SACRO CUORE DON CALABRIA
I R C C S

MEMBER OF
ECI
Organisation of European Cancer Institutes - EEIG

SCUOLA DI METODOLOGIA DELLA RICERCA CLINICA

2023 - 9ª EDIZIONE

MODULI SPECIALISTICI - S4

NEGRAR DI VALPOLICELLA • 12-13 MAGGIO 2023

Centro Formazione IRCCS "Sacro Cuore - Don Calabria"

Classificazione degli studi osservazionali: descrittivi *Vs* analitici

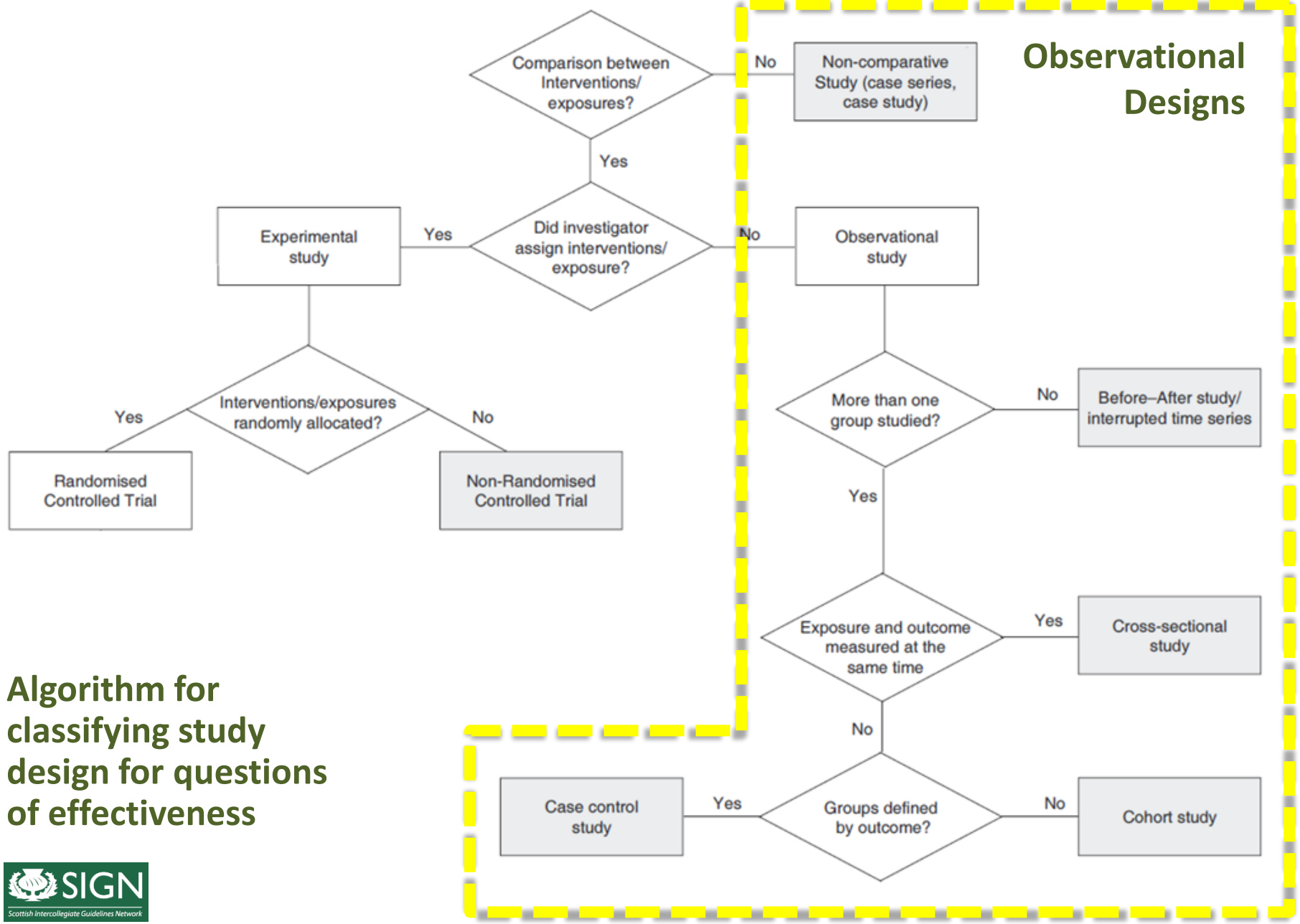
(G.L. Pappagallo)

OBSERVATIONAL STUDY: A DEFINITION

An observational study draws inferences from a sample to a population where the independent variable is **not under the control** of the researcher.

The term observational study covers a wide range of study designs, a common feature of which is that they are noninterventional, in the sense that the **study protocol does not determine the** precise **features of any therapy** given to the participants in the study.

Observational Designs



Algorithm for classifying study design for questions of effectiveness

SIGN
Scottish Intercollegiate Guidelines Network

Observational Studies

“EPIDEMIOLOGIC” Vs “THERAPEUTIC”

Before-After

Surveys disease status before and after an intervention

Cross-Sectional

Provide information on prevalence of a particular condition at a single time point (time window)

Case-Control

Identify predictors of a particular outcome

Cohort

Identify the incidence of a particular outcome over time

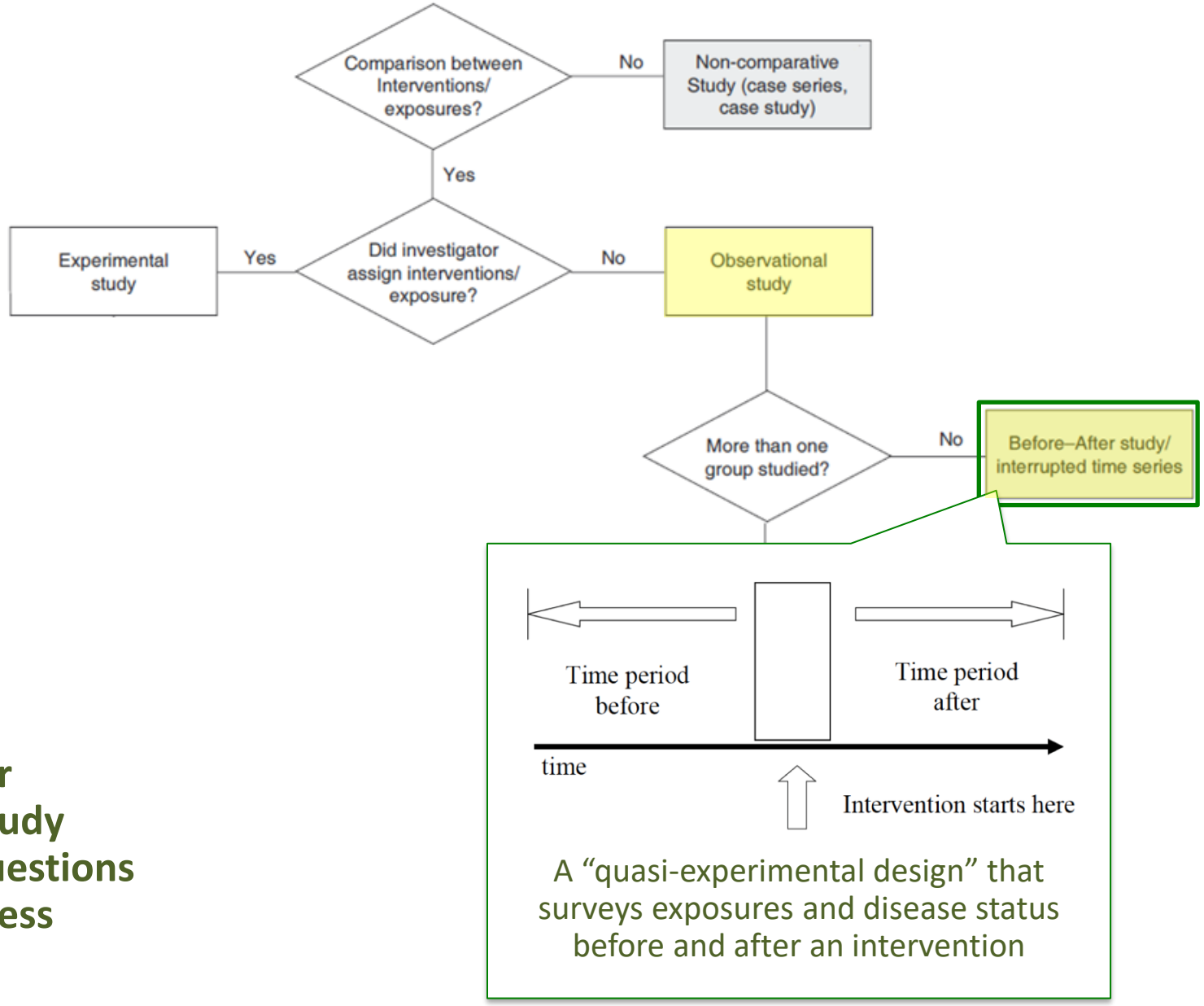
Non-comparative case series

Report outcomes of patients who received a specific intervention

Comparative case series

Compare outcomes between patients who received different interventions

Algorithm for classifying study design for questions of effectiveness



A "quasi-experimental design" that surveys exposures and disease status before and after an intervention

Observational Studies

“EPIDEMIOLOGIC” Vs “THERAPEUTIC”

Before-After

Surveys disease status before and after an intervention

Cross-Sectional

Provide information on prevalence of a particular condition at a single time point (time window)

Case-Control

Identify predictors of a particular outcome

Cohort

Identify the incidence of a particular outcome over time

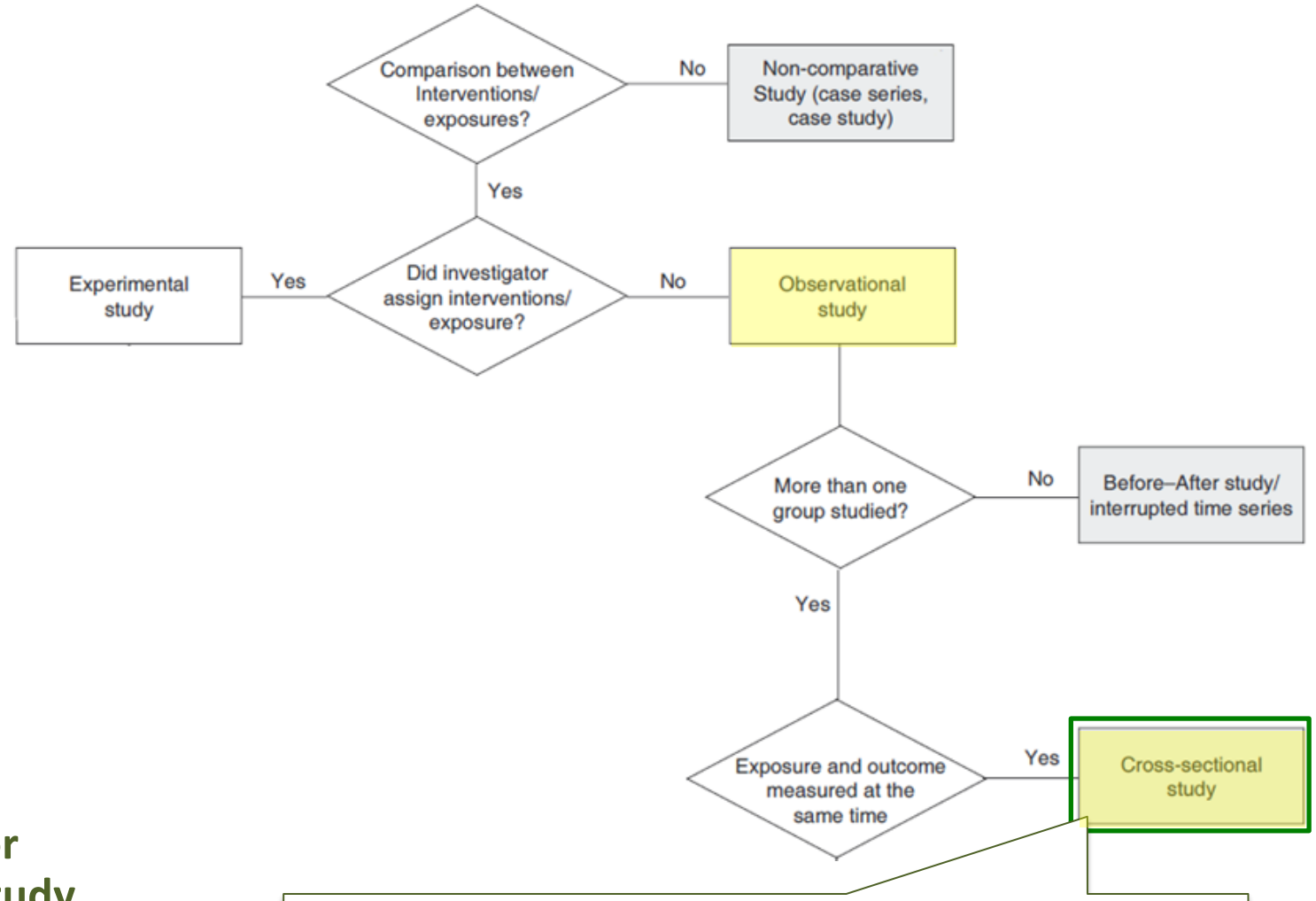
Non-comparative case series

Report outcomes of patients who received a specific intervention

Comparative case series

Compare outcomes between patients who received different interventions

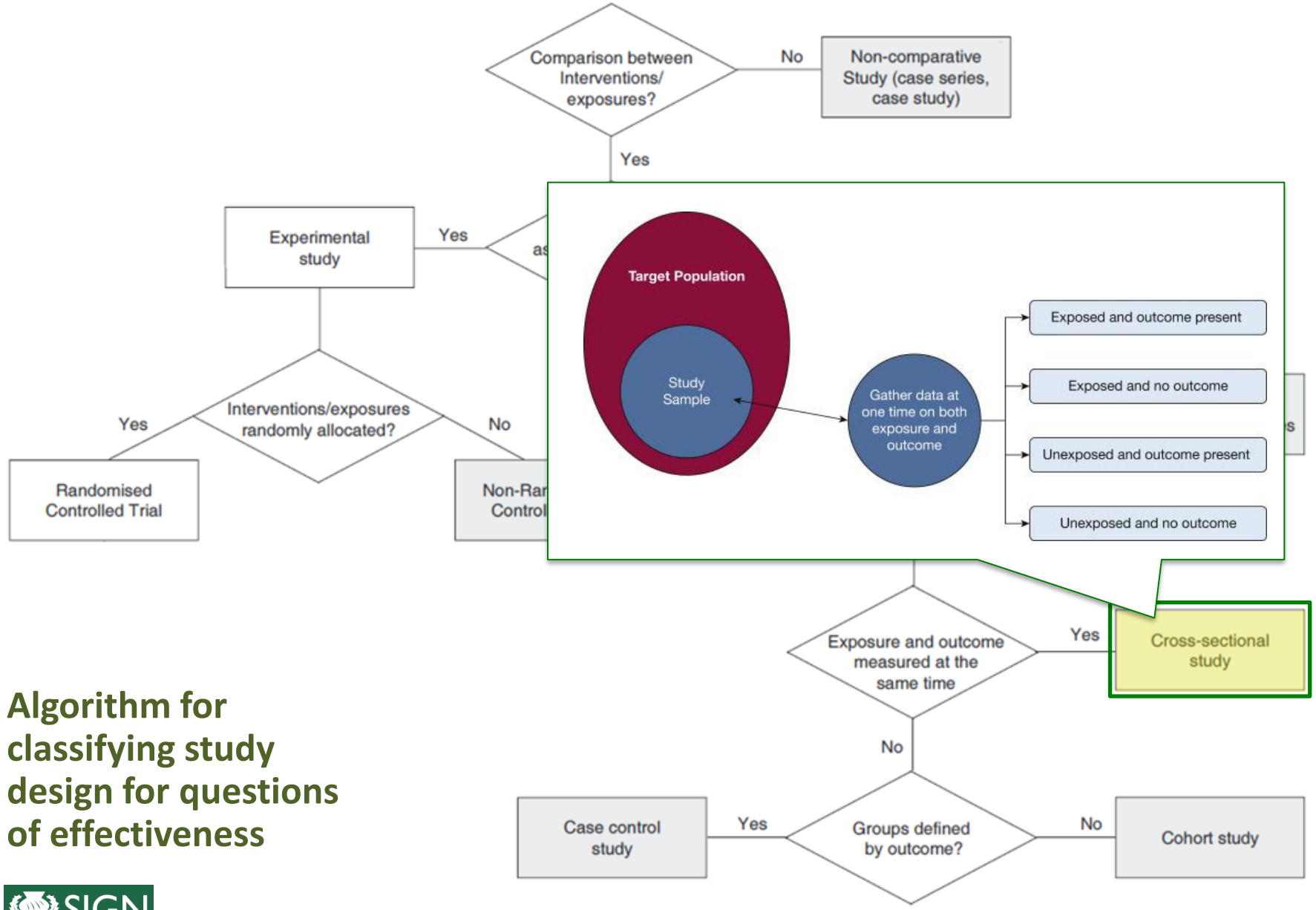
Algorithm for classifying study design for questions of effectiveness



Subjects selected irrespective of the presence or absence of the characteristics of interest. Similar to a case series, except that the purpose of the analysis is to record associations between variables, rather than merely to report frequencies of their occurrence

SIGN
Scottish Intercollegiate Guidelines Network

Algorithm for classifying study design for questions of effectiveness



SIGN
Scottish Intercollegiate Guidelines Network

Observational Studies

"EPIDEMIOLOGIC" Vs "THERAPEUTIC"

Before-After

Surveys disease status before and after an intervention

Cross-Sectional

Provide information on prevalence of a particular condition at a single time point (time window)

Case-Control

Identify predictors of a particular outcome

Cohort

Identify the incidence of a particular outcome over time

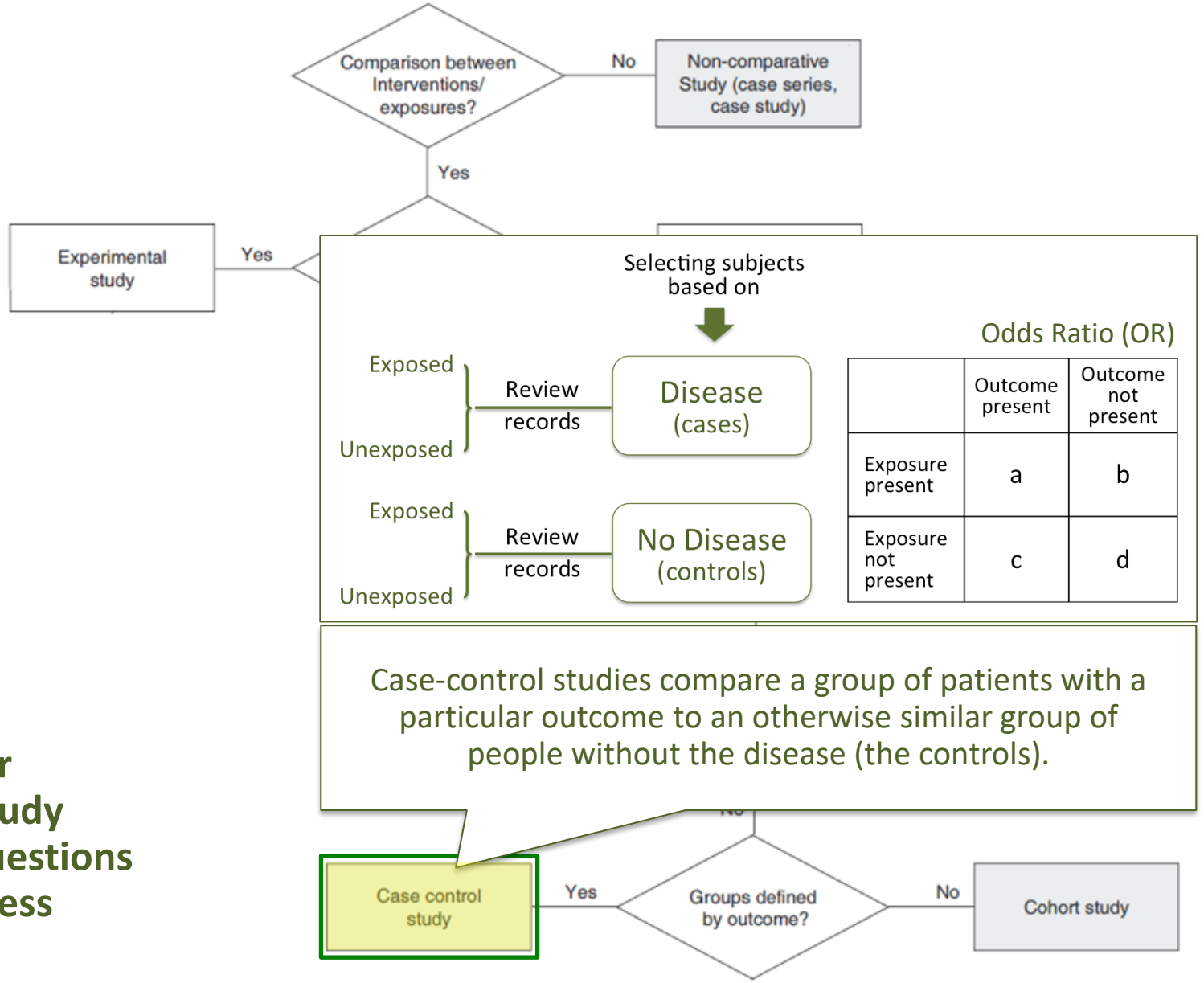
Non-comparative case series

Report outcomes of patients who received a specific intervention

Comparative case series

Compare outcomes between patients who received different interventions

Algorithm for classifying study design for questions of effectiveness



Observational Studies

“EPIDEMIOLOGIC” Vs “THERAPEUTIC”

Before-After

Surveys disease status before and after an intervention

Cross-Sectional

Provide information on prevalence of a particular condition at a single time point (time window)

Case-Control

Identify predictors of a particular outcome

Cohort

Identify the incidence of a particular outcome over time

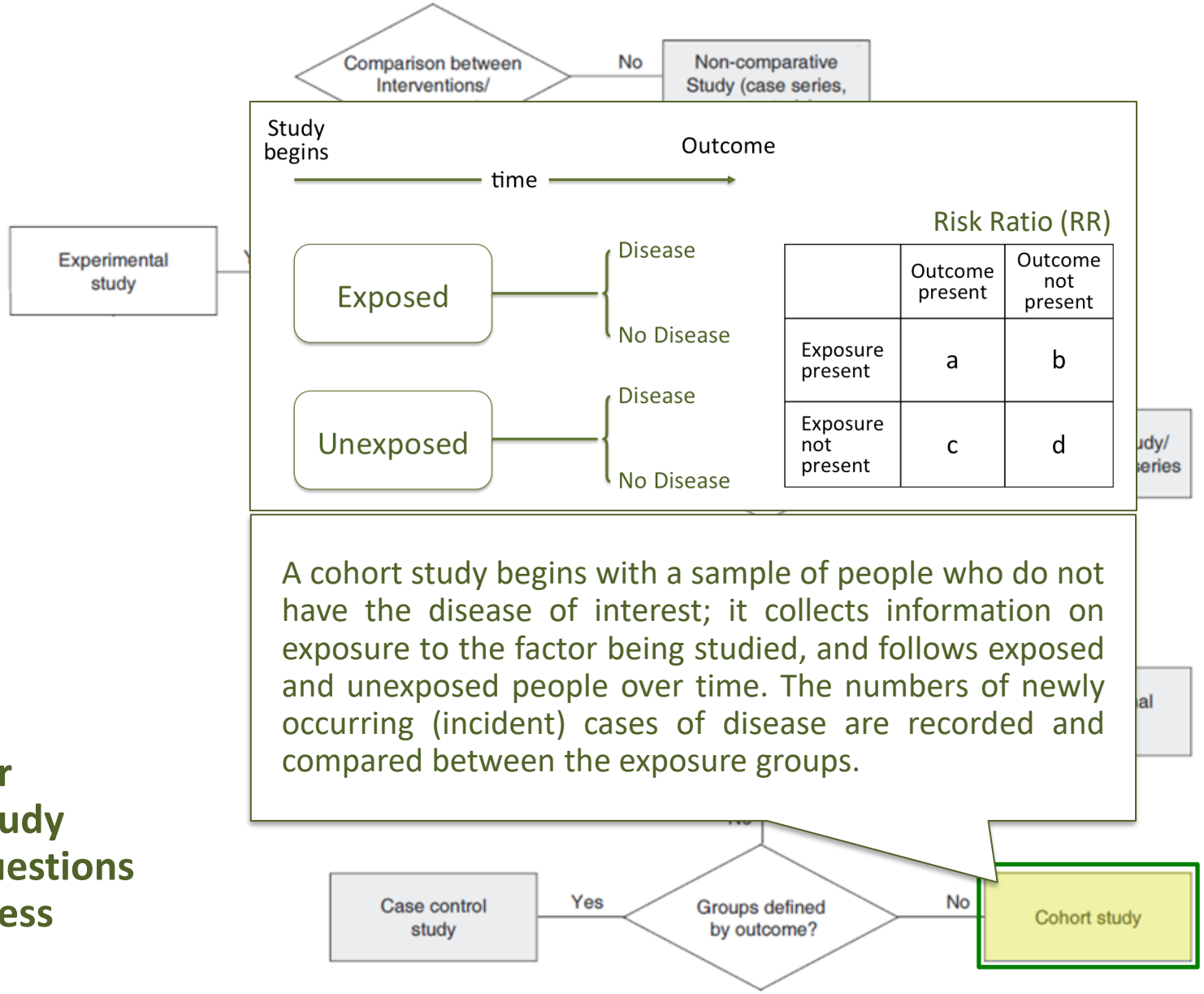
Non-comparative case series

Report outcomes of patients who received a specific intervention

Comparative case series

Compare outcomes between patients who received different interventions

Algorithm for classifying study design for questions of effectiveness



Time matters…

Cross-Sectional Studies
(exposure and outcome measured at the same time)

Case-Control Studies
(groups defined by the outcome)

Cohort Studies
(groups not defined by the outcome)

Observational Studies

“EPIDEMIOLOGIC” Vs “THERAPEUTIC”

Before-After

Surveys disease status before and after an intervention

Cross-Sectional

Provide information on prevalence of a particular condition at a single time point (time window)

Case-Control

Identify predictors of a particular outcome

Cohort

Identify the incidence of a particular outcome over time

Non-comparative case series

Report outcomes of patients who received a specific intervention

Comparative case series

Compare outcomes between patients who received different interventions

The Value of Observational Cohort Studies for Cancer Drugs

Randomized controlled trials — the gold standard for clinical drug evaluation — can't always predict adverse events in real-world settings. For the new cancer therapies, observational cohort studies (OCSs) can help evaluate their effects in broader populations and provide valuable information for future clinical trials.

BY DAVID R. SPIGEL, MD **BIOTECHNOLOGY HEALTHCARE** · SUMMER 2010

WHAT IS AN OCS?

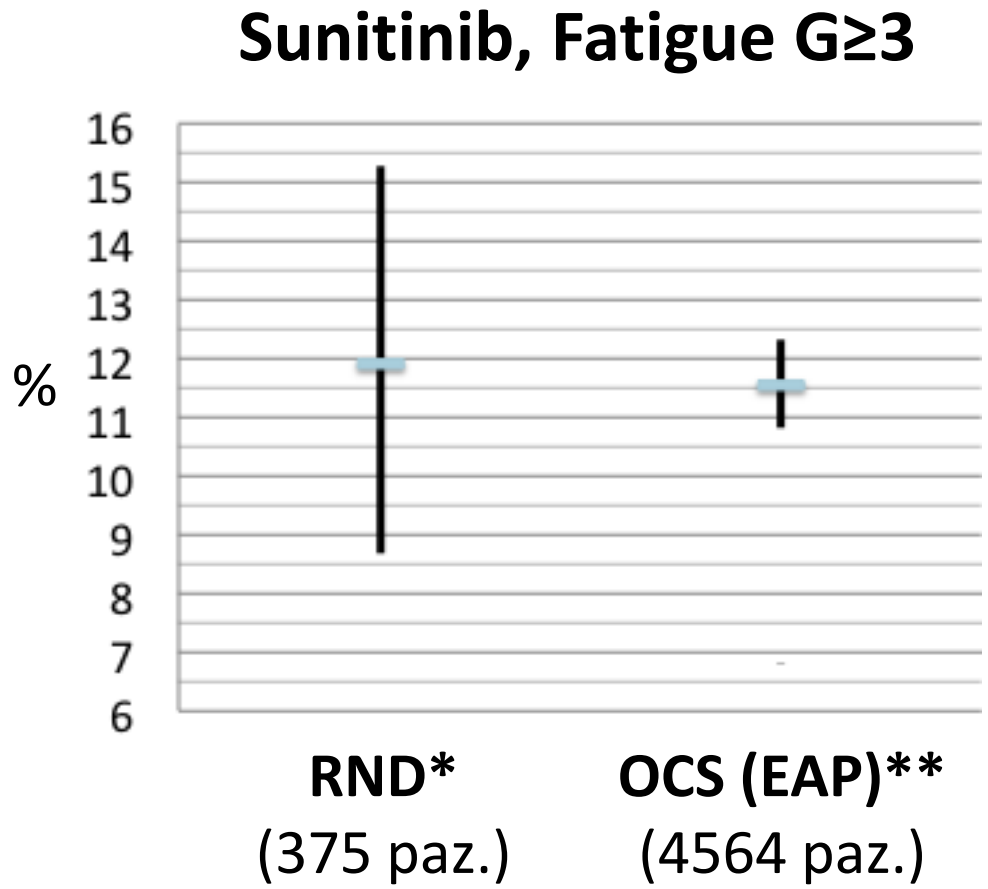
An OCS is an analysis of a group of individuals who have specific features in common and who are followed over a defined period of time.

Prospective OCSs are designed to examine predefined primary outcomes.

Post-approval OCSs generally follow a single cohort, although patient subgroups may be analyzed separately.

To represent a broad and diverse patient base and to detect rare adverse events, large community-based, multicenter OCSs are useful in the post-approval setting for new therapeutics.

Studio RND registrativo vs OCS (EAP)



Quale dei due studi è più UTILE per la Clinica?

* Motzer, NEJM 2007; ** Gore, Lancet Oncol 2009

Observational Studies

“EPIDEMIOLOGIC” Vs “THERAPEUTIC”

Before-After

Surveys disease status before and after an intervention

Cross-Sectional

Provide information on prevalence of a particular condition at a single time point (time window)

Case-Control

Identify predictors of a particular outcome

Cohort

Identify the incidence of a particular outcome over time

Non-comparative case series

Report outcomes of patients who received a specific intervention

Comparative case series

Compare outcomes between patients who received different interventions

Effectiveness Versus Efficacy: More Than a Debate Over Language

Julie M. Fritz, PT, PhD, ATC[1] *Joshua Cleland, PT, DPT, OCS*[2]

To some, the best evidence may be viewed as research that minimizes bias to the greatest extent possible, while others may prioritize research that is deemed most pertinent to clinical practice.

Qualità = eliminare (ridurre) le fonti di bias?

Le tre regole d'oro della sperimentazione clinica

Randomize!

Randomize!

Randomize!

Ratain MJ, ASCO oral presentation, 2006

Effectiveness Versus Efficacy: More Than a Debate Over Language

Julie M. Fritz, PT, PhD, ATC[1] *Joshua Cleland, PT, DPT, OCS*[2]

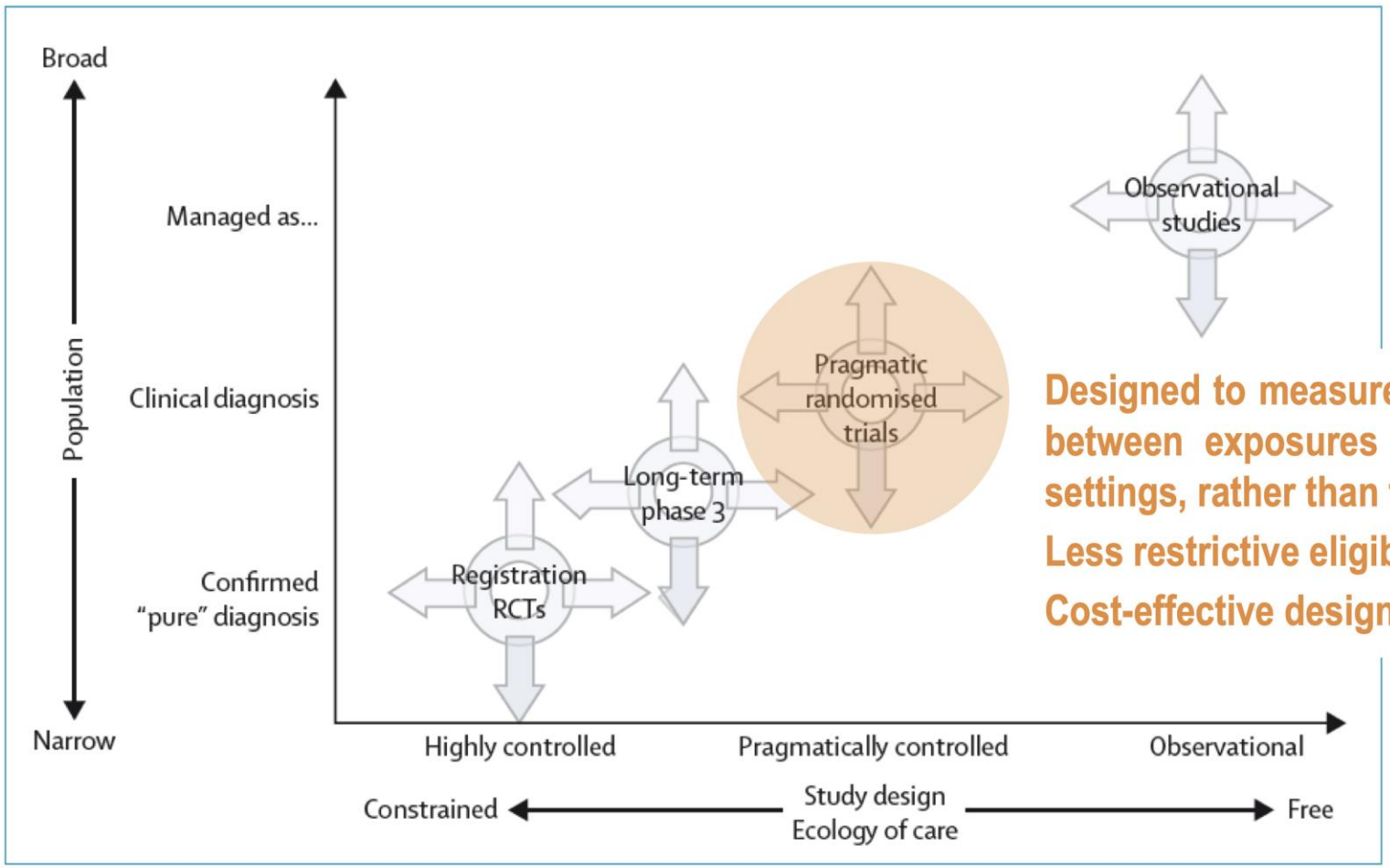
To some, the best evidence may be viewed as research that minimizes bias to the greatest extent possible, while others may prioritize research that is deemed most pertinent to clinical practice.

Qualità = trasferibilità alla pratica clinica?

Integrating real-life studies in the global therapeutic research framework

*Nicolas Roche, Helen K Reddel, Alvar Agusti, Eric D Bateman, Jerry A Krishnan, Richard J Martin, Alberto Papi, Dirkje Postma, Mike Thomas, Guy Brusselle, Elliot Israel, Cynthia Rand, Alison Chisholm, David Price, on behalf of the Respiratory Effectiveness Group

www.thelancet.com/respiratory Vol 1 December 2013



Designed to measure the strength of associations between exposures and outcomes in real-world settings, rather than to prove causal relationships.

Less restrictive eligibility criteria

Cost-effective design

Figure 1: A conceptual framework for therapeutic research

Integrating real-life studies in the global therapeutic research framework

*Nicolas Roche, Helen K Reddel, Alvar Agusti, Eric D Bateman, Jerry A Krishnan, Richard J Martin, Alberto Papi, Dirkje Postma, Mike Thomas, Guy Brusselle, Elliot Israel, Cynthia Rand, Alison Chisholm, David Price, on behalf of the Respiratory Effectiveness Group

www.thelancet.com/respiratory Vol 1 December 2013

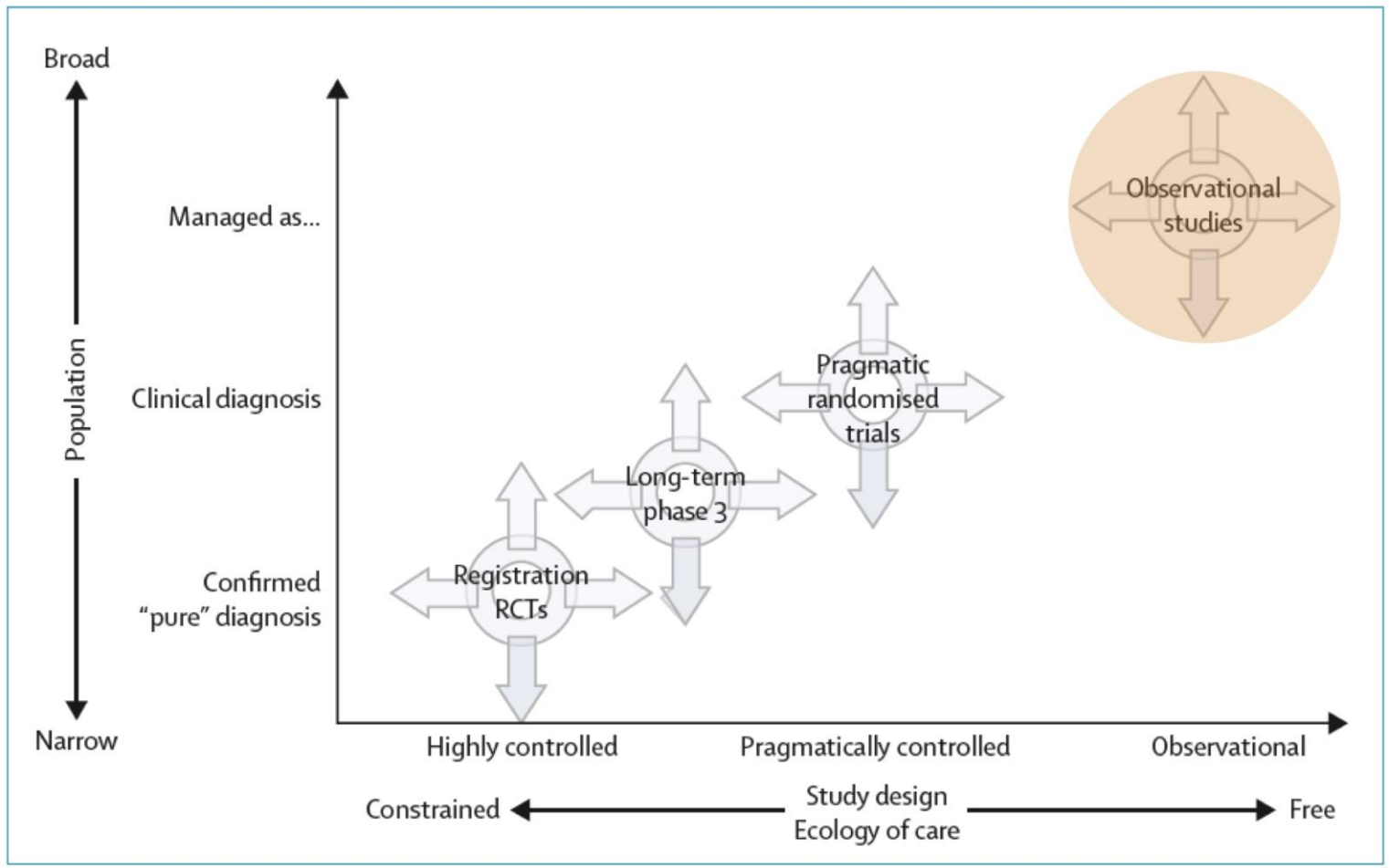
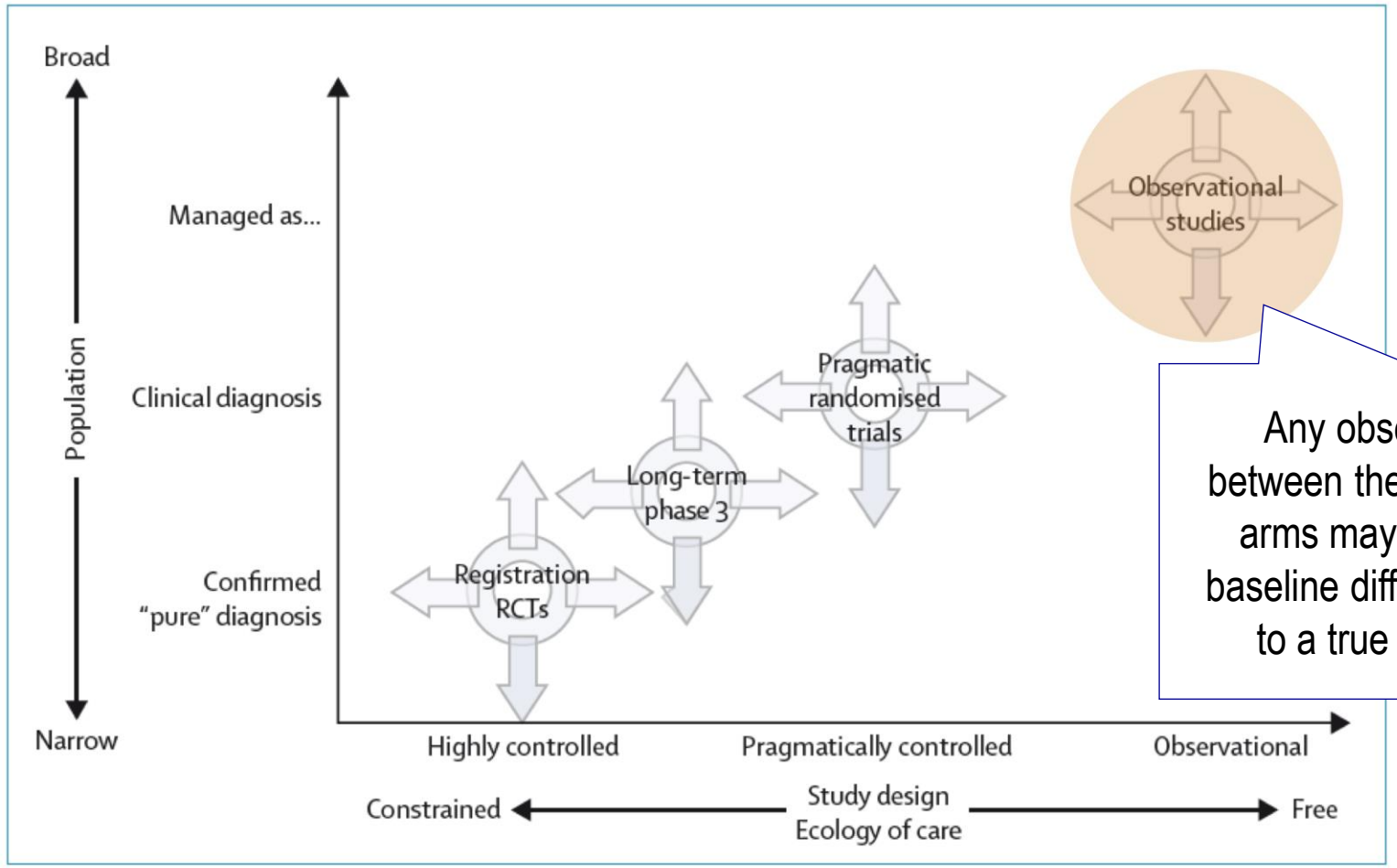


Figure 1: A conceptual framework for therapeutic research

Integrating real-life studies in the global therapeutic research framework

*Nicolas Roche, Helen K Reddel, Alvar Agusti, Eric D Bateman, Jerry A Krishnan, Richard J Martin, Alberto Papi, Dirkje Postma, Mike Thomas, Guy Brusselle, Elliot Israel, Cynthia Rand, Alison Chisholm, David Price, on behalf of the Respiratory Effectiveness Group

www.thelancet.com/respiratory Vol 1 December 2013



Selection Bias
Confounding
Missing Data
Measurement Bias
Inflated Type1 error

Any observed difference between the outcomes of study arms may be attributable to baseline differences rather than to a true treatment effect.

Potential Solutions:

Careful Study Design
Appropriate Statistical Methods

Figure 1: A conceptual framework for therapeutic research

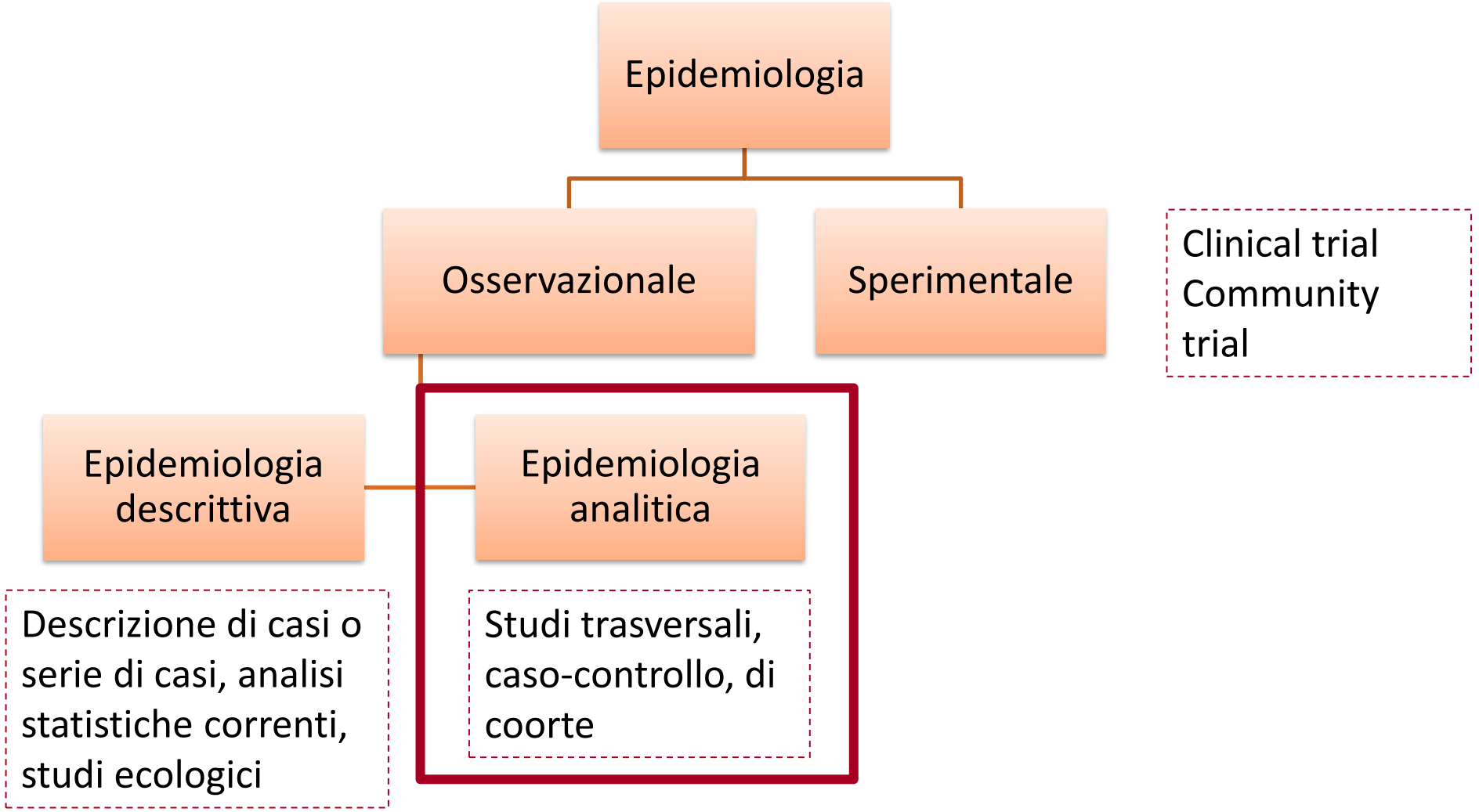
Studi trasversali

(C. Bosetti)

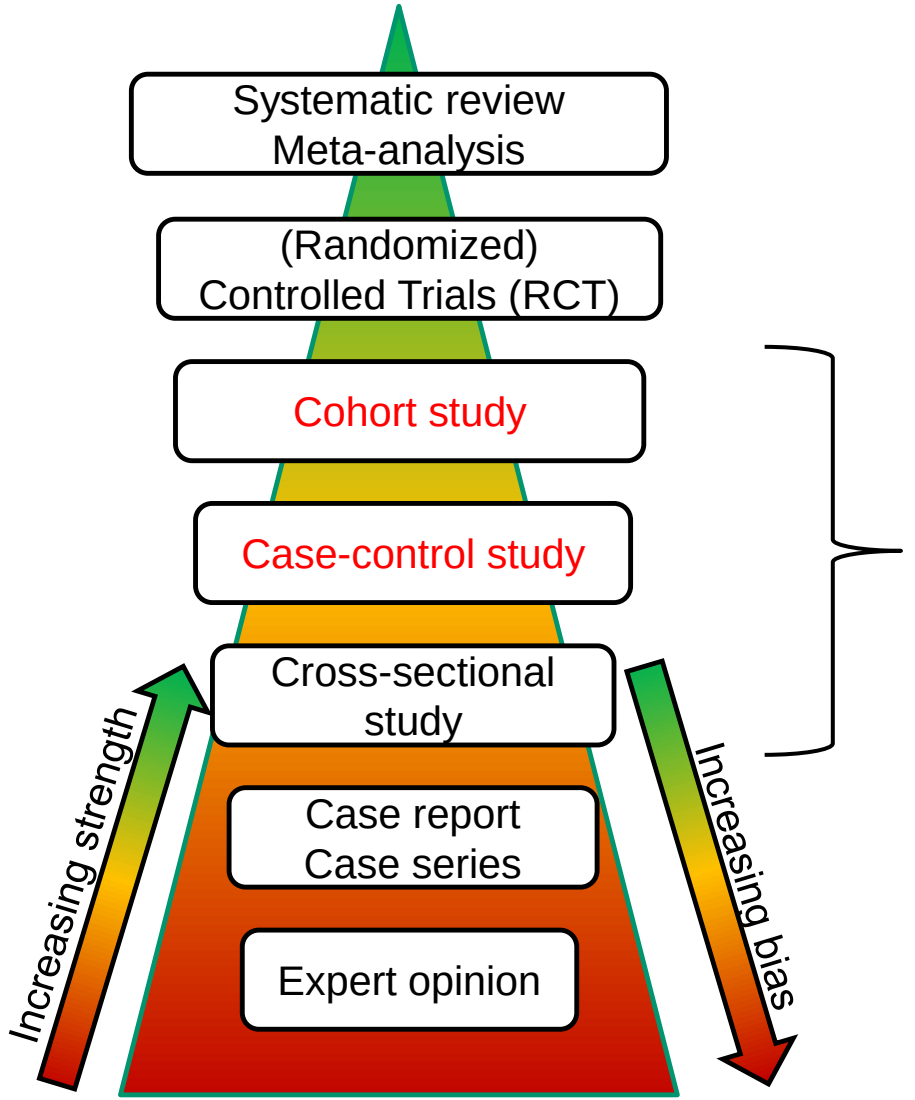
Outline

- Definizione
- Metodologia di pianificazione, conduzione e analisi
- Punti di forza/debolezza

Classificazione degli studi



Gerarchia degli studi clinici



Studi osservazionali analitici

STUDI TRASVERSALI

Definizione

STUDI TRASVERSALI

- Gli studi trasversali (*o cross-sectional o survey*) esaminano una **popolazione definita** in un **preciso istante temporale**, in cui si rileva la presenza di una o più condizioni o eventi, quali ad esempio lo stato di malattia, l'esposizione ad un particolare fattore di rischio
- Possono essere considerati come una **«fotografia»** istantanea di un gruppo di individui

Relazione del tempo e gli studi analitici

STUDI TRASVERSALI: FINALITÀ

Possono avere finalità solo descrittive o anche analitiche e possono consentire di stimare, in una popolazione:

- la **prevalenza** di un fenomeno (mallattia o fattore di rischio)
- le **associazioni** fra fattori di esposizione e condizioni di salute o di malattia, e anche le associazioni di patologie tra di loro (co-morbosità) o di fattori di rischio o protettivi

Studi trasversali: esempi di settori di ricerca

- Prevalenza delle malattie croniche disabilitanti
- Consumi di tabacco, alcol, droghe e farmaci
- Consumi alimentari e fabbisogni dietetici in relazione ai consumi
- Differenze nello stato di salute
- Valutazione della qualità della vita
- Valutazione dei servizi socio-assistenziali
- Giornate di inabilità lavorativa temporanea e permanente
- Salute mentale
- Indici di crescita dei neonati e modalità di allattamento

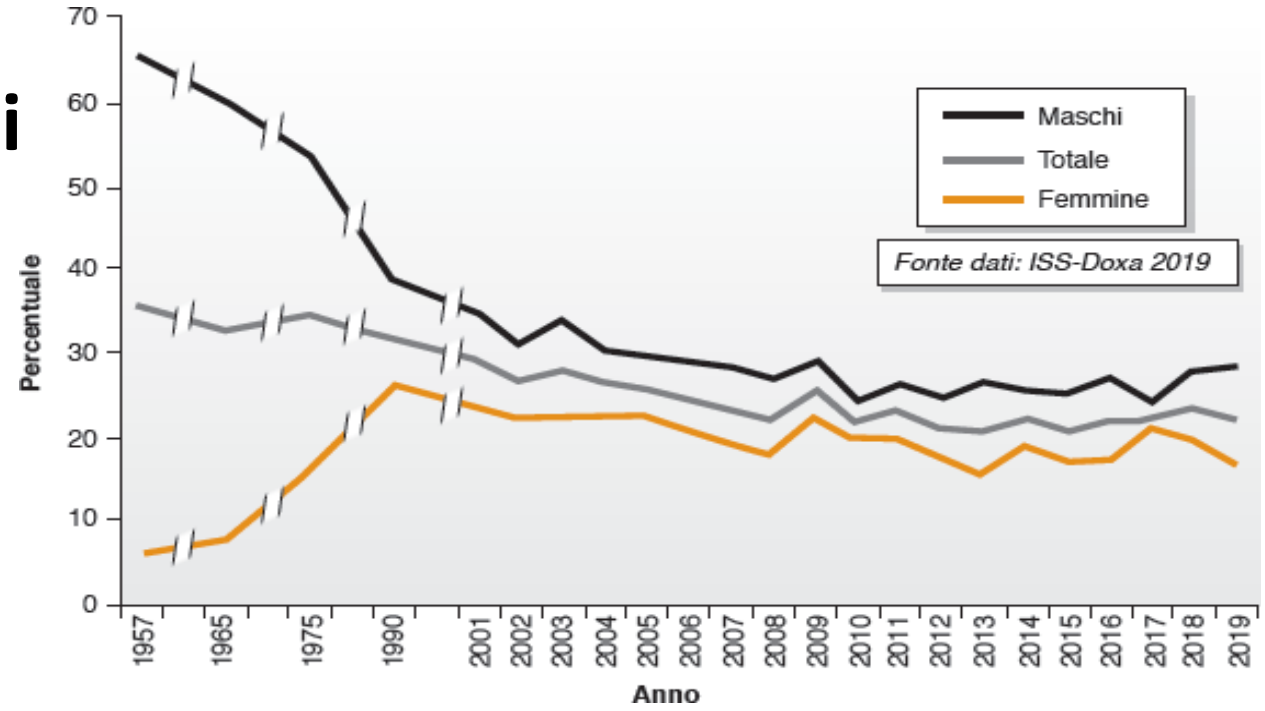
STUDI TRASVERSALI: QUANDO USARLI?

- Le indagini trasversali trovano impiego soprattutto per studiare la distribuzione di **condizioni frequenti**, di **lunga durata**, a **bassa letalità**

- Adatti a studiare la prevalenza delle malattie croniche
- Non adatti a studiare le malattie infettive, soprattutto a ricorrenza epidemica come l'influenza

STUDI TRASVERSALI

Se **ripetute nel tempo** con le medesime modalità, le **indagini trasversali** consentono di **confrontare la prevalenza** di determinati fenomeni in **momenti diversi** e, quindi, di valutare l'andamento temporale degli stessi



Prevalenza del fumo di tabacco in Italia dalle indagini DOXA, dal 1957 al 2019 (fonte dati: ISS-DOXA 2019)

STUDI TRASVERSALI

Metodologia

Studio cross-sectional – Esempio

Prevalence and Determinants of Tinnitus in the Italian Adult Population

Gallus at al, 2015; doi: 10.1159/000431376

Background: Limited, outdated, and poor quality data are available on the prevalence of tinnitus, particularly in Italy. ***Methods:*** A face-to-face survey was conducted in 2014 on 2,952 individuals, who represented the Italian population aged 18 or more (50.6 million). Any tinnitus was defined as the presence of ringing or buzzing in the ears lasting for at least 5 min in the previous 12 months. ***Results:*** Any tinnitus was reported by 6.2% of Italian adults, chronic tinnitus (i.e. for more than 3 months) by 4.8%, and severe tinnitus (i.e. which constitutes a big or very big problem) by 1.2%. The corresponding estimates for the population aged ≥45 years were 8.7, 7.4 and 2.0%, respectively. Multivariable analysis on population aged ≥45 years revealed that old age (odds ratio (OR) = 4.49 for ≥75 vs. 45–54 years) and obesity (OR = 2.14 compared to normal weight) were directly related to any tinnitus, and high monthly family income (OR = 0.50) and moderate alcohol consumption (OR = 0.59 for <7 drinks/week vs. non-drinking) were inversely related. ***Conclusions:*** This is the first study on tinnitus prevalence among the gen-

Studi trasversali: le fasi di pianificazione

1. **Identificazione** della popolazione da sottoporre allo studio (definita *target* o bersaglio) e del **campione rappresentativo**, mediante idonee tecniche di campionamento
2. Definizione del **tempo** nel quale la ricerca deve essere effettuata (se un solo istante → *prevalenza puntuale,* se un arco di tempo → *prevalenza periodale*)
3. Definizione delle **variabili** da indagare e messa a punto degli strumenti di rilevazione, utilizzando quanto più possibile definizioni precise e standard riconosciuti a livello internazionale

DEFINIZIONE DELLA POPOLAZIONE: RAPPRESENTATIVITÀ

- Idealmente, il modo migliore per stimare la prevalenza di una condizione di salute o l'esposizione a una determinata malattia in una popolazione sarebbe quello di acquisire e registrare sistematicamente le informazioni di tutti i membri di tale **popolazione** (censimento)
- In realtà, raramente il tempo e/o i finanziamenti sono sufficienti per raccogliere i dati di tutti i membri di una popolazione. Pertanto, è **necessario trovare un campione** piccolo ma scelto con cura che possa essere utilizzato per rappresentare la popolazione
- A volte siamo interessati solo a un **sottogruppo specifico** (per esempio, solo uomini, giovani o anziani) **della popolazione**. Si può quindi decidere di prendere in considerazione un campione (rappresentativo) di un sottoinsieme della popolazione

METODI DI CAMPIONAMENTO

- **Campionamento probabilistico**: si tratta di un campione all'interno del quale ogni unità della popolazione ha una determinata possibilità di essere selezionata nel campione. Il campionamento probabilistico comprende: campionamento casuale semplice, campionamento stratificato e campionamento a più stadi
- **Campione non probabilistico**: si tratta di un campione in cui alcuni elementi della popolazione non hanno alcuna possibilità di essere selezionati o in cui la probabilità di selezione non può essere determinata con precisione. I metodi di campionamento non probabilistico includono il campionamento per convenienza e il campionamento per quote

MODALITÀ DI RACCOLTA DATI

- Uno studio trasversale può essere condotto usando diverse **modalità di raccolta dati** (con o senza l'utilizzo di intervistatori):
  - CATI (Computer Assisted Telephone Interview)
  - PAPI (Paper and Pen Interview)
  - CAPI (Computer Assisted Personal Interview)
  - CASI (Computer Assisted Self Interview)
  - CAWI (Computer Assisted Web Interview)
  - Mixed mode (or multimode) survey
  - Online panel

Strumento di indagine - Validità e riproducibilità

- Il **disegno del questionario** è **estremamente importante**. Molte indagini si concentrano su obiettivi simili, eppure utilizzano questionari completamente diversi. Questo spesso impedisce ai ricercatori di confrontare i risultati di diversi studi
- Nuovi questionari dovrebbero essere testati per la **riproducibilità e la validità**, al fine di garantire che catturino accuratamente ciò che intendono misurare
- La formulazione delle domande è fondamentale e dovrebbe prendere in considerazione l'adeguatezza del contenuto e il livello di sofisticazione del linguaggio

Studio cross-sectional – Esempio

- **Fieldwork:** DOXA
- **Year:** 2014
- **Sample population:** sample **representative** of the general adult Italian population (aged 18 years or more; 50.6 million)
- **Data collection mode:** Computer-assisted personal interview (CAPI)
- **Sampling method:** multi-stage random sampling (3 stages)
- **Sample size:** 2952 adults

Strumento di indagine – Traduzione da altre lingue

- La **traduzione di un questionario** è **indispensabile** se uno strumento non è disponibile nella lingua richiesta
- La traduzione non è un lavoro meccanico e non dovrebbe essere implementata parola per parola in tutte le lingue. È importante comprendere le interrelazioni tra una lingua specifica (o la scelta delle parole) e il suo contesto locale, problemi specifici e significati culturali
- La **traduzione inversa** è preziosa per valutare la qualità della traduzione. La lingua di partenza viene tradotta in un'altra lingua e poi ritradotta nella lingua di partenza originale da un altro traduttore, che non è stato esposto alla versione originale

Studio cross-sectional – Esempio

- **Presence of tinnitus** : “In the past 12 months, have you been bothered by ringing or buzzing in your ears or head that lasted for 5 minutes or more?”
  - No
  - Yes ➔ **Any tinnitus**
- **Chronic tinnitus**: “How long have you been bothered by this ringing or buzzing in your ears or head?”
  - less than 3 months ➔ Acute tinnitus
  - 3 months or more ➔ **Chronic tinnitus**
- **Severity of tinnitus**: “How much of a problem is this ringing or buzzing in one or both ears or in your head?”
  - “no problem” ➔ Mild tinnitus
  - “a small problem” ➔ Mild tinnitus
  - “a moderate problem” ➔ Moderate tinnitus
  - “a big problem”
  - “a very big problem” ➔ **Severe tinnitus**

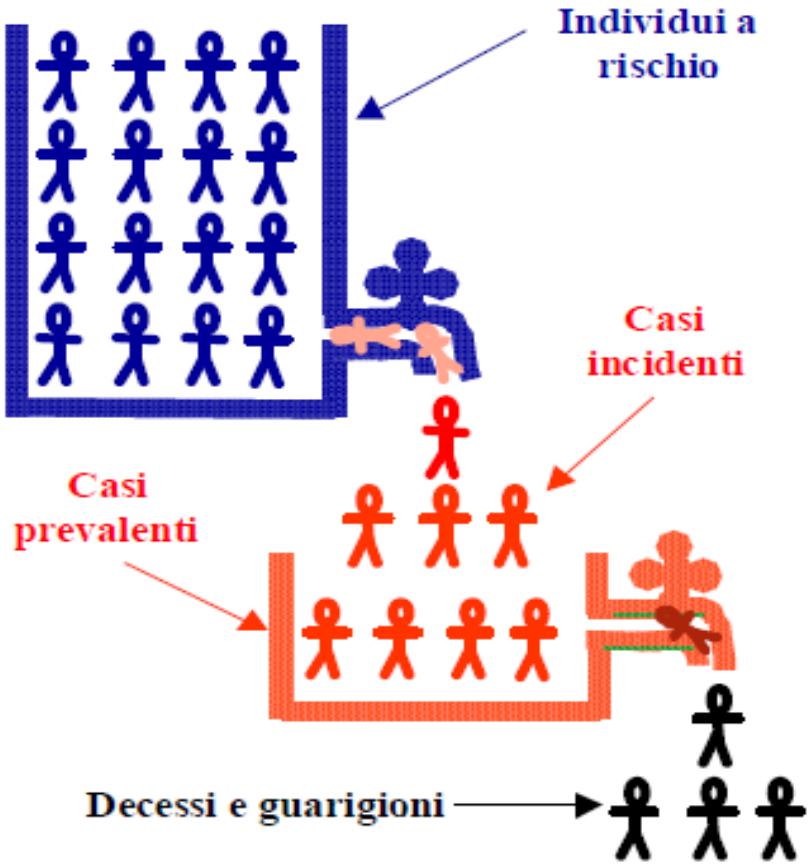
Indicatore epidemiologico - Prevalenza

***Prevalenza* (assoluta)** = N. persone vive con la malattia in un certo istante di tempo

***Prevalenza* (%)** = N. persone vive con la malattia/Popolazione totale

Prevalenza

I casi di malattia presenti in un determinato istante in una popolazione, indipendentemente da quando sono insorti, sono definiti **CASI PREVALENTI**



Prevalenza

Il numero di casi prevalenti dipende:

- da **quante persone si ammalano** (casi incidenti)
- da **quanto tempo i casi**, una volta insorti, **rimangono nello stato di malattia** o vi escono per effetto delle guarigione o del decesso

✓ Le misure di prevalenza non forniscono forte evidenza di causalità

✓ Sono tuttavia utili per valutare la necessità di assistenza sanitaria e la pianificazione dei servizi sanitari

Prevalenza

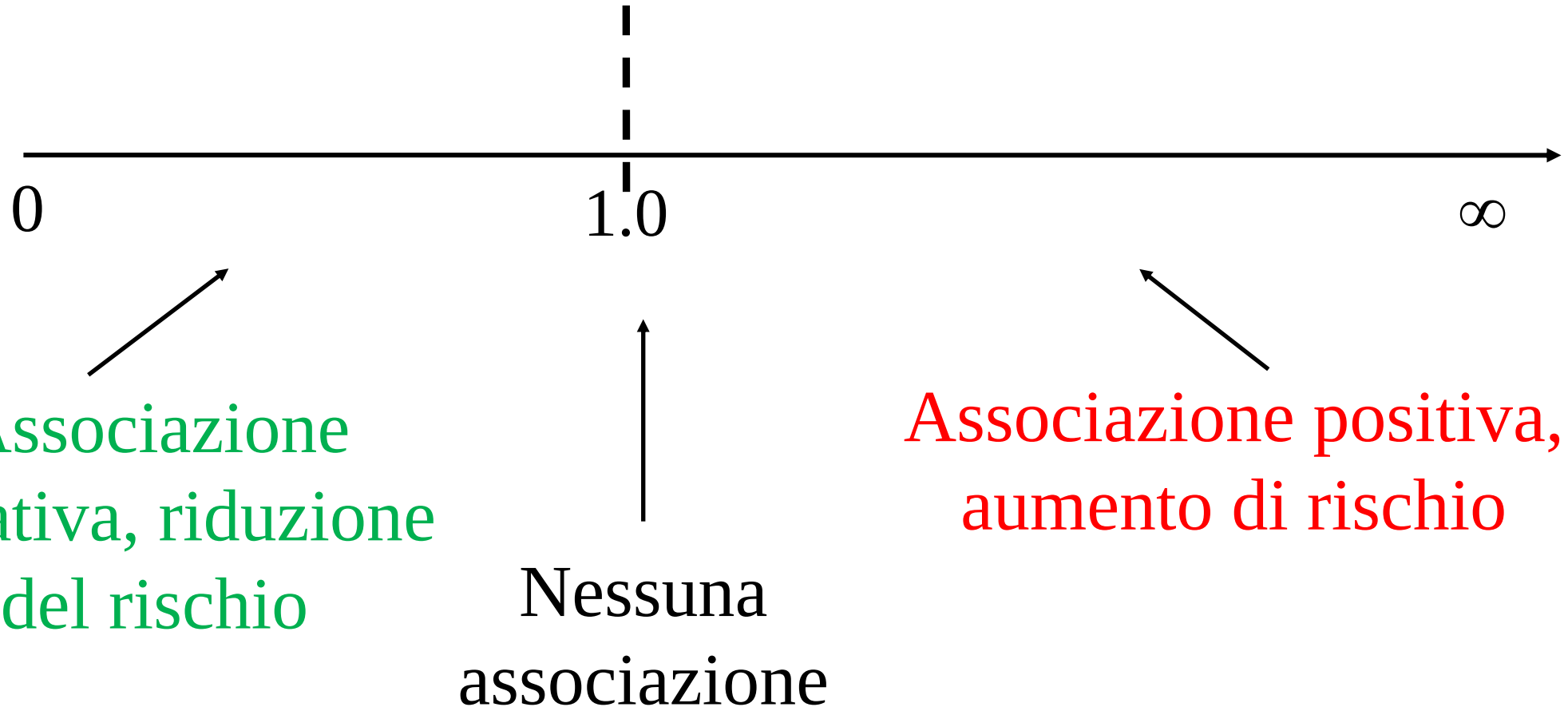
- Rende l’idea di **quanto è presente** una malattia nella popolazione
- E’ solitamente **elevata per malattie croniche** con bassa letalità
- E’ solitamente **bassa per malattie acute**, da cui si può guarire o con alta letalità

Studio cross-sectional – Esempio

Table 1. Percent prevalence[a] of tinnitus, and corresponding 95% CI, overall and by sex and age group, according to the duration of symptom and its severity, among 2,952 adults, Italy, 2014

	Overall		Sex				Age group, years					
	%	95% CI	men		women		18–44		45–64		≥65	
			%	95% CI	%	95% CI	%	95% CI	%	95% CI	%	95% CI
Any tinnitus	**6.2**	**5.3–7.0**	**6.0**	**4.7–7.2**	**6.4**	**5.2–7.6**	**2.7**	**1.8–3.6**	**5.9**	**4.5–7.4**	**12.3**	**9.9–14.6**
Tinnitus duration												
Acute tinnitus (<3 months)	1.3	0.9–1.7	1.1	0.5–1.6	1.6	1.0–2.2	1.4	0.7–2.1	1.3	0.6–2.1	1.2	0.4–2.0
Chronic tinnitus (≥3 months)	**4.8**	**4.1–5.6**	**4.9**	**3.8–6.0**	**4.8**	**3.7–5.9**	**1.3**	**0.7–1.9**	**4.6**	**3.3–5.9**	**11.1**	**8.8–13.3**
Tinnitus severity												
Mild tinnitus	2.9	2.3–3.5	3.1	2.2–4.1	2.6	1.8–3.4	2.1	1.3–2.9	2.8	1.8–3.9	4.3	2.8–5.7
No problem	0.5	0.2–0.7	0.2	0.0–0.5	0.7	0.3–1.1	0.6	0.1–1.0	0.3	0.0–0.6	0.7	0.1–1.3
Small problem	2.4	1.8–3.0	2.9	2.0–3.8	1.9	1.2–2.6	1.5	0.8–2.2	2.6	1.6–3.6	3.6	2.3–5.0
Moderate tinnitus												
Moderate problem	2.1	1.6–2.6	2.0	1.3–2.8	2.1	1.4–2.8	0.5	0.1–0.9	2.0	1.1–2.8	4.8	3.2–6.3
Severe tinnitus	**1.2**	**0.8–1.6**	**0.8**	**0.3–1.3**	**1.6**	**1.0–2.3**	**0.1**	**0.0–0.3**	**1.2**	**0.5–1.8**	**3.2**	**1.9–4.5**
Big problem	1.0	0.6–1.3	0.4	0.1–0.7	1.5	0.9–2.1	0.0	–	0.9	0.3–1.5	2.7	1.5–3.8
Very big problem	0.3	0.1–0.5	0.4	0.1–0.8	0.1	0.0–0.3	0.1	0.0–0.3	0.3	0.0–0.6	0.6	0.0–1.1

Rischio relativo (RR)



RR = 1.5 *Rischio aumentato del 50%*

RR = 2.0 *Rischio doppio ovvero aumento del rischio del 100%*

RR = 0.7 *Rischio ridotto del 30%*

Studio cross-sectional – Esempio

Table 2. Percent prevalence of any, chronic, and severe tinnitus among 1,724 adults aged ≥45 years, according to selected socio-demographic characteristics, with corresponding OR[a] and 95% CI, Italy, 2014

	n	Any tinnitus		Chronic tinnitus		Severe tinnitus	
		%	OR (95% CI)	%	OR (95% CI)	%	OR (95% CI)
Total	1,724	8.7	–	7.4	–	2.0	–
Sex							
Men	792	8.3	1[b]	7.0	1[b]	1.3	1[b]
Women	932	9.0	1.43 (0.92–2.20)	7.8	1.45 (0.91–2.32)	2.7	**3.26 (1.28–8.31)**
Age group, years							
45–54	536	3.8	1[b]	2.5	1[b]	0.5	1[b]
55–64	445	8.6	**2.18 (1.20–3.95)**	7.1	**2.63 (1.32–5.23)**	2.0	3.57 (0.87–14.66)
65–74	492	10.8	**2.80 (1.55–5.08)**	9.9	**3.79 (1.93–7.45)**	2.4	**4.21 (1.03–17.17)**
≥75	251	15.2	**4.49 (2.34–8.62)**	13.4	**5.87 (2.81–12.26)**	4.9	**9.59 (2.24–40.96)**
p for trend			**<0.001**		**<0.001**		**0.002**
Level of education							
Low	934	10.1	1[b]	8.6	1[b]	2.6	1[b]
Intermediate	596	7.9	1.30 (0.85–1.99)	7.2	1.55 (0.98–2.43)	1.7	1.34 (0.59–3.04)
High	193	3.9	0.61 (0.28–1.35)	2.2	0.43 (0.16–1.19)	0.5	0.29 (0.03–2.52)
p for trend			0.727		0.788		0.587

Studio cross-sectional – Esempio

Table 3. Percent prevalence of any, chronic and severe tinnitus among 1,724 adults aged ≥45 years, according to smoking status, alcohol consumption, and BMI, and corresponding OR[a] and 95% CI, Italy, 2014

	n	Any tinnitus		Chronic tinnitus		Severe tinnitus	
		%	OR (95% CI)	%	OR (95% CI)	%	OR (95% CI)
Smoking status							
Never smokers	1,097	8.4	1[b]	7.2	1[b]	2.1	1[b]
Current smokers	312	5.2	0.64 (0.34–1.20)	3.8	0.59 (0.29–1.19)	1.1	0.94 (0.29–3.09)
Ex-smokers	314	13.0	1.50 (0.96–2.34)	11.7	1.53 (0.96–2.46)	2.6	1.51 (0.61–3.76)
Alcohol drinking[c]							
Non-drinkers	727	9.7	1[b]	8.5	1[b]	2.8	1[b]
<7 drinks/week	483	5.0	**0.59 (0.36–0.98)**	4.4	0.60 (0.35–1.03)	0.4	**0.19 (0.04–0.89)**
≥7 drinks/week	500	10.7	1.38 (0.87–2.21)	8.8	1.29 (0.78–2.14)	2.6	1.69 (0.72–3.97)
p for trend			0.307		0.479		0.447
BMI categories[c]							
Under/normal weight (BMI <25 kg/m^2)	741	6.1	1[b]	5.0	1[b]	1.4	1[b]
Overweight (25≤ BMI <30 kg/m^2)	642	10.5	1.49 (0.99–2.25)	9.1	**1.56 (1.00–2.44)**	3.2	1.99 (0.91–4.39)
Obesity (BMI ≥30 kg/m^2)	175	14.7	**2.14 (1.25–3.67)**	13.3	**2.31 (1.30–4.10)**	2.7	1.50 (0.48–4.75)
p for trend			**0.004**		**0.003**		0.241

STUDI TRASVERSALI

Punti di forza/debolezza

Punti di forza

- Di **breve durata**, **poco costosi**
- Se il campionamento viene effettuato correttamente, i risultati sono **generalizzabili** al tutta la popolazione
- **Libera scelta sia della popolazione da cui selezionare il campione**, sia dei metodi di rilevazione (es. questionari, accertamenti clinici, ecc.), sia dei criteri diagnostici da applicare

Punti di debolezza

- **Non adatti** a studiare **malattie rare** o **di breve durata**
- I dati relativi a **esposizioni pregresse** possono essere **poco attendibili**
- **Non** fornisce indicazioni sull'**incidenza** del fenomeno in studio
- **Non** fornisce alcuna **informazione sulla natura «causale» dell'associazione**, in quanto non consente di ottenere la sequenza temporale corretta

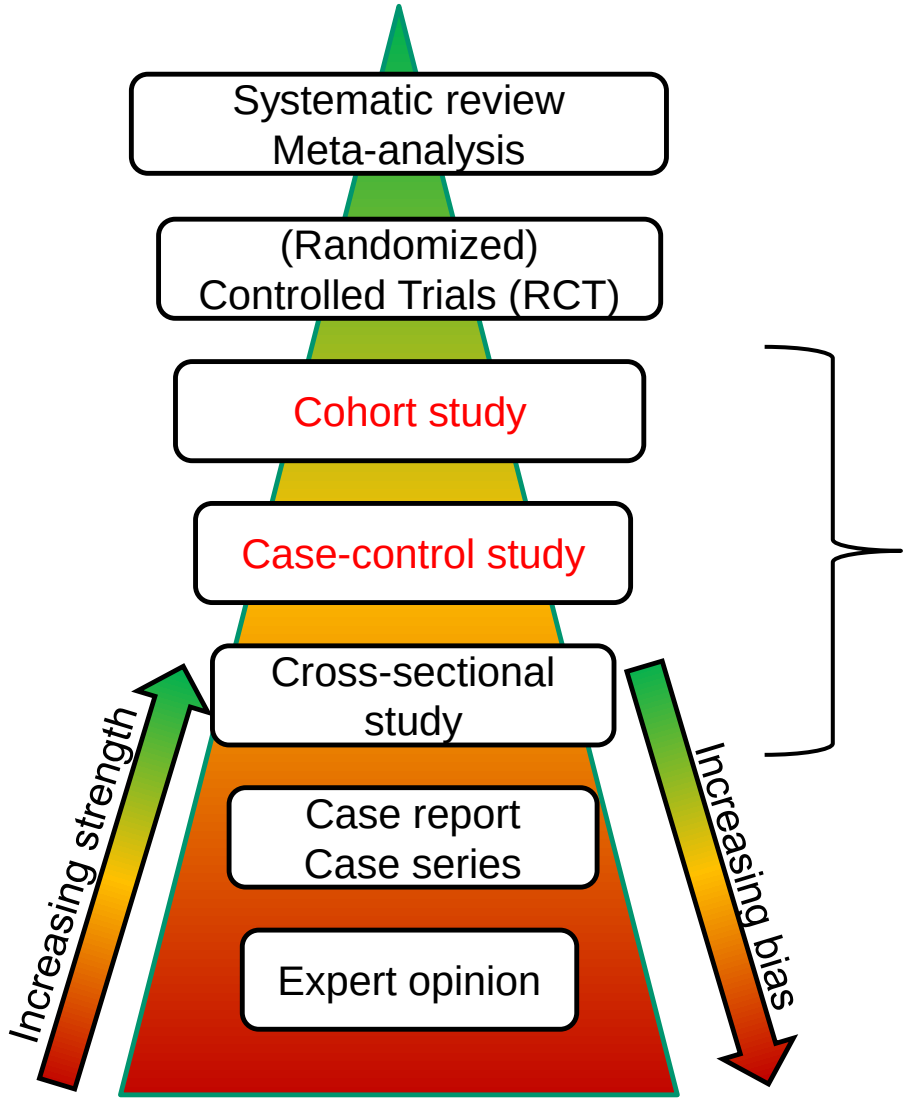
Studi caso-controllo

(C. Bosetti)

Outline

- Definizione
- Metodologia di pianificazione, conduzione e analisi
- Punti di forza/debolezza
- Qualità metodologica

Gerarchia degli studi clinici



Studi osservazionali analitici

STUDI CASO-CONTROLLO

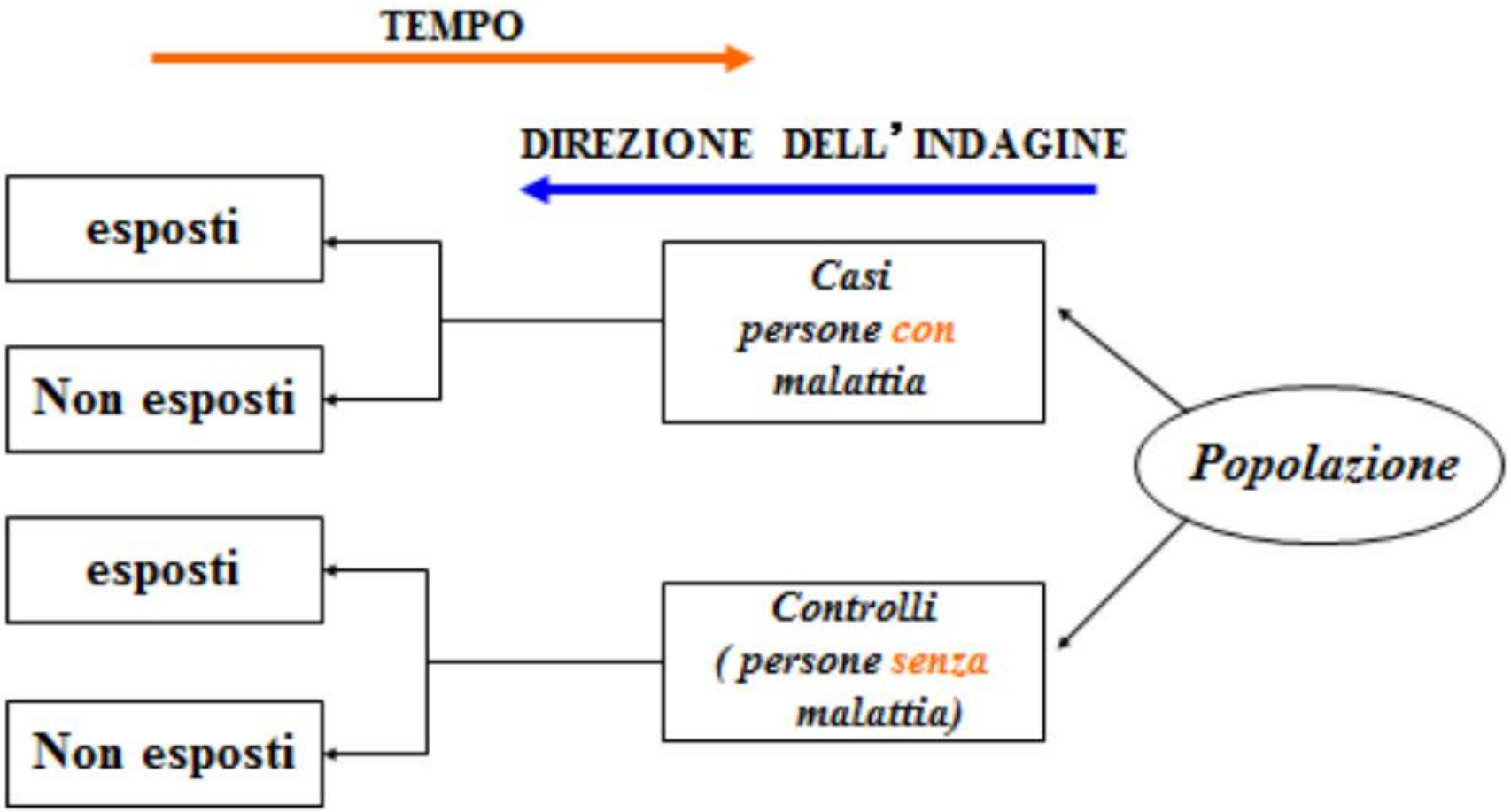
Definizione

Che cosa è uno studio caso-controllo?

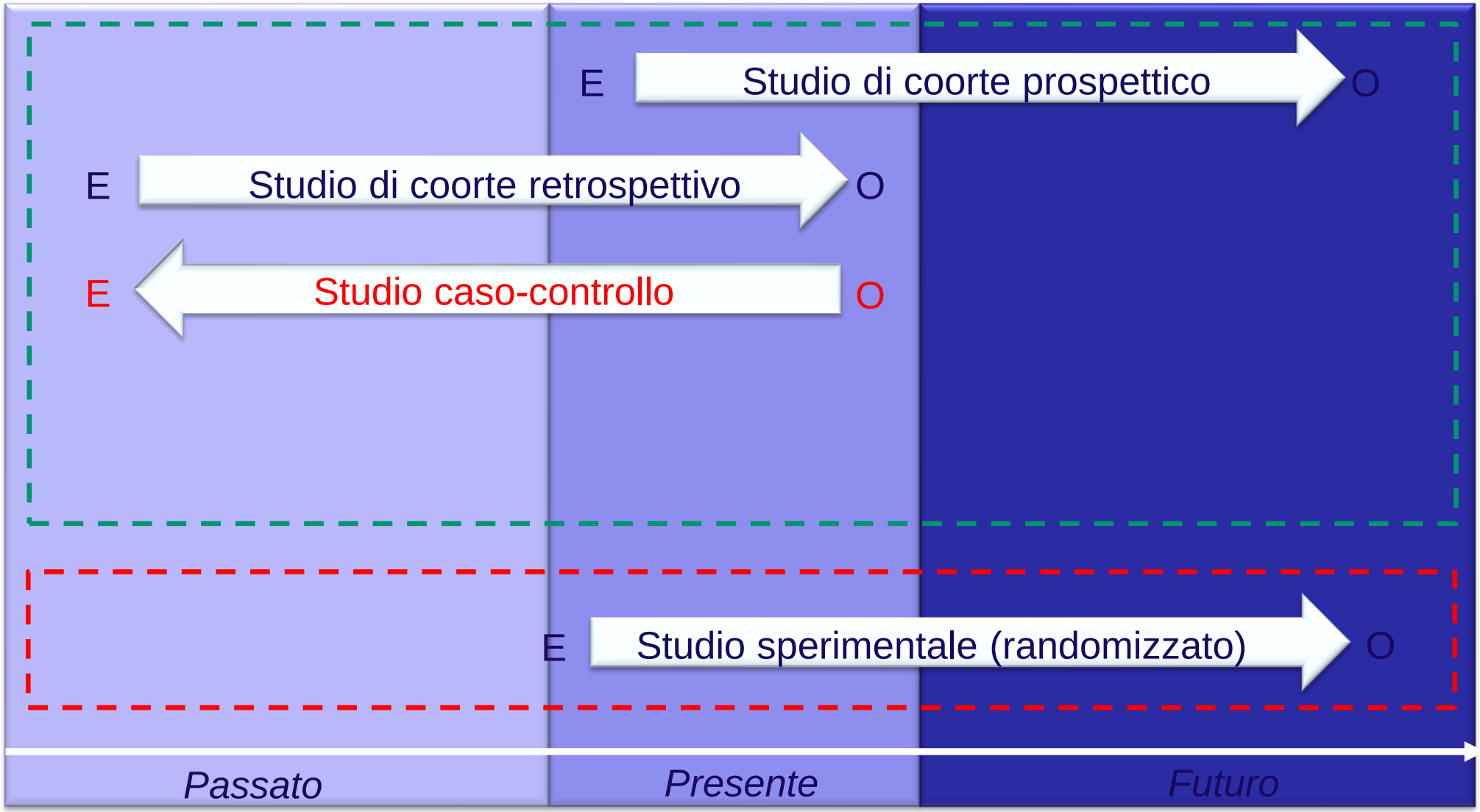
E’ uno studio osservazionale in cui vi sono individui con una malattia di interesse (casi) e un gruppo adeguato di individui senza la malattia in studio (controllo).

La potenziale relazione causale tra un fattore di rischio viene esaminata confrontando i soggetti malati e i non malati per quanto riguarda quanto frequentemente è presente il fattore di rischio in ciascuno dei due gruppi *(A Dictionary of Epidemiology, Porta et al. 2014)*

Studio caso-controllo



Studi prospettici o retrospettici



E = Esposizione al fattore in studio
O = Outcome di interesse (malattia)

Studio caso-controllo

OBIETTIVI

- Valutare il ruolo di uno o più fattori di rischio nell'eziopatogenesi di una malattia
- Stimare l'associazione dei singoli fattori di rischio attraverso l'odds ratio e della loro eventuale interazione

TIPI DI DATI

- Dati individuali ottenuti tramite documenti personali (e. cartelle cliniche, registri, ecc.) o contatto diretto con i soggetti (es. interviste, questionari, ecc.)

STUDI CASO-CONTROLLO

Metodologia

Studio caso-controllo – Esempio

Type 2 Diabetes, Antidiabetic Medications, and Colorectal Cancer Risk: Two Case–Control Studies from Italy and Spain

Rosato et al, 2016; doi: 10.3389/fonc.2016.00210

Background: Type 2 diabetes mellitus has been associated with an excess risk of colorectal cancer, although the time–risk relationship is unclear, and there is limited information on the role of antidiabetic medications.

Aim: We examined the association between type 2 diabetes, antidiabetic medications, and the risk of colorectal cancer, considering also duration of exposures.

Methods: We analyzed data derived from two companion case–control studies conducted in Italy and Spain between 2007 and 2013 on 1,147 histologically confirmed colorectal cancer cases and 1,594 corresponding controls. Odds ratios (ORs) and 95% confidence intervals (CIs) were estimated by unconditional multiple logistic regression models, adjusted for socioeconomic factors and major potential confounding factors.

Results: Overall, 14% of cases and 12% of controls reported a diagnosis of diabetes, corresponding to an OR of colorectal cancer of 1.21 (95% CI 0.95–1.55). The OR was 1.49 (95% CI 0.97–2.29) for a duration of diabetes of at least 15 years. The OR was 1.53 (95% CI 1.06–2.19) for proximal colon cancer, 0.94 (95% CI 0.66–1.36) for distal colon cancer, and 1.32 (95% CI 0.94–1.87) for rectal cancer. In comparison with no use, metformin use was associated with a decreased colorectal cancer risk (OR 0.47, 95% CI 0.24–0.92), while insulin use was associated with an increased risk (OR 2.20, 95% CI 1.12–4.33); these associations were stronger for longer use (OR 0.36 and 8.18 for ≥10 years of use of metformin and insulin, respectively).

Conclusion: This study shows evidence of a positive association between diabetes and colorectal cancer, mainly proximal colon cancer. Moreover, it indicates a negative association between colorectal cancer and metformin use and a positive association for insulin use.

Selezione dei casi

- **Definizione della malattia**
- Definizione delle **caratteristiche dei soggetti** che siamo interessati a studiare (sesso, età, ….)

Selezione dei casi

- **Scelta di provenienza dei casi**

  Casi **ospedalieri**: sono più collaborativi e più semplici da intervistare, ma difficile identificare la popolazione che li ha generati
  Casi di **popolazione**: si evitano distorsioni nella selezione dei casi, ma serve valido sistema di registrazione dei casi della malattia in studio

Selezione dei casi

- **Casi incidenti/prevalenti**

  Casi **incidenti** (soggetti con nuova diagnosi)

  Casi **prevalenti** (diagnosi avvenuta in passato)

Selezione dei casi

CASI INCIDENTI

- Minore il tempo che intercorre tra il momento dell'esposizione e la diagnosi → si riesce a **quantificare meglio l'esposizione** al fattore di rischio
- Più semplice **distinguere tra i fattori con reale significato causale** da quelli legati alla sopravvivenza del paziente
- Il tempo di r**eclutamento** è **più lungo**

CASI PREVALENTI

- Risulta **più probabile la selezione di malati meno gravi**, con malattia a più lunga durata
- La rilevazione **dell'esposizione può essere condizionata dalla presenza della malattia**

Selezione dei controlli

- Devono essere selezionati dalla **stessa popolazione da cui sono estratti i casi**
- Devono essere **confrontabili con i casi** (possibile appaiamento o *matching*, ad esempio per età, sesso, …)
- La scelta dei controlli deve essere **indipendente dall'esposizione**

Controlli appaiati (*matched*)

- I controlli possono essere **appaiati** per età (e sesso) (rapporto 1:1, 1:2, …, 1:5)
- L'appaiamento rende simili casi e controllo rispetto alla caratteristica in questione in modo tale che questa non possa influenzare l'effetto di interesse

Selezione dei controlli

- **Provenienza dei controlli**

  Controlli **ospedalieri**

  Controlli di **popolazione**

Selezione dei controlli

CONTROLLI OSPEDALIERI

- Sono **più facili da rintracciare**
- Sono in genere **più collaborativi**
- Sono **più confrontabili** con i casi

Selezione dei controlli

CONTROLLI OSPEDALIERI

- Sono **più facili da rintracciare**
- Sono in genere **più collaborativi**
- Sono **più confrontabili** con i casi

ma….

- Devono provenire da **ospedali con bacini di utenza coincidenti con quelli dei casi**
- Devono essere **ricoverati per patologie che non hanno a che fare con la patologia in studio**
- In quanto malati, **non sempre comparabili con la popolazione sorgente dei casi** in termini di esposizione

Selezione dei controlli

CONTROLLI DI POPOLAZIONE

- Sono **più rappresentativi della popolazione** da cui derivano i **casi**

Selezione dei controlli

CONTROLLI DI POPOLAZIONE

- Sono **più rappresentativi della popolazione** da cui derivano i **casi**

ma….

- **Più complicati e dispendiosi**
- Per la scelta **richiedono elenchi precostituiti dei soggetti eligibili**
- **Alte percentuali tendono a rifiutare** la collaborazione
- Sono **meno attenti e precisi** nel riportare le informazioni all'intervista

Studio cross-sectional – Esempio

- **Study area:** Italy and Spain
- **Year:** 2007-2013
- **Cases:** 1,147 incident and histologically confirmed colorectal cancers admitted to hospital in the study areas
- **Controls**: Italy: patients admitted to the same hospitals as cases for a wide spectrum of acute, non-neoplastic conditions; Spain: population-based (identified from the lists of selected family practitioners and contacted by telephone)
- **Data collection:** interviewed by *ad hoc* trained interviewers using similar structured questionnaires to collect sociodemographic factors, lifestyle habits, anthropometric measures, a problem-oriented medical history, and family history of cancer

Confronto tra casi e controlli

Il **confronto** tra i due gruppi - **casi e controlli** – sulla base della distribuzione dell'esposizione ad un fattore consente di misurare l'associazione tra un fattore di rischio/protettivo e una determinata malattia

Se il fattore di esposizione è **presente maggiormente tra i casi** rispetto ai controlli ➔ il fattore è **associato direttamente/positivamente** alla malattia

Se il fattore di esposizione è **presente in maniera minore tra i casi** rispetto ai controlli ➔ il fattore è **associato inversamente/negativamente** alla malattia

Misure di associazione

Odds Ratio (OR): rapporto tra la l'odds di esposizione nei casi e l'odds di esposizione nei controlli

Se la malattia è rara OR ~ RR

Rischio relativo (RR)

RR = Rischio di sviluppare la malattia tra gli esposti /
Rischio di sviluppare la malattia tra i non esposti

	MALATTIA		
ESPOSIZIONE	**SI (Malati)**	**NO (Non malati)**	**Totale**
SI (Esposti)	A	B	**A+B**
NO (Non esposti)	C	D	**C+D**

RR = [A /(A+B)] / [C/(C+D)]

Odds ratio (OR)

ESPOSIZIONE	MALATTIA		Totale
	SI (Casi)	NO (Controlli)	
SI (Esposti)	A	B	A+B
NO (Non esposti)	C	D	C+D
Totale	A+C	B+D	

Odds ratio (OR)

	MALATTIA	
ESPOSIZIONE	**SI (Casi)**	**NO (Controlli)**
SI (Esposti)	A	B
NO (Non esposti)	C	D
Totale	**A+C**	**B+D**

$Odds_{malati}$ = (A/A+C)/(C/A+C) = **A/C**

$Odds_{non\ malati}$ = (B/B+D)/(D/B+D) = **B/D**

→ **$OR = Odds_M/Odds_{NM}$**

= (A/C)/(B/D)

=AxD/BxC

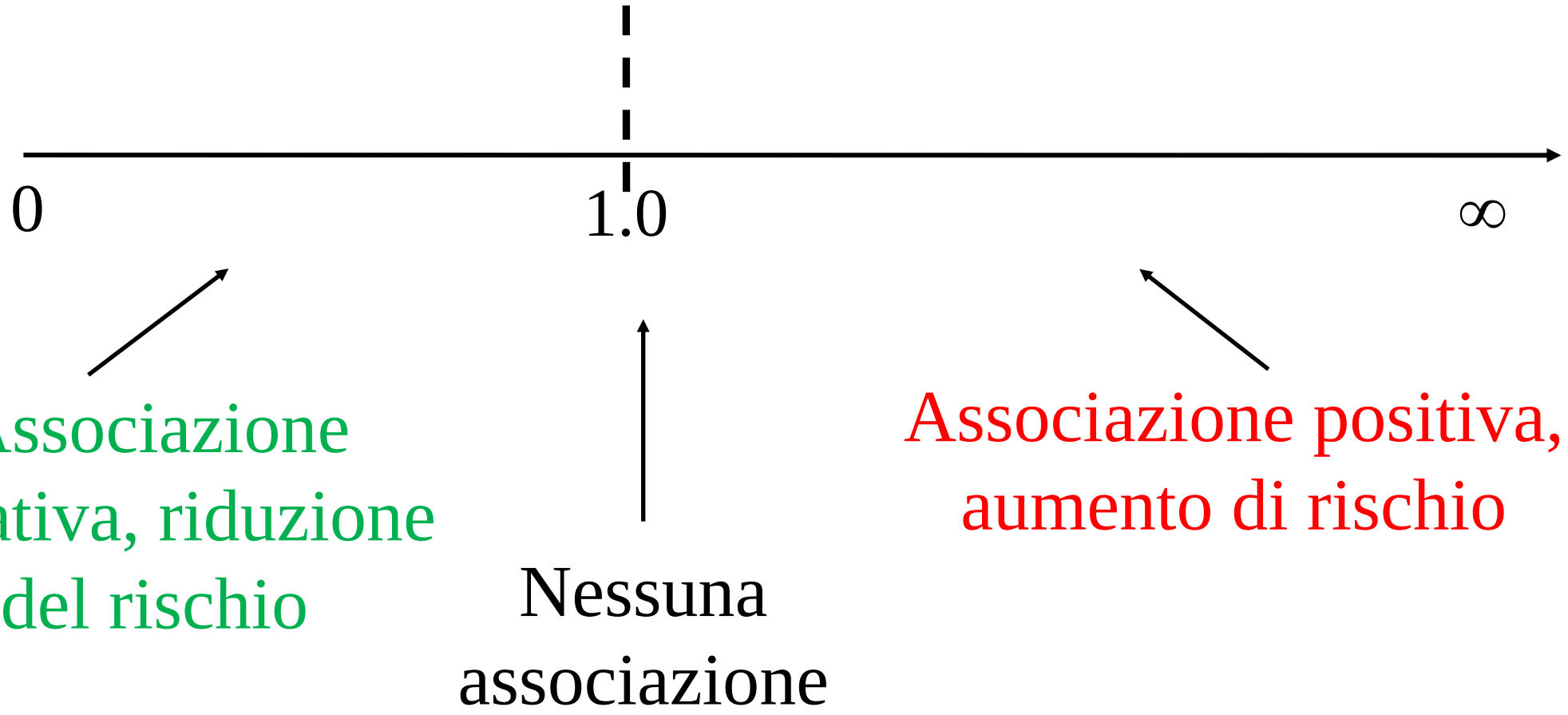
Odds ratio (OR)

	MALATTIA	
ESPOSIZIONE	**SI (Casi)**	**NO (Controlli)**
SI (Esposti)	A	B
NO (Non esposti)	C	D

Rapporto incrociato

OR = AxD/BxC

Rischio relativo (RR)



RR = 1.5 *Rischio aumentato del 50%*

RR = 2.0 *Rischio doppio ovvero aumento del rischio del 100%*

RR = 0.7 *Rischio ridotto del 30%*

Odds ratio (OR) - Esempio

ESPOSIZIONE **Uso di insulina**	**MALATTIA** **Tumore del colonretto**	
	SI	**NO**
SI	36	27
NO	73	108
Totale	**109**	**135**

OR = (36x108)/(27x73) = **1.97**

Studio caso-controllo – Esempio

TABLE 2 | Distribution of 1,147 colorectal cancer cases and corresponding 1,594 controls, with odds ratios (ORs) and 95% confidence intervals (CIs), according to history of diabetes by country and overall.

	Cases *N* (%)	Controls *N* (%)	OR[a] (95% CI)
History of diabetes			
No	988 (86.1)	1,406 (88.2)	1.00[b]
Yes	159 (13.9)	188 (11.8)	1.21 (0.95–1.55)

[a]*Estimates from multiple logistic regression models including terms for study center, sex, age, education, tobacco smoking, alcohol drinking, body mass index, physical activity, statin use, and aspirin use.*

[b]*Reference category.*

Rosato et al, 2016; doi: 10.3389/fonc.2016.00210

Studio caso-controllo – Esempio

TABLE 4 | Distribution of 109 diabetic colorectal cancer cases and 135 diabetic controls, with corresponding odds ratios (ORs) and 95% confidence intervals (CIs), according to use of antidiabetic medications.

Antidiabetic medications	Cases		Controls		OR[a] (95% CI)
	N	(%)	*N*	(%)	
Metformin					
No[d]	36	(33.0)	32	(23.7)	1[c]
Yes	73	(67.0)	103	(76.3)	0.47 (0.24–0.92)
Duration (years)					
1 to <5	34	(46.6)	34	(33.0)	0.74 (0.33–1.66)
5 to <10	11	(15.1)	27	(26.2)	0.25 (0.09–0.69)
≥10	20	(27.4)	34	(33.0)	0.36 (0.15–0.85)
Missing	8	(11.0)	8	(7.8)	
p-Value for trend					0.005
Insulin					
No[d]	73	(67.0)	108	(80.0)	1[c]
Yes	36	(33.0)	27	(20.0)	2.20 (1.12–4.33)
Duration (years)					
1 to <5	15	(41.7)	16	(59.3)	1.52 (0.63–3.71)
5 to <10	5	(13.9)	4	(14.8)	1.86 (0.43–8.04)
≥10	14	(38.9)	4	(14.8)	8.18 (2.06–32.50)
Missing	2	(5.6)	3	(11.1)	
p-Value for trend					0.002

Rosato et al, 2016; doi: 10.3389/fonc.2016.00210

STUDI CASO-CONTROLLO

Punti di forza/debolezza

Punti di forza

- **Durata limitata**, **relativamente economici** e logisticamente **facili da condurre**
- Possono essere l'unico modo **per studiare patologie rare o a lunga induzione**
- Permettono di raccogliere **informazioni specifiche** e **dettagliate** per ciascun soggetto
- Permettono di studiare un **ampio spettro di potenziali fattori eziologici**

Punti di debolezza

- Per definizione, **permettono di studiare una sola patologia**
- Sono **soggetti a errori sistematici** (*bias*)
- Non è presente esatta sequenza temporale dell'associazione (possibile ***reverse causation***)
- Consentono di stimare solo i **rischi relativi**
- **Non** consentono di **studiare fattori di rischio che possono essere alterati dalla patologia**

Errori sistematici (bias)

- Bias nella selezione dei soggetti (***selection bias***): distorsione dovuta alle modalità di selezione della popolazione in studio
- Bias di informazione (***recall bias***): distorsione dovuta a non corretta classificazione dell'esposizione di interesse

Confondimento

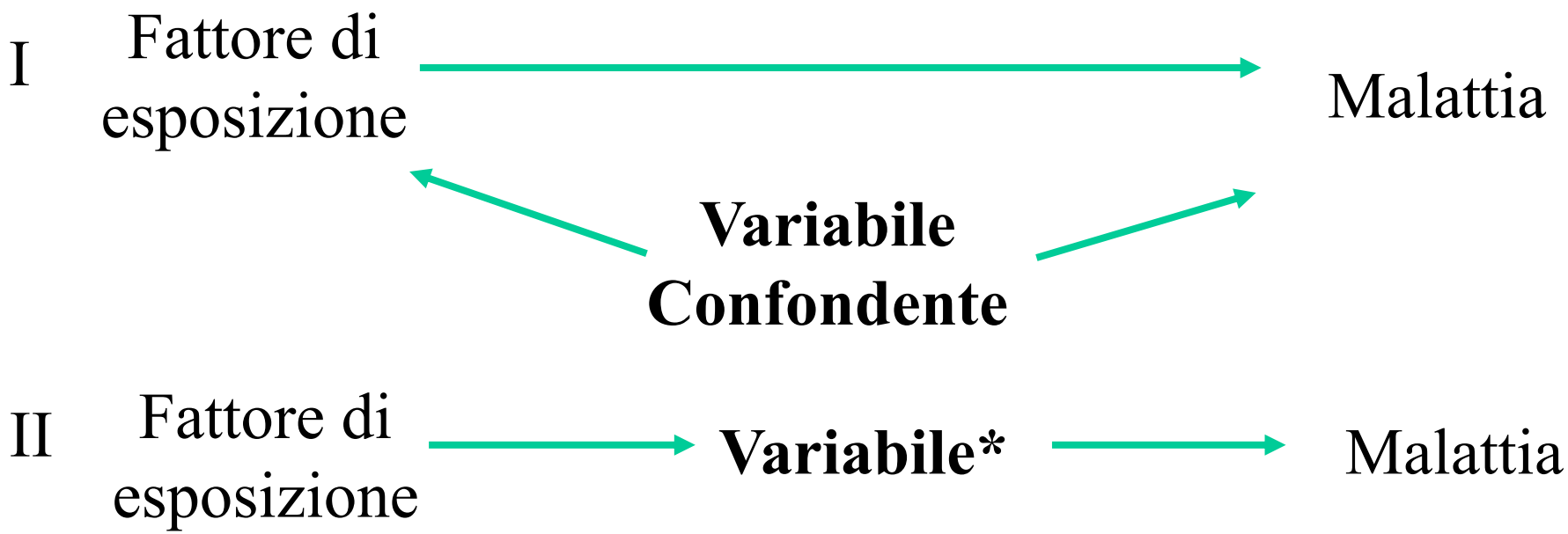
Distorsione del rapporto causa-effetto dovuto **all'associazione dell'esposizione con un altro fattore (confondente)** che influenza l'occorrenza dell'outcome

Confondente deve soddisfare le due condizioni seguenti:

- Deve essere **estraneo alle due variabili** (malattia e esposizione) ma può distorcere la loro relazione
- Deve essere **associato** sia con la **malattia** che con il fattore di **esposizione** studiati

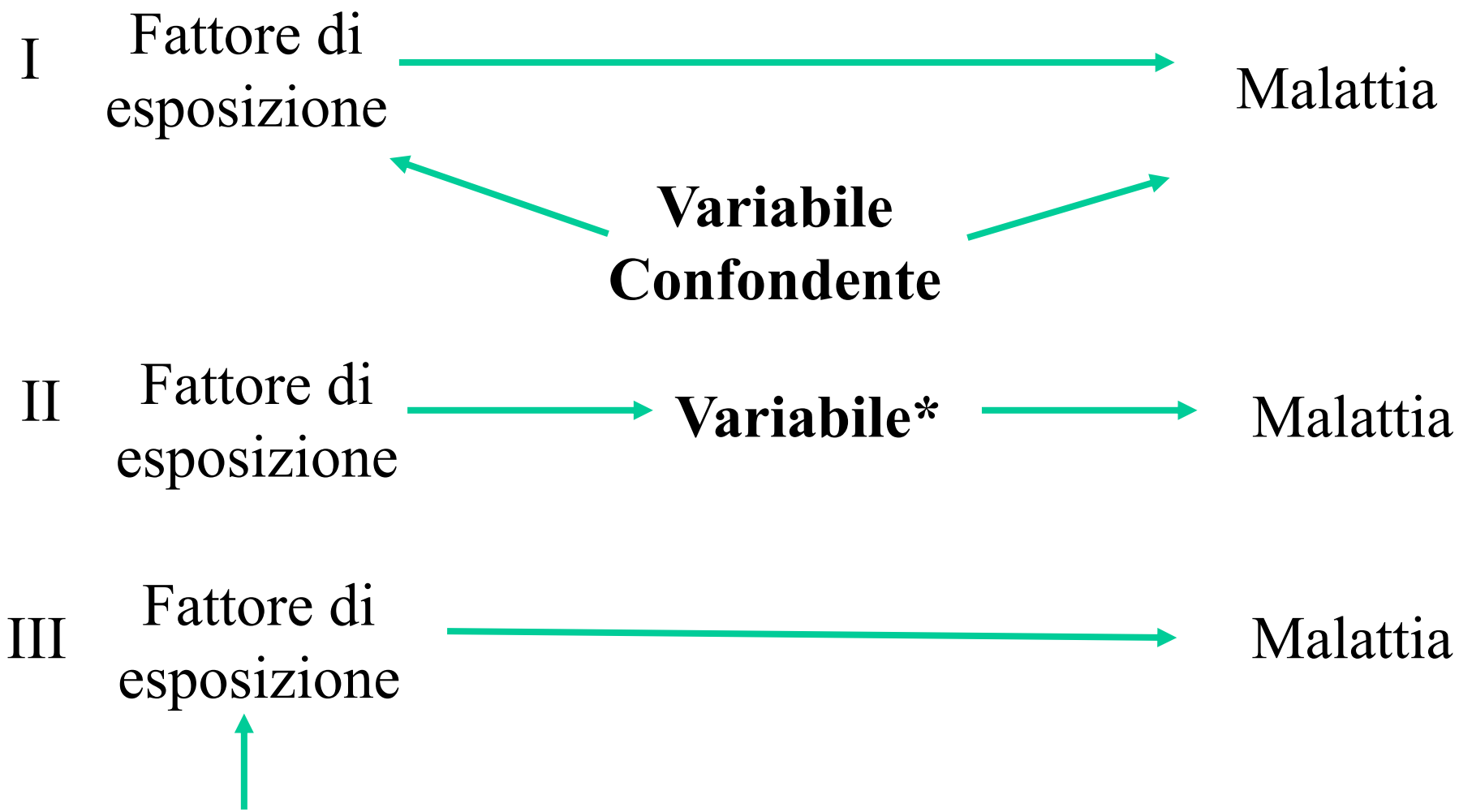
Variabile confondente

Variabile confondente



*Variabile **NON** confondente

Variabile confondente



*Variabile **NON** confondente

Variabile confondente - Esempio

Alcol	**Casi (Tumore polmone)**	**Controlli**
Si	71	52
No	29	48
Totale	100	100

OR= (71x48)/(52x29) = **2.26**

Variabile confondente - Esempio

Alcol	Casi	Controlli
Si	71	52
No	29	48

Non fumatori

Alcol	Casi	Controlli
Si	8	16
No	22	44

OR = (8x44)/(16x22) = 1.0

Fumatori

Alcol	Casi	Controlli
Si	63	36
No	7	4

OR = (63x4)/(36x7) = 1.0

STUDI CASO-CONTROLLO

Qualità metodologica

La qualità di uno studio osservazionale

The Newcastle-Ottawa Quality Assessment Scale (NOS)

Wells, G.A., Shea, B., O'Connell, D.A., Peterson, J., Welch, V., Losos, M., et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses 2013
https://www.ohri.ca/programs/clinical_epidemiology/oxford.asp

NEWCASTLE-OTTAWA SCALE (NOS) – Studi caso-controllo

Vengono proposti **8 domande** classificate in **3 categorie**:

- **Selezione** (max 4 punti)
- **Confrontabilità** tra i gruppi (max 2 punti)
- **Esposizione** (max 3 punti)

Valutazione della qualità di uno studio caso-controllo (1)

SELEZIONE

1) Definizione di caso
   - **a) Casi ben definiti con validità della diagnosi nota e riportata** ★
   - b) Casi ben definiti ma con selezione basata su fonti ufficiali o su *self reports*
   - c) Nessuna descrizione

2) Rappresentatività dei casi
   - **a) Serie consecutiva di casi insorti da una ben definita popolazione** ★
   - b) Potenzialmente affetti da distorsione da selezione (ad esempio serie di casi ospedalieri)
   - c) Nessuna descrizione

3) Definizione di controllo
   - **a) Controlli liberi dalla malattia oggetto di studio** ★
   - b) Nessuna descrizione

4) Selezione dei controlli
   - **a) Controlli di popolazione** ★
   - b) Controlli ospedalieri
   - c) Nessuna descrizione

Valutazione della qualità di uno studio caso-controllo (2)

CONFRONTABILITÀ

5) Procedure per assicurare la confrontabilità dei casi e dei controlli in fase di pianificazione (es. appaiamento) o analisi (es. stratificazione)

a) Confrontabilità assicurata da: _____ (indicare il fattore più importante) ★

b) Confrontabilità assicurata da: _____ (elencare i fattori secondari) ★

c) Nessuna procedura utilizzata

d) Nessuna descrizione

Valutazione della qualità di uno studio caso-controllo (3)

ESPOSIZIONE

6) Accertamento dell'esposizione

a) Record di buona qualità (ad esempio cartelle chirurgiche) ★

b) Interviste strutturate effettuate in condizioni di cecità dello stato di caso o controllo ★

c) Interviste non effettuate in condizioni di cecità

d) *Self reports* o *record* di qualità incerta (ad esempio cartelle cliniche)

e) Nessuna descrizione

7) Stesso metodo di accertamento dell'esposizione nei casi e nei controlli

a) Sì ★

b) No

8) Proporzione di non rispondenti

a) Stessa proporzione in ambedue i gruppi ★

b) Diversa proporzione nei due gruppi

c) Non riportato

NEWCASTLE-OTTAWA SCALE (NOS) – Studi caso-controllo

Vengono proposti **8 domande** classificate in **3 categorie**:

- **Selezione** (max 4 punti)
- **Confrontabilità** tra i gruppi (max 2 punti)
- **Esposizione** (max 3 punti)

Punteggio totale: ≤5 - Basso rischio bias

6-7 - Medio rischio bias

≥8 - Alto rischio bias

Studi di coorte, con coorte parallela

(M. Cinquini)

1) STUDI OSSERVAZIONALI

Studi *di coorte o longitudinali (prospettici)*

Caratteristiche:

-Sono i **più precisi**

-**affidabili**

-richiedono **alti costi**

-**Tempo elevato di studio**

Permettono di:

-Rilevare l'associazione tra **fattore di rischio e malattia**

-Calcolare l'**incidenza.**

1) STUDI OSSERVAZIONALI
Studi *di coorte o longitudinali (prospettici)*

- **Solo questi studi consentono il calcolo dell'incidenza** ovvero la probabilità di ammalarsi seguendo una coorte di persone nel tempo registrando tutte le patologie e i tempi di insorgenza, per questo sono studi **prospettici ovvero proiettati al futuro.**
- **Raramente si eseguono studi di coorte retrospettivi.**

1) STUDI OSSERVAZIONALI

Studi *di coorte o longitudinali, prospettici*

Obiettivi principali:

- Descrivere il ***cambiamento nel tempo di variabili*** quantitative in rapporto alla *intensità di esposizione* a possibili fattori di rischio
- Analizzare l'***associazione*** **di un possibile *fattore di rischio*** con l'incidenza futura della *malattia*
- **Indagare il destino** a distanza di tempo di pazienti trattati da differenti istituzioni sanitarie

1) STUDI OSSERVAZIONALI
Studi *di coorte o longitudinali (prospettici)*

- Si individua un determinato gruppo di soggetti da studiare e lo si divide in 2 gruppi (tassisti in città diverse)
- esposti e non esposti seguiti secondo un codificato **follow-up**
- si valuta nel tempo quanti soggetti tra questi 2 gruppi si ammalano di una patologia e quali sono le variabili responsabili della malattia

ES.

Possiamo verificare la sequenza temporale tra esposizione

e patologia: possiamo escludere il fumo come causa di un

tumore se il soggetto ha iniziato a fumare una settimana

prima di ammalarsi

1) STUDI OSSERVAZIONALI

Studi *di coorte o longitudinali (prospettici)*

SORGENTI di DISTORSIONE:

- **Perdite al follow-up**. Cercare di ottenere una % di follow-up di almeno il **90%**
- Conoscenza dell'esposizione o meno ai fattori di rischio può influenzare l'accertamento della malattia.
- **Non è utile per le malattie rare**
- Evitare fattori di confondimento: ***selection bias***

1) STUDI OSSERVAZIONALI
Studi *di coorte o longitudinali (prospettici)*

Vantaggi: ***metodo migliore*** per le indagini eziologiche:

- ***Tutti i casi di malattia*** o di complicazioni che si verificano in un periodo di tempo definito possono essere ***accertati***,
- Si possono ***calcolare*** direttamente ***i tassi di incidenza*** nei gruppi esposti in modo differente ai fattori di rischio in esame;
- La ***rilevazione*** dei ***fattori di rischio non*** può essere ***distorta*** dalla presenza della malattia e le loro modificazioni possono essere misurate.
- Si possono considerare **più fattori contemporaneamente**
- **Sono in grado di verificare la sequenza temporale di avvenimento tra esposizione e patologia**

1) STUDI OSSERVAZIONALI
Studi *di coorte o longitudinali (prospettici)*

Svantaggi

- ***lunga durata***, difficile e ***costoso***. È difficile mantenere costanti nel tempo le modalità di rilevazione

  – ***Non può saggiare ipotesi suggerite recentemente***

  – ***Non adatto*** per ***malattie rare*** nella coorte in esame.

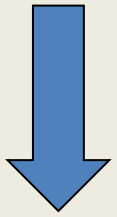
STUDI di COORTE:

	Malati (patologia respiratoria)	Sani	Totale
Esposti (residenti in città inquinate)	100	300	400
Non esposti (residenti in città meno inquinate)	60	240	300
Totale	160	540	700

1) STUDI OSSERVAZIONALI

Studi *di coorte o longitudinali (prospettici)*

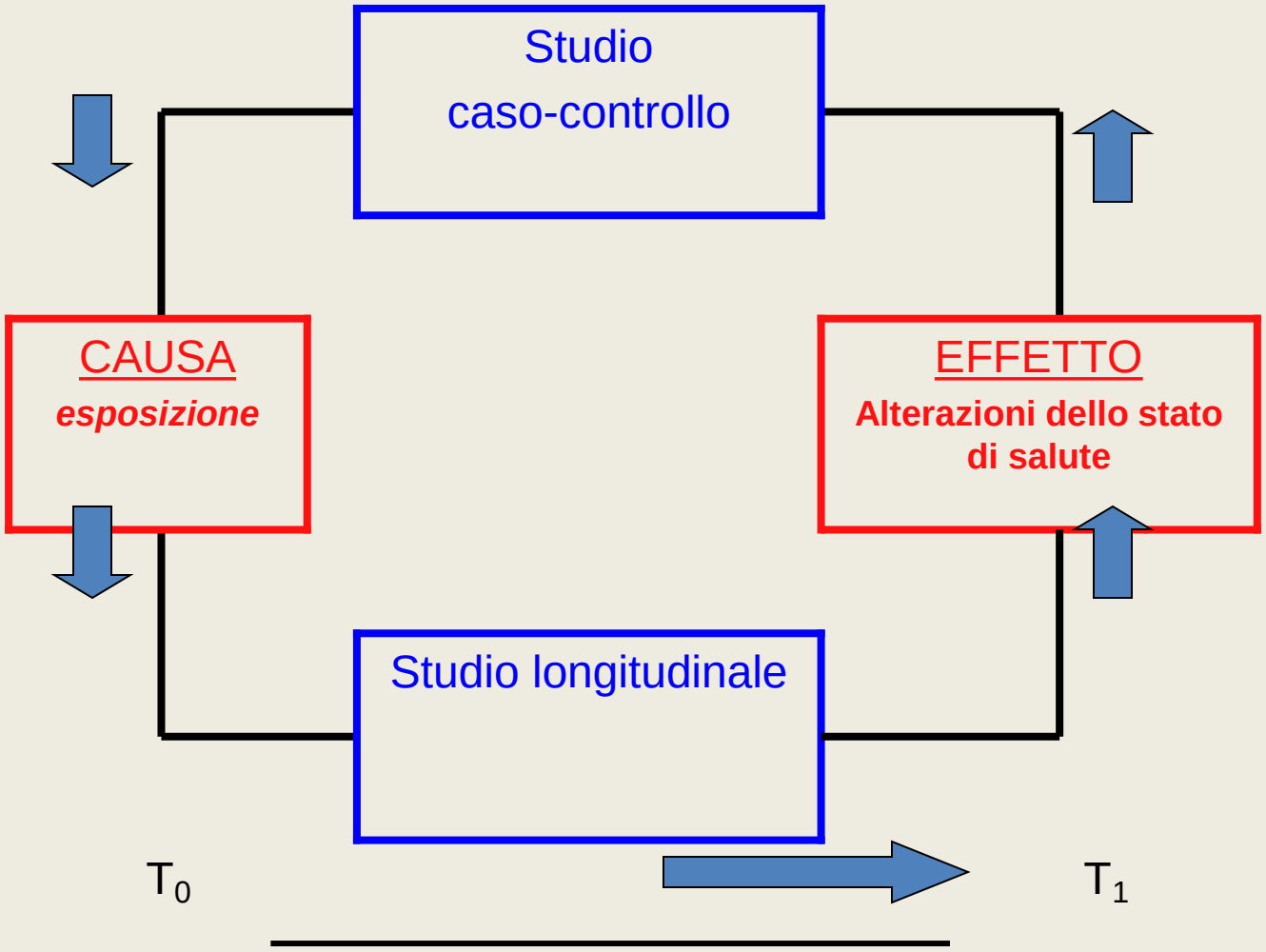
Tra tutti gli studi osservazionali solo gli studi di coorte permettono il calcolo dell'incidenza



Un gruppo di persone (COORTE) viene seguito nel tempo

e vengono registrate man mano tutte le patologie e il momento in cui sono state diagnosticate.

Relazione causa effetto in studi epidemiologici



Apriamo 2 nuovi capitoli

Per chiarire le misure epidemiologiche fondamentali che ricaviamo dalle tabelle 2x2 e che rappresentano i risultati finali degli studi epidemiologici ovvero le

EVIDENZE SCIENTIFICHE

Evidence-Based

- misure di frequenza e misure di associazione

Evidenze scientifiche = Evidence-bsaed

MISURE DI FREQUENZA:

Servono a **quantificare la frequenza delle malattie**

- PREVALENZA
- INCIDENZA
- Frequenza assoluta/relativa
- Tasso

MISURE DI ASSOCIAZIONE:

Quantificano l'associazione **tra esposizione e malattia**

- RISCHIO RELATIVO
- AFFINI

MISURE DI FREQUENZA: Servono a quantificare la frequenza delle malattie

Frequenza assoluta = es. numero di malati in città

numero di posti letto in UTIN

Frequenza relativa = es. numero di malati tra 2 città relativamente al numero di abitanti per le 2 città mediante l'uso di un denominatore e motiplicando il risultato per 100.

Tasso = studio epidemiologico in un determinato **periodo di tempo** (generalmente un anno) rapportato alla popolazione di riferimento

MISURE DI FREQUENZA: **PREVALENZA**

probabilità di essere malati in un momento o periodo permette di programmare ad es. il materiale necessario nelle corsie nel tempo successivo a quello studiato

PREVALENZA PUNTUALE
Di una patologia è il rapporto tra il numero di **malati presenti** in una popolazione in un certo **momento** ed il numero totale di individui di quella Popolazione (sani e malati) espressa in %

$$\frac{\text{nuovi malati}}{\text{malati+sani}}$$

PREVALENZA PERIODALE
È la prevalenza puntuale lungo un determinato **periodo** di tempo del numero dei malati vecchi e nuovi dell'anno considerato in %

$$\frac{\text{vecchi+nuovi malati}}{\text{malati+sani}}$$

MISURE DI FREQUENZA: **INCIDENZA**

come evolve la malattia

INCIDENZA CUMULATIVA

Stima della **probabilità di ammalarsi** in un determinato periodo di tempo. Misura solo i nuovi casi in rapporto alla popolazione totale dalla quale si escludono i vecchi malati che come tale non sono più a rischio di ammalarsi in %

$$\frac{\text{nuovi malati}}{\text{sani nel periodo in esame}}$$

TASSO o densità DI INCIDENZA

Stima della **probabilità di ammalarsi** in un determinato periodo di **tempo** in cui il campione dei malati non sia costante. **Fattore Tempo legato alle variabili** Individuali es. cambio residenza, Nuova malattia, rifiuto di visite etc.

$$\frac{\text{nuovi malati}}{\text{nei periodi presi in esame per tutta la popolazione}}$$

***MISURE DI ASSOCIAZIONE*:**

Quantificano l'associazione tra esposizione e malattia:

Rischio Relativo e affini

RISCHIO RELATIVO

Patologia o **evento** (**outcome**) **riferito a un fattore di rischio** o di esposizione
es. Tumore al polmone e fumo

Il Rischio relativo di sviluppare un tumore è 9 volte > che un non fumatore.

Così per un tumore tiroideo e l'esposizione alle radiziono ionizzanti

$$RR=\frac{\text{Incidenza di malattia negli esposti}}{\text{Incidenza di malattia nei non esposti}}$$

Nei fumatori RR>1 dei non fumatori il RR di ammalarsi sarà più alto **fattore di rischio**

Negli africani RR<1 degli europei per il melanoma **fattore protettivo**

Se RR=1 non c'è rischio di ammalarsi tra gli esposti e i non esposti **fattore indifferente**

MISURE DI ASSOCIAZIONE:

Quantificano l'associazione tra esposizione e malattia:

Rischio Relativo e affini

Rischio attribuibile

Espresso in % rappresenta
il **numero reale di malattia**
dovuto all'esposizione.

- Se l'incidenza negli esposti era il 25% e quella dei non esposti era il 20% l'incidenza vera dovuta all'esposizione è pari a 25-20%=5%

Errori di valutazione dovuti al

Bias e al confondimento

non esistono studi perfetti del tutto privi di Bias e/o confondimento e devono sempre essere considerati

BIAS errore puro in uno studio epidemiologico tra esposizione e malattia

CONFONDIMANTO quando l'associazione tra fattore di rischio e malattia è dovuta ad un altro fattore e non a quello da noi studiato.

BIAS=errore

- **BIAS di Selezione**

  Quando il campione selezionato per lo studio differisce notevolmente dalla popolazione di riferimento e ciò altera i risultati dello studio

- **Bias di informazione**

  Quando la malattia o altro è misurata diversamente nei gruppi studiati e ciò altera i risultati dello studio.
  Bias sul ricordo (**recall bias**)
  Bias x perdita dati di follow-up
  Bias dell'intervistatore
  Misclassification Bias quando un malato o un sano vengono confusi in sani o malati e non lo sono

Confondimento o Bias ecologico

- Quando una **variabile annulla, riduce o aumenta** l'associazione tra esposizione e patologia.
- es. Caffè/fumo malattia cardiovascolare il fumo è un fattore di confondimento perché è un fattore di rischio x mal. Cardiovascolari ed è più diffuso tra i bevitori di caffè.

Controllo x Bias e Confondimento

Scelta del campione possibilmente a caso **l'1% del totale**

- Es. ospedalizzati nello stesso istituto
- Individui simili rappresentativi della popolazione in esame

- **Raccolta dati** mediante un **questionario possibilmente già testato e riportato in letteratura** oppure adottarne uno **nuovo** motivandolo. Inoltre sarebbe ottimale **raccogliere dati oggettivi** come ricerca Acantivirali nel sangue … e fidarsi poco delle interviste.

Controllo x Bias e Confondimento

- **Randomizzazione** del campione ovvero scelta casuale
- **Restrizione** del campionamento in rapporto all'età
- **Appaiamento o matching** controlli appaiati ai casi per età, sesso, attitudini …
- **Analisi statistica stratificata** si suddivide il gruppo x più variabili analizzate separatamente

- **Analisi statistica multivariata** prende in esame più di 2 variabili simultaneamente con un modello di analisi che viene adattato tramite complessi **algoritmi statistici**. Non è utilizzabile per studi di pochi campioni

Risk o bias

- **Quali checklist scegliere per la valutazione della qualità metodologica/rischio di bias?**

The variety of study designs classified as NRS, and their varying susceptibility to different biases, makes it difficult to produce a generic robust tool that can be used to evaluate risk of bias. Inclusion of a knowledgeable methodologist in the team is essential to identify the key areas of weakness in the included study designs. (Cochrane Handbook Higgins 2011)

doi:10.1093/ije/dym018

Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography

Simon Sanderson,[1]* Iain D Tatt[2,4] and Julian PT Higgins[3]

Methods Tools were identified from a search of three electronic databases, bibliographies and an Internet search using Google®. Two reviewers extracted data using a pre-piloted extraction form and strict inclusion criteria. Tool content was evaluated for domains potentially related to bias and was informed by the STROBE guidelines for reporting observational epidemiological studies.

Results A total of 86 tools were reviewed, comprising 41 simple checklists, 12 checklists with additional summary judgements and 33 scales. The number of items ranged from 3 to 36 (mean 13.7). One-third of tools were designed for single use in a specific review and one-third for critical appraisal. Half of the tools provided development details, although most were proposed for future use in other contexts. Most tools included items for selection methods (92%), measurement of study variables (86%), design-specific sources of bias (86%), control of confounding (78%) and use of statistics (78%); only 4% addressed conflict of interest. The distribution and weighting of domains across tools was variable and inconsistent.

Conclusion A number of useful assessment tools have been identified by this report. Tools should be rigorously developed, evidence-based, valid, reliable and easy to use. There is a need to agree on critical elements for assessing susceptibility to bias in observational epidemiology and to develop appropriate evaluation tools.

Check list per risk of bias of NRS

- Cohort studies: New Castle Ottawa scale (Wells 2008) for cohort studies;
- Case control studies: New Castle Ottawa scale for case control
- Cross sectional surveis: New Castle Ottawa scale for cross sectional
- Controlled before after studies: criteria of the Cochrane EPOC group (Cochrane Effective Practice and Organisation of Care) (revised 2015)
- Interrupted time series analysis: criteria of the Cochrane EPOC group

New Castle Ottawa Scale – cohort

Selection

1) **Representativeness of the exposed cohort**

- a) truly representative of the average _____ (describe) in the community +
- b) somewhat representative of the average _____ in the community +
- c) selected group of users eg nurses, volunteers
- d) no description of the derivation of the cohort

2) **Selection of the non exposed cohort**

- a) drawn from the same community as the exposed cohort +
- b) drawn from a different source
- c) no description of the derivation of the non exposed cohort

3) **Ascertainment of exposure**

- a) secure record (eg surgical records) +
- b) structured interview +
- c) written self report
- d) no description

4) **Demonstration that outcome of interest was not present at start of study**

- a) yes +
- b) no

New Castle Ottawa Scale – cohort

Comparability

1) Comparability of cohorts on the basis of the design or analysis

- a) study controls for _____ (select the most important factor) +
- b) study controls for any additional factor +

Outcome

1) Assessment of outcome

- a) independent blind assessment +
- b) record linkage +
- c) self report
- d) no description

2) Was follow-up long enough for outcomes to occur

- a) yes (select an adequate follow up period for outcome of interest)
- b) no

3) Adequacy of follow up of cohorts

- a) complete follow up - all subjects accounted for
- b) subjects lost to follow up unlikely to introduce bias - small number lost - > ____ % (select an adequate %) follow up, or description provided of those lost)
- c) follow up rate < ____% (select an adequate %) and no description of those lost
- d) no statement

Nuovo Risk of bias tool for NRS (ROBINS-I)

Sterne JA, Hernán MA, Reeves BC et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016 Oct 12;355:i4919. doi: 10.1136/bmj.i4919.

Può essere utilizzato per più disegni di studi (studi di coorte, studi controllati prima dopo, studi caso controllo, studi cross sectional)

Va bene solo per studi che hanno come obbiettivo quello di valutare l'effetto (efficacia) di un intervento

1° step: define the 'ideal' RCT

The Cochrane 'risk of bias' (RoB) tool for NRS is concerned with evaluating the risk of bias in the results of non-randomized studies that compare the health effects of two or more interventions.

Facilitated by considering each NRS as an attempt to emulate (mimic) a hypothetical randomized trial that compares the health effects of two or more interventions.

- If confounding is successfully controlled, the effect estimates from the observational study will be identical, except for sampling variation, to those from a target trial that randomly assigns individuals in the same study population to either intervention A or B.
- **The risk of bias arising from the observational design is a function of how imperfectly the observational study emulates the target trial.**

1° step: define the 'ideal' RCT

We refer to such a hypothetical randomized trial as the "**target randomized trial"**.

At the protocol stage

define hypothetical "**target randomized trial"**, the RCT that would be "ideal " to answer the review question

Participants

Intervention

Comparator

Outcomes (benefits and/or harms)

2° step: Specify whether interested in the effect of initiating (ITT) or initiating and adhering to (per protocol) intervention

- When the effect of interest is that of **assignment to the intervention** at baseline (randomized trials) or starting intervention at baseline (NRSs), **risk of bias assessments for both types of study need not be concerned with post-baseline departures from intended interventions** that reflect the natural course of events
- When the effect of interest is the **per protocol effect**, risk of bias assessments of both randomized and nonrandomized studies may **have to consider intervention discontinuation, switches between interventions, or departures from intended interventions**.

3°step: Identify possible confounding domains

- A confounding domain is a pre-intervention prognostic factor that predicts whether an individual receives one or the other intervention of interest. Some common examples are severity of pre-existing disease, physician prescribing practices, health care utilization, adiposity, and socioeconomic status.
- We recommend that **subject-matter experts be included in the team** writing the review protocol, and encourage the **listing of confounding domains in the review protocol**, based on initial discussions among the review authors
  - **At protocol stage list the confounding domains relevant to all or most studies eligible for the review**

4°step: Identify possible co-interventions

- Relevant co-interventions are the interventions or exposures that individuals might receive after or with initiation of the intervention of interest, which are related to the intervention received and which are prognostic for the outcome of interest.
- These are also likely to be identified through the expert knowledge of members of the review group, via initial (scoping) reviews of the literature, and after discussions with health professional
  - **At protocol stage list the possible co-interventions that could differ between intervention groups and have an impact on study outcomes**.

Risk of bias tool - 7 domains

Pre-intervention

1. Bias due to baseline confounding
2. Bias in selection of participants into the study

At intervention

3. Bias in measurement of the interventions

Post-intervention

4. Bias due to departures from intended interventions (performance bias)
5. Bias due to missing data (attrition bias)
6. Bias in measurement of outcomes or Interventions (detection bias)
7. Bias in selection of the reported result (outcome reporting bias)

ROBINS-I

Pre-intervention

1. Bias due to baseline confounding
2. Bias in selection of participants into the study

At intervention

3. Bias in classification of the interventions

Post-intervention

4. Bias due to departures from intended interventions (performance bias)
5. Bias due to missing data (attrition bias)
6. Bias in measurement of outcomes or Interventions (detection bias)
7. Bias in selection of the reported result (outcome reporting bias)

Pre or at intervention features for which consideration of bias in NRS are mainly distinct from those in RCTs

Post intervention features for which many considerations are similar to those in RCTs

Signalling questions

To help reviewer… for each domain some signalling question have been proposed

- **Signalling questions**: possible answers:

Yes

Probably yes

Probably no

No

No information

Responses of ‘Yes’ and ‘Probably yes’ (also of ‘No’ and ‘Probably no’) have similar implications.

Judgment of risk of risk of bias

Assessment must be done **at the outcome level**: (e.g. 5 outcomes in the review and 10 included studies: for each study you should assess risk of bias separately for each outcome , i.e. 5 times; total 50 risk of bias table….)

- **5°step**: assess risk of bias for a given outcome for each of the 7 domain
- **6° step**: make an overall judgment of risk of bias for that outcome at the study level
- **7° step**: make an overal judgment of risk of bias for that outcome across all the studies
- **8° 9° 10° etc step…** repeat all of these for each outcome

6° step : Judgments at each domain level

Low risk of bias (the study is comparable to a well-performed randomized trial with regard to this domain);

Moderate risk of bias (the study is sound for a non-randomized study with regard to this domain but cannot be considered comparable to a well-performed randomized trial);

Serious risk of bias (the study has some important problems in this domain);

Critical risk of bias (the study is too problematic in this domain to provide any useful evidence);

No information on which to base a judgment about risk of bias for this domain.

6° step : overall judgment at the study level for each outcome

RESPONSE OPTION	CRITERIA
Low risk of bias (the study is comparable to a well-performed randomized trial);	The study is judged to be at **low risk** of bias for **all domains**.
Moderate risk of bias (the study appears to provide sound evidence for a non-randomized study but cannot be considered comparable to a well-performed randomized trial);	The study is judged to be at **low or moderate risk of bias for all domains**.
Serious risk of bias (the study has some important problems);	The study is judged to be at **serious risk of bias** in at least one domain, but not at critical risk of bias in any domain.
Critical risk of bias (the study is too problematic to provide any useful evidence on the effects of intervention);	The study is judged to be at **critical risk of bias in at least one domain.**
No information on which to base a judgement about risk of bias.	There is no clear indication that the study is at serious or critical risk of bias *and* there is a lack of information in one or more key domains of bias (*a judgement is required for this*).

Table 2. Reaching an overall RoB judgement for a specific outcome.

Declaring a study to be at a particular level of risk of bias for an individual domain will mean that the study as a whole has a risk of bias at least this severe (for the outcome being

7° step: overall judgment across all studies for the given outcome (following the GRADE approach)

- Outcome specific
- Do not average risk quality across the studies
- Evaluate the extent to which each trial contributes toward the estimate of magnitude of effect. This contribution will usually reflect study sample size and number of outcome events -larger studies with many events will contribute more, much larger studies with many more events will contribute much more (look at the weight of each study in the forest plot)

8°and further step: overall judgment across all studies if you have several outcomes

Domain	O1	O2	O3
Bias due to confounding	Serious risk	Moderate risk	Serious risk
Bias in selection of participants into the study	Low risk	Low risk	Low risk
Bias in measurement of interventions	Low risk	Low risk	Low risk
Bias due to departures from intended interventions	Moderate risk	Moderate risk	Moderate risk
Bias due to missing data	Low risk	No info	No info
Bias in measurement of outcomes	Low risk	Low risk	Serious risk
Bias in selection of the reported result	Moderate risk	Moderate risk	Serious risk
Overall*	*Serious risk*	*Moderate risk*	*Serious risk*

Studi di *real world*

(O. Nanni)

- Scaletta 45 min

- Rct gold standrad
- Validita interna esterna e generalizzazione
- Random itt
- Ciclo statsitica
- Pop obiettivo bersagliodifferenza reclutamento studi real world
- % reclutamneto
- Come incrementare reclutamento TRINETX
- Disponibilita dati
- Competenze
- Data scientist
- Privacy

Demographic recruitment bias

Randomized Clinical Trial

- Recognized as the gold standard for conclusive evidence generation by scientists, clinicians, regulators, healthcare decision makers
  - Requires at least two groups (treatment and control) with randomized assignments and crossovers
  - Requires that the clinicians are blinded to remove bias and deliver objective results
  - Requires rigorous statistical analysis for appropriate design and accurate analysis of results
  - Requires significant funding, it takes a long time to enroll and follow
  - It has inherent risks to patients, to clinics, to sponsors

Quando non è possibile fare un RCT

- Verifica di tossicità di una esposizione/trattamento
- gli esiti da valutare sono troppo lontani nel tempo
- gli esiti da valutare sono rari (servirebbe un campione troppo ampio)
- selezione dei pazienti: preferenza per uno dei trattamenti, con conseguente rifiuto alla randomizzazione
- fondi non sufficienti

A. Liberati Master 2002

RESEARCH

the bmj | BMJ 2018;363:k5094 | doi: 10.1136/bmj.k5094

Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial

Robert W Yeh,[1] Linda R Valsdottir,[1] Michael W Yeh,[2] Changyu Shen,[1] Daniel B Kramer,[1] Jordan B Strom,[1] Eric A Secemsky,[1] Joanne L Healy,[1] Robert M Domeier,[3] Dhruv S Kazi,[1] Brahmajee K Nallamothu[4] On behalf of the PARACHUTE Investigators

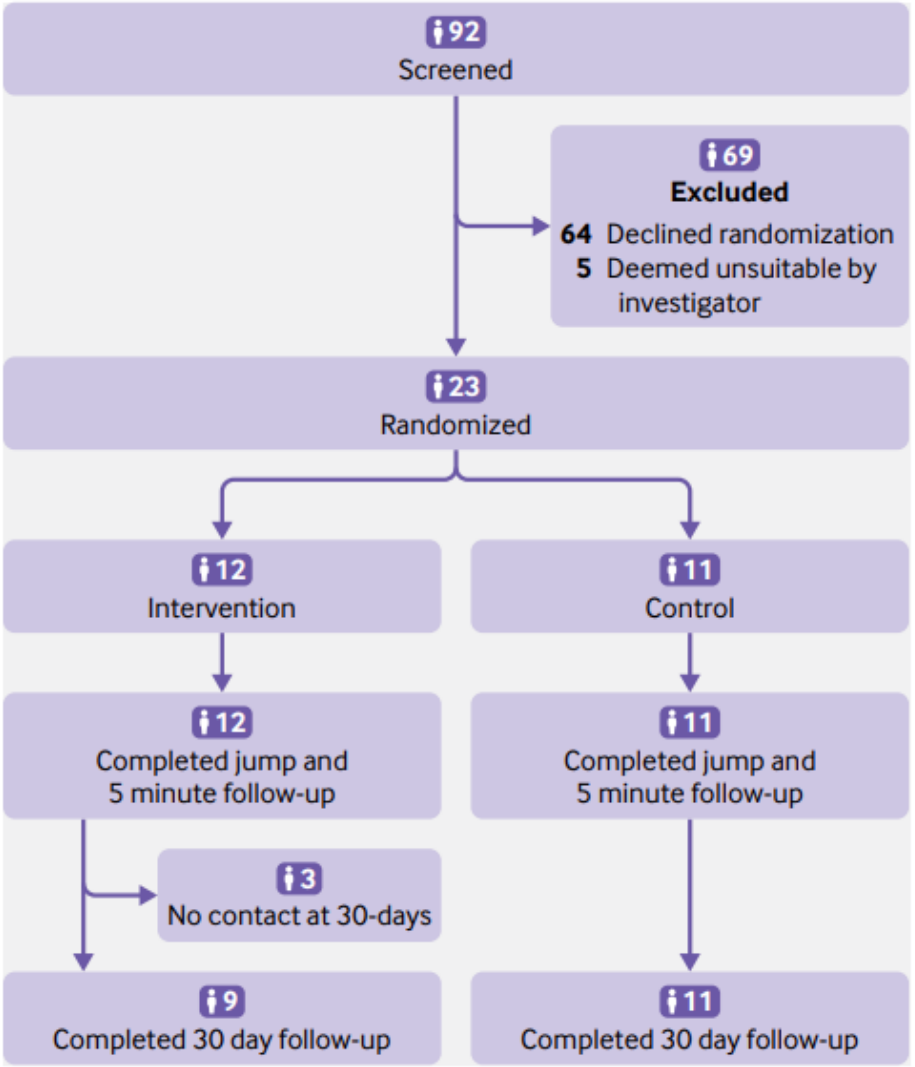
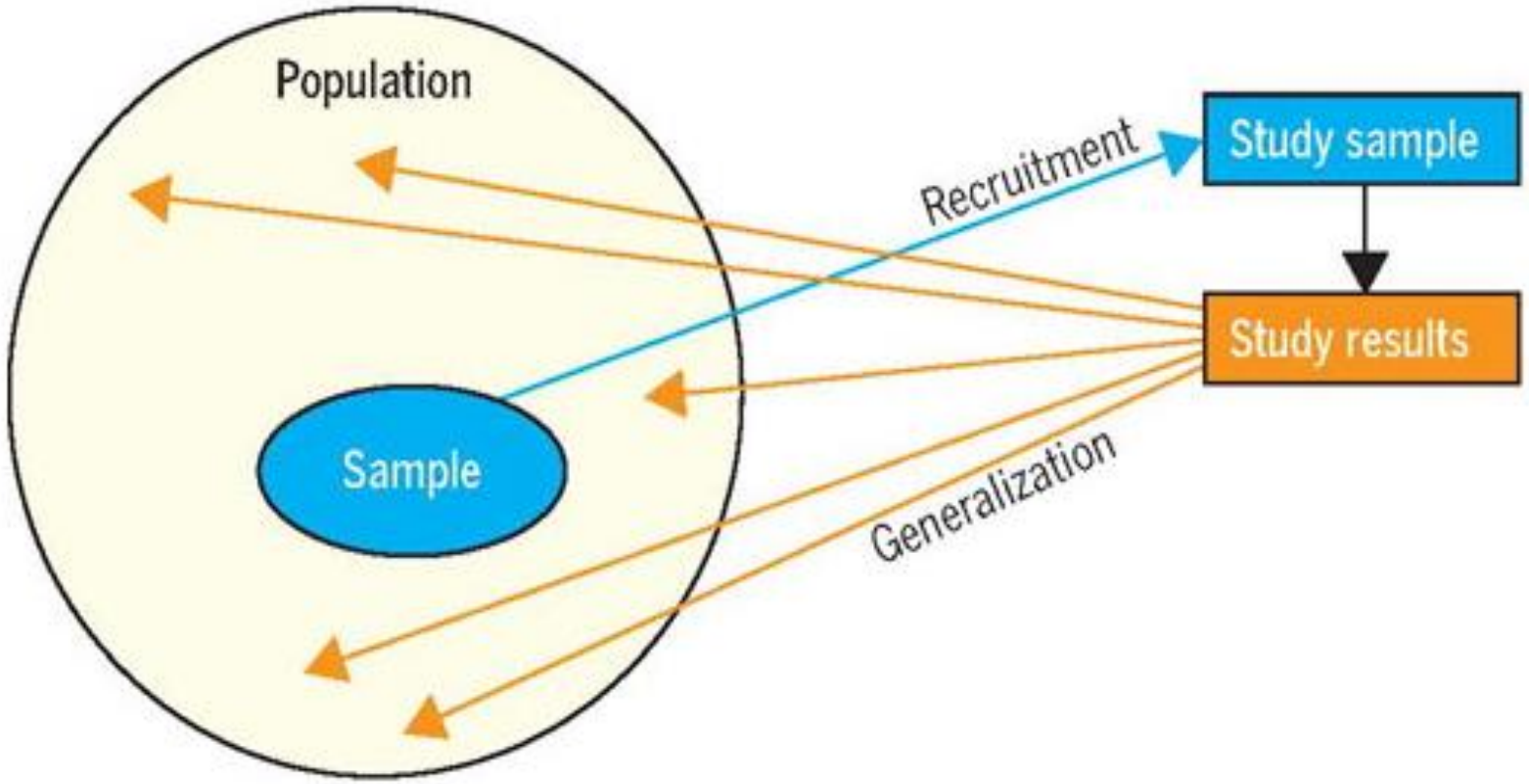


Fig 1 | Study flow diagram


Population
Sample
Recruitment
Study sample
Study results
Generalization

La randomizzazione …

Non assicura che i gruppi siano equivalenti dal punto di vista medico ma che i fattori prognostici noti e non siano distribuiti a caso

Senza la randomizzazione "una differenza statisticamente significativa" può essere il risultato di differenze non casuali nella distribuzione dei fattori prognostici sconosciuti

=

Garantisce la validità dei metodi statistici

Intention to treat

Tutti i pazienti randomizzati nell'analisi ed assegnati al gruppo definito dalla randomizzazione

Indipendentemente da:

- **Criteri di eligibilità**
- **Trattamento effettivamente ricevuto**
- **Fuoriuscite o deviazioni dal protocollo**

Pazienti con carcinoma mammario operabile

Pazienti eleggibili per lo studio

(T_{1-3a} N_+ M_0)

Pazienti non eleggibili

Pazienti che danno il consenso allo studio

Pazienti che rifiutano il consenso

Registrazione delle pazienti allo studio

a. **Pazienti che ricevono il trattamento correttamente**

b. **Pazienti che rifiutano di proseguire il trattamento**

c. **Pazienti che ricevono il trattamento in modo discontinuo**

d. **Pazienti perse al follow-up**

ANALISI DEI RISULTATI

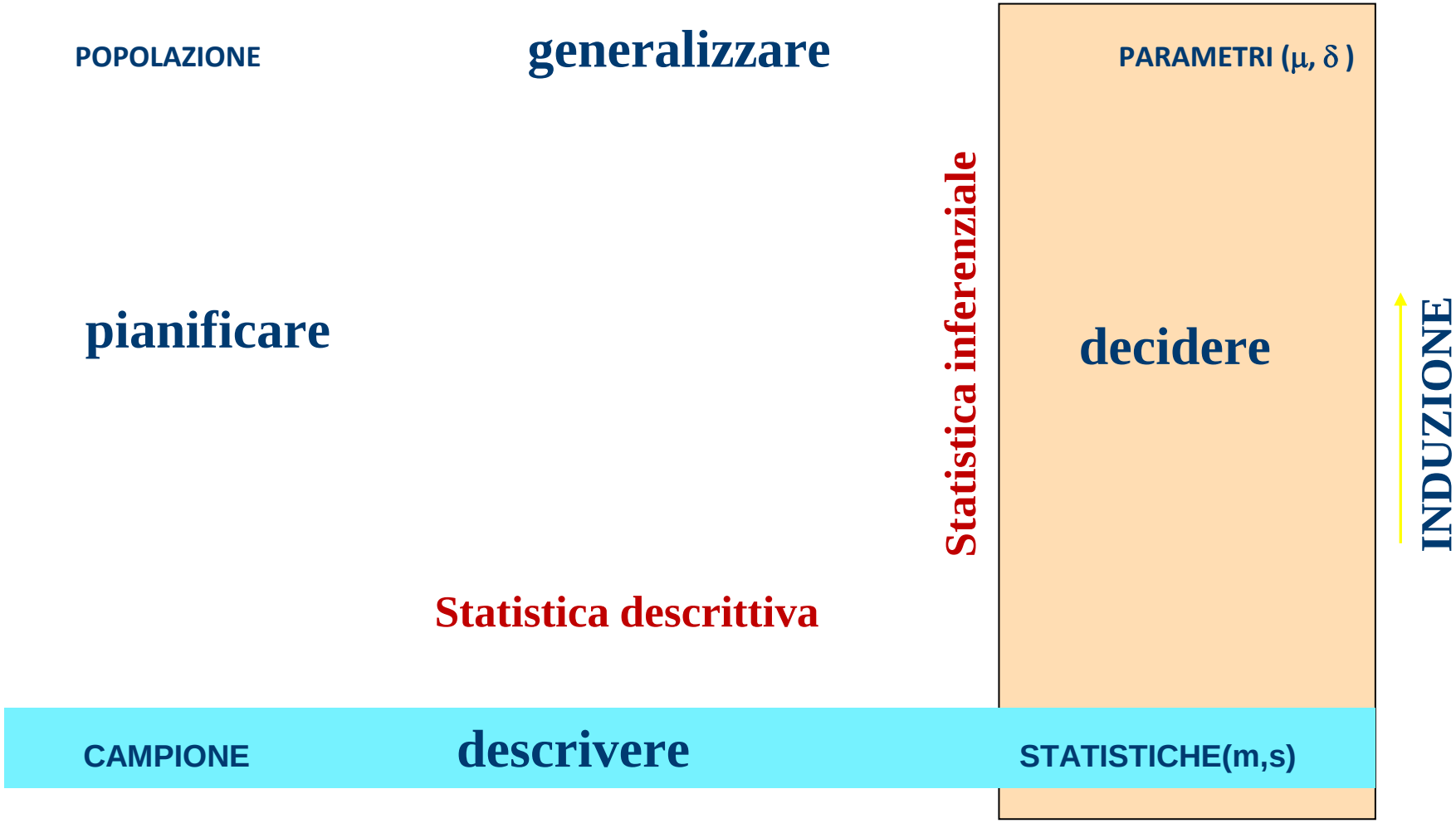
Intention To Treat : (a + b + c + d)

Per Protocol: a

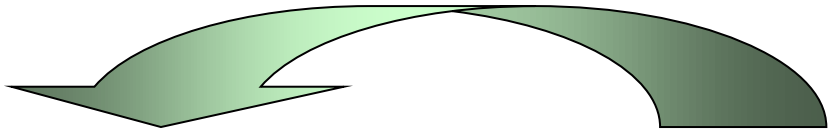
Modificato da Bonadonna

Bonadonna-Valagussa

Il ciclo logico dell'analisi statistica (GA Maccacaro)



I due grandi rami della Statistica



Statistica Descrittiva

• Raccolta dei dati

• Presentazione dei risultati e sintesi dei dati di un campione (più raramente di una popolazione)

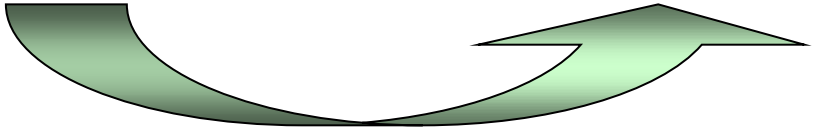
Statistica Inferenziale

• Metodo induttivo

(dal particolare al generale)

• Stima dei parametri di una popolazione ignota

• Verifica delle ipotesi

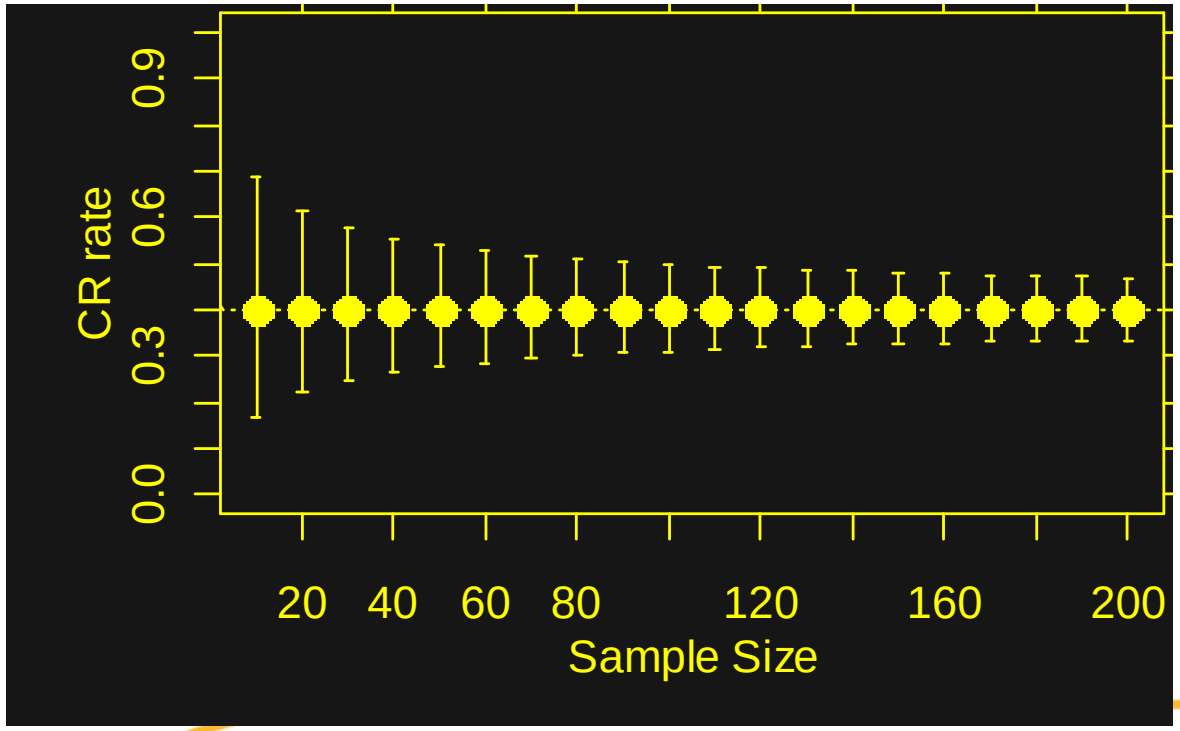


Estimation: a point estimate and some measure of precision

Example: “The CR rate in each arm will be estimated with its confidence interval.”

- **this provides us with an estimate of the CR rate**
- **it also provides us with a measure of precision about the estimate 95% confidence interval , an interval that contains the true value of the parameter of interest 95% of the time. “we are 95% confident that the true CR lies in this interval”**

Example: observed CR rate is 0.40. 95% confidence interval width depends on the sample size



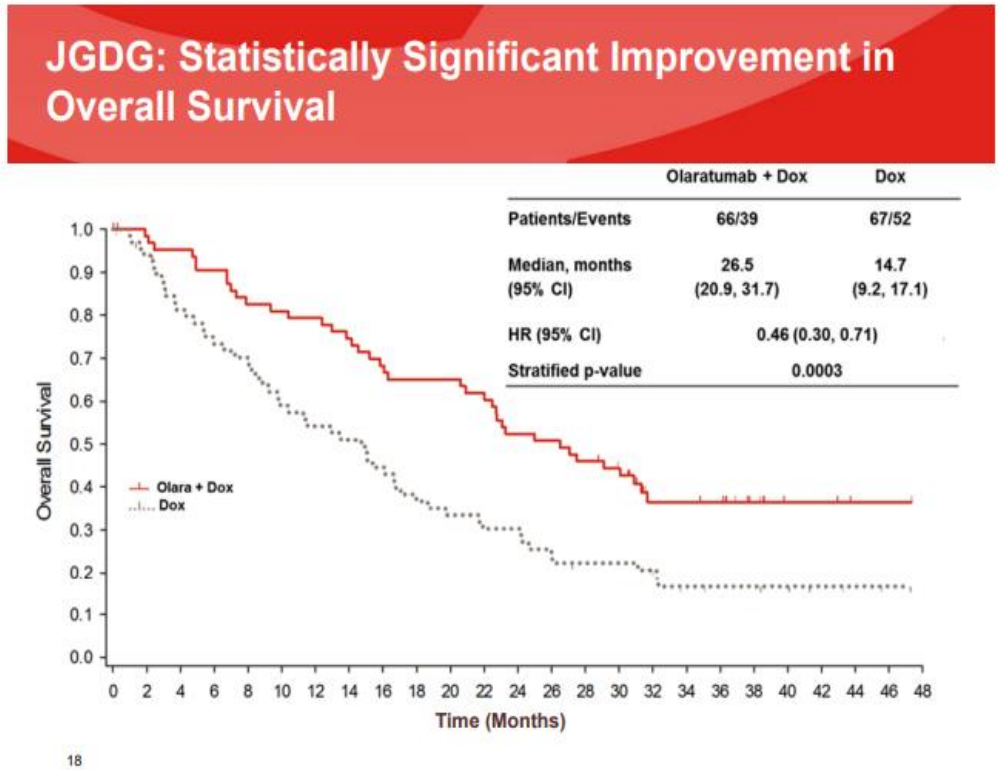
Hypothesis-Testing

Example: Overall survival (Kaplan Meier curve)

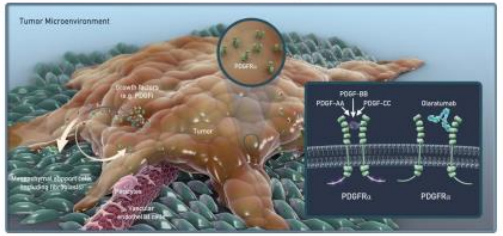
H0: Hazard Ratio =1 (null)

H1: HR= 1 (alternative)

- This test is performed using log rank
- The result is a p-value that provides “evidence” to either reject or fail to reject the null hypothesis



Mechanism of Action of Olaratumab



- Olaratumab is a fully human monoclonal antibody (mAb) of immunoglobulin G class 1 (IgG1) that binds PDGFRα and blocks PDGF-AA, -BB, and -CC, inhibiting ligand-induced receptor phosphorylation and downstream signaling

10

JGDG: Open-label, Multicenter, Phase 1b/2 Trial (led to accelerated approval)

Phase 1b

- Advanced STS, not amenable to surgery or radiotherapy
- Age ≥ 18 years; ECOG PS ≤ 2
- Any number of prior treatments; no doxorubicin

N=15; 21-day cycles

Cycles 1-8: olaratumab 15 mg/kg D1,8 + doxorubicin 75 mg/m² D1 for 8 cycles

Subsequent cycles: olaratumab monotherapy if benefit

Primary endpoint: Safety

Phase 2

- New patients, same entry criteria as in Phase 1b
- Dynamic minimization used to balance arms with respect to PDGFRα, ECOG PS, line of treatment, and histology (LMS, synovial, other)

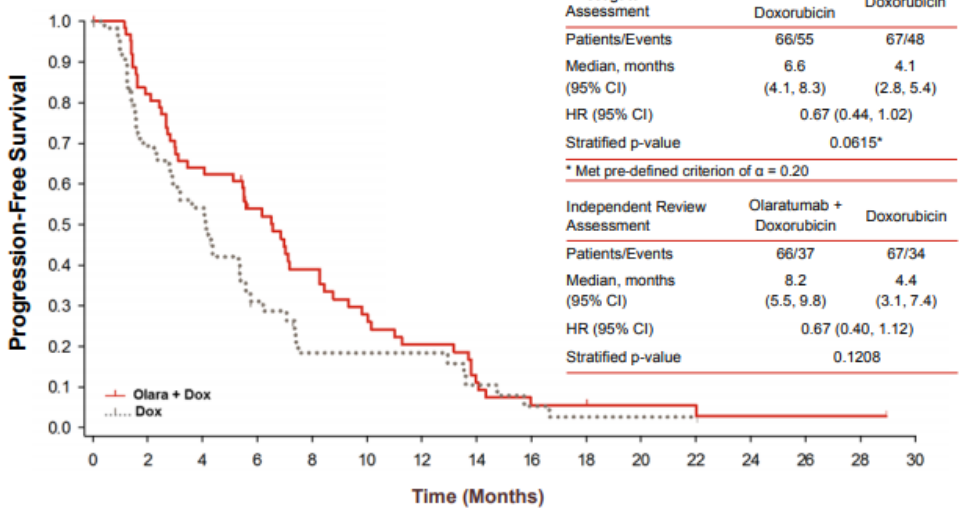
N=130; 21-day cycles

RANDOMIZE

Olaratumab 15 mg/kg D1,8 + doxorubicin 75 mg/m² D1 for 8 cycles → Olaratumab monotherapy until progression

Doxorubicin 75 mg/m² D1 for 8 cycles → Option to receive olaratumab monotherapy after progression

JGDG: Study Met Protocol-defined Significance Level for PFS per Investigator Assessment



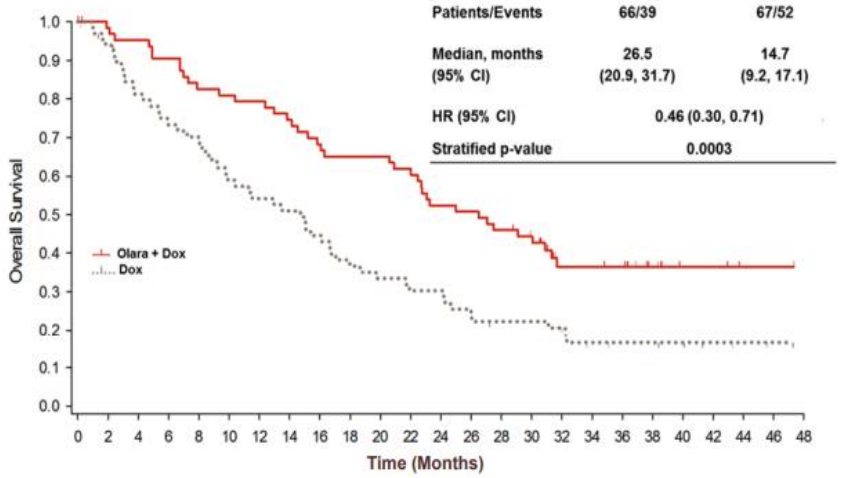
Investigator Assessment	Olaratumab + Doxorubicin	Doxorubicin
Patients/Events	66/55	67/48
Median, months (95% CI)	6.6 (4.1, 8.3)	4.1 (2.8, 5.4)
HR (95% CI)	0.67 (0.44, 1.02)	
Stratified p-value	0.0615*	

* Met pre-defined criterion of α = 0.20

Independent Review Assessment	Olaratumab + Doxorubicin	Doxorubicin
Patients/Events	66/37	67/34
Median, months (95% CI)	8.2 (5.5, 9.8)	4.4 (3.1, 7.4)
HR (95% CI)	0.67 (0.40, 1.12)	
Stratified p-value	0.1208	

17

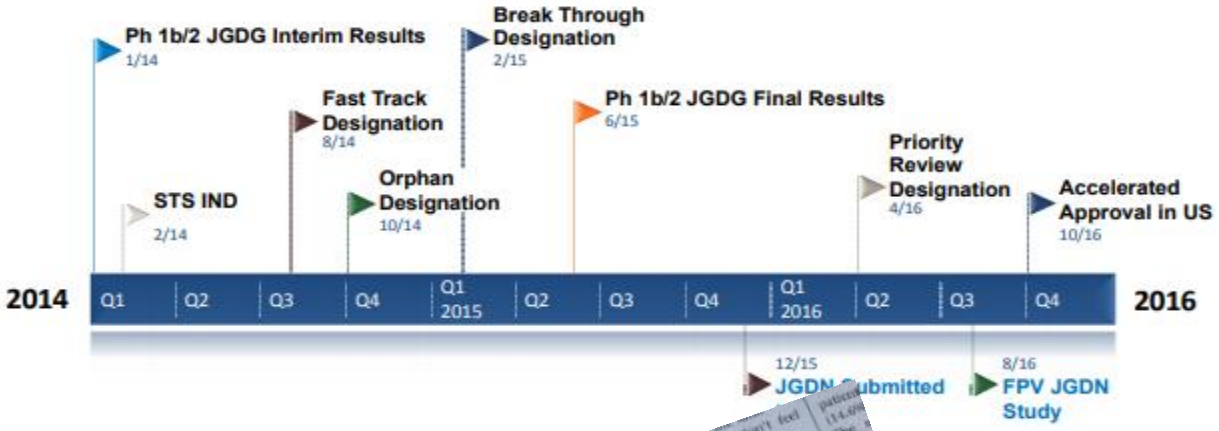
JGDG: Statistically Significant Improvement in Overall Survival



	Olaratumab + Dox	Dox
Patients/Events	66/39	67/52
Median, months (95% CI)	26.5 (20.9, 31.7)	14.7 (9.2, 17.1)
HR (95% CI)	0.46 (0.30, 0.71)	
Stratified p-value	0.0003	

18

Extensive FDA Interactions Leading to Accelerated Approval



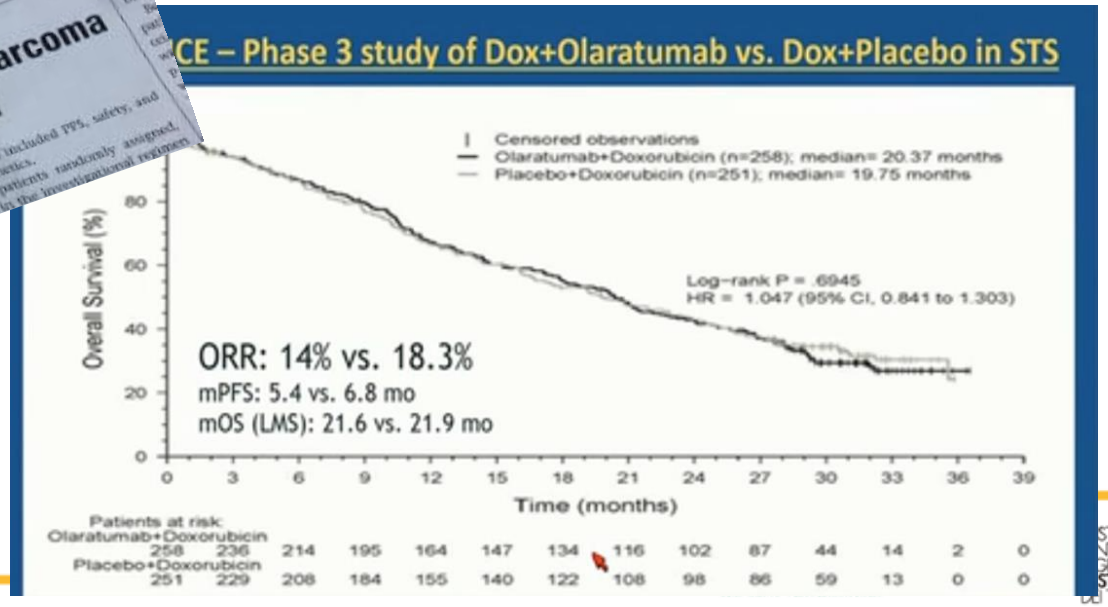
Olaratumab Phase III Trial in Soft Tissue Sarcoma Fails to Meet Endpoint of Overall Survival

The ANNOUNCE trial, assessing the overall survival (OS) benefit of olaratumab in patients with advanced soft tissue sarcoma (STS)

discrepancy between results in OS and PFS, the study's primary endpoint; and that the drug had an unknown mechanism of action in sarcoma subtypes.

ary endpoints included PFS, safety, and pharmacokinetics.

Of 509 patients randomly assigned, 258 were in the investigational regimen



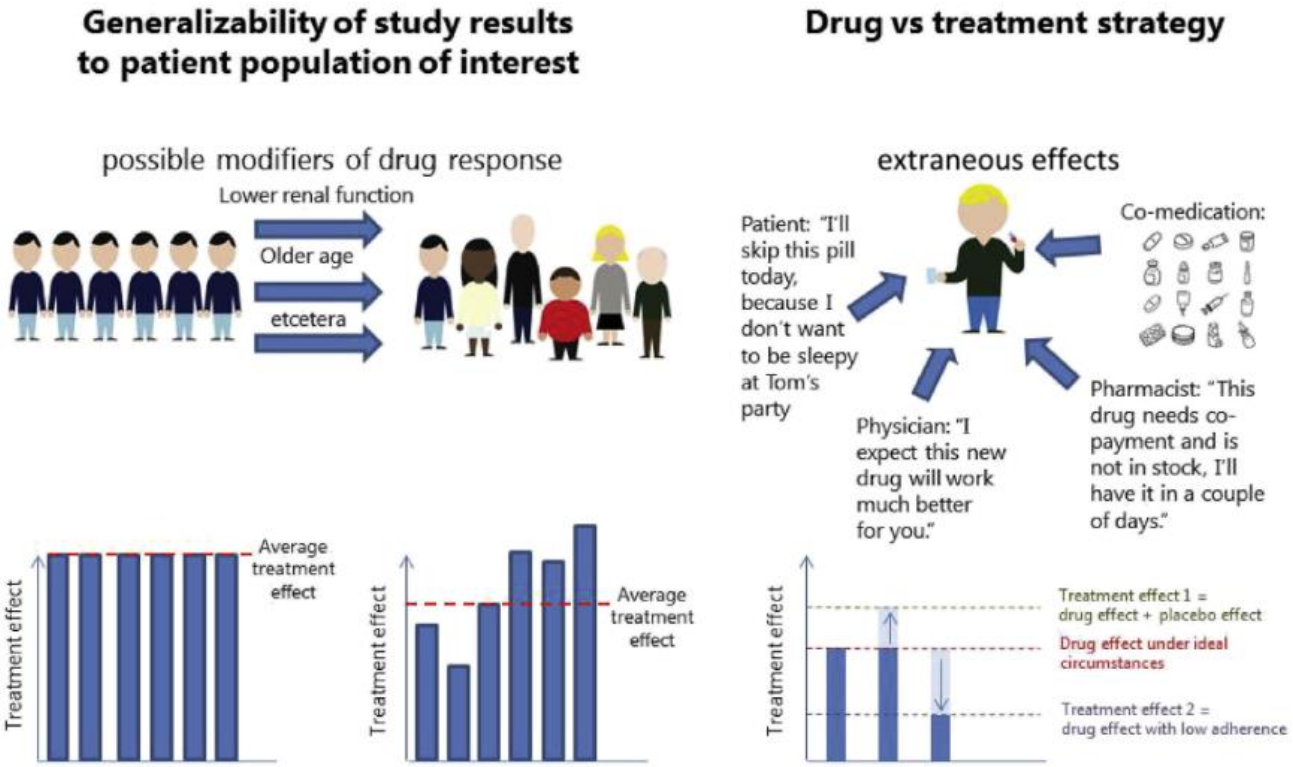


Fig. 1. Pragmatic trial design.

complies with a pragmatic approach (choosing between Although this approach maximizes the possibility to reveal

Why Diverse Clinical Trial Participation Matters

Aaron L. Schwartz, M.D., Ph.D., Marcella Alsan, M.D., Ph.D., Alanna A. Morris, M.D., and Scott D. Halpern, M.D., Ph.D.

n engl j med 388;14 nejm.org April 6, 2023

Goals of Increasing Diversity in Clinical Trials.

Goal	Key Challenges	Implications
Building trust in medical research and institutions	Distrust of medical and scientific professions can be an important obstacle to receiving effective medical care.	The effect on public trust of the design and conduct of clinical trials can be as important to public health as trials' results. Investments should be made in elucidating how clinical trial practices affect public trust.
Promoting fairness for potential participants and their communities	Opportunities to participate in trials are limited. Preferences, resources, and trust all affect willingness to participate in trials. Health systems' capacities to conduct trials vary among communities.	Overcoming unjust barriers to participation for disenfranchised groups will require affirmative outreach and recruitment actions. Grading trials on inclusive outreach and recruitment practices, rather than solely enrollment demographics, may better reflect recruitment equity. Investing in trial capacity in marginalized communities may benefit such communities broadly by improving adoption of innovations.
Generating biomedical knowledge	Sample sizes are often too small to permit assessment of treatment efficacy within particular subgroups. Clinically significant differences in treatment efficacy between groups that are underrepresented and those that are overrepresented in trials may not be common. Efforts to diversify trials address only some of the barriers to efficient patient recruitment.	Investigators should acknowledge that more inclusive trials may not show whether a treatment is effective for certain patient subgroups or meaningfully shift estimates of the treatment's efficacy. Shifting the focus of trials to diseases that disproportionately affect marginalized groups may more effectively generate knowledge benefiting these groups. Future meta-research could clarify the importance and detectability of heterogeneous treatment effects.

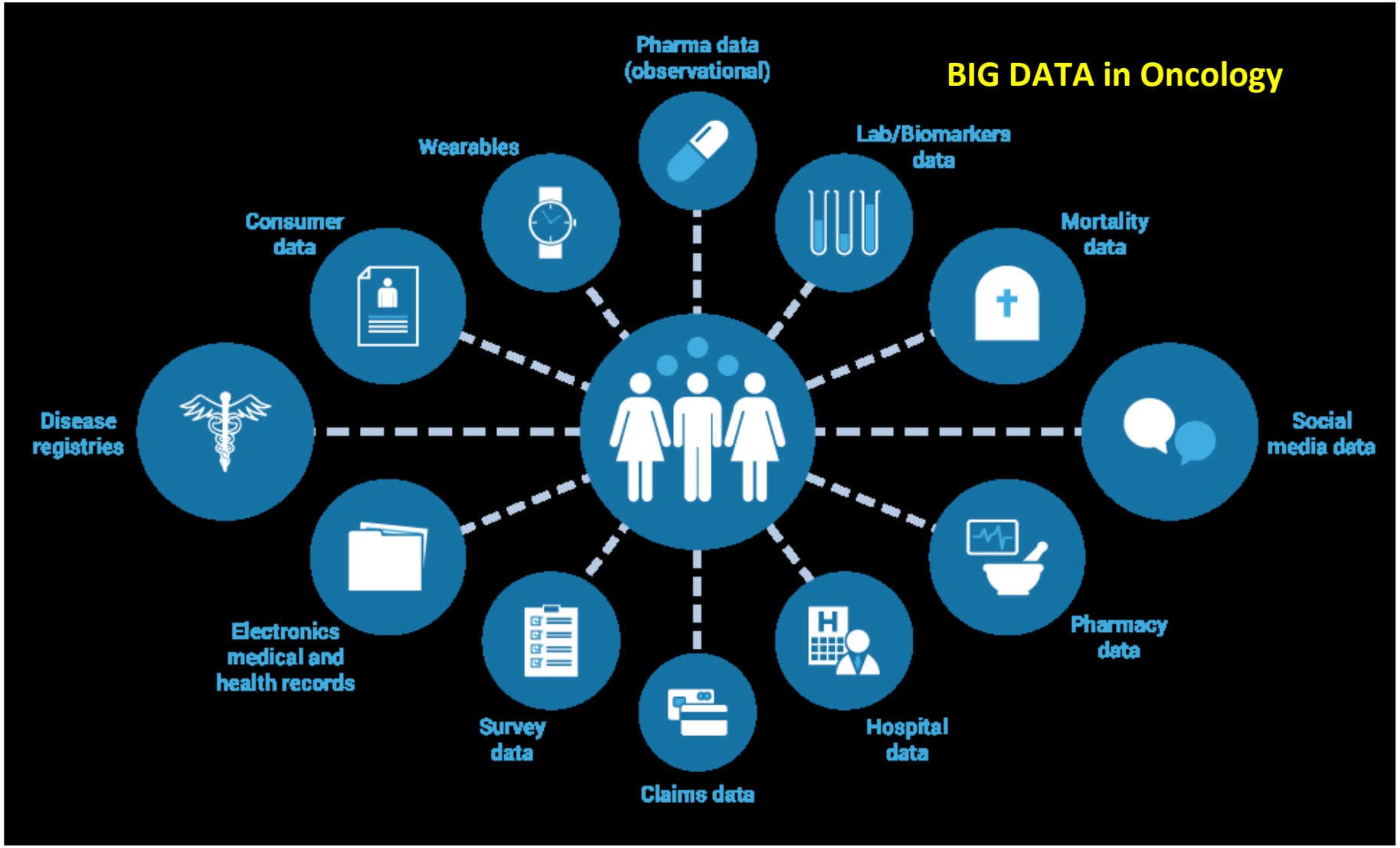
RCT are used to evaluate treatment outcomes.
An important, complementary source of knowledge is “real-world” data.
RWD contribute to our understanding of treatment effectiveness and safety, disease and treatment patterns, and patient behaviors in everyday clinical practice.
These data are collected outside of the controlled environment of conventional RCTs.


FIRST OPINION

Real-world evidence is changing the way we study drug safety and effectiveness

By NANCY A. DREYER / JANUARY 29, 2019

ADOBE


BIG DATA in Oncology
Pharma data (observational)
Wearables
Lab/Biomarkers data
Consumer data
Mortality data
Disease registries
Social media data
Electronics medical and health records
Pharmacy data
Survey data
Hospital data
Claims data

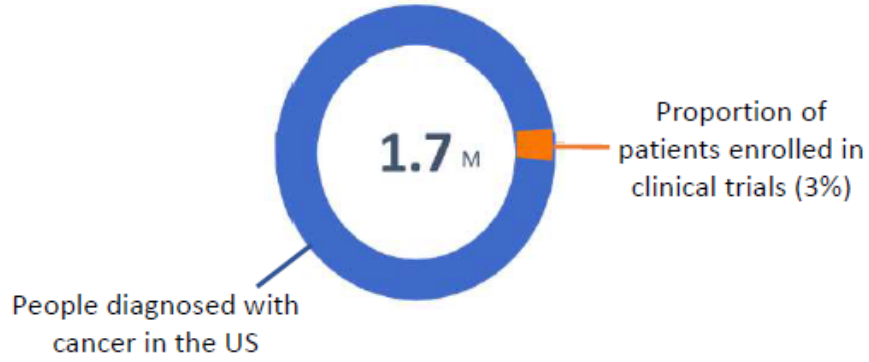
CLINICAL TRIALS AND THE REAL WORLD

Most clinical trials are strictly controlled with specific inclusion and exclusion criteria and investigate carefully defined patient populations.

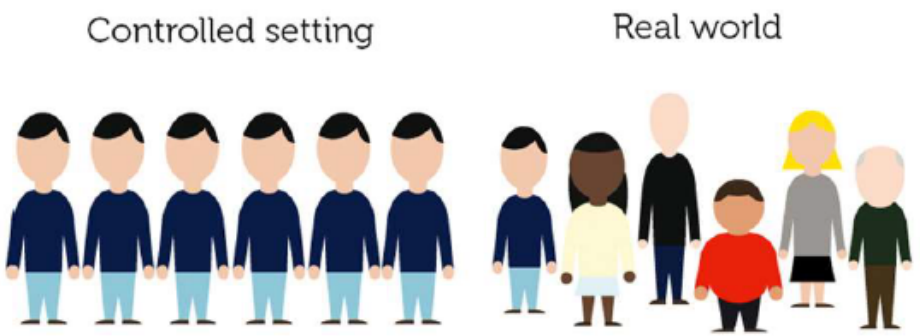
In clinical trials, inclusion and exclusion criteria for the targeted patient population may influence outcomes.[5] They excel at answering scientific questions about how well a treatment works and its potential side effects in controlled conditions. Participants are often randomly assigned to different groups in the study, which is the most effective means of minimizing biases that could lead to false conclusions.[5]

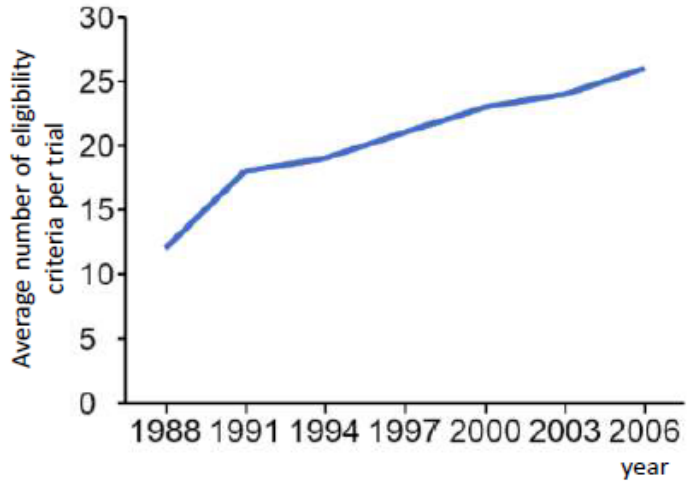
The patients who receive treatment in the real world can differ in important ways from the patients enrolled in the clinical trials for that treatment.[6] Real-world patients may be older, have more medical conditions, and have more advanced disease.[6]

From "*Randomized Controlled Trials*" to "*Real World Data*"



Patients enrolled in clinical trials are highly selected. The 'real-world' patient is unlikely to be fully represented.





NIH. Cancer statistics. March 2017. Available from: https://www.cancer.gov/about-cancer/understanding/statistics;
Institute of Medicine (US) Forum on Drug Discovery, Development, and Translation. Transforming Clinical Research in the United States: Challenges and Opportunities: Workshop Summary. Washington (DC): National Academies Press (US)

Limiti degli RCTs

- Molto costosi (tempo e denaro)

- Condotti su popolazioni di soggetti selezionati, in settings protetti

Elevata validità interna ma bassa validità esterna

- **Dubbia generalizzabilità dei risultati a popolazioni con caratteristiche lievemente diverse (pazienti in condizioni cliniche più gravi o meno gravi, fasce protette)**
- **Pochi dati su interazioni con malattie e terapie concomitanti**
- **Pochi dati sulla reale compliance alle terapie**
- **Condotti in centri ultraspecializzati**

From "*Randomized Controlled Trials*" to "*Real World Data*"

Table 1 – ISPOR, ABPI, RAND Corporation, and IMI-GetReal definitions for RWD.

Term and source	Definition
RWD (ISPOR [7])	Data used for decision making that are not collected in conventional RCTs.
RWD (ABPI [8])	For the purposes of this guidance, "RWD" will refer to data obtained by any non-interventional methodology that describe what is happening in normal clinical practice.
RWD (RAND [9])	"RWD" is an umbrella term for different types of health care data that are not collected in conventional RCTs. RWD in the health care sector come from various sources and include patient data, data from clinicians, hospital data, data from payers, and social data.
RWD (IMI-GetReal [10])	An umbrella term for data regarding the effects of health interventions (e.g., benefit, risk, and resource use) that are not collected in the context of conventional RCTs. Instead, RWD are collected both prospectively and retrospectively from observations of routine clinical practice. Data collected include, but are not limited to, clinical and economic outcomes, patient-reported outcomes, and health-related quality of life. RWD can be obtained from many sources including patient registries, electronic medical records, and observational studies.

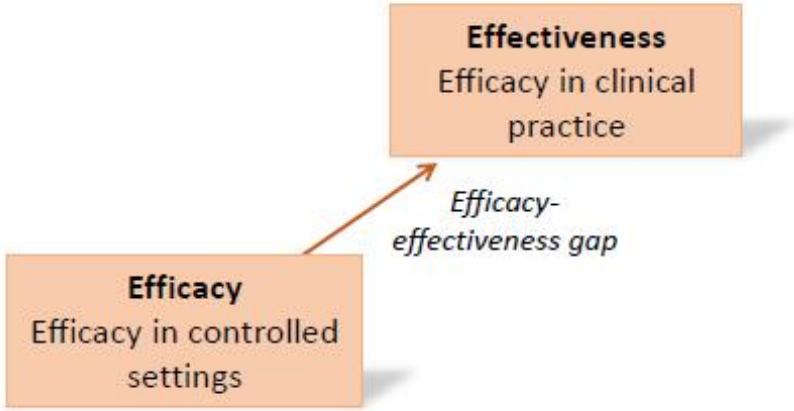
ABPI, Association of the British Pharmaceutical Industry; ISPOR, International Society for Pharmacoeconomics and Outcomes Research; RCT, randomized controlled trial; RWD, real-world data.

Makady et al. Value in health 2017

RWD definition

- Heterogeneous nature
- Multiple sources
- Uncontrolled collection, both retrospective and prospective

	RCTs	RWD
Type	Interventional/experimental	Non interventional/Observational
Outcomes	*Efficacy* and safety	*Effectiveness* and economic assessment
Population	Narrow and selected	Wide and unselected
Randomization and blinding	Yes	No
Cost	High	Low
Relevance	Internal validity	Clinical practice



From "*Real World Data*" to "*Real World Evidence*": the case of Alectinib

Alectinib received a conditional approval from FDA for the treatment of ALK+ NSCLC. US approval was based on data from two single-arm Phase II trials (NCT01801111 and NCT01871805).

National Health Technology Assessment (NHTA) authorities in EU requested additional evidence of the effectiveness of alectinib relative to the SoC (ceritinib) in order to provide a coverage decision.

To meet this requirement, Roche collaborated with Flatiron Health to conduct a retrospective analysis of EHRs of patients treated with ceritinib. The real-world external control arm was compared to the phase II single-arm and submitted to NHTA authorities, satisfying coverage requirements

Media Release

Basel, 21 February 2017

Roche receives EU approval of Alecensa (alectinib) for people with previously treated ALK-positive non-small cell lung cancer

- **Alecensa provides an efficacious, systemic treatment option for people with ALK-positive NSCLC, which is also active in patients with brain metastases, who have been previously treated with crizotinib**

Conversation with the Cancer Letter

Roche CEO O'Day: Our investment in Flatiron will accelerate their mission

Roche's purchase of Flatiron Health will accelerate the development of real-world data suitable for supporting regulatory decisions, said Daniel O'Day, CEO of Roche Pharmaceuticals.

98P
Retrospective indirect comparison of alectinib phase II data vs ceritinib real-world data in ALK+ NSCLC after progression on crizotinib FREE

J. Davies, M. Martinec, R. Martina, P. Delmar, M. Coudert, W. Bordogna, S. Golding, G. Crane

Annals of Oncology, Volume 28, Issue suppl_2, April 2017, mdx091.018, https://doi.org/10.1093/annonc/mdx091.018
Published: 03 April 2017

Contains Nonbinding Recommendations

Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices

Guidance for Industry and Food and Drug Administration Staff

Document issued on August 31, 2017.

The draft of this document was issued on July 27, 2016

For questions about this document regarding CDRH-regulated devices, contact the Office of Surveillance and Biometrics (OSB) at 301-796-5997 or CDRHClinicalEvidence@fda.hhs.gov. For questions about this document regarding CBER-regulated devices, contact the Office of Communication, Outreach, and Development (OCOD) at 1-800-835-4709 or 240-402-8010.

U.S. FOOD & DRUG ADMINISTRATION

U.S. Department of Health and Human Services
Food and Drug Administration

Center for Devices and Radiological Health

Center for Biologics Evaluation and Research

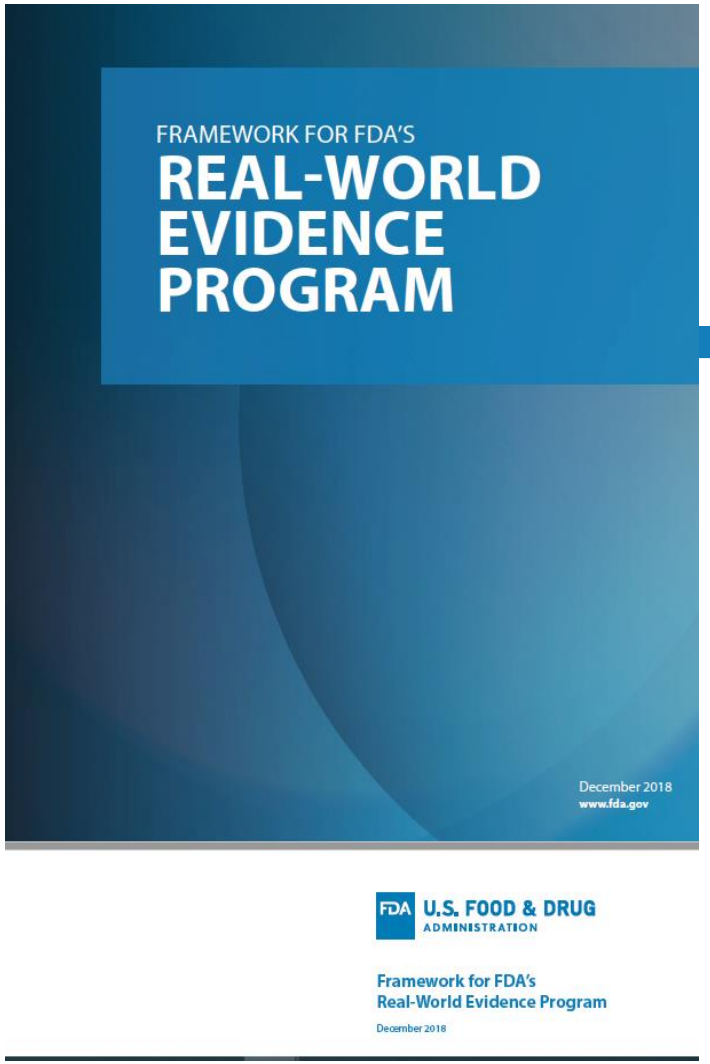


Table of Contents

Introduction ... 3
Definitions of Real-World Data and Real-World Evidence ... 4
Scope of RWE Program Under 21st Century Cures Act ... 7
Current Use of RWD for Evidence Generation ... 8
Generating Evidence Regarding Safety and Effectiveness ... 8
Supporting FDA's Regulatory Decisions of Effectiveness ... 9
Trial Designs Using RWD to Generate Evidence ... 11
Framework for Evaluating RWD/RWE for Use in Regulatory Decisions ... 13
Using Trials or Studies with RWD/RWE for Effectiveness Decisions ... 13
Assessing Fitness of RWD for Use in Regulatory Decisions ... 14
Potential for Study Designs Using RWD to Support Effectiveness ... 19
Regulatory Considerations for Study Designs Using RWD ... 22
Data Standards – Appropriate Data Standards for Integration and Submission to FDA ... 24
Stakeholder Engagement ... 26
Conclusion ... 27
Glossary ... 28
References ... 30
Appendix ... 34

SERVIZIO SANITARIO REGIONALE EMILIA-ROMAGNA
Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori

ISTITUTO SCIENTIFICO ROMAGNOLO PER LO STUDIO E LA CURA DEI TUMORI

DIVERSE REAL-WORLD DATA

Real-world data can come from many sources. It includes prospective observational studies designed to collect data on real-world patients. It can also retrospectively draw on existing patient registries, insurance databases, and electronic medical records.[4] There is a lot of room for innovation in collecting real-world data. As with RCTs, real-world data must be collected with careful consideration for ethics and patient privacy.

COMPLEMENTING, NOT REPLACING CLINICAL TRIALS

Used together, both real-world studies and RCTs contribute to the understanding of a treatment or disease. They are complementary, rather than substitutes for each other, because they provide data from different settings.

Definitions

- **Real-World Data (RWD)** are data relating to patient health status and/or the delivery of health care routinely collected from a varie

RWD include data derived from electronic health records (EHRs), claims and billing data, data from product and disease registries, patient-generated data including in home-use settings, and data gathered from other sources that can inform on health status, such as mobile devices.

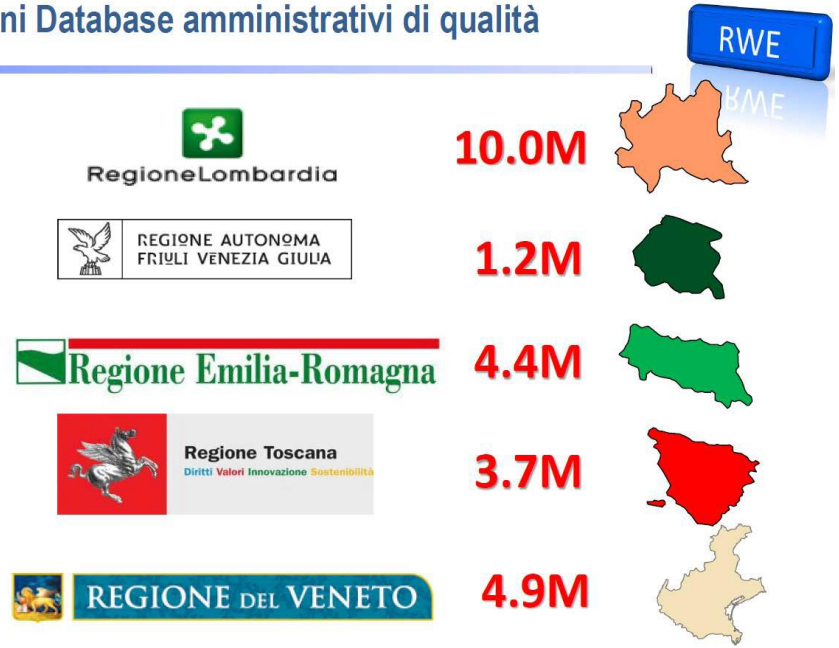
- **Real-World Evidence (RWE)** is the clinical evidence regarding the usage a
derived fro

RWE can be generated by different study designs or analyses, including but not limited to, randomized trials, such as large simple trials, pragmatic trials, and observational studies (prospective and/or retrospective).

RWE – BIG DATA

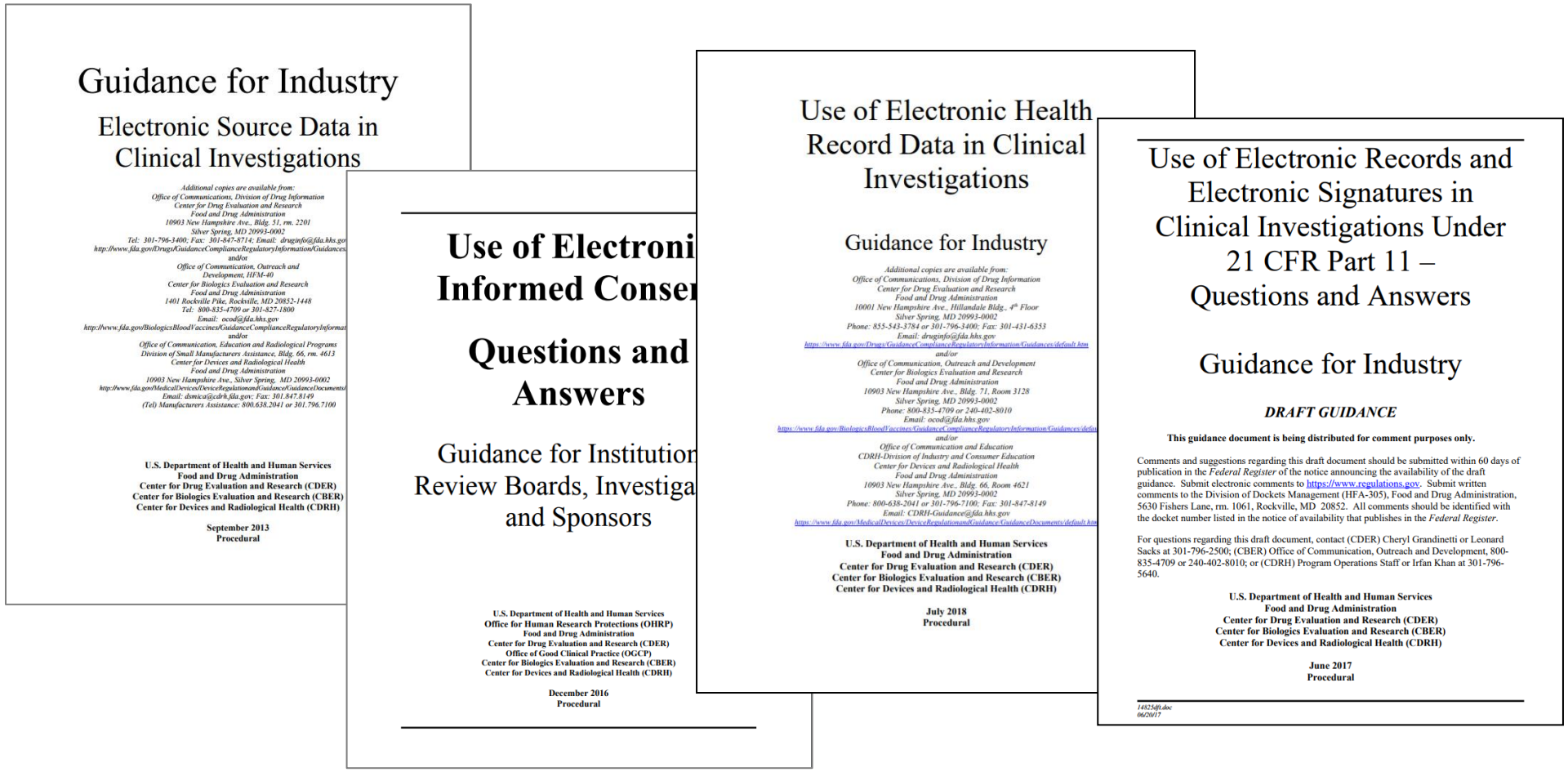
Il percorso di cura e la registrazione di informazioni cliniche

Alcuni Database amministrativi di qualità



Immagini, ngs……

Foundation for use of Electronic Data



Guidance for Industry

Electronic Source Data in Clinical Investigations

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)

September 2013
Procedural

Use of Electroni
Informed Conse

Questions and Answers

Guidance for Institution
Review Boards, Investiga
and Sponsors

U.S. Department of Health and Human Services
Office for Human Research Protections (OHRP)
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Office of Good Clinical Practice (OGCP)
Center for Biologics Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)

December 2016
Procedural

Use of Electronic Health Record Data in Clinical Investigations

Guidance for Industry

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)

July 2018
Procedural

Use of Electronic Records and Electronic Signatures in Clinical Investigations Under 21 CFR Part 11 – Questions and Answers

Guidance for Industry

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 60 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to https://www.regulations.gov. Submit written comments to the Division of Dockets Management (HFA-305), Food and Drug Administration, 5630 Fishers Lane, rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document, contact (CDER) Cheryl Grandinetti or Leonard Sacks at 301-796-2500; (CBER) Office of Communication, Outreach and Development, 800-835-4709 or 240-402-8010; or (CDRH) Program Operations Staff or Irfan Khan at 301-796-5640.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)

June 2017
Procedural

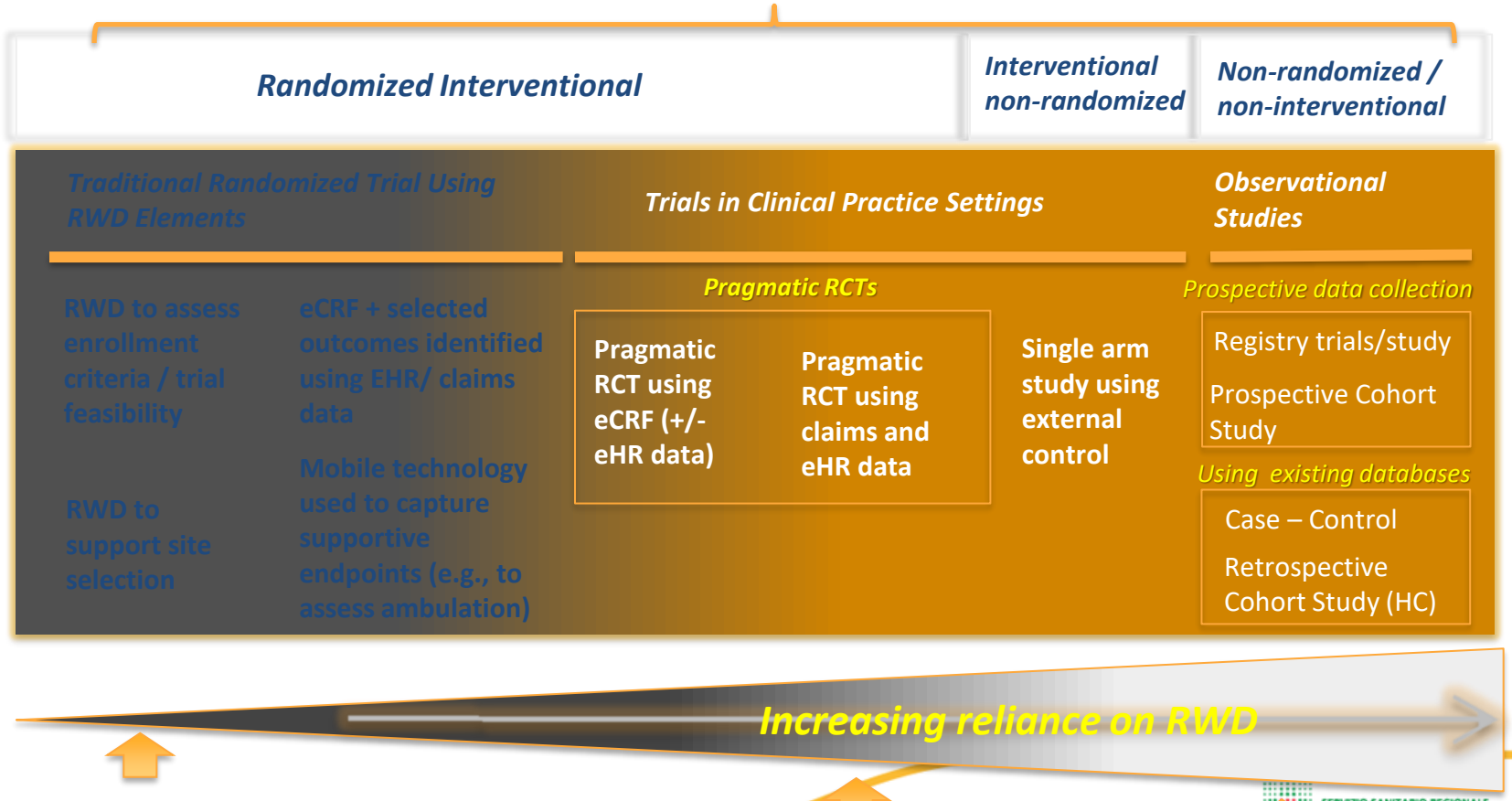
14825dft.doc
06/2017

SERVIZIO SANITARIO REGIONALE
EMILIA-ROMAGNA
Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori

ISTITUTO SCIENTIFICO ROMAGNOLO PER LO STUDIO E LA CURA DEI TUMORI

Wide Spectrum of Potential Uses of RWD / RWE in Clinical Studies

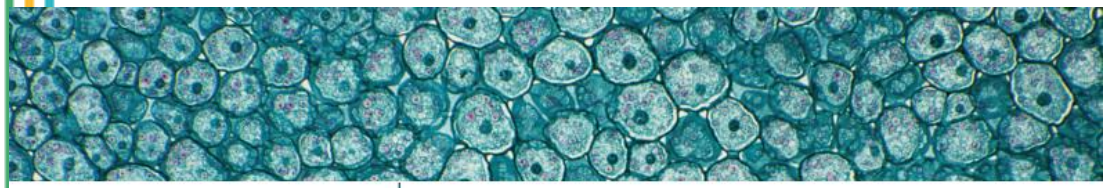
Different Challenges and Opportunities for Each Approach



Courtesy of Peter Stein, OND

DRUG	INDICATION	STATUS	DATA
Lutathera *(lutetium 177 dotate)*	GEP-NET Gastropanc. Neuroendo tumors	Approved 2017	▪ Open label clinical trial ▪ **Analysis of 360 patients in an investigator sponsored, open-label**, single-arm, single institution study of 1214 patients*
Voraxaze *(glucarpidase)*	Treatment of MTX toxicity	Approved 2012	▪ Approval based on open-label, **NIH compassionate Use Protocol**
Uridine Triacetate	Treatment of 5 FU overdose	Approved 2015	▪ Two single-arm, open label expanded access trial of **135 patients compared to case history control**
Defitelio *(defibrotide sodium)*	Severe hepatic veno-occlusive disorder	Approved 2016	▪ Two prospective clinical trials enrolling 179 patients and **an expanded access study with 351 patients**
Blincynto *(Blinatumomab)*	Treatment of Acute Lymphoblastic Leukemia	Approved 2014	▪ Single arm trial ▪ Reference group weighted analysis of patient level **data on chart review of 694 patients at EU and US study sites***
Carbaglu carglumic acid	Treatment of NAGS deficiency	Approved 2010	▪ Retrospective, non-random, un-blinded **case series of 23 patients** compared to **historical control group**
Myozyme (alglucosidase alfa)	Treatment of Pompe disease	Approved 2004	▪ Open-label, non-randomized study of 18 patients compared to **historical control group of 62 untreated patients**
Refludan	Anti-coagulation in heparin-induced thrombocytopenia	Approved 1998	▪ Two non-randomized, open-label multicenter trials using **historical control comparator group from HIT Registry**

Bold = RWE NOT EXHAUSTIVE *https://www.nature.com/bcj/journal/v6/n9/full/bcj201684a.html



grace PRINCIPLES

A Validated Checklist

for Evaluating the Quality of Observational Cohort Studies for Decision-Making Support

GRACE: Good ReseArch for Comparative Effectiveness

The GRACE Checklist is designed for the assessment of observational studies of comparative effectiveness in terms of their quality and usefulness for decision-making. The checklist was developed from a review of the literature with guidance from recognized experts in this field. The content includes questions about data and methods. Validation activities have documented the usefulness of all 11 questions in this checklist. Approaches to scoring are under consideration.

The GRACE Initiative has been spearheaded by Quintiles Real-World & Late Phase Research in collaboration with the National Pharmaceutical Council. GRACE contributors represent perspectives from academic, government, and private sectors in the US, Europe, Asia and Africa. More information, including a list of contributors and collaborators can be found at www.graceprinciples.org. More information is available in the American Journal of Managed Care 2010; 16(6): 467-471 (Dreyer NA, Schneeweiss S, McNeil B et al.) and the Journal of Managed Care & Specialty Pharmacy 2014; 20(3):301-8.

To join the GRACE Initiative or for more information, please contact us at coordinator@graceprinciples.org. Feedback welcomed.

Nancy A. Dreyer
Leader, GRACE Initiative

ALWAYS CONSIDER: *Is this adequate for the study purpose?*

Data

D1 Were treatment and/or important details of treatment exposure adequately recorded for the study purpose in the data source(s)? *Note: not all details of treatment are required for all research questions.*

- ☐ Yes—reasonably necessary information to determine treatment or intervention was adequately recorded for study purposes (e.g., for drugs, sufficient detail on dose, days supplied, route or other data important. For vaccines, consider the importance of batch, dose, route and site of administration, etc. For devices, consider type of device, placement, surgical procedure used, serial number, etc.).
- ☐ No—data source clearly deficient *or* not enough information in article

Comments:

D2 Were the primary outcomes adequately recorded for the study purpose (e.g., available in sufficient detail through data source(s))?

- ☐ Yes—information to ascertain outcomes were adequately recorded in the data source (e.g., if clinical outcomes were ascertained using ICD-9 diagnosis code(s) in an administrative database, the level of sensitivity and specificity captured by the code(s) was sufficient for assessing the outcome of interest.)
- ☐ No—data source clearly deficient (e.g., the code(s) captured a range of conditions that was too broad or narrow, and supplementary information such as that from medical charts was not available), *or* not enough information in article

Comments:

D3 Was the primary *clinical* outcome(s) measured objectively rather than subject to clinical judgment (e.g., opinion about whether the patient's condition has improved)?

- ☐ Yes—clinical outcomes were measured objectively (e.g., hospitalization, mortality)
- ☐ N/A—primary outcome not clinical (e.g., PROs)
- ☐ No—e.g., clinical opinion about whether patient's condition improved, *or* not enough information in article

Comments:

SERVIZIO SANITARIO REGIONALE EMILIA-ROMAGNA
Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori

ISTITUTO SCIENTIFICO ROMAGNOLO PER LO STUDIO E LA CURA DEI TUMORI

ALWAYS CONSIDER: *Is this adequate for the study purpose?*

D4 Were primary outcomes validated, adjudicated, or otherwise known to be valid in a similar population?

- ☐ **Yes**—outcomes were validated, adjudicated, or based on medical chart abstractions with clear definitions, e.g. a validated instrument was used to assess patient-reported outcomes (e.g., SF-12 Health Survey); a clinical diagnosis via ICD-9 code was used, with formal medical record adjudication by committee to confirm diagnosis or other procedures to achieve reasonable sensitivity and specificity; billing data were used to assess health resource utilization, etc.
- ☐ **No, *or* not enough information in article**

Comments:

D5 Was the primary outcome(s) measured or identified in an equivalent manner between the treatment/intervention group and the comparison group(s)?

- ☐ **Yes**
- ☐ **No, *or* not enough information in article**

Comments:

D6 Were important covariates that may be known confounders or effect modifiers available and recorded? *Important covariates depend on the treatment and/or outcome of interest, (e.g., body mass index should be available and recorded for studies of diabetes; race should be available and recorded for studies of hypertension and glaucoma).*

- ☐ **Yes**—most if not all important known confounders and effect modifiers available and recorded, e.g., measures of medication dose and duration.
- ☐ **No**—at least one important known confounder or effect modifier not available and recorded (as noted by authors or as determined by user's clinical knowledge), ***or* not enough information in article**

Comments:

Methods

M1 Was the study (or analysis) population restricted to new initiators of treatment or those starting a new course of treatment? *Efforts to include only new initiators may include restricting the cohort to those who had a washout period (specified period of medication nonuse) prior to the beginning of study follow-up.*

- ☐ **Yes**—only new initiators of the treatment of interest were included in the cohort, or for surgical procedures and devices, only patients who never had the treatment before the start of study follow-up were included.
- ☐ **No, *or* not enough information in article**

Comments:

M2 If one or more comparison groups were used, were they concurrent comparators? If not, did the authors justify the use of historical comparisons group(s)?

- ☐ **Yes**—data were collected during the same time period as the treatment group ("concurrent") or historical comparators were used with reasonable justification, e.g., when it is impossible for researchers to identify current users of older treatments or when a concurrent comparison group is not valid—(i.e., uptake of new product is so rapid that concurrent comparators differ greatly on factors related to the outcome)
- ☐ **No**—historical comparators used without being scientifically justifiable, ***or* not enough information in article**

Comments:

M3 Were important covariates, confounding and effect modifying variables taken into account in the design and/or analysis? *Appropriate methods to take these variables into account may include: restriction, stratification, interaction terms, multivariate analysis, propensity score matching, instrumental variables or other approaches.*

- ☐ **Yes**—most if not all important covariates that would be likely to change the effect estimate substantially were accounted for, e.g., measures of medication dose and duration.
- ☐ **No**—some important covariates were available for analysis but not analyzed appropriately, ***or* at least one important covariate was not measured, *or* not enough information in article**

Comments:

M4 Is the classification of exposed and unexposed person-time free of "immortal time bias"? *Immortal time in epidemiology refers to a period of cohort follow-up time during which death (or an outcome that determines end of follow-up) cannot occur.*

- ☐ **Yes**
- ☐ **No, *or* not enough information in article**

Comments:

M5 Were any meaningful analyses conducted to test key assumptions on which primary results are based? *E.g., were some analyses reported to evaluate the potential for a biased assessment of exposure or outcome, such as analyses where the impact of varying exposure and/or outcome definitions was tested to examine the impact on results.*

- ☐ **Yes**—and primary results ***did not*** substantially change
- ☐ **Yes**—and primary results changed substantially
- ☐ **None reported, *or* not enough information in article**

Comments:

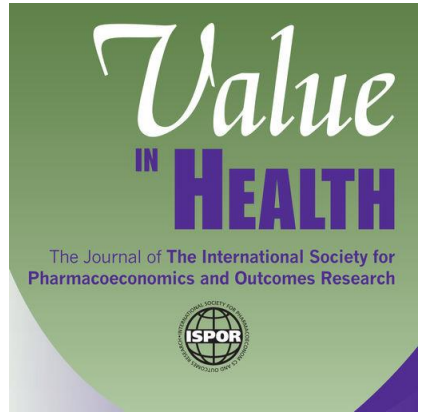
grace | Good ReseArch for Comparative Effectiveness

Observational Designs for Efficacy?

Why Did the Randomized Clinical Trial Become the Primary Focus of My Career?

*David L. Sackett, OC, MD, FRSC, FRCP (Canada, England, and Scotland)**

Department of Clinical Epidemiology & Biostatistics, McMaster University, Hamilton, ON, Canada



- Clinicians might preferentially give new treatments to patient with better prognosis
- Compliant patients might have better prognosis, regardless of their treatment
- Patients who liked their Rx might report better outcomes unrelated to the true efficacy of their treatments
- Clinicians who like their RX might report spuriously better outcomes among patients who receive them

Efforts to Enhance Transparency

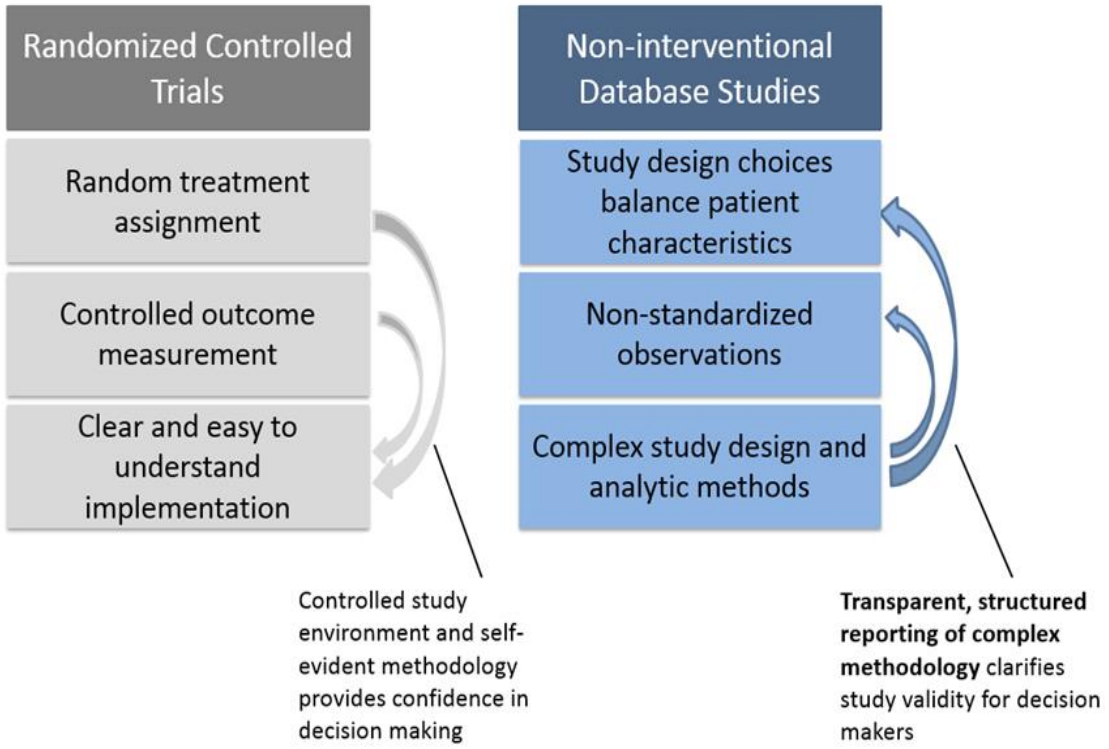


Figure courtesy of S. Schneeweiss

Good Practices for Real-World Data Studies of Treatment and/or Comparative Effectiveness: Recommendations from the Joint ISPOR-ISPE Special Task Force on Real-World Evidence in Health Care Decision Making

Marc L. Berger[1,], Harold Sox[2], Richard J. Willke[3], Diana L. Brixner[4], Hans-Georg Eichler[5], Wim Goettsch[6], David Madigan[7], Amr Makady[6], Sebastian Schneeweiss[8], Rosanna Tarricone[9], Shirley V. Wang[8], John Watkins[10], C. Daniel Mullins[11]*

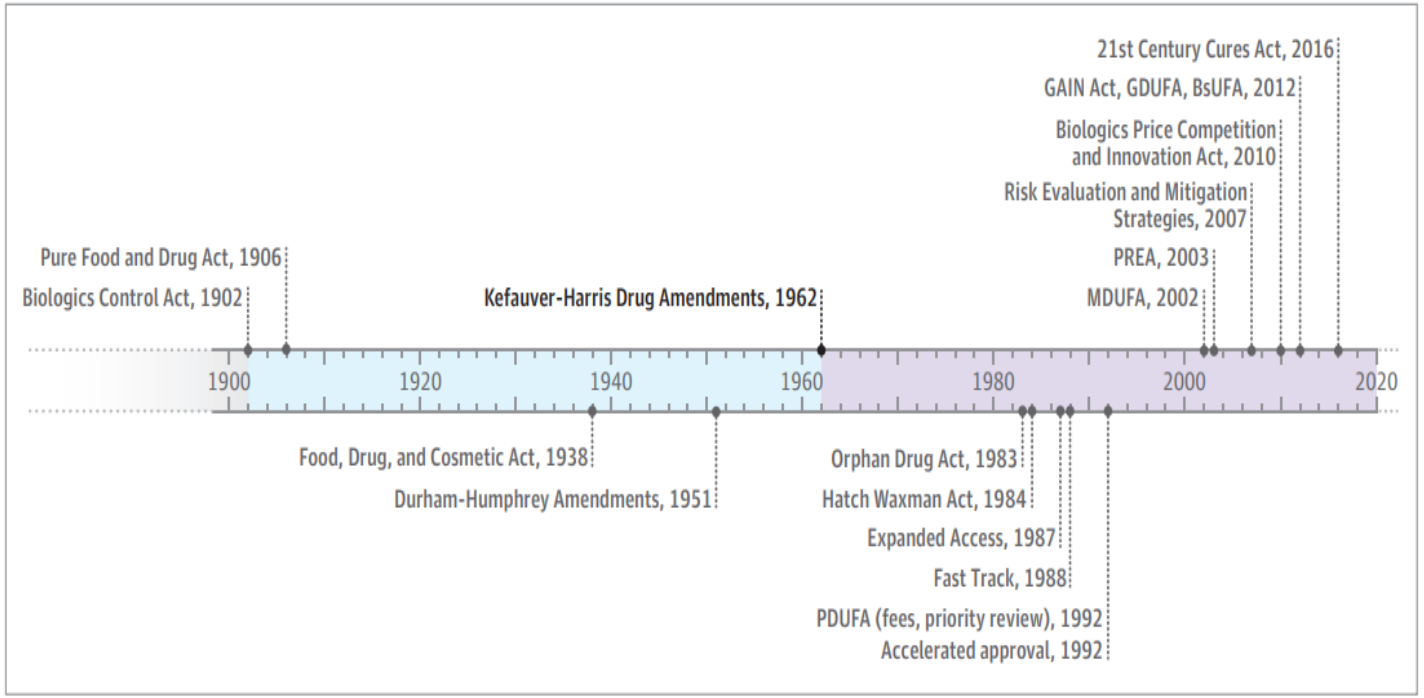
1. *A priori*, determine and declare that a study is a Hypothesis Evaluation Treatment Effectiveness (HETE) study or an Exploratory study based on conditions outlined below
2. Post a HETE study protocol and analysis plan on a public study registration site prior to conducting the study analysis.
3. Publish HETE study results with attestation to conformance and/or deviation from the study protocol and original analysis plan. Possible publication sites include a medical journal, or a publicly available web-site.
4. Enable opportunities to replicate HETE studies (i.e., for other researchers to be able to reproduce the same findings using the same data set and analytic approach). The ISPE companion paper lists information that should be reported in order to make the operational and design decisions behind a RWD study transparent enough for other researchers to reproduce the conduct of the study.
5. Perform HETE studies on a different data source and population than the one used to generate the hypotheses to be tested unless it is not feasible (e.g., another data set is not available)
6. Authors of the original study should work to publicly address methodological criticisms of their study once it is published.
7. Include key stakeholders (patients, caregivers, clinicians, clinical administrators, HTA/payers, regulators, manufacturers) in designing, conducting, and disseminating HETE studies.

JAMA | Special Communication

FDA Approval and Regulation of Pharmaceuticals, 1983-2018

Jonathan J. Darrow, SJD, JD, MBA; Jerry Avorn, MD; Aaron S. Kesselheim, MD, JD, MPH

Figure 1. Timeline of Landmark Legislation and Regulations Relating to FDA Authority to Regulate Drugs



After the 1962 Kefauver-Harris Drug Amendments (left), a series of legislative enactments and regulatory programs (right) progressively increased the flexibility of evidence requirements and imposed expanding user fees to fund the drug approval process. BsUFA indicates Biosimilar User Fee Act; FDA, US Food and Drug Administration; GAIN, Generating Antibiotic Incentives Now; GDUFA, Generic Drug User Fee Act; MDUFA, Medical Device User Fee Act; PDUFA, Prescription Drug User Fee Act; PREA, Pediatric Research Equity Act. Sources: Hein Online (statutes); Federal Register (FDA regulations).

Box 2. Special Approval Programs

Orphan Drug Act (1983). US legislation creating incentives for the development of rare disease treatments, defined in 1984 as diseases or conditions affecting fewer than 200 000 people in the United States.

Fast-Track (1987). A program intended to expedite the development, evaluation, and marketing of new therapies for serious and life-threatening conditions by, among other things, eliminating phase 3 trials.

Accelerated approval (1992). A program intended to expedite the development and marketing of new therapies for serious and life-threatening conditions by allowing the use of surrogate measures only reasonably likely to predict clinical benefit as end points for the pivotal clinical trials forming the basis for drug approval.

Priority review (1992). Under the Prescription Drug User Fee Act, the FDA committed to first-cycle review deadlines for new drug applications of 6 months for priority applications and 12 months for standard applications (shortened to 10 months by 2002).

Breakthrough Therapy (2012). Experimental therapies designated in this program are eligible for greater FDA attention and expedited response timelines during the clinical development process.

Abbreviation: FDA, US Food and Drug Administration.

	89-98	08-18
	18%	41%
	11%	34%
	9%	13%
	8 anni	4.8 anni

SERVIZIO SANITARIO REGIONALE EMILIA-ROMAGNA
Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori

ISTITUTO SCIENTIFICO ROMAGNOLO PER LO STUDIO E LA CURA DEI TUMORI

A retrospective review of FDA-approved drug indications in oncology from 2006 to 2018.

- **55 indications for 59 cancer drugs**
- **32 (38%) regular approval, 53 (62%) accelerated approval**
- **29 (55%) accelerated approvals converted to regular approval.**
- **6 (21%) approvals showed overall survival benefit,**
- **16 (55%) progression-free survival benefit,**
- **7 (24%) continued to use RR but gained regular approval.**

The median sample size 117 (IQR, 76-182; range, 18-1052 participants).
Confirmatory trials

JAMA Intern Med. 2019 May 28. doi: 10.1001/jamainternmed.2019.0583. [Epub ahead of print]

An Overview of Cancer Drugs Approved by the US Food and Drug Administration Based on the Surrogate End Point of Response Rate.

Chen EY[1], Raghunathan V[1], Prasad V[1,2].

Post-marketing drug surveillance by EMA: the HMA-EMA Joint Big Data Taskforce

HMA
Heads of Medicines Agencies

13 February 2019
EMA/105321/2019

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

14 National Competent Authorities (NCAs)
EMA representatives

HMA-EMA Joint Big Data Taskforce

Objective:
USE OF BIG DATA FOR
MEDICINE REGULATION

- map relevant **sources** of big data and define the main **format**, in which they can be expected to exist and through a regulatory lens describe the current landscape, the future state and challenges;
- identify areas of **usability and applicability** of emerging data sources;
- perform a gap analysis to determine the current state of **expertise** across the European regulatory network, future **needs and challenges**;
- generate a list of **recommendations** and a Big Data Roadmap.

HMA-EMA Joint Big Data Taskforce – summary report 13 February 2019 EMA/105321/2019

Post-marketing drug surveillance by EMA: the HMA-EMA Joint Big Data Taskforce

Relevant Big Datasets → 7 subgroups

WS1: DATASET CHARACTERIZATION

Clinical trial and Imaging
Observational data
Spontaneous adverse drug reports
Social media and m-health
Genomics
Bioanalytical 'omics
Data analytics *(ongoing)*

WS3: STATE OF THE ART, CHALLENGES AND EXPECTATIONS

WS2: USABILITY AND APPLICABILITY

WS4: LIST OF RECOMMENDATIONS AND BIG DATA ROADMAP

WS: worksheet

HMA-EMA Joint Big Data Taskforce – summary report 12 February 2019 EMA/105321/2019

Conclusions

- Vast amounts of healthcare data are continually being generated, offering huge opportunities to improve knowledge and increase effectiveness of therapies.
- The use of Big Data in healthcare needs to face to many technical obstacles, because of the lack of harmonised and structured data, interoperability standards, and semantical alignment between different datasets.
- Research methodology improvement and its standardisation is essential to differentiate co-incidence from real causality.
- Privacy and security must always be assured. Patients empowering, trust and direct involvement is essential to exploit Big Data as its best.
- Use of Big Data for pharmacovigilance and regulatory purpose offers the opportunity to increase reporting of ADRs, continuous and real time monitoring of drugs efficacy/toxicity (with a special attention to rare and long term toxicities), and finally increase knowledge about the real effectiveness of drugs.

Statistico vs data scientist? Le nuove sfide della professione

Domenica Fioredistella Iezzi

Dipartimento di Scienze storiche, filosofico-sociali, dei Beni culturali e del Territorio
Università di Roma "Tor Vergata"

La statistica tradizionale si è da sempre confrontata con la scarsità di dati, per questo motivo ha dovuto predisporre tecniche per inferire regolarità, probabilità, previsioni.

Nell'era dei Big Data, solo negli ultimi due anni è stato creato il 90% dei dati in circolazione, le informazioni sono cresciute in maniera esponenziale, quindi, dalla mancanza di dati si è passati a una sovrabbondanza di informazioni, spesso non strutturate, da codificare e ripulire, analizzare e interpretare. Questo nuovo scenario ha spinto lo statistico a confrontarsi non solo con le competenze tipiche della sua disciplina, ma ad acquisire nuove skill, soprattutto nel campo dell'informatica e delle Scienze Sociali.

La disponibilità di diverse fonti di dati, strutturati e non, richiede che lo statistico abbia delle competenze informatiche per raccogliere informazioni e organizzarle in modo tale da rendere utilizzabili per un'analisi statistica; conoscenze nel campo della Linguistica per disambiguare le informazioni non strutturate; pratica nelle Scienze Sociali per compenetrare il settore in cui si opera.

Negli Stati Uniti, l'esperto di big data è stato denominato "Data Scientist" ossia lo "scienziato del dato".

BIG DATA: Common definitions

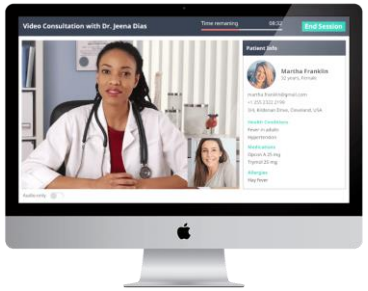
"Big data **exceeds the reach of commonly used hardware** environments **and software** tools to capture, manage, and process it with in a tolerable elapsed time for its user population." *Teradata Magazine article, 2011*

"Big data refers to data sets whose **size is beyond the ability of typical** database **software** tools to capture, store, manage and analyze. "The *McKinsey Global Institute, 2012*

"Big data is data sets that are **so voluminous and complex that traditional** data processing application **software are inadequate** to deal with them. "Wikipedia

CONTESTO: L'ERA DEI BIG DATA IN SANITA'

L'era dei **BIG DATA** apre a nuove opportunità per la medicina personalizzata, le cure preventive, la gestione delle malattie croniche e la gestione dei pazienti (non face-to-face meeting, telemonitoraggio, digital health, protesi e impianti, robotica, …).

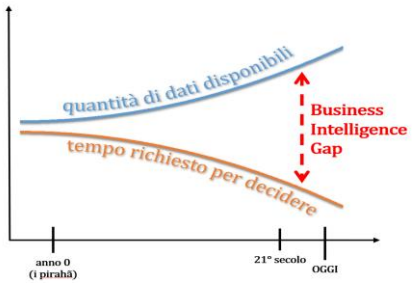


La ricchezza dei dati accumulati (dentro e fuori dalle aziende) permette di "osservare al microscopio" i comportamenti umani a larga scala e aiuta a formulare **domande completamente nuove** riguardo i legami tra i comportamenti e la gestione dei servizi sanitari.

Data Driven Healthcare: from patterns to actions

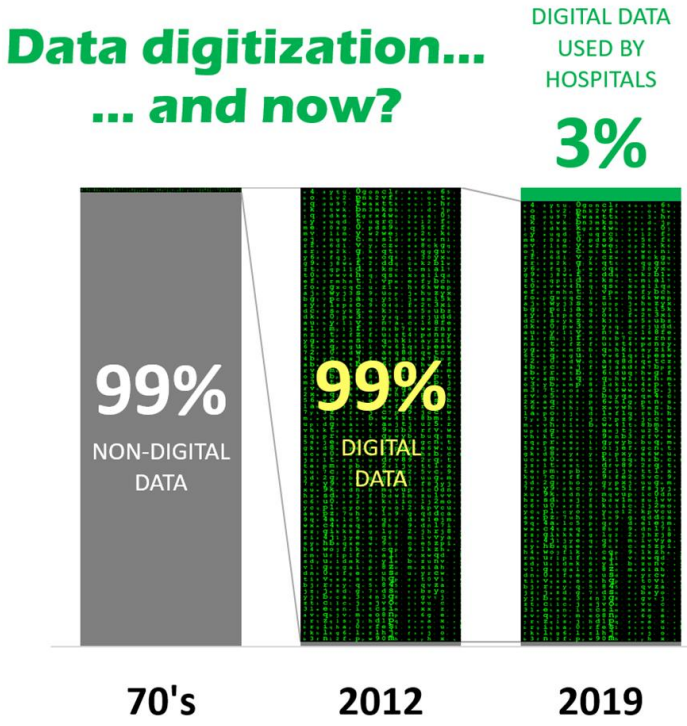
L'informatica moderna ha sviluppato, grazie alle tecnologie e a Internet, la capacità di disporre di dati in tutti gli ambiti umani permettendo di individuare nuove soluzioni a problemi esistenti e di riconoscere nuovi problemi da affrontare.

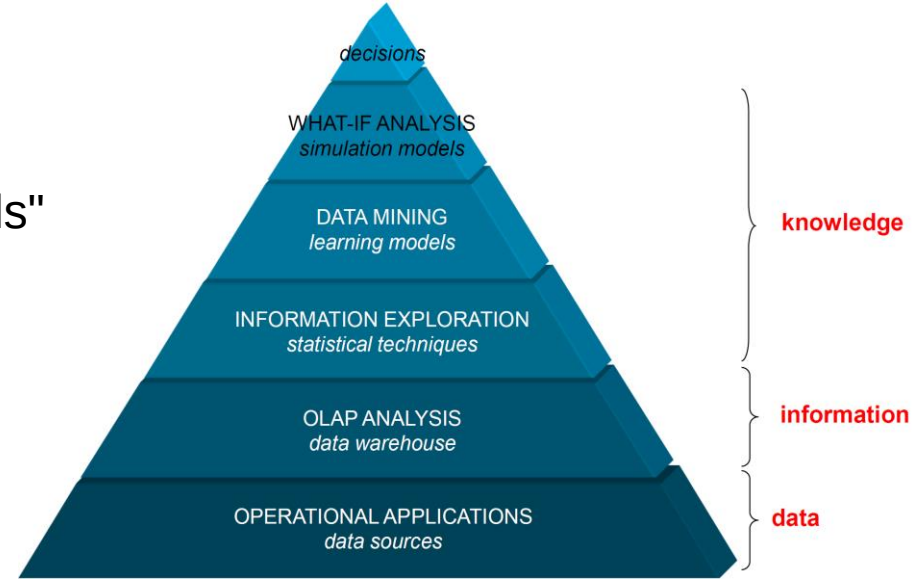
La chiave di queste opportunità risiede nei DATI, nella loro gestione e capacità di trasformarli in informazioni per il management e nell'uso di queste informazioni in maniera intelligente in modo che le decisioni aziendali dettate dall'esperienza e dalla visione siano sempre supportate da una conoscenza oggettiva.

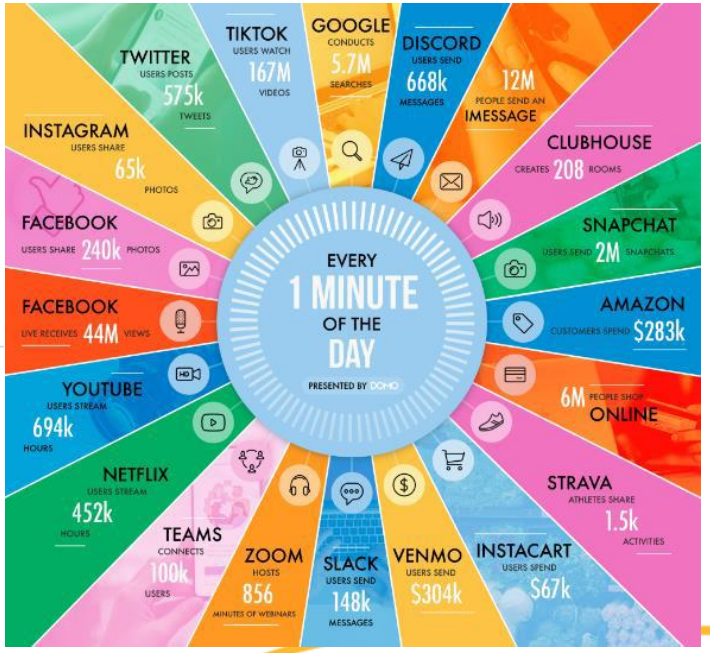


BIG DATA:

Noun: “We have *big data*"
Adjective: “We use *big data* tools"







BIG DATA – The V’s

Volume
Grandi quantità di dati

Velocità
Velocità di produzione dei dati…
e rapidità nell’utilizzo dei dati

Varietà
Dati strutturati, non strutturati, multimediali, …

Veracità
Affidabilità dei dati, gestione della non completezza, inconsistenza, ambiguità, obsolecenza, approssimazione

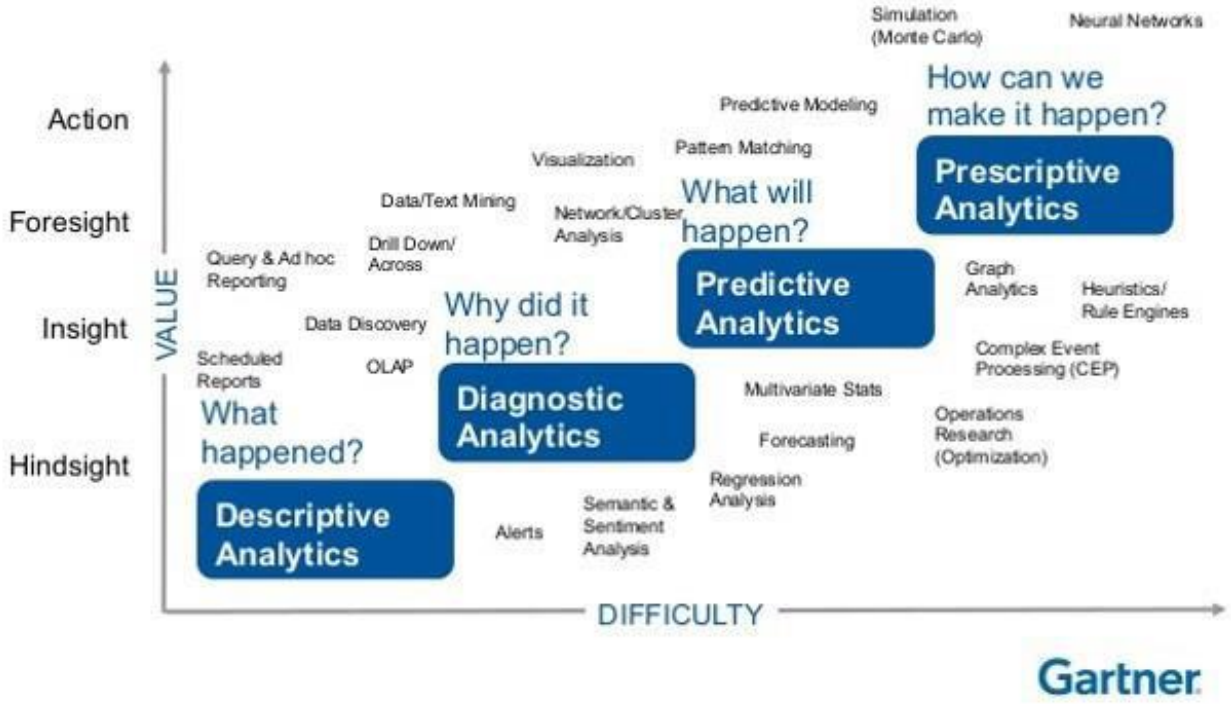
Analytics

Descriptive analytics
Cosa è accaduto

Diagnostic analytics
Perché?

Predictive analytics
Cosa è probabile che accada?

Prescriptive analytics
Cosa fare per aumentare la probabilità di far accadere qualcosa



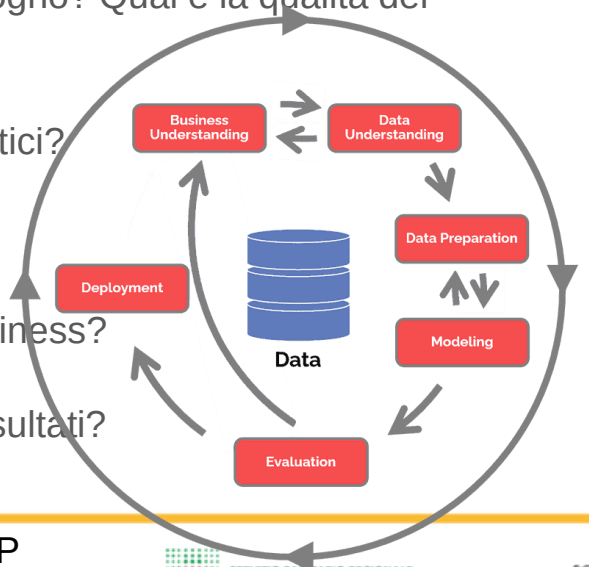
Cosa fa un data scientist

Data Scientist: The Sexiest Job of the 21st Century [Harward Business Review 2013]
Data scientist? A guide to 2015's hottest profession [Mashable 2015]
"It's official –data scientist is the best job in America" [Forbes, 2016]

Il data scientist è la figura professionale che si occupa della gestione, analisi e interpretazione di dati numerici e informazioni.

- L'analisi di questi dati serve a dare informazioni utili per prevedere un fenomeno o avere una visione chiara e reale della situazione in esame.
- **L'attività del data scientist è un percorso strutturato che comprende:**

1. Business understanding – Quali sono gli obiettivi / bisogni del business?
2. Data understanding – Quali dati abbiamo, di quali dati abbiamo bisogno? Qual è la qualità dei dati?
3. Data preparation – Come organizzare i dati per creare modelli analitici?
4. Modeling – Quali metodi e tecnologie di modellazione applicare?
5. Evaluation – Qual è il modello che soddisfa meglio i bisogni del business?
6. Deployment – In che modo gli stakeholders possono accedere ai risultati?



Metodologia CRISP
https://www.datascience-pm.com/crisp-dm-2/

Data Scientist: COMPETENZE

Laurea in Informatica, Ingegneria Informatica, Matematica e statistica, Economia.

Il **Data Scientist** possiede particolari competenze tecniche, tra le quali si possono ritrovare:

- Conoscenza database NoSQL (Not only SQL): SQL, Graph, documentali, …
- Conoscenza di linguaggi di programmazione (tipici: Python, R4)
- Conoscenza approfondita in ambito statistico;
- Conoscenza strumenti per la gestione di fogli di lavoro;
- Conoscenza dei principali metodi di gestione e analisi dei Big Data;
- Capacità avanzate nella modellazione dei dati;
- Capacità implementazione tecniche di analytics (classificazione, clustering,…);
- Conoscenza delle tecniche di Data Visualization;

Tra le principali soft skill che solitamente questa figura professionale possiede, invece, vi sono:

- Propensione al pensiero matematico e statistico;
- Predisposizione all'investigazione;
- Problem solving;
- Capacità di analisi;
- Capacità di organizzazione;
- Doti comunicative - per poter riportare in modo esaustivo e chiaro le informazioni ottenute;
- Lavoro di squadra: **la conoscenza è sempre distribuita**.

«Tipi» di Data Scientist

Data Architect

Develops data architecture to effectively capture, integrate, organize, centralize and maintain data. Core responsibilities include:

- Data Warehousing Solutions
- Extraction, Transformation and Load (ETL)
- Data Architecture Development
- Data Modeling

Data Engineer

Develop, test and maintain data architectures to keep data accessible and ready for analysis. Key tasks are:

- Extraction Transformation and Load (ETL)
- Installing Data Warehousing Solutions
- Data Modeling
- Data Architecture Construction and Development
- Database Architecture Testing

Data Analyst

Processes and interprets data to get actionable insights for a company. Responsibilities include:

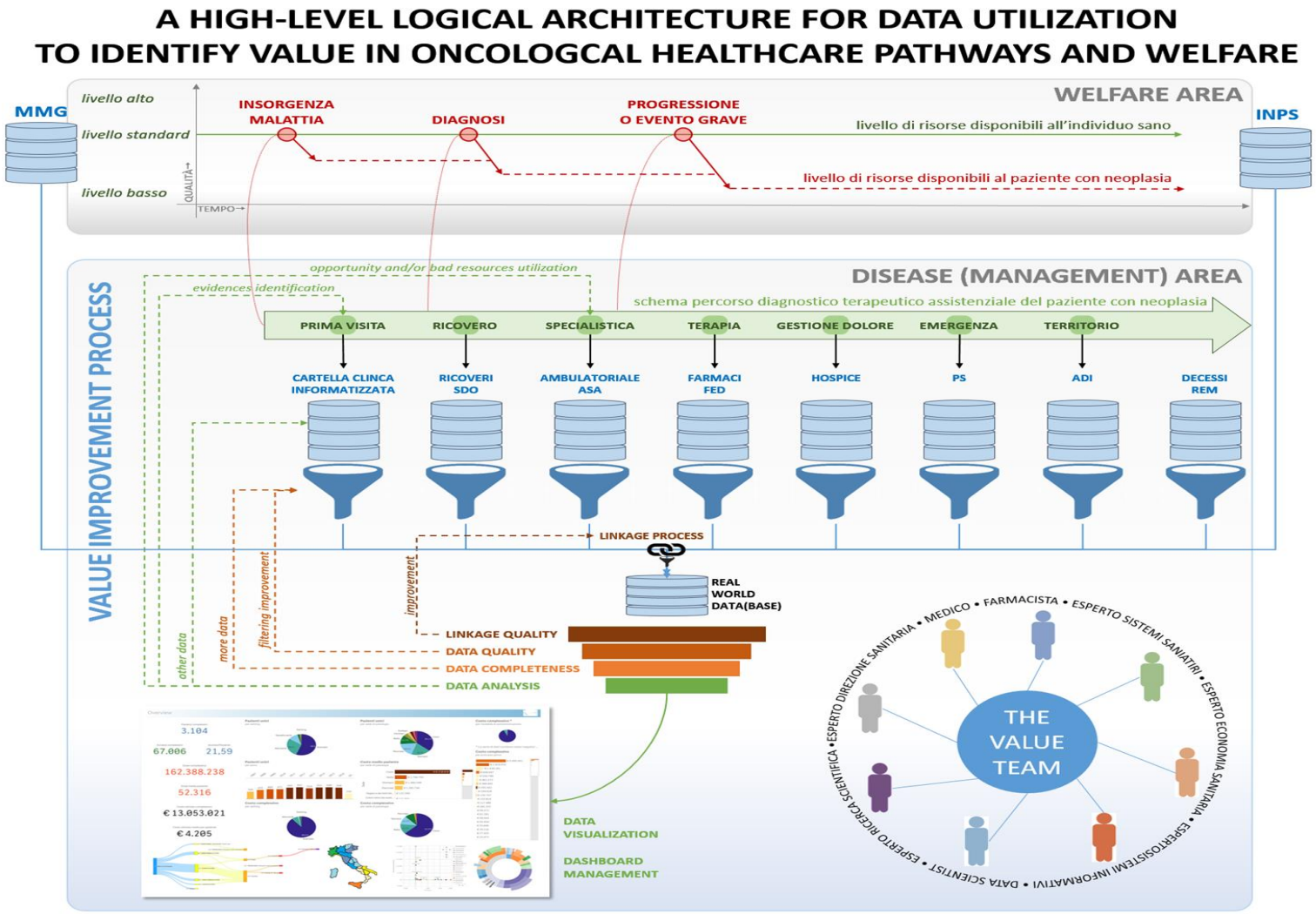
- Data Collection and Processing
- Programming
- Machine Learning
- Data Munging
- Data Visualization
- Applying Statistical Analysis

Data Scientist

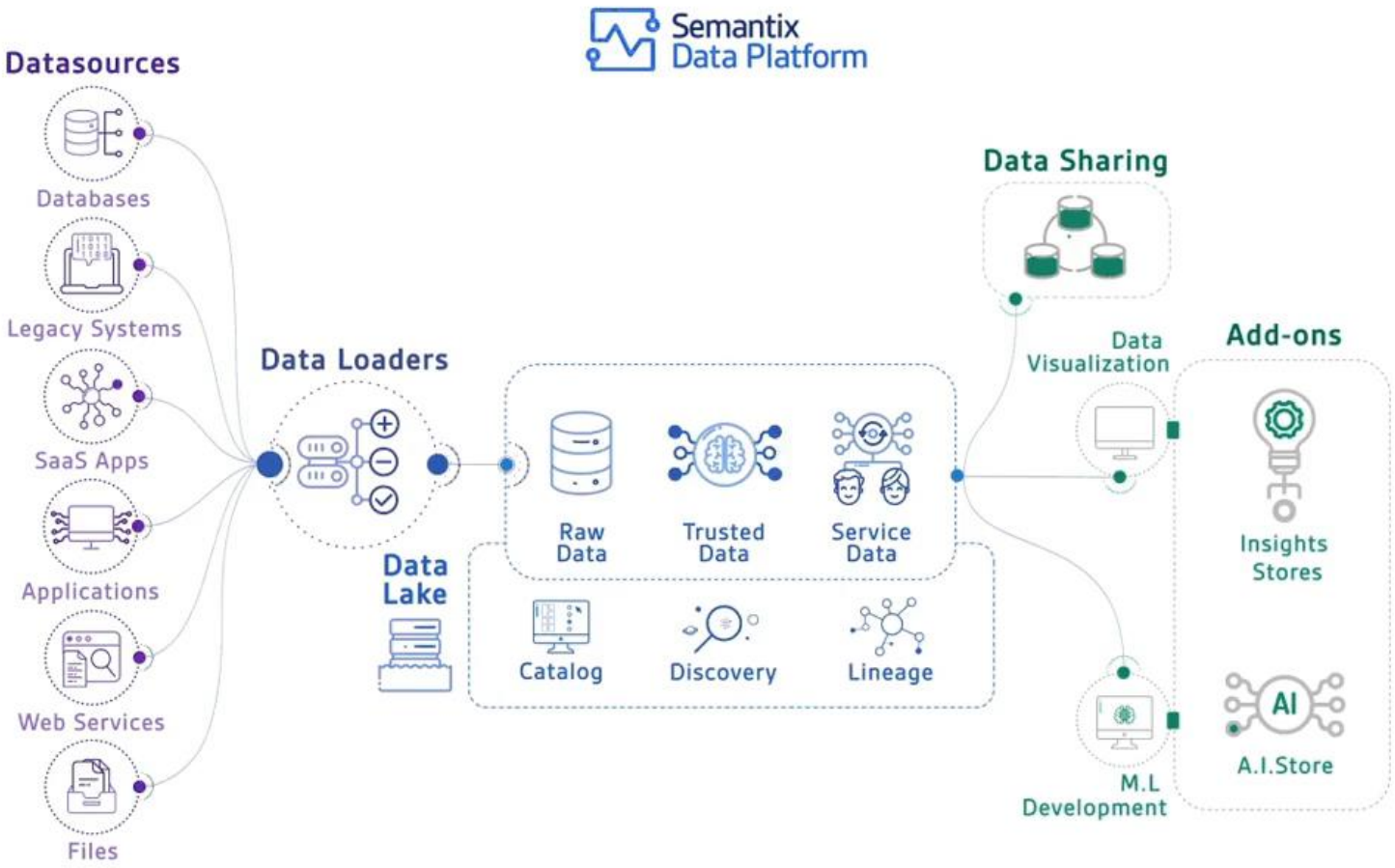
Data analysis once data volume and velocity reaches a level requiring sophisticated technical skills. Core tasks are:

- Data Cleansing and Processing
- Predictive Modeling
- Machine Learning
- Identifying Questions
- Running Queries
- Applying Statistical Analysis
- Correlating Disparate Data
- Storytelling and Visualization

VISION: ARCHITETTURA PER L'USO DEI DATI



VISION: ARCHITETTURA PER L'USO DEI DATI



CONTESTO: OSTACOLI PER LE NUOVE TECNOLOGIE

ELEMENTI CHE CONTRASTANO L'ADOZIONE DELLE NUOVE TECNOLOGIE RALLENTANDONE L'INTRODUZIONE
(Why Healthcare has been slow to adopt Technology?)

- **AVVERSIONE AL RISCHIO**
  - il danno potenziale da utilizzo di software e trattamento di dati è potenzialmente molto più alto rispetto a qualsiasi altro settore: un cattivo risultato può determinare peggioramenti degli esiti di salute
- **COMPLESSITÀ**
  - tecnologie, varietà e velocità di produzione dei dati, interoperabilità tra strumenti professionali e personali quali ad esempio strumenti di self measuring
- **COSTO**
  - sia in termini di spesa per l'introduzione di un'innovazione, sia in termini di investimento aziendale per la formazione e l'attivazione
- **SOLUZIONI GENERAL PURPOSE**
  - in sanità sono spesso necessarie soluzioni estremamente specifiche ma le proposte disponibili sono talvolta troppo generiche
- **PROTEZIONE DEI DATI**
  - GDPR, violazioni, minacce, data protection e i relativi impatti sulla visibilità e reputazione aziendale

CONTESTO: I DATI IN IRST

- IRST: Particolare attenzione per la misurazione dei risultati, è un istituto “affamato” di dati e su questi ha sempre basato le proprie strategie.
- Grande pervasività delle informazioni in ogni ambito umano; non è più sufficiente analizzare flussi informativi classici (SDO, ASA, …) sempre più necessario integrare ogni informazione aziendale, studiare le conseguenze delle azioni intraprese, individuare connessioni tra le informazioni, ma anche produrre valore (in termini di salute ed in termini economici) individuando «bolle informative» da sfruttare (analytics insights).
- Costituzione di una Data Unit che abbia come principio fondamentale la valorizzazione del patrimonio informativo.
- Messa a valore dell’integrazione dati sanitari prodotti in ospedale, a casa del paziente, da device wearable generando veri e propri Big Data o Data Lake capaci di raccogliere un’infinità di informazioni da esplorare per estrarre conoscenza in termini di informazioni o di *esperienza* per i modelli di apprendimento e di intelligenza artificiale.

Alcuni esempi di progetti in corso:

- Harmony Alliance
- InSite Platform
- Oncorelief
- …

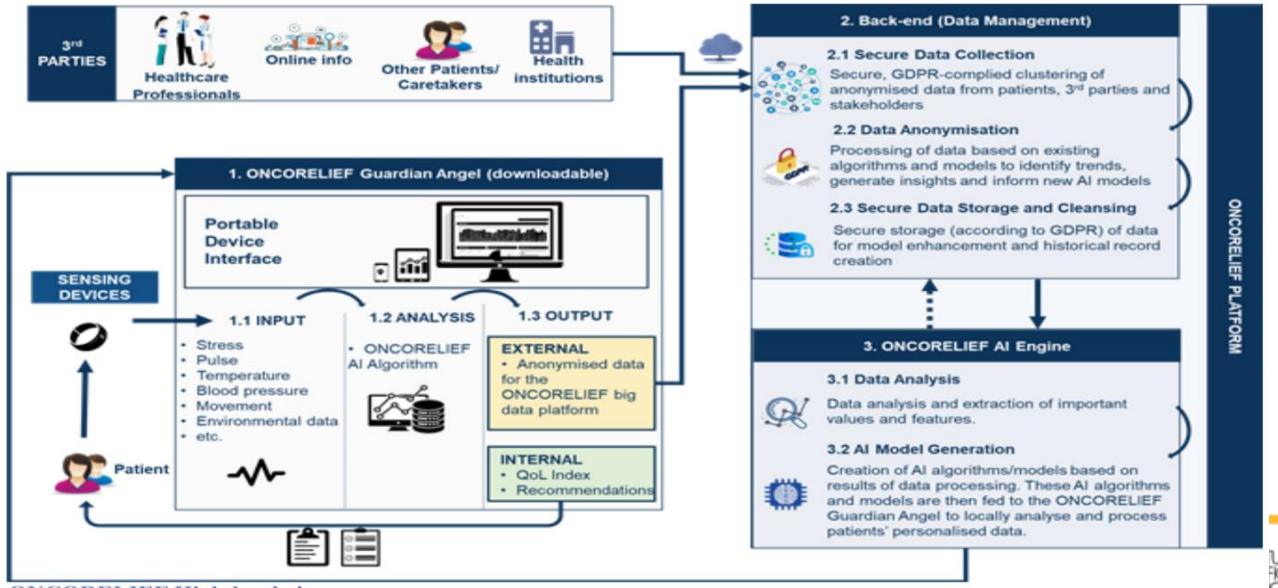


Figure 3: ONCORELIEF High-level view.

DATA UNIT: STRUTTURA

La **DATA UNIT** è strutturata in 4 ambiti complementari.

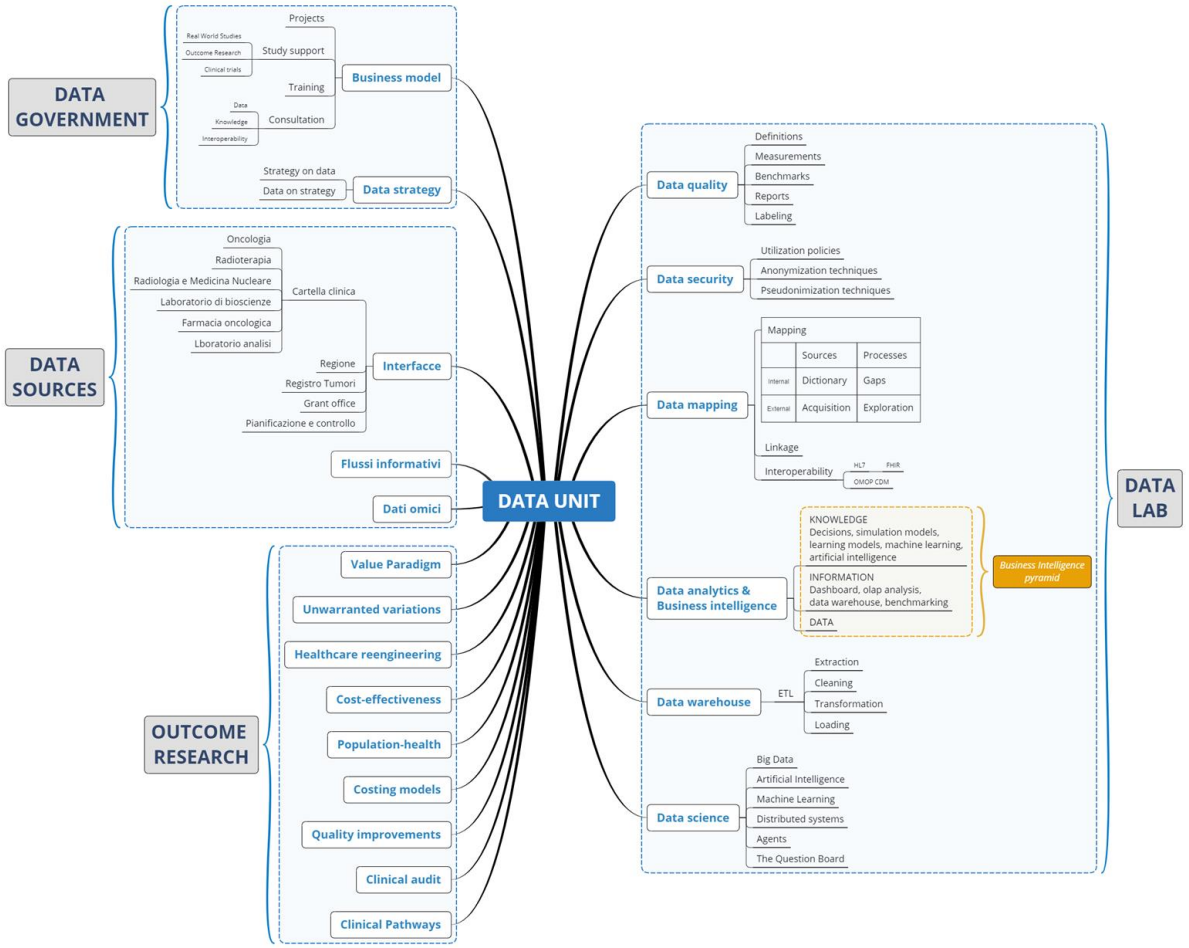
DATA GOVERNMENT

DATA SOURCES

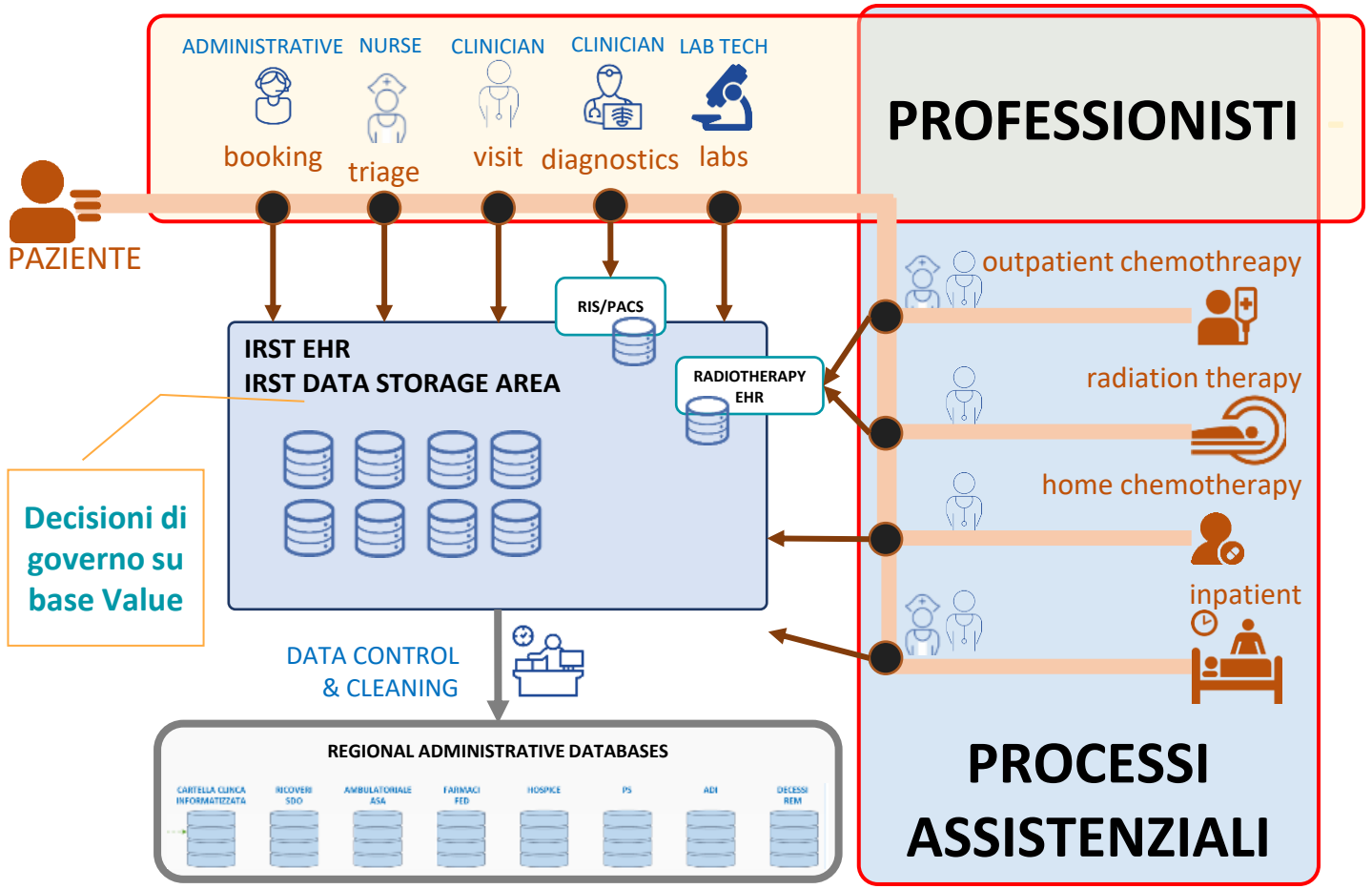
DATA LAB

OUTCOME RESEARCH

La **DATA UNIT** è nativamente interfacciata con il Servizio Informatico, il Data Protection Officer, la Biostatistica, le Direzioni.



CONTESTO: I PROCESSI PRODUTTORI DI DATI



DATA QUALITY

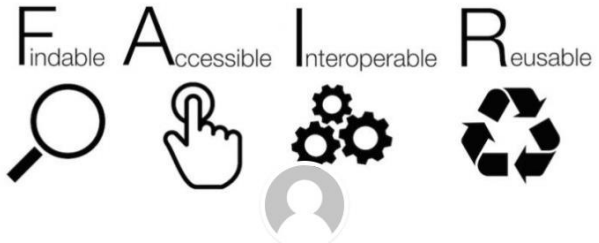
Data quality definition from ISO 9000:2015:

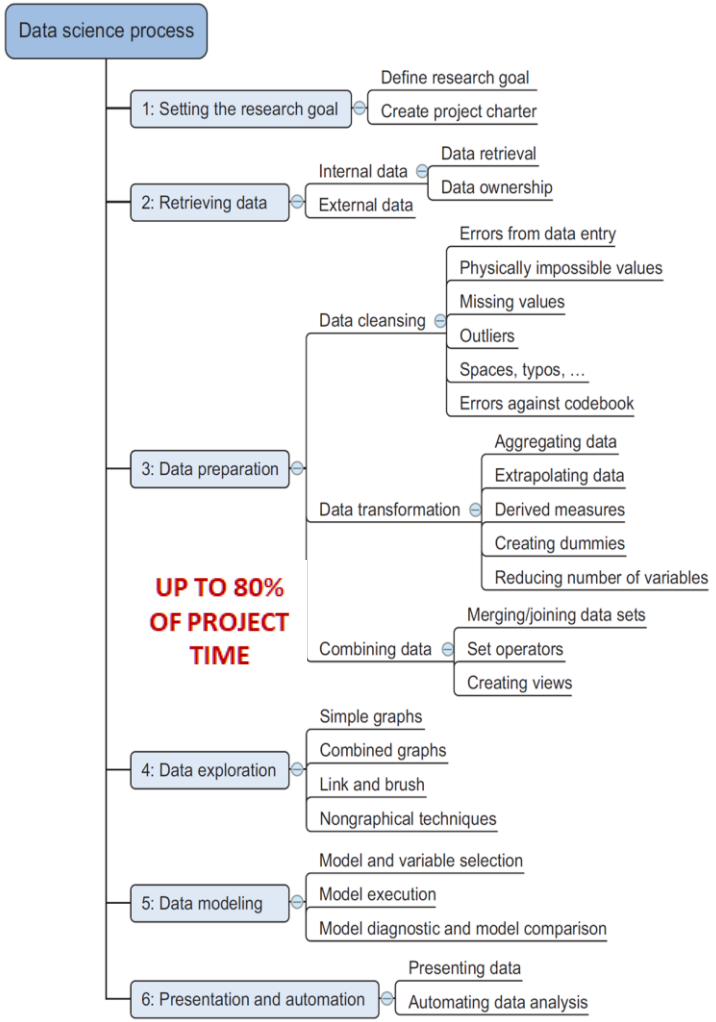
"**degree to which a set of characteristics of data fulfills requirements"**.

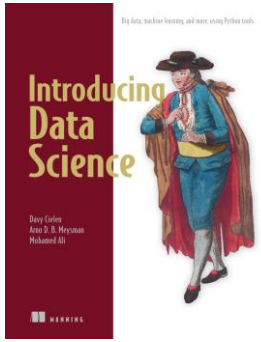
Requirements are defined as the need or expectation that is stated, generally implied or obligatory.

Some **Data quality** *suggestions*:

- Degree of excellence in the data
- Completeness, validity, consistency, timeliness, accuracy
- Ability of data to satisfy a given purpose







EHR2EDC

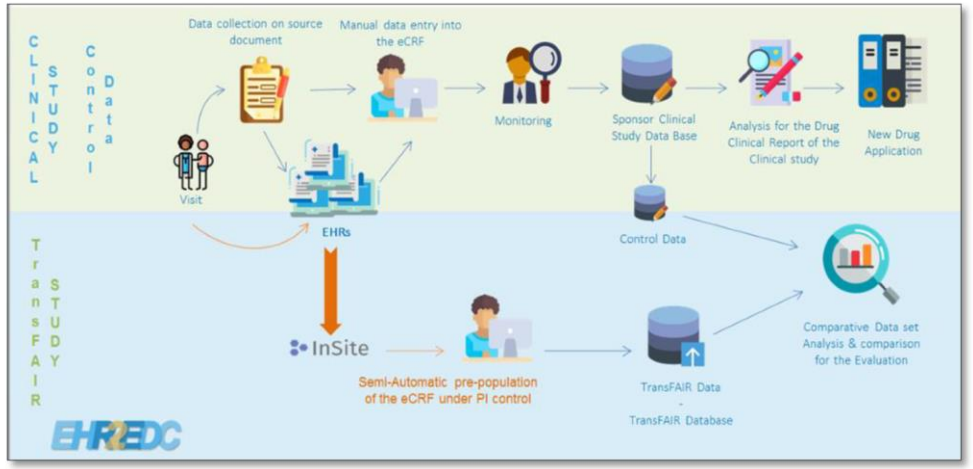
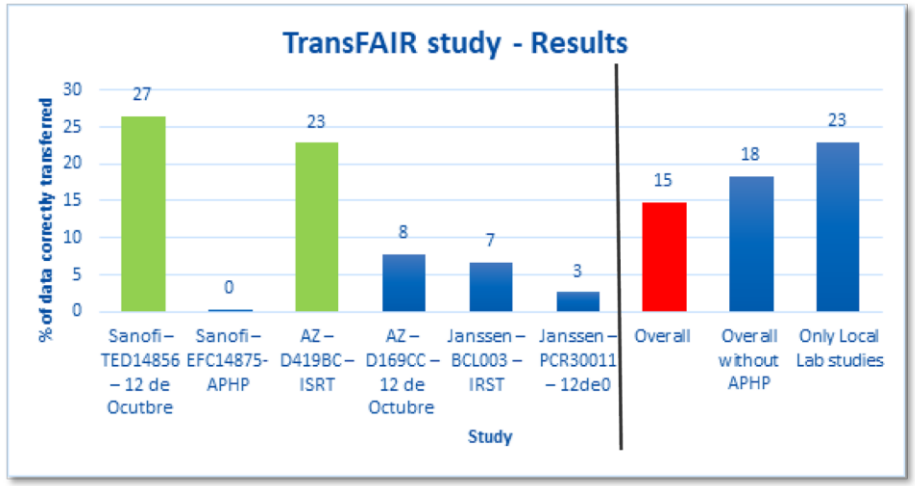
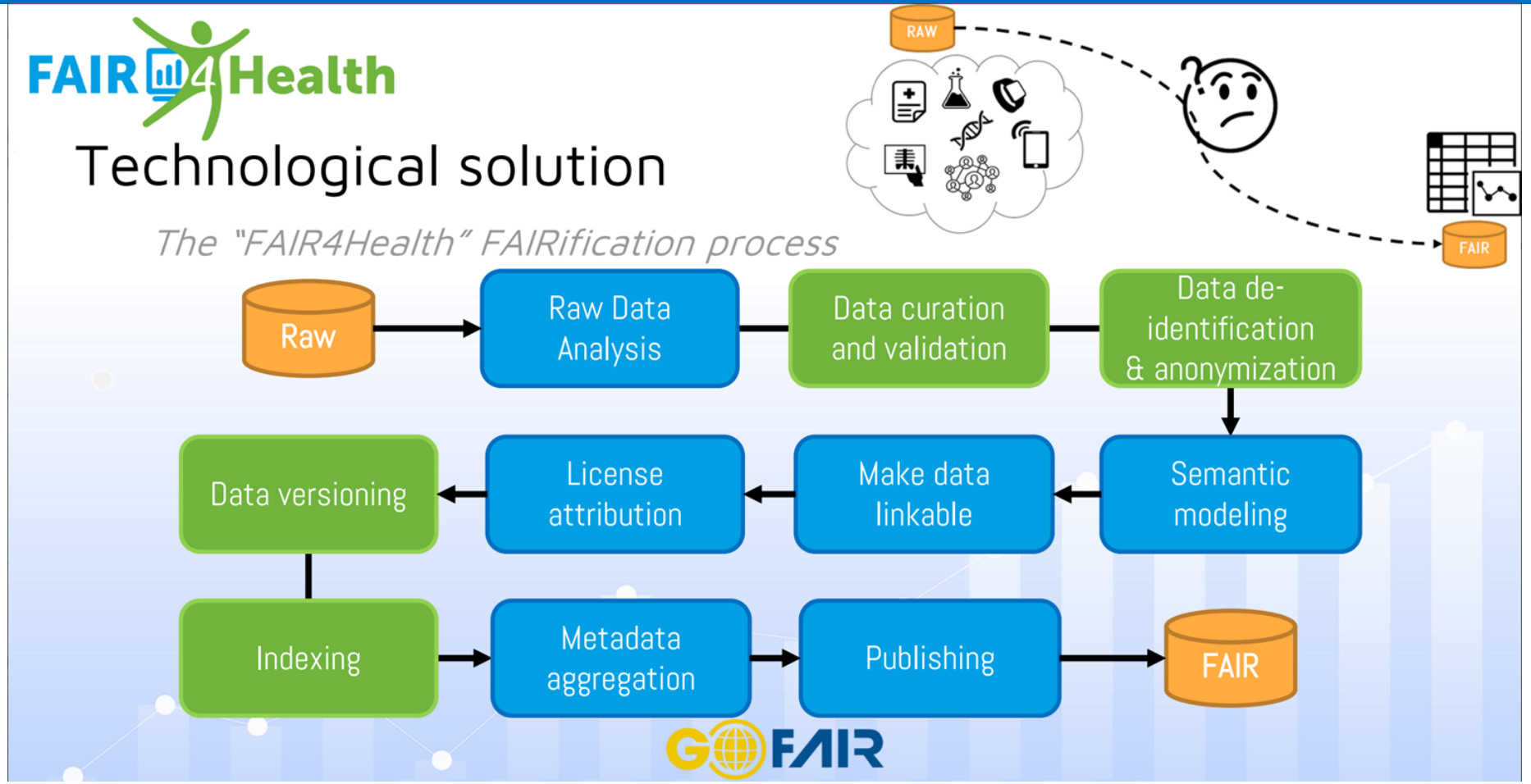


Figure 2: Clinical Study vs. TransFAIR study Design



DATA JOURNEY



DATA JOURNEY

FAIR4Health Data Privacy Tool

- Medical data usually needs to be de-identified/anonymized for Data Sharing
  - Patient rights and data protection
  - Anonymization isn‘t [yes|no]
  - But there is an information loss, which can be problematic for statistical analyses!
- The are various algorithms and best practices for anonymization
  - Suppression
  - Generalization
  - K-Anonymity
- The data anonymization tool will transform data according to specified parameters

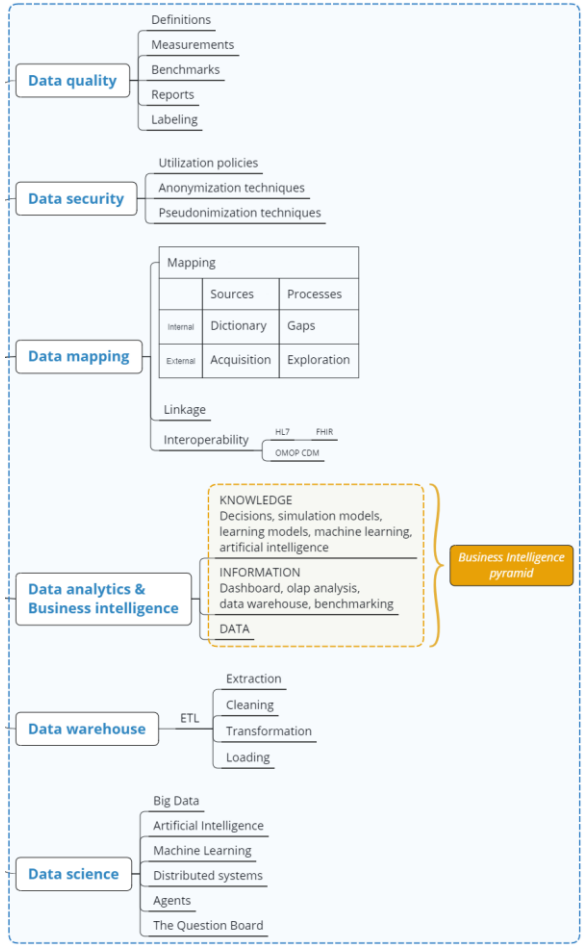
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

DATA LAB

Il **DATA LAB** è la frontiera innovativa della DATA UNIT orientata allo studio e all'applicazione delle tecniche di raccolta e gestione dei dati, di miglioramento della qualità, della sicurezza e dell'estrazione di conoscenza: si occupa di ottenere profonda competenza e padronanza dei sistemi di business intelligence, data warehousing e dei progetti di data science.

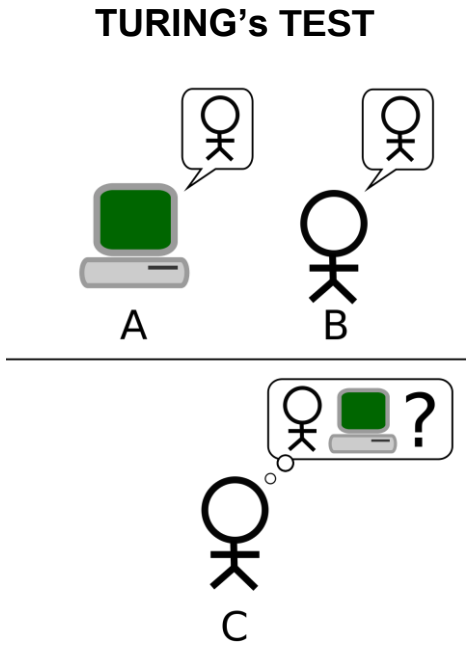
Produce strumenti e metodi per:

- **DATA QUALITY** ⇢ dimensioni e misurazione della qualità
- **DATA SECURITY** ⇢ politiche di protezione e pseudonimizzazione
- **DATA MAPPING** ⇢ semantica, linkage, basi per l'interoperabilità
- **DATA ANALYTICS** ⇢ business intelligence: from data to knowledge
- **DATA WAREHOUSE** ⇢ infrastrutture dati, ETL
- **DATA SCIENCE** ⇢ Big Data, Artificial Intelligence, sistemi distribuiti, agenti, the 'Question Board'



Artificial Intelligence

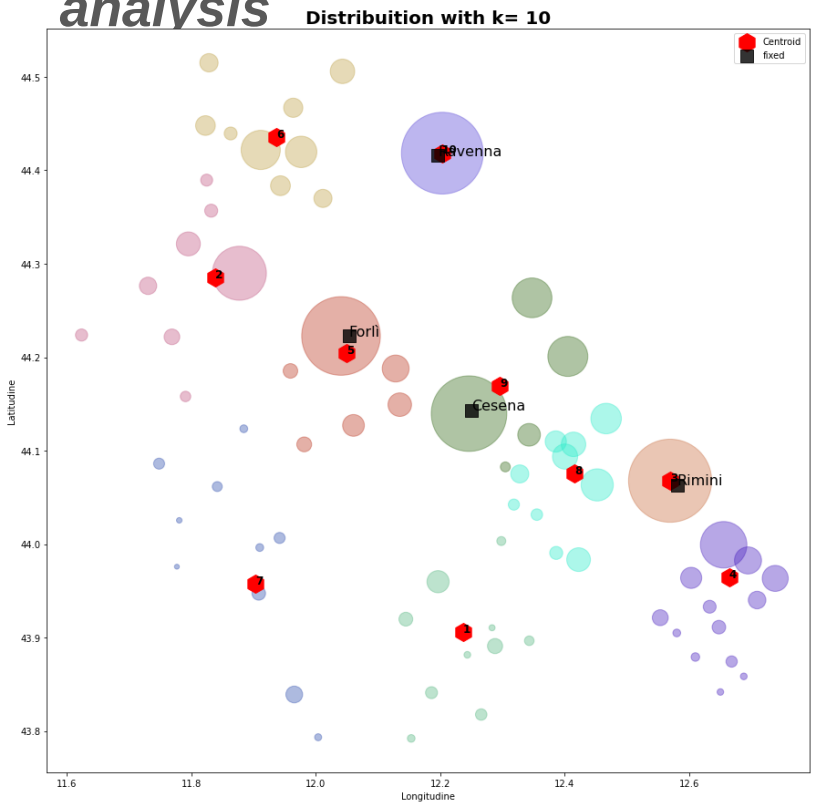
L'intelligenza artificiale (AI) è l'intelligenza dimostrata dalle macchine, al contrario dell'intelligenza degli esseri umani e di altri animali.
Esempi: riconoscimento vocale, visione artificiale, traduzione tra lingue (naturali)...



- *Abilità di una macchina di mostrare capacità umane quali ragionamento, apprendimento, pianificazione e creatività*
- *Disciplina che studia la progettazione, lo sviluppo e la realizzazione di sistemi capaci di simulare le abilità, il ragionamento e il comportamento umano in modo autonomo*

Machine Learning and Artificial Intelligence

How to reduce the burden of travel for the patients? A proof of concept (PoC) for oncological oral drugs distribution points analysis



Applicazione algoritmo di **Clustering** (creazione gruppi da dati in base a caratteristiche comuni – *apprendimento non supervisionato*) a partire da dati geografici e dati amministrativi regionali sull'erogazione dei farmaci oncologici (FED).

Obiettivo: minimizzazione gruppi e ottimizzazione delle distanze dal centro geografico identificato.

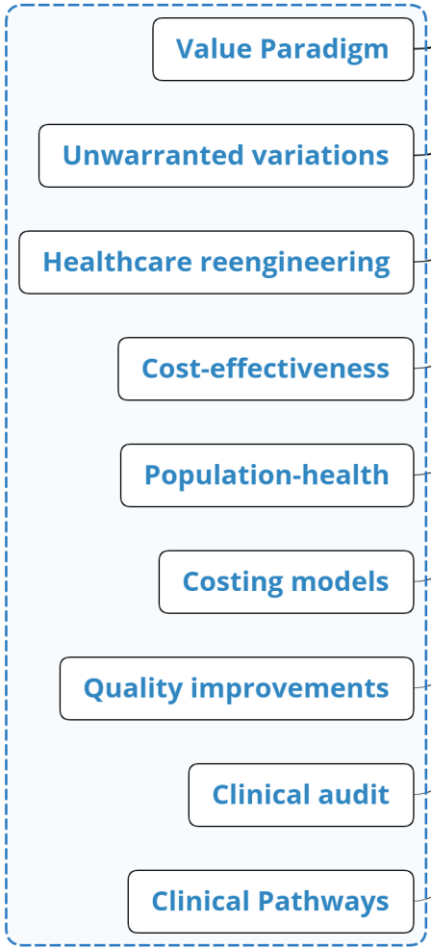
Ogni bolla rappresenta un comune, la grandezza il numero di pazienti e lo stesso colore rappresenta un cluster. I punti rossi sono i centroidi dei cluster rappresentati.

OUTCOME RESEARCH

L'unità di **Outcome Research** costituisce la frontiera operativa per la valutazione degli esiti: sfrutta i dati forniti dall'area **DATA SOURCES** e le metodologie e gli strumenti studiati ed implementati dal **DATA LAB**.

Le linee di analisi riguardano il paradigma del Value, le Unwarranted variations, la reingegnerizzazione dei processi sanitari, la cost-effectiveness, la population health, i modelli di costing, i progetti di miglioramento della qualità, i clinical audit, l'analisi dei percorsi e dei timing di cura, dei Key Performance Indicator predittivi di esito e legati all'appropriatezza, all'universalità, all'accessibilità, all'omogeneità.

Tra i suoi ambiti di lavoro nel breve medio termine ricade la valutazione delle modalità di coinvolgimento strutturato dei pazienti nell'analisi dei percorsi e nella valutazione degli esiti, ad esempio facendosi carico per l'Istituto assieme all'URP e allo IOR del supporto alla creazione, formazione, aggiornamento e di un'associazione pazienti e dell'implementazione delle interfacce di comunicazione tra l'Istituto, i pazienti e i caregiver al fine di migliorare la qualità dei servizi e gli esiti assistenziali.



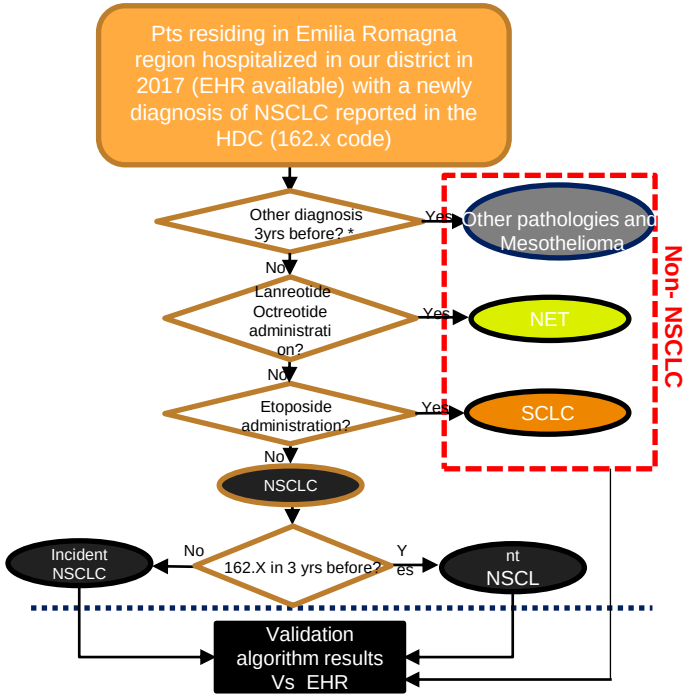
How to discriminate NSCLC cases from administrative databases?

Objectives and Method

To evaluate the performance of an algorithm developed to identify NSCLC incident cases from among a pool of patients with a diagnosis of ICD-9-CM 162.x code reported in the hospital discharge cards (HDC). Algorithm discrimination capacity to select both NSCLC or non-NSCLC was carried out on a sample for which electronic health record (EHR) diagnosis was available.

Results

A total of 430 patients were identified as lung cancer based on ICD-9 diagnosis. Focusing on incident case (n=314) algorithm had an overall accuracy of 82.8%. with a sensitivity of 88.8%. Algorithm confirmed a high level of PPV (90.2%), but lower specificity (53.7%) and NPV (50%). Higher LOS (>5 days) seemed to be associated with a correct classification. Hospitalization regimen and a supply of antiblastic therapy seemed to increase the level of PPV.



* (ICD-9- CM :140.x-161.x, 163.x-195.x, 200.x-208.x or V. 10. xx except V10.11 and V10.12)

Balzi W. et al. BMJ Open 2021; 11:1–7

EHR verified

Algorithm Classification	NSCLC	Non-NSCLC	Total
NSCLC	231	25	**256**
Non-NSCLC	29	29	**58**
Total	**260**	**54**	**314**

Conclusion

The algorithm demonstrated a strong validity for identify NSCLC patient in hospital administrative database. Modest specificity is due to misclassification of clinical diagnoses and coding errors in HDC.

SERVIZIO SANITARIO REGIONALE EMILIA-ROMAGNA
Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori

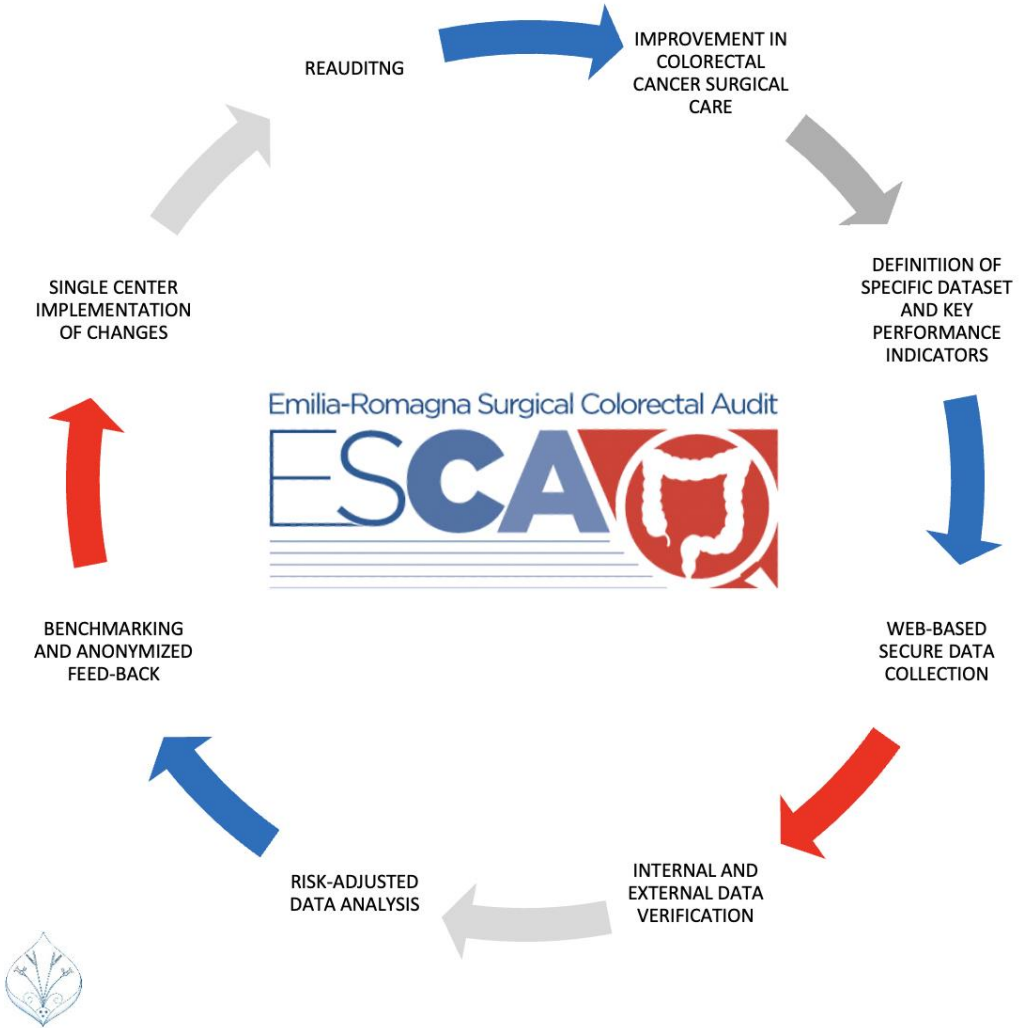
ISTITUTO SCIENTIFICO ROMAGNOLO PER LO STUDIO E LA CURA DEI TUMORI

Emilia-Romagna Surgical Colorectal Audit (ESCA)

"*Improving outcomes in cancer patients*

throughout a collaborative

and systematic auditing activity"

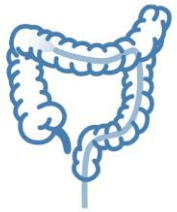


Emilia-Romagna Surgical Colorectal Audit

ESCA

1 PREOPERATIVE ASSESSMENT

2 DIAGNOSTIC WORK UP



3 SURGERY

4 POST-OPERATIVE COURSE

5 PATHOLOGY

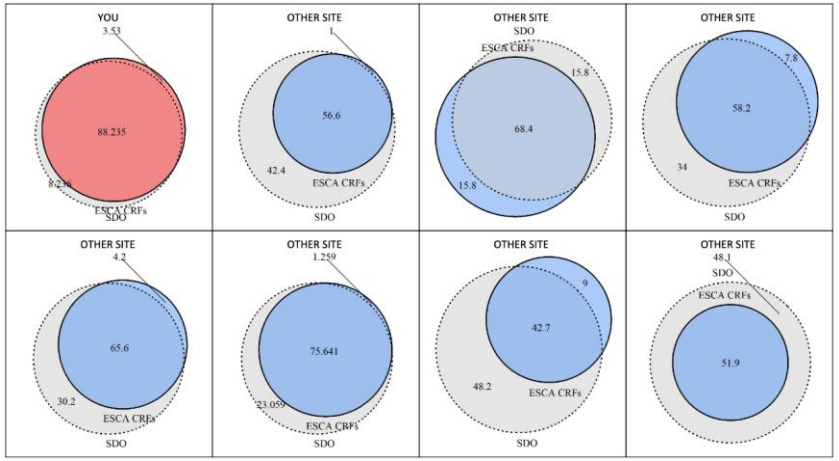
6 POST-OPERATIVE TREATMENT

7 FOLLOW-UP

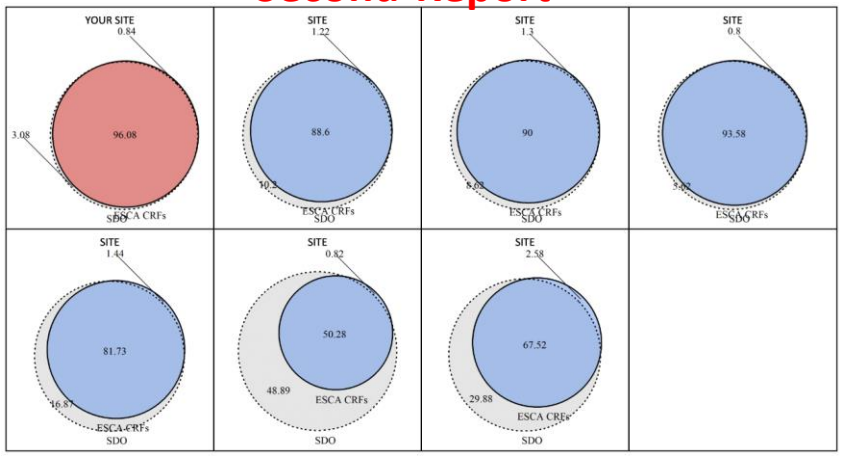
SERVIZIO SANITARIO REGIONALE EMILIA-ROMAGNA
Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori

ISTITUTO SCIENTIFICO ROMAGNOLO PER LO STUDIO E LA CURA DEI TUMORI

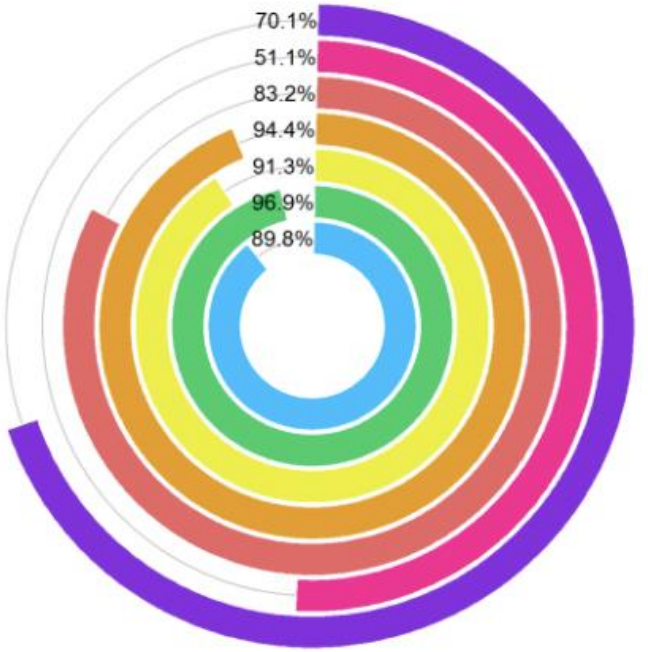
First Report



Second Report



Overall enrollment rate



74.6%

Feedback on enrollment

82.8%

SERVIZIO SANITARIO REGIONALE EMILIA-ROMAGNA
Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori

ISTITUTO SCIENTIFICO ROMAGNOLO PER LO STUDIO E LA CURA DEI TUMORI

Misure di accuratezza diagnostica e validazione di un test

(V. Torri)

Biomarker

Biomarker: definition

- A characteristics objectively measured and used as an indicator of normal biological processes, pathogenetic, or pharmacological response to therapeutic interventions.
  - Ex: plasma cholesterol concentration
- Although not all known biomarkers lead to validated diagnostic tests, there are, to date, biomarkers useful to determine the best therapeutic path for patients

Use of biomarkers

There are several purposes:

- as an outcome measure of the effectiveness of a treatment
- as an indicator of diagnosis
- as an indicator of prognosis
- as a predictor for the effectiveness of specific treatments

}

Diagnostic Tests

Overview

- Medical tests are procedures intended to detect, diagnose, characterize, or monitor a specific medical condition.

- The development of a medical test involves several investigative phases,
  - from an evaluation of its technical merits to
  - an assessment of its clinical (and even societal implications)

Clinical validation of a medical test

- During this step, the goal is to assess the test's ability to perform a certain clinical task (eg, to detect disease or provide a diagnosis).

- Studies in this arena can take many forms depending on the intended use and clinical role of the test.
  - used for
    - diagnosis,
    - screening,
    - staging
  - replace or supplement an existing test

Common features

- Although studies of medical tests can vary widely,
  - determining the study objective,
  - selecting a patient (and possibly reader) sample,
  - establishing a reference standard,
  - planning for statistical analysis.

  are building blocks shared by all study designs.

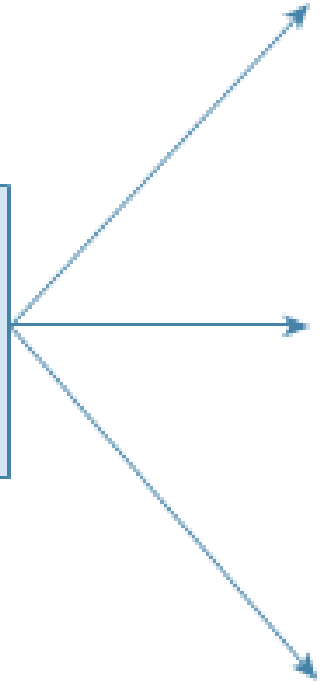
Steps in the development of a medical test

Technical efficacy
Evaluate technical performance (eg, spatial resolution of imaging, repeatability of biomarker measurement).

Clinical validation
Evaluate ability to perform a clinical task (eg, performance in detection, localization, or diagnosis of disease).

Clinical utility
Evaluate impact on clinical outcomes (eg, influence on diagnostic thinking, patient management, or patient outcomes).

Social efficacy
Evaluate impact on societal outcomes (eg, influence on population health outcomes, economic costs).



Exploratory study
Retrospectively assess performance in a less challenging population (eg, obvious cases of disease and healthy control subjects).

Challenge study
Retrospectively assess performance in a more challenging population (eg, patients with comorbidities).

Clinical study
Prospectively assess performance in intended use population.

Suptypes of study design

- Many medical tests involve a human reader who provides an interpretation.
- Multi-reader, multi-case studies are common in this scenario.
- When multiple readers are involved in a study comparing two medical tests, there are various designs for collecting the readers' interpretations.

Paired and unpaired designs

- In a paired-reader design, the same readers provide interpretations for both medical tests.
- In a paired-patient design, the same patients undergo each medical test.
  - The advantage of both of these designs is increased statistical power because the readers and patients are serving as their own controls.
- In an unpaired-reader design, a different set of readers provides interpretations for each test.
- In an unpaired-patient design, a different set of patients undergoes each medical test.

- The most common multi-reader, multi-case design is paired-reader, paired-patient, but this is not always practical

Designs for head to head comparison

- New medical tests are often compared head-to-head with existing medical tests.
- A common example is the assessment of computer-aided devices (CADs), algorithms specifically developed to assist human readers during their interpretation of a medical image.
- There are two common study designs in this scenario.
- sequential design,
  - the reader first gives his or her interpretation without the CAD (unaided read).
  - Then, the CAD is applied, and the reader immediately gives a second interpretation after incorporating the information provided by the device (aided read).
- crossover design,
  - the aided and unaided interpretations are given during separate reading sessions, typically with a washout period in between.

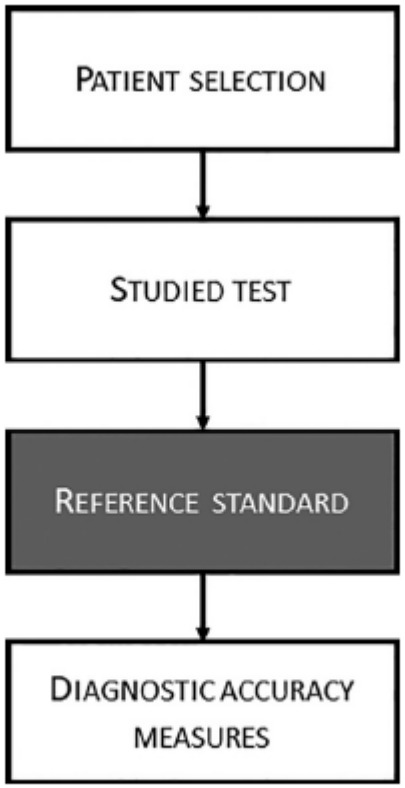
Sequential and crossover designs: pros & cons

- Sequential design
  - minimizes variability between the aided and unaided interpretations and makes for a shorter study, as there is only one reading occasion and no washout period.
  - there is a potential for bias both because the unaided interpretations are always given first and because readers may pay more or less attention during the unaided reads as they know they are about to get the CAD results.
- Crossover design
  - reduces the potential for these biases
  - requires a washout period, increases the duration of the study, and results in more variability in the interpretations.

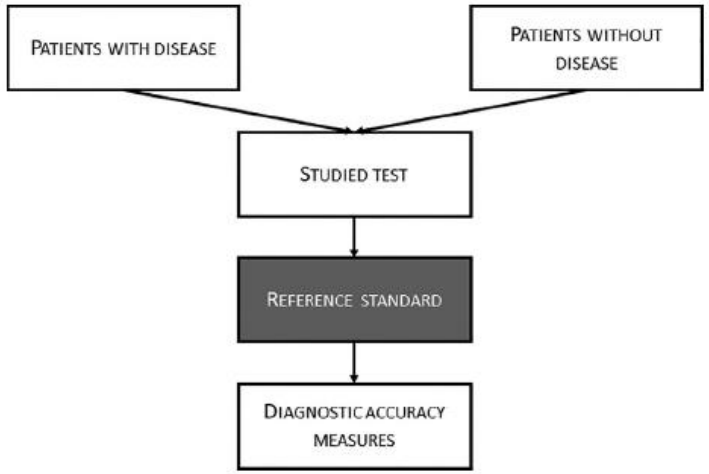
Use Cases of Study Design

- Even with a focus on clinical validation, studies of medical tests can vary widely.
- Possible goals
  - determine the performance of a single test
    - Ex: to assess the accuracy of V_/Q_ scans for acute pulmonary embolism [PIOPED study]
  - compare the performance of multiple tests
- Tests can be compared head-to-head or in an adjunct framework
- It is possible also to extend the goals beyond clinical validation, with interest in studying also the clinical utility

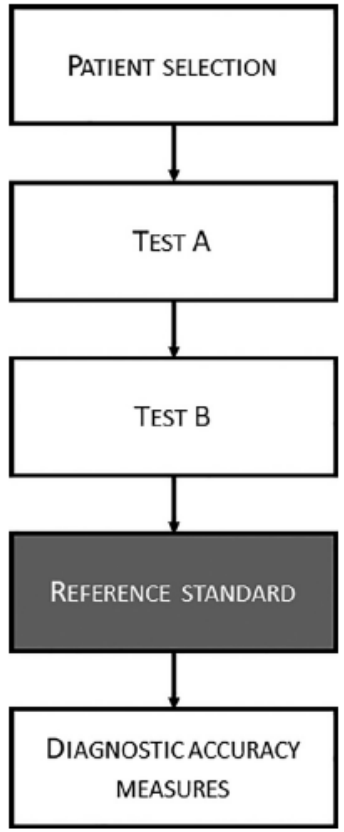
Example of study design



Diagnostic accuracy cross-sectional studies.

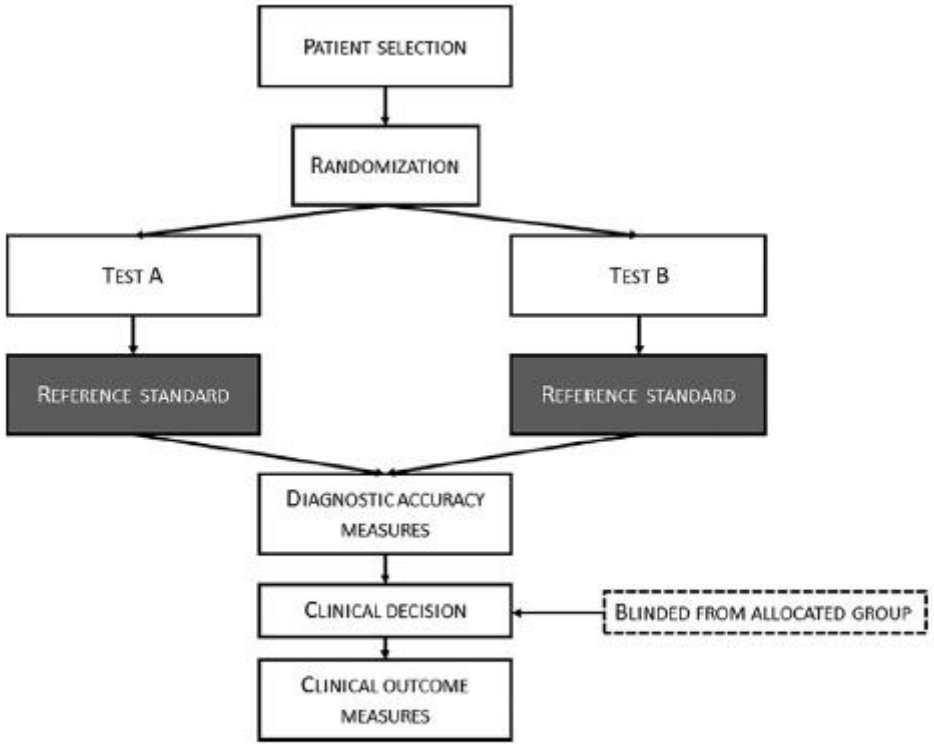


Diagnostic accuracy case-control studies.

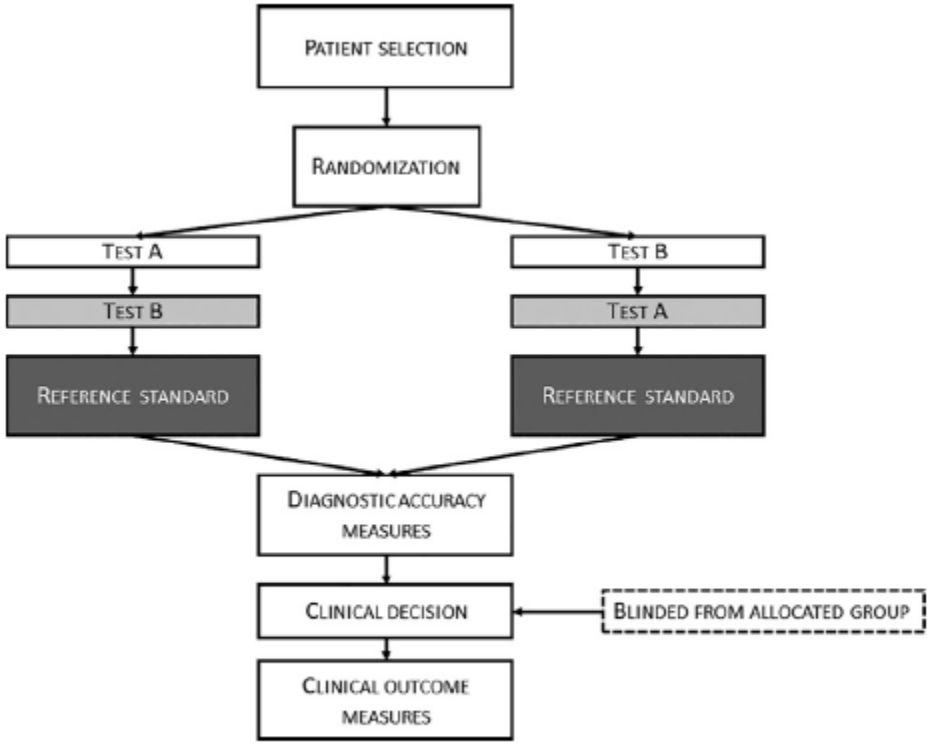


Nonrandomized comparative diagnostic accuracy study.

Example of study design



Randomized diagnostic accuracy study—option 1.



Randomized diagnostic accuracy study—option 2.

Benefits of study design

- Assessing the accuracy of medical tests is a critical part of informed decision-making in the world of health care.
- Accuracy studies allow investigators to report metrics such as the test's
  - sensitivity,
  - specificity
  - the negative and positive predictive values.
- Accuracy studies allow investigators to describe the types of errors a test may be prone to making and how often they happen.
- This sort of information is crucial for clinicians who must ultimately incorporate the results from medical tests into patient management

Common design goals and potential biases

Goal	What Can Go Wrong?
Select an appropriate sample	**Verification bias**
	Can happen when the medical test result influences the probability that a subject has a reference standard result and the sample is limited to subjects with the reference standard result
	Selection bias
	Can happen when an external factor has influenced the selection process so that the sample is not representative of the target population
Establish the ground truth for each subject at the time of the medical test	**Treatment paradox bias**
	Can happen when the treatment is administered between the medical test and the reference standard
	Disease progression bias
	Can happen when the disease has had time to progress (or regress) between the medical test and reference standard
	Imperfect gold standard bias
	Can happen when the reference standard is not 100% accurate
Determine the medical test result and the reference standard result independently	**Test review bias**
	Can happen when the readers interpreting the medical test are not blinded to the reference standard results
	Diagnostic review bias
	Can happen when the reference standard was performed with knowledge of the medical test results
	Incorporation bias
	Can happen when the results from the medical test are used, even in part, to establish the reference standard

Study Subject Considerations

- Patient Sample
  - In a retrospective design, the true disease state of the patient is known at the time of selection.
  - In a prospective design, the true disease state is not known at the time of selection.
  - with retrospective studies, the sample prevalence is often intentionally greater than the population prevalence so that enough diseased subjects are included to estimate sensitivity with reasonable precision.

Study Subject Considerations

- Patient Sample in exploratory studies
  - Exploratory studies are often retrospective, with the patient sample being selected from medical records or disease-specific registries.
  - The subjects with the condition are often typical, obvious cases of the disease, and those without the condition are typically normal volunteers

Study Subject Considerations

- Patient Sample in challenge studies
- Challenge studies tend to be retrospective, so that an adequate range of subjects can be selected.
- The subjects with the condition should include examples from the range of pathologic, clinical, and comorbid characteristics of the population.
- The subjects without the condition should be similar on these parameters

Study Subject Considerations

- Patient Sample in clinical studies
- In clinical studies, a retrospective or prospective design can be used to select subjects that closely represent the target population. In this phase, it is important to avoid common biases
  - selection bias, which can occur if an external factor has influenced the selection process so that the sample is not representative of the target population.

Reader Sample

- Studies of medical tests that require interpretation by a human reader effectively have two samples to consider:
  - patient and reader.

Reader Sample

- Exploratory and challenge studies may target a narrower reader population
  - readers at a single institution
- convenience sampling is often used
  - selection is based on reader expertise and willingness to participate.

Reader Sample

- During clinical studies,
  - it is important to select a sample of readers that is representative of the readers who will be interpreting the test if it is shown to be sufficiently accurate.
  - this ideally means random sampling or stratified random sampling of readers across relevant medical centers.

Sample Size

- In the context of diagnostic accuracy studies there are methods for any study objectives including
  - the estimation of a single test's accuracy
  - the assessment of superiority (or noninferiority) of one medical test compared to another.
  - the combined effect of reader and patient sample sizes
  - the estimation of accuracy endpoints
    - area under the receiver-operating characteristic [roc] curve
    - sensitivity
    - specificity

Selecting a statistical method in the context of an accuracy study

Scale of ground truth results	Scale of medical test results
What scale is the reference standard on? • For most accuracy studies, there will be two ground truth states (eg, with disease, without disease). • However, ground truth results may occasionally have more than two categories or even be on an ordinal or continuous scale.	What scale is the medical test on? • A binary test can take one of two values (eg, positive for condition, negative for condition). • An ordinal test can take one of any number of ordered categories (eg, unlikely to have condition, equivocal, likely to have condition). • A continuous test results in a numerical value (eg, percent confidence in the presence of condition).
No. of samples	**Accuracy measure**
How many samples are there? • A one-sample study may estimate the accuracy of a single reader using one modality. • A two-sample study may compare two readers using the same modality or one reader using two different modalities. • A multi-sample study may involve multiple readers and one or more modalities.	What measure will be used to characterize the accuracy of the medical test? • If the medical test is on a binary scale, options include sensitivity, specificity, and predictive values. • If the medical test is on an ordinal or continuous scale, options include the full (or partial) area under the ROC curve, sensitivity at a fixed false-positive rate, and specificity at a fixed true positive rate.
Independent/clustered data	**Parametric/nonparametric approach**
Are the medical test results independent or clustered? • If there is more than one observation per subject (eg, the unit of analysis is the lung and both of a patient's lungs are included in the sample or the unit of analysis is the lesion and multiple lesions from the same patient are included in the sample), the data are clustered.	Will the analysis take a parametric or nonparametric approach? • There are pros and cons to each approach and the decision is nuanced. • Generally, nonparametric approaches are simpler, make fewer assumptions, and perform better in smaller samples.

Selecting an endpoint

Accuracy measures assess the performance of a medical test relative to a gold standard.

Agreement measures assess the concordance between two medical tests and do not require a ground truth

Count table (2×2) for agreement data

	Medical test 2	
Medical test 1	*Positive*	*Negative*
Positive	*a*	*b*
Negative	*c*	*d*

Example measures of agreement

Percent agreement

$$\frac{a+d}{a+b+c+d}$$

Cohen's kappa

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

where...

$$p_o = \frac{a+d}{a+b+c+d}$$

$$p_e = p_{e1} + p_{e2}$$

$$p_{e1} = \frac{c+d}{a+b+c+d} * \frac{b+d}{a+b+c+d}$$

$$p_{e2} = \frac{a+b}{a+b+c+d} * \frac{a+c}{a+b+c+d}$$

Count table (2×2) for accuracy data

	Reference Standard	
Medical test	*Positive*	*Negative*
Positive	*w*	*x*
Negative	*y*	*z*

Example measures of accuracy

Sensitivity

$$\frac{w}{w+y}$$

Specificity

$$\frac{z}{x+z}$$

Positive predictive value

$$\frac{w}{w+x}$$

Negative predictive value

$$\frac{z}{y+z}$$

Positive likelihood ratio

$$\frac{Sensitivity}{1 - Specificity}$$

Negative likelihood ratio

$$\frac{1 - Sensitivity}{Specificity}$$

Diagnostic OR

$$\frac{Positive\ likelihood\ ratio}{Negative\ likelihood\ ratio}$$

Youden's index

$$Sensitivity + Specificity - 1$$

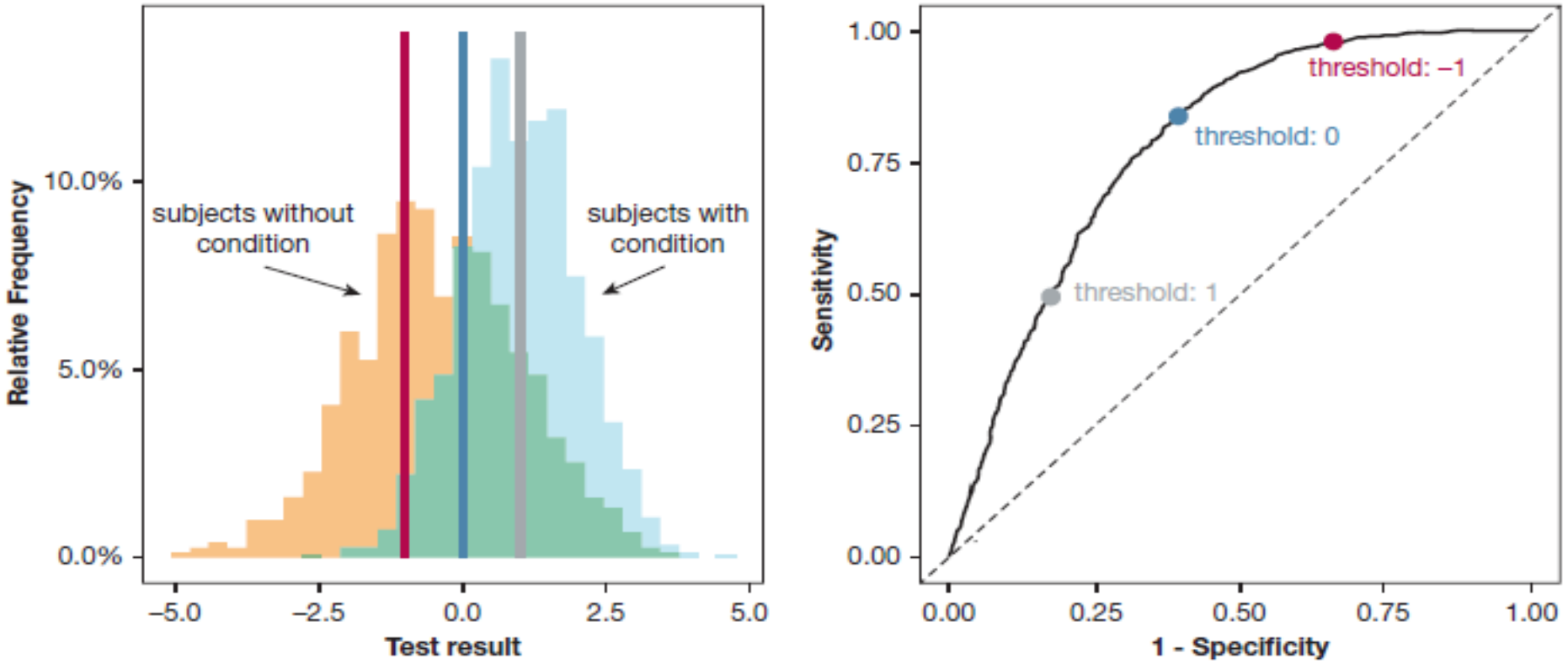
Accuracy endpoints

- Sensitivity, specificity, and the area under the ROC curve are the most common accuracy measures.
  - ROC curve can be used when the medical test results are ordinal or continuous.
    - It summarizes the sensitivity and specificity across all possible thresholds for defining a positive test result

Accuracy endpoints

- The area under the ROC curve
  - in exploratory and challenge studies is often used because it addresses the primary question of whether the test can be used to distinguish patients with vs. without the condition of interest.
  - For clinical studies, the area under the ROC curve is typically not recommended as a primary end point because it is too global.
  - A clinically interpretable measure that is relevant for the application should be selected.
    - Eg: sensitivity, specificity, positive predictive value, negative predictive value, and partial area under the ROC curve.

Sensitivity, specificity and ROC curve



Relation between sensitivity, specificity, and ROC curve. On the left, histogram of simulated (continuous scale) test results. Subjects without the condition are in orange, and subjects with the condition are in blue (overlap in histograms is in green). The vertical lines represent three potential thresholds for determining what signifies a positive test result. As the threshold moves to the right, more of the subjects without the condition are correctly specified (ie, specificity increases), but fewer of the subjects with the condition are correctly specified (ie, sensitivity decreases). On right, the ROC curve for the simulated data is illustrated. The three sensitivity/specificity pairs defined by the thresholds in the plot on the left are indicated with a dot on the ROC curve

STARD checklist

Objective	• Is this an exploratory, challenge, or clinical study? • What is your primary objective? What is the primary end point?
Sample	• What is the target patient population? How will you sample patients from this population (consecutive, random, or convenience series)? • If the medical test involves interpretation by a human reader, what is the target reader population? How will you sample readers from this population? • How many patients/readers do you need to have enough statistical power for your primary objective? • How will you avoid common biases during sample selection (eg, selection bias)?
Data collection	• How will you collect the medical test results? What format will they take (binary, ordinal, or continuous)? • If the medical test involves interpretation by a human reader, how will you train the readers and in what environment will they provide the interpretations? • What will you use as a reference standard? How will you collect the reference standard?
Analysis	• How will you avoid common biases during data collection (eg, verification bias, treatment paradox bias, disease progression bias, imperfect gold standard bias, test review bias, diagnostic review bias, incorporation bias, reading order bias)? • What is your statistical hypothesis? • What is the unit of analysis? Are special methods for clustered data required? • Is covariate adjustment needed?
Reporting	• Have you consulted the STARD reporting checklist?

Appraisal of a diagnostic test results

How to assess the results of the test

- There are two types of result commonly reported in diagnostic test studies.

- One concerns the accuracy of the test and is reflected in the sensitivity and specificity, often defined as the test's ability to find true positives for the disorder (sensitivity) or true negatives for the disorder (specificity).
  - An ideal diagnostic test finds no false positives but at the same time misses no one with the disease (finds no false negatives).

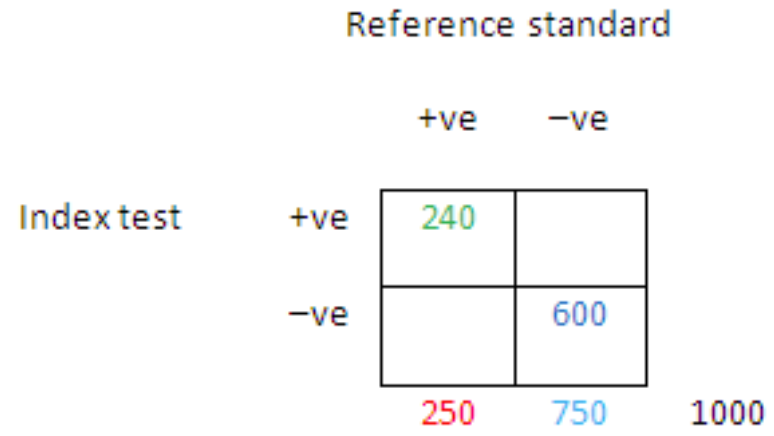
- The other concerns how the test performs in the population being tested and is reflected in predictive values (also called post-test probabilities) and likelihood ratios.

Example

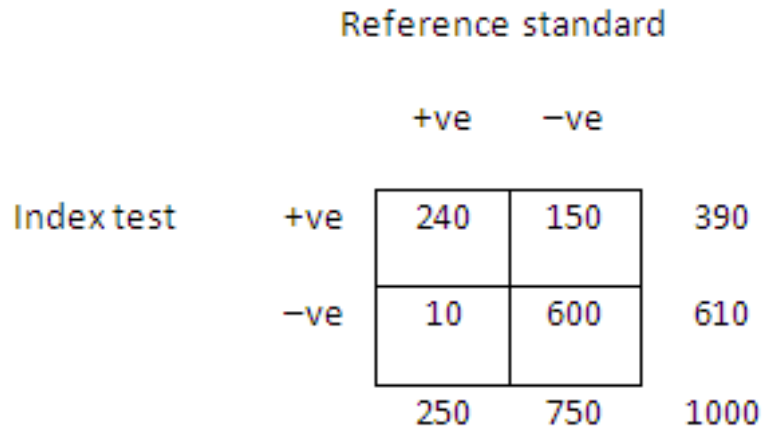
- 1000 elderly people with suspected dementia undergo an index test and a reference standard. The prevalence of dementia in this group is 25%.
- 240 people tested positive on both the index test and the reference standard and
- 600 people tested negative on both tests.
- The remaining 160 people had inaccurate test results.

Example

- The first step is to draw a 2x2 table.
- We are told that the prevalence of dementia is 25%; therefore, we can fill in the last row of totals - 25% of 1000 people is 250 - so 250 people will have dementia and 750 will be free of dementia. We also know the number of people testing positive and negative on both tests and so we can fill in two more cells of the table.



By subtraction the table can easily be completed:



From the 2x2 table the following measures can be calculated

Term	Definition	Example
Pre-test probability = (true positive + false negative)/total number of people	This measure tells us the probability of having a target condition before a diagnostic test	In this example: 250/1000 = 0.25 What does this mean: The probability of a patient in this study having dementia before the tests are run
Sensitivity (Sn) = the proportion of people with the condition who have a positive test result	The sensitivity tells us how well the test identifies people with the condition. A highly sensitive test will not miss many people	In our example, the Sn = 240/250 = 0.96 What does that mean? 10 (4%) people with dementia were falsely identified as not having it, as opposed to the 240 (96%) people who were correctly identified as having dementia. This means the test is fairly good at identifying people with the condition
Specificity (Sp) = the proportion of people without the condition who have a negative test result	The specificity tells us how well the test identifies people without the condition. A highly specific test will not falsely identify many people as having the condition	In our example, the Sp = 600/750 = 0.80 What does that mean? 150 (20%) people without dementia were falsely identified as having it. This means the test is only moderately good at identifying people without the condition

From the 2x2 table the following measures can be calculated:

Term	Definition	Example
Positive predictive value (PPV) = the proportion of people with a positive test who have the condition	This measure tells us how well the test performs in this population. It is dependent on the accuracy of the test (primarily specificity) and the prevalence of the condition	In our example, the PPV = 240/390 = 0.62 What does that mean? Of the 390 people who had a positive test result, 62% will have dementia
Negative predictive value (NPV) = the proportion of people with a negative test who do not have the condition	This measure tells us how well the test performs in this population. It is dependent on the accuracy of the test and the prevalence of the condition	In our example, the NPV = 600/610 = 0.98 What does that mean? Of the 610 people with a negative test, 98% will not have dementia

From the 2x2 table the following measures can be calculated

Term	Definition	Example
Positive likelihood ratio (LR+) = sensitivity / (1- specificity)	This measure tells us how much the odds of a specific diagnosis increase when a test is positive. The larger the LR+, the more likely it is that the person with a positive test result has the condition. An LR+ of 10 indicates a 10-fold increase in the odds of the patient having the condition (i.e., a large increase in probability), whereas an LR+ of 2 would indicate a modest increase in the odds of the patient having the condition. An LR+ of 1 would mean that the test provides no new information regarding the odds of the patient having the condition.	In this example the LR+ = 96%/20% = 4.8 What does that mean? There is a 4.8 fold increase in the odds of having dementia in a person with a positive test (i.e., a moderate increase in the probability that they have dementia)
Negative likelihood ratio (LR–) = (1-sensitivity) / specificity	This measure tells us how much the odds of a specific diagnosis decrease when a test is negative. The smaller the LR-, the more likely it is that the person with a negative test result does not have the condition. An LR- of 0.5 indicates a 2 fold decrease in the odds of the patient having the condition (i.e., a modest decrease in probability), whereas an LR- of 0.1 indicates a 10-fold decrease in the odds of having the condition (i.e., a large decrease in probability).	In this example LR– =4%/80% = 0.05 What does that mean? There is a 20-fold decrease in the odds of having dementia in a person with a negative test result (i.e., a large decrease in the probability that they have dementia)

Appraisal of results

- How to apply the diagnostic test to a specific patient
  - Is the test available, affordable, and accurate in your setting?
  - Can a clinically sensible estimate of the pre-test probabilities of the patient be made from personal experience, prevalence statistics, practice databases, or primary studies?
  - Are the study patients similar to the patient in question?
  - How current is the study we are analysing - has evidence moved on since the publication of the study?

Appraisal of results

- **Will the post-test probability affect the management of the specific patient?**
  - Could the result move the clinician across a test-treatment threshold: for example, could the results of the test stop all further testing? That is, rule the target disorder out so the clinician would stop pursuing that possibility, or make a firm diagnosis of the target disorder and move onto choosing appropriate treatment options.
  - Will the patient be willing to have the test carried out?
  - Will the results of the test help the patient reach their goals?

Appraisal of results

- **Will the post-test probability affect the management of the specific patient?**
  - Could the result move the clinician across a test-treatment threshold: for example, could the results of the test stop all further testing? That is, rule the target disorder out so the clinician would stop pursuing that possibility, or make a firm diagnosis of the target disorder and move onto choosing appropriate treatment options.
  - Will the patient be willing to have the test carried out?
  - Will the results of the test help the patient reach their goals?

List of questions

- The study population.
  - How close did the study population represent the overall intended use population?
  - Did the study population include relevant subgroups?
  - Were there potential biases resulting from the subject selection process?
- The accuracy of the test results.
  - Was a reasonable reference standard identified, described, and applied?
  - Was the test accuracy internally vali- dated?
  - Was a clinically meaningful test result threshold(s) identified?
  - Was the accuracy of the test assessed in an independent sample of representative subjects after establishing the diagnostic threshold?
  - Was variability of the test result re- ported or discussed?
- The potential utility of the test.
  - Were the test results compared vs current standard of care, alone or in combination?
  - Was a potentially clinically relevant increment in accuracy identified, and were tradeoffs in changes in sensitivity and specificity discussed?
  - Was the frequency with which the result of the test could affect a decision reported?
  - If the study claims clinical utility, was use of the test proven to affect clinical decisions in a manner resulting in more benefit than harm in the intended use population?

Screening

Outcome measure for screening

- the appropriate measure is mortality, calculated on the entire population studied.
  - consider all potential screened patients, not just the minority that develops disease of interest.
  - is not biased by early detection

$$\frac{\text{Deaths for disease}}{\text{persons-year}}$$

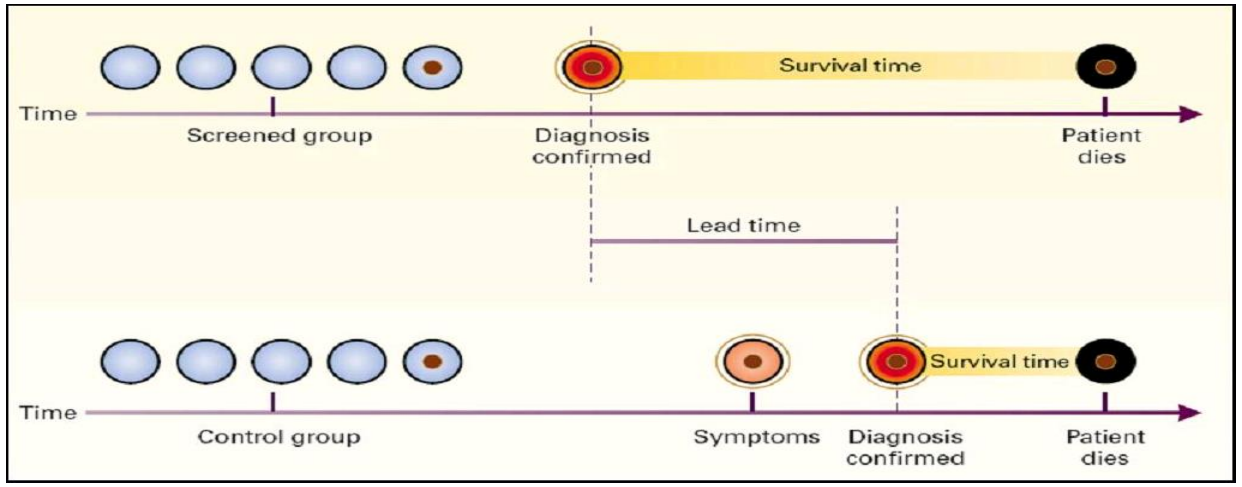
Screening

- Facilitates early diagnosis and allows a timely treatment
  - PAP test - uterine cervix
  - Colonoscopy - right colon
  - Mammography - breast

Why RCTs are needed for evaluating screeening efficacy

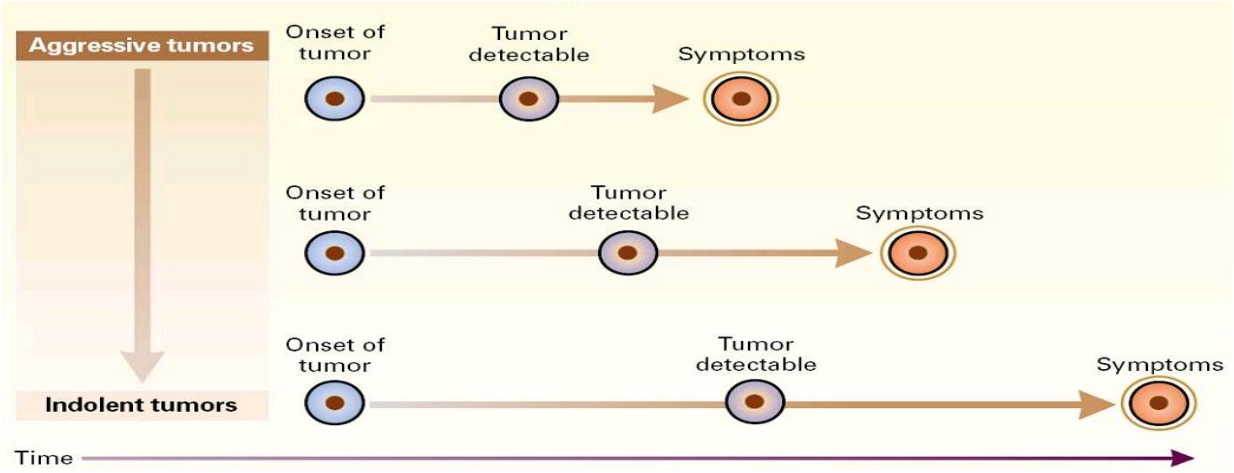
- Comparing the prognosis of a group of cases diagnosed in a screening with that of cases diagnosed normally does not provide information on the effectiveness of screening
  - lead time bias
  - length time bias
  - Overdiagnosis bias

Lead time bias



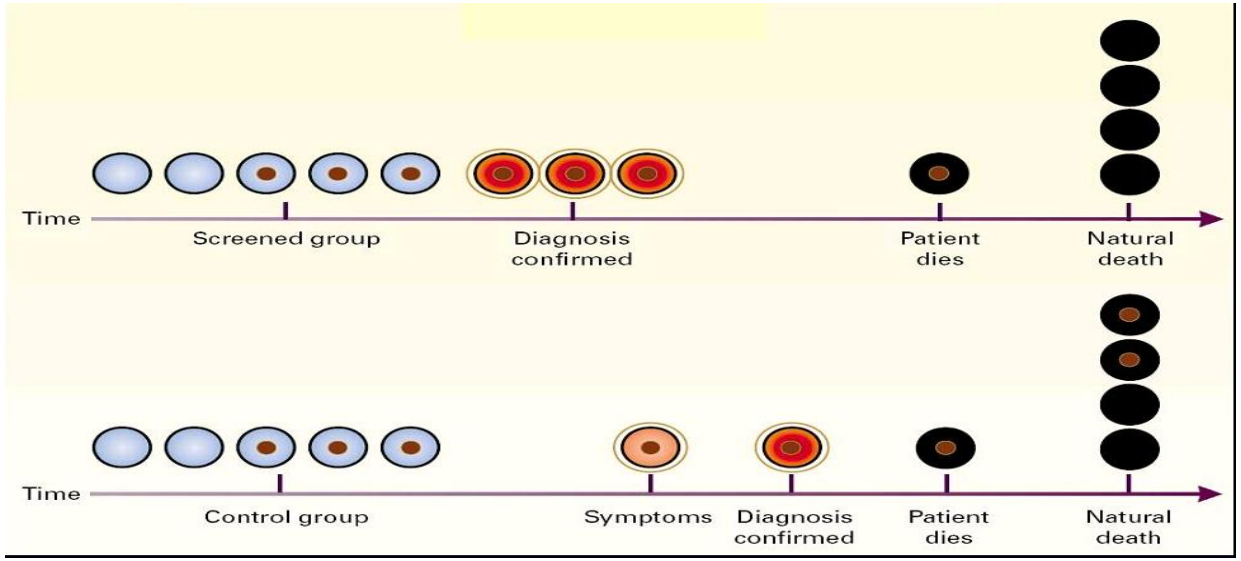
- diagnostic anticipation results in an (apparent) lengthening of the survival period, even if death is not postponed

Lenght bias



- screening tends to detect slow-growing cases, however destined for a better prognosis, or not to give symptoms in the course of life

Overdiagnosis bias



- screening identifies proportionately more conditions that would not have become clinically significant had they not been identified

What factors determine the effectiveness of screening?

- The prevalence (risk) of disease.
- The effectiveness of screening in preventing illness or death.
  - Is the test any good at detecting disease/precursor (sensitivity of the test)?
  - Is the test detecting a clinically relevant condition?
  - Is there anything we can do if disease (or pre-disease) is detected (cures, treatments)?
  - Does detecting and treating disease at an earlier stage really result in a better outcome?
- The risks of screening, such as false positives and radiation.

Example: Mammography

- Mammography utilizes ionizing radiation to image breast tissue.
- The examination is performed by compressing the breast firmly between a plastic plate and an x-ray cassette that contains special x-ray film.
- Mammography can identify breast cancers too small to detect on physical examination.
- Early detection and treatment of breast cancer (before metastasis) can improve a woman's chances of survival.
- Studies show that, among 50-69 year-old women, screening results in 20-35% reductions in mortality from breast cancer.

Mammography

- Controversy exists over the efficacy of mammography in reducing mortality from breast cancer in 40-49 year old women.
- Mammography has a high rate of false positive tests that cause anxiety and necessitate further costly diagnostic procedures.
- Mammography exposes a woman to some radiation, which may slightly increase the risk of mutations in breast tissue.

Calculating sensitivity and specificity from a 2x2 table

Truly have disease	Screening Test +	Screening Test -	
+	**a**	**b**	a+b
-	**c**	**d**	c+d

Sensitivity $= \frac{a}{a+b}$

Among those with true disease, how many test positive?

Specificity $= \frac{d}{c+d}$

Among those without the disease, how many test negative?

Hypothetical Example

Breast cancer (on biopsy)	Mammography +	Mammography -	
+	**9**	**1**	10
-	**109**	**881**	990

Sensitivity=9/10=.90

1 false negatives out of 10 cases

Specificity= 881/990 =.89

109 false positives out of 990

Example

- A 60-year old woman has an abnormal mammogram; what is the chance that she has breast cancer? E.g., what is the positive predictive value?

Calculating PPV and NPV from a 2x2 table

Screening Test

Truly have disease

	+	-
+	a	b
-	c	d
	a+c	b+d

PPV $= \frac{a}{a+c}$

Among those who test positive, how many truly have the disease?

NPV $= \frac{d}{b+d}$

Among those who test negative, how many truly do not have the disease?

Hypothetical Example

Breast cancer (on biopsy)	Mammography +	Mammography -
+	**9**	**1**
-	**109**	**881**
	118	882

PPV= 9/118= 7.6%

NPV=881/882=99.9%

Prevalence of disease = 10/1000 =1%

What if disease was twice as prevalent in the population?

Breast cancer (on biopsy)	Mammography +	Mammography -	
+	**18**	**2**	20
-	**108**	**872**	980

sensitivity= 18/ 20=.90

specificity=872/980=.89

Sensitivity and specificity are characteristics of the test, so they don't change!

What if disease was more prevalent?

Mammography

Breast cancer (on biopsy)

	+	-
+	**18**	**2**
-	**108**	**872**
	126	874

PPV= 18/126=14.3%

NPV=872/874=99.8%

Prevalence of disease = 20/1000 =2%

Relationship of predictive value with prevalence and test performance: Bayes rule

$$\text{Positive predictive value (PPV)} = \frac{\text{Sn}\times\text{Prev}}{\text{Sn}\times\text{Prev}+(1-\text{Sp})\times(1-\text{Prev})}$$

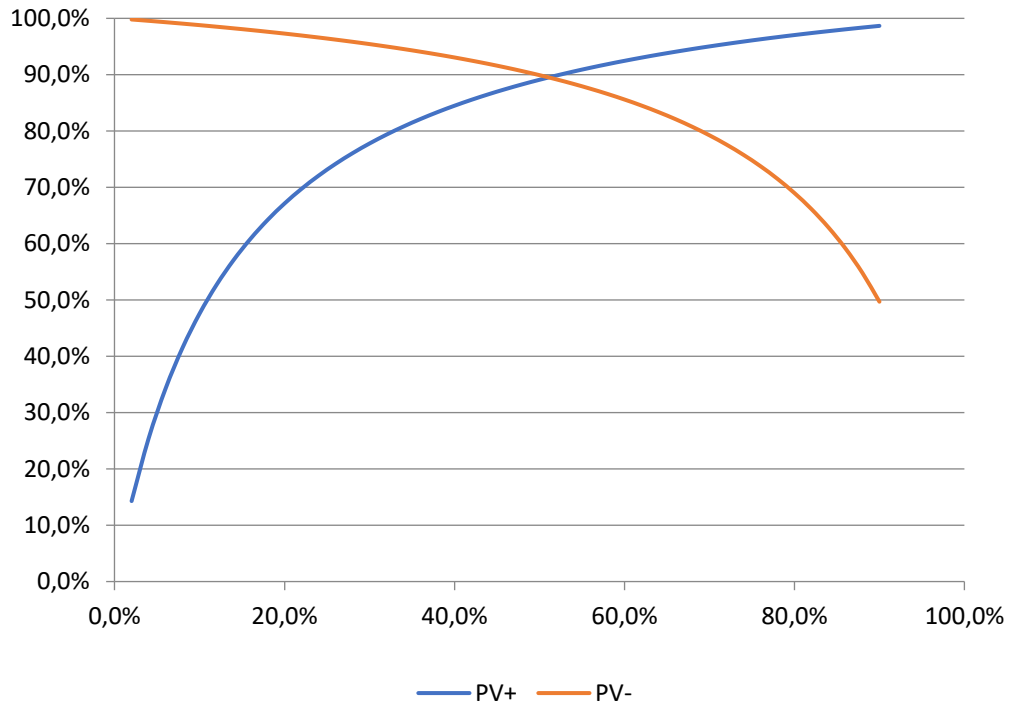
$$\text{Negative predictive value (NPV)} = \frac{\text{Sp}\times(1-\text{Prev})}{(1-\text{Sn})\times\text{Prev}+\text{Sp}\times(1-\text{Prev})}$$

$Odds_{PPV} = Odds_{Prev} * LH^{+}$ [$LH^{+}=Sn/(1-Sp)$]

$Odds_{NPV} = Odds_{1-Prev} * LH^{-}$ [$LH^{-}=Sp/(1-Sn)$]

Sensitivity 90%

Specificity 89%



Sviluppo e validazione di un modello prognostico

(V. Torri)

Introduction

"Validation is one of those words... that is constantly used and seldom defined". Alvan Fainstein

- Prognostic models are used in medicine for investigating patient outcome in relation to patient and disease characteristics.
- Such models do not always work well in practice, so it is widely recommended that they need to be validated.
- The idea of validating a prognostic model is generally taken to mean establishing that it works satisfactorily for patients other than those from whose data it was derived.

Definition and purpose of prognostic models

- Prognostic models are regression models relating to patients outcome
- They are widely used in cancer and other medical specialties for investigating patient outcome in relation to patient and disease characteristics

- In some studies the aim is determining
  - which variables are associated with prognosis,
  - whether a particular variable is prognostic after allowance for other, previously identified prognostic variable

- The clinical purpose may be
  - to inform treatment or other clinical decisions for individual patients;
  - to inform patients and their families;
  - to create clinical risk groups for informing treatment or for stratifying patients by disease severity in clinical trials.

Assessing the usefulness of a prognostic model

"Any classification system, be it nominal, ordinal, or scalar, should be proved to be a workable tool before it is used in a discriminatory or predictive manner." Burstein AH

Usefulness is determined by how well a model works in practice, not by how many zeros there are in the associated *P*-values

Usefulness of a prognostic model

- There are two broad ways in which a model may be useful.
  - allows the reliable classification of patients into two or more groups with different prognoses.
    - Such classification schemes can be used to influence therapy or save patients from unnecessary referrals or tests.
  - can estimate the prognosis of individual patients.

- While in some sense these are two different ways of looking at the same information, they differ fundamentally.

Limitation of a prognostic models

- A basic issue relating to predicting events is that binary outcome data have considerable irreducible variability.

- The data are all 0 and 1, but the predictions are probabilities lying between these extremes.

- Need to distinguish between how well a model may predict for groups of patients, and how well for individual patients.
  - Although with an excellent model we may successfully distinguish between high and low risk patients, and can estimate group survival probabilities with precision,
    our ability to provide informative prognoses at the individual level is almost always limited.

Validating a prognostic model

- A vital aspect of prediction is to consider whether a model derived from an analysis of the original data set is transportable to similar patients in another location.
  - The concept is sometimes referred to as generalizability or validity, and a model which is found to pass such a test is said to have been validated.
    - what or should be meant by validating a model
    - why we need to validate models
    - how we should attempt to validate a model

Validating a prognostic model

- two types of validated model [not necessarily associated]:
  - A *statistically validated model* is one which passes all appropriate statistical checks, including goodness-of-fit on the original data set and unbiased prediction on a new data set.

  - A *clinically validated model* is one which performs satisfactorily on a new data set according to context-dependent statistical criteria laid down for it.

Deficiencies of standard modelling methods

- It is known that analyses which are not prespecified but are data-dependent are liable to lead to overoptimistic conclusions
- For clinical purposes it is usually necessary for a predictive model to be based on a small number of variables, and it is arguable that parsimony is a desirable feature of a good model.
- In most cases, however, there is a large number of ‘candidate' variables available for consideration, and thus there is a need to select the “important' ones.
  - The data-dependent aspect of most models stems from this variable selection.
- Of the various strategies for producing a prognostic model, by far the most common is multiple regression using a stepwise selection algorithm (backward or forward).
- The choice of variables included in the final model is based on multiple sequential hypothesis testing of individual variables, usually with $P \leq 0.05$ as the inclusion criterion

Deficiencies of standard modelling methods

- An alternative approach is all subsets regression, although it is rarely used.
- This method can discover combinations of variables that explain more variation in patient outcome than the model achieved by stepwise selection.
  - . A popular approach is to choose the model which optimizes a measure of goodness-of-fit penalized for including each extra variable, such as the Akaike information criterion
  - the approach has some major drawbacks, including the possibility of selecting models which omit important predictors
- Recent alternative methodologies include regression trees (CART) and neural networks. These methods also use the data to determine the model, although in rather different ways

Deficiencies in the design of prognostic studies

- The most reliable observational studies are those that attempt to emulate the careful design standards used in clinical trials, with the goal of achieving the same answer as if an experimental study had been performed
- There are various weaknesses of prognostic factor studies which could result in misleading findings, creating overoptimism and/or bias.
  - absence of clear inclusion and exclusion criteria,
  - many patients excluded through missing data (which may not be missing at random),
  - unclear rationale for the choice of treatments
  - inadequate sample size.

Deficiencies in the design of prognostic studies

- The definition of the characteristics of the sample is of clear importance to the clinician who wishes to know whether a model is relevant to a particular patient.
- The problems that arise from data-dependent selection are exacerbated by small sample size.
- With a small sample there will be a low signal-to-noise ratio, with an increased risk of selecting unimportant variables and failing to include important ones.
- Another consideration is the number of events (for example, deaths) per variable (EPV) considered for inclusion in the model.
  - EPV is suggested to be ≥10 (Harrel, Peduzzi) or, better, ≥ 20 (Feinstein)
  - Most published studies do not meet either criterion.

How to validate a model

- There are several main considerations in validating a model:
  - study design;
  - measuring the intrinsic prognostic information;
  - comparing predictions with observations;
  - quantifying the performance of a model;
  - prespecifying adequate performance.

Study design

- We will consider a hierarchy of increasingly stringent validation strategies:
  - internal procedures restricted to a single data set;
  - temporal evaluation on a second data set from the same centre(s);
  - external evaluation on data from one or more other centres, perhaps by different investigators.

Internal validation

- One common way of establishing how well a model might perform for further patients is data splitting or cross-validation.
  - The original sample is split into two parts before the modelling begins.
    The model is derived on the first portion of the data (often called the &training' set) and then its ability to predict outcome is evaluated on the second or &test' data set.
- A variation is to carry out the modelling procedure on each portion of the data and to evaluate each model on the other portion.
  - An issue is how to split the data set: authors rarely consider what proportion of patients should be in the test and training sets (or fail to justify any recommendation),
  - Random splitting must lead to data sets that are the same other than for chance variation and is thus a weak procedure
  - Estimates of predictive accuracy from data-splitting procedures, though unbiased, tend to be imprecise
- A tougher test is to split the data in a non-random way. For example, we might take groups of patients seen in different time periods.
- Rather different, and better, approaches are to use bootstrapping or “leave-one out' cross-validation.
- From these analyses shrinkage factors can be estimated and applied to the regression coefficients to counter over-optimism

Temporal validation

- evaluate the performance of a model on subsequent patients within the same centre(s).

- This approach is no different in principle from the idea already mentioned of splitting a single data set into two cohorts seen in different time periods.
- it is at least a prospective evaluation, independent of the original data and model-fitting process.

- A disadvantage of the approach when the outcome is survival time is the need to wait several years to accrue an adequate number of events in a further cohort.

External validation

- Neither internal nor temporal evaluation addresses the wider issue of the generalizability of the model.
- As the goal of validation is to demonstrate satisfactory performance for patients from a different population from the original, it is clearly desirable to evaluate a model on new data collected from an appropriate patient population in a different centre.
- Important design issues such as sample selection and sample size have been largely neglected in the literature.
- External evaluation can be based on retrospective data and so is viable for validating survival models needing long follow-up.

Measuring intrinsic prognostic information

- prognoses are to be framed as predicted probabilities of a particular event, implicitly or explicitly linked to a specific time-point
  - i.e.: chance of surviving for 5 years following initial treatment.

- The predicted probabilities are obtained as outputs from a prognostic model.

Measuring intrinsic prognostic information

- Intuitively, the idea of prognostic information relates to the spread of predicted probabilities.
- For example, in an analysis unadjusted for other factors, the estimated chance of surviving for 3 years following initial treatment
  - for node-positive breast cancer may be about 90% for patients with 1-3 affected lymph nodes compared with about 60% for those with 10 or more affected nodes.
  - the corresponding figures for pre- and post- menopausal patients may be about 84% and 82%, respectively.
  - The prognostic information contained in lymph node status is clearly much greater than that in menopausal status, since the spread of probabilities is 0.3 as against only 0.02.

Measuring intrinsic prognostic information

- The spread of probabilities depends on
  - how finely the prognostic factor or index is graded: the finer, the greater the spread.
  - the prevalence of the event. In a survival study, the spread usually increases with the length of follow-up.
  - by the amount of overoptimism in the statistical model used to estimate the probabilities.

An index intrinsic prognostic information: PSEP

- PSEP
  - P_{worst} predicted probability of dying for a patient in the group with the worst prognosis
  - P_{best} predicted probability of dying for a patient in the group with the best prognosis.
  - Then the predicted prognostic information can be measured by the separation
    PSEP= P_{worst} - P_{best}

  - With just two groups, P_{worst} and P_{best} are closely related to the familiar concepts of the positive (PPV) and negative (NPV) predictive value of a diagnostic test by the relations P_{worst} =PPV and P_{best} =1-NPV.
  - Thus PSEP=PPV+NPV-1.

- Given the overall prevalence of events, PPV and NPV and hence PSEP may be calculated from the sensitivity and specificity by standard formulae.

Quantifying the performance of a model

- Evaluation consists of comparing the appropriate observed and predicted measure, an aspect of model calibration
- The comparison between predicted and observed probabilities may be made in several ways, such as at the patient level by using the Brier score,
  - mean squared difference between observed patient outcomes in the validation sample and the corresponding probabilities predicted by the model
  - lacks an obvious interpretation other than in general terms: the bigger the score, the worse the quality of the prediction
- A more interpretable statistic is the difference between observed and predicted probabilities at the group level (PSEP)
- Validation cannot be determined by statistical criteria alone but must be considered in relation to the clinical aims

Prespecifying adequate performance

- Prognostic studies may fall into two categories: pragmatic and explanatory
  - Pragmatic studies are driven by explicit clinical aims.
    The idea is to prejudge the quality of predictions from a prognostic model that may or may not be acceptable.
  - This is the notion of a clear, quantitative aim, guided by statistical principles, and is reminiscent of predefining the desired size of a treatment effect or treatment difference in a clinical trial.
  - For example, if the aim was to identify patients with a three-year survival rate of 80%, a validation study showing that the chance was actually 60% would cause the model to be rejected for that purpose, even though strong prognostic information might be present.

Prespecifying adequate performance

- Explanatory studies are mainly concerned with scientifc understanding and hypothesis generation, to answer such questions as:
  - What factors are important to predict the course of disease X?
  - Can we discriminate reproducibly between good and bad prognosis for disease X?

- In a explanatory validation study we would want to examine general qualitative and quantitative aspects such as:
  - Are the same variables still important?
  - Is the functional form of the prognostic model correct?
  - Are the estimated regression coe$cients compatible?
  - How well does the model "t the new data?
  - Is the correct ordering of the prognostic groups preserved?
  - Are the event rates between the prognostic groups signi"cantly di!erent?

Prespecifying adequate performance

- It is possible also to compare summaries such as
  - PSEP, p_{worst} and p_{best} for the original groups with those in the validation study in order to get estimates of overoptimism and whether there was a need to calibrate the model to reduce prediction bias

Case study: Predicting relapse in asthma

- Fischl *et al*. described a prognostic index to predict the chance of a patient with acute bronchial asthma relapsing or requiring hospitalization following initial treatment in the emergency room.
  - Assessment of patients with acute asthma by this multi-factorial approach should allow the physician to identify rapidly those patients who are at risk for relapse and in need of hospitalisation.
    - patients with predictor index scores of 4 or higher (calculated before therapy) should be considered for prompt hospitalisation.
- Comparison with validation study

Prognostic classification of relapse or hospitalization in acute asthma. Values in table are proportions (and numbers) of patients who relapsed or were admitted to hospital.

Index	Fischl *et al.*		Centor *et al.*	
$\geqslant 4$	0.95	(81/85)	0.52	(11/21)
< 4	0.03	(4/120)	0.28	(18/65)
Total	0.41	(85/205)	0.34	(29/86)
PSEP	0.92		0.25	
(95 per cent CI)	(0.86 to 0.98)		(0.01 to 0.49)	

Case study: Predicting death from acute myocardial infarction

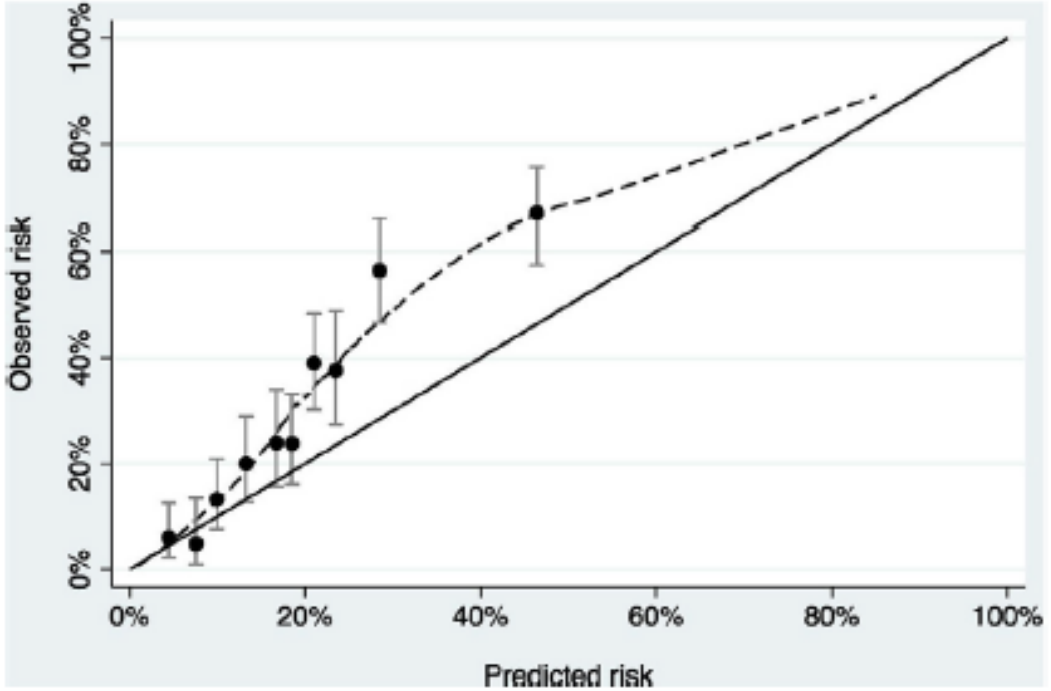
- Woo described a prognostic index to predict the chance of dying in hospital after acute myocardial infarction (MI) among a Chinese population.
  - the performance of the index in the original centre and in two additional cities was evaluated
  - The intention was to provide an objective guide for the assessment of patients with acute MI and stratify different grades of clinical severity: a high index would conceivably unmask the high-risk patients who deserve more attention and a more energetic regimen, while a low index would identify a low-risk group for earlier ambulisation and discharge from the coronary care unit and from the hospital
- Comparison with validation study

Validation of prognostic index in acute myocardial infarction. Values are proportions (and number of deaths/number of patients) in each subgroup.

Index	Original sample		Validation samples			
	Hong Kong 1971–1980		Hong Kong 1977–1986		Guangzhou 1977–1982 and Shanghai 1978–1986	
⩾ 8	0.724	(63/87)	0.613	(49/80)	0.792	(38/48)
6–7	0.539	(55/102)	0.392	(38/97)	0.397	(27/68)
4–5	0.214	(39/182)	0.159	(20/126)	0.203	(35/172)
⩽ 3	0.051	(14/273)	0.061	(8/132)	0.031	(5/163)
Total	0.266	(171/644)	0.264	(115/435)	0.233	(105/451)
PSEP (95 per cent CI)	0.673 (0.575 to 0.770)		0.552 (0.438 to 0.606)		0.761 (0.643 to 0.879)	

How do we know if a prediction model is a good one? Problems with traditional measures

- Discrimination:
  - Roc curve analysis
- Calibration:
  - Calibration plot
    - The data are first split up into groups, typically 10, in terms of predicted risk. The observed risk in each group is then calculated along with a 95% confidence interval.



How do we know if a prediction model is a good one?

- Discrimination and calibration provide valuable information for researchers.
  - Poor discrimination, demonstrates the predictors in a model are not strongly associated with outcome, such that researchers should consider additional predictors.
  - Poor calibration of a model suggests genuine differences between the data set used to generate a model and the evaluation data set

  - Difficulties in interpretation:
    - AUC of 0.77 is high enough to warrant using the model or would the AUC need to be closer to 0.85?
    - how much miscalibration would be “too much” and suggest that a model should not be used?

  - Calibration and discrimination may not be associated

Decision curve analysis

- Method to determine whether use of a prediction model in the clinic to inform clinical decision-making would do more good than harm

DCA: working example

- Suppose there are two prediction models, *model A* and *model B*, that have been developed by two different teams to predict a patient's risk of pathologic fractures within the next 6 months using variables such as SINS and history of osteoporosis as predictors.
- A study is conducted to evaluate both models on an independent cohort of 1,000 patients who were eligible for surgery on the basis of SINS or imaging but in fact never underwent an operation.
- The results show that calibration is superior for model A but discrimination is better for model B.
- It is hard to tell whether the miscalibration shown for model B offsets the advantages of superior AUC.

DCA: working example

- in order to know whether the benefits of a model outweigh the harms, we have to put some numbers on benefit and harm.

- To do so, we need to think about the *threshold probability of disease*, defined as the minimum probability of disease (in this case, future pathologic fracture) at which a decision-maker – doctor or patient – would opt for an intervention (in this case, surgery).

    - Consider that, if a patient were told that the probability of pathologic fractures was 1%, the discomfort and risks of surgery would certainly outweigh any benefit from reduced risk of fracture
    - Conversely, if the patient were told the risk of fractures was 99%, they would certainly choose to have surgery.
    - If we were to gradually increase the probability of pathologic fractures from 1% to 99%, there would come a point where the patient would be unsure whether or not to have surgery.

    - We call this point p_t, the threshold probability and it is directly linked to how the consequences of the decision are weighted.

DCA: working example

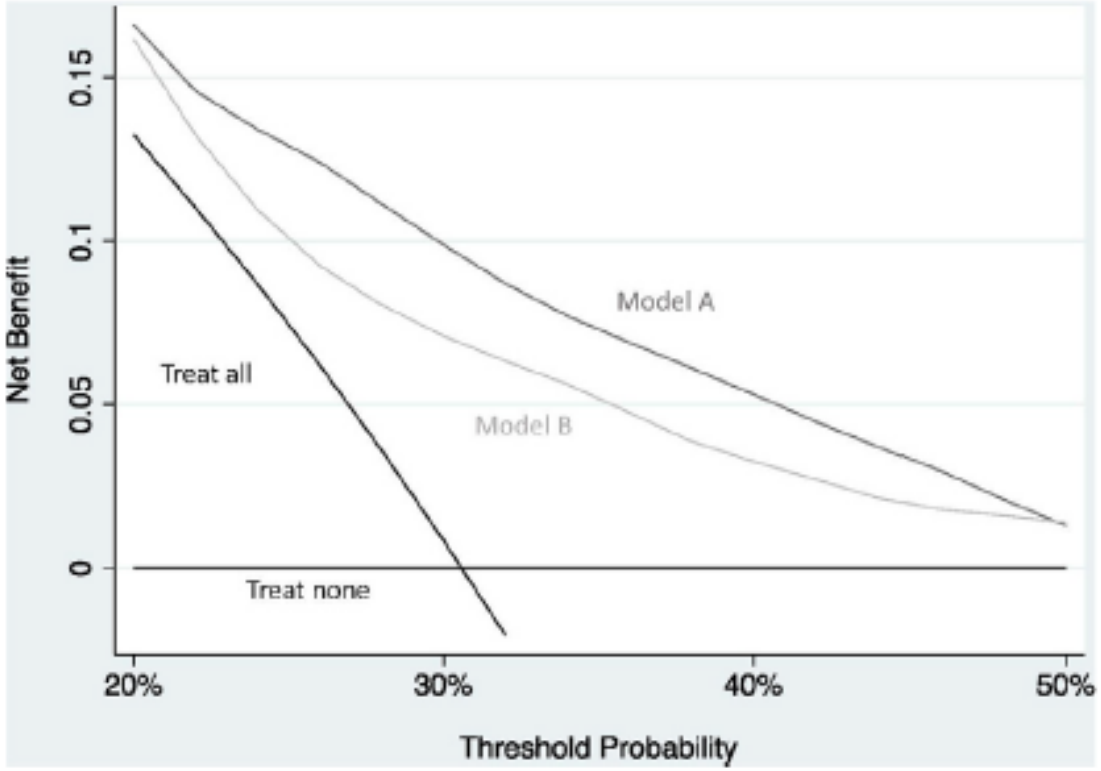
$$Net\ Benefit = \frac{True\ positives - False\ positives \times \frac{p_t}{1-p_t}}{N}$$

Net benefit of the two models for predicting facture

Strategy	True positives: patients recommended for surgery who would otherwise get a fracture	False positives: patients recommended for surgery who will not get a fracture	Net benefit
Recommend surgery for all patients	306	694	$(306 - 694 \times (0.25 \div 0.75)) \div 1000 = 0.0747$
Recommend surgery if risk ≥25% according to Model A	240	335	$(240 - 335 \times (0.25 \div 0.75)) \div 1000 = 0.128$
Recommend surgery if risk ≥25% according to Model B	136	84	$(136 - 84 \times (0.25 \div 0.75)) \div 1000 = 0.108$
Surgery for no patients	0	0	$(0 - 0 \times (0.25 \div 0.75)) \div 1000 = 0$

Net benefit is given at a threshold probability of 25% along with that for the clinical alternatives of recommending surgery for all or no patients.

DCA: working example



Decision curve analysis for two hypothetical models predicting pathologic spinal fracture in patients with metastatic disease. Model B has better discrimination (0.715 vs. 0.758) but is miscalibrated The decision curve shows that the miscalibration offsets improved discrimination: model A has a higher net benefit compared to model B, as well in comparison to the clinical default strategies of "treat all" or "treat none", over the entire range of reasonable threshold probabilities. Using model A to decide which patients should receive surgery would therefore lead to the best clinical outcomes.

DCA: working example

A Two-Step Frailty Assessment Strategy in Older Patients With Solid Tumors: A Decision Curve Analysis

Adolfo González Serrano, MD, MSc[1]; Marie Laurent, MD, PhD[1,2]; Thomas Bamay, PhD[3]; Claudia Martínez-Tapia, PhD[1]; Etienne Audureau, MD, PhD[1,4]; Pascaline Boudou-Rouquette, MD[5]; Thomas Aparicio, MD, PhD[6]; Florence Rollot-Trad, MD[7]; Pierre Soubeyran, MD, PhD[8]; Carine Bellera, PhD[9,10]; Philippe Caillet, MD[1,11,12]; Elena Paillaud, MD, PhD[1,11,12]; and Florence Canouï-Poitrine, MD, PhD[1,4]

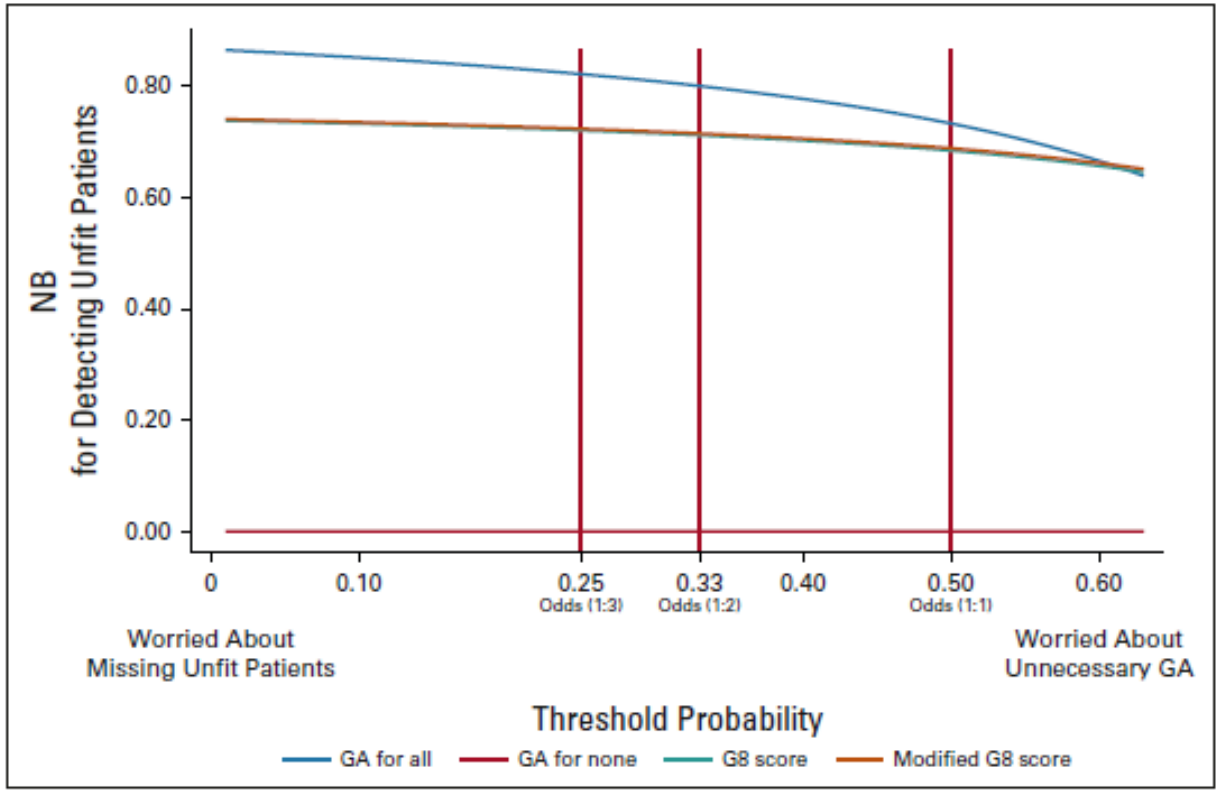
PURPOSE The intended clinical value of frailty screening is to identify unfit patients needing geriatric assessment (GA) and to prevent unnecessary GA in fit patients. These hypotheses rely on the sensitivity and specificity of screening tests, but they have not been verified.

METHODS We performed a cross-sectional analysis of outpatients age ≥ 70 years with prostate, breast, colorectal, or lung cancer included in the ELCAPA cohort study (ClinicalTrials.gov identifier: NCT02884375) between February 2007 and December 2019. The diagnostic accuracy of the G8 Geriatric Screening Tool (G8) and modified G8 scores for identifying unfit patients was determined on the basis of GA results. We used decision curve analysis to calculate the benefit of frailty screening for detecting unfit patients and avoiding unnecessary GA in fit patients across different threshold probabilities.

RESULTS We included 1,648 patients (median age, 81 years), and 1,428 (87%) were unfit. The sensitivity and specificity were, respectively, 85% (95% CI, 84 to 87) and 59% (95% CI, 57 to 61) for G8, and 86% (95% CI, 84 to 87) and 60% (95% CI, 58 to 63) for the modified G8 score. For decision curve analysis, the net benefit (NB) for identifying unfit patients were 0.72 for G8, 0.72 for the modified G8, and 0.82 for GA at a threshold probability of 0.25. At a threshold probability of 0.33, the NBs were 0.71, 0.72, and 0.80, respectively. At a threshold probability of 0.5, the NBs were 0.68, 0.69, and 0.73, respectively. No screening tool reduced unnecessary GA in fit patients at predefined threshold probabilities.

CONCLUSION Although frailty screening tests showed good diagnostic accuracy, screening showed no clinical benefits over the GA-for-all strategy. NB approaches, in addition to diagnostic accuracy, are necessary to assess the clinical value of tests.

J Clin Oncol 41:826-834. © 2022 by American Society of Clinical Oncology



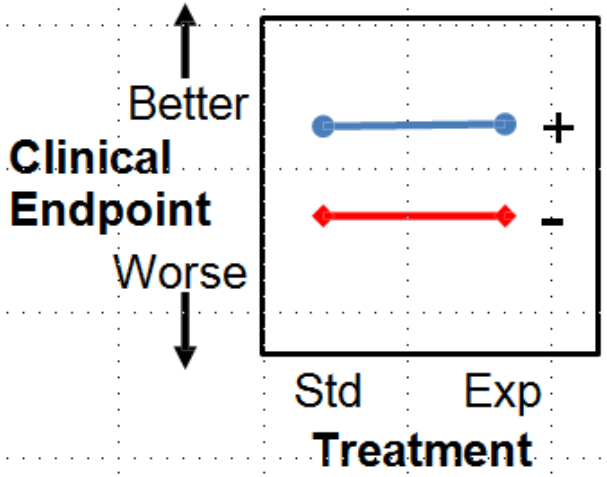
Decision curve for the NB of three different strategies (relative to not doing GA for any patient), with different threshold probabilities. The odds of 1:3 and 1:2 mean that missing an unfit patient is, respectively, three and two times as bad as exposing a fit patient to an unnecessary GA. The odds of 1:1 mean missing an unfit patient is the same as exposing a fit patient to an unnecessary GA. The NB represents the number of patients (per 100) who will be found to be unfit using a given frailty screening tool. G8, G8 Geriatric Screening Tool; GA, geriatric assessment; NB, net benefit.

Predictive models

Prognostic vs. Predictive biomarker

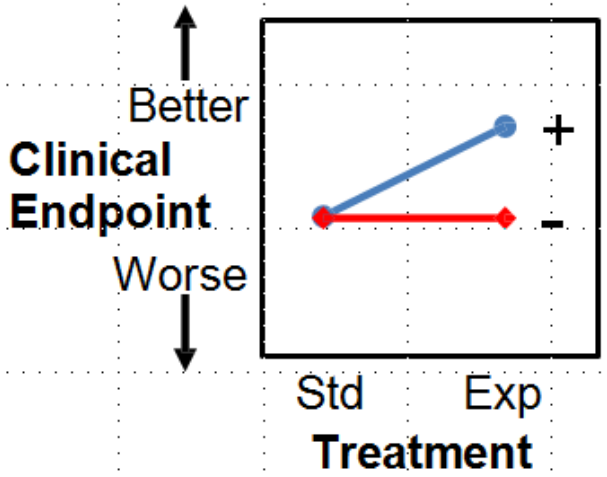
- prognostic biomarkers
  - used to identify likelihood of a clinical event, disease recurrence or progression in patients who have the disease or medical condition of interest,
- predictive biomarkers
  - used to identify individuals who are more likely than similar individuals without the biomarker to experience a favorable or unfavorable effect from exposure to a medical product or an environmental agent.
- Prognostic biomarkers and predictive biomarkers cannot generally be distinguished when only patients who have received a particular therapy are studied.
- Prognostic biomarkers are often identified from observational data and are regularly used to identify patients more likely to have a particular outcome.
- The best design for prognostic marker identification is a prospective cohort study

Biomarkers – classifications and uses



Prognostic:

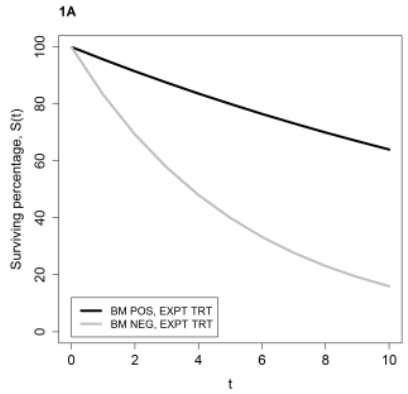
- associated with **disease outcome**
- risk assess (+,-) to stratify for treatment

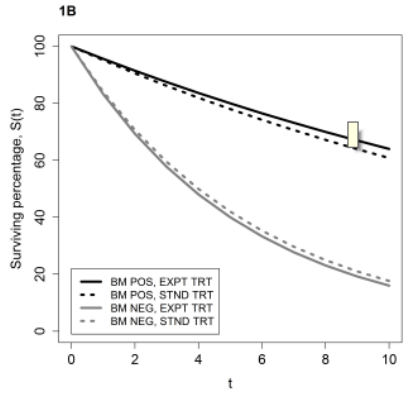


Predictive:

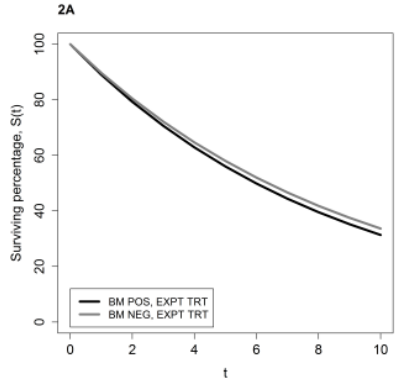
- associated with **treatment response**
- M+ benefit from experimental tmt
- individualise therapy
- personalised medicine

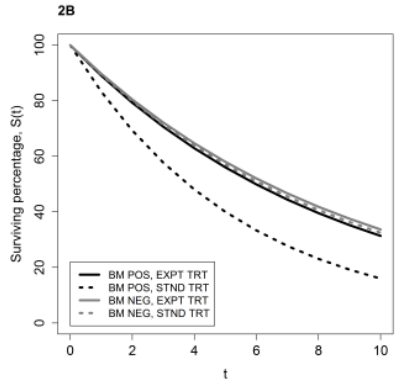
Prognostic vs. Predictive biomarker





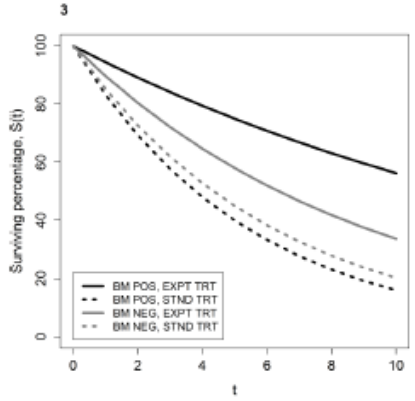
Example of a biomarker that is prognostic but not predictive. Assume that patients have been randomized to the experimental and standard therapies. A) For patients receiving the experimental therapy, those who are positive for the biomarker (black curve) survive longer than those who are negative for the biomarker (gray curve). B) The biomarker is associated with the same difference in survival for those patients receiving the standard therapy (black dashed curve versus gray dashed curve); therefore, it is prognostic. The biomarker is not predictive for benefit of the experimental therapy (solid curves) relative to the standard therapy (dashed curves) because within each biomarker subgroup the survival distribution is the same regardless of treatment received.



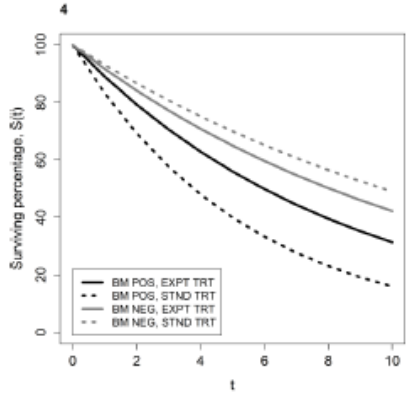


Example of a biomarker that is both prognostic (negatively) and predictive. Assume that patients have been randomized to the experimental and standard therapies. A) For patients receiving the experimental therapy, those who are positive for the biomarker (black curve) have survival similar to those who are negative for the biomarker (gray curve). B) For patients receiving the standard therapy, those who are positive for the biomarker have shorter survival (black dashed curve) compared to those who are negative for the biomarker (gray dashed curve); therefore, the biomarker is negatively prognostic. The biomarker is also predictive for benefit of the experimental therapy (solid curves) relative to the standard therapy (dashed curves) because survival for patients who are positive for the biomarker is substantially longer for those receiving the experimental therapy (black solid curve) compared to standard therapy (black dashed curve); whereas, for patients who are negative for the biomarker (gray curves) the survival distribution is the same regardless of treatment received.

Prognostic vs. Predictive biomarker



Example of a predictive biomarker that exhibits a quantitative treatment-by-biomarker statistical interaction. Assume that patients have been randomized to the experimental and standard therapies. Within each biomarker subgroup (black curves for biomarker positive and gray curves for biomarker negative), survival is substantially longer for patients who receive the experimental therapy (solid curves) compared to standard therapy (dashed curves). The magnitude of the increase in survival for those receiving experimental therapy compared to standard therapy is numerically larger for those who are positive for the biomarker than for those who are negative for the biomarker.



Example of a predictive biomarker that exhibits a qualitative treatment-by-biomarker statistical interaction. Assume that patients have been randomized to the experimental and standard therapies. For patients who are positive for the biomarker (black curves), survival is substantially longer for patients who receive the experimental therapy (solid black curve) compared to standard therapy (dashed black curve); whereas, for patients who are negative for the biomarker (gray curves), survival is about the same or slightly shorter for patients who receive the experimental therapy (solid gray curve) compared to standard therapy (dashed gray curve).

Predictive biomarker

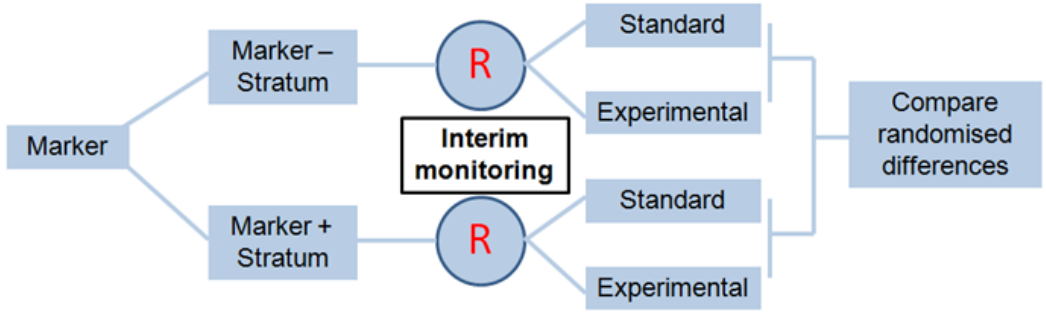
- To identify a predictive biomarker, there generally should be a comparison of a treatment to a control in patients with and without the biomarker.

- However, there are circumstances in which preclinical and early clinical data provide such compelling evidence that a new treatment will not work in patients without the biomarker that definitive clinical trials are performed only in populations enriched for the putative predictive biomarker.

Biomarkers – designs

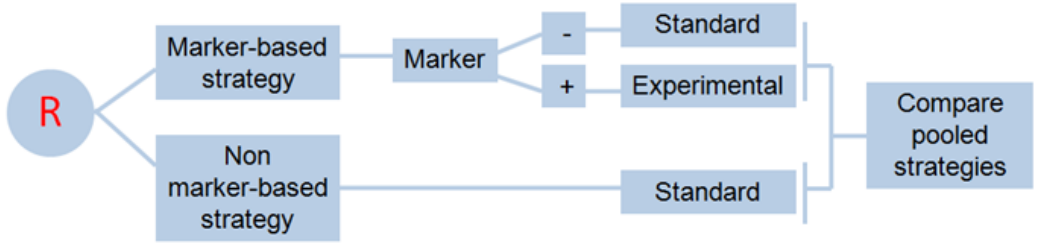
Biomarker-Stratified Design (Full specification)

Recommended when **preliminary evidence of effect** is less robust

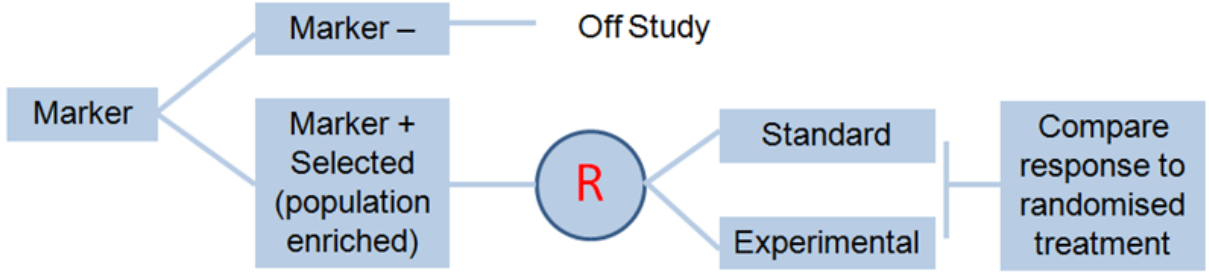


Biomarker-Strategy Design ("Use" vs "Ignore" biomarker)

Less feasible with low **M+ prevalence**



Requires evidence of **lack of benefit** of experimental treatment in M-



Metrics performance of predictive biomarker

OXFORD

JNCI J Natl Cancer Inst (2015) 107(8): djv153

doi:10.1093/jnci/djv153
First published online June 24, 2015
Brief Communication

BRIEF COMMUNICATION

Sensitivity, Specificity, PPV, and NPV for Predictive Biomarkers

Richard Simon

$$ppv_i = \frac{\Delta_+}{1+\Delta_+} \quad (1)$$

where Δ_+ is the hazard ratio of C vs T (>1) for biomarker-positive patients. The negative predictive value (NPV) is the probability that a biomarker-negative patient will not have longer survival on T rather than C.

$$npv_i = \frac{1}{} \quad (2)$$

PPV_i can be interpreted as the probability that a marker-positive individual receiving treatment T will have longer survival than that for a randomly chosen individual with the same covariates and marker value who receives C.

For example, Amado et al. reported hazard ratios for progression-free survival of best supportive care vs panitumumab in second- or later-line therapy of patients with metastatic colorectal cancer. For patients with wild-type KRAS, the hazard ratio was 2.22, favoring panitumumab with a 95% confidence interval (CI) of 1.69 to 2.94. The hazard ratio for the patients with mutated KRAS was 1.01 with a 95% CI of 0.73 to 1.37. For this data, calling wild-type KRAS marker positive, the PPV_i and NPV_i as calculated from (1) and (2) are 0.69 and 0.50, respectively. We note that expression (2) indicates that an NPV of 0.5 results when there is no treatment difference in the marker-negative stratum.

Metrics performance of predictive biomarker

OXFORD

JNCI J Natl Cancer Inst (2015) 107(8): djv153

doi:10.1093/jnci/djv153
First published online June 24, 2015
Brief Communication

BRIEF COMMUNICATION

Sensitivity, Specificity, PPV, and NPV for Predictive Biomarkers

Richard Simon

The sensitivity is the probability that the biomarker is positive for patients who benefit from T relative to C. Specificity is the probability that the biomarker is negative for patients who do not benefit from T relative to C. The usual relationships between sensitivity, specificity, PPV, NPV, and prevalence can be written

$$sensitivity_i = 1 / \left\{ 1 + \frac{1 - npv_i}{ppv_i} \frac{\Pr[B-]}{\Pr[B+]} \right\} \quad (3)$$

$$specificity_i = 1 / \left\{ 1 + \frac{1 - ppv_i}{npv_i} \frac{\Pr[B+]}{\Pr[B-]} \right\} \quad (4)$$

where Pr[B+] denotes the prevalence of biomarker-positive patients and Pr[B-]=1-Pr[B+].

For the Amado et al. data, the prevalence of wild-type KRAS was 0.62 and so the sensitivity and specificity as calculated from (3) and (4) are 0.62 and 0.50, respectively. Predicting treatment outcome is generally more difficult than distinguishing diagnostic categories; consequently, the very high performance indices commonly observed for diagnostic markers should not be expected for predictive biomarkers.

Rete Oncologica Veneta
Ricerca, innovazione, assistenza

L' impatto degli studi osservazionali nella pratica clinica

Pierfranco Conte
Coordinatore, Advisory Board CRAO
Direttore Scientifico, S Camillo Hospital, IRCCS

Rete Oncologica Veneta
Ricerca, innovazione, assistenza

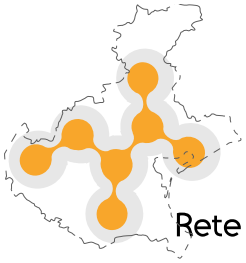
SAN CAMILLO IRCCS

Rete Oncologica Veneta
Ricerca, innovazione, assistenza

PierFranco Conte

Disclosure of potential conflicts of interests

- Consultant:
  Novartis, EliLilly, Astra Zeneca, Tesaro, Daiichi-Sankyo, Gilead, Reveal Genomics
- Honoraria:
  BMS,Roche, EliLilly, Novartis, AstraZeneca
- Research Funding from profit organizations:
  Novartis, Roche, EliLilly, BMS, Merck-KGa
- Funding from non profit organizations:
  National Research Council, Ministry of Education and Research, Italian Association for Cancer Research, Italian Drug Agency (AIFA), EmiliaRomagna Secretary of Health, Veneto Secretary of Health, University of Padova, Ministry of Health
- Founder & Chairman:
  Periplo Foundation



Rete Oncologica Veneta

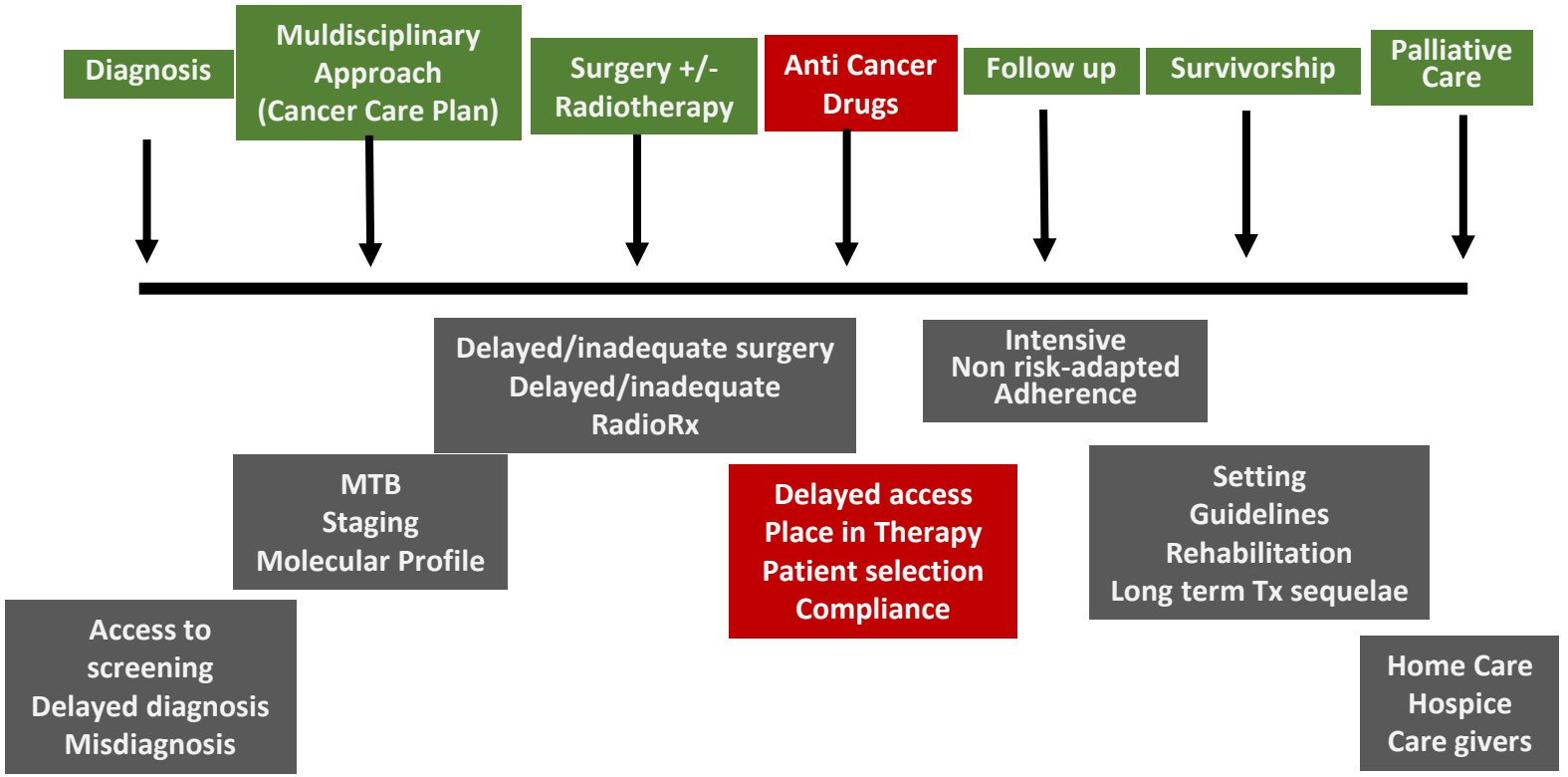
Ricerca, innovazione, assistenza

Oncology at the Cross Roads: the Role of Oncology Networks

- Diagnostic-therapeutic Pathways: markers and outcomes
- Innovation is sustainable; way of funding the NHS is NOT
- Evidence- based medicine: from efficacy to effectiveness

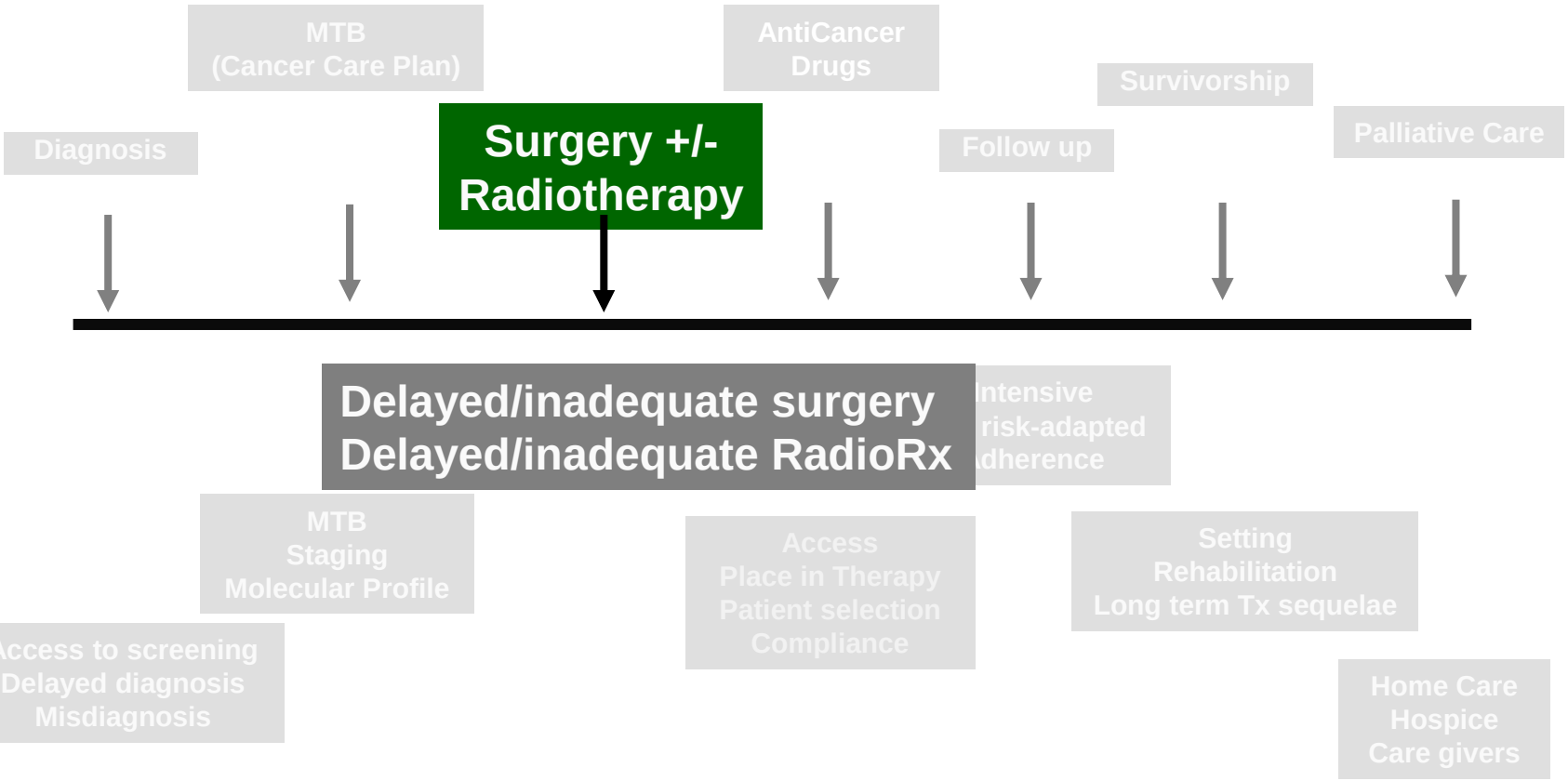
Rete Oncologica Veneta
Ricerca, innovazione, assistenza

Patients' Journey in Oncology



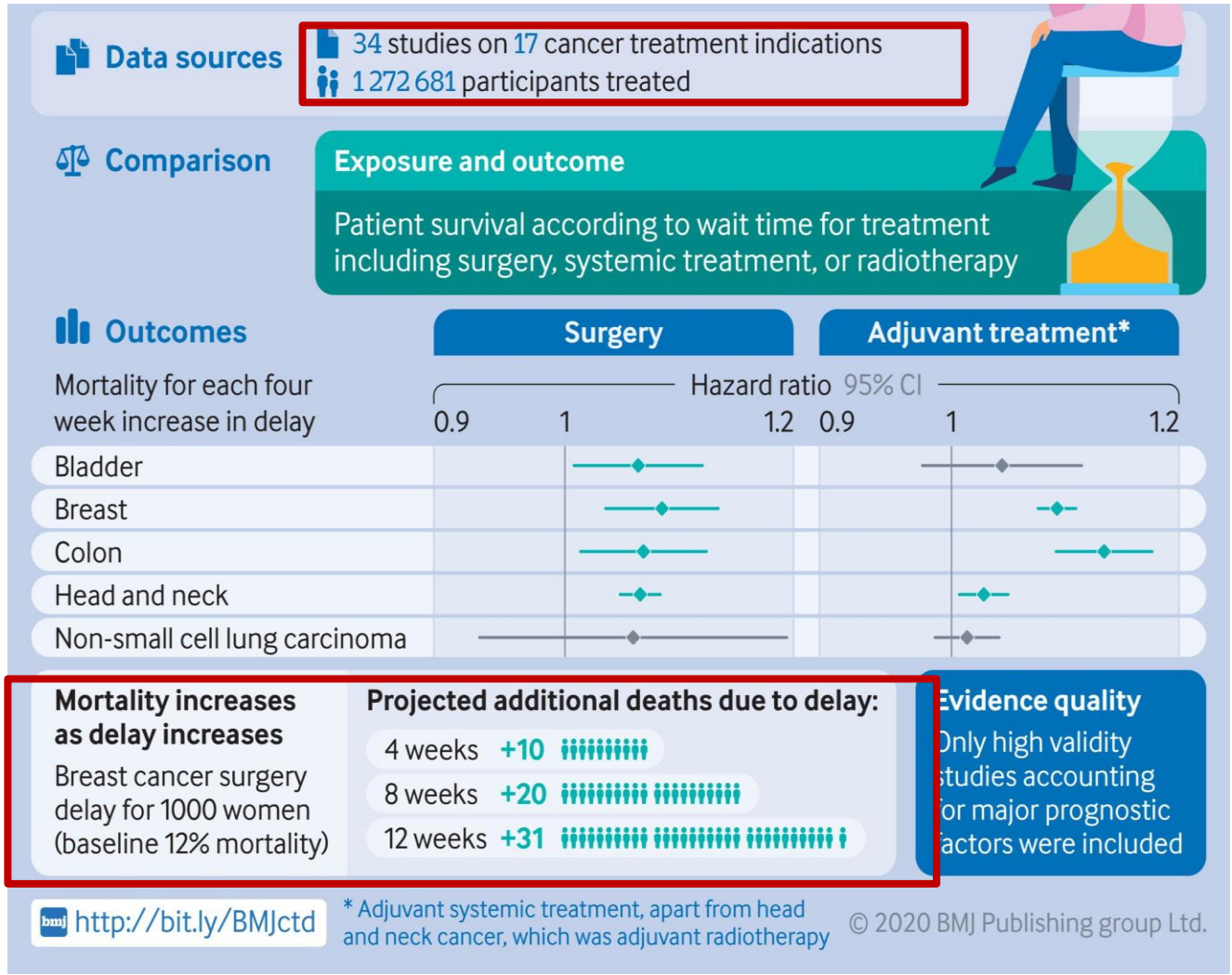
Rete Oncologica Veneta
Ricerca, innovazione, assistenza

Patients' Journey in Oncology



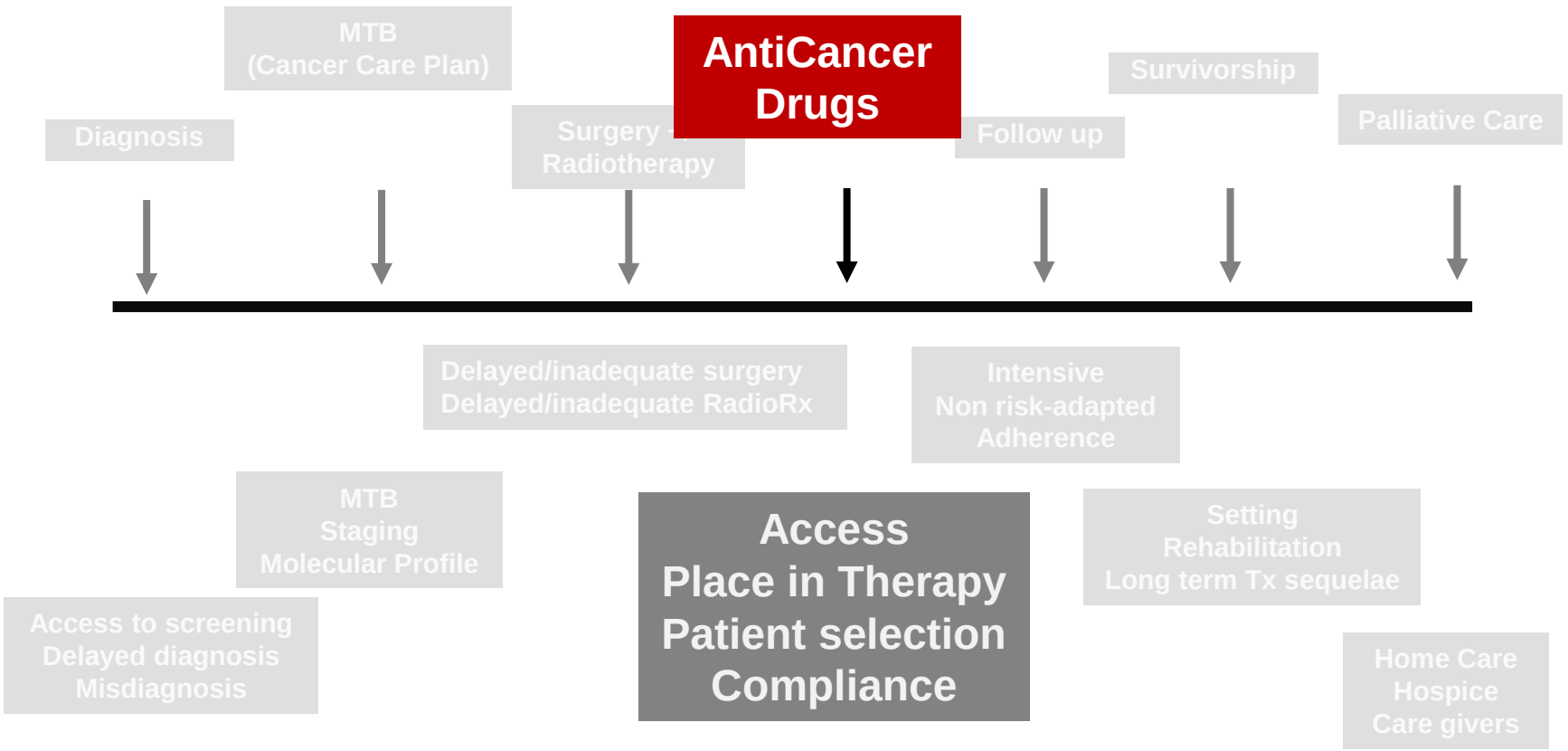
Rete Oncologica Veneta
Ricerca, innovazione, assistenza

Treatment delay and outcome



Rete Oncologica Veneta
Ricerca, innovazione, assistenza

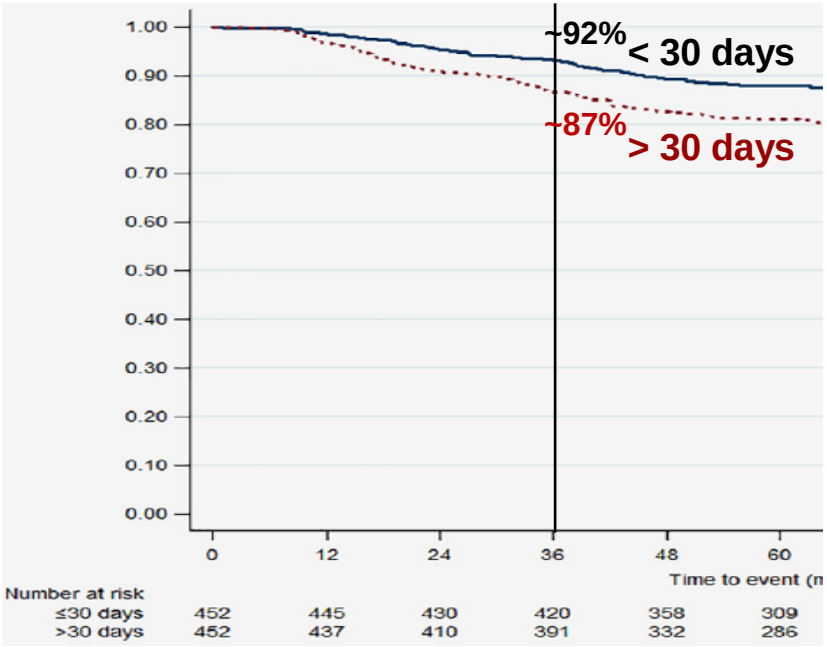
Patients' Journey in Oncology



Rete Oncologica Veneta
Ricerca, innovazione, assistenza

Time to adjuvant chemotherapy for eBC and outcome

eTNBC:
3 yrs OS according to time to chemotherapy



Heeg E et al, Int J Cancer 2020

FOCUS ON QUALITY ReCAP

Use of Electronic Administrative Databases to Measure Quality Indicators of Breast Cancer Care: Experience of Five Regional Oncology Networks in Italy

Valentina Guarneri, PhD, MD[1,2]; Paolo Pronzato, MD[3,4]; Oscar Bertetto, MD[5]; Fausto Roila, MD[6]; Gianni Amunni, MD[7,8]; Alberto Bortolami, PharmD[2,9]; Sandro Tognazzo, MS[2,9]; Gaia Griguolo, MD[1,2]; Eva Pagano, MEcon[10]; Fabrizio Stracci, PhD, MD[11]; Fortunato Bianconi, PhD, MS[12]; Fabrizio Gemmi, MD[13]; Letizia Bachini, MS[13]; Giovannino Ciccone, MS[10]; Gabriella Paoli, MEng[14]; Laura Paleari, PhD, MS[14]; and Pier Franco Conte, MD[1,2] on behalf of the Periplo Association

Adjuvant therapy within 8 weeks from surgery % of patients					
Veneto	Liguria	Toscana	Piemonte	Umbria	Benchmark
73.7 %	66.7 %	NA	71.8 %	69.9 %	≥ 80%

No data available on breast cancer subtypes

The VERO Study:
The VEneto and ROmagna Breast project

IRST Meldola:
Ilaria Massa, Roberta Maltoni, Valentina Danesi, William Balzi, Andrea Roncadori
AUSL Romagna:
Mattia Altini, Roberto Grilli
Rete Oncologica Veneta:
Pierfranco Conte, Alberto Bortolami, Sandro Tognazzo, Valentina Guarneri
Azienda Zero:
Manuel Zorzi

Rete Oncologica Veneta
Ricerca, innovazione, assistenza

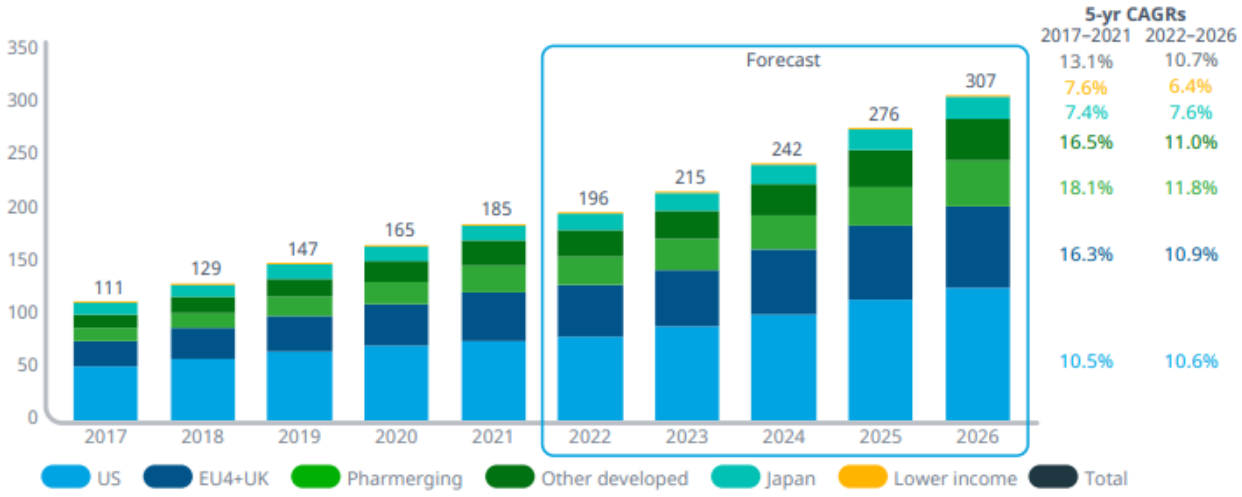
Oncology at the Cross Roads: the Role of Oncology Networks

- Diagnostic-therapeutic Pathways: markers and outcomes
- Innovation is sustainable; way of funding the NHS is NOT
- Evidence- based medicine: from efficacy to effectiveness

Rete Oncologica Veneta
Ricerca, innovazione, assistenza

Spesa per farmaci oncologici globale

Exhibit 42: Oncology spending by region, US$Bn



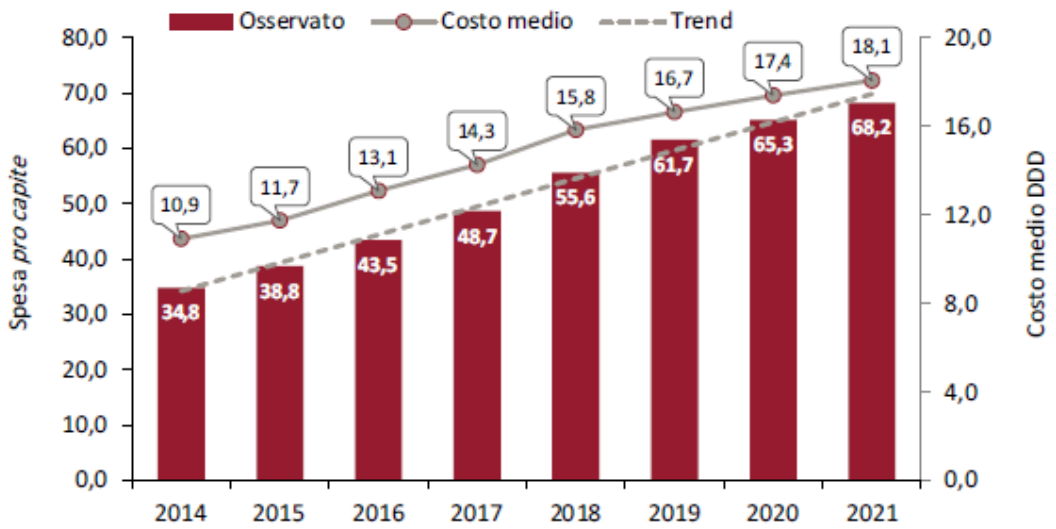
Source: IQVIA Oncology Link, Apr 2022.

La spesa per farmaci oncologici è salita a $ 185 miliardi a livello globale nel 2021, con il 74% concentrato nei principali mercati (Stati Uniti, EU4+UK e Giappone).

Rete Oncologica Veneta
Ricerca, innovazione, assistenza

Spesa per farmaci oncologici in Italia

Figura 3.1.1a Farmaci oncologici, andamento temporale 2014-2021 della spesa *pro capite* e del costo medio per giornata di terapia



L'uso dei Farmaci in Italia
Rapporto Nazionale
Anno 2021

Rete Oncologica Veneta
Ricerca, innovazione, assistenza

PDTA per i pazienti affetti da tumore del polmone

Thoracic Cancer
Open Access
Thoracic Cancer ISSN 1759-7706

ORIGINAL ARTICLE

Estimated direct costs of non-small cell lung cancer by stage at diagnosis and disease management phase: A whole-disease model

Alessandra Buja[1], Michele Rivera[1], Anna De Polo[1], Eugenio di Brino[6], Marco Marchetti[6], Manuela Scioni[2], Giulia Pasello[4], Alberto Bortolami[7], Vincenzo Rebba[3], Marco Schiavon[1], Fiorella Calabrese[1], Giovanni Mandoliti[5], Vincenzo Baldo[1] & PierFranco Conte[4,8]

Table 3 Estimates of average (and confidence interval) per-patient costs of care for NSCLC by disease stage (€) during the first year after diagnosis

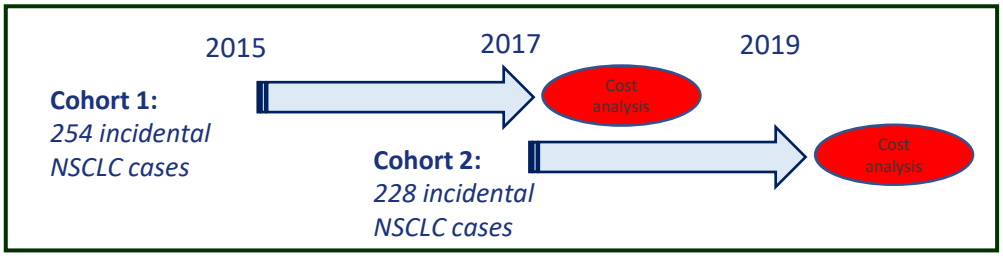
	Average total costs	Cost ratio vs. stage I
Stage I	16 291 (95% CI: 15 284–17 505)	1
Stage II	19 530 (95% CI: 18 263–21 091)	1.19
Stage III	21 938 (95% CI: 20 271–25 252)	1.34
Stage IV	22 175 (95% CI: 22 127–22 190)	1.36
Pancoast	28 711 (95% CI: 27 711–29 890)	1.79
TOTAL	21 328 (95% CI: -20 897–22 322)	

Rete Oncologica Veneta
Ricerca, innovazione, assistenza

VALUE IN CANCER CARE ReCAP

Non–Small-Cell Lung Cancer: Real-World Cost Consequence Analysis

Alessandra Buja, MD, PhD[1]; Giulia Pasello, MD[2]; Giuseppe De Luca, MD[1]; Alberto Bortolami, PharmD[3]; Manuel Zorzi, MD[4]; Federico Rea, MD[1]; Carlo Pinato, MStat[3]; Antonella Dal Cin, BS[4]; Anna De Polo, MD[1]; Marco Schiavon, MD[1]; Andrea Zuin, MD[1]; Marco Marchetti, MD[4]; Giovanna Scroccaro, PharmD[5]; Vincenzo Baldo, MD[1]; Massimo Rugge, MD[4]; Valentina Guarneri, MD, PhD[2,6]; and PierFranco Conte, MD[2,6]; on behalf of Rete Oncologica Veneta



Regression models dependent variable	Coefficient 2017 (ref 2015)	95% CI	p-value
Hospitalization costs	343.9	383.7 ; 0.9	0.37
Outpatient visits costs	192.0	314.1 ; 0.6	0.541
Emergency room costs	39.8	27.6 ; 1.4	0.149
Hospice costs	-911.3	397.0 ; -2.3	0.022
Hospital delivered drugs costs	2976	1116.0 ; 2.7	0.008
Medical devices costs	522.6	371.9 ; 1.4	0.160
Other Drugs costs	-55.1	44.84; -1.2	0.219
Total costs	3006	1148.0 ; 2.6	0.009

- Total costs adjusted for age, stage at diagnosis, sex, cohort, at 2 yrs after cancer diagnosis
- significant **increase in the average costs** of patients in the 2017 cohort
- significant **decrease** in the average cost of **hospice care**
- significant **increase** in the average cost of **drugs**
- **Significant increase in OS for stage III & IV**

The proportion of patients treated with targeted agents or ICPi increased by 523% for stage III and by 250% for stage IV disease.

Buja A et al, JOP 2021

Rete Oncologica Veneta
Ricerca, innovazione, assistenza

Registro Tumori Veneto

Rete Oncologica Veneta
Ricerca, innovazione, assistenza

Università degli Studi di Padova

O-CO-CA Project

(evaluation of impact of the introduction of the clinical pathway on health outcome, costs and quality of care in NSCLC patients)
PI V Guarneri

Studio osservazionale di coorte che prevede la raccolta di tutti i casi di NSCLC diagnosticati nella Regione Veneto in tre anni differenti (2015,2017,2019) per valutare:

1) Qualità delle cure
2) Costi della presa in carico globale
3) Esiti in termini di sopravvivenza globale

Al 3/2/2023 registrati e inclusi 960 pazienti

Rete Oncologica Veneta
Ricerca, innovazione, assistenza

Materiali e metodi

Identificazione dei casi

Casi di tumore al polmone NSCLC

incidenti nel 2017

RTV Registro Tumori Veneto

Fonti dei dati per il calcolo degli indicatori

Flussi amministrativi
SDO, SPS, Registro mortalità, farmaceutica, Hospice, Assistenza protesica, PS, Device, Assistenza domiciliare

Flusso di anatomia patologica
referti istologici, referti citologici

RTV - *Registro Polmone*
archivi di registro, collegamenti SIL delle ASL, cartelle cliniche

Rete Oncologica Veneta
Ricerca, innovazione, assistenza

Oncology at the Cross Roads: the Role of Oncology Networks

- Diagnostic-therapeutic Pathways: markers and outcomes
- Innovation is sustainable; way of funding the NHS is NOT
- Evidence- based medicine: from efficacy to effectiveness

Rete Oncologica Veneta
Ricerca, innovazione, assistenza

163/544 trials with OS improvement
108,344 patients included in these trials
14.2 millions years of life gained

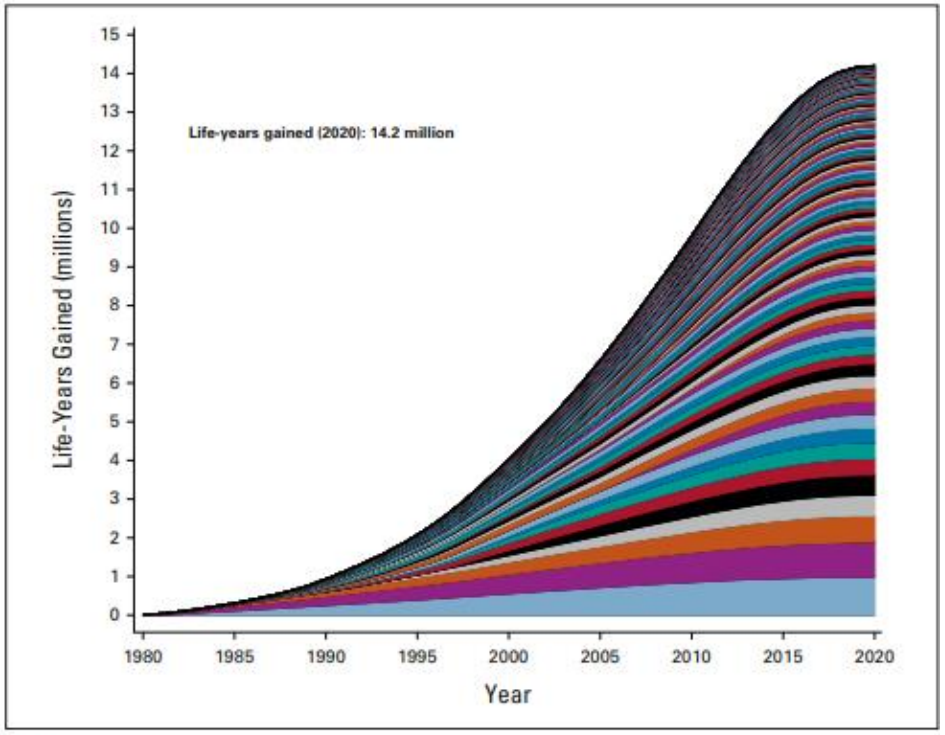
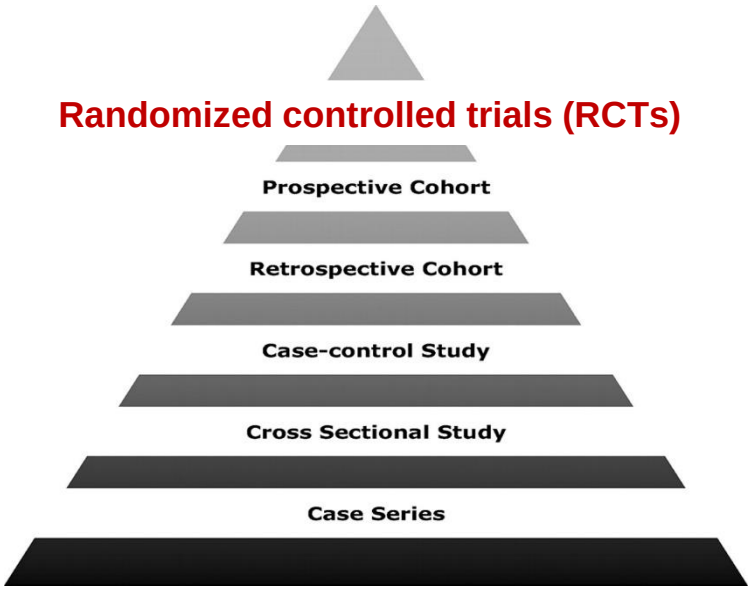


FIG 1. Cumulative life-years gained through 2020 by study. Each color-coded area represents cumulative life-years for 1 of 133 studies for which life-year gains were estimated.

Unger JM et al, JCO 2023

EBM



RCTs are the backbone of an application for marketing authorization.
However they operate in an idealised experimental environment (estimate of efficacy rather than a true measure of effectiveness):

- may lack external validity \\\\\\\\\\\\\\\\\\\\
- include selected patients:
  - not entirely representative of real-life population: elderly, poor PS patients, those with comorbidities are under-represented or excluded
  - differences in ethnic/racial composition
  - lacking data on budget impact

Rete Oncologica Veneta
Ricerca, innovazione, assistenza

The Oncologist®
Commentary

Real-World Evidence in Oncology: Opportunities and Limitations

Massimo Di Maio.[a] Francesco Perrone.[b] Pierfranco Conte[c]

Trial patient **Real life patient**

Large proportion of new treatments only show a globally modest efficacy within RCTs

↓

Effect in clinical practice might be further diluted

↓

Real value of results may fall under an acceptable threshold of relevance

↓

Post marketing studies could be useful to **confirm or refute the drug's benefit** on survival in real-world populations

↓

RWE analysis may challenge the magnitude of the efficacy previously shown in RCTs

Di Maio, Perrone, Conte; The Oncologist *2020*

Rete Oncologica Veneta
Ricerca, innovazione, assistenza

Evidence-based Medicine vs Real World Evidence

We believe that these findings raise the idea that ***overall survival in registration trials should be considered a surrogate for overall survival in the real world****, along with other surrogates, such as response rate and progression-free survival.*

Mailankody S & Prasad V, JAMA Oncology 2017; 3(7): 889-890

Rete Oncologica Veneta
Ricerca, innovazione, assistenza

The NEW ENGLAND JOURNAL of MEDICINE

SOUNDING BOARD

The Magic of Randomization versus the Myth of Real-World Evidence

Rory Collins, F.R.S., Louise Bowman, M.D., F.R.C.P., Martin Landray, Ph.D., F.R.C.P., and Richard Peto, F.R.S.

Observational studies cannot be trusted when the effect of treatment is moderate (i.e. less than a two-fold difference in the incidence of the health outcome).
Replacement of randomized trials with non randomized observational studies is a false solution to a serious problem.

Examples quoted:
«false effect» of statins and aspirin in the reduction of cancer incidence

Rete Oncologica Veneta
Ricerca, innovazione, assistenza

Precision Cancer Medicine - Change of Paradigm ?

New Paradigm	
PRESENT	**FUTURE**
Histology	Biomarker
Population – Biomarker	Drug
Drug	Indications
Indications	Regardless cancer site
PARADIGM CHANGE: WHEN A BIOMARKER DEFINES THE INDICATIONS	

Rete Oncologica Veneta
Ricerca, innovazione, assistenza

Agnostic FDA & EMA drug approval based on basket trials

- **Pembrolizumab for MSI-H or mismatch-repair deficient tumors**

prevalence of MSI-H:

15% in CRC, 1.9% in Pancreatic Cancer

nb of tumors evaluated:

90 CRC
5 each endometrium and gastric
3 each biliary tract, pancreatic, small intestine, breast
1 each prostate, esophageal, small cell lung, retroperitoneal adenoca

- **Larotrectinib & Entrectinib for TRK-fusion positive cancers**

prevalence of TRK-fusion mutations:

90% in infantile fibrosarcoma, < 1% in CRC and lung cancer

nb of tumors evaluated (larotrectinib + entrectinib):

24 STS
19 Salivary Gland tumors
14 Lung cancers
10 Thyroid cancers
8 CRC
7 each BC, infantile fibrosarcoma
4 each Pancreas, Melanoma
3 each Neuroendocrine, GIST, Cholangiocarcinoma
1 each endometrtial, ovary, appendix

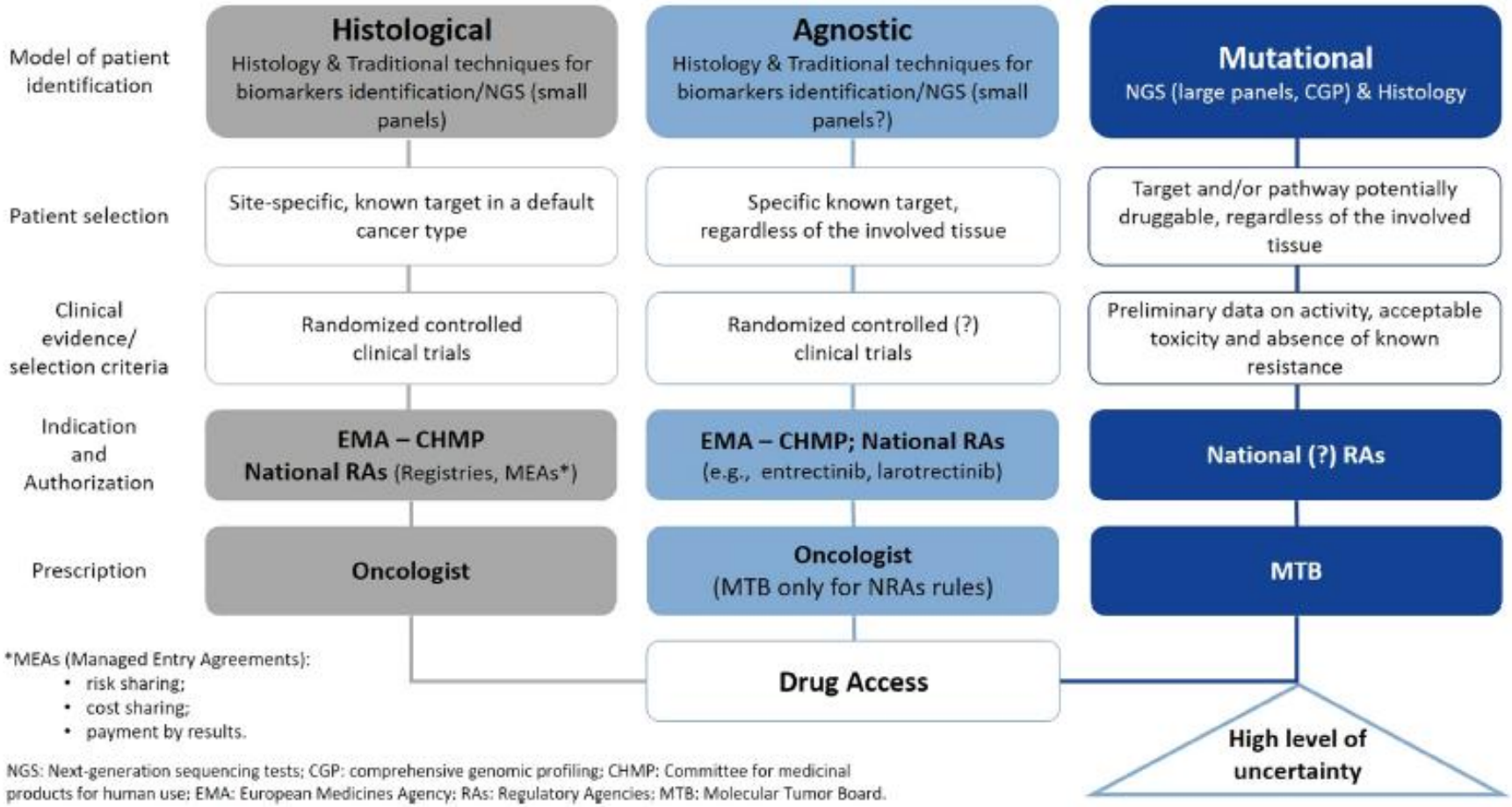
Rete Oncologica Veneta
Ricerca, innovazione, assistenza

Same target, same drug, similar statistical impact, different clinical benefit …….what if PARPi would have been approved agnostically?

PARPi in BRCAm Tumors						
	setting	drug	PARPi PFS mo.	Control PFS mo.	Δ	HR
Ovary	1st L maintenance	olaparib	> 49 mo.	13.8 mo.	**> 30 mo.**	0.30 (0.23-0.41)
Breast	Advanced	olaparib	7.0	4.8	**2.2 mo.**	0.58 (0.43-0.80)
Breast	Advanced	talazoparib	8.6	5.6	**2.0 mo.**	0.54 (0.41-0.71)
Pancreas	1st L maintenance	olaparib	7.4	3.8	**3.6 mo.**	0.53 (0.35-0.82)
Prostate	Advanced	olaparib	7.4	3.6	**3.8 mo.**	0.34 (0.25-0.47)

Rete Oncologica Veneta
Ricerca, innovazione, assistenza

Three different models in Precision Oncology



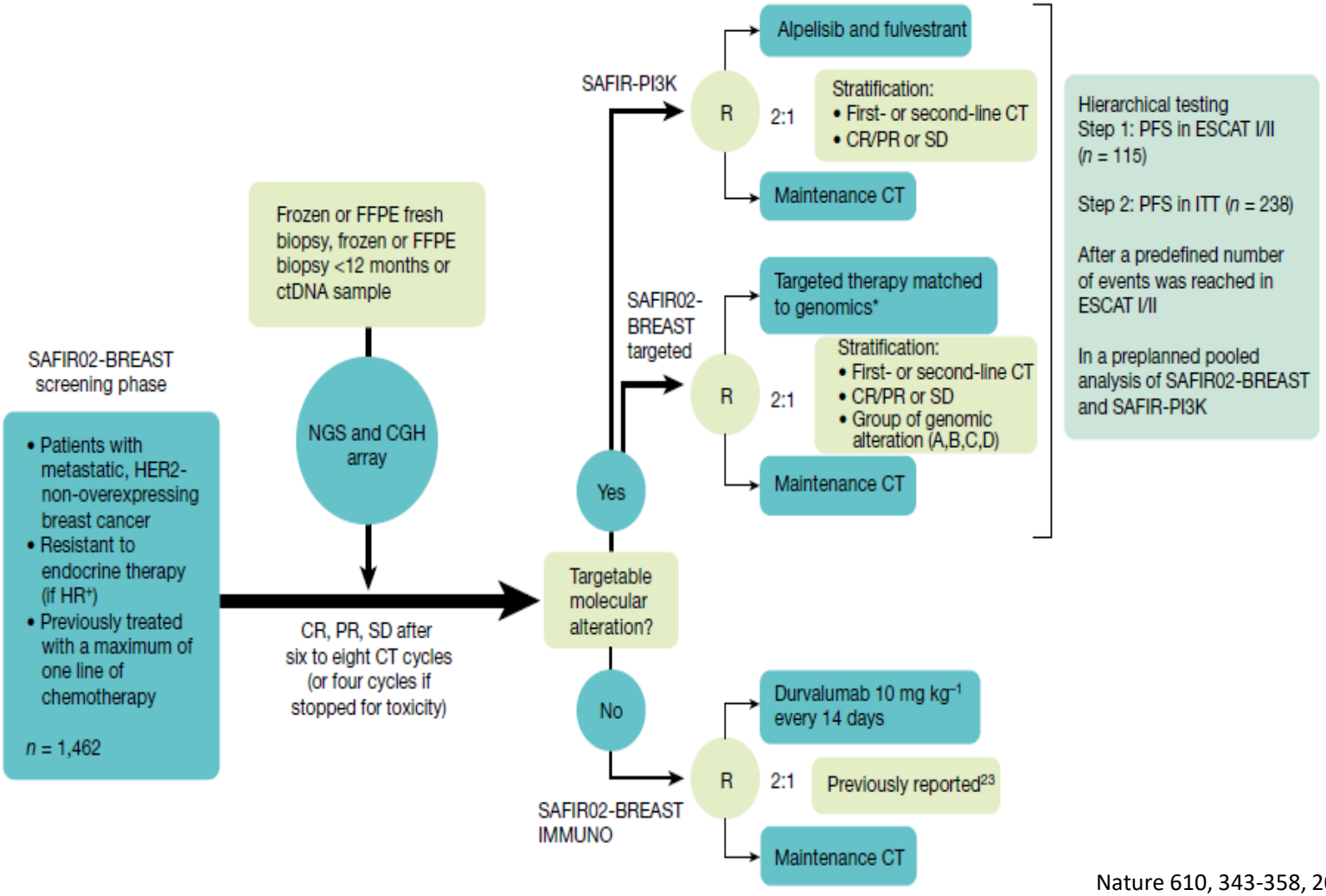
P Marchetti et al, EJC 2023

Article

Genomics to select treatment for patients with metastatic breast cancer

https://doi.org/10.1038/s41586-022-05068-3
Received: 24 October 2021
Accepted: 3 July 2022
Published online: 7 September 2022
Check for updates

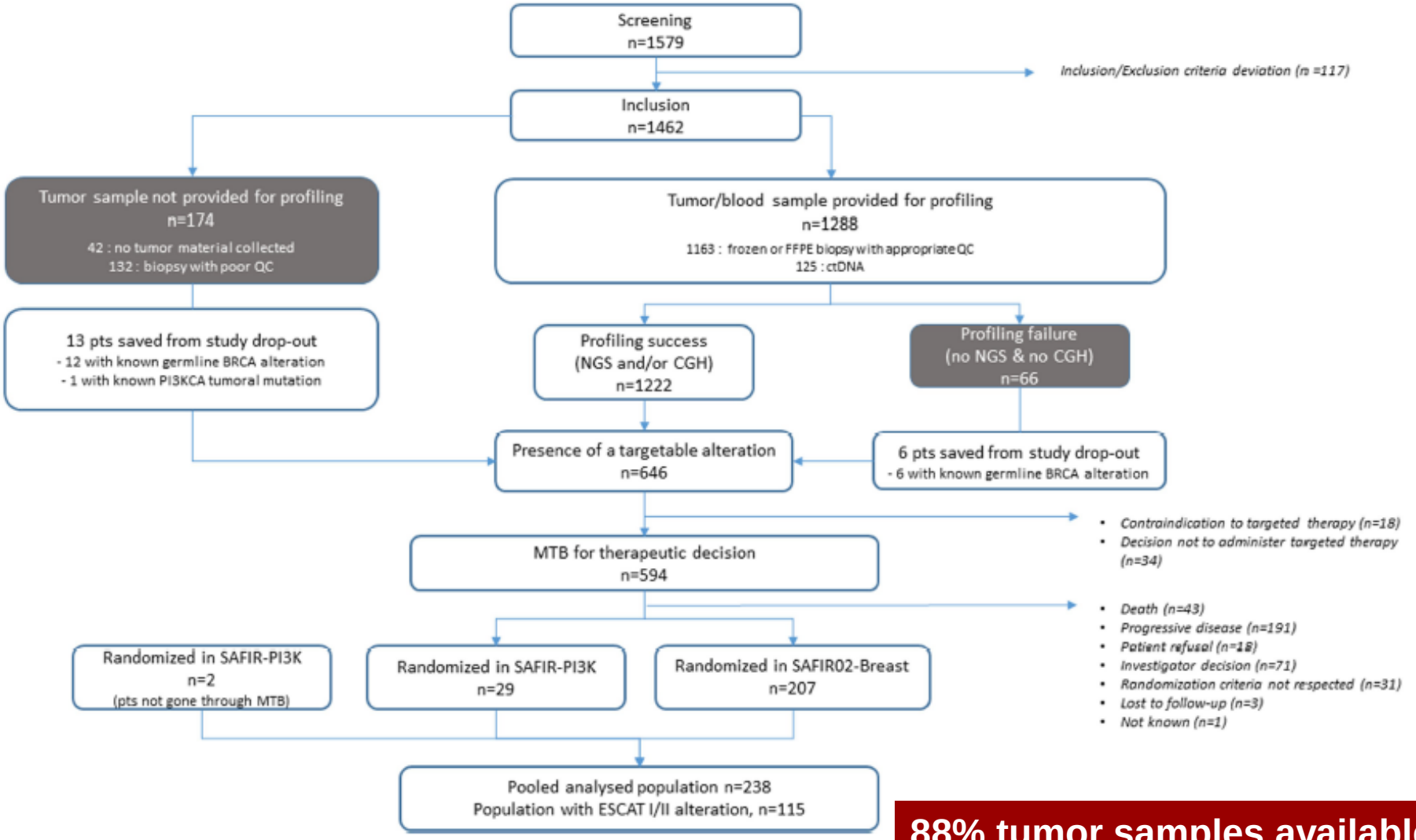
Fabrice Andre[1,2,3,4,5✉], Thomas Filleron[6,3✉], Maud Kamal[6,3✉], Fernanda Mosele[7], Monica Arnedos[1], Florence Dalenc[7], Marie-Paule Sablin[6,8], Mario Campone[9], Hervé Bonnefoi[10], Claudia Lefeuvre-Plesse[11], William Jacot[12], Florence Coussy[13], Jean-Marc Ferrero[14], George Emile[15], Marie-Ange Mouret-Reynier[16], Jean-Christophe Thery[17], Nicolas Isambert[18], Alice Mege[19], Philippe Barthelemy[20], Benoit You[21], Nawale Hajjaji[22], Ludovic Lacroix[23], Etienne Rouleau[23], Alicia Tran-Dien[2,3,4✉], Sandrine Boyault[24], Valery Attignon[24], Pierre Gestraud[25], Nicolas Servant[26], Christophe Le Tourneau[5], Linda Larbi Cherif[5], Isabelle Soubeyran[27], Filippo Montemurro[28], Alain Morel[29], Amelie Lusque[6], Marta Jimenez[30], Alexandra Jacquet[30], Anthony Gonçalves[31,2✉], Thomas Bachelot[32,3✉] & Ivan Bieche[33,2✉]



*olaparib, capivasertib, vistusertib, AZD8931, vandetanib, bicalutamide, AZD4547, selumetinib

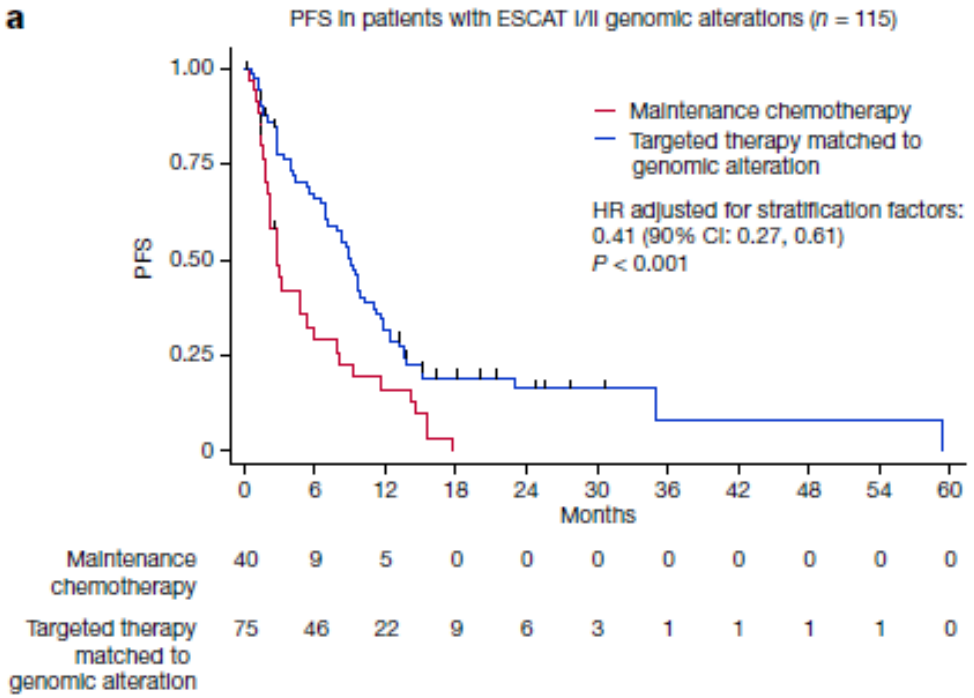
Nature 610, 343-358, 2022

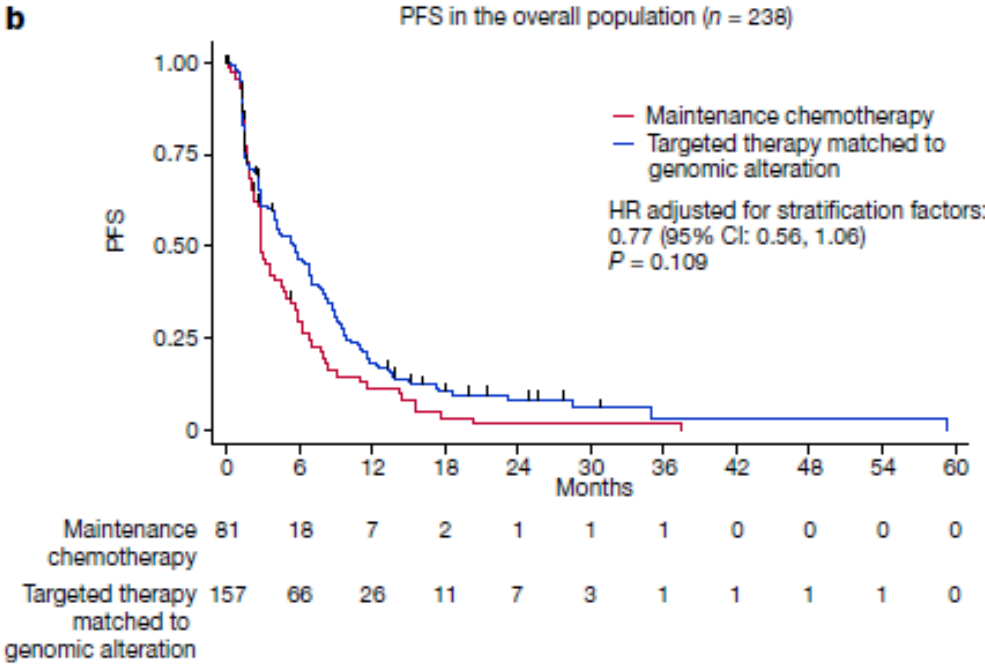
CONSORT Diagram



88% tumor samples available
83.5% successful profiling
44.2% targetable alterations
7.9 % with ESCAT I/II

Andre F et al, Nature 610, 343-358, 2022





mPFS 9.1 vs 2.8 months

ESCAT I/II:
57 BRCAm
3 PALB2m
31 PI3Km
16 AKTm
5 PTENm/deletion
3 ERBB2m

Beyond ESCAT I/II:
no advantage for therapies matched to genomic alterations

Andre F et al, Nature 610, 343-358, 2022

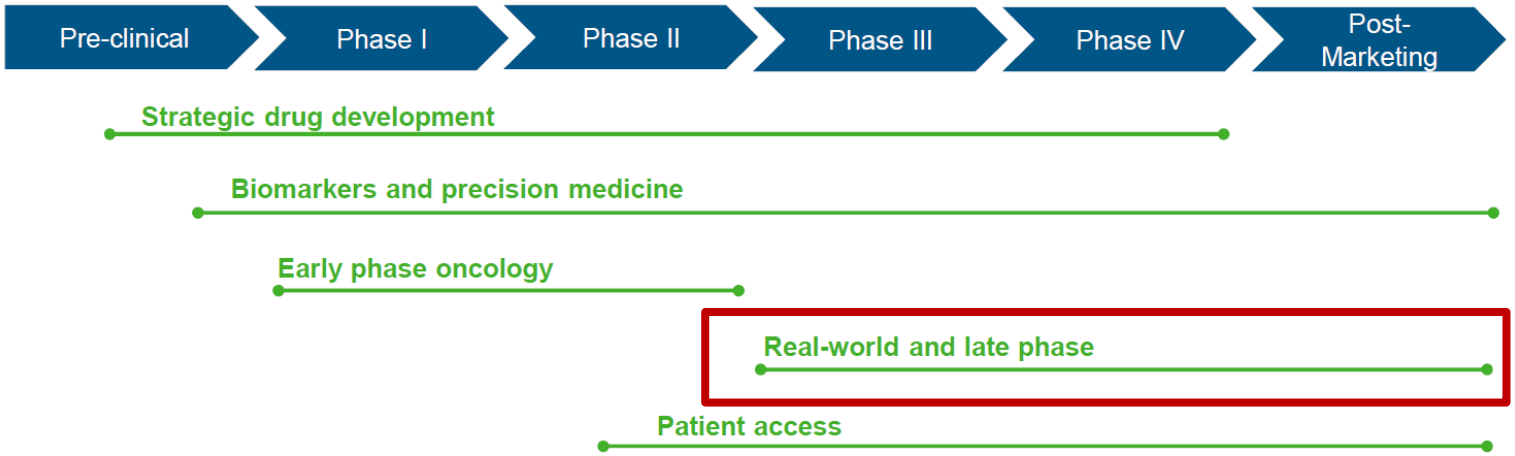
Rete Oncologica Veneta
Ricerca, innovazione, assistenza

Innovative drugs and clinical research revolution

Genomic-driven trials are focused on rare tumors or subgroups with highly unmet needs and can lead to a rapid agnostic approval.

However:

- data aquisition and interpretation can be an issue
- analitycal and biological relaibility can be an issue
- centralised labs and companion diagnostics are key
- multidisciplinarity and multiprofessionality are mandatory
- external validity is necessary.



NGS in Italia: un'opportunità, molti quesiti da risolvere

Ripartizione Fondi NGS		
	Italia	Veneto
Adenocarcinoma polmonare	4348/9876	343/780
Colangiocarcinoma	174/1695	14/142

Problemi:

Criteri di selezione (dati iniziali progetto O.CO.CA indicano che il 24% dei pazienti con tumore polmonare non ha diagnosi istologica)

Reflex test: NGS deve venire richiesto da UO Oncologia; tempo alla diagnosi istologica + richiesta dell'oncologo + analisi e interpretazione NGS sono compatibili on storia clinica di questi tumori?

Rete Oncologica Veneta
Ricerca, innovazione, assistenza

NGS in Italia: opportunità e sfide

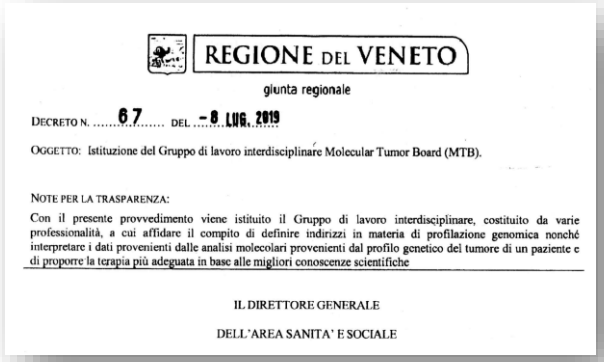
Ripartizione Fondi NGS		
	Italia	Veneto
Adenocarcinoma polmonare	**4348/9876**	**343/780**
Colangiocarcinoma	**174/1695**	**14/142**

Problemi:

Criteri di selezione (dati iniziali progetto O.CO.CA indicano che il 24% dei pazienti con tumore polmonare non ha diagnosi istologica)
Reflex test: NGS deve venire richiesto da UO Oncologia; tempo alla diagnosi istologica + richiesta dell'oncologo + analisi e interpretazione NGS sono compatibili con la storia clinica di questi tumori?

Rete Oncologica Veneta
Ricerca, innovazione, assistenza

Criteri selezione pazienti
Test da eseguire
Registro per il monitoraggio
Individuazione laboratori
PDTA dedicato
Definizione delle tariffe
Analisi e valutazione casi sottoposti



REGIONE DEL VENETO

giunta regionale

DECRETO N. 67 DEL - 8 LUG. 2019

OGGETTO: Istituzione del Gruppo di lavoro interdisciplinare Molecular Tumor Board (MTB).

NOTE PER LA TRASPARENZA:

Con il presente provvedimento viene istituito il Gruppo di lavoro interdisciplinare, costituito da varie professionalità, a cui affidare il compito di definire indirizzi in materia di profilazione genomica nonché interpretare i dati provenienti dalle analisi molecolari provenienti dal profilo genetico del tumore di un paziente e di proporre la terapia più adeguata in base alle migliori conoscenze scientifiche

IL DIRETTORE GENERALE

DELL'AREA SANITA' E SOCIALE

REGIONE DEL VENETO

ALLEGATO A **Proposta n. 1346 / 2021**

c) Unico Molecular Tumor Board (MTB) multidisciplinare

Il coordinamento del MTB è affidato al Coordinatore della Rete Oncologica del Veneto

Inoltre il MTB deve essere dotato di una segreteria scientifica composta da un clinico, un patologo ed un case manager dedicato con specifiche competenze in oncologia.

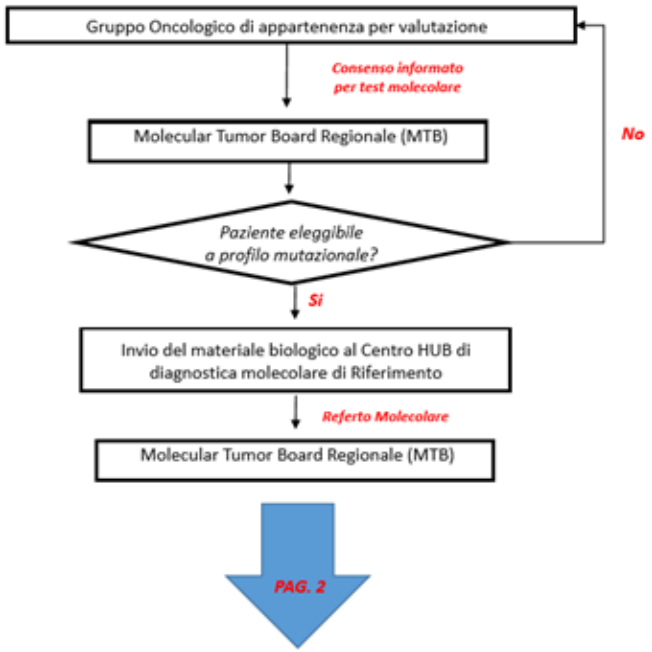
Le figure professionali "fisse "che devono essere rappresentate nel MTB regionale sono:

- ✓ oncologo
- ✓ anatomopatologo
- ✓ bioinformatico
- ✓ biostatistico
- ✓ genetista
- ✓ farmacista ospedaliero
- ✓ patologo molecolare
- ✓ farmacologo
- ✓ ematologo
- ✓ bioeticista

Rete Oncologica Veneta
Ricerca, innovazione, assistenza

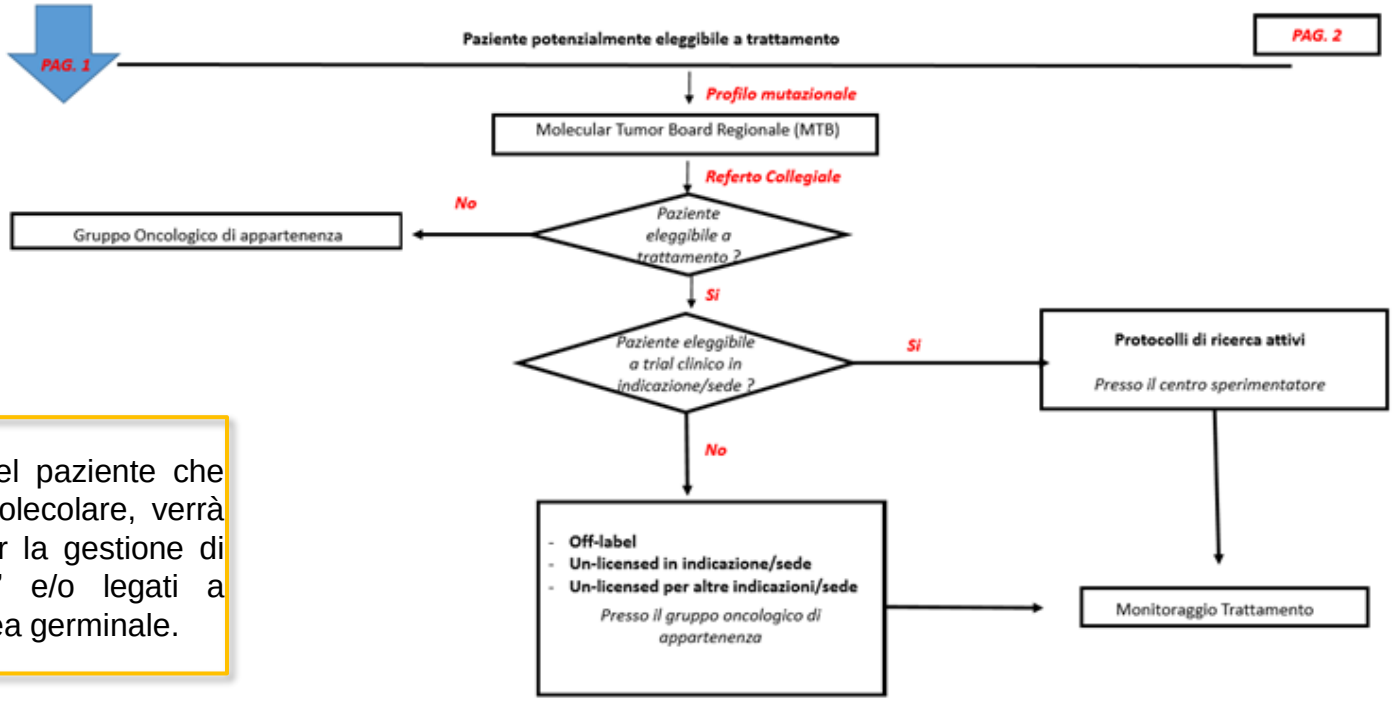
Paziente potenzialmente eleggibile a profilo mutazionale

PAG. 1



*Tutto il Percorso richiesta di valutazione, referto molecolare,referto collegiale e utilizzo farmaco e follow-up deve essere documentato all'interno della piattaforma dedicata

Rete Oncologica Veneta
Ricerca, innovazione, assistenza



Oltre al PDTA standard del paziente che accede alla profilazione molecolare, verrà identificata una politica per la gestione di possibili risultati “inattesi” e/o legati a possibili mutazioni della linea germinale.

*Tutto il Percorso richiesta di valutazione, referto molecolare,referto collegiale e utilizzo farmaco e follow-up deve essere documentato all’interno della piattaforma dedicata

Rete Oncologica Veneta

Ricerca, innovazione, assistenza

LA RETE A PROTEZIONE DEL PAZIENTE ONCOLOGICO