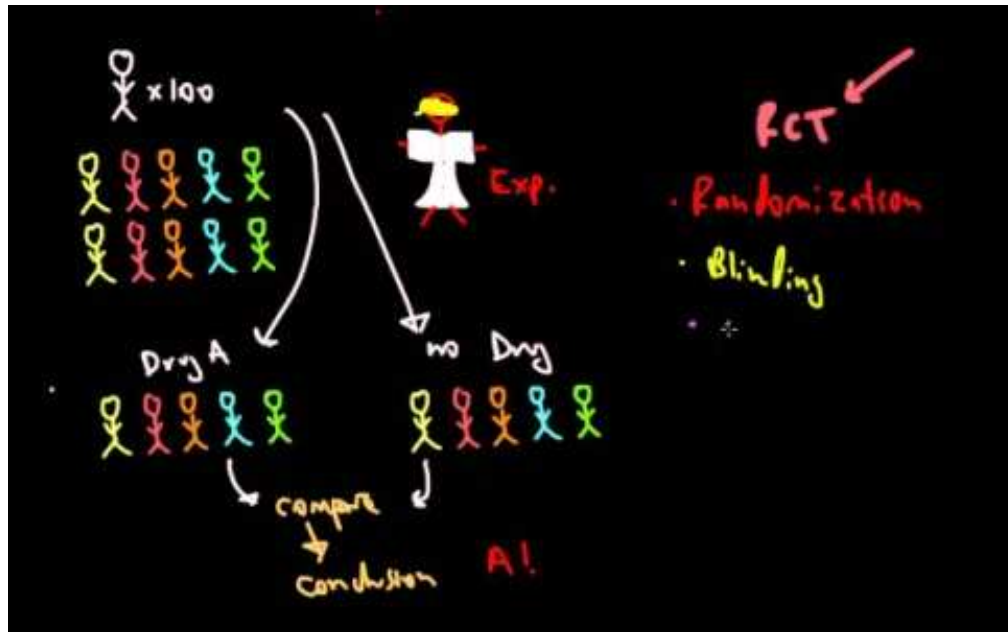

Studi clinici randomizzati: teoria e pratica



Marco Massari, Roberto Raschetti

Inferenze statistiche: usate per mostrare che una causa produce una differenza negli effetti (modelli probabilistici, controfattuali, ecc)

Statistics and Causal Inference

PAUL W. HOLLAND*

© 1986 American Statistical Association
Journal of the American Statistical Association
December 1986, Vol. 81, No. 396, Theory and Methods

The reaction of many statisticians when confronted with the possibility that their profession might contribute to a discussion of causation is immediately to deny that there is any such possibility. “That correlation is not causation is perhaps the first thing that must be said” (Barnard 1982,

Some authors focus on the ultimate meaningfulness of the notion of causation. Others are concerned with deducing the causes of a given effect. Still others are interested in understanding the details of causal mechanisms. The emphasis here will be on *measuring the effects of causes* because this seems to be a place where statistics, which is concerned with measurement, has contributions to make. It is my opinion that an emphasis on the effects of causes rather than on the causes of effects is, in itself, an important consequence of bringing statistical reasoning to bear on the analysis of causation and directly opposes more traditional analyses of causation.

Teoria probabilistica della causalità

Patrick Suppes (1970)

Roughly speaking, the modification of Hume's analysis I propose is to say that one event is the cause of another if the appearance of the first event is followed with a high probability by the appearance of the second, and there is no third event that we can use to factor out the probability relationship between the first and second events.

Dati due istanti di tempo: $t_1 < t_2$ e $P(C_{t_1}) > 0$

Se: $P(E_{t_2} | C_{t_1}) > P(E_{t_2} | \text{non-}C_{t_1})$

C è una causa "*prima facie*" dell'evento E

*L'idea centrale è che le cause aumentano le probabilità dei loro effetti, a parità di altre condizioni (**ceteris paribus**).*

L'approccio controfattuale

C_{t1} vs non- **C_{t1}**

L'approccio controfattuale

(Sliding doors)

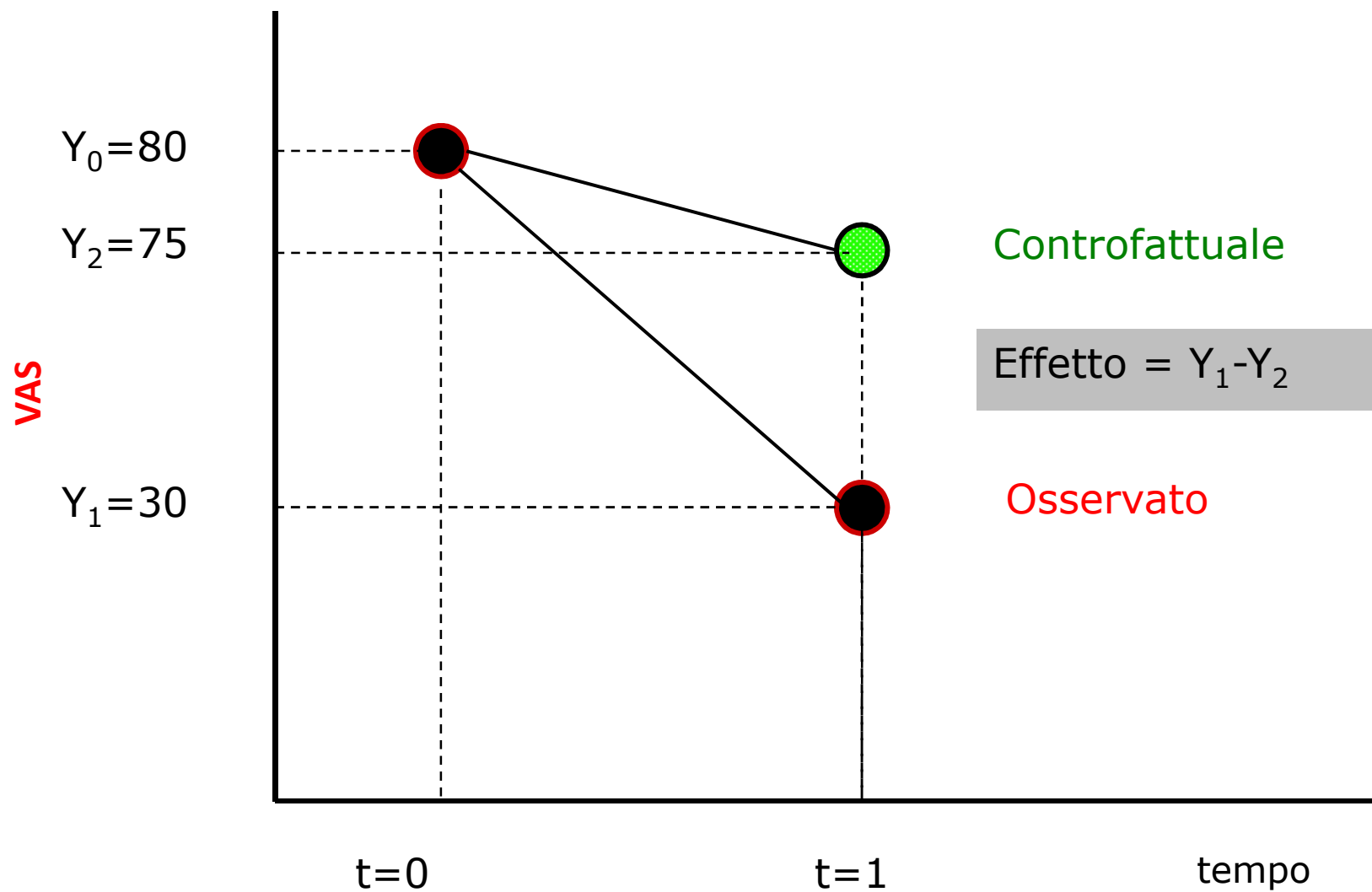
"Possiamo definire una causa come essere un oggetto seguito da un altro... dove, se il primo oggetto non ci fosse stato il secondo non sarebbe mai esistito" David Hume [1748]

Se C non fosse avvenuto non avrei osservato E

- " La prima guerra mondiale non sarebbe scoppiata se non ci fosse stato l'attentato di Sarajevo"*
- " se non mi fossi distratto con il telefonino non avrei tamponato"*
- " Se un'ora fa non avessi preso due aspirine, il mio mal di testa non sarebbe passato"*

La definizione che ipotizza cosa sarebbe avvenuto in condizioni contrarie a quelle effettivamente osservate è chiamata

argomentazione controfattuale



L'effetto causale a livello del singolo individuo, è definito come la differenza nella variabile di risposta Y che potrebbe essere osservata se l'unità fosse esposta al trattamento e al non trattamento nella stessa identica situazione e sotto le medesime circostanze.

[differenza delle differenze]

Il problema fondamentale dell'inferenza causale

Tutte le unità sono osservate (osservabili) solo in uno dei due stati possibili: per ciascuna possiamo osservare Y_1 o Y_2 ma non ambedue.

Di conseguenza l'effetto causale individuale
non può essere direttamente osservato.

Come possiamo operare per valutare un relazione causale?

L'effetto causale medio (Average Treatment Effect, ATE) in una popolazione di soggetti con una misura indiretta effettuata tramite il confronto tra due gruppi.

Il principale scopo dei gruppi di controllo è quello di consentire di discriminare nei pazienti gli esiti causati dal trattamento in studio dagli esiti causati da altri fattori, come la progressione naturale della malattia, le attese degli osservatori e dei pazienti o da altri trattamenti. L'esperienza derivata dal gruppo di controllo ci dice **cosa sarebbe avvenuto ai pazienti se essi non avessero ricevuto il trattamento in studio** o se avessero ricevuto un differente trattamento riconosciuto come efficace.

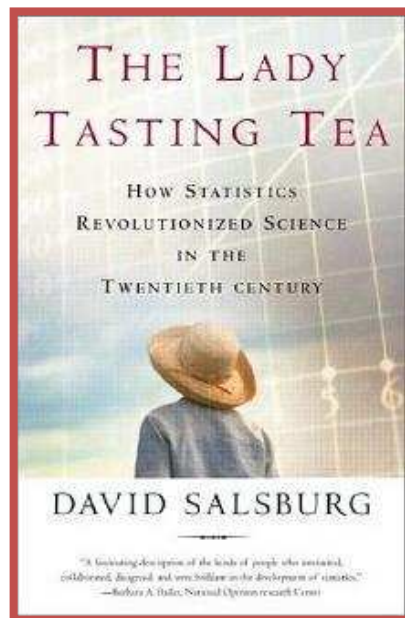
International Conference on Harmonisation (ICH : E10)

Come formare i gruppi da
mettere a confronto ?

The Lady tasting Tea

Una annosa questione che tuttora impegna la società inglese

Quando deve essere aggiunto il latte nella tazza, prima o dopo aver versato il the ?

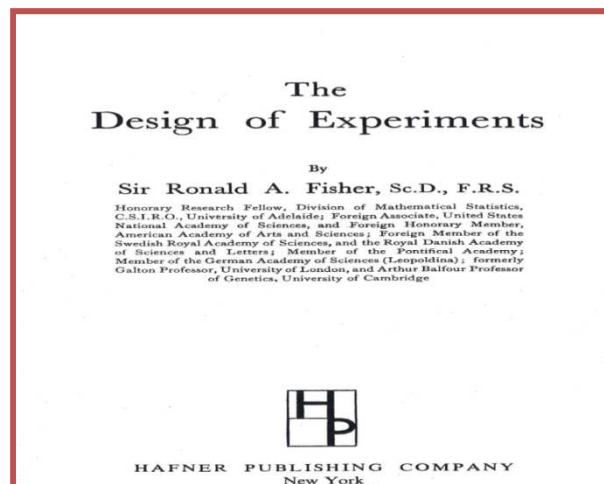


The Lady tasting Tea

Immaginiamo che una gentildonna dichiari di essere in grado, assaggiando una tazza di the con il latte, di discriminare se il latte sia stato aggiunto prima o dopo il the.

Come possiamo verificare questa presunta capacità di "discriminazione"?

Fisher (1935) progetta un "esperimento".



CONTENTS	
I. INTRODUCTION	
1. The Grounds on which Evidence is Disputed	PAGE 1
2. The Mathematical Attitude towards Induction	3
3. The Rejection of Inverse Probability	6
4. The Logic of the Laboratory	7
II. THE PRINCIPLES OF EXPERIMENTATION, ILLUSTRATED BY A PSYCHO-PHYSICAL EXPERIMENT	
5. Statement of Experiment	11
6. Interpretation and its Reasoned Basis	12
7. The Test of Significance	13
8. The Null Hypothesis	15
9. Randomisation; the Physical Basis of the Validity of the Test	17
10. The Effectiveness of Randomisation	19
11. The Sensitiveness of an Experiment. Effects of Enlargement and Repetition	21
12. Qualitative Methods of increasing Sensitiveness	22
12-1. Scientific Inference and Acceptance Procedures	25
III. A HISTORICAL EXPERIMENT ON GROWTH RATE	
13.	27
14. Darwin's Discussion of the Data	27
15. Galton's Method of Interpretation	29
16. Pairing and Grouping	32
17. " Student's " t Test	34
18. Fallacious Use of Statistics	38
19. Manipulation of the Data	40
20. Validity and Randomisation	41
xi	

II

THE PRINCIPLES OF EXPERIMENTATION, ILLUSTRATED BY A PSYCHO-PHYSICAL EXPERIMENT

5. Statement of Experiment

A LADY declares that by tasting a cup of tea made with milk she can discriminate whether the milk or the tea infusion was first added to the cup. We will consider the problem of designing an experiment by means of which this assertion can be tested. For this purpose let us first lay down a simple form of experiment with a view to studying its limitations and its characteristics, both those which appear to be essential to the experimental method, when well developed, and those which are not essential but auxiliary.

Our experiment consists in mixing eight cups of tea, four in one way and four in the other, and presenting them to the subject for judgment in a random order. The subject has been told in advance of what the test will consist, namely that she will be asked to taste eight cups, that these shall be four of each kind, and that they shall be presented to her in a random order, that is in an order not determined arbitrarily by human choice, but by the actual manipulation of the physical apparatus used in games of chance, cards, dice, roulettes, etc., or, more expeditiously, from a published collection of random sampling numbers purporting to give the actual results of such manipulation. Her task is to divide the 8 cups into two sets of 4, agreeing, if possible, with the treatments received.

Randomisation: the physical basis of the validity of the test

It is no sufficient remedy to insist that "all the cups must be exactly alike" in every respect except that to be tested. For this is a totally impossible requirement in our example, and equally in all other forms of experimentation.

The element in the experimental procedure which contains the essential safeguard is that the two modifications of the test beverage are to be prepared "in random order." This, in fact, is the only point in the experimental procedure in which the laws of chance, which are to be in exclusive control of our frequency distribution, have been explicitly introduced. The

Vi sono $70 = \binom{8}{4}$ possibili modalità di preparazione delle 8 tazze organizzate in due gruppi di 4

T= prima il the **L= prima il latte**



T **L** **L** **T** **T** **L** **L** **T**



L **T** **T** **L** **T** **L** **L** **T**



T **L** **L** **T** **L** **T** **L** **T**



Qual è la probabilità di “azzeccare” per caso la corretta sequenza?

Come sono state preparate

Latte The

Le risposte della Lady

“Latte”	4	0
“The”	0	4

La probabilità è 1/70 (0,014)

E ammettendo un errore di identificazione?

Come sono state preparate

Latte The

Le risposte della Lady

“Latte”	3	1
“The”	1	3

Vi sono $4 \times 4 = 16$ modi di commettere per errore uno scambio.
La probabilità di “azzeccare” per caso diventa $16/70$ ($p=0,229$)

Complessivamente la probabilità è 0,24 ($1/70 + 16/70 = 17/70$)

Come sono state preparate

Le risposte della Lady

a	b	a+b
c	d	c+d
a+c	b+d	n

Fisher dimostrò che la probabilità di ottenere per effetto del caso uno specifico set di valori è data da una distribuzione ipergeometrica:

$$p = \frac{\binom{a+b}{a} \binom{c+d}{c}}{\binom{n}{a+c}} = \frac{(a+b)! (c+d)! (a+c)! (b+d)!}{a! b! c! d! n!}$$

Il modello di Neyman-Rubin

esiti potenziali

controfattualità

Le radici del modello risalgono agli anni 1920-1930 grazie ai contributi di Ronald A. **Fisher** (1890-1962) e di Jerzy **Neyman** (1894-1981) nell'ambito degli studi sperimentali.

Il modello è stato poi espanso da Donald **Rubin** negli anni '70 per ricomprendere il caso degli studi osservazionali.

Neyman describe gli effetti causati dai trattamenti come confronto di esiti potenziali (potential outcomes) sotto trattamenti alternativi.

Statistical Science
1990, Vol. 5, No. 4, 465-480

On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9.

Jerzy Splawa-Neyman

Translated and edited by D. M. Dabrowska and T. P. Speed from the Polish original, which appeared in *Roczniki Nauk Rolniczych Tom X (1923) 1-51 (Annals of Agricultural Sciences)*

1923

Potential Outcomes (esiti potenziali): I valori assunti da una misura di interesse dopo l'applicazione del trattamento e la non applicazione del trattamento alla medesima unità

ESTIMATING CAUSAL EFFECTS OF TREATMENTS IN RANDOMIZED AND NONRANDOMIZED STUDIES¹

DONALD B. RUBIN²

Educational Testing Service, Princeton, New Jersey

A discussion of matching, randomization, random sampling, and other methods of controlling extraneous variation is presented. The objective is to specify the benefits of randomization in estimating causal effects of treatments. The basic conclusion is that randomization should be employed whenever possible but that the use of carefully controlled nonrandomized data to estimate causal effects is a reasonable and necessary procedure in many cases.

Let $y(E)$ be the value of Y measured⁵ at t_2 on the unit, given that the unit received the experimental Treatment E initiated at t_1 ;

Let $y(C)$ be the value of Y measured at t_2 on the unit given that the unit received the control Treatment C initiated at t_1 ;

Then $y(E) - y(C)$ is the causal effect of the E versus C treatment on Y for that trial, that is, for that particular unit and the times t_1, t_2 .

**Rubin describe il suo
modello**

1974

La struttura del modello: $\{U_i, T, Y, W\}$

Unità (U_i)	: una persona, cosa, luogo sul quale opererà un trattamento ad un particolare istante
Trattamento (T)	: un intervento, l'effetto del quale (attraverso una specifica misura sulle unità) il ricercatore vuole valutare in relazione al non-intervento (controllo)
Esiti potenziali (Y)	: I valori di una misura di interesse dopo l'applicazione del trattamento $Y(1)$ e la non applicazione del trattamento $Y(0)$ alla medesima unità
Assignment Mechanism (W)	: processo attraverso il quale si decide quali unità ricevono il trattamento e quali ricevono il controllo



Causal Effect (effetto causale) Per ciascuna unità U_i il confronto dell'esito potenziale in due situazioni: con il trattamento e senza il trattamento (controllo).

$$Y_i(1) - Y_i(0)$$

Quando:

- a) non vi è interferenza tra le unità
- b) vi è un solo tipo di trattamento ed un solo tipo di controllo.

l'insieme completo degli esiti potenziali in una popolazione di N unità può essere rappresentato in una matrice del tipo:

Unità	Covariate	Trattamento Y(1)	Controllo Y(0)
1	X_1	$Y_1(1)$	$Y_1(0)$
.	.	.	.
.	.	.	.
J	X_j	$Y_j(1)$	$Y_j(0)$
.	.	.	.
N	X_N	$Y_N(1)$	$Y_N(0)$

Lo stato di trattamento di una qualsiasi unità non influenza gli esiti potenziali delle altre unità.

Stable unit treatment value assumption (SUTVA)

Quale meccanismo (W) di assegnazione del trattamento alle unità?

E' il processo secondo il quale le unità sono assegnate al trattamento e al controllo.

W fornisce la probabilità di ciascuna sequenza di assegnazione:

$$P_r (W | Y(0), Y(1), X)$$

Quando l'assegnazione è **Random** la probabilità di ricevere il trattamento (*propensity score*) è uguale per ogni unità e non dipende dal valore di qualche esito potenziale non osservato:

$$P_r (W | Y(0), Y(1), X) = P_r (W | X)$$

Statistics and Causal Inference

PAUL W. HOLLAND*

© 1986 American Statistical Association
Journal of the American Statistical Association
December 1986, Vol. 81, No. 396, Theory and Methods

Fundamental Problem of Causal Inference. It is impossible to *observe* the value of $Y_t(u)$ and $Y_c(u)$ on the same unit and, therefore, it is impossible to *observe* the effect of t on u .

Unità	Covariate	Trattamento $Y(1)$	Controllo $Y(0)$
1	X_1	$Y_1(1)$?
2	X_2	?	$Y_2(0)$
3	X_3	?	$Y_3(0)$
4	X_4	$Y_4(1)$?
5	X_5	?	$Y_5(0)$
6	X_6	$Y_6(1)$?

$$\bar{Y}_1 - \bar{Y}_0 = \sum_{i=1}^N \frac{Y_i(1) - Y_i(0)}{N}$$

Lo sviluppo di uno studio randomizzato

- 1- Una dimostrazione per assurdo attraverso l'ipotesi Nulla (H_0 : assenza di effetto).
- 2- Le unità sono assegnate random al trattamento e al controllo, generando una delle possibili assegnazioni, tutte equiprobabili. $\binom{N}{N/2}$
- 3- Si osservano le risposte al trattamento e si calcola l'effetto causale *osservato*
- 4- Per ciascuna possibile assegnazione si calcola l'effetto causale che si sarebbe potuto osservare sotto questa assegnazione
- 5- Si determina quanto raro è il valore dell'effetto causale osservato al passo 3 (livello di significatività)

Unità	Trattamento Y(1)	Controllo Y(0)
1	?	55.0
2	?	72.0
3	?	72.7
4	70.0	?
5	66.0	?
6	78.9	?

$$\text{Effetto osservato} = \frac{7 + 6 + 7}{3} - \frac{5 + 7 + 7}{3} = 5$$

1

Si derivano gli esiti *potenziali* non osservati usando l'ipotesi nulla e gli esiti osservati ($Y(1) = Y(0) = Y_{obs}$)

Unità	Trattamento Y(1)	Controllo Y(0)
1	55.0	55.0
2	72.0	72.0
3	72.7	72.7
4	70.0	70.0
5	66.0	66.0
6	78.9	78.9

La sequenza di randomizzazione osservata è una delle 20
potenzialmente osservabili per 6 unità

$$\binom{6}{3}$$

Per ciascuna sequenza si calcola l'effetto causale
che si sarebbe potuto osservare

Sequenze di randomizzazione (S)	Prob. Di S	Y _m (1)-Y _m (0)
111000	1/20	-5.1
110100	1/20	-6.9
110010	1/20	-9.5
110001	1/20	-0.9
101100	1/20	-6.4
101010	1/20	-9.1
101001	1/20	-0.5
100110	1/20	-10.9
100101	1/20	-2.3
100011	1/20	-4.9
011100	1/20	4.9
011010	1/20	2.3
011001	1/20	10.9
010110	1/20	0.5
010101	1/20	9.1
010011	1/20	6.4
001110	1/20	0.9
001101	1/20	9.5
001011	1/20	6.9
000111	1/20	5.1

$$= \frac{55.0 + 72.0 + 72.7}{3} - \frac{70.0 + 66.0 + 78.9}{3}$$

$$= \frac{55 + 72.7 + 78.9}{3} - \frac{72 + 70 + 66}{3}$$

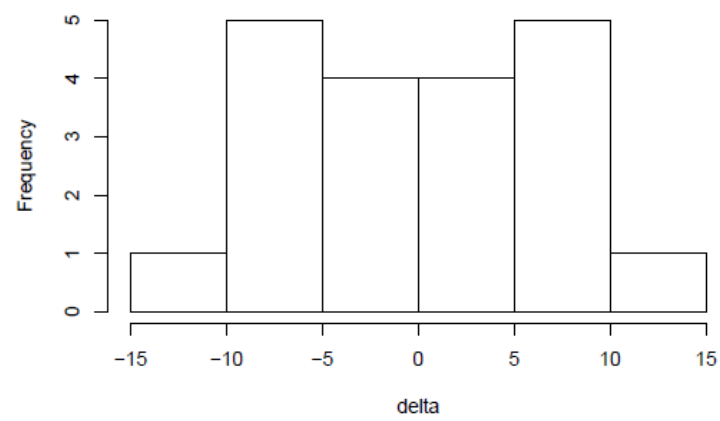
$$= \frac{72 + 72.7 + 78.9}{3} - \frac{55 + 70 + 66}{3}$$

$$= \frac{70.0 + 66.0 + 78.9}{3} - \frac{55.0 + 72.0 + 72.7}{3}$$

1= trattato
0=controllo

Sequenze di randomizzazione (S)	Prob. Di S	$Y_m(1)-Y_m(0)$
011001	1/20	10.9
001101	1/20	9.5
010101	1/20	9.1
001011	1/20	6.9
010011	1/20	6.4
000111	1/20	5.1
011100	1/20	4.9
011010	1/20	2.3
001110	1/20	0.9
010110	1/20	0.5
101001	1/20	-0.5
110001	1/20	-0.9
100101	1/20	-2.3
100011	1/20	-4.9
111000	1/20	-5.1
101100	1/20	-6.4
110100	1/20	-6.9
101010	1/20	-9.1
110010	1/20	-9.5
100110	1/20	-10.9

} 6/20=0.3



Prob (dati_{osservati} | H₀)

~~Prob (H₀ | dati_{osservati})~~

Meccanismi di assegnazione

Without a model for how treatments get assigned to units, formal causal inference, as least using probabilistic statements, is impossible.

Donald B. Rubin

<i>Units</i>	<i>Covariates</i> X	<u><i>Potential outcomes</i></u>		<i>Unit-level</i> <i>Causal effects</i>	<i>Summary</i> <i>Causal effects</i>
		<i>Treatment</i> $Y(1)$	<i>Control</i> $Y(0)$		
1	X_1	$Y_1(1)$	$Y_1(0)$	$Y_1(1)$ v. $Y_1(0)$	Comparison of $Y_i(1)$ v. $Y_i(0)$ for a common set of units
⋮	⋮	⋮	⋮	⋮	
i	X_i	$Y_i(1)$	$Y_i(0)$	$Y_i(1)$ v. $Y_i(0)$	
⋮	⋮	⋮	⋮	⋮	
N	X_N	$Y_N(1)$	$Y_N(0)$	$Y_N(1)$ v. $Y_N(0)$	

Figure 1. “Science”—The Causal Estimand.

Randomizzazione

La randomizzazione introduce deliberatamente un elemento di casualità nell'assegnazione del trattamento ai soggetti di una sperimentazione clinica.

Durante la successiva analisi dei dati fornisce una base statistica solida per la valutazione quantitativa della evidenza relativa agli effetti del trattamento. Essa inoltre tende a costituire gruppi di trattamento nei quali la distribuzione dei fattori prognostici, noti e non noti, sono simili.

International Conference on Harmonisation (ICH : E9)

Tipi di randomizzazione

Semplice

Sequenza singola di assegnazioni casuali.

Ristretta

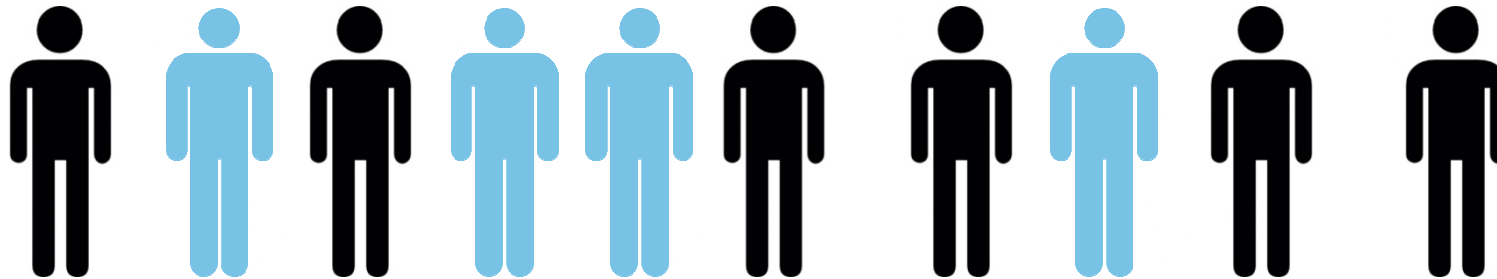
Tramite procedure controllate per raggiungere il bilanciamento tra gruppi in termini di dimensioni o di altre caratteristiche

- * a blocchi (ad esempio di 4 pazienti)
- * stratificata (per età, gravità, etc)

Randomizzazione semplice

Una sequenza randomizzata tra le possibili 2^N (disposizioni con ripetizioni)

[0 1 0 1 1 0 0 1 0 0]



1= trattato 0=controllo

Possibili sbilanciamenti

Randomizzazione a blocchi

Assegniamo i trattamenti non ad un singolo paziente ma ad un “**blocco**” di **N** pazienti consecutivi.

In ogni blocco **N/2** pazienti sono assegnati al trattamento e **N/2** al controllo.

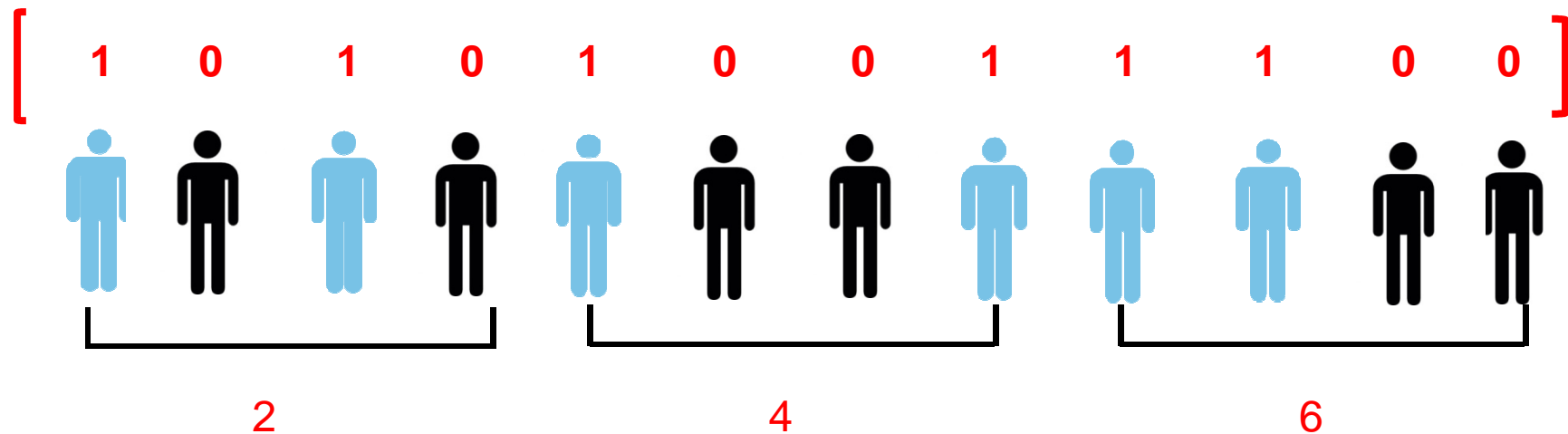
Nel caso ad esempio di $N=4$ si avranno 6 possibili configurazioni (permutazioni con ripetizione).

$$\frac{N!}{N_1!N_2!\dots N_i!} = \frac{4!}{2!2!} = 6$$

Le possibili configurazioni sono:

- 1- TTCC
- 2- TCTC
- 3- TCCT
- 4- CTTC
- 5- CTCT
- 6- CCTT

Estraendo, ad esempio, casualmente i blocchi 2, 4, 6 si avrà la seguente sequenza:



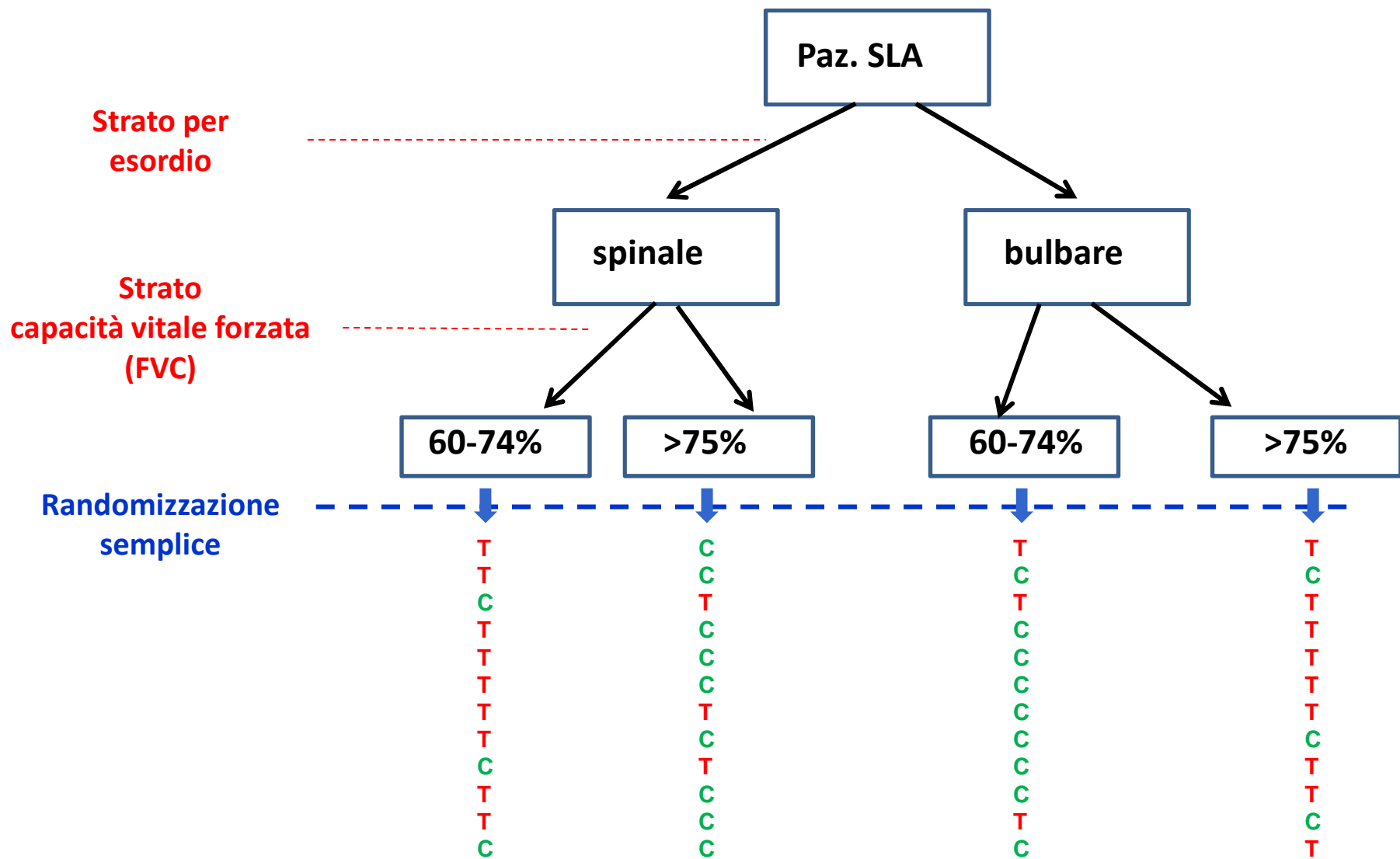
1= trattato 0=controllo

Utile nei trial multicentrici.

Maggiore predicibilità con blocchi di piccole dimensioni.

Cecità sulla dimensione del blocco.

Randomizzazione stratificata



Randomizzazione stratificata

