (RR-1)}{P_{exp} (RR-1) + 1} * 100$

Where P_{exp} = Prevalence of exposure in the population

Attributable risk vs. Relative Risk

□ Relative risk

- Provides a measure of the strength of an association between an exposure and a disease
- Helps to evaluate the causal relationship between an exposure and a disease
- Magnitude of relative risk does not predict magnitude of attributable risk

□ Attributable risk

- Provides a measure of the public health impact of an exposure on the exposed group: if the exposure were removed, how much of the disease burden will be reduced?
- Assumes the exposure is causal
- Attributable risks for different risk factors do not add up to 100% (because multiple causes interact to cause disease – see later)

Attributable Risk in Cohort Studies

□ Attributable risk (AR)

- Synonym (and conceptually): AR = risk difference (RD)
- Provides information about absolute effect of an exposure removal

□ The excess risk of disease in the exposed

$$\begin{aligned} \text{AR} &= I_{\text{exposed}} - I_{\text{nonexposed}} \\ &= I_e - I_o \end{aligned}$$

The incidence measure can be either CI (cumulative incidence) or ID (incidence density)

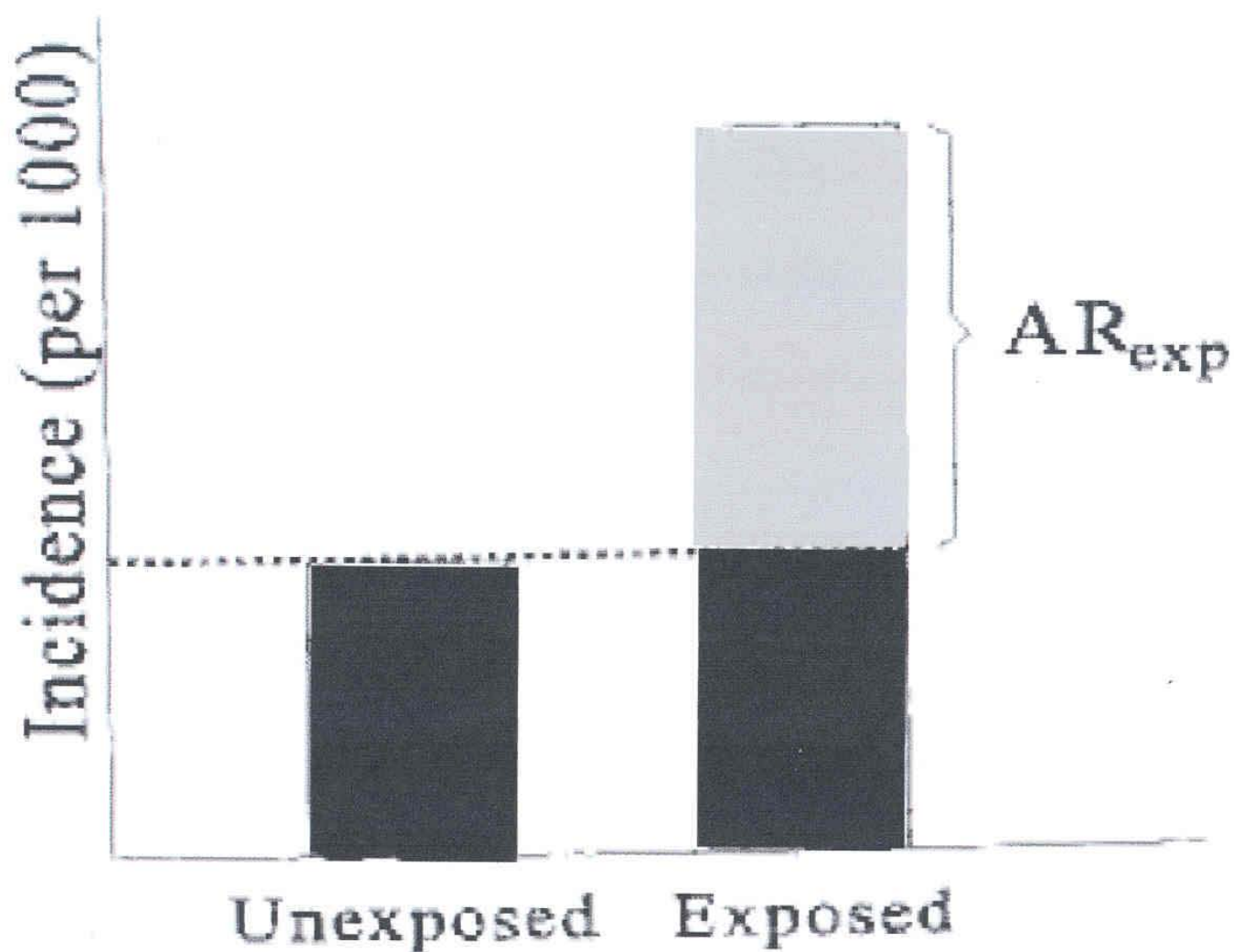


Figure 3-1 Attributable risk in the exposed.

from Szklo and Nieto: Epidemiology Beyond the Basics, 2000.

Attributable Risk

$$\text{Incidence in the } \underline{\text{exposed}} \text{ group} = \text{Incidence } \textit{not} \text{ due to the exposure (background incidence)} + \text{Incidence due to the exposure}$$

$$\text{Incidence in the } \underline{\text{nonexposed}} \text{ group} = \text{Incidence } \textit{not} \text{ due to the exposure (background incidence)}$$

Gordis: Epidemiology, 4th Edition.

Copyright © 2008 by Saunders, an imprint of Elsevier, Inc. All rights reserved

The concept of attributable risk.

Measures of effect: Attributable Risk example (Cohort data)

	SIDS	No SIDS	Column total
Prone	9	837	846
Non-prone	6	1755	1761
Row total	15	2592	2607

Cumulative incidence among exposed (prone) = 1.06% [10.6 per 1000]

Cumulative incidence among unexposed = 0.34% [3.4 per 1000]

Risk difference or attributable risk or excess risk (AR) = 0.7% or 7 per 1000

AR has same units as the incidence measure used (dimensionless if CI; time⁻¹ if ID)

Interpretation: Among every 1000 babies that sleep prone, there are 7 excess cases of SIDS attributable to prone sleeping.

Note: the interpretation of the AR is dependent upon the assumption that the risk factor (in this case prone sleeping position) is causal

The AR is a useful measure of the public health impact of a particular exposure: **if prone babies were made to sleep on their back, then 7 SIDS cases will be averted for every 1000 babies that sleep prone**

Measures of effect: AR% in a cohort study

- Suppose you wish to estimate, among exposed babies (prone), what proportion of cases of SIDS are due to prone posture?
- In other words, among all of the cases among the exposed, what proportion is represented by the excess cases?
- The appropriate measure is the attributable risk % (AR%)

$$AR\% = (AR) / I_e \times 100$$

$$AR\% = (I_e - I_o) / I_e \times 100$$

- Alternative formulation (very helpful when incidence measures are not available – e.g. case-control study):

$$AR\% = (RR - 1) / RR \times 100$$

AR% is also known as “Attributable Fraction (Exposed)” – defined as “**the proportion by which the incidence rate of the outcome among those exposed would be reduced if the exposure were eliminated**” [Porta, 2008]

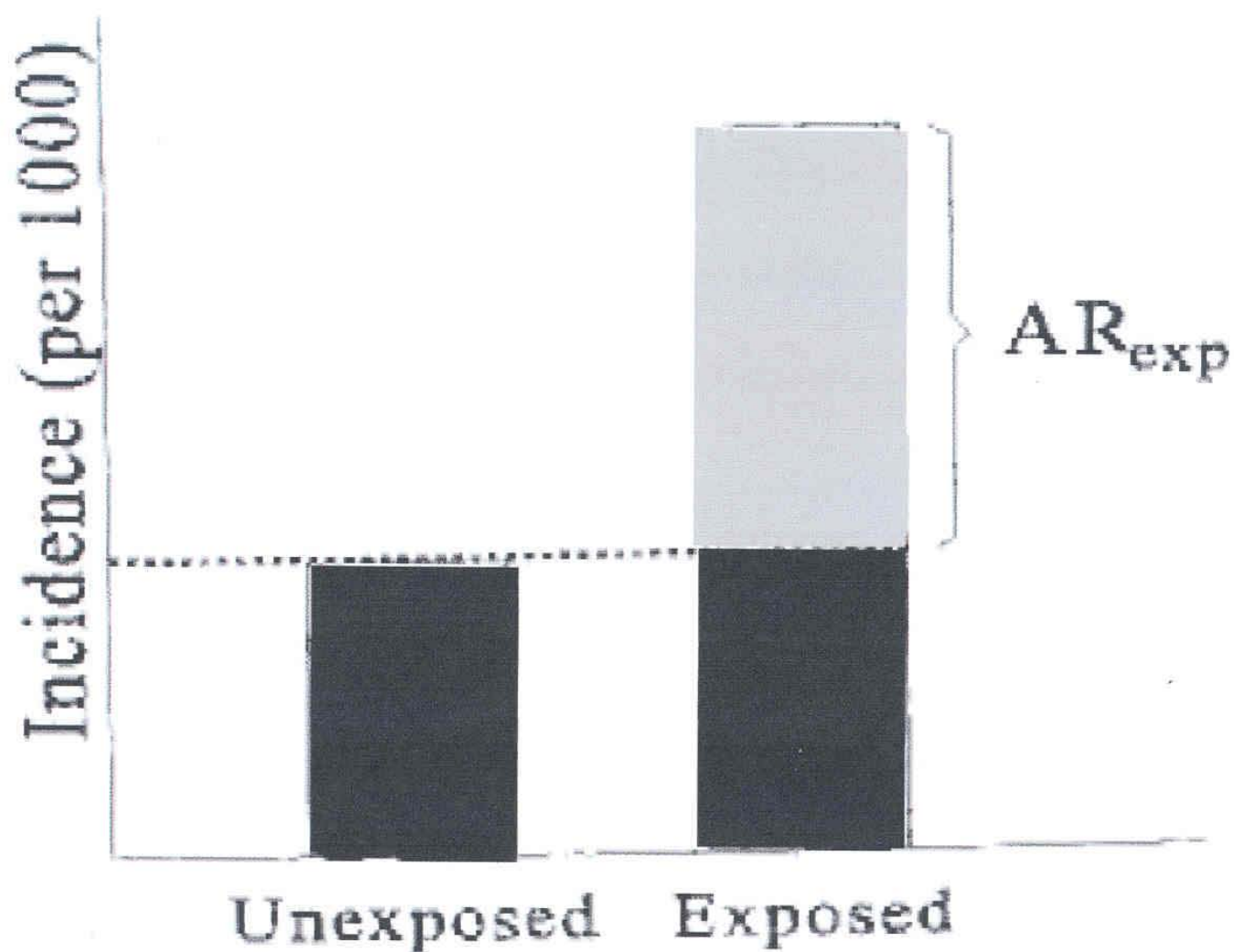


Figure 3-1 Attributable risk in the exposed.

from Szklo and Nieto: Epidemiology Beyond the Basics, 2000.

Measures of effect: Attributable Risk

Percent in a cohort study

- Suppose you wish to estimate, among exposed babies (prone), what proportion of cases of SIDS are due to prone posture?

- The appropriate measure is the attributable risk % (AR%)

$$\begin{aligned} \text{AR\%} &= (\text{AR}) / I_e \times 100 \\ &= \{(I_e - I_o) / I_e\} \times 100 \\ &= (0.7\%) / (1.06\%) \times 100 \\ &= 66\% \end{aligned}$$

- Interpretation: Among the prone sleeping babies, 66% of the cases of SIDS are attributable to the prone sleeping posture [so, not all SIDS is due to prone posture; in about a third of the cases, something other than prone posture is responsible]
- Alternative formulation:

$$\begin{aligned} \text{AR\%} &= (\text{RR} - 1) / \text{RR} \times 100 \\ &= \{(3.12 - 1) / 3.12\} \times 100 = 67\% \end{aligned}$$

Measures of effect: Attributable Risk in a case-control study

- ❑ In case control studies, it is (generally) not possible to estimate incidence rates.
- ❑ Therefore, it is not possible to directly estimate attributable risk (AR) in case control studies.
- ❑ However...it is possible to estimate attributable risk percent (AR%) using the following alternative expression for AR% (Note: this formula may also be used in cohort or in case control studies).

$$AR\% = (RR - 1) / (RR) \times 100$$

- ❑ In a case control study [previous example on SIDS], it will be:

- ❑ $(OR - 1) / (OR) \times 100$

$$AR\% = (4.13 - 1) / (4.13) \times 100 = 76\%$$

- ❑ Interpretation: **Among babies that sleep prone, 76% of cases of SIDS are due to prone sleeping posture.** [note 76% is higher than the estimate in the cohort study (66%) – because the case-control study reported a stronger association (OR = 4.13) than the cohort study (RR = 3.12)]
- ❑ Question: why did the case-control study report a stronger effect??

Measures of effect: Population attributable risk

- Population attributable risk (PAR): basic goal is to estimate the burden of disease due to the exposure on the entire population (not just among the exposed as is done with “attributable risk”)

$$PAR = I_{\text{total}} - I_{\text{unexposed}} = I_t - I_o$$

- Another formulation:

$$PAR = (AR)(P_e)$$

Utility of the measure: to determine which exposures have the most relevance to the health of a community: **if the exposure was removed from the population, then how much of the disease in the population will be averted?**

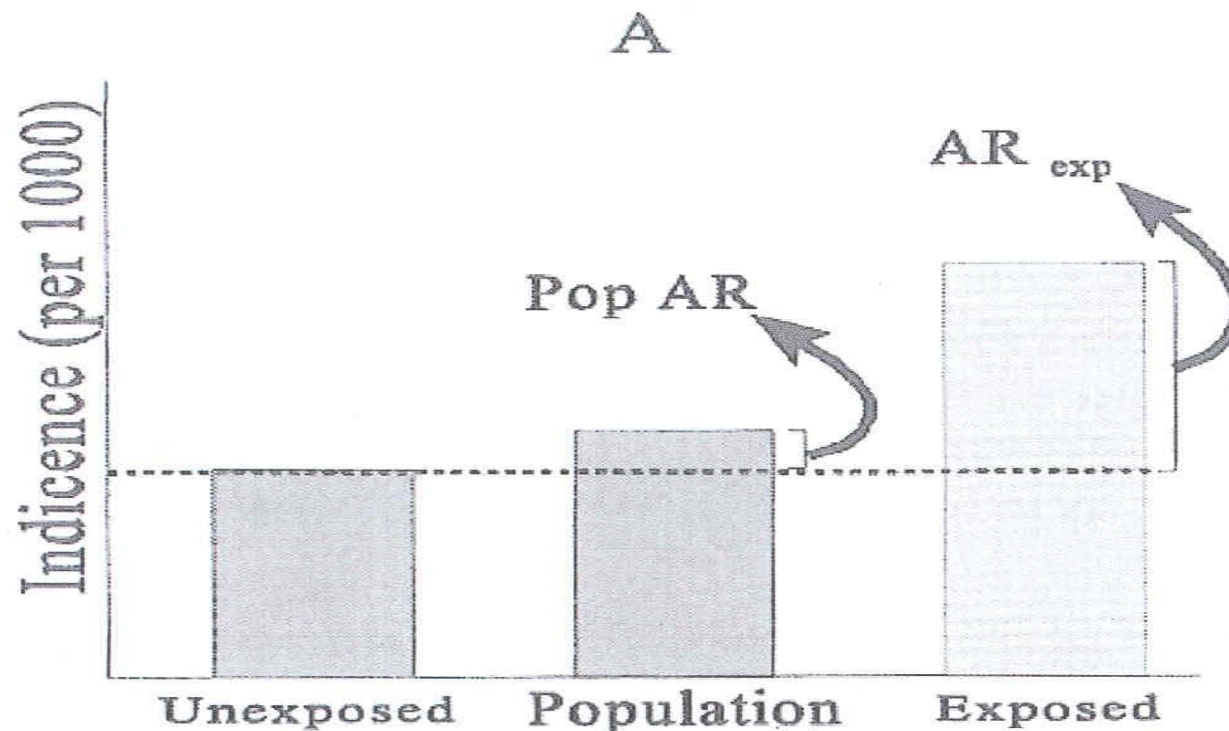


Figure 3-2: Szklo and Nieto,
Epidemiology Beyond the Basics, 2000

Measures of effect: Population attributable risk

	SIDS	No SIDS	Column total
Prone	9	837	846
Non-prone	6	1755	1761
Row total	15	2592	2607

Cumulative incidence among exposed (prone) = 10.6 per 1000

Cumulative incidence among unexposed = 3.4 per 1000

Cumulative incidence in the population = $15/2607 = 5.8$ per 1000

Risk difference or attributable risk or excess risk (AR) = 7 per 1000

$$PAR = I_t - I_0 = 5.8 - 3.4 = 2.4 \text{ per 1000}$$

- Interpretation: Among every 1000 babies in a population, there are 2 excess cases of SIDS attributable to prone sleeping: **if all babies in a population were made to sleep on their backs, then 2 SIDS cases can be averted for every 1000 babies**
- Note: PAR will always be less than AR. Why?

Measures of effect: Population attributable risk percent

- Population attributable risk percent (PAR%)
 - Definition: the proportion of disease in the study population that is attributable to the exposure and that, theoretically, would be eliminated if the exposure were removed.
 - $PAR\% = (PAR / I_t) \times 100$
 - PAR% is also known as “Attributable Fraction (Population)” – defined as “the proportion by which the incidence rate of the outcome in the entire population would be reduced if the exposure were eliminated” [Porta, 2008]
- SIDS example
 - $PAR = 2.4 \text{ per } 1000$
 - $PAR\% = 2.4 / 5.8 \times 100 = 41\%$
 - Interpretation: Making all babies sleep on their back would eliminate 41% of all cases of SIDS in the population.
 - Recall that the AR% was 66% (earlier slides)—obviously, the impact on the exposed group (measured by AR%) is greater than on the whole population (PAR%). Why?

Relative risk vs. population attributable risk

- Relative risk
 - Provides a measure of the strength of an association between an exposure and a disease
 - Helps to evaluate the causal relationship between an exposure and a disease
 - Magnitude of relative risk does not predict magnitude of attributable risk
- PAR%
 - Provides a measure of the public health impact of an exposure on the entire population
 - Assumes the exposure is causal
 - A strong RR may not translate to a large PAR% if the exposure is not widely prevalent in the population
 - Conversely, a weak RR may have a big PAR% if the exposure is very common (e.g. smoking, obesity, air pollution)
 - To appreciate this, see the alternative formula for PAR%

$$\text{PAR\%} = \frac{P_{\text{exp}} (RR - 1)}{P_{\text{exp}} (RR - 1) + 1} * 100$$

Where P_{exp} = Prevalence of exposure in the population

So, if P_{exp} is large, then even if the RR is small, it will still work out to a large PAR%

How PAR% is dependent on prevalence of exposure and RR

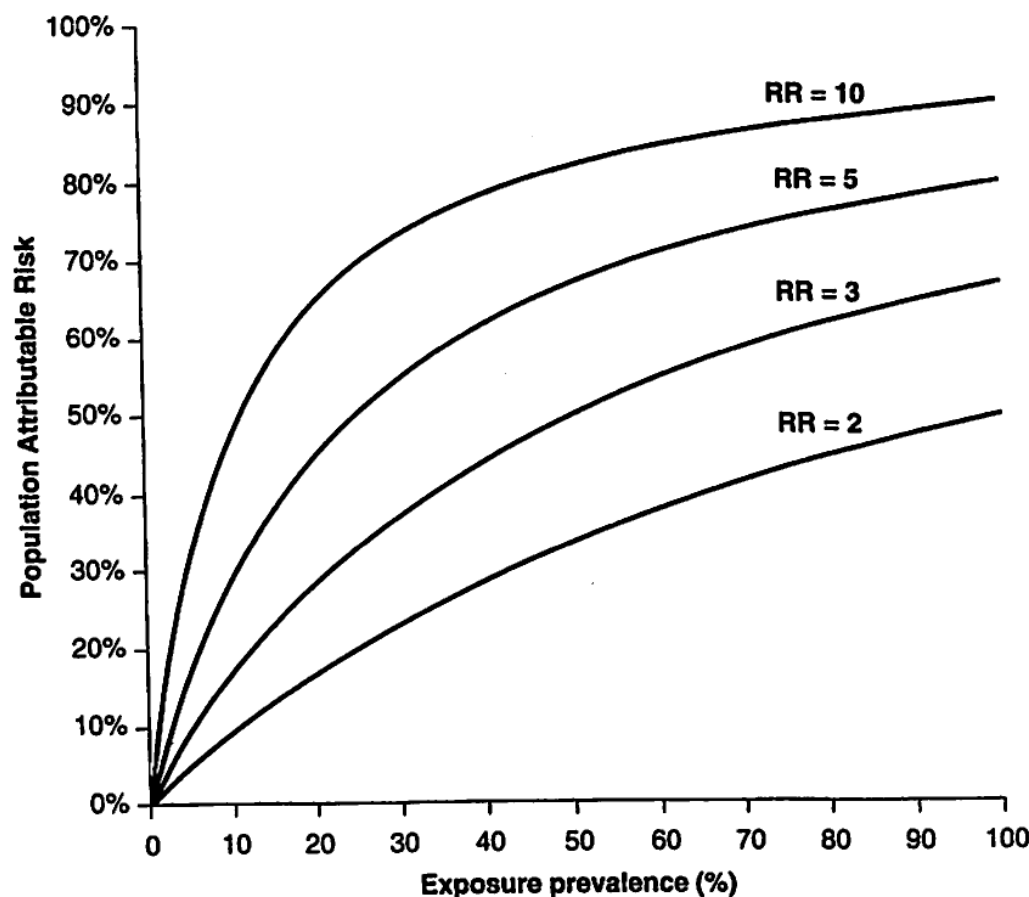


Figure 3–3 Population attributable risk: dependence on prevalence of exposure and relative risk.

How PAR% is dependent on prevalence of exposure and RR

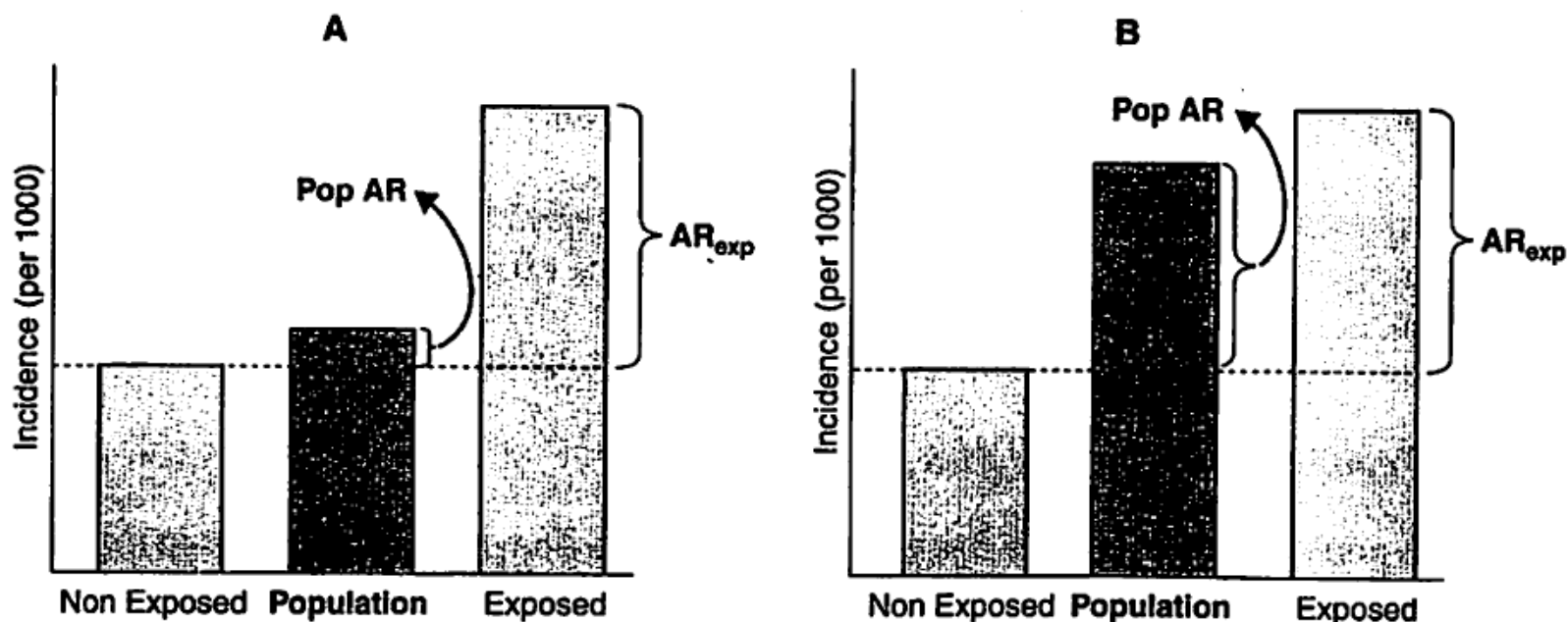


Figure 3–2 Population attributable risk and its dependence on the population prevalence of the exposure. As the population is composed of exposed and unexposed individuals, the incidence in the population is similar to the incidence in the unexposed when the exposure is rare (A) and is closer to that in the exposed when the exposure is common (B). Thus, for a fixed relative risk (eg, $RR \approx 2$ in the figure) the population attributable risk is heavily dependent on the prevalence of exposure.

How PAR% is dependent on prevalence of exposure and RR: example from TB

Table 1 Relative Risk, Prevalence and Population Attributable Risk of Selected Risk Factors for TB, in 22 High TB Burden Countries

Risk Factor (reference for relative risk and prevalence estimates, respectively)	Relative Risk for Active TB Disease (Range)^a	Weighted Prevalence, Total Population, 22 TB High Burden Countries^b	Population Attributable Fraction (Range)^c
HIV infection ^{53,54}	8.3 (6.1–10.8)	1.1%	7.3% (5.2–9.6)
Malnutrition ^{46,55,d}	4.0 (2.0–6.0)	17.2%	34.1% (14.7–46.3)
Diabetes ^{51,56,e}	3.0 (1.5–7.8)	3.4%	6.3% (1.6–18.6)
Alcohol use > 40g/day ^{50,f}	2.9 (1.9–4.6)	7.9%	13.1% (6.7–22.2)
Active smoking ^{48,57,g}	2.6 (1.6–4.3)	18.2%	22.7% (9.9–37.4)
Indoor pollution ^{47,49,h}	1.5 (1.2–3.2)	71.1%	26.2% (12.4–61.0)

But the PAR% estimates add up to >100%!
How is that possible??

How to sum up attributable fractions?

- ❑ This has been a source of controversy
- ❑ In the 1970s, scientists from the NIH proposed that 40% of all cancer is attributable to occupational exposures.
- ❑ Some argued that this was an over-estimate because of x% of cancer is due to smoking, y% is due to diet, z% is due to alcohol, and so on, all of these add up to greater than 100%
- ❑ Rothman pointed out that this rebuttal is fallacious because it assumes that a case of disease has only one cause [Rothman KJ, 2002]
- ❑ The causal pie model shows that each case of cancer could be attributed repeatedly to many separate component causes
- ❑ So, the sum of disease attributable to various component causes has no upper limit (i.e. they can exceed 100%)

Sum of attributable fractions: example

Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study

Salim Yusuf, Steven Hawken, Stephanie Öunpuu, Tony Dans, Alvaro Avezum, Fernando Lanas, Matthew McQueen, Andrzej Budaj, Prem Pais, John Varigos, Liu Lisheng, on behalf of the INTERHEART Study Investigators*

Summary

Background Although more than 80% of the global burden of cardiovascular disease occurs in low-income and middle-income countries, knowledge of the importance of risk factors is largely derived from developed countries. Therefore, the effect of such factors on risk of coronary heart disease in most regions of the world is unknown.

Methods We established a standardised case-control study of acute myocardial infarction in 52 countries, representing every inhabited continent. 15 152 cases and 14 820 controls were enrolled. The relation of smoking, history of hypertension or diabetes, waist/hip ratio, dietary patterns, physical activity, consumption of alcohol, blood apolipoproteins (Apo), and psychosocial factors to myocardial infarction are reported here. Odds ratios and their 99% CIs for the association of risk factors to myocardial infarction and their population attributable risks (PAR) were calculated.

Findings Smoking (odds ratio 2.87 for current *vs* never, PAR 35.7% for current and former *vs* never), raised ApoB/ApoA1 ratio (3.25 for top *vs* lowest quintile, PAR 49.2% for top four quintiles *vs* lowest quintile), history of hypertension (1.91, PAR 17.9%), diabetes (2.37, PAR 9.9%), abdominal obesity (1.12 for top *vs* lowest tertile and 1.62 for middle *vs* lowest tertile, PAR 20.1% for top two tertiles *vs* lowest tertile), psychosocial factors (2.67, PAR 32.5%), daily consumption of fruits and vegetables (0.70, PAR 13.7% for lack of daily consumption), regular alcohol consumption (0.91, PAR 6.7%), and regular physical activity (0.86, PAR 12.2%), were all significantly related to acute myocardial infarction ($p < 0.0001$ for all risk factors and $p = 0.03$ for alcohol). These associations were noted in men and women, old and young, and in all regions of the world. Collectively, these nine risk factors accounted for 90% of the PAR in men and 94% in women.



Lancet 2004; 364: 937–52

Published online
September 3, 2004
<http://image.thelancet.com/extras/04art8001web.pdf>
See [Comment](#) page 912

*Listed at end of report.

Population Health Research
Institute, Hamilton General
Hospital, 237 Barton Street
East, Hamilton, Ontario,
Canada L8L 2X2
(Prof S Yusuf DPhil,
S Öunpuu PhD, S Hawken MSc,
T Dans MD, A Avezum MD,
F Lanas MD, M McQueen FRCP,
A Budaj MD, P Pais MD,
J Varigos BSc, L Lisheng MD)

Correspondence to:
Prof Salim Yusuf
yusufs@mcmaster.ca

This paper,
apparently,
lead to “loss of
faith”
in new CVD risk
factors!

New cardiovascular risk determinants do exist and are clinically useful

Yvo M. Smulders^{1*}, Abel Thijs¹, and Jos W. Twisk²

¹Department of Internal Medicine, VU University Medical Center, PO Box 7057, Amsterdam 1007MB, The Netherlands; and ²Department of Clinical Epidemiology and Biostatistics, VU University Medical Center, PO Box 7057, Amsterdam 1007MB, The Netherlands

Received 13 July 2007; revised 5 November 2007; accepted 12 November 2007

See page 441 for the editorial comment on this article (doi:10.1093/eurheartj/ehm644)

Can we improve our understanding of cardiovascular disease (CVD) causality and prediction? Intuitively, we can. Recent publications, however, could be misinterpreted as suggesting the opposite. First, the Interheart study, which concluded that nine conventional risk factors explain >90% of premature myocardial infarction, is at risk for being interpreted as saying that other, 'new' cardiovascular risk factors can only cause a small remaining fraction of disease of at most 10%. Secondly, papers addressing the predictive value of new risk factors or markers of early CVD risk have concluded that risk prediction does not improve by adding these variables to risk models. In this paper, we will explain that searching for 'new causes' of CVD is still highly relevant, and that improvement of risk prediction is often assessed using inappropriate statistical methodology.

Keywords

Cardiovascular disease • Risk factors • Risk estimation • Epidemiology

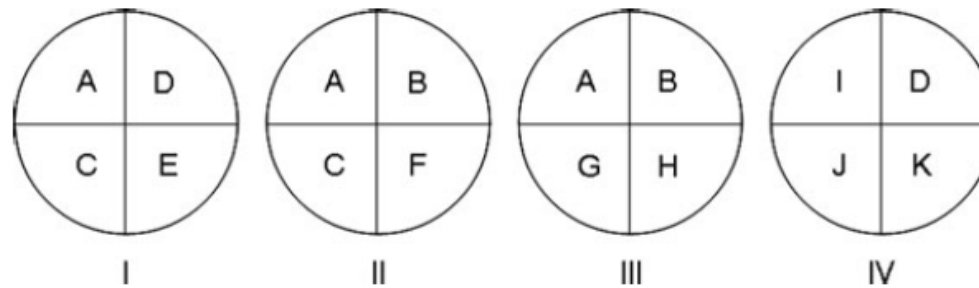


Figure I Why attributable fractions of disease causes add up to $>100\%$

Consider a complex disease with component causes A to K, and four possible constellations of these component causes forming sufficient causes for disease, each responsible for 25% of disease cases after mutual adjustment. Elimination of component cause A would render constellations I–III insufficient and could thus prevent 75% of disease cases. Likewise, elimination of component cause B–D could each prevent 50% of disease, and elimination of component causes E–K would each prevent 25% of disease cases. In this theoretical model of 10 component causes and only four sufficient causal constellations, the sum of the fractions of disease occurrence attributable to each of the component causes adds up to 400%.⁷ In reality, the situation is more complex, as the distribution of attributable fractions of component causes within each of the causal constellations is variable, depending on the sequence of inclusion into risk models.⁸ However, the principle that attributable fractions can add up to $>100\%$ remains valid just the same

Measures of impact when exposure is protective (e.g. RCT)

- ARR (absolute risk reduction): $I_o - I_e$
 - absolute arithmetic difference (reduction) in rates of bad outcomes between experimental and control participants in a trial
- Relative risk reduction (RRR): $(I_o - I_e) / I_o$
 - proportional reduction in rates of bad outcomes between experimental and control participants in a trial, expressed as a fraction of the incidence in the control group
- $NNT = 1 / AR$
 - number of patients who need to be treated over a specific period of time in order to prevent one additional bad outcome

Example

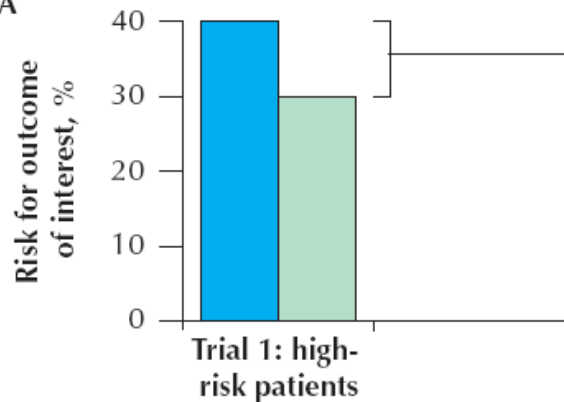
Diabetes Control and Complications Trial (DCCT):

- Effect of intensive diabetes therapy on the development and progression of neuropathy
- neuropathy occurred in 9.6% of patients randomized to usual care and 2.8% of patients randomized to intensive therapy.

Occurrence of endpoint		ARR	RRR	NNT
Io (control)	Ie (intervention)	Io - Ie	$(Io - Ie) / Io$	$1/ARR$
9.6%	2.8%	$9.6\% - 2.8\% = 6.8\%$ (CI: 4.3% to 9.3%)	$\frac{9.6\% - 2.8\%}{9.6\%} = 71\%$	$1/6.8\% = 15$ pts, (CI: 11 to 23)

Which measure to report in a trial?

A



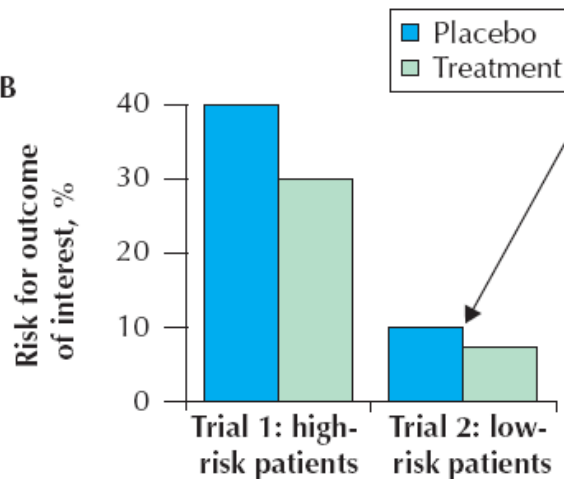
Among high-risk patients in trial 1, the event rate in the control group (placebo) is 40 per 100 patients, and the event rate in the treatment group is 30 per 100 patients.

Absolute risk reduction (also called the risk difference) is the simple difference in the event rates ($40\% - 30\% = 10\%$).

Relative risk reduction is the difference between the event rates in relative terms. Here, the event rate in the treatment group is 25% less than the event rate in the control group (i.e., the 10% absolute difference expressed as a proportion of the control rate is $10/40$ or 25% less).

$$\text{NNT in trial 1} = 10$$

B



Among low-risk patients in trial 2, the event rate in the control group (placebo) is only 10%. If the treatment is just as effective in these low-risk patients, what event rate can we expect in the treatment group?

The event rate in the treated group would be 25% less than in the control group or 7.5%. Therefore, the absolute risk reduction for the low-risk patients (second pair of columns) is only 2.5%, even though the relative risk reduction is the same as for the high-risk patients (first pair of columns).

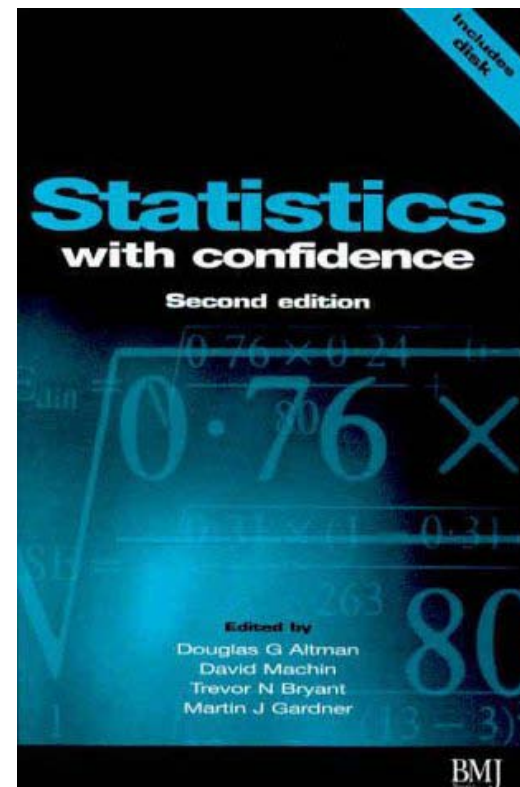
$$\text{NNT in trial 2} = 40$$

Importance of reporting confidence intervals

Always report confidence intervals for all measures of disease frequency, effect and impact!

General format:

95% CI = $\ln(RR) \pm 1.96 SE$
SE = standard error



CI example: case-control study (SIDS study)

	Case	Control	Column total
Prone	27	35	62
Non-prone	20	107	127
Row total	47	142	189

$$OR = (27/20) / (35/107) = 4.13 \text{ [95\% CI } 2.07 \text{ to } 8.22]$$

$$\ln(OR) = 1.42$$

$$\begin{aligned} SE \text{ of } \ln(OR) &= \sqrt{(1/a + 1/b + 1/c + 1/d)} \\ &= \sqrt{(1/27) + (1/35) + (1/20) + (1/107)} \\ &= \sqrt{0.037 + 0.028 + 0.05 + 0.009} \\ &= \sqrt{0.124} = 0.35 \end{aligned}$$

$$\begin{aligned} 95\% \text{ CI for } \ln(OR) &= \ln(OR) \pm 1.96SE \\ &= 1.42 \pm 1.96 (0.35) \\ &= 1.42 \pm 0.69 \\ &= 0.73 \text{ to } 2.1 \end{aligned}$$

$$\begin{aligned} 95\% \text{ CI for } OR &= e^{0.73} \text{ to } e^{2.1} \\ &= 2.07 \text{ to } 8.2 \end{aligned}$$

Confidence intervals of risk ratio measures: Null value is 1

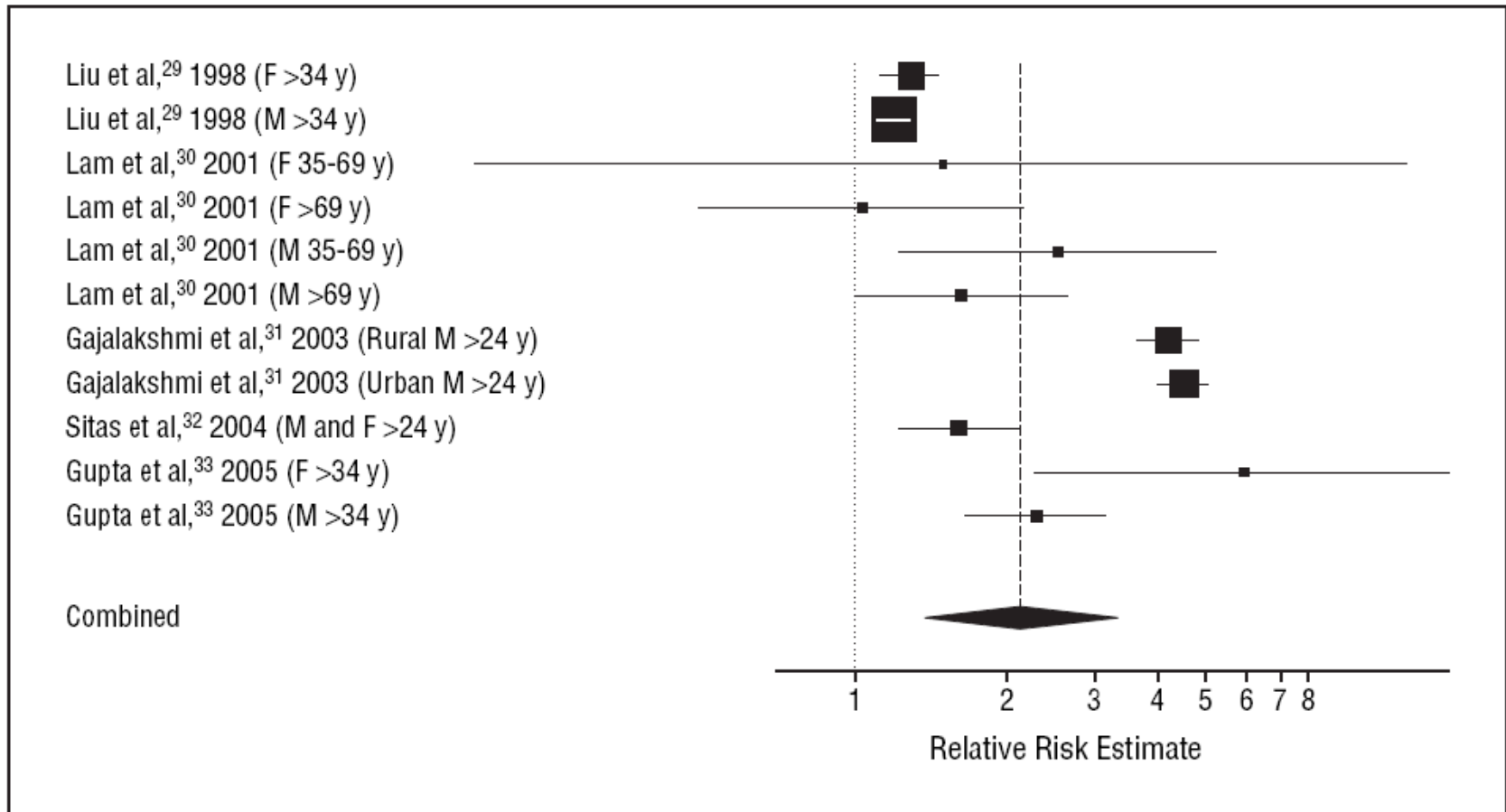
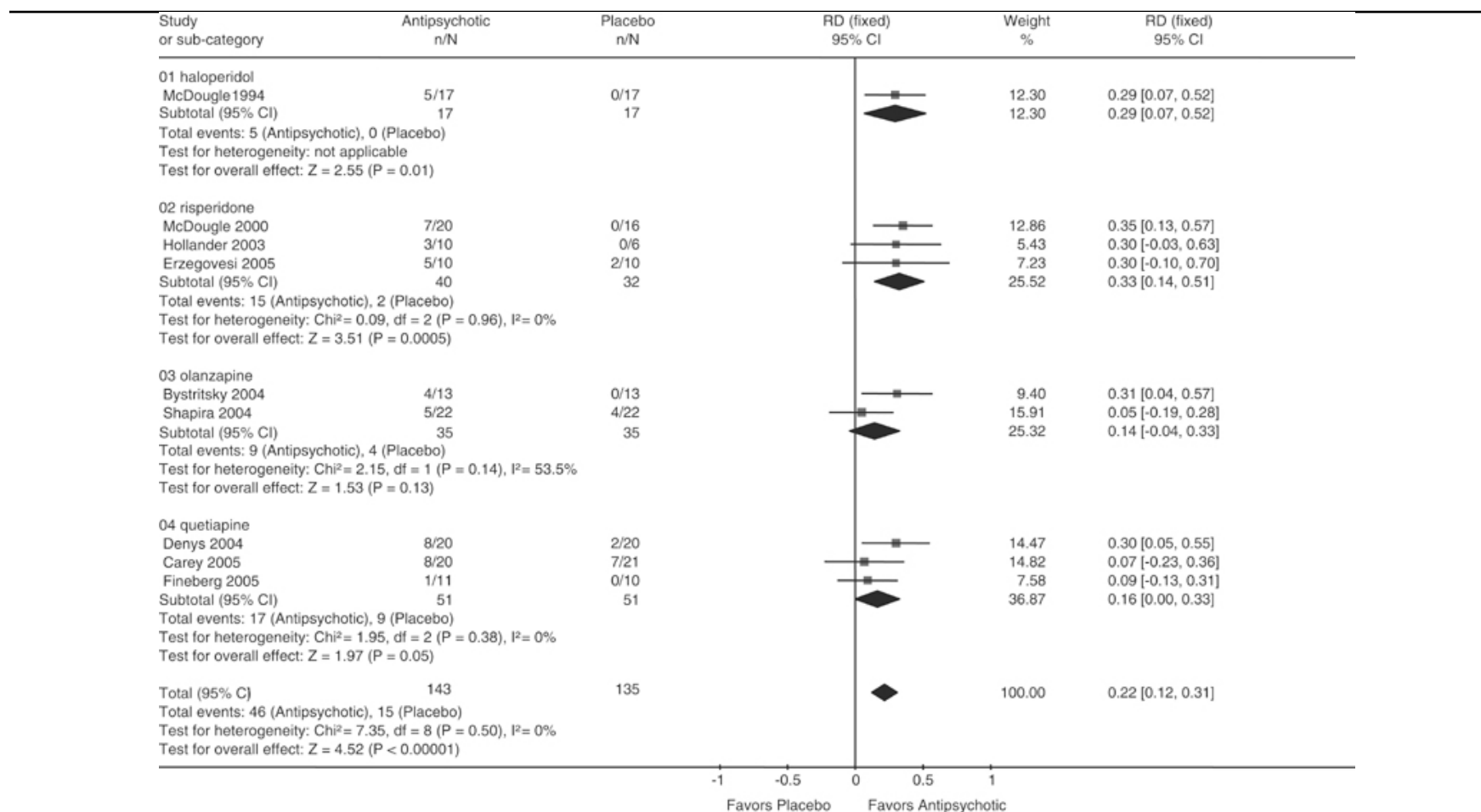


Figure 5. Forest plot of studies²⁹⁻³³ that examined smoking and tuberculosis mortality. The sex and age of the study population are shown on the y-axis.

Confidence intervals of risk difference measures: Null value is 0



Readings for this week

- Rothman text:

- Chapter 3: Measuring disease occurrence and causal effects

- Gordis text:

- Chapter 11: Estimating risk: is there an association?
- Chapter 12: More on risk: estimating the potential for prevention



NEWTON'S THREE LAWS OF GRADUATION

Though famous for his seminal work in Mechanics, Isaac Newton's theories on the prediction of a doctoral graduation formulated while still a grad student at Cambridge remain his most important contribution to academia.

FIRST LAW

"A grad student in procrastination tends to stay in procrastination unless an external force is applied to it"

This postulate is known as the "**Law of Inertia**" and was originally discovered experimentally by Galileo four years before Newton was born when he threatened to cut his grad student's funding. This resulted in a quickening of the student's research progress.

Galileo's observations were later perfected by Descartes through the application of "Weekly Meetings."

Before Galileo's time, it was wrongfully thought that grad students would rest only as long as no work was required of them and that in the absence of external forces, they would graduate by themselves.

(From *Encyclopaedia Britannica*)

PHD.STANFORD.EDU
JORGE CHAM @THE STANFORD DAILY



NEWTON'S THREE LAWS OF GRADUATION

First published in 1679, Isaac Newton's "*Procrastinare Unnaturalis Principia Mathematica*" is often considered one of the most important single works in the history of science. Its Second Law is the most powerful of the three, allowing mathematical calculation of the duration of a doctoral degree.

SECOND LAW

"The age, a , of a doctoral process is directly proportional to the flexibility, f , given by the advisor and inversely proportional to the student's motivation, m "

Mathematically, this postulate translates to:

$$age_{PhD} = \frac{flexibility}{motivation}$$

$$a = F / m$$

$$\therefore F = m a$$

This Law is a quantitative description of the effect of the forces experienced by a grad student. A highly motivated student may still remain in grad school given enough flexibility. As motivation goes to zero, the duration of the PhD goes to infinity.

PHD.STANFORD.EDU
JORGE CHAM @THE STANFORD DAILY



NEWTON'S THREE LAWS OF GRADUATION

Having postulated the first two Laws of Graduation, Isaac Newton the grad student was still perplexed by this paradox: If indeed the first two Laws accounted for the forces which delayed graduation, why doesn't explicit awareness of these forces allow a grad student to graduate?

It is believed that Newton practically abandoned his graduate research in Celestial Mechanics to pursue this paradox and develop his Third Law.

THIRD LAW

"For every action towards graduation there is an equal and opposite distraction"

This Law states that, regardless of the nature of the interaction with the advisor, every force for productivity acting on a grad student is accompanied by an equal and opposing useless activity such that the net advancement in thesis progress is zero.

Newton's Laws of Graduation were ultimately shown to be an approximation of the more complete description of Graduation Mechanics given by Einstein's Special Theory of Research Inactivity.

Einstein's theory, developed during his graduate work in Zurich, explains the general phenomena that, relative to the grad student, time slows down to nearly a standstill.

PHD.STANFORD.EDU
JORGE CHAM @THE STANFORD DAILY 4