Epidemiologia III

Fabrizio Stracci

Confounding

Grimes DA, Schulz KF. Bias and causal associations in observational research. Lancet. 2002; 359:248-52.

- A researcher **attempts to relate an exposure to an outcome**,
- **but** actually **measures** the effect of a third factor, termed **a confounding variable**.
- A confounding variable is associated with the exposure and it affects the outcome,
- but it is not an intermediate link in the chain of causation between exposure and outcome

Variabile confondente

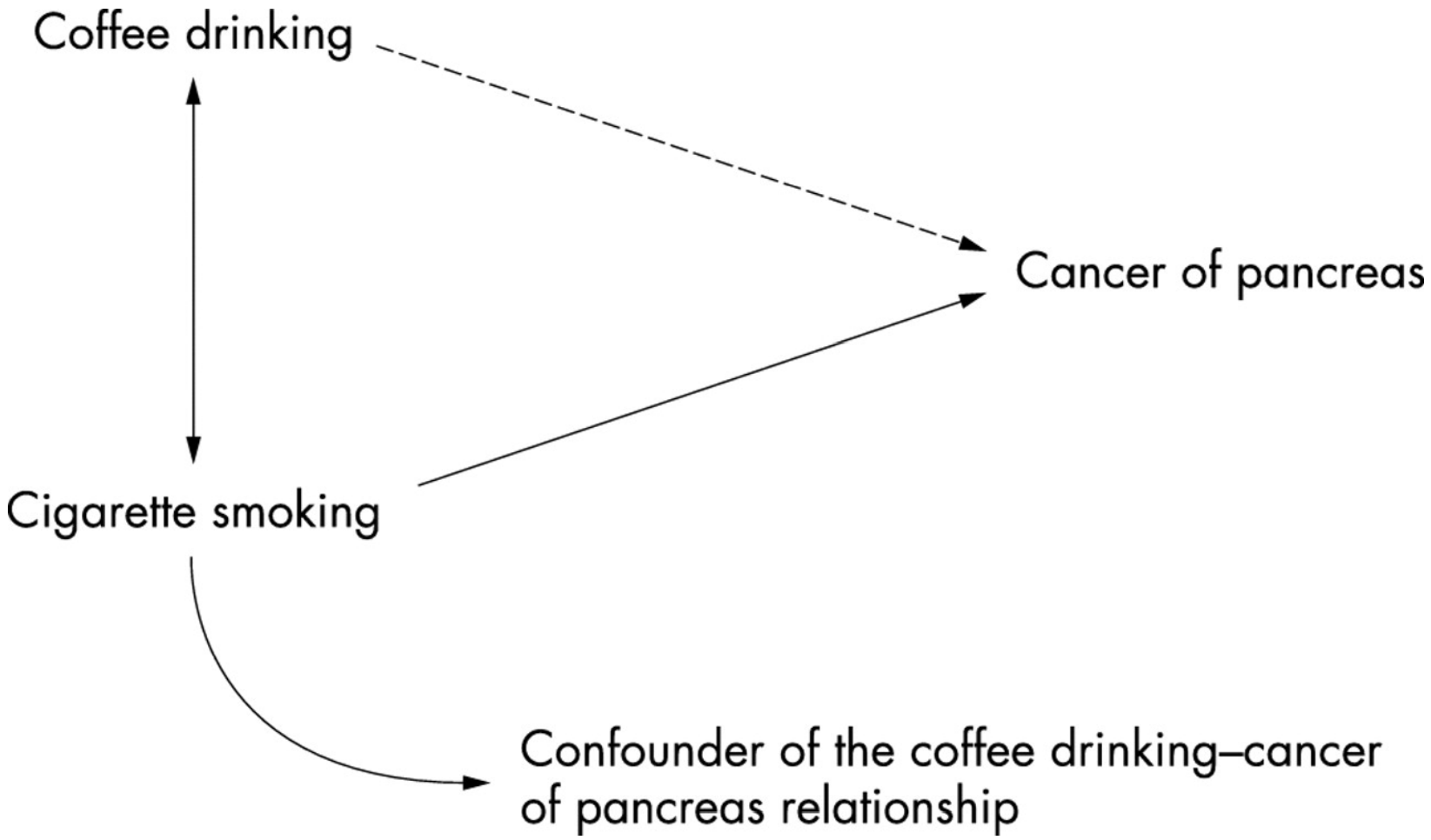
A confounding variable **is**

- **associated with** the **exposure** and
- it affects the **outcome**,

but it **is not**

- an **intermediate** link in the chain of causation between exposure and outcome

Confondimento



Confondimento

- In uno studio intendiamo stimare l'influenza di un **intervento x_1** (esposizione, trattamento) su un parametro μ relativo alla distribuzione delle risposte *y* di una popolazione A, rispetto alla applicazione di **x_0.**
- **x_1** è definito anche trattamento indice e
- **x_0** trattamento di controllo o di riferimento

Confondimento

- Nella popolazione A (popolazione indice o bersaglio), il parametro μ sarà eguale a:
- μ_{A1} In caso di applicazione di x_1
- μ_{A0} In caso di applicazione di x_0
- L'effetto causale può essere misurato, ad esempio, come $\mu_{A1} - \mu_{A0}$
- che tuttavia non è osservabile

Confondimento

- Siamo costretti ad assumere che μ_{A0} sia eguale a μ_{B0} , il valore assunto dal parametro in seguito alla applicazione dell'intervento x_0 ad una **popolazione B**, detta anche **di riferimento o di controllo**
- Il **confondimento è una violazione dell'assunto di equivalenza delle popolazioni A e B rispetto ad x_0**

Confondimento

- E’ presente confondimento se $\mu_{A0} \neq \mu_{B0}$
- In tal caso è presente una qualche differenza tra le popolazioni A e B che è diversa dall’intervento e che è causa della discrepanza tra i parametri
- Ne consegue che la stima del parametro $\mu_{A1} - \mu_{B0}$ che misura l’effetto dell’intervento tra popolazioni è confusa (sistematicamente diversa) dalla misura causale $\mu_{A1} - \mu_{A0}$

Sperimentale randomizzato > Osservazionale

…In the case of a large high quality randomised trial, this means that the results can be assumed to provide an unbiased estimate of the treatment effect in the population studied.

This is not so in observational epidemiology. A well conducted case control or cohort study might still produce misleading results if, for example, important confounders were not known, not measured, or imprecisely measured.

[von Elm E. The scandal of poor epidemiological research. BMJ 2004]

Confondimento

- It is an unfortunate but true fact that many important causal questions are simply not answerable,
- at least not without employing assumptions that are untestable given ethical considerations or limitations of current knowledge and technology.

Examples include

- assumptions of no confounding (the focus of this paper),
- assumptions about independence of unit-specific susceptibilities or responses, and
- various distributional assumptions

Greenland S and Morgenstern H. CONFOUNDING IN HEALTH RESEARCH Annu. Rev. Public Health 2001. 22:189–212

Example of confounding in a <u>hypothetical</u> cohort study of intrauterine device use and salpingitis *[Grimes D, Lancet 2003]*

All women (n=2000)

Use of IUD	Salpingitis: Yes	Salpingitis: No	Total	Proportion with salpingitis
Yes	45	955	1000	4·5%
No	15	985	1000	1·5%

$$\text{Crude RR} = \frac{4{\cdot}5\%}{1{\cdot}5\%} = 3{\cdot}0 \text{ (95\% CI } 1{\cdot}7\text{–}5{\cdot}4\text{)}$$

Women with 1 sexual partner (n=1200)

Use of IUD	Salpingitis: Yes	Salpingitis: No	Total	Proportion with salpingitis
Yes	3	297	300	1·0%
No	9	891	900	1·0%

$$RR=\frac{1\cdot0\%}{1\cdot0\%}=1\cdot0$$

Women with >1 sexual partner (n=800)

Use of IUD	Salpingitis: Yes	Salpingitis: No	Total	Proportion with salpingitis
Yes	42	658	700	6·0%
No	6	94	100	6·0%

$$RR=\frac{6\cdot0\%}{6\cdot0\%}=1\cdot0$$

When the crude relative risk is controlled for the confounding effect of number of sexual partners, the raised risk disappears.

Type of intrauterine device and the risk of pelvic inflammatory disease.

- Lee NC, Rubin GL, Ory HW, Burkman RT. Obstet Gynecol. 1983 Jul;62(1):1-6.

- Abstract
- To study the association of pelvic inflammatory disease and various types of intrauterine devices (IUDs), data from the **Women's Health Study** were analyzed. The analysis included data from interviews of 622 women hospitalized with an initial episode of pelvic inflammatory disease and 2369 hospitalized control subjects reporting no history of pelvic inflammatory disease. Compared to the risk in women using no contraception, the relative risk of pelvic inflammatory disease in women currently using the Dalkon Shield was **8.3 (95% confidence limits 4.7 to 14.5)**. This represented a fivefold increase in risk compared to women currently using other types of IUDs. In this study, only 10% of women wearing an IUD were using the Dalkon Shield, yet they accounted for almost 20% of the excess risk of pelvic inflammatory disease occurring among all the IUD users. Most of the increased risk of pelvic inflammatory disease for women currently using other IUDs (excluding the Dalkon Shield) occurred in the first four months after insertion. These associations were not explained by differences between cases and controls in demographic variables, level of sexual activity, or medical history. The authors recommend that women still using a Dalkon Shield have it removed.

Stratificazione: The intrauterine device and pelvic inflammatory disease revisited: new results from the Women's Health Study

Lee NC, Rubin GL, Borucki R. *Obstet Gynecol 1988;* ***72: 1–6.***

Abstract

To examine whether the risk of pelvic inflammatory disease associated with intrauterine device (IUD) use varies with a woman's sexual behavior, we analyzed data from the Women's Health Study, a hospital-based, case-control study carried out in the United States from 1976-1978. The cases were 657 women hospitalized with pelvic inflammatory disease; controls were 2566 women hospitalized with non gynecologic conditions. **After controlling for confounding factors, we found no consistent differences in the risk of pelvic inflammatory disease associated with IUD use among women in different categories of gonorrhea history, frequency of intercourse, or number of recent sexual partners**. However, **among women with only one sexual partner, married and cohabiting women had little appreciable increased pelvic inflammatory disease risk associated with IUD use** compared with those using no contraception, whereas previously and never-married women using IUDs had relative risk estimates of 1.8 and 2.6, respectively. These results suggest that women at low risk of acquiring sexually transmitted infections have little increase in the risk of pelvic inflammatory disease from use of an IUD.

Rianalisi da parte di altri ricercatori

Abstract

The Women's Health Study (WHS) was a large, widely accepted and influential **case-control study** of the relationship between the use of intrauterine contraceptive devices (IUDs) and pelvic inflammatory disease (PID). The data were collected at 16 hospitals in 9 cities across the U.S.A. from October 1976 through August 1978. **The first paper** on this research was published in 1981 and **concluded that IUDs increase the risk of PID**. The report cited an estimated **RR (relative risk) of PID for current IUD users vs nonIUD users of 1.6 with a 95% confidence interval of (1.4, 1.9).** However, careful examination of the report reveals that the data support conclusions antithetical to those at which the author arrived. When the second report on the WHS was published in 1983, it was anticipated that many of the shortcomings of the first report would be corrected, but they were not. In 1983 we undertook a complete reanalysis of the same WHS data using more appropriate criteria and the results were compared to the first two published reports. **The reanalysis revealed an RR of 1.02 (0.86, 1.21) for current IUD users compared to noncontraceptors.** The conclusion of the WHS should have been that IUDs do not increase the risk of PID.

Kronmal RA, Whitney CW, Mumford SD. The intrauterine device and pelvic inflammatory disease: the Women's Health Study reanalyzed. J Clin Epidemiol. 1991;44:109-22.

Reconsidering the IUD.

- Critical comment is provided on the risk of pelvic inflammatory disease (PID) with the use of IUDs, and specifically the Dalkon Shield. The reopening of the issue of IUD and PID was precipitated by the reanalysis of the Women's Health Study (WHS) data by Kronmal et al. The focus of the discussion is on the description of an epidemiological case control study, the WHS, assessing the criticisms of the study, the possible biases, and an interpretation of the conflicting analyses. The WHS was conducted in 8 health care centers in the US between 1976-78 and included hospitalized PID cases. Controls were other hospitalized patients. Burkman in 1981 and Lee et al, in 1983 published results of their analyses of the WHS data. Burkman estimated a 1.6 risk of PID among IUd users relative to all other contraceptive users and 1 or 2.1 relative to women with no history of PID. Lee estimated risk among Dalkon Shield users as 8.3 relative to women not using any contraceptive method prior to hospitalization, and 1.6 among users of other IUDs; analysis excluded women with a prior history of PID. Recently Kronmal et al, found fault with the WHS on 10 counts, which are provided in detail. Their reanalysis indicates there is no increased risk of PID for IUd users. **However, the question is raised as to the appropriate reference group, -all other users of contraceptives or nonusers. Another question pertains to whether subjects should all, or just those without a prior PID history, be included**. Due to the protective effect of oral contraceptives, the reference group should be nonusers of any type as analyzed by Lee et al. Lee et al. also excludes those with a prior PID history, which is the best solution to the 2nd question. The Kronmal analysis allowed those with a prior PID history, and the reference group was all women regardless of use. The Kronmal analysis also describes ways in which selection bias is involved in the higher risk of PID among Dalkon Shield users relative to nonusers. Some of the reasons are that women using the Dalkon Shield may have been less likely to seek early medical care, to have a greater likelihood of having their diseases properly diagnosed, and to identify more easily the Dalkon Shield as their IUD type. This reasoning is found to be excessive and unrealistic to explain a difference between a risk of 8.3 vs. 1.6, and the Dalkon Shield was deservedly removed as an unsafe product. Support for the IUD is better served by discussing the careful selection of users to reduce PID, meticulous insertion techniques, and close monitoring.

- Fam Plann Perspect. 1992 Jan-Feb;24(1):33-5.
- Petitti DB.

The Checkered History and Bright Future of Intrauterine Contraception in the United States

Early research examining PID

- used inappropriate comparison groups,
- overdiagnosed PID in IUD users and
- Did not control for the confounding effects of sexual behavior

IUD e PID : revisione

Gareen IF, Greenland S, Morgenstern H. Intrauterine devices and pelvic inflammatory disease: meta-analyses of published studies, 1974-1990. *Epidemiology*. 2000;11:589-97.

We observed consistent, positive associations of IUD use with both symptomatic and asymptomatic PID. These associations were largest for the Dalkon Shield.

TABLE 3. Summary (Average) Rate Ratio Estimates with Random-Effects, 95% Confidence Limits and Data Descriptions for Intrauterine Device (IUD) and Dalkon Shield (DS) Contrasts, by Type of Cases of Pelvic Inflammatory Disease (Symptomatic *vs* Asymptomatic) (IUD Contrasts Only) and Type of Reference Group

Type of Cases/Reference Group*	No. of Studies	Rate Ratio†	95% CL	*P*-Value‡	Percentile			
					Minimum	25%	75%	Maximum
Intrauterine Device Comparisons								
All PID cases								
All studies	32	4.1	2.9, 5.8	<0.001	1.1	2.2	8.8	132
No contraception	11	2.4	1.8, 3.3	0.007	1.5	2.2	6.5	7.9
Any non-IUD contraception	29	4.3	2.9, 6.3	<0.001	1.1	2.1	9.0	132
Studies with symptomatic cases								
All studies	22	3.0	2.4, 3.7	<0.001	1.1	2.1	4.4	9.3
No contraception	8	2.5	1.7, 3.7	<0.001	1.5	1.9	3.7	7.3
Any non-IUD contraception	18	3.0	2.4, 3.8	<0.001	1.1	2.1	4.9	9.3
Studies with asymptomatic cases								
All studies	10	9.2	3.6, 24	<0.001	1.5	6.7	12	132
No contraception	3	2.9	0.81, 10	0.90	2.6	2.6	7.9	7.9
Any non-IUD contraception	10	9.2	3.6, 24	<0.001	1.5	6.7	12	132
Dalkon Shield Comparisons								
All PID cases								
All studies	12	3.2	1.4, 7.1	<0.001	0.32	1.5	7.8	25
No contraception	2	13	6.7, 26	0.87	8.3	8.3	11	13
Any non-IUD contraception	4	13	4.0, 42	0.12	4.7	7.8	25	25
Other IUDs	12	2.2	1.5, 3.4	0.026	0.32	1.5	3.5	5.0

IUD in STI may increase PID

Mohllajee AP, Curtis KM, Peterson HB. Does insertion and use of an intrauterine device increase the risk of pelvic inflammatory disease among women with sexually transmitted infection? A systematic review. Contraception. 2006;73:145-53.

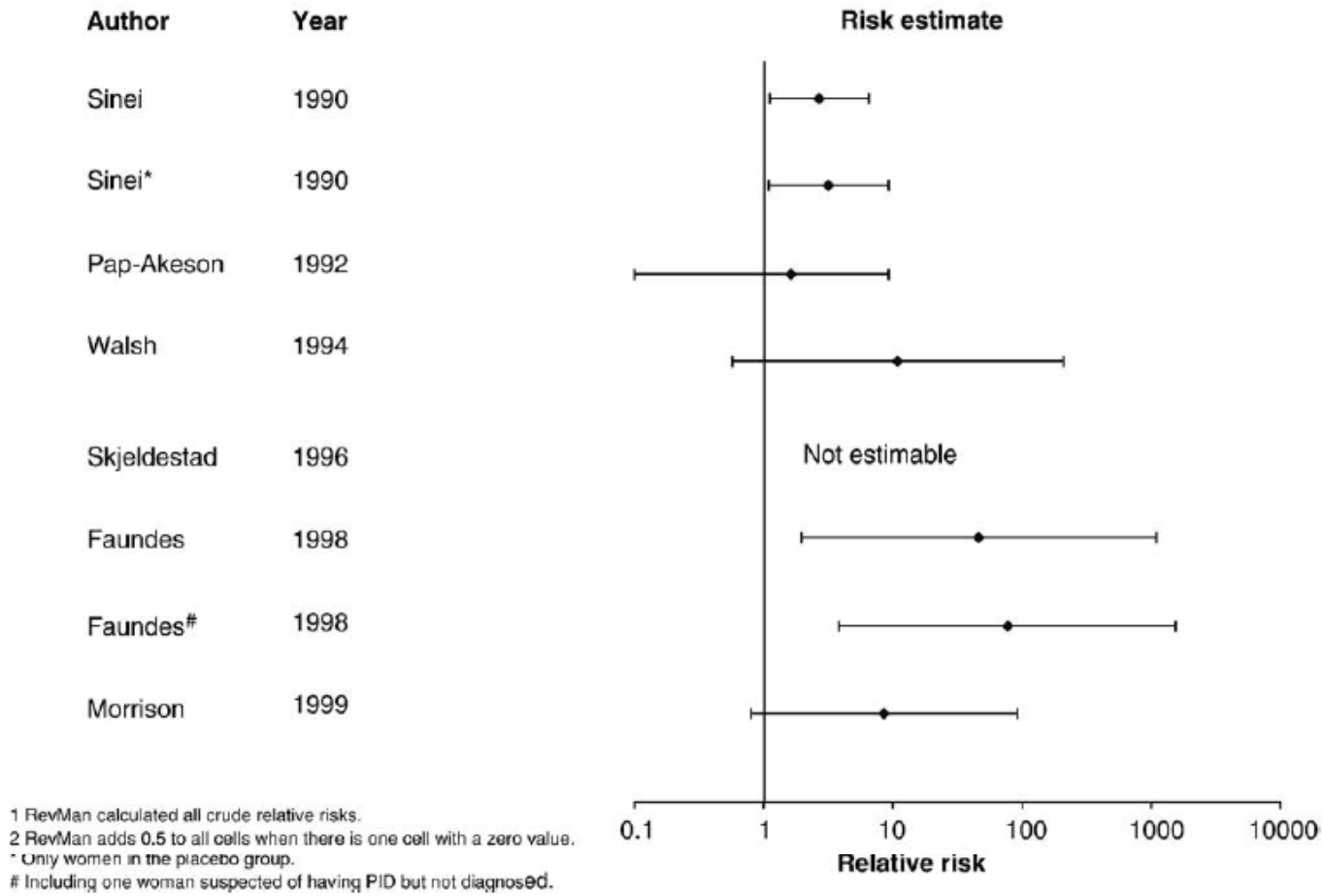


Fig. 1. Crude relative risk of PID among women with and without STI at insertion [1,2].

Controllo della distorsione

- Mediante disegno dello studio
  - Restrizione (criteri di inclusione)
  - Appaiamento (matching , difficile per numerose variabili)
  - Assegnazione casuale (equivalenza probabilistica tra A e B)
- Mediante ‘aggiustamento’, correzione, in fase di analisi dei dati
  - Stratificazione
  - Propensione al trattamento
  - Modello di regressione

Restrizione

- also called exclusion or specification
- Although this tactic avoids confounding, it also hinders recruitment (and thus power) [Nonetheless, restriction on many factors can reduce the number of available subjects to unacceptably low levels]
- and precludes extrapolation to the excluded group.
- Restriction might increase the internal validity of a study at the cost of poorer external validity
- *Ad esempio nel caso del cancro del pancreas, si sarebbero potuti escludere I fumatori e condurre lo studio sui soli bevitori di caffè non fumatori*

Appaiamento

This approach, although often used by investigators, has two drawbacks:

- If matching is done on several potential confounding factors, the recruitment process can be cumbersome, and, by definition,
- one cannot examine the effect of a matched variable
- differential losses to observation may undo the initial covariate balances produced by matching
- *In a case-control study in which smoking is deemed a confounding factor, cases and controls can be matched by smoking status. For each case who smokes, a control who smokes is found*

Assegnazione casuale

- E’ l’unico mezzo per trattare confondenti non considerati esplicitamente in fase di disegno (ad esempio fattori non noti). Tuttavia
- La soluzione è solo probabilistica e soggetta a restrizioni
- Ad esempio violazioni del protocollo e perdite al follow-up possono produrre una distribuzione delle covariate sbilanciata tra i gruppi (i trial di dimensioni ridotte sono più suscettibili a questo problema)

Stratificazione

- Stratification can be considered a form of post hoc restriction, done during the analysis rather than during the accrual phase of a study.
- It would seem natural, then, to control confounding due to measured factors by simply stratifying on them all. Unfortunately, one would then confront the well-known sparse-data problem: Given enough factors, few if any strata would have subjects in both treatment groups, thereby making comparisons biased, inefficient, or impossible.

Propensione al trattamento

- Subject to any modeling restrictions used for score estimation, balance in probability for a set of covariates could be achieved by exact stratification on the estimated **propensity score**,
- Where the propensity score is defined as the probability of treatment given the *(observed)* covariates in the combined (treated and untreated) study population

Propensity score

- The propensity score is the conditional probability of receiving the treatment given the observed covariates
- In an observational study, the true propensity score is unknown and must be estimated from the data.
- There are two important results on propensity scores: first, if there is no hidden bias, that is, if treatment assignment is related only to measured subject characteristics, then treatment assignment, conditional on the propensity score, is random. Thus, within strata of subjects matched on the propensity score, standard multivariate adjustment methods can be used to obtain unbiased estimates of average treatment effects.
- Second, treated and control subjects in strata or matched sets that are homogeneous in the propensity score tend to have the same distribution of covariates that were used in estimating the propensity score.
- Of important note is the fact that propensity score methods only reduce bias due to differences in measured covariates between the two treatment groups. It makes no claim to eliminate bias due to unmeasured covariate differences between the two treatment groups

Stima e utilizzo del punteggio di propensione

1. after having modeled the distribution of the treatment indicator variable given the observed covariates,
2. the ensuing propensity score can be used to reduce selection bias through

- matching,
- stratification (i.e. subclassification),
- regression adjustment, or some combination of all three

Modelli di regressione

- In multivariate techniques, mathematical modelling examines the potential effect of one variable while simultaneously controlling for the effect of many other factors.
- Possono essere controllati i fattori misurati (noti) correttamente e correttamente introdotti nell'analisi
- Il successo del controllo dipende anche dalla correttezza del modello utilizzato

Propensity score vs adjustement via regression

- **OBJECTIVE:** To determine whether adjusting for confounder bias in observational studies using propensity scores gives different results than using traditional regression modeling.
- **METHODS:** Medline and Embase were used to identify studies that described at least one association between an exposure and an outcome using both traditional regression and propensity score methods to control for confounding. From 43 studies, 78 exposure-outcome associations were found. Measures of the quality of propensity score implementation were determined. The statistical significance of each association using both analytical methods was compared. The odds or hazard ratios derived using both methods were compared quantitatively.
- **RESULTS:** Statistical significance differed between regression and propensity score methods for only 8 of the associations (10%), kappa = 0.79 (95% CI = 0.65-0.92). In all cases, the regression method gave a statistically significant association not observed with the propensity score method. The odds or hazard ratio derived using propensity scores was, on average, 6.4% closer to unity than that derived using traditional regression.
- **CONCLUSIONS:** Observational studies had similar results whether using traditional regression or propensity scores to adjust for confounding. Propensity scores gave slightly weaker associations; however, many of the reviewed studies did not implement propensity scores well
- Shah BR, Laupacis A, Hux JE, Austin PC. Propensity score methods gave similar results to traditional regression modeling in observational studies: a systematic review. J Clin Epidemiol. 2005;58:550-9. Epub 2005 Apr 19.

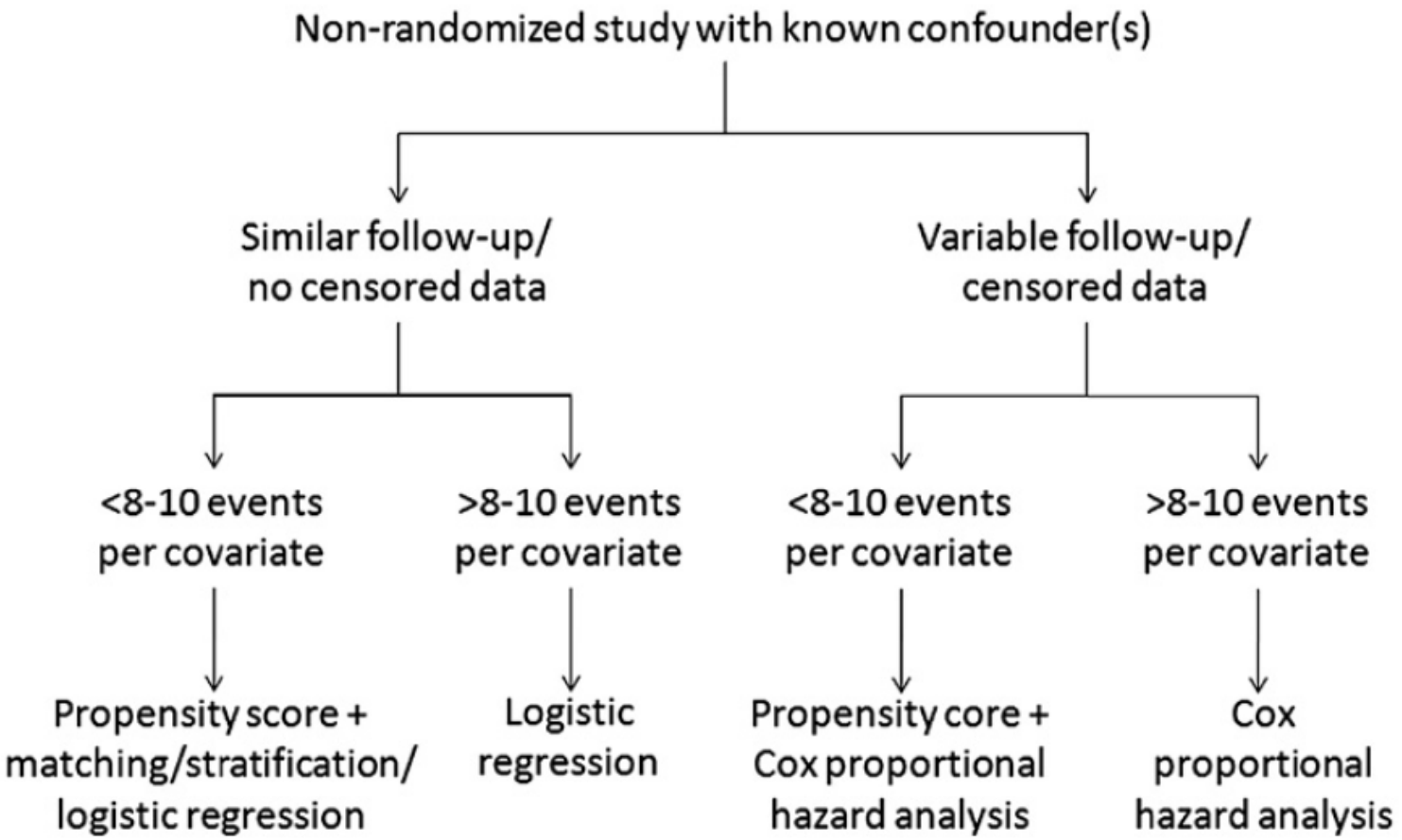
Alcuni risultati del confronto

Associations between exposures and outcomes using both traditional regression and propensity score methods, with application of propensity score and any check for confounder balance

Report	Patients studied	Intervention or exposure	Outcome[b]	Traditional regression			Propensity score methods			Propensity score application	Check for balance[a]
				OR or HR	95% CI	*P*-value	OR or HR	95% CI	*P*-value		
Abramov et al. [3]	Coronary artery bypass graft	Pulsatile vs. nonpulsatile perfusion	Mortality			NS			NS	Modeling	None
			Mortality; MI; stroke; low output			NS			NS		
			Stroke	1.91	1.11–3.30	<.05	2.22	1.13–4.37	.01		
			Renal dysfunction			NS			NS		
Aronow et al. [4]	Acute coronary syndrome	Post-discharge lipid-lowering drugs	Mortality	0.6	0.46–0.78	<.0002	0.67	0.49–0.92	.012	Modeling	Reported
	Exclude propensity deciles 1–3			0.61	0.46–0.81	.0006	0.67	0.48–0.93	.017		
Barosi et al. [5]	Myelofibrosis	Splenectomy	Blast transformation	2.61	1.38–4.95	.003			.008	Stratification	None
Chan et al. [6]	Elective PCI	Beta blockers	Mortality	0.62	0.45–0.87	.0048	0.68	0.50–0.93	.0164	Modeling	Reported
Chan et al. [7]	Elective PCI	Statins	Mortality	0.65	0.42–0.99	.045	0.64	0.42–0.97	.034	Modeling	Reported
Cook & Goldman [8]	ICU admission	Closed vs. open unit	Mortality	0.62		.013	0.75		.047	Stratification	None
Drake & Fisher [9]	Newborns	Maternal smoking	Low birth weight	2.45	1.15–5.21		1.61	0.70–3.71		Stratification	Reported
Earle et al. [10]	Stage IV NSCLC	Chemotherapy	Mortality	0.81	0.76–0.85		Similar[c]			Stratification	Displayed
Elad et al. [11]	Late-presentation acute MI	Initial invasive therapy	Mortality	0.67	0.49–0.92		0.69		.036	Matching	Displayed

Are propensity scores really superior to standard multivariable analysis?

Biondi-Zoccai G, et al. *Contemp Clin Trials.* 2011 ;32:731-40.



Altri caffè: effetti sul feto della assunzione di caffè in gravidanza

- Caffeine is the most commonly used psychoactive substance in the world.
- It is found in a range of beverages and food, mainly in tea, coffee, cola, chocolate bars and some medications.
- There has been a concern that maternal consumption of caffeine in pregnancy may be associated with adverse pregnancy outcomes
- There are more than 80 published observational studies on the effects of caffeine during pregnancy, and the results are conflicting, controversial, and rarely evidence based.
- These studies have focused on whether ingestion of caffeine causes spontaneous abortion, congenital abnormalities, fetal growth restriction, low birth weight, and preterm birth.
- It is significant and worth emphasizing that only one randomized controlled trial (RCT) was available for this Cochrane Review, and that trial showed that there was no difference in birth weight or length of gestation between women who drank 3 or more cups of coffee daily compared to those who did not.
- Jahanfar S, Sharifah H.Effects of Restricted Caffeine Intake by Mother on Fetal, Neonatal, and Pregnancy Outcome. *Obstetrics & Gynecology,* 2009; 114:161-162

Measuring Exposure—How and When, and What Is the Reference Period?

- Women can change their caffeine intake considerably over the course of pregnancy.
- In particular, pregnant women generally decrease their caffeine intake markedly during the first trimester
- In case-control studies, differential timing of interviews between cases and controls may lead to intractable bias
- A substantial amount of measurement error is therefore present in published studies.

Severe limitazioni

- For example, the fundamental bias that arises because of the interrelationship among nausea, caffeine consumption, and fetal viability, which no prior study can claim to have overcome, would create or inflate an association between caffeine and spontaneous abortion.
- Other biases that would have tended to overestimate the effect of caffeine in prior studies are recall bias, selection bias, differential timing of data collection, confounding by smoking…

Control of Confounding

- Many potentially confounding factors have either been matched on or statistically adjusted for in the studies under review.
- These include maternal age,[8–16,27–30] gestational age at enrollment or interview,[12,13,27,28,30] cigarette smoking, [9,10,12,15,27–29] alcohol consumption,[9,10,12,13,15,27–29] pregnancy symptoms,[13,15] parity or gravidity,[9,13,15,29] history of spontaneous abortion,[9,10,13,15,27,29,30] history of therapeutic abortion,10 prior gynecologic surgery,[30] uterine abnormality, [12] race or ethnic group,[9,10,27,30] marital status,[10,27] insurance coverage or payment group,[10,11] education,[9,12,15,29] socioeconomic status or income,[27,29] and employment status or work schedule.[9,12,27]

- Signorello LB, McLaughlin JK. Maternal caffeine consumption and spontaneous abortion: a review of the epidemiologic evidence. *Epidemiology*. 2004;15:229-39.

Fumo come confondente

- There is a distinct possibility of residual confounding by smoking—the effect of which would be to exaggerate any association between caffeine and spontaneous abortion.
- Many studies did not adjust for smoking, because this variable did not meet a significance criterion, because adjustment for smoking was found to have little effect on the relative risk for caffeine or for unstated reasons
- However, smoking *should be a confounder, because smoking* behavior is linked to caffeine intake

Caffeine intake and the risk of first-trimester spontaneous abortion.

- Cnattingius S, Signorello LB, Annerén G, Clausson B, Ekbom A, Ljunger E, Blot WJ, McLaughlin JK, Petersson G, Rane A, Granath F. N Engl J Med. 2000 Dec 21;343(25):1839-45.

- Abstract
- BACKGROUND: Some epidemiologic studies have suggested that the ingestion of caffeine increases the risk of spontaneous abortion, but the results have been inconsistent.
- METHODS: We performed **a population-based, case-control study** of early spontaneous abortion in Uppsala County, Sweden. The subjects were 562 women who had spontaneous abortion at 6 to 12 completed weeks of gestation (the case patients) and **953 women who did not have spontaneous abortion and were matched to the case patients according to the week of gestation (controls**). Information on the ingestion of caffeine was obtained from in-person interviews**. Plasma cotinine was measured as an indicator of cigarette smoking,** and fetal karyotypes were determined from tissue samples. Multivariate analysis was used to estimate the relative risks associated with caffeine ingestion after adjustment for smoking and symptoms of pregnancy such as nausea, vomiting, and tiredness.
- RESULTS: Among nonsmokers, more spontaneous abortions occurred in women who ingested at least 100 mg of caffeine per day than in women who ingested less than 100 mg per day, with the increase in risk related to the amount ingested (100 to 299 mg per day: odds ratio, 1.3; 95 percent confidence interval, 0.9 to 1.8; 300 to 499 mg per day: odds ratio, 1.4; 95 percent confidence interval, 0.9 to 2.0; and 500 mg or more per day: odds ratio, 2.2; 95 percent confidence interval, 1.3 to 3.8). Among smokers, caffeine ingestion was not associated with an excess risk of spontaneous abortion. When the analyses were stratified according to the results of karyotyping, the ingestion of moderate or high levels of caffeine was found to be associated with an excess risk of spontaneous abortion when the fetus had a normal or unknown karyotype but not when the fetal karyotype was abnormal.
- CONCLUSIONS: The ingestion of caffeine may increase the risk of an early spontaneous abortion among non-smoking women carrying fetuses with normal karyotypes.

TABLE 3. ADJUSTED ODDS RATIOS FOR SPONTANEOUS ABORTION ASSOCIATED WITH THE INGESTION OF CAFFEINE DURING PREGNANCY AMONG NONSMOKERS AND SMOKERS, ACCORDING TO FETAL KARYOTYPE.*

VARIABLE	KARYOTYPE		
	ABNORMAL	NORMAL	UNKNOWN
Nonsmokers			
No. of case patients/no. of control subjects	115/811	74/811	212/811
		odds ratio (95% CI)	
Daily intake of caffeine			
0–99 mg†	1.0	1.0	1.0
100–299 mg	1.0 (0.6–1.6)	2.0 (1.0–3.7)	1.5 (0.9–2.3)
300–499 mg	0.9 (0.5–1.7)	1.8 (0.8–3.8)	1.5 (0.9–2.4)
≥500 mg	1.8 (0.8–3.9)	2.2 (0.8–6.4)	2.6 (1.3–5.1)
P value	0.39	0.11	0.001
Smokers			
No. of case patients/no. of control subjects	29/121	20/121	66/121
		odds ratio (95% CI)	
Daily intake of caffeine			
0–99 mg†	1.0	1.0	1.0
100–299 mg	1.4 (0.2–11.1)	0.5 (0.1–4.5)	0.7 (0.2–2.4)
300–499 mg	2.1 (0.3–15.9)	2.6 (0.4–16.7)	1.5 (0.5–4.7)
≥500 mg	0.5 (0.1–4.0)	1.9 (0.3–11.3)	0.5 (0.2–1.7)
P value	0.32	0.22	0.42

Caffeinated beverages, decaffeinated coffee, and spontaneous abortion.

- Fenster L, Hubbard AE, Swan SH, Windham GC, Waller K, Hiatt RA, Benowitz N. Epidemiology. 1997 ;8:515-23.

- Abstract
- We examined the relations between spontaneous abortion and the consumption of caffeine, individual caffeine-containing beverages (coffee, tea, and soda), and decaffeinated coffee **in a prospective study of 5,144 pregnant women**. We collected information about potential risk factors for spontaneous abortion, including consumption of caffeinated beverages and decaffeinated coffee before and during pregnancy, by interview in the first trimester. Neither total estimated caffeine nor individual caffeinated beverage consumption during the first trimester was associated with an appreciable increase in risk for spontaneous abortion. The adjusted odds ratio for consumption of greater than 300 mg per day of caffeine was 1.3 [95% confidence interval (CI) = 0.8-2.1] after adjustment for maternal age, pregnancy history, cigarette and alcohol consumption, employment, race, gestational age at interview, and marital and socioeconomic status. The adjusted odds ratio for spontaneous abortion related to consumption of three or more cups of decaffeinated coffee during the first trimester was 2.4 (95% CI = 1.3-4.7) in the same model. Although we could not demonstrate this with available data, we suspect that this association was due to bias resulting from the relations among fetal viability, symptoms of pregnancy such as nausea, and consumption patterns during pregnancy.

Confondimento derivante dalla indicazione al trattamento in studi osservazionali

- The confounder is the indication, as it is related to the intervention and is a risk indicator for the disease
- The indication for treatment (i.e. the reason why a specific medication was given to a given individual) will be related to choice of treatment and may also be related to the risk of future health outcomes.
- The indication for treatment (which is often a mix of several reasons to start or withhold treatment) may be a very strong prognostic indicator.
- This results in confounding and the resulting imbalance in the underlying risk (prognostic) profile between treated and comparison groups can generate biased results

Effetti paradossali

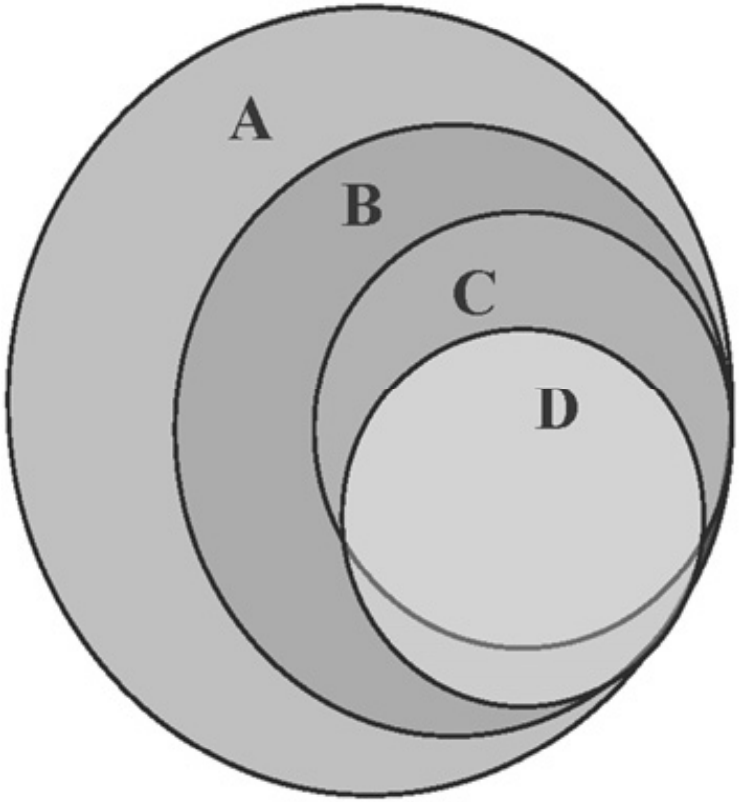
- For example, a drug might look ineffective or even harmful because of poor outcomes among those taking that drug; this, however, could be merely because the drug was given to highly selected individuals who needed the drug because of their poor prognosis.
- In other words, the sickest patients were given the drug and it is not surprising that their outcomes were poor.
- [l'effetto del trattamento in questo caso dovrebbe essere misurato non in termini di beneficio rispetto ai non trattati ma di riduzione della differenza di esito]
- On the other hand, suppose that a drug contains a warning that it must never be given to people with hypertension.
- A crude analysis of the association between the drug and hypertension status in the patient population will appear as though the drug is highly protective against this disease

A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies.

- Bosco JL, et al. *J Clin Epidemiol*. 2010;63:64-74.
- Abstract
- OBJECTIVE: To evaluate the effectiveness of methods that control for confounding by indication, we compared breast cancer recurrence rates among women receiving adjuvant chemotherapy with those who did not.
- STUDY DESIGN AND SETTING: In a medical record review-based study of breast cancer treatment in older women (n=1798) diagnosed between 1990 and 1994, our crude analysis suggested that adjuvant chemotherapy was positively associated with recurrence (hazard ratio [HR]=2.6; 95% confidence interval [CI]=1.9, 3.5). We expected a protective effect, so postulated that the crude association was confounded by indications for chemotherapy. We attempted to adjust for this confounding by restriction, multivariable regression, propensity scores (PSs), and instrumental variable (IV) methods.
- RESULTS: After restricting to women at high risk for recurrence (n=946), chemotherapy was not associated with recurrence (HR=1.1; 95% CI=0.7, 1.6) using multivariable regression. PS adjustment yielded similar results (HR=1.3; 95% CI=0.8, 2.0). The IV-like method yielded a protective estimate (HR=0.9; 95% CI=0.2, 4.3); however, imbalances of measured factors across levels of the IV suggested residual confounding.
- CONCLUSION: Conventional methods do not control for unmeasured factors, which often remain important when addressing confounding by indication. PS and IV analysis methods can be useful under specific situations, but neither method adequately controlled confounding by indication in this study.

Unadjusted analysis

Using Cox proportional hazards regression on the unrestricted cohort, we estimated the hazard ratio (HR) associating receipt of adjuvant chemotherapy vs. not receiving adjuvant chemotherapy.



A Unrestricted* N=1798

B Restricted† N=946

C Propensity Score Trimmed‡ N=723

D Instrumental Variable§ N=539

Restriction and multivariable regression

- Within the unrestricted cohort, we identified a restricted subset of women as at high risk for recurrence
- Using the restricted cohort, we adjusted for
- demographic characteristics (age group, race/ethnicity, health care system, baseline Charlson Comorbidity Index score),
- tumor characteristics (tumor size, node positivity, histologic grade, ER expression, and PR expression),
- and treatment characteristics (primary therapy, tamoxifen prescription)
- to estimate the HR of breast cancer recurrence comparing those who received chemotherapy with those who did not.

Propensity score method

- Using logistic regression with the restricted cohort, we modeled the probability of receiving adjuvant chemotherapy as a function of the variables included in the multivariable adjusted model.
- We used Cox proportional hazards regression to model the association between adjuvant chemotherapy and recurrence,
- using three PS adjustment approaches.
- First we divided the trimmed sample into PS quintiles. We adjusted for PS quintiles and used the lowest quintile as the reference.
- Second, we adjusted for the continuous PS measure in the Cox proportional hazards model.
- Last, we used a doubly robust adjustment, in which we adjusted for the continuous PS and the variables used to predict the probabilities of receiving adjuvant chemotherapy

IV, Instrumental Variable

- Specifically, the use of our IV-like approach was intended to control for the confounding by unmeasured indications for chemotherapy
- we used each patient's surgeon's chronologically preceding patient's receipt of adjuvant chemotherapy (preceding patient within our data set) as the IV within strata of stage and ER expression
- to estimate the effect of receipt of adjuvant chemotherapy on time to breast cancer recurrence.
- it has been argued that the application of this method is limited because of its strong assumptions, making it difficult in practice to find a suitable instrumental variable

Interazione

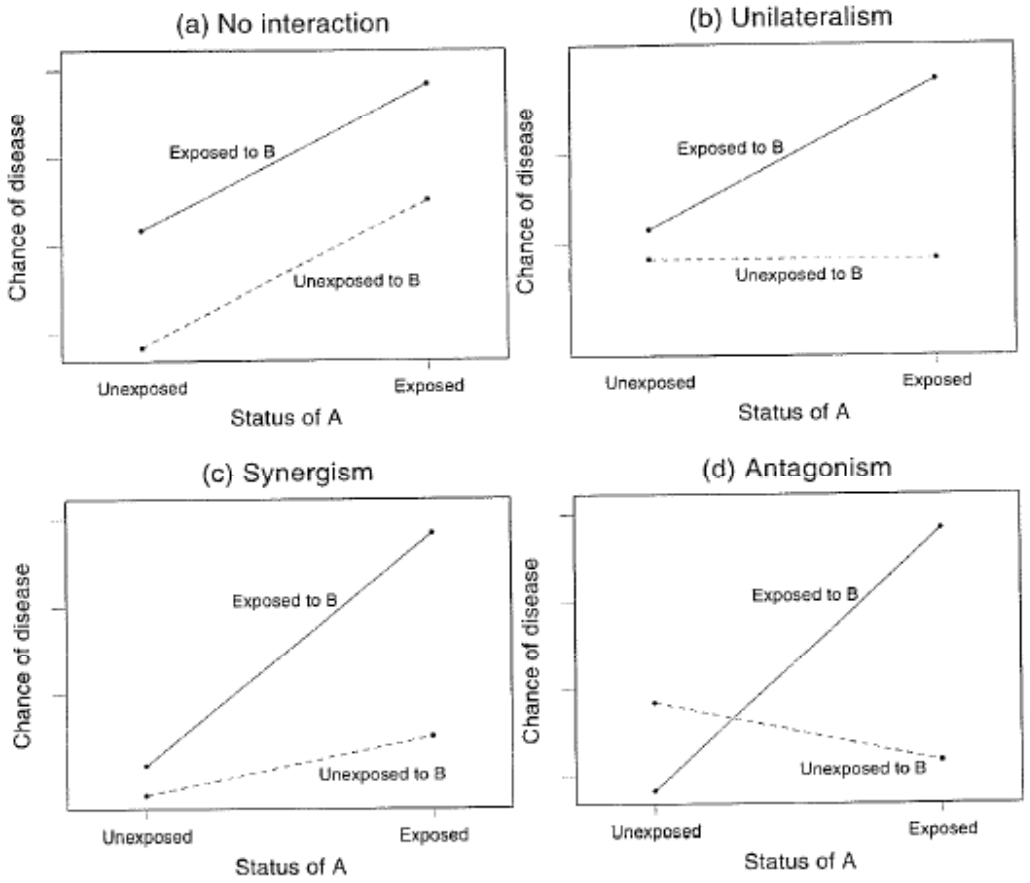


Figure 4.7 Interaction diagrams, illustrating four different situations for risk factors A and B.

Interaction is the situation whereby the association of one risk factor with a certain outcome variable differs across strata of another risk factor