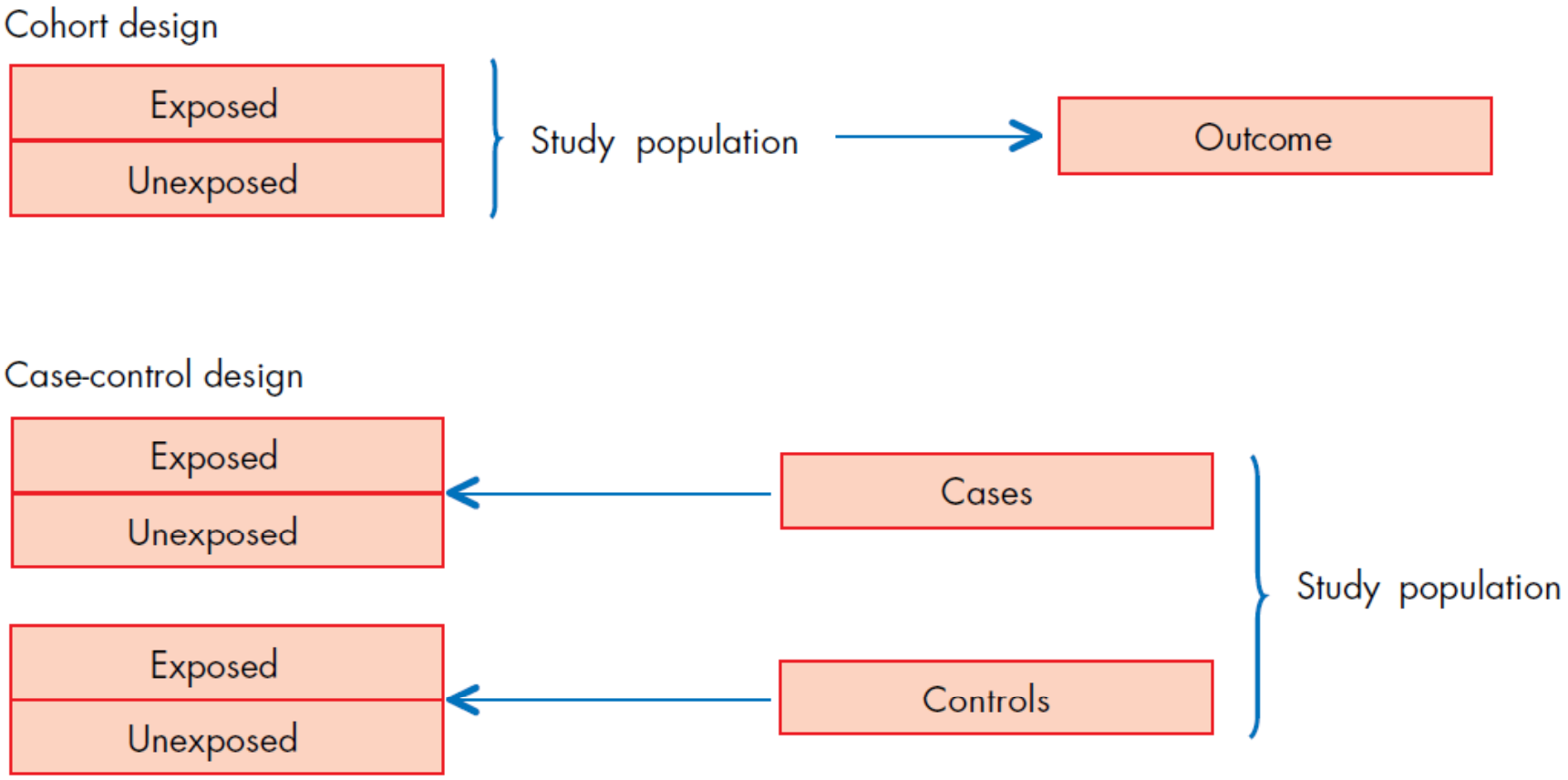
Disegno dello studio caso-controllo

Fabrizio Stracci

Studio caso-controllo

- Classificazione: Studi epidemiologici > osservazionali > analitici
- Altri nomi: case-comparison, case-referent, retrospective studies
- 1. Persone ammalate (casi) e persone non affette dalla malattia in studio (controlli) sono incluse nello studio
- 2. la proporzione di casi con una data caratteristica o con una storia di esposizione viene determinata tra i casi e tra i controlli e confrontata

The three main study designs used in observational studies: cohort (follow up), case–control, and cross sectional.



Jepsen P et al. Heart 2004;90:956-960

Perché utilizzare un disegno caso controllo

- By comparison with other study types,
- case-control studies can yield important findings in a relatively short time,
- and with relatively little money and effort.

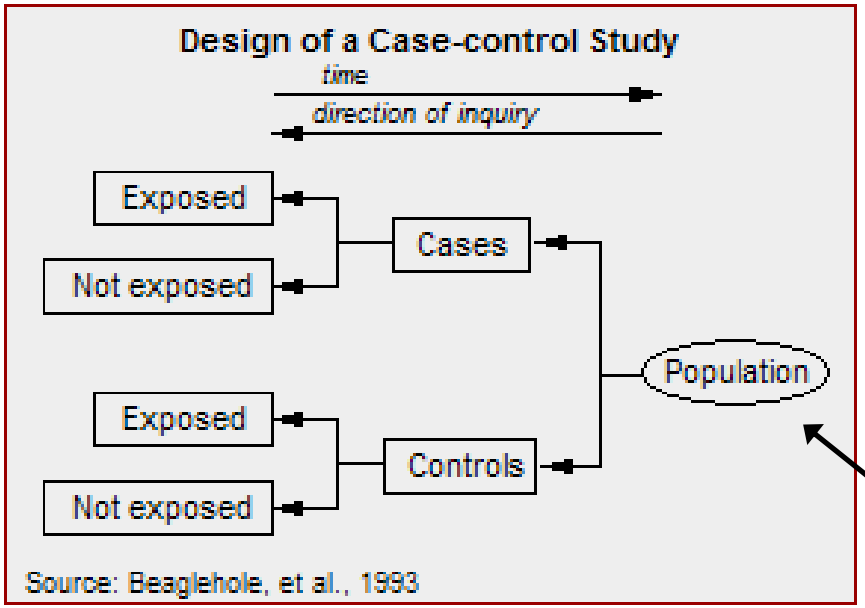
- In particolare quando si intende studiare una malattia rara, lo studio caso-controllo risulta molto più economico dello studio di coorte (il numero di unità sperimentali necessario è molto inferiore rispetto allo studio di coorte)

Perché non utilizzare un disegno caso-controllo

- Se l'esposizione di interesse è rara lo studio è inefficiente
- Lo studio apparentemente più semplice da realizzare presenta difficoltà tecniche maggiori ed è maggiormente soggetto a distorsione
- L'informazione sulla esposizione può non essere accurata o egualmente accurata per casi e controlli; lo stesso problema può riguardare fattori confondenti importanti; il reclutamento dei casi e dei relativi controlli in relazione con una base dello studio definita può essere difficile

La popolazione

Wacholder S. Design issues in case-control studies.
Stat Methods Med Res. 1995; 4:293-309.



Identificabile?

- A case-control study should always be considered in reference to the corresponding full cohort study that might have been undertaken in the same *study base*

Base dello studio

- I casi dello studio sono quelli rilevati tra t1 e t2
- La base dello studio è la coorte dinamica da cui i casi hanno avuto origine

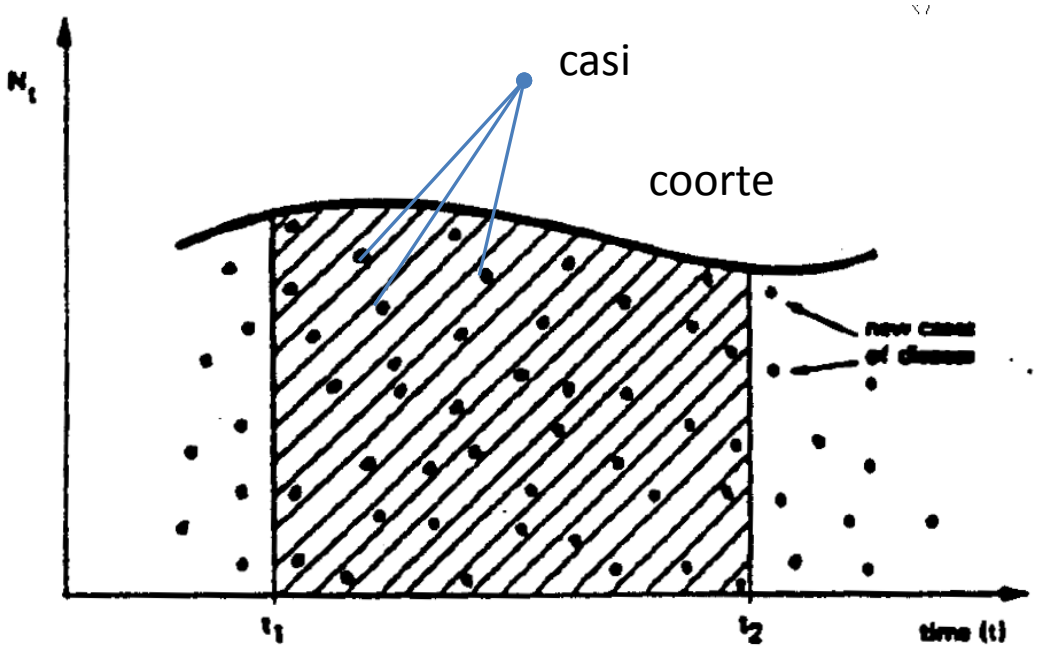
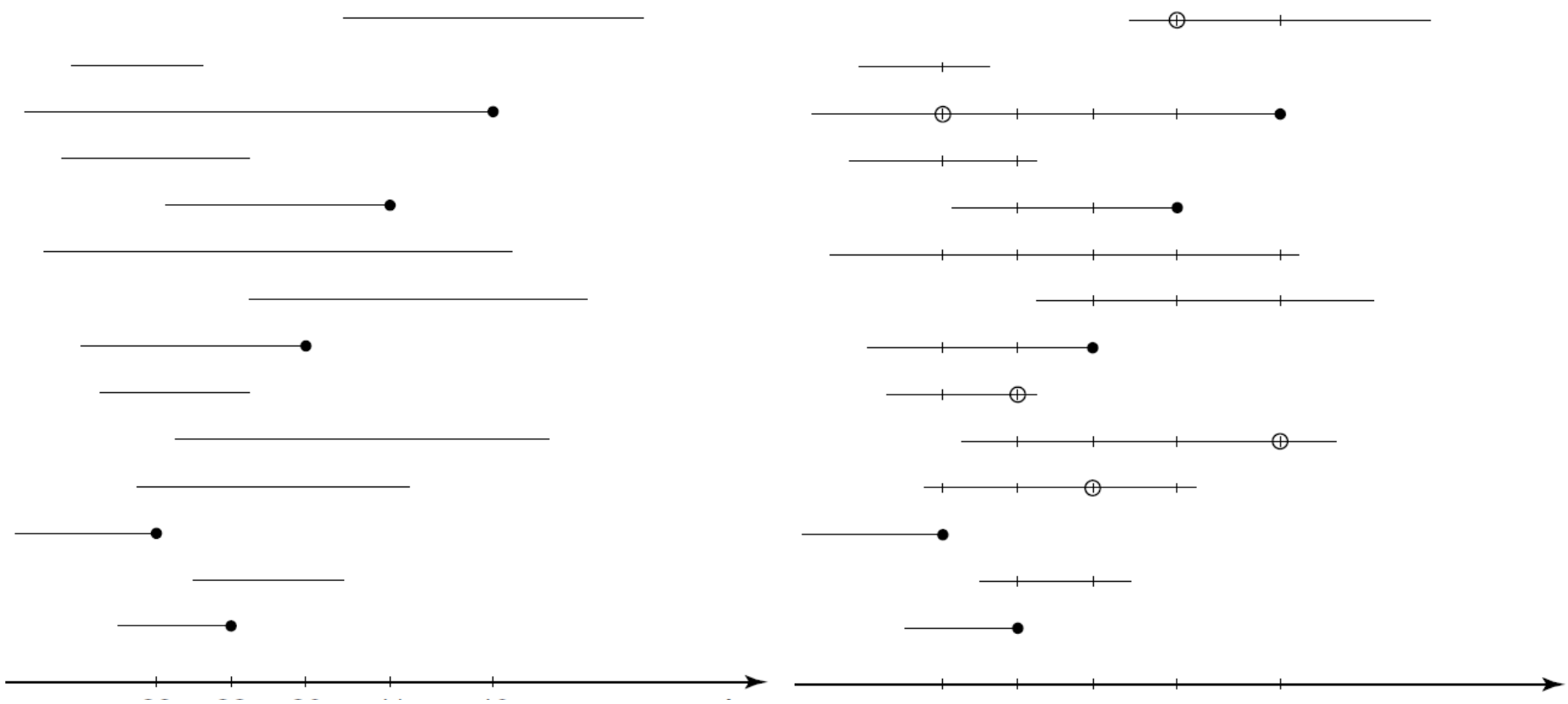


FIGURE 2 *Graphical illustration of the occurrence of new (incident) cases over time in a candidate population (of size N_t at time t)*

Persona-tempo

- L’unità di misura della base dello studio è persona-tempo
- La relazione dello studio caso controllo con la sottostante coorte è fondamentale per il disegno dello studio
- Il parametro di interesse è, solitamente, il rapporto tra i tassi di incidenza e, di conseguenza, lo studio si basa in genere sui casi incidenti (nuovi) ed esclude i casi prevalenti



A sinistra: coorte di 14 individui e rispettivo tempo in studio. Il cerchio pieno rappresenta l'evento. A destra: i tagli individuano l'insieme a rischio in corrispondenza di ogni evento (tempo continuo); i cerchi vuoti rappresentano controlli selezionati

Coorte ≠ popolazione generale

- Come è possibile e utile condurre studi di coorte su coorti speciali, selezionate
- Così può essere utile e possibile condurre studi caso controllo su sottoinsiemi di casi e sulla speciale coorte (base dello studio) che li ha generati
- In altre parole i casi non devono coincidere o rappresentare tutti i casi originati nella popolazione generale

I casi e i controlli

- I casi, idealmente, dovrebbero essere quelli generati nella base dello studio in un intervallo di tempo definito o un loro campione rappresentativo
- I controlli dovrebbero essere estratti casualmente dalla base dello studio o scelti in modo tale che l'esposizione in questo gruppo sia la stesso della base dello studio
- Le unità della base dello studio non incluse dovrebbero essere equivalenti a dati mancanti completamente casuali

Control group

- The control group provides the background proportion of exposure expected in the case group.
- Controls should, therefore, be free of the disease (outcome) being studied, but should be representative of those individuals who would have been selected as cases had they developed the disease.
- In other words, controls should represent the population at risk of becoming cases.
- Selection of controls must be independent of the exposure being investigated.

Perché l'esposizione tra i controlli deve essere la stessa della base dello studio?

- In questo caso possiamo utilizzare i controlli al denominatore e costruire una misura di frequenza della malattia tra gli esposti e tra i non esposti
- Pseudo tassi di Incidenza (PI):
- $PI_{exp+} = casi_{exp+} / controlli_{exp+}$
- $Pi_{exp-} = casi_{exp-} / controlli_{exp-}$
- Queste misure non hanno interpretazione se non nel caso di un tasso di campionamento eguale per controlli esposti e non esposti (ciò che è atteso se i controlli sono selezionati in modo indipendente dalla exp):
- $controlli_{exp+} /anni\ persona_{exp+} = controlli_{exp-} /anni\ persona_{exp-} = r$
- Se dividiamo tra loro i due pseudo tassi abbiamo:
- $PI_{exp+} /PI_{exp-} = (casi_{exp+} /controlli_{exp+}) /(casi_{exp-} /controlli_{exp-}) =$

Dividendo entrambi i termini per gli Anni persona (AP)/AP

- $(casi_{exp+} /((controlli_{exp+} /AP_{exp+})*AP_{exp+})/ (casi_{exp-} / controlli_{exp-} /Ap_{exp-})*Ap_{exp-})) =$
- $=(casi_{exp+} /(r*AP_{exp+}))/ (casi_{exp-} / (r*Ap_{exp-})) =$
- $=(casi_{exp+} /AP_{exp+})/ (casi_{exp-} /Ap_{exp-}) = I_{exp+} / I_{exp-}$.

- In altre parole il rapporto tra gli pseudo-tassi è una stima del rapporto tra tassi di incidenza nella popolazione di origine dei casi posto che il tasso di campionamento dei controlli è indipendente dall'esposizione

Identificazione della base dello studio

- Studi con **base primaria** (su base di popolazione): la popolazione è facilmente identificabile e a partire dalla popolazione si costruisce un sistema per registrare i casi
- Studi con **base secondaria**: lo studio origina dalla identificazione dei casi e dipende dalla individuazione di controlli in grado di rappresentare la base dello studio che ha dato origine ai casi.
- Ad esempio in uno studio in cui i casi sono identificati in un dato ospedale i controlli dovrebbero essere selezionati dalla popolazione che in caso di malattia si sarebbe rivolta a quel medesimo ospedale, una popolazione non direttamente identificabile.

Tipi di controlli utilizzati negli studi

- Selezione casuale dalla base dello studio negli studi di popolazione (o appaiamento)
- Negli studi con base secondaria:
  - Vicini di casa (complicato da realizzare in pratica, possibile collegamento con esposizione o differenza rispetto a base dello studio)
  - Telefonate a numeri composti a caso (*random-digit dialing;* i casi debbono disporre di telefono, mancano elenchi dei numeri mobili, la probabilità di selezione è legata alle utenze e non al numero di residenti)
  - Controlli ospedalieri (difficile identificare la base dello studio: diversa attrazione per diverse patologie, persone ospedalizzate scelte come controlli possono non rappresentare l'esposizione nella base dello studio)

Matching in Wacholder S. Design issues in case-control studies. *Stat Methods Med Res.* 1995;4:293-309.

Matching has been commonly used in studies with and without a roster. Controls are selected so that the value of a covariate, believed to be a confounder, is the same for the case and the controls. The main advantage of matching is the additional efficiency that can sometimes be achieved relative to random sampling when the control and case distributions are substantially different. But the efficiency advantages for matching are often too slight[88] to compensate for: any additional cost or extra effort required to identify controls[89,90]; possible exclusion of cases for whom no match is found[89]; and reduced flexibility in the analysis.[36] Less often, and not always successfully, matching is used in an attempt to capture a set of unmeasured risk factors, such as social class or access to a particular health care facility, in a single variable that is easy to measure, such as neighbourhood.[36]

Sometimes matching hurts rather than helps. *Overmatching* is the term for counter-productive matching, i.e. matching that can cause bias or reduce precision.[36,91,92] Matching on a variable in the pathway between exposure and disease can lead to bias. An example would be matching on endometrial hyperplasia in a study of oestrogen and endometrial cancer.[91,92] Matching on a variable that is not itself a strong risk factor can lead to reduction in precision if it reduces the variability of the exposure *conditional on the matching variable*, i.e. the variability that is a strong determinant of the precision of the estimate of effect. Finally, the analysis of a matched study needs to account for the matching, in contrast to unmatched studies, where a decision about stratification can be made at the analysis stage.

Relationship between long durations and different regimens of hormone therapy and risk of breast cancer.

Li CI, et al. JAMA. 2003;289:3254-63

- **Abstract**
- CONTEXT: Women using combined estrogen and progestin hormone replacement therapy (CHRT) have an increased risk of breast cancer; however, data on use for long durations and on risk associated with patterns of use are lacking. OBJECTIVE: To evaluate relationships between durations and patterns of CHRT use and risk of breast cancer by histological type and hormone receptor status.
- DESIGN: **Population-based case-control study**. SETTING: Three counties in western Washington State. PARTICIPANTS: Nine hundred seventy-five women 65-79 years of age diagnosed with invasive breast cancer from April 1, 1997, through May 31, 1999 (histology: 196 lobular cases, 656 ductal cases, 114 cases with other histological type, and 9 cases with an unspecified histological type; estrogen receptor (ER)/progesterone receptor (PR) status: 646 ER+/PR+ cases, 147 ER+/PR- cases, and 101 ER-/PR- cases [6 ER-/PR+ cases and 75 cases with unknown ER/PR status were not included in the analyses herein]) and 1007 population controls. MAIN OUTCOME MEASURES: **Risks of invasive lobular, ductal, ER+/PR+, ER+/PR-, and ER-/PR- breast carcinomas.**
- RESULTS: Women using unopposed estrogen replacement therapy (ERT) (exclusive ERT use), even for 25 years or longer, had no appreciable increase in risk of breast cancer, although the associated **odds ratios** were not inconsistent with a possible small effect. Ever users of CHRT (includes CHRT users who also had used ERT) had a 1.7-fold (95% confidence interval [CI], 1.3-2.2) increased risk of breast cancer, including a 2.7-fold (95% CI, 1.7-4.3) increased risk of invasive lobular carcinoma, a 1.5-fold (95% CI, 1.1-2.0) increased risk of invasive ductal carcinoma, and a 2.0-fold (95% CI, 1.5-2.7) increased risk of ER+/PR+ breast cancers. The increase in risk was greatest in those using CHRT for longer durations (users for 5-14.9 years and >or=15 years had 1.5-fold [95% CI, 1.0-2.3] and 1.6-fold [95% CI, 1.0-2.6] increases in risk of invasive ductal carcinoma, respectively, and 3.7-fold [95% CI, 2.0-6.6] and 2.6-fold [95% CI, 1.3-5.3] increases in risk of invasive lobular carcinoma, respectively. Associations of similar magnitudes were seen among users of both sequential and continuous CHRT. Risks of ER+/PR- and ER-/PR- tumors were not increased by use of any form of hormone replacement therapy; however, small numbers of these tumors limited power to detect possible associations.
- CONCLUSION: These data suggest that use of CHRT is associated with an increased risk of breast cancer, particularly invasive lobular tumors, whether the progestin component was taken in a sequential or in a continuous manner.

Cases

- Women aged 65 to 79 years with no prior history of in situ or invasive breast cancer when diagnosed with invasive breast cancer from April 1, 1997, through May 31, 1999, were eligible as cases. The Cancer Surveillance System (CSS**), the population-based tumor registry** that serves the Seattle-Puget Sound region of Washington State and participates in the Surveillance, Epidemiology, and End Results program of the National Cancer Institute, was used to identify these women.

Partecipazione

- Of the 1210 **eligible cases** identified, 975 (**80.6%**) were interviewed.
- Eligible cases were approached through their physicians.
- Patients for whom physicians gave permission to contact were invited to **participate in the study** through a mailed letter describing the study as an investigation of causes of breast cancer in older women.
- Fourteen percent of eligible cases refused to be interviewed, 4% died before an interview could be conducted, 1% moved away from the area, and the physicians treating 1% of cases refused to allow contact with their patients.

Controls

- The HCFA records were used to identify women from the general population of female residents of King, Pierce, and Snohomish counties who were the same ages as cases to serve as controls.
- Once identified, eligible controls were sent a letter similar to the one sent to cases that described the study and invited them to participate.
- Of the 1365 eligible women selected as controls, 1007 (**73.8%**) were interviewed. Twenty-two percent of eligible controls refused to be interviewed, 2% died after selection but before they could be interviewed, 2% moved away, and 1% could not be located.

Intervista

- Cases and controls were interviewed in person in their homes by a trained interviewer, and a standardized structured questionnaire was used to ask them about their reproductive history, body size, medical history, and family history of cancer
- Additionally, detailed histories of all episodes of HRT use, including beginning and ending dates, total duration, brand, dose, and pattern of use (number of days per month) were obtained.

Analisi

- We compared all breast cancer cases with controls using **unconditional logistic regression**[22] and compared ILC and IDC cases with controls, and cases with different ER/PR profiles with controls, using polytomous logistic regression.[23]
- All analyses were conducted using Stata version 7.0 (Stata Corp, College Station, Tex).
- Both statistical approaches were used to calculate **odds ratios (ORs)** as an estimate of the relative risk and to compute 95% confidence intervals (CIs) and associated *P* values; $P<.05$ was used to determine statistical significance.
- Multiple variables were evaluated as potential confounders, including family history of breast cancer (first-degree, no first-degree), type of menopause (natural, induced, simple hysterectomy [hysterectomy without a bilateral oophorectomy]), age at menopause (5-year categories), parity, body mass index 1 year prior to reference date (quartiles of control population), mean daily alcohol use during the 20 years prior to reference date (none, ≤8.1 g, ≥8.2 g), and oral contraceptive use (never, <5 years, ≥5 years).
- Only adjustment for type of menopause changed the risk estimates of the ORs of interest by more than 10%. Type of menopause was likely a confounder because ERT is associated with an increased risk of endometrial cancer but CHRT is not, and therefore ERT is more likely to be considered for women who have had a hysterectomy and CHRT for those with an intact uterus.[24] Thus, all analyses were adjusted both for type of menopause and for age (continuous), since cases and controls were matched on age.

From: **Relationship Between Long Durations and Different Regimens of Hormone Therapy and Risk of Breast Cancer**

JAMA. 2003;289(24):3254-3263.
doi:10.1001/jama.289.24.3254

Table 1. Distribution of Demographic Characteristics and Risk Factors for Women With Invasive Breast Carcinoma and for Controls

Characteristic	No. (%) Controls (n = 1007)	No. (%) All Cases (n = 975)	*P* Value, Controls vs Cases	No. (%) IDC Cases (n = 656)	No. (%) ILC Cases (n = 196)	*P* Value, IDC vs ILC
Reference age, y						
65-69	330 (32.8)	300 (30.8)		204 (31.1)	58 (29.6)	
70-74	381 (37.8)	381 (39.1)	.63	252 (38.4)	85 (43.4)	.44
75-79	296 (29.4)	294 (30.2)		200 (30.5)	53 (27.0)	
Race						
White	925 (91.9)	929 (95.3)		623 (95.0)	188 (95.9)	
Black	37 (3.7)	16 (1.6)	.01	11 (1.7)	3 (1.5)	.06
Asian/Pacific Islander	29 (2.9)	19 (1.9)		18 (2.7)	1 (0.5)	
Other/unknown	16 (1.6)	11 (1.1)		4 (0.6)	4 (2.0)	
Income, $						
≤14 999	191 (21.7)	177 (21.3)		124 (22.1)	30 (17.9)	
15 000-24 999	214 (24.3)	198 (23.9)	.82	139 (24.8)	39 (23.2)	.25
25 000-49 999	296 (33.6)	296 (35.7)		204 (36.4)	60 (35.7)	
≥50 000	180 (20.4)	159 (19.2)		94 (16.8)	39 (23.2)	
Missing data	126	145		95	28	
Marital status						
Married	536 (54.6)	517 (54.4)		343 (53.8)	103 (53.4)	
Widowed	315 (32.1)	301 (31.7)	.92	201 (31.6)	65 (33.7)	.90
Divorced/separated	121 (12.3)	125 (13.1)		87 (13.7)	23 (11.9)	
Single	10 (1.0)	8 (0.8)		6 (0.9)	2 (1.0)	
Missing data	25	24		19	3	
Education						
<High school	153 (15.2)	126 (12.9)		87 (13.3)	19 (9.7)	
High school graduate	395 (39.3)	376 (38.6)	.25	251 (38.3)	84 (42.9)	.44
Some college	286 (28.4)	312 (32.0)		210 (32.0)	59 (30.1)	
College graduate	172 (17.1)	161 (16.5)		108 (16.5)	34 (17.3)	
Missing data	1	0		0	0	
Age at menarche, y						
8-11	173 (17.2)	182 (18.8)		126 (19.4)	35 (17.9)	
12-13	520 (51.7)	525 (54.2)	.12	357 (54.8)	104 (53.1)	.65
≥14	313 (31.1)	261 (27.0)		168 (25.8)	57 (29.1)	
Missing data	1	7		5	0	
Parity						
Nulliparous	94 (9.3)	88 (9.0)	.81	57 (8.7)	20 (10.2)	.52
Parous	913 (90.7)	887 (91.0)		599 (91.3)	176 (89.8)	
Age at first birth, y						
14-19	187 (20.5)	152 (17.2)		98 (16.4)	31 (17.6)	
20-24	435 (47.7)	432 (48.9)	.32	302 (50.7)	84 (47.7)	.90
25-29	205 (22.5)	206 (23.3)		136 (22.8)	41 (23.3)	
≥30	85 (9.3)	93 (10.5)		60 (10.1)	20 (11.4)	
Missing data	95	92		60	20	
Type of menopause						
Natural	607 (61.6)	583 (61.4)		400 (62.8)	113 (59.8)	
Induced	148 (15.0)	129 (13.6)	.19	78 (12.2)	29 (15.3)	.37
Simple hysterectomy	231 (23.4)	237 (25.0)		159 (25.0)	47 (24.9)	
Missing data	21	26		19	7	
Age at menopause, y						
23-39	64/647 (9.9)	38/574 (6.6)		29/382 (7.6)	4/112 (3.6)	
40-44	99/647 (15.3)	77/574 (13.4)		40/382 (10.5)	19/112 (17.0)	
45-49	172/647 (26.6)	165/574 (28.7)	.18	116/382 (30.4)	27/112 (24.1)	.14
50-54	222/647 (34.3)	217/574 (37.8)		143/382 (37.4)	47/112 (42.0)	
≥55	90/647 (13.9)	77/574 (13.4)		54/382 (14.1)	15/112 (13.4)	

(continued)

Figure Legend:

From: **Relationship Between Long Durations and Different Regimens of Hormone Therapy and Risk of Breast Cancer**

JAMA. 2003;289(24):3254-3263. doi:10.1001/jama.289.24.3254

Table 3. Use of Combined Estrogen and Progestin Hormone Replacement Therapy (CHRT) and Risk of Overall and Specific Histological Types of Invasive Breast Carcinoma*

Regimen	Controls, No. (%) (n = 1007)	All Cases (n = 975)			IDC Cases (n = 656)			ILC Cases (n = 196)		
		No. (%)	OR (95% CI)	*P* Value	No. (%)	OR (95% CI)	*P* Value	No. (%)	OR (95% CI)	*P* Value
Exclusive Ever Use of CHRT†										
Never	339 (33.7)	284 (29.1)	Reference		199 (30.3)	Reference		47 (24.0)	Reference	
Ever	96 (9.5)	136 (13.9)	1.8 (1.3-2.5)	<.001	89 (13.6)	1.6 (1.1-2.3)	.01	29 (14.8)	2.5 (1.4-4.3)	.002
6 mo-4.9 y	29 (2.9)	30 (3.1)	1.3 (0.8-2.2)	.35	23 (3.5)	1.4 (0.8-2.5)	.27	5 (2.6)	1.4 (0.5-3.9)	.52
5-14.9 y	37 (3.7)	57 (5.8)	2.0 (1.3-3.2)	.004	33 (5.0)	1.6 (1.0-2.7)	.07	15 (7.7)	3.4 (1.7-7.0)	.001
≥15 y	30 (3.0)	49 (5.0)	2.0 (1.2-3.3)	.01	33 (5.0)	1.9 (1.1-3.2)	.02	9 (4.6)	2.4 (1.1-5.5)	.04
Ever Use of CHRT†										
Never	339 (33.7)	284 (29.1)	Reference		199 (30.3)	Reference		47 (24.0)	Reference	
Ever	165 (16.4)	232 (23.8)	1.7 (1.3-2.2)	<.001	148 (22.6)	1.5 (1.1-2.0)	.01	58 (29.6)	2.7 (1.7-4.3)	<.001
6 mo-4.9 y	60 (6.0)	65 (6.7)	1.3 (0.9-2.0)	.18	46 (7.0)	1.3 (0.9-2.1)	.20	14 (7.1)	1.8 (0.9-3.6)	.09
5-14.9 y	63 (6.3)	101 (10.4)	1.9 (1.3-2.8)	<.001	58 (8.8)	1.5 (1.0-2.3)	.03	30 (15.3)	3.7 (2.0-6.6)	<.001
≥15 y	42 (4.2)	66 (6.8)	1.8 (1.2-2.7)	.004	44 (6.7)	1.6 (1.0-2.6)	.02	14 (7.1)	2.6 (1.3-5.3)	.01
Recency of CHRT‡										
Never	339 (33.7)	284 (29.1)	Reference		199 (30.3)	Reference		47 (24.0)	Reference	
Former	20 (2.0)	32 (3.3)	2.0 (1.1-3.6)	.02	23 (3.5)	2.0 (1.1-3.7)	.03	5 (2.6)	2.0 (0.7-5.7)	.19
Current	115 (11.4)	178 (18.3)	1.9 (1.4-2.6)	<.001	113 (17.2)	1.7 (1.2-2.4)	<.001	44 (22.4)	3.1 (1.9-5.2)	<.001
6 mo-4.9 y	32 (3.2)	31 (3.2)	1.2 (0.7-2.1)	.44	24 (3.7)	1.3 (0.8-2.3)	.33	5 (2.6)	1.3 (0.5-3.6)	.60
5-14.9 y	50 (5.0)	87 (8.9)	2.2 (1.5-3.3)	<.001	49 (7.5)	1.7 (1.1-2.7)	.02	27 (13.8)	4.6 (2.5-8.5)	<.001
≥15 y	33 (3.3)	60 (6.2)	2.2 (1.4-3.5)	.001	40 (6.1)	2.0 (1.2-3.4)	.01	12 (6.1)	3.0 (1.4-6.3)	.004

Abbreviations: CI, confidence interval; IDC, invasive ductal carcinoma; ILC, invasive lobular carcinoma; OR, odds ratio.

*All models are adjusted for age of women at their reference date (see "Methods" section) and type of menopause. Analyses among all cases (including IDC cases, ILC cases, and cases with other or unspecified histological types) compared with controls were conducted using unconditional logistic regression. Analyses comparing IDC and ILC cases with controls were conducted using polytomous logistic regression. Separate categories of 15-24.9 years and ≥25 years are not given because of small numbers. Evaluation of recency of CHRT use is based on all ever users of CHRT, rather than being restricted only to exclusive ever users of CHRT.

†Never users defined as women never using any type of hormone replacement therapy (HRT); ever users defined as those using CHRT for ≥6 months, with exclusive CHRT users including only those who ever used CHRT but never used ERT for ≥6 months.

‡Never users defined as women never using any type of HRT; former users defined as those using CHRT for ≥6 months with last use >6 months prior to reference date who are not current ERT users; current users defined as those using CHRT for ≥6 months with last use within the 6 months prior to reference date.

Certain limitations of our study should be considered

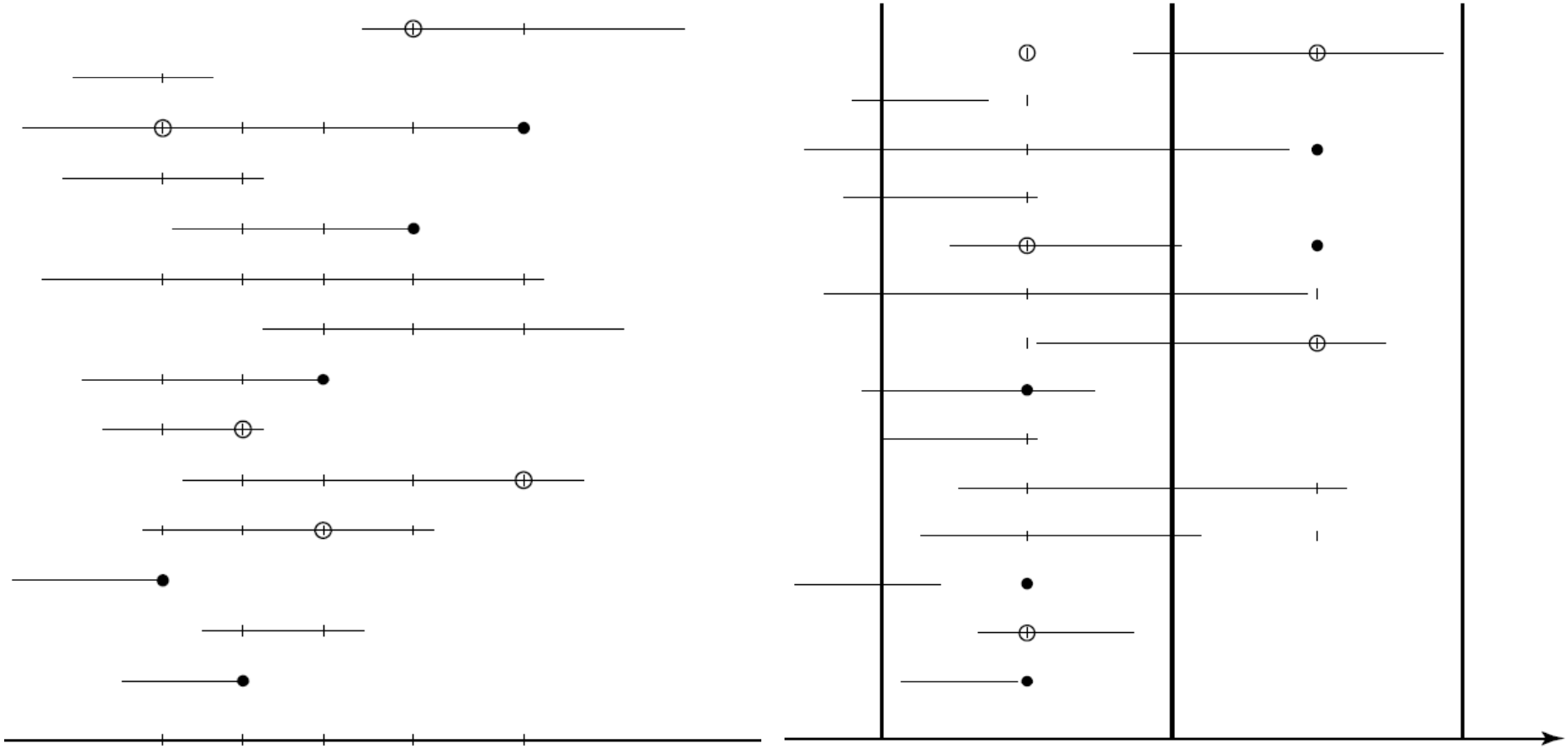
- We did not perform independent or centralized pathology reviews or hormone receptor evaluations of the tumors, …**Misclassification** of tumor histological type and ER/PR status may have resulted in some instances.
- Additionally, we were able to interview only 80.6% of all eligible cases and 73.8% of all eligible controls. Our results could be biased if the women we were unable to interview differed from those who did participate with regard to type or patterns of HRT use *(**selection bias**).*
- We also relied on participants' **recall** of the types of HRT used as well as the timing and duration of use. However, studies have shown reasonable agreement between reports from postmenopausal women and physicians' or medical records.25 - 28

Studi caso-controllo prospettici

- “Standard” case–control studies, the most common study design in epidemiologic research, may often be viewed as nested case–control studies in which a portion of underlying cohort (usually among the nondiseased) has not been identified
- Lo studio caso controllo innestato in una coorte (*nested case-control* design) è uno studio che origina da una coorte enumerata e seguita nel tempo.
- Un sottoinsieme della coorte (ed eventualmente dei casi) viene estratto con selezione casuale e ulteriori informazioni vengono raccolte per i casi e il sottoinsieme della coorte selezionato
- Lo studio nested case-control è impiegato per indagini in cui non è conveniente raccogliere l’informazione per l’intera coorte, ad esempio quando deve essere somministrato un test costoso o quando la coorte è molto numerosa

Cohort sampling designs

- Cohort sampling designs are used in follow-up studies
- when large cohorts are needed to observe enough cases
- but it is not feasible to collect data on all covariates for the whole cohort



Due schemi di estrazione dei controlli negli studi innestati in una coorte; a sinistra la selezione dei controlli avviene tra gli esposti a rischio al momento di un evento (appaiamento), a destra tra gli individui a rischio in un intervallo di tempo definito (dati non appaiati). Da Langholz B. *Case–Control Study, Nested*. Vol 1, pp. 646–655 In Encyclopedia of Biostatistics

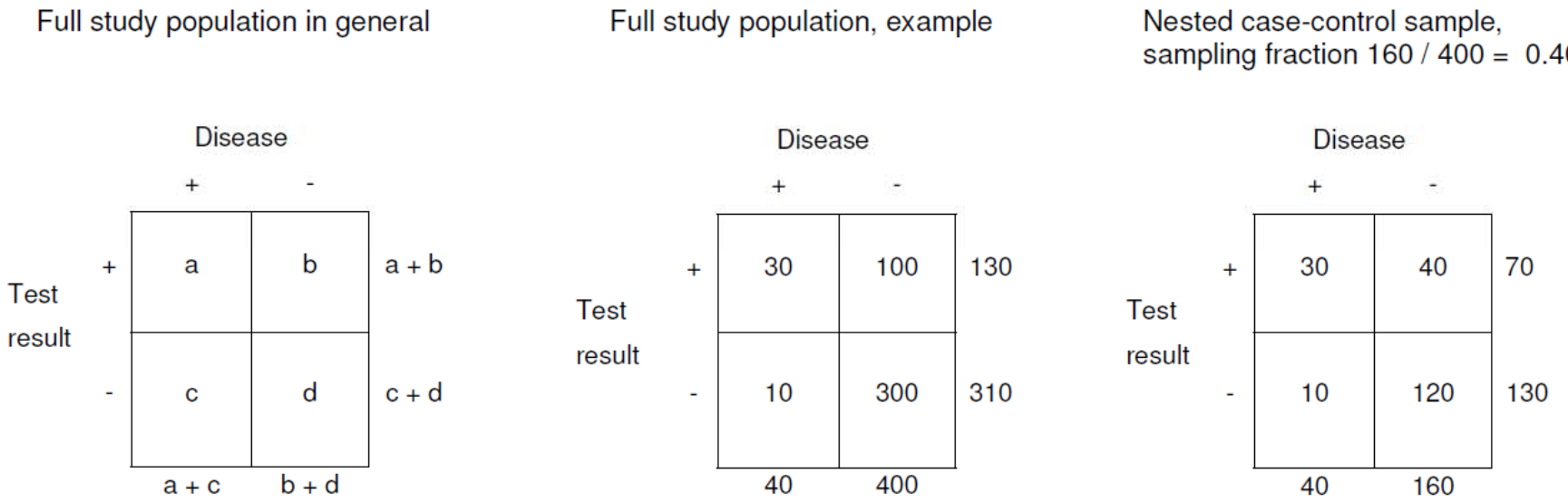


Figure 1

Theoretical example of a full study population and a nested case-control sample. The index test result and the outcome are obtained for all patients of the study population. The case-control ratio was 1:4 (sampling fraction (SF) = 160/400 = 0.40). Valid diagnostic accuracy measures can be obtained from the nested case-control sample, by multiplying the controls with 1/sampling fraction. For example, the positive predictive value (PPV) of a full study population can be calculated with a/(a + b), in this example 30/(30 + 100) = 0.23. In a nested case-control sample the PPV is calculated with a/(a + (1/SF)*b), in this example: 30/(30 + 2.5*40) = 0.23. In a case-control sample however, the controls are sampled from a source population with unknown size. Therefore, the sample fraction is unknown and valid estimate of the PPV cannot be calculated.

Studio caso-coorte

- Come nello studio caso controllo innestato, ulteriori informazioni sono necessarie per i casi e per un sottoinsieme della coorte
- Diversamente dallo studio *nested,* non vi è appaiamento tra casi e controlli presenti in studio al momento della diagnosi del caso
- I controlli possono essere utilizzati per diversi insiemi di casi

Case-cohort design:

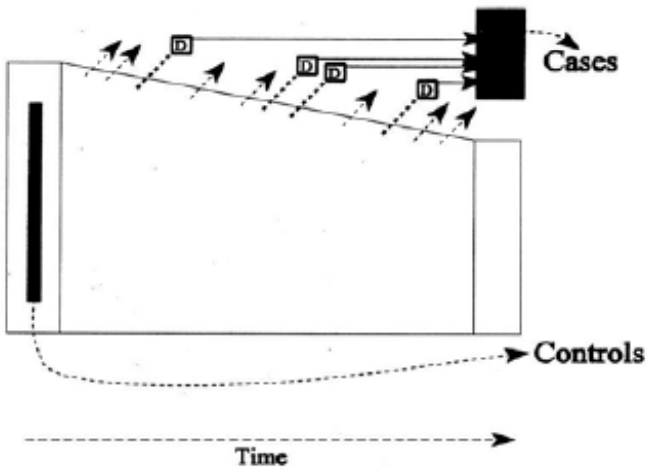


Figure 1–20 Case-control study in which the controls are selected from the baseline cohort (case-cohort study). Cases are represented by "D" boxes. Broken diagonal lines with arrows represent losses to follow-up.

Controls sample from the baseline cohort (regardless of future disease status)

Nested case-control design:

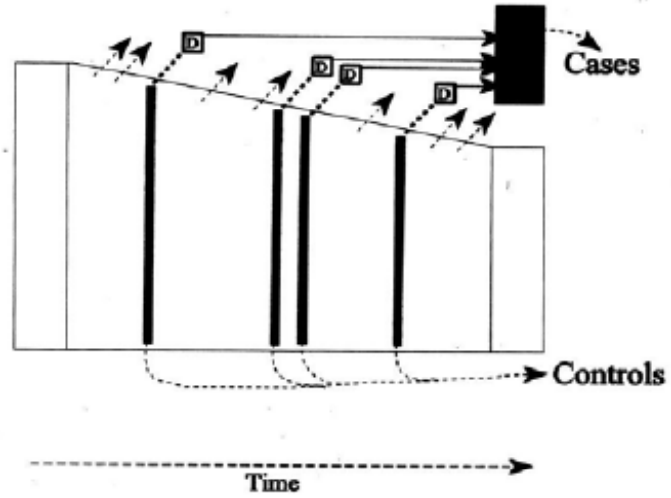


Figure 1–21 Nested case-control study in which the controls are selected at each time when a case occurs (incidence density sampling). Cases are represented by "D" boxes. Broken diagonal lines with arrows represent losses to follow-up.

Controls sampled from people currently at risk - in the risk set at the time an incident case occurs in the study base

Figures from: Szklo & Nieto. Epidemiology: beyond the basics. Aspen Publishers, 2000

Case-cohort design in practice - experiences from the MORGAM Project.

Kulathinal S, et al. *Epidemiol Perspect Innov*. 2007;4:15.

- **Abstract**
- When carefully planned and analysed, the case-cohort design is a powerful choice for follow-up studies with multiple event types of interest. While the literature is rich with analysis methods for case-cohort data, little is written about the designing of a case-cohort study. Our experiences in designing, coordinating and analysing the MORGAM case-cohort study are potentially useful for other studies with similar characteristics. The motivation for using the case-cohort design in the MORGAM genetic study is discussed and issues relevant to its planning and analysis are studied. We propose solutions for appending the earlier case-cohort selection after an extension of the follow-up period and for achieving maximum overlap between earlier designs and the case-cohort design. Approaches for statistical analysis are studied in a simulation example based on the MORGAM data.

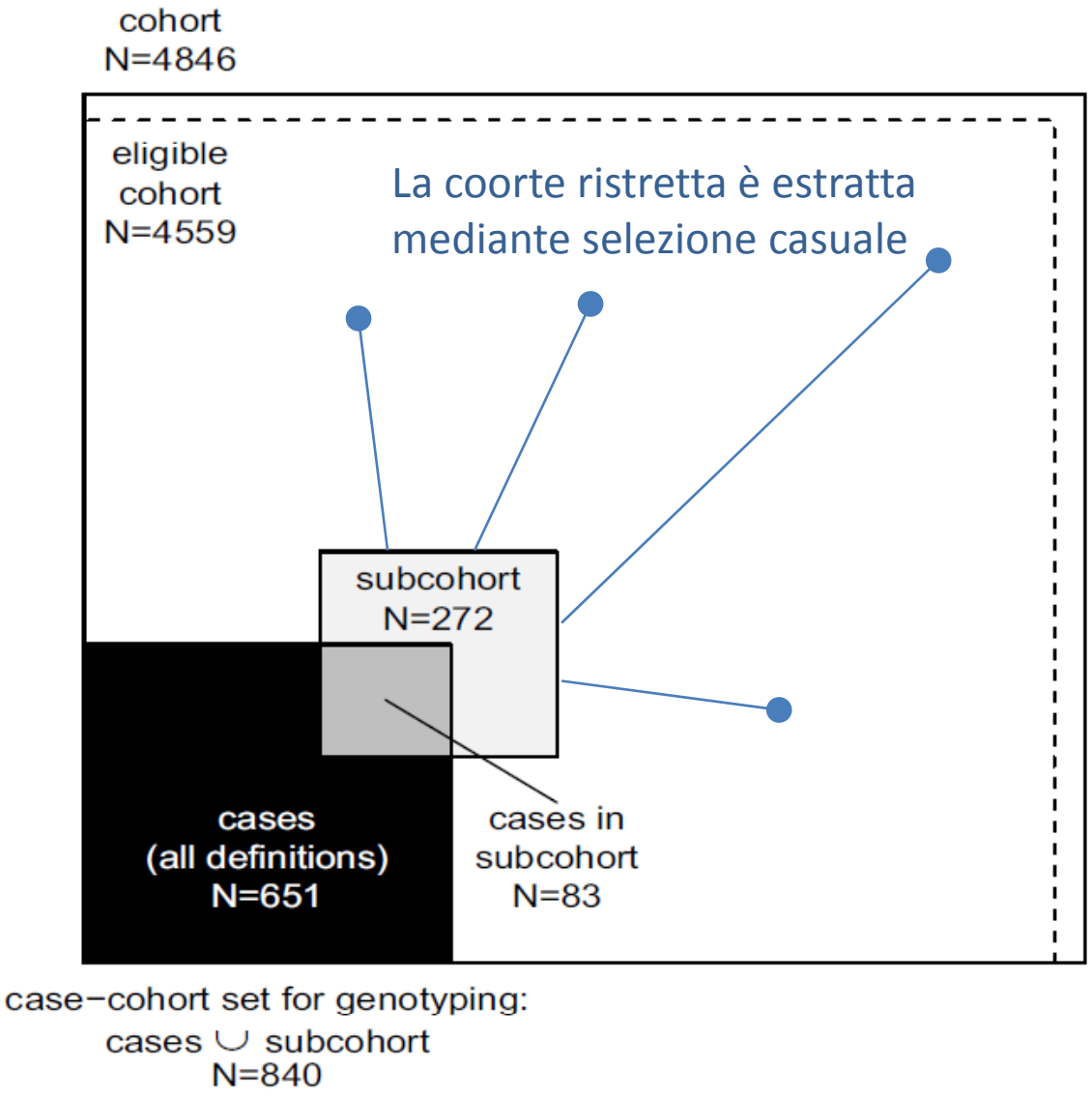


Figure 1

Conceptual illustration of the case-cohort design in the example cohort. Areas are proportional to numbers of observations.

E’ possibile confrontare i dati disponibili della coorte con il sottoinsieme della sub-coorte selezionata

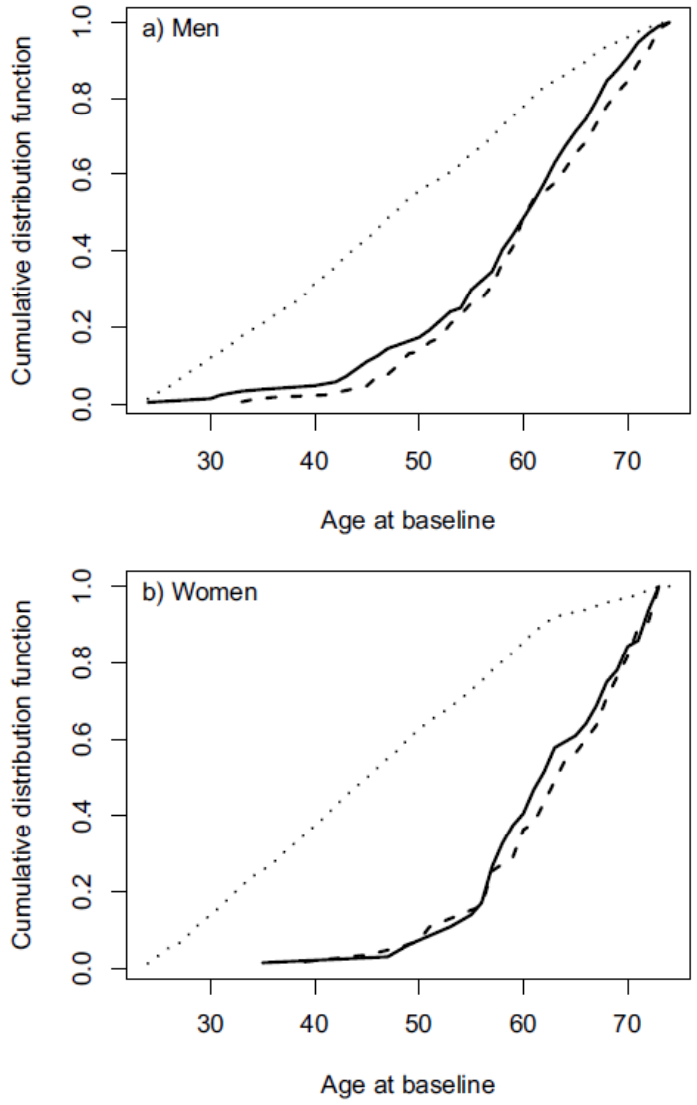


Figure 2
Age distribution of the cohort (dotted line), the subcohort (solid line) and the CHD cases (dashed line) in the example cohort.

Motivi della scelta

The study aims at exploring the relationships between the development of cardiovascular diseases and their classic and genetic risk factors.

MORGAM opted for a case-cohort design for its genetic study because genotyping of the entire cohorts is not viable due to the cost consideration and because there is interest in several definitions of a case.

Cohort sampling designs are used in follow-up studies when large cohorts are needed to observe enough cases but it is not feasible to collect data on all covariates for the whole cohort.